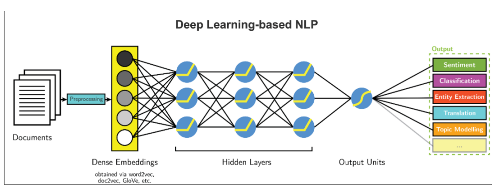
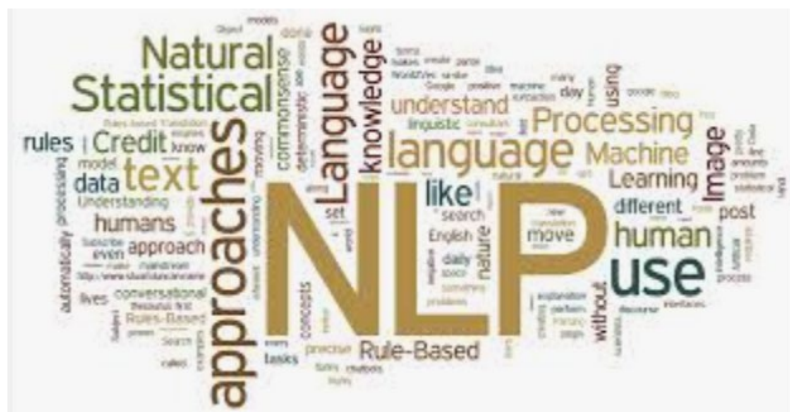
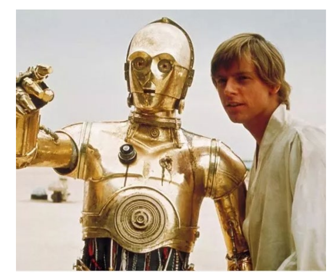
Natural Language Processing (NLP) in Medicine

Li Zhou, MD, PhD, FACMI, FIAHSI, FAMIA

Professor of Medicine
Division of General Internal Medicine and Primary Care
Brigham and Women's Hospital, Harvard Medical School

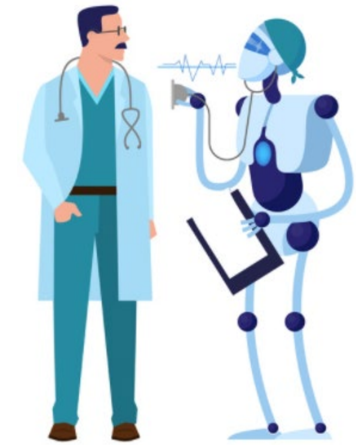






Email: lzhou@bwh.harvard.edu
Website: http://mterms.bwh.harvard.edu/

What technology can help me?



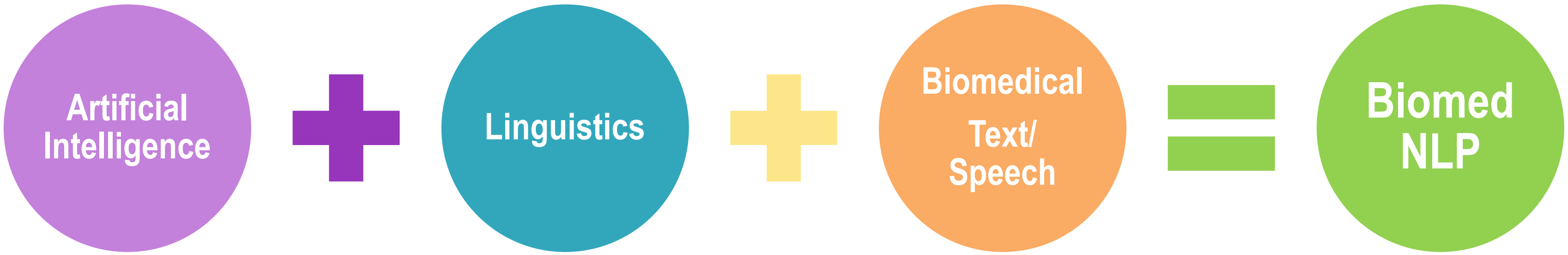
- Identify the long-term symptoms resulting from COVID-19?
- Generate a cohort of patients who had severe cutaneous adverse reactions caused by vancomycin?
- Find “need to know” clinical information from Epic EHR relevant to the patient’s chief complaints?
- Read/interpret a pathology report to find abnormal cancer screening results and tell me when to follow up?
- Transcribe and summarize my conversation with the patient?
- Detect and correct errors in my notes dictated by Dragon?
- ……

Natural Language Processing (NLP)

Today's Outline

- **Definitions** - What is natural language processing (NLP)?
- **Significance** – Why NLP is important in Biomedicine?
- **Challenges** – What are the challenges in biomedical NLP?
- **Methods** - How to conduct NLP?
- **Systems and Applications** – MTERMS NLP as an example
- **Future Directions** – The future use of NLP in biomedicine

Biomedical NLP



- NLP aims to creates intelligent agents to interpret, understand, analyze and generate human language.
- A biomedical NLP system refers to any system that processes and/or generate biomedical **text** or **speech.**

NLP Tasks

- Information extraction and encoding
- Information retrieval

- Named entity recognition
- Text/document classification
- Grammar and spelling checking and correction
- Relationship extraction
- Sentiment analysis
- Predictive risk analytics

- Speech recognition
- Speech synthesis

- Question-answering
- Language translation

- Text generation
- Text summarization

Natural Language Understanding	Natural Language Generation

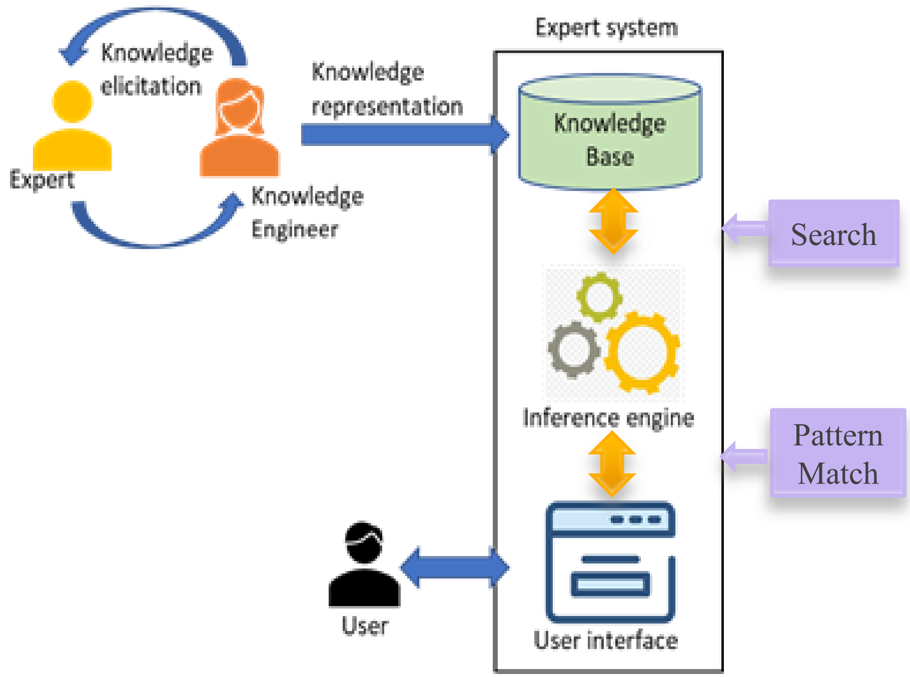
Eliza

An early example of NLP system that simulates a Rogerizn Psychotherapist

A Conversation with Eliza

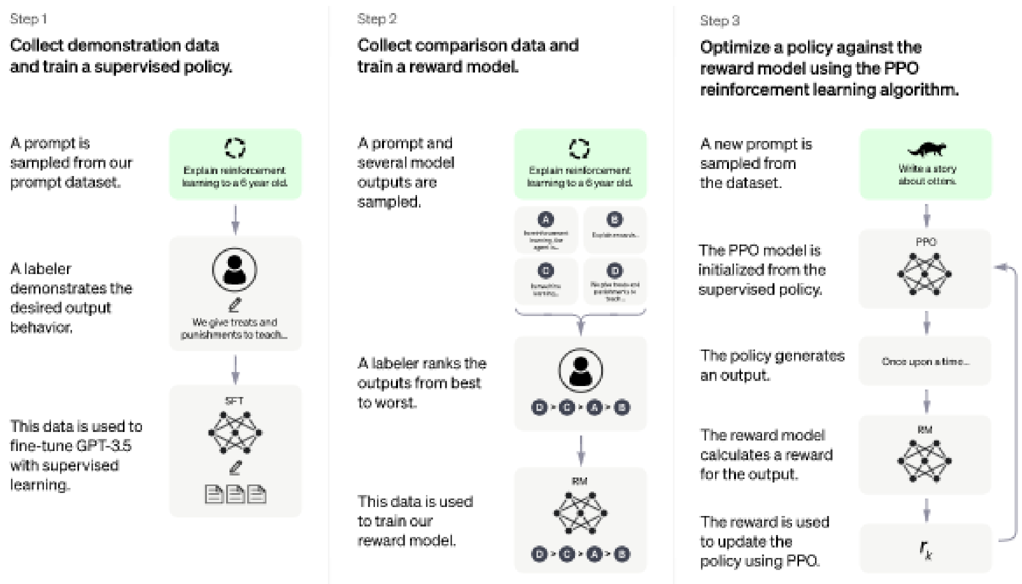
Eliza: Hello. My name is Eliza. How may I help you?
Li: I do not feel well today.
Eliza: Ahh. Please continue…
Li: I suffer from allergies.
Eliza: Say, do you have any psychological problems?
Li: I have bad allergies, so I feel sad…
Eliza: Do you often feel sad?
Li: No.
Eliza: Why not?
Li: Why should I?
Eliza: What do you think?
Li: I want to hear your suggestions?
Eliza: What if you never got to hear my suggestions?
Li: To see a real doctor?
Eliza: I am not sure I understand your fully.
Li: You're just a computer program, aren't you?
Eliza: What makes you think I am just a computer program, aren't me?
Li: You don't fully understand my problem.
Eliza: Oh… don't fully understand your problem?

Knowledge-based NLP System



One of the first chatbots developed in the mid-1960s by Joseph Weizenbaum.

ChatGPT



- An AI chatbot developed by OpenAI and launched in Nov 2022.
- Built on top of OpenAI’s GPT-3.5 and GPT-4 (generative pre-trained transformer) families of large language models (LLMs).
- Fine-tuned using both supervised and reinforcement learning techniques in a process called reinforcement learning from human feedback (RLHF).
  - An initial model was trained using supervised fine-tunning. Human AI trainers provided conversations in which they played both sides—the user and an AI assistant.
  - In the reinforcement learning step, human trainers first ranked responses that the model had created in a previous conversation. These rankings were used to create 'reward models' that the model was further fine-tuned on using several iterations of Proximal Policy Optimization (PPO).
  - OpenAI also continues to gather data from ChatGPT users (upvote/downvote) that could be used to further train and fine-tune ChatGPT.
- The models were trained in collaboration with Microsoft on their Azure supercomputing infrastructure,, using Nvidia GPUs.

ChatGPT

LZ I do not feel well today.

I'm sorry to hear that you're not feeling well today. If you are comfortable sharing more about what's going on, I may be able to offer some suggestions for how to feel better.

In general, if you're not feeling well, it's important to take care of yourself. This can include getting enough rest, eating a healthy diet, staying hydrated, and ......

If you're feeling physically unwell, it's a good idea to rest and take care of yourself. If your symptoms persist or worsen, or if you're experiencing severe or concerning symptoms, it's important to seek medical attention......

If you're feeling emotionally unwell, it may be helpful to talk to someone about how you're feeling. This could be a friend, family member, or mental health professional. You can also consider engaging in self-care activities that help you feel better......

Remember, it's okay to not feel well sometimes. It's important to take care of yourself and seek help if you need it.

LZ I suffer from allergies

I'm sorry to hear that you suffer from allergies. Allergies can be a frustrating and uncomfortable condition to deal with, but there are steps you can take to manage your symptoms.

The first step in managing allergies is to identify what triggers your symptoms. Common allergens include pollen, dust mites, pet dander, mold, and certain foods. Once you have identified your triggers, you can take steps to avoid them or minimize your exposure.

Some strategies for managing allergies include:

1. Avoiding allergens: .....
2. Taking medications: ....
3. Using allergy shots: ......

......

It's important to work with a healthcare provider to develop a personalized treatment plan for your allergies. They can help you identify your triggers and develop a plan to manage your symptoms.

LZ I have bad allergies, so I feel sad...

I'm sorry to hear that your allergies are making you feel sad. Allergies can be frustrating and uncomfortable to deal with, and it's understandable that they can affect your mood. **In addition to the strategies I mentioned earlier** for managing allergies, there are also some things you can do to improve your mood and reduce your stress levels. Here are a few ideas:

1. Practice relaxation techniques: Deep breathing, progressive muscle relaxation, ...
2. Engage in physical activity: Exercise can be a great way to boost your mood and reduce stress......
3. Connect with others: Spending time with loved ones or participating in social activities......
4. Get enough sleep: ......
5. Consider talking to a mental health professional: ....

Remember, allergies can be challenging to deal with, but there are things you can do to manage your symptoms and improve your mood. Don't hesitate to seek help if you need it.

LZ You are just a computer program, not a real doctor.

That's correct, I am not a real doctor. I'm an artificial intelligence language model designed to provide information and answer questions to the best of my ability based on the data I have been trained on. While I can offer helpful suggestions and information, I am not a substitute for medical advice or treatment from a qualified healthcare provider. If you have any concerns about your health, it's important to seek advice from a doctor or other healthcare professional who can provide a diagnosis and recommend appropriate treatment.

LZ You don’t fully understand my problem.

I'm sorry to hear that I may not fully understand your problem. I'm here to help in any way I can, and I'll do my best to provide you with accurate and helpful information based on the details you provide. If you're comfortable sharing more about your situation, I may be able to offer more specific suggestions for how to address your concerns. Alternatively, if you feel that your situation requires more specialized attention or care than I am able to provide, I encourage you to seek the advice of a healthcare professional or other qualified expert.

- It is amazing!
- It remembers previous prompts given to it in the same conversation
- While it can be verbose
- AI hallucination - ChatGPT "sometimes writes plausible-sounding but incorrect or nonsensical answers"
- Moderation API is used to warn or block certain types of unsafe content.

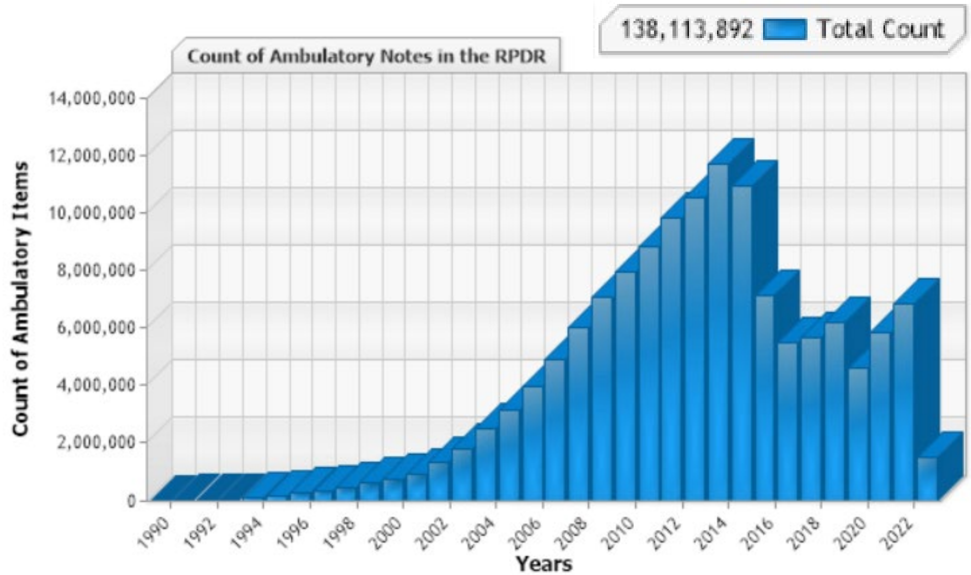
Today's Outline

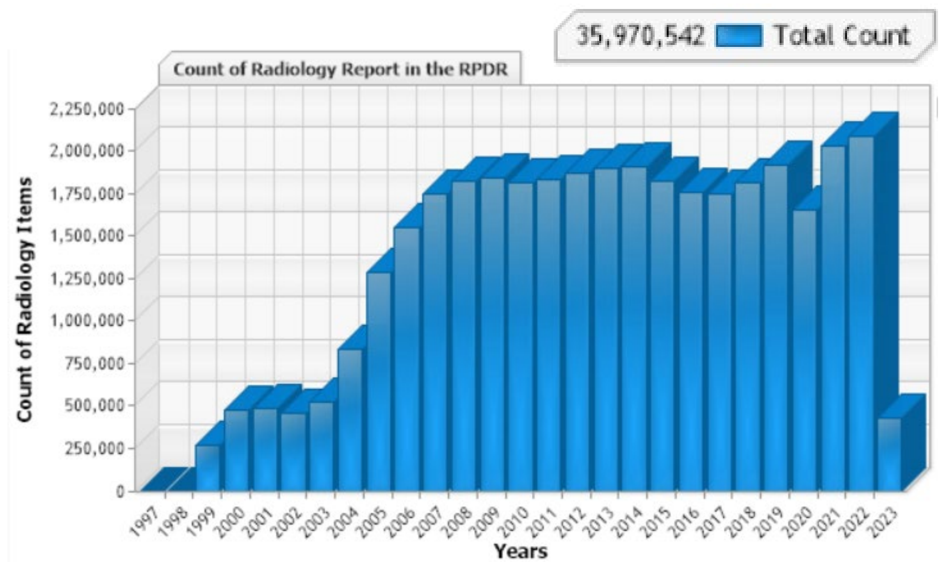
- **Definitions** - What is natural language processing (NLP)?
- **Significance – Why NLP is important in Biomedicine?**
- **Challenges** - What are the challenges in biomedical NLP?
- **Methods** - How to conduct NLP?
- **Systems and Applications** – MTERMS NLP as an example
- **Future Directions** – The future use of NLP in biomedicine

Free-text Clinical Data in EHR

- A significant portion of biomedical information is stored in textual form.

- **Electronic Health Records**
  - Ambulatory notes
  - Admission notes
  - Progress notes
  - Discharge summaries
  - Radiology reports
  - Pathology reports
  - Free-text entries and comments
  - ……





Radiology reports include CT, ECG, MRI, x-Ray, ultrasound, mammography, etc

The Cloud, the Crowd, and Big Data

- Biomedical Literature
- Books, guidelines, surveys, emails
- Wikipedia
- Social media (Twitter, Facebook, blogs)
- News, reports
- Open innovation contests

- Blumenthal K, Topaz M, Zhou L, et al. Mining Social Media Data to Assess the Risk of and Soft Tissue Infections from Allergen Immunotherapy. J Allergy Clin Immunol. 2019
- Hua Y, Jiang H, Lin S, Yang J, Plasek JM, Bates DW, Zhou L. Using Twitter Data to Understand Public Perceptions of Approved versus Off-label Use for COVID-19-related Medications. J Am Med Inform Assoc. 2022. PMID: 35775946;.

Tang C, Zhou L, et al;. Comment Topic Evolution on a Cancer Institution's Facebook Page. Applied clinical informatics 2017.

Speech Recognition

- Dictation

- Voice-enabled care (Virtual medical assistant)

- Speech and Diseases

Zhou L, et al. Analysis of Errors in Dictated Clinical Documents Assisted by Speech Recognition Software and Professional Transcriptionists. JAMA Network Open. 2018

Blackley SV, et al. Speech Recognition for Clinical Documentation from 1990 to 2018: A Systematic Review. JAMIA 2019

Goss FR, et al. A Clinician Survey of Using Speech Recognition for Clinical Documentation in the Electronic Health Record. Int J Med Inform (IJMI). 2019.

Blackley SV, et al. Physician Use of Speech Recognition versus Typing in Clinical Documentation: A Controlled Observational Study. IJMI, 2020.

Today's Outline

- **Definitions** - What is natural language processing (NLP)?
- **Significance** – Why NLP is important in Biomedicine?
- **Challenges - What are the challenges in biomedical NLP?**
- **Methods** - How to conduct NLP?
- **Systems and Applications** – MTERMS NLP as an example
- **Future Directions** – The future use of NLP in biomedicine

NLP Challenges

Challenging Areas		Examples
Synonym	A concept may be expressed in many different ways	*Congestive heart failure, heart failure, CHF*
Polysemy/ Homonym	One word may have multiple meanings	*Discharge from hospital* vs *discharge from wound*
Ambiguity	Interpretation of an expression can be uncertain	*010202* can be a number, a date, a symbol, etc.
Medical Grammar /Format	It often ignores many restrictions that are required in the general English grammar	More compact and largely expressed by nouns or noun phrases: *c/o: fever; Labs: K+ 4.1*
Abbreviation/ Acronym	A shortened form of a word or phrase	*CAP*: community-acquired pneumonia, catabolite activator protein, capsule
Pragmatics /Context	Context dependence	*Diabetes* mentioned in Family History section
Others	Ellipsis, anaphora, coreference, conjunctions, heterogeneous formats, spelling errors, granularity, uncertainty, implicit info, etc.	

Deciphering Abbreviations / Acronyms
(An Example)

“94 yo LOL in NAD with dat + hx chf aodm2 pvd s/p almi w/ ef 20 s/p tah bso appy BII w/ b12 def now at * ecf comes to ED w/ temp”**

“94 year-old elderly woman ("Little Old Lady") in no apparent distress with dementia, Alzheimer's Type + history of congestive heart failure, adult onset diabetes mellitus (Type 2), peripheral vascular disease, status post anterolateral myocardial infarction with ejection fraction 20%, status post total abdominal hysterectomy and bilateral salpingo-oophorectomy, appendectomy, Bilroth II [gastric bypass procedure] with Vitamin B12 deficiency, now at *** extended care facility comes to emergency department with temperature.”

Adopted from AMIA CIS working group email discussion; Omar Bouhaddou, Clayton Curtis, et al. Jan 13, 2012

Semi-structured vs Narrative

Identifying Allergies

Allergies:
Cephalosporins – Hives
Pencillin– Rash
LISINOPRIL – Anaphylaxis 5/12/17

History of present illness:
She developed a rash that was probably due to metoprolol. The rash resolved when she switched to atenolol.

CDA_Section	LOINC_Code	Section_Name
Allergies Section	48765-2	Adverse drug reactions/drug allergies
Allergies Section	48765-2	ALL
Allergies Section	48765-2	Allergies
Allergies Section	48765-2	Allergies and adverse reactions
Allergies Section	48765-2	Allergies or drug sensitivities
Allergies Section	48765-2	Allergies to Medications
Allergies Section	48765-2	Allergies Updated
Allergies Section	48765-2	Allergies, Adverse Reactions
Allergies Section	48765-2	Allergies/drug sensitivity
Allergies Section	48765-2	Allergies/Hay Fever
Allergies Section	48765-2	ALLERGIES/REACTIONS
Allergies Section	48765-2	allergy
Allergies Section	48765-2	Allergy History
Allergies Section	48765-2	Detergent allergies
Allergies Section	48765-2	Drug Allergies
Allergies Section	48765-2	Environmental allergies
Allergies Section	48765-2	Food Allergies
Allergies Section	48765-2	Food Allergies/Intolerances
Allergies Section	48765-2	List Allergies
Allergies Section	48765-2	Medication allergies
Allergies Section	48765-2	Medicine Allergies
Allergies Section	48765-2	Meds/Allergies
Allergies Section	48765-2	New Allergies
Allergies Section	48765-2	Ocular allergies
Allergies Section	48765-2	Patient has had history of allergies
Allergies Section	48765-2	PHM/FH/Meds/Allergies/ROS
Allergies Section	48765-2	Reactions to dental/other local anesthesia
Allergies Section	48765-2	Scents, oils, lotion allergies
Allergies Section	48765-2	Seasonal Allergies
Allergies Section	48765-2	Signs of allergies

Challenges – Domain Level

- **Sublanguage** – a subset of language used in a particular field or by a particular social group and characterized esp. by distinctive vocabulary (Zellig Harris)
  - Informational categories (semantic types):
    - **body location** (*chest, leg, heart*)
    - **symptom** (*pain, fever, cough*)
    - **severity** (*severe, mild*)
  - Co-occurrence patterns: **severity** + **body location** + **symptom** (*severe chest pain*)
  - Contextual information
- Methods and tools trained on general English may receive decreased performance on biomedical texts

Challenges – Application Level

- Availability of clinical text
- Privacy and confidentiality
- Determining types of information to capture
- Rare events/ imbalanced data
- Intelligence (reasoning/ inference)
- Incorporating medical knowledge into NLP systems
- Interpreting clinical information
- Integrating narrative data with structured data
- Generalizability, intra- and inter-operability
- Good performance
- Evaluation methods and measures

Challenges – Clinical Environments

- Organizational factors
  - Regulations, rules and culture
  - Disciplines, departments and groups
- Users
  - Diverse roles, backgrounds and needs
- Clinical domains and subspecialties
- Workflows

☞ *Although natural language is easy for human to use and understand, it may be challenging for computers to understand, interpret and generate.*

Today's Outline

- **Definitions** - What is natural language processing (NLP)?
- **Significance** – Why NLP is important in Biomedicine?
- **Challenges** - What are the challenges in biomedical NLP?
- **Methods** - How to conduct NLP?
- **Systems and Applications** – MTERMS NLP as an example
- **Future Directions** – The future use of NLP in biomedicine

NLP Methods

	Techniques	Tasks	Pros and Cons
Linguistic rules	"Symbolic"; use experts' knowledge that is explicitly represented by rules for the manipulation of symbols. A sentence is either well formed or not Often work top-down by imposing known grammatical patterns and meaning associations upon texts	Information extraction	"Good Old-Fashioned AI" More interpretable; limited ability handling complex language structures & nuances.
Statistical	"Empirical"; derive language data from large text corpora Identify common patterns that occur in language use by applying statistics (probability theory) and machine learning methods Work bottom-up from the texts themselves, looking for patterns and associations to model	Text classification, sentiment analysis, named entity recognition	Some patterns may not correspond to purely syntactic or semantic relations Often requires hand-crafted feature engineering
Deep learning	Train deep neural networks (e.g., RNNs, CNNs and transformer models) on large datasets to learn complex patterns and structures	Language modeling, text classification, machine translation, language generation	Automatically learns the relevant features from the data; Requires more data and computational resources to train

NLP Timeline

1950
- Syntactic structures
- Generative grammars by Chomsky (explicitly/often recursive rules)

1960
- ELIZA
- Automatic Language Processing Advisory Committee (ALPAC) report – machine translation

- First AI Winter

1970
- Semantic networks
- Conceptual dependency theory

1980
- Symbolic approaches, expert systems
- Ontologies (knowledge base, concepts and relationships)
- LISP machines

1990
- Statistical models
- RNNs and LSTMs

- NVIDIA GPUs

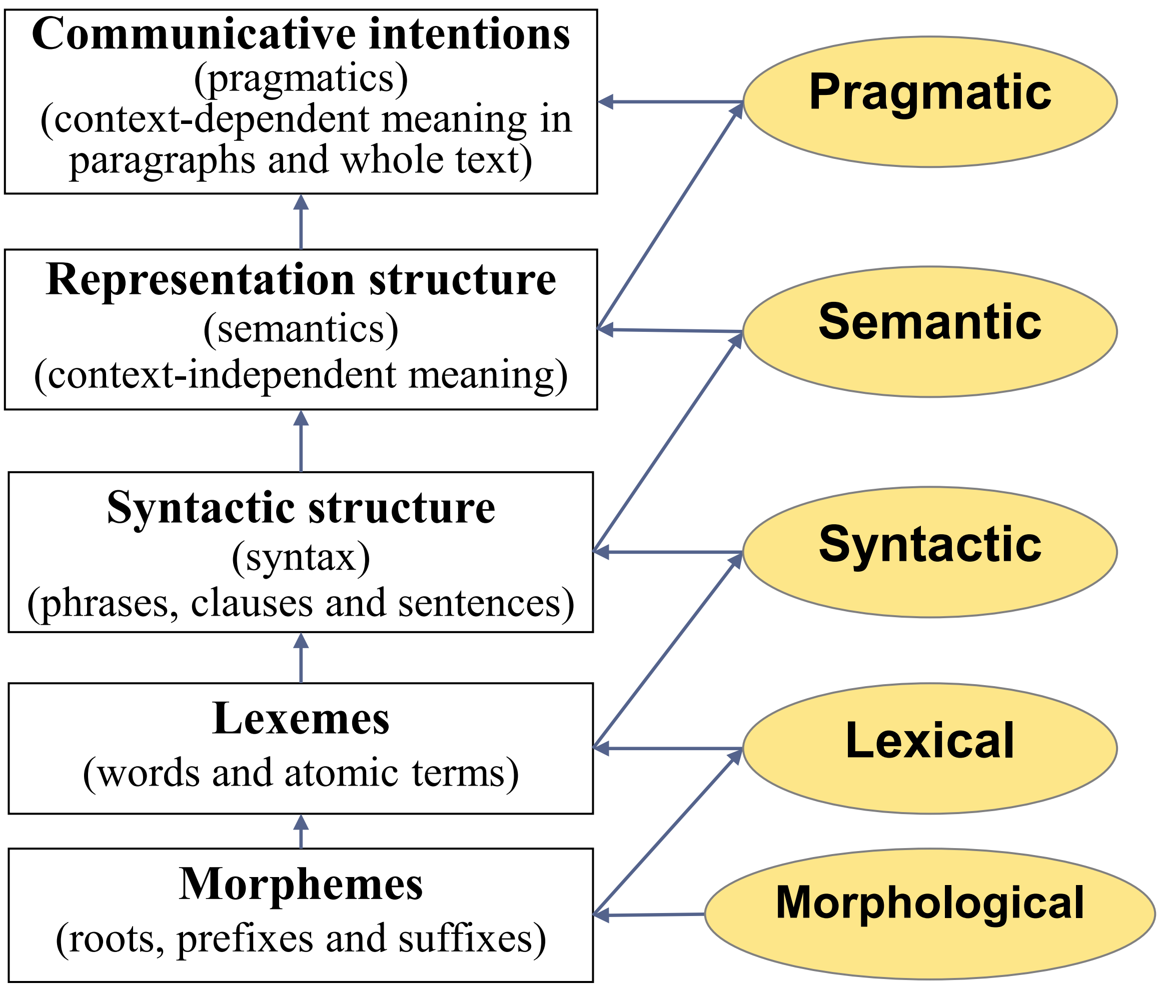
2000
- Neural language modeling
- Word embeddings (representing words with dense vectors of numbers)
- Multi-task learning
- Google translate

2010
- Word2Vec
- Neural networks for NLP, rise of LSTMs and CNNs
- Sequence-to-sequence problems (machine translation)
- Encoder-decoder
- Attention and transformers
- Pre-trained models & transfer learning (Elmo, BERT)
- OpenAI funded

2020
- Large language models
- GPT-3, 3.5, 4
- Google Language model for dialogue application (LaMDA)
- Hugging Face BLOOM (Bigscience large open-science open-access multilingual) language model
- Google BARD chatbot

Linguistic Representation and Analysis

Structure | **Analysis**



Pragmatic: Determine how sentences combine to form discourse and how this context affects the interpretation of the text (e.g., co-reference resolution, discourse analysis, relation analysis)

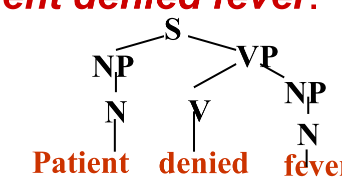
Semantic: Interpret the meaning of the words and how the words combine to form the meaning of phrases and sentences (e.g., encoding, semantic relations)

Syntactic: Combine multiple words and determine the structure of phrases and sentences (e.g., parsing, pattern recognition)

Lexical: Determine the categorization of lexemes (e.g., part-of-speech tagging, named entity recognition)

Morphological: Determine the sequences of morphemes into words (e.g., tokenization)

Linguistic Analysis

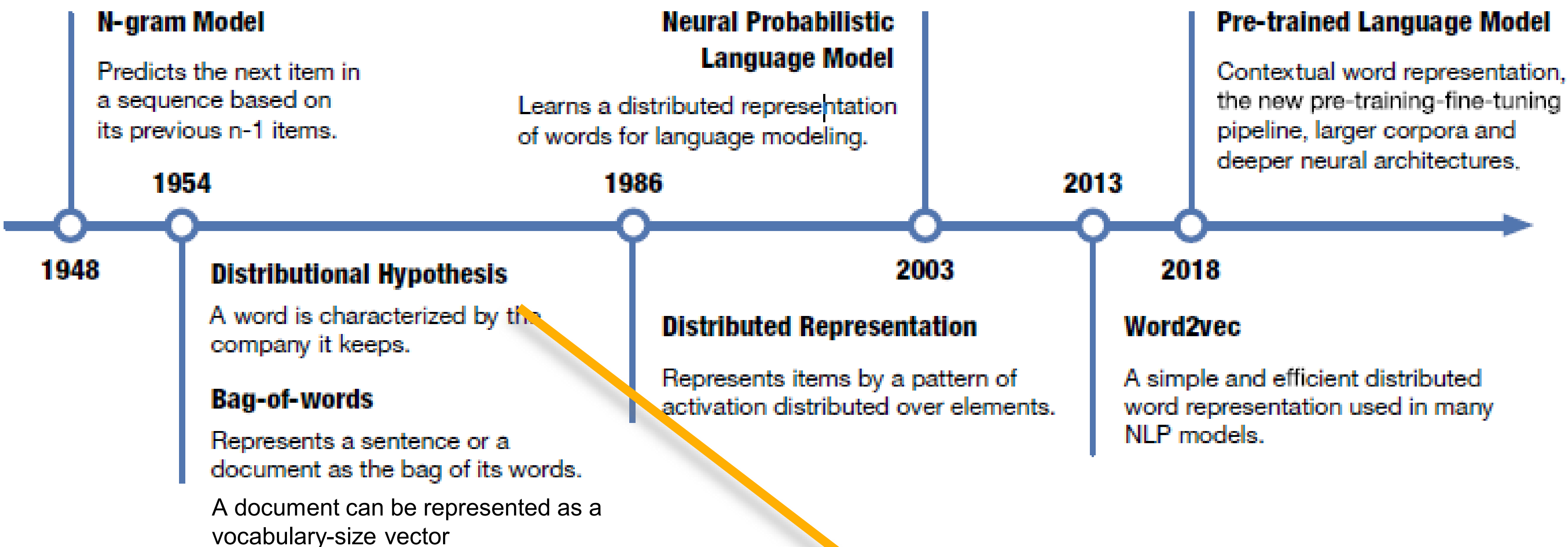
Modules	Methods and Examples	
Pragmatic/ Discourse Analysis (co-reference resolution)	Study context-dependent meaning - machine learning - spatial/temporal reasoning	"*He was given antibiotics at that time*" - the interpretation of "*he*" and "*at that time*" require prior expressions elsewhere in the text
Semantic Analysis (semantic relation)	Identify meaning - statistical/machine learning - Semantic grammar - (cascading) FSA	1. "patient c/o ***excruciating pain in left leg***." <problem v= "***pain***" SNOMED="22253000"> <severity v="***severe***" SNOMED="67849003"/> <location v="***leg***" SNOMED="51185008"/> <laterality v="***left***" SNOMED="7771000"/> </problem> 2."patient ***left*** hospital"
Syntactic Analysis (parser, pattern recognition)	Identify structure (parsing) - regular expression (**JJ*NN→"*acute pain*"**) - context-free grammar - statistical/machine learning approaches	"***Patient denied fever***." 
Lexical Analysis	Identify words/phrases to determine their categories - part of speech tagging: noun (***chest***), adjective (***mild***) - stemming: ***interact*** *(stem)* - ***interacts, interaction*** - named entity recognition: disease (***diabetes***)	
Morphological Analysis (tokenization)	Separate the text into individual units (sentence, word) - regular expression (finite state automata - FSA) **[a-z]** indicates a lowercase letter, **-otomy** indicates a procedure	

Representation Learning / Language Modeling

Local Representation or Symbol-based Representation
Generative models: model the joint probability distribution p (x, c)

→

Distributed Representation
Discriminative models: predict the conditional probability P(x|c) or P(c|x)



John Firth, 1957. Fundamental ideal of many language model. Linguistic items with similar meanings tend to occur in similar context.

Language modeling – predicts the probability of a sequence of words occurring in a given context
Language models can be built using a variety of techniques

Adopted from Z. Liu et al. Representation Learning for Natural Language Processing. 2020

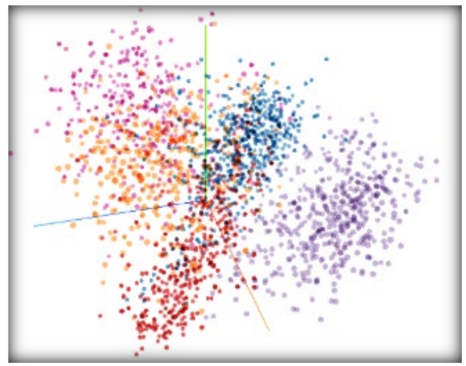
Distributed Representation

Entities & Structure

- Knowledge
- Network
- Document
- Sentence
- Phrase
- Word
- Character

→

Distributed representation



→

Analysis

- Application
- Pragmatic
- Semantic
- Syntactic
- Lexical
- Morphological

- **Challenges**: multiple granularities and multiple tasks
- **Distributed representation**
  - Represent words, phrases, or sentences as vectors in a high-dimensional space, where each dimension captures a different aspect of meaning
  - Represent the semantics of these language entities in a unified semantic space.
    - Geometric distance between two objects in the space indicates their semantic relatedness.
  - Build a unified representation space to support different linguistic analyses.
  - Build representation automatically from large-scale domain data (compared to specific feature extraction for each specific domain).

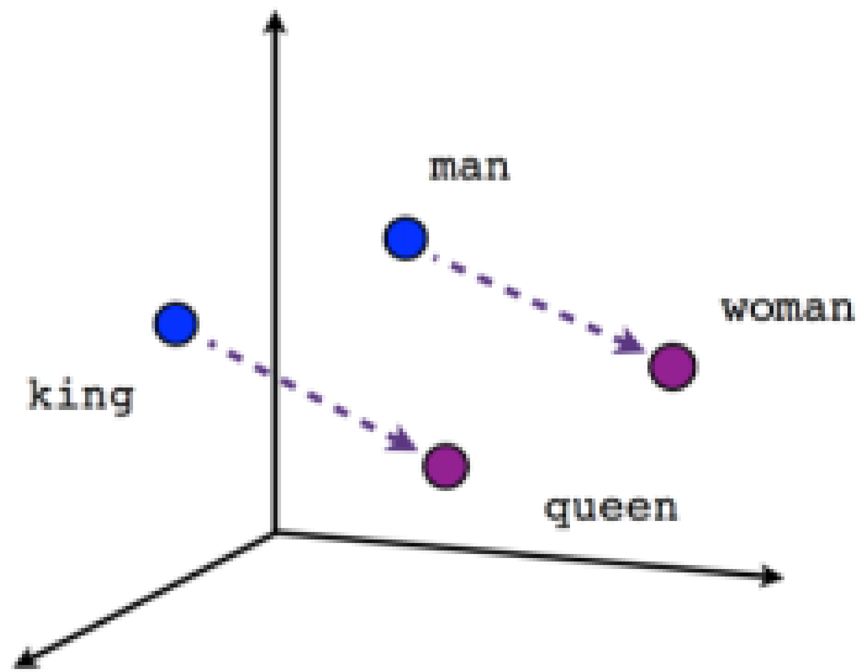
A neural probabilistic language model (Bengio, 2003)

Word Embedding

- “A word is characterized by the company it keeps” (Firth, JR. 1957)
  - Penicillin is antibiotics. Vancomycin is antibiotics. Penicillin and vancomycin tend to have a similar context – antibiotics.
- Quantify and categorize **semantic similarities** between linguistic items based on their distributional properties in large corpora (e.g., using cosine similarity).
- Linguistic items can be represented as real number vectors of co-occurring words and linguistic contexts in which the words occur.
- Methods to generate word embeddings include neural networks, dimensionality reduction on the word co-occurrence matrix, probabilistic models, etc.

A neural probabilistic language model (Bengio, 2003)

Word Embedding



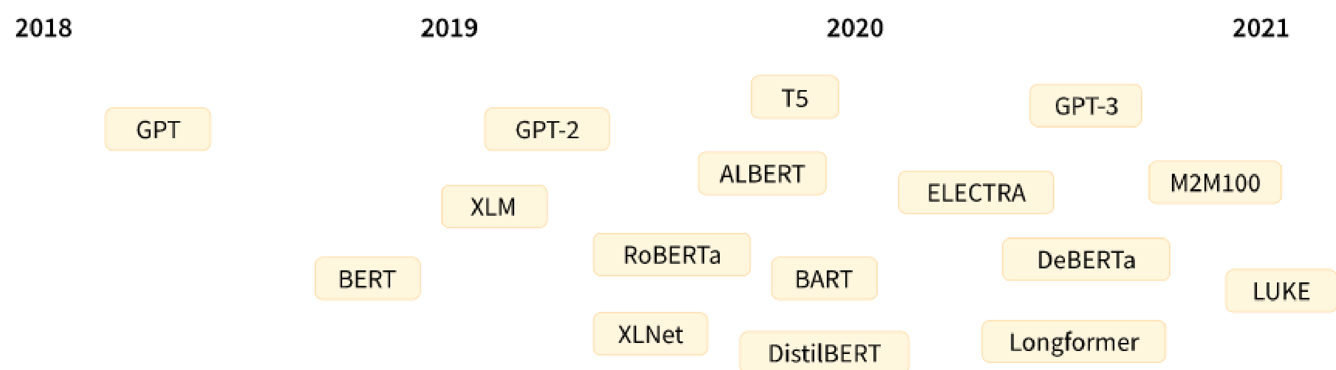
This Figure illustrates the output of a word embedding model where individual words are plotted in 3-dimensional space generated by the model.
By examining the adjacency (context) of words in this space, word embedding models can identify analogies such as “Man is to woman as king is to queen.”

Word Embeddings

- Methods for learning of word embeddings
  - Word2vec (Mikolov, Google; 2013)
  - GloVe (Pennington, et al; 2014)
  - fastText (Bojanowski, et al, 2017)
- Trained using large scale corpora
- Map words into informative low-dimensional vectors
- Word embeddings are used as input representation for other downstream NLP models, often as the first data processing layer in a deep learning model, which have been shown to boost the performance of NLP tasks such as translation, sentiment analysis, etc.
- Limitations
  - As embeddings are learned within a small window of surrounding words, sometimes words (e.g., good and bad) share almost the same embedding, which is problematic for sentiment analysis.
  - Traditional word embedding algorithms assign a distinct vector to each word, which can't account for polysemy.

Transformer

- "Attention is All You Need" by Vaswani et al. in 2017.
- Transformer-based language models are based on the transfer architecture, a neural network that uses self-attention mechanisms to process input sequences of tokens and produce output sequences.
- Self-attention mechanism allows the model to weigh the importance of different input tokens when generating the output sequence.
- Unlike traditional recurrent neural networks (RNNs), which process sequences in a sequential manner, Transformers can process sequences in parallel, making them much faster and more efficient.



https://huggingface.co/course/

Transformer

- Pretraining: transformer models are trained on large amount of raw text in a self-supervised fashion
  - The objective is automatically computed from the inputs of the model; don’t need humans to label the data
- Fin-tunning, transfer learning: the model is fine-tuned in a supervised way using human-annotated labels on a specific practical task.
- Transformer architecture consists of an encoder and a decoder.
  - Encoder receives an input and builds a produce a sequence of hidden representations of it (its features). This means that the model is optimized to acquire understanding from the input.
  - Decoder uses the encoder’s representation (features) along with other inputs to generate a target sequence. This means that the model is optimized for generating outputs.

Transformer

- Encoder Models
  - Good for tasks that requires understanding of the input, such as sentence classification, named entity recognition, and extractive question answering.
  - BERT (Bidirectional Encoder Representations from Transformers)
- Decoder Models
  - Good for generative tasks such as text generation
  - GPT (Generative Pre-trained Transformer)
- Sequence-to-sequence models
  - Good for generative tasks that require an input, such as translation, summarization, or generative question answering
  - BART (Bidirectional and Auto-Regressive Transformers)
  - T5 (Text-to-Text Transfer Transformer)

Transformers are big models

NLP in Biomedicine

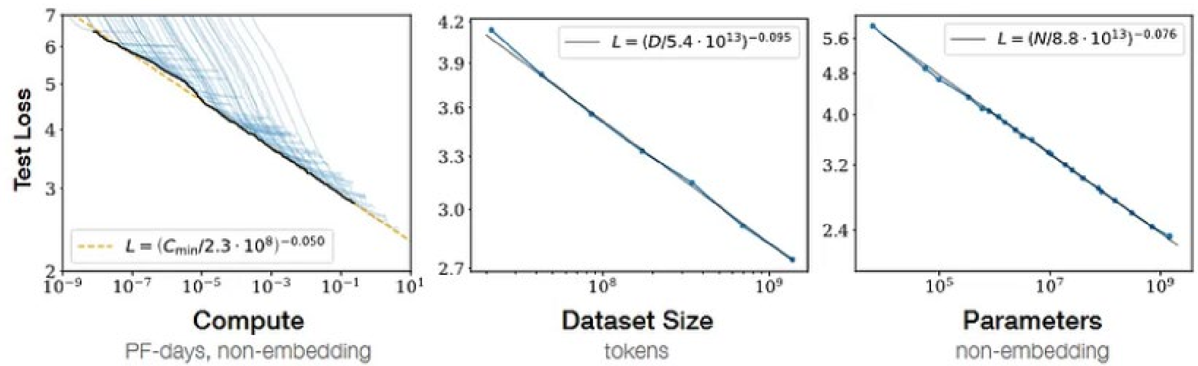
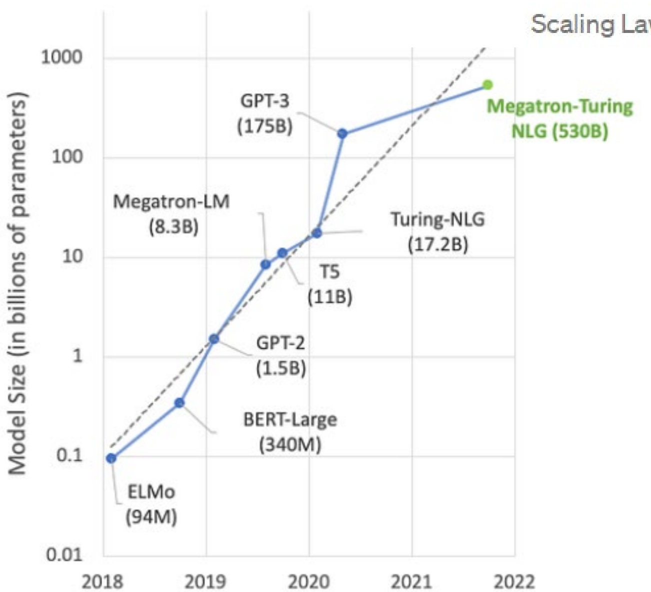
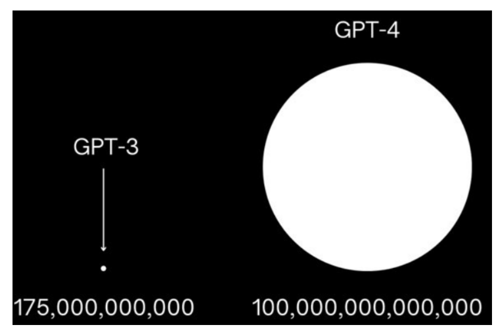


Figure 1 Language modeling performance improves smoothly as we increase the model size, datasetset size, and amount of compute[2] used for training. For optimal performance all three factors must be scaled up in tandem. Empirical performance has a power-law relationship with each individual factor when not bottlenecked by the other two.

Scaling Laws for Neural Language Models. Image by Cornell University.



Model size (in billions of parameters). Image by Nvidia.



Deep Learning Methods in Clinical NLP

Deep Learning Methods in Clinical NLP

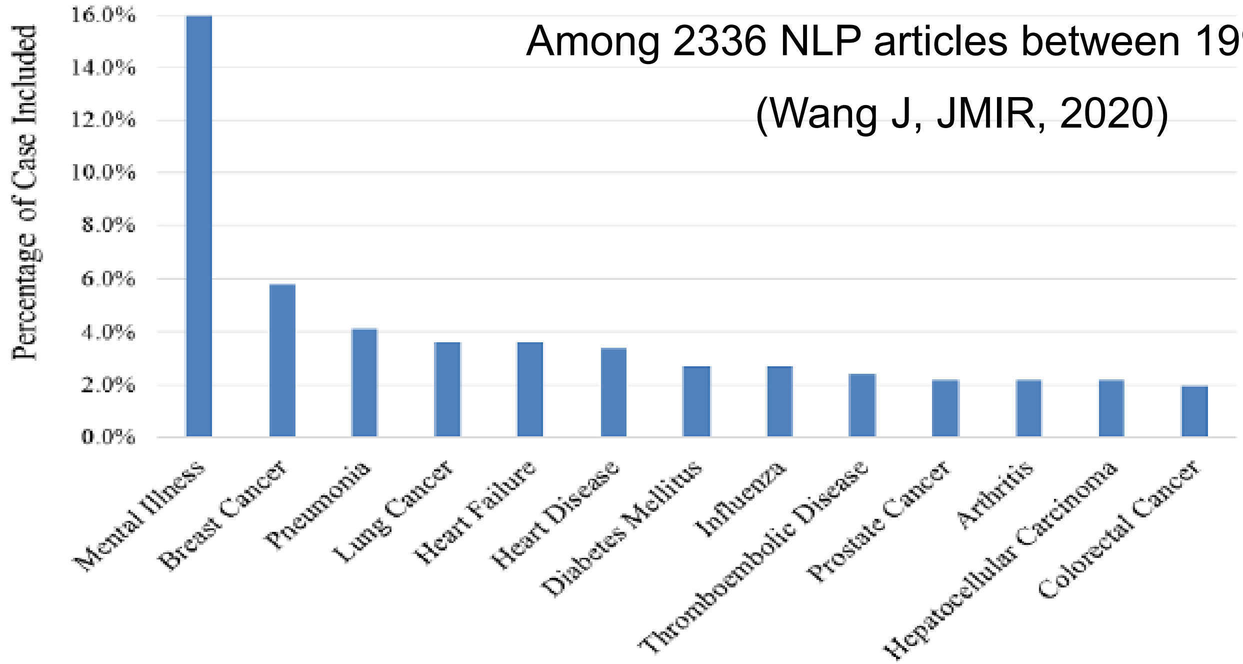
(n=212 articles)

(Wu S, JAMIA 2019)

Architecture	Method	Freq.
RNN	LSTM	109
	GRU	16
	Vanilla RNN	5
	Tree-LSTM	1
	CNN-LSTM	3
CNN	CNN	80
	CNN-LSTM	3
FFNN	NN	22
Embeddings Only	Embeddings	21
Other	Autoencoder	3
	DBN	3
	Other DL	3
	Capsule	1
	Memory Network	1
	RecursiveNN	1
	Transformer	1

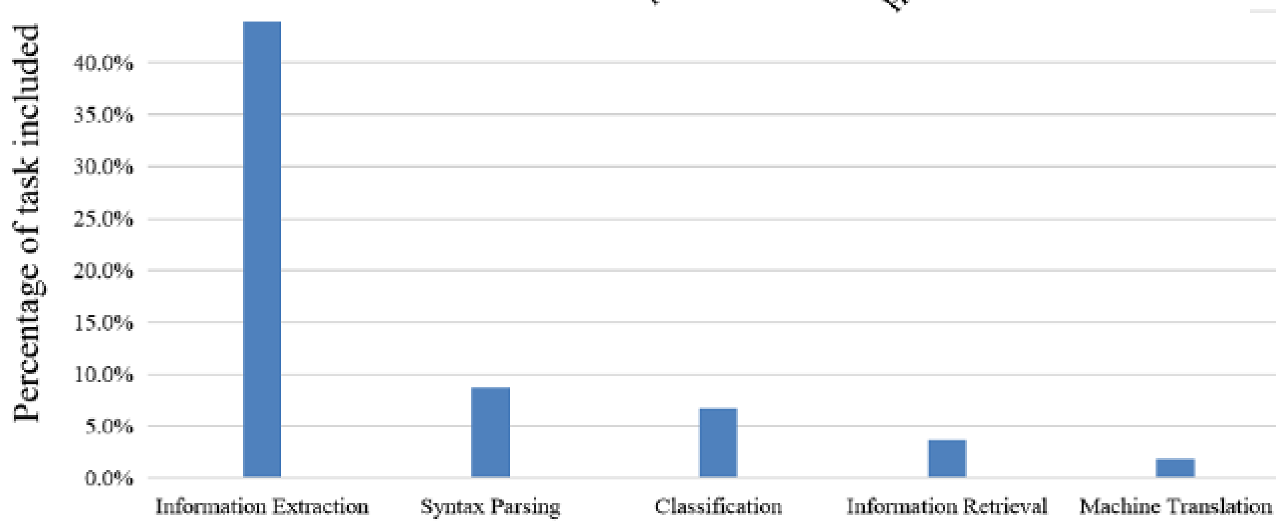
Clinical Domains and Tasks

NLP in Biomedicine



Among 2336 NLP articles between 1999-2018

(Wang J, JMIR, 2020)



I2b2/n2c2 NLP Challenges

- 2006, 1) De-identification and 2) Smoking
- 2008, Obesity
- 2009, Medication extraction
- 2010, Relations (of medical problems, tests, treatments)
- 2011, 1) Co-reference (anaphora) resolution and 2) Sentiment classification (emotions in suicide notes)
- 2012, Temporal relations
- 2014, 1) De-identification and 2) Identifying risk factors for heart disease over time
- 2016, 1) De-identification and 2) RDoc classification (determine symptom severity based on a patient's initial psychiatric evaluation)
- 2018, 1) Cohort selection for clinical trials and 2) Adverse drug events and medication extraction in EHRs
- 2019, 1) Clinical semantic textual similarity 2) Family history 3) Clinical concept normalization 4) Novel data use
- 2022, 1) Contextualized medication event extraction, 2) Social determinants of health, 3) Progress note understanding: assessment and plan reasoning,

https://n2c2.dbmi.hms.harvard.edu/

The challenges and data sets are now administered through the DBMI Data Portal.

MTERMS Research Areas

MTERMS NLP

Clinical Documentation
- Speech recognition / clinical dictation
- Document quality
- Template / Similarity

Data Extraction Case Identification
- Clinical info extraction
  - e.g. diagnosis, social behavioral factors, functional status, ADEs
- Case/cohort identification
  - e.g., opioid use disorder, psychosis
- Case encoding
  - e.g., Malpractice claim case encoding and auditing

Data Mining Knowledge Discovery
- Document classification
- Relation identification
- Topic modeling
- Clustering
- Active learning

Clinical Decision Support
- Information reconciliation
  - Medication, problem and allergy reconciliation
- Prediction
  - Hospital readmission
  - Mortality
  - Clinical deterioration
- Cancer screening abnormal result identification

Research / Innovation
- New method development
- Application development
- Multimodal data
- Visualization
- Other areas: Pharmacovigilance

http://mterms.bwh.harvard.edu/

MTERMS Applications

- Real-time pilots (integrated with Epic)
  - Allergy reconciliation module
    - Medication reconciliation module (in LMR)
  - Cancer Screening Follow-up (primary care)
  - Patient clinical deterioration based on nursing notes and EHR (inpatient)
- Near real-time
  - Patient mortality predication to improve palliative care intervention
- Research projects
  - Allergic and adverse reactions
  - Opioid use disorder patient identification
  - Gunshot intension classification
  - Malpractice cases (coding + similar cases)
  - Psychosis identification
  - Confounding factors for pharmacoepidemiology studies
  - Dementia/cognitive decline
  - PASCLex: Post-Acute Sequelae of COVID-19 (PASC) Symptom
  - Using Twitter data to understand public perceptions of approved vs. off-label user for COVID-19-related medications
  - Examination of Stigmatizing Language in the Electronic Health Record

Allergic and adverse drug reactions

- Reconcile medication and allergy information within EHR
- Identify allergic reactions in hospital safety reports
- Compare ADRs reported in EHR and social media
- Identify ADRs from EHR

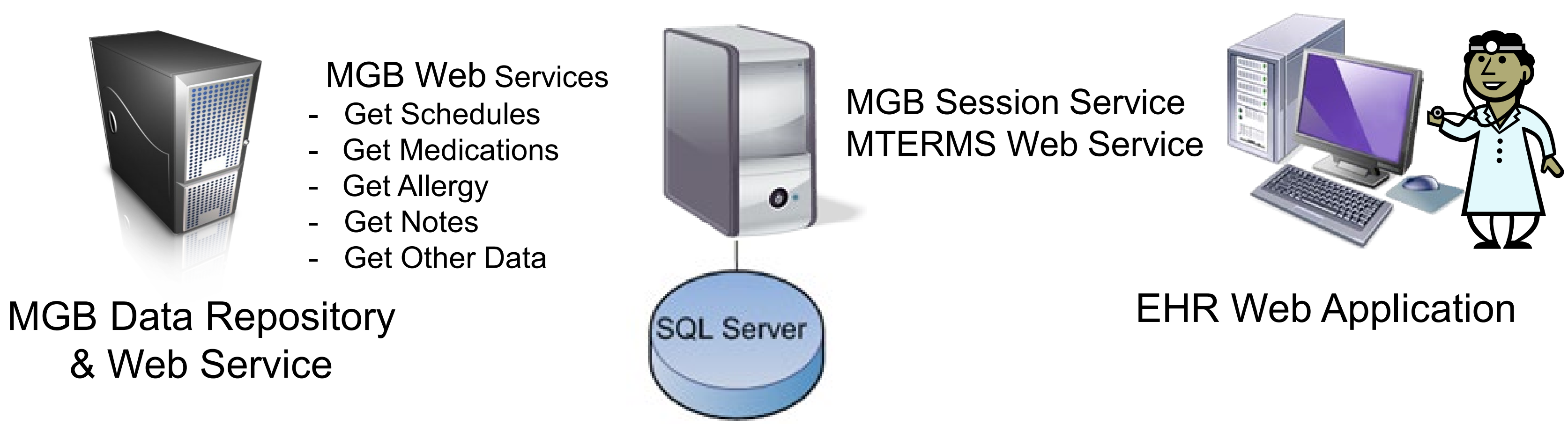
NLP service to support real-time(or near real-time) applications: System Architecture

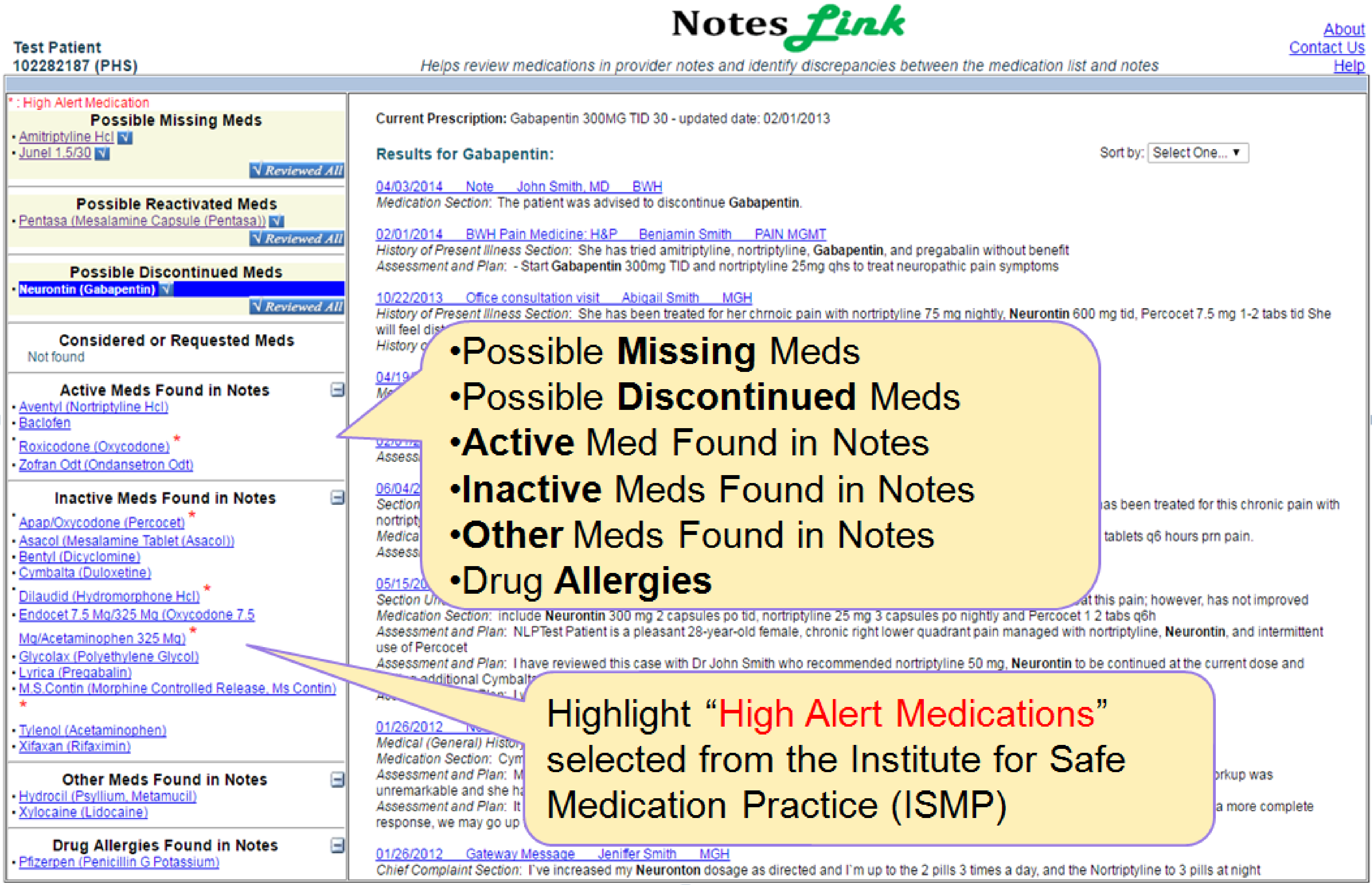
Objectives

- Improving data interoperability
- Integrating NLP with EHRs
- Providing clinical decision support
- Improving patient medication safety

MTERMS NLP Services

- MTERMS Natural Language Processing
- Batch Processing & Summarization
- Knowledge Base
- Web Application

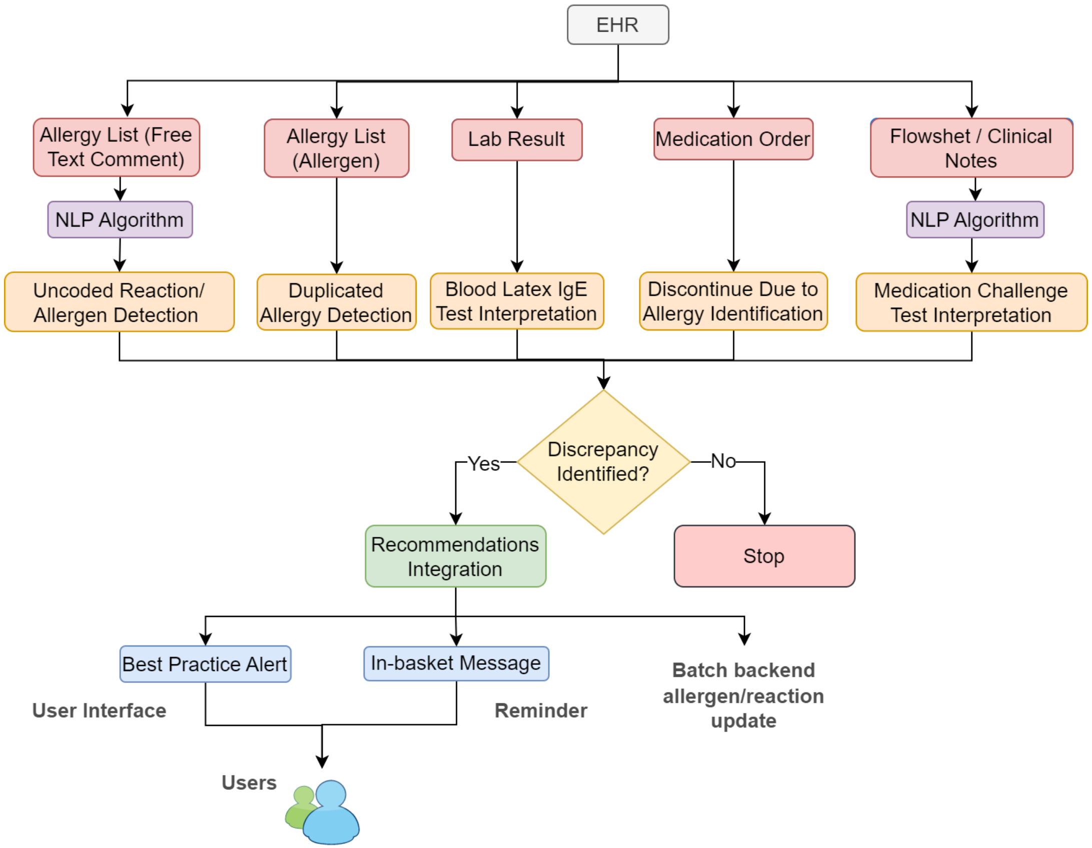




Notes Link
Test Patient
102282187 (PHS)
Helps review medications in provider notes and identify discrepancies between the medication list and notes
About
Contact Us
Help
* : High Alert Medication
Possible Missing Meds
Amitriptyline Hcl
Junel 1.5/30
Reviewed All
Possible Reactivated Meds
Pentasa (Mesalamine Capsule (Pentasa))
Reviewed All
Possible Discontinued Meds
Neurontin (Gabapentin)
Reviewed All
Considered or Requested Meds
Not found
Active Meds Found in Notes
Aventyl (Nortriptyline Hcl)
Baclofen
Roxicodone (Oxycodone) *
Zofran Odt (Ondansetron Odt)
Inactive Meds Found in Notes
Apap/Oxycodone (Percocet) *
Asacol (Mesalamine Tablet (Asacol))
Bentyl (Dicyclomine)
Cymbalta (Duloxetine)
Dilaudid (Hydromorphone Hcl) *
Endocet 7.5 Mg/325 Mg (Oxycodone 7.5 Mg/Acetaminophen 325 Mg) *
Glycolax (Polyethylene Glycol)
Lyrica (Pregabalin)
M.S.Contin (Morphine Controlled Release, Ms Contin) *
Tylenol (Acetaminophen)
Xifaxan (Rifaximin)
Other Meds Found in Notes
Hydrocil (Psyllium, Metamucil)
Xylocaine (Lidocaine)
Drug Allergies Found in Notes
Pfizerpen (Penicillin G Potassium)
Current Prescription: Gabapentin 300MG TID 30 - updated date: 02/01/2013
Results for Gabapentin:
Sort by: Select One...
04/03/2014 Note John Smith, MD BWH
Medication Section: The patient was advised to discontinue Gabapentin.
02/01/2014 BWH Pain Medicine: H&P Benjamin Smith PAIN MGMT
History of Present Illness Section: She has tried amitriptyline, nortriptyline, Gabapentin, and pregabalin without benefit
Assessment and Plan: - Start Gabapentin 300mg TID and nortriptyline 25mg qhs to treat neuropathic pain symptoms
10/22/2013 Office consultation visit Abigail Smith MGH
History of Present Illness Section: She has been treated for her chrnoic pain with nortriptyline 75 mg nightly, Neurontin 600 mg tid, Percocet 7.5 mg 1-2 tabs tid She
Medication Section: include Neurontin 300 mg 2 capsules po tid, nortriptyline 25 mg 3 capsules po nightly and Percocet 1 2 tabs q6h
Assessment and Plan: NLPTest Patient is a pleasant 28-year-old female, chronic right lower quadrant pain managed with nortriptyline, Neurontin, and intermittent use of Percocet
Assessment and Plan: I have reviewed this case with Dr John Smith who recommended nortriptyline 50 mg, Neurontin to be continued at the current dose and
01/26/2012 Gateway Message Jenifer Smith MGH
Chief Complaint Section: I've increased my Neuronton dosage as directed and I'm up to the 2 pills 3 times a day, and the Nortriptyline to 3 pills at night
•Possible Missing Meds
•Possible Discontinued Meds
•Active Med Found in Notes
•Inactive Meds Found in Notes
•Other Meds Found in Notes
•Drug Allergies
Highlight "High Alert Medications" selected from the Institute for Safe Medication Practice (ISMP)

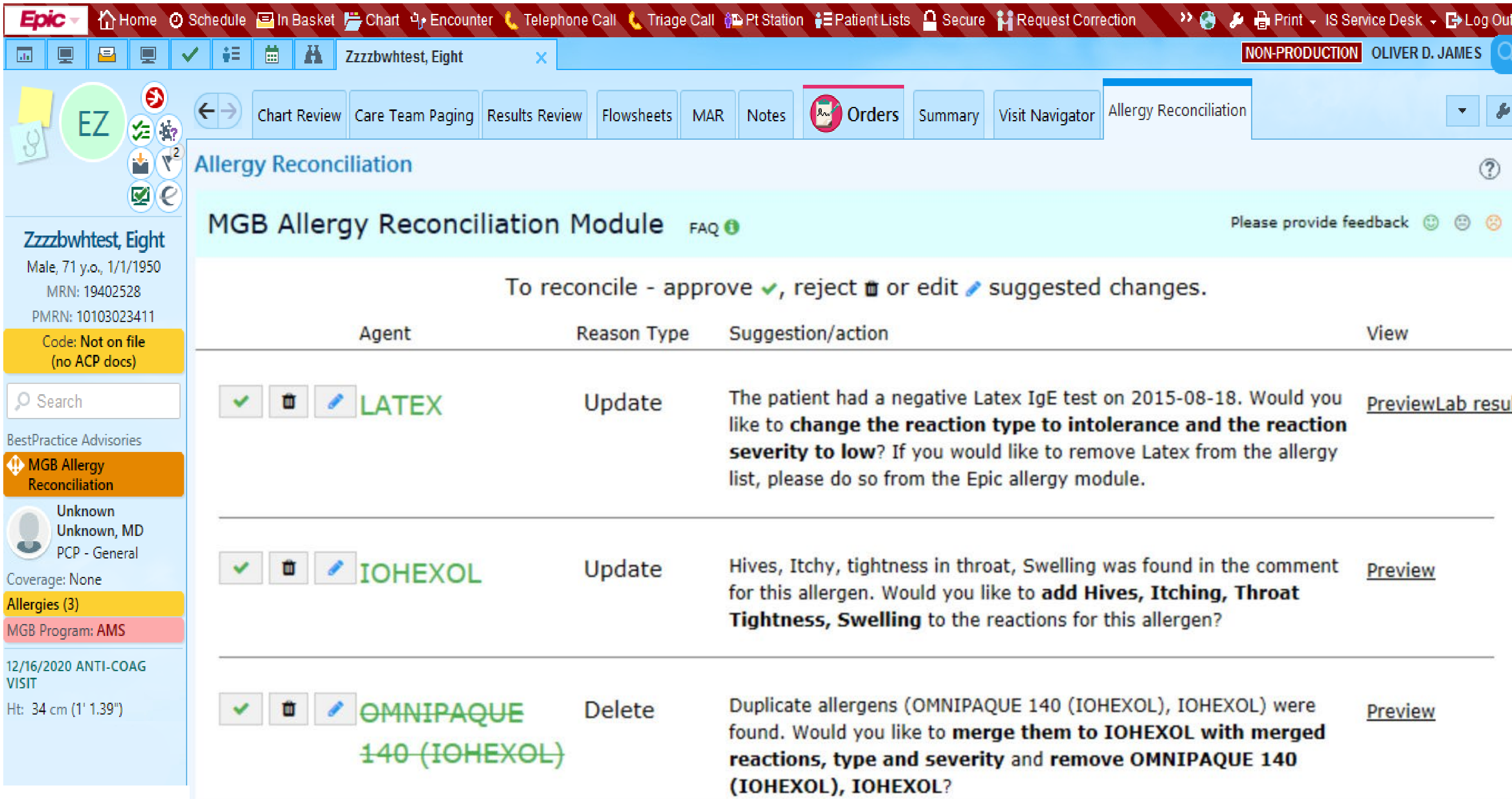
Allergy Reconciliation across the EHR

- Accurate and complete allergy documentation in the EHR is essential to guide clinical decision-making.
- Patient allergy information often exists in several locations in the EHR, and patients' allergy lists are often inaccurate or incomplete
- Automatically identify discrepancies in allergy information from across the EHR
- User interface – providers can accept, reject, or modify suggested changes
  - Automatically add free-text reactions to allergy list
- Integrated with Epic in pilot study with 111 BWH providers



*Preliminary unpublished data

EHR Allergy Reconciliation Module



Unpublished; Please don't distribute

› Front Allergy. 2022 May 10;3:904923. doi: 10.3389/falgy.2022.904923. eCollection 2022.

Reconciling Allergy Information in the Electronic Health Record After a Drug Challenge Using Natural Language Processing

Ying-Chih Lo [1 2], Sheril Varghese [1], Suzanne Blackley [3], Diane L Seger [1 3], Kimberly G Blumenthal [2 4], Foster R Goss [5], Li Zhou [1 2]

Affiliations + expand

PMID: 35769562 PMCID: PMC9234873 DOI: 10.3389/falgy.2022.904923

Free PMC article

Abstract

Background: Drug challenge tests serve to evaluate whether a patient is allergic to a medication. However, the allergy list in the electronic health record (EHR) is not consistently updated to reflect the results of the challenge, affecting clinicians' prescription decisions and contributing to inaccurate allergy labels, inappropriate drug-allergy alerts, and potentially ineffective, more toxic, and/or costly care. In this study, we used natural language processing (NLP) to automatically detect discrepancies between the EHR allergy list and drug challenge test results and to inform the clinical recommendations provided in a real-time allergy reconciliation module.

Methods: This study included patients who received drug challenge tests at the Mass General Brigham (MGB) Healthcare System between June 9, 2015 and January 5, 2022. At MGB, drug challenge tests are performed in allergy/immunology encounters with routine clinical documentation in notes and flowsheets. We developed a rule-based NLP tool to analyze and interpret the challenge test results. We compared these results against EHR allergy lists to detect potential discrepancies in allergy documentation and form a recommendation for reconciliation if a discrepancy was identified. To evaluate the capability of our tool in identifying discrepancies, we calculated the percentage of challenge test results that were not updated and the precision of the NLP algorithm for 200 randomly sampled encounters.

Results: Among 200 samples from 5,312 drug challenge tests, 59% challenged penicillin reactivity and 99% were negative. 42.0%, 61.5%, and 76.0% of the results were confirmed by flowsheets, NLP, or both, respectively. The precision of the NLP algorithm was 96.1%. Seven percent of patient allergy lists were not updated based on drug challenge test results. Flowsheets alone were used to identify 2.0% of these discrepancies, and NLP alone detected 5.0% of these discrepancies. Because challenge test results can be recorded in both flowsheets and clinical notes, the combined use of NLP and flowsheets can reliably detect 5.5% of discrepancies.

Conclusion: This NLP-based tool may be able to advance global delabeling efforts and the effectiveness of drug allergy assessments. In the real-time EHR environment, it can be used to examine patient allergy lists and identify drug allergy label discrepancies, mitigating patient risks.

- Allergy labels are common, often incorrect, and potentially harmful.

  Up to 15% of hospitalized patients, 6% to 10% of the general population report a penicillin allergy. Of these individuals, 94% can tolerate penicillin after formal allergy testing

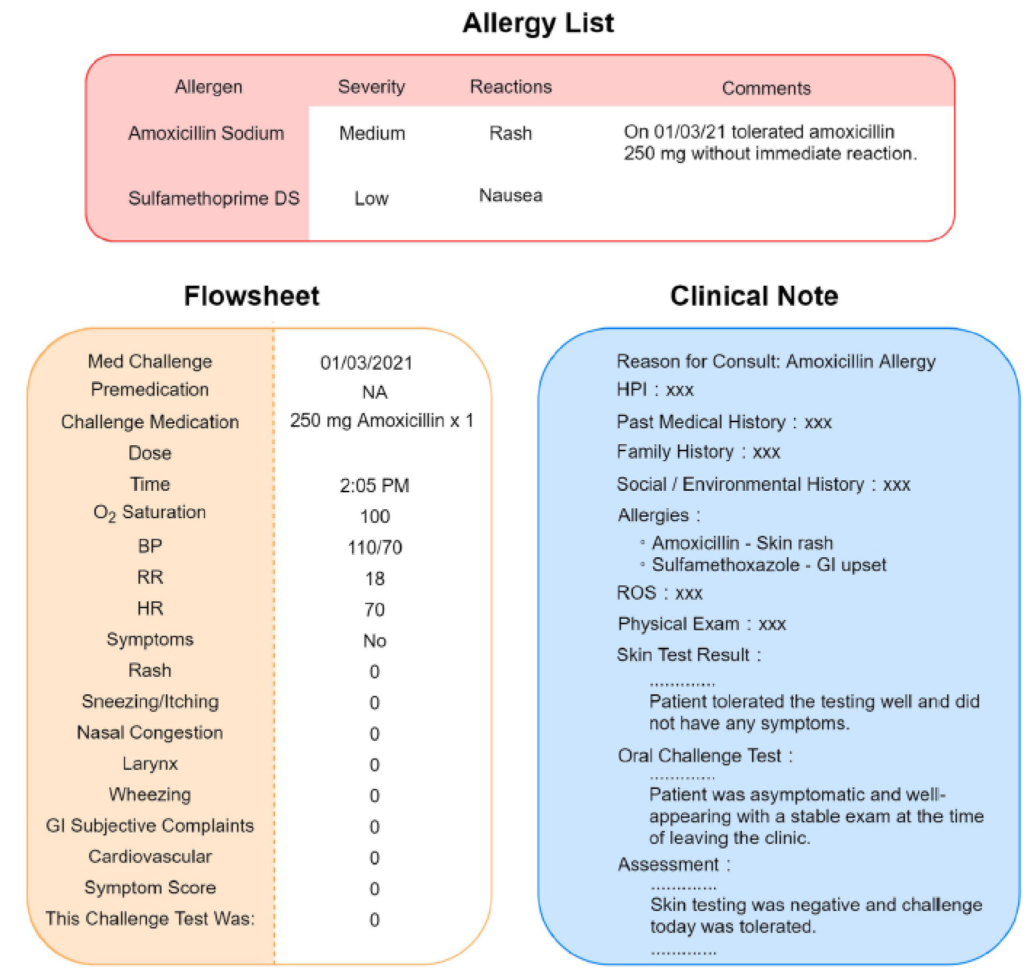


FIGURE 1 | Scenario of allergy information discrepancy in EHR. Allergy list, flowsheets, and clinical notes are different locations in the EHR that store allergy information. A negative result of drug challenge test may not be updated to the allergy list accordingly. Sometimes, the physician would leave a comment instead of removing the allergen from the list, as pictured in this figure.

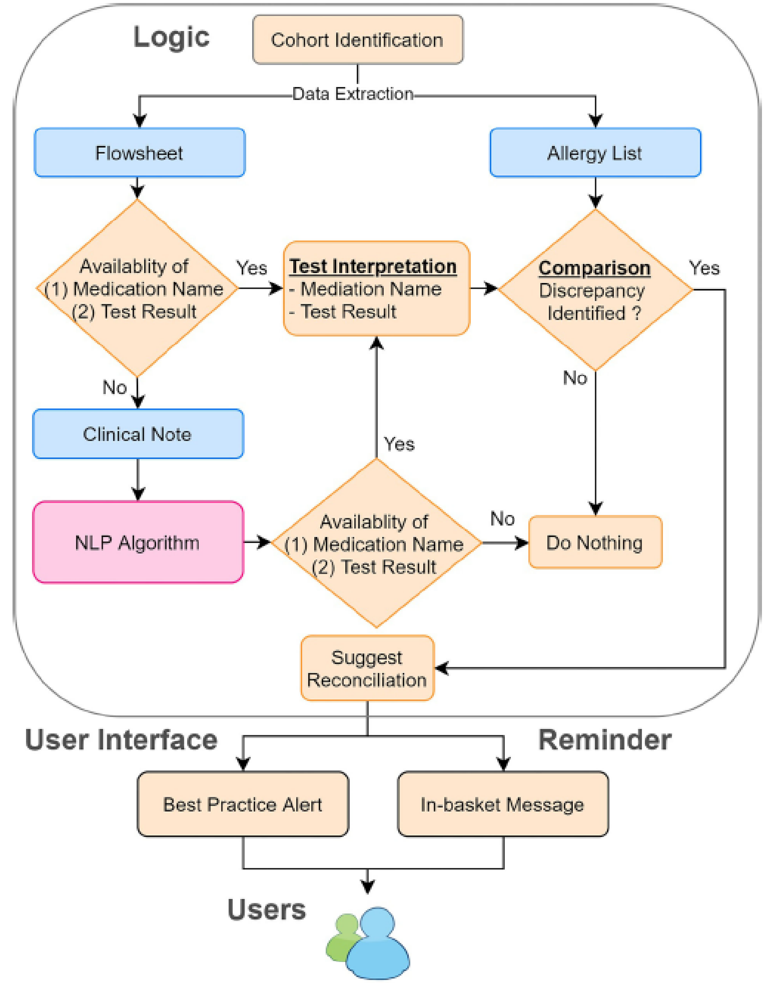


FIGURE 2 | System architecture of the reconciliation module. We combined the information derived from the flowsheets and clinical notes. We then compared this information to the allergy list to identify the discrepancies. If any discrepancies were found, we sent in-basket messages weekly to remind the physician to reconcile the allergy discrepancies by using our tool.

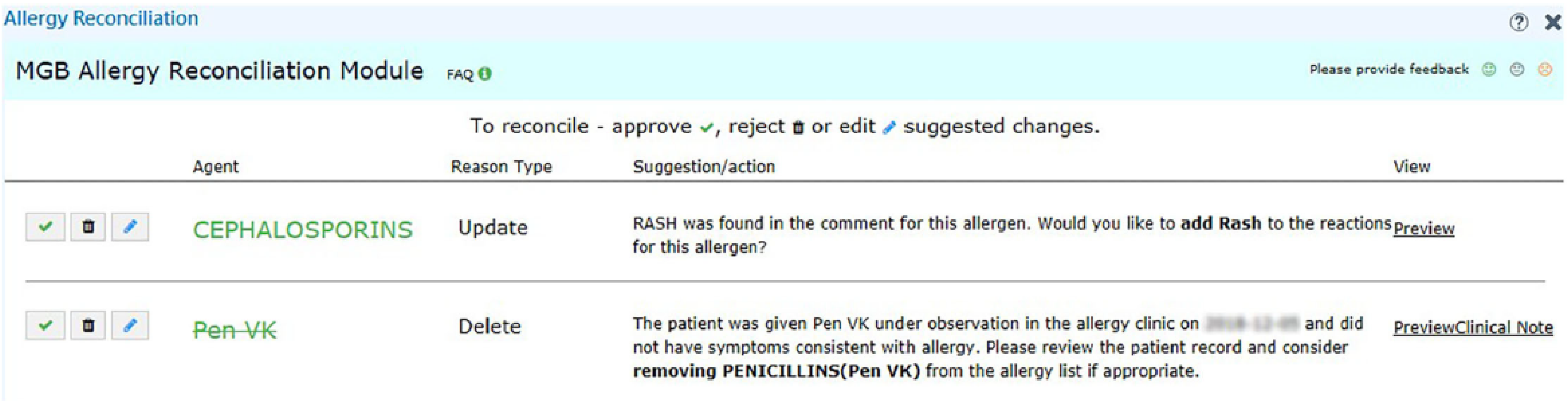


FIGURE 4 | User interface of the recommendation for challenge tests. We provide the reason in addition to the suggested action, such as "Add" and "Delete" for the user to make decision. We also include a hyperlink (right-hand side) to the clinical notes in case the user wants to know more about the reaction.

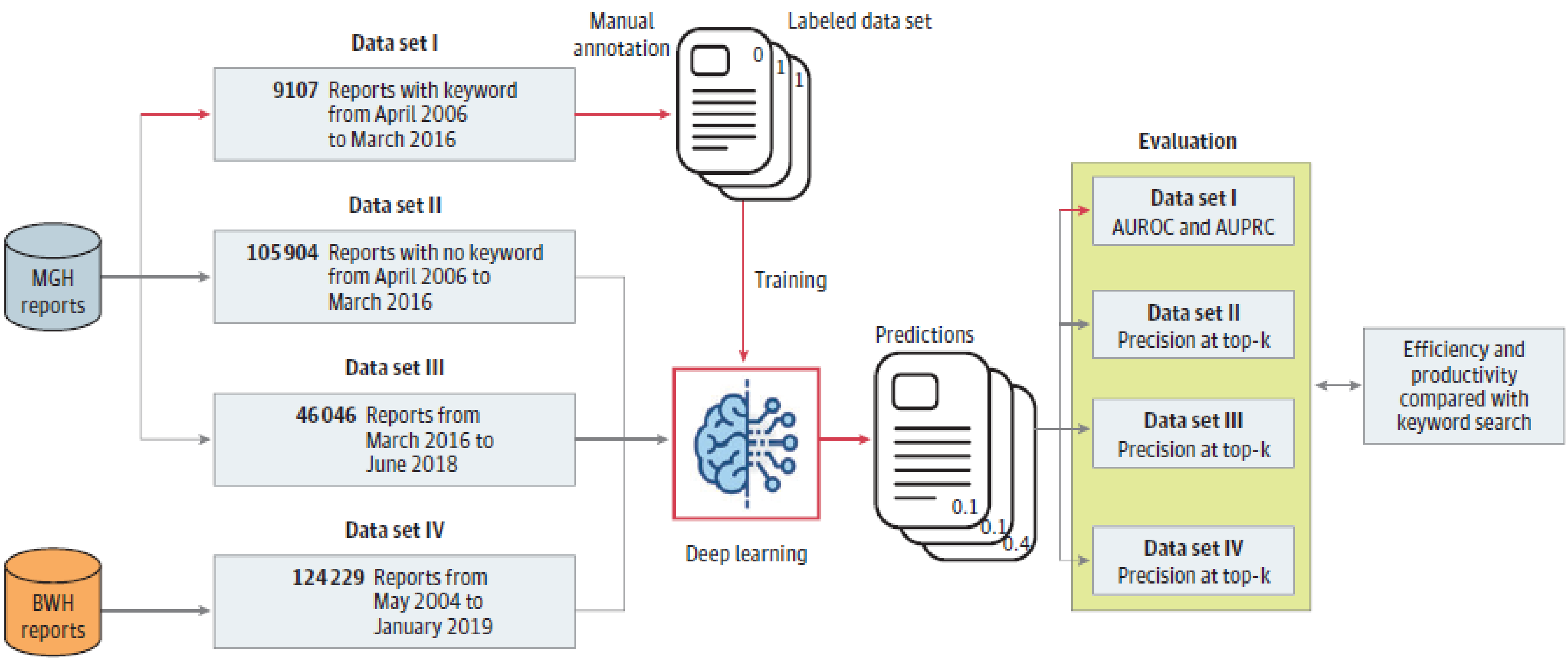
Results: Among 200 samples from 5,312 drug challenge tests, 59% challenged penicillin reactivity and 99% were negative. 42.0%, 61.5%, and 76.0% of the results were confirmed by flowsheets, NLP, or both, respectively. The precision of the NLP algorithm was 96.1%. Seven percent of patient allergy lists were not updated based on drug challenge test results. Flowsheets alone were used to identify 2.0% of these discrepancies, and NLP alone detected 5.0% of these discrepancies. Because challenge test results can be recorded in both flowsheets and clinical notes, the combined use of NLP and flowsheets can reliably detect 5.5% of discrepancies.

Deep Learning to Detect Allergy Events from Hospital Safety Reports

- Allergy safety knowledge is limited by case identification challenges.
- Hospital safety event reporting systems are integral to the detection of patient safety signals in health care, but still lacking are processes to analyze them in a manner that allows for timely feedback to health care professionals.
- We developed an AI method, a hierarchical attention-based deep neural network (DNN), that automatically reads the free-text description of safety reports and identifies cases describing allergic reactions.

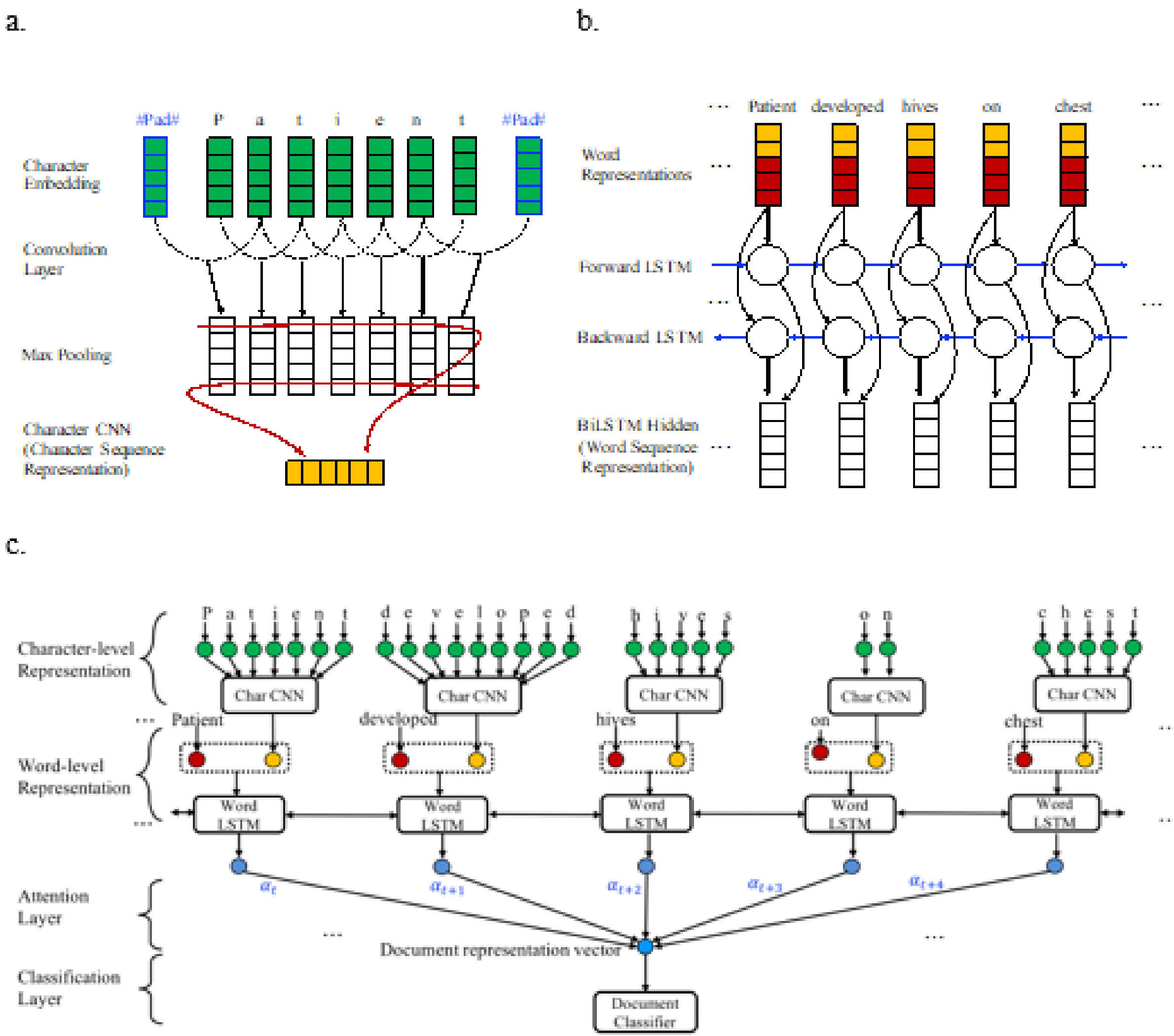
Yang J, Wang L, Phadke NA, Wickner PG, Mancini CM, Blumenthal KG, Zhou L. Development and Validation of a Deep Learning Model for Detection of Allergic Reactions Using Safety Event Reports Across Hospitals. *JAMA Network Open*. 2020 Nov 2;3(11):e2022836

Study Design and Datasets



- Our model was trained on the free-text descriptions of 9,107 labeled reports extracted using expert-curated keywords from MGH's safety event reporting system.
- We then used the model to automatically identify allergy events from nearly 300,000 reports from MGH and BWH across 15 years.

Deep Learning Model



The *first* layer is a character-level encoder, which aims to capture lexical variations (e.g., misspelling) of a word. It encoded the character sequence within each word using a single layer CNN. Each character of a word was represented using randomly initialized character embedding, which was fed as the input of the character-level CNN. The output of the CNN was then fed into a max-pooling function to create a fix-dimension vector for the word.

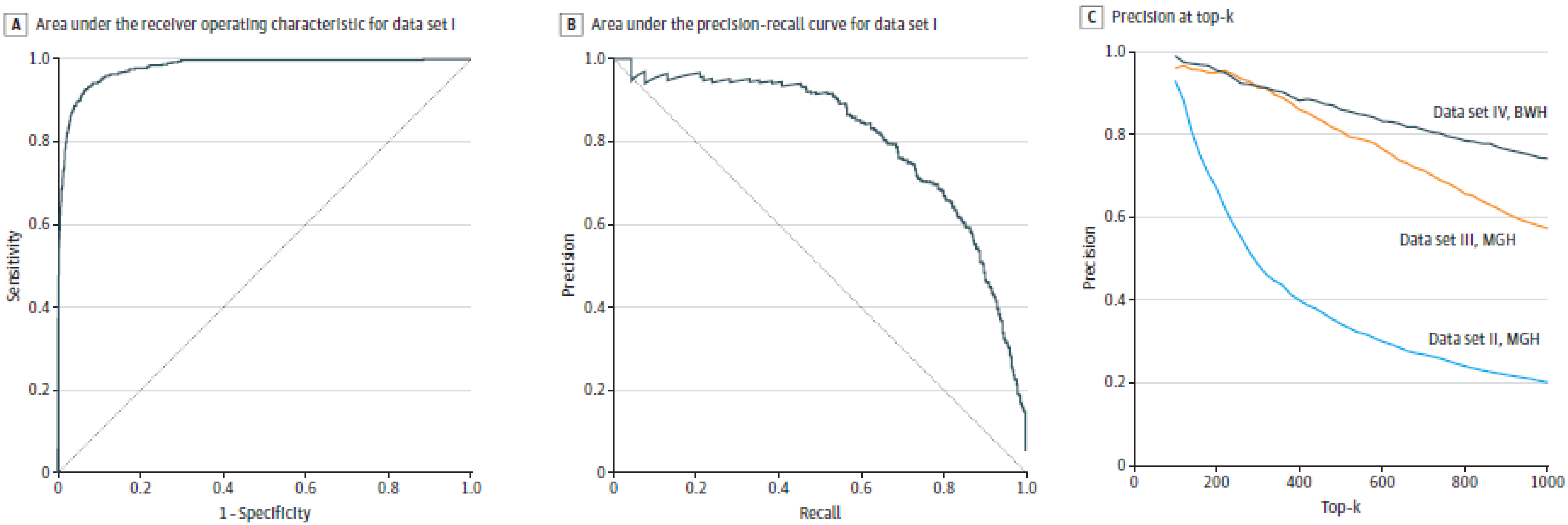
In the *second* layer, each word vector was concatenated with the word's embedding that was pretrained on all MGH reports using word2vec. On top of the concatenated word representation, a LSTM network was built to utilize the contextual information of the whole report and generate an output vector for each word.

Because different words within a report may have different levels of contribution in distinguishing the report, we added an attention model as a *third* layer to assign a unique weight for each word, which was calculated based on the LSTM output vector.

We computed a weighted sum of the LSTM output vectors of all the words in the report to generate a report representation vector, which was then fed into the *fourth* layer, the classifier. The classifier was trained using the cross-entropy loss function and the Stochastic Gradient Descent (SGD) optimizer.

The output of the classifier was a vector representing the probability of whether or not a report described an allergy event.

Model Performance



The deep learning model achieved an AUROC of 0.979 (95%CI, 0.973-0.985) and an area under the precision-recall curve of 0.809 (95%CI, 0.773- 0.845).

Efficiency and Productivity

Data set	Measures	Keyword-search approach	Attention-based DNN model
II	Cases to review	0	1627
	True cases	0	184
	Precision, %	NA	11.3
III	Cases to review	10 131	1984
	True cases	570	625
	Precision, %	5.6	31.5
IV	Cases to review	15 896	5800
	True cases	1344	1569
	Precision, %	8.5	27.1
Total	Cases to review	26 027	9411
	True cases	1914	2378
	Precision, %	7.4	25.3

Compared with the keyword-search approach, the deep learning model **reduced the number of cases for manual review by 64%** and **identified 24% more cases** of confirmed allergic reactions.

Interpretation

A Attention to words contributing to the prediction of positive cases

PT immediately started sneezing x2 after injection of isovue 300. He developed a stuffy nose 4 minutes after injection. Took him to RN station to be examined by RN and Radiologist. He developed a hive on left arm and heavy 15 min after injection.
PT received 100cc's of Isovue 300. Immediately following the injection the PT experienced itchy eyes, face and throat. He also had a racing heart and difficulty breathing. Dr's Smith and Brown responded immediately. Vitals BP 130/80, HR 95. 200ml normal saline was hung and 60mg po Benedryl was administered by Dr. Smith. He was monitored here for approximately 20 minutes. As symptoms resolved we notified his nurse.
Gd injected at 17:20. Pt started coughing and c/o throat. Radiologist evaluated pt. and asked that patient be given oxygen 3 L, BP 120/80, 60mg benadry p.o given.

B Attention to words contributing to the prediction of negative cases

An order to rule out a groin hematoma was written for both Radiology as well as the vascular lab. It was performed in the vascular lab and we did not realize that we were about to repeat the exam until the patient told us that she had already had the exam earlier in the day.
Medication came up in the tube system, there was one 10ml syringe with the appropriate dose and amount (was labled correctly), the second Medication (same as the first) had the same lable but it was on a 60ml bag of D5W. Medication not given, pharmacy called and safety report filed.
Pt ambulated to BR with RN. Once in BR, pt was instructed to pull string when ready. Pt pulled string, RN was in BR with pt, PT stated she felt SOB but not light headed or dizzy. Pt stood up with RN and was walking out of the BR and started to fall forward, RN caught pt and directed to the floor. BP 87/51 SPO2 100% HR 76. MD at bedside. After some IVF, pt SBP 116 and was able to walk assisted to bed with staff.

Regarding interpretability of the model, these attention heatmaps demonstrate how much attention the model gives and to which words when making positive and negative allergy event predictions. Darker color represents a higher attention weight.

ADRs: EHR vs. Social Media

- **Introduction**: Large databases of clinician reported (e.g., EHRs) and patient reported (e.g., social media) adverse drug reactions (ADRs) exist; however, whether patients and clinicians report the same concerns is not clear.
- **Objectives**: Compare EHRs and social media data to better understand differences and similarities between clinician-reported ADRs and patients' concerns
- **Data**:

	EHRs	Social media data
Aspirin	31,817	19,186
Atorvastatin	15,047	23,408

Topaz, M, Lai, K, Dhopeshwarkar, N, Seger, D, Sa'adon, R., Goss, F, Rozenblum, R, **Zhou L**. Clinicians' Reports in Electronic Health Records versus Patients' Concerns in Social Media: A Pilot Study of Adverse Drug Reactions of Aspirin and Atorvastatin. *Drug Safety*. 2016 Mar 39:241-250. doi: 10.1007/s40264-015-0381-x. PMID: 26715498.

EHR vs. Social Media Data

Table 1 Aspirin adverse reactions from clinical and social media databases

Adverse drug reaction	Electronic health record data					Social media data		
	Structured (*n*)	Free-text (*n*)	Structured + free text (*n*)	Structured + free text (%)	Frequency ranking	*n*	%[a]	Frequency ranking
Hives or other rash	4653	2304	6957	21.87	1	2647	13.8	1
GI upset	4732	1869	6601	20.75	2	2193	11.43	2
Bleeding (free text, except nose bleeding)	–	3429	3429	10.78	3	624	3.25	11
Swelling	532	1464	1996	6.27	4	35	0.18	40
Anaphylaxis	1658	55	1713	5.38	5	443	2.31	15
Angioedema	1126	297	1423	4.47	6	117	0.61	15
Bronchospasm or wheezing	885	157	1042	3.27	7	415	2.16	16
Nausea or vomiting	382	504	886	2.78	8	–	–	–
Hypotension	34	9	43	0.14	21	1047	5.46	6
Reye's Syndrome (free text)	–	21	21	0.07	27	1359	7.08	5
Hypoglycemia (social media only)	–	–	–	–	–	457	2.38	14

The most frequently reported ADRs matched the most frequent patients' concerns. However, several less frequently reported reactions were more prevalent on social media. Overall, we found a relatively strong positive and statistically significant correlation between the frequency ranking of reactions and patients' concerns for atorvastatin (Pearson's r = 0.61, p<0.001) but not for aspirin (Pearson's r = 0.1, p = 0.69).

MTERMS Applications

- Real-time pilots (integrated with Epic)
  - Allergy reconciliation module
    - Medication reconciliation module (in LMR)
  - Cancer Screening Follow-up (primary care)
  - Patient clinical deterioration based on nursing notes and EHR (inpatient)
- Near real-time
  - **Patient mortality predication to improve palliative care intervention**
- Research projects
  - Allergic and adverse reactions
  - Opioid use disorder patient identification
  - Gunshot intension classification
  - Malpractice cases (coding + similar cases)
  - Psychosis identification
  - Confounding factors for pharmacoepidemiology studies
  - Dementia/cognitive decline
  - PASCLex: Post-Acute Sequelae of COVID-19 (PASC) Symptom
  - Examination of Stigmatizing Language in the Electronic Health Record

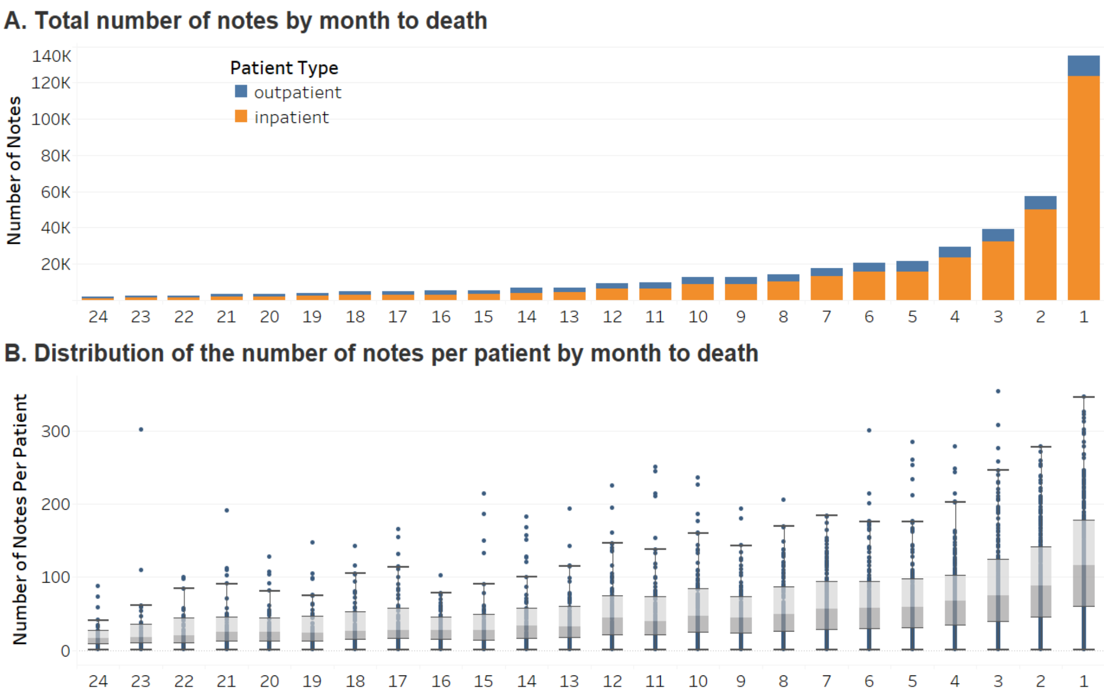
Deep Learning for Mortality Prediction in Selecting Patients for Earlier Palliative Care Interventions

One of the largest challenges in expanding palliative care is identifying those patients who can benefit the most

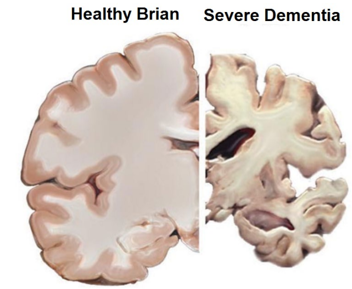
- Which patients will benefit from which interventions and when?
  - ✓ Predict patients' clinical trajectories
  - ✓ Identify those who need palliative care
  - ✓ Determine the right time to start the interventions
- Existing population management algorithms generally target patients with high healthcare utilization
- Most clinicians hesitate to provide prognosis information to patients

There is an urgent need to leverage information technology and the EHR to provide decision support for healthcare providers

Large amount of free-text EHR data



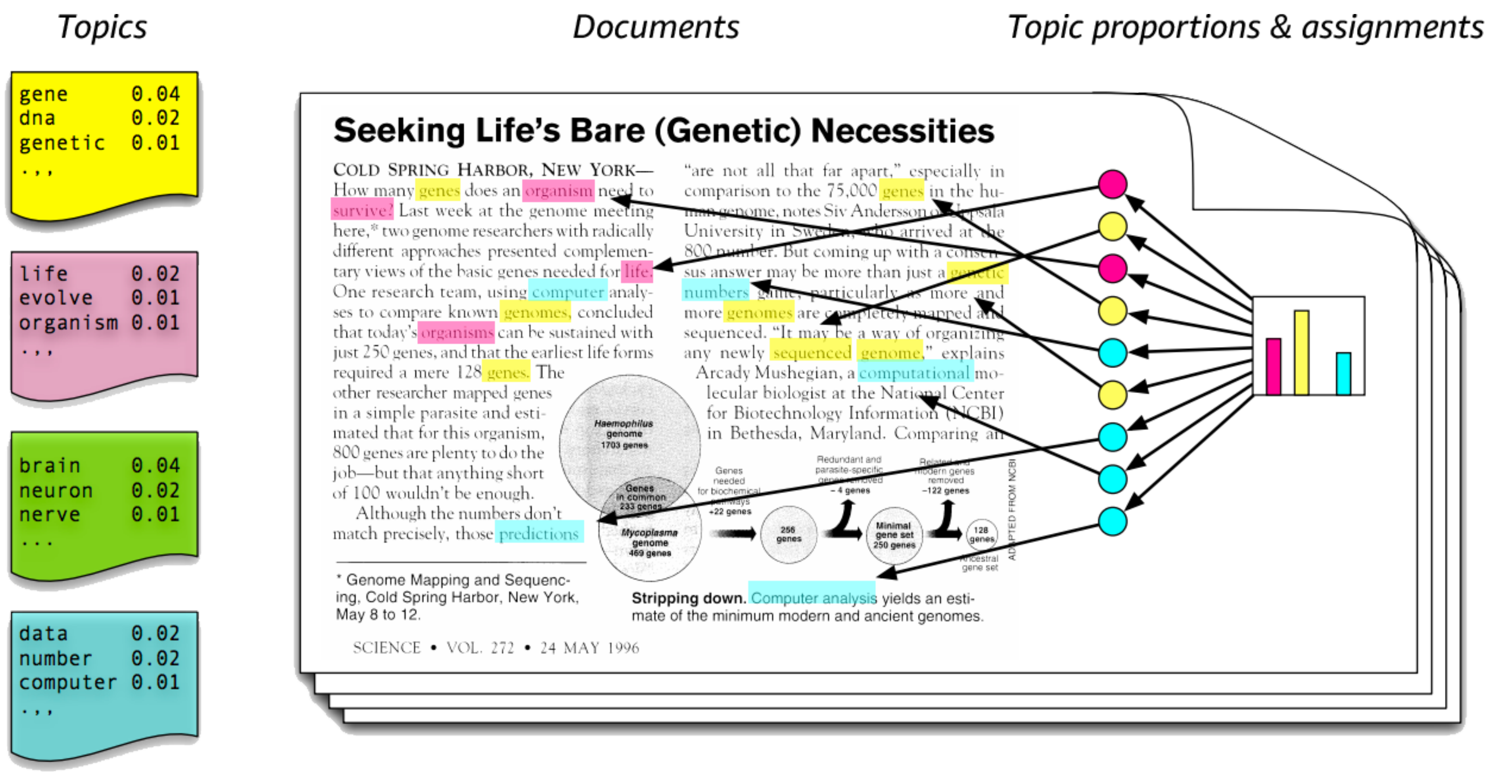
Dementia



- > 5.5 million Americans in 2017
- Sixth leading cause of death
- One of the costliest disease
  - $259 billion for elderly per year

Number of clinical notes per patient with dementia by month over the last two years of life (a total of 432,007 notes of 7,875 patients) (Wang L, Zhou L, et al. AMIA 2019)

Latent Topic Modeling

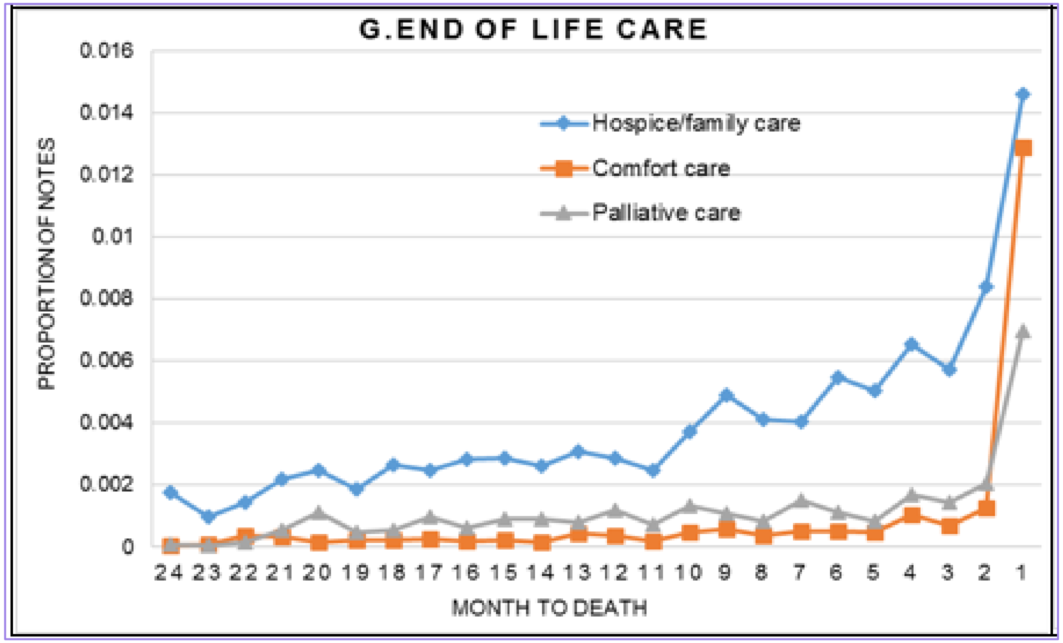


Topic modeling applies statistical-based unsupervised machine learning approaches to discover abstract topics that occur in a collection of documents. The topics are clusters of similar words. Each document may have multiple topics with different proportions.

Blei, D.M., 2012. Probabilistic topic models. *Communications of the ACM*, *55*(4), pp.77-84.

Topics

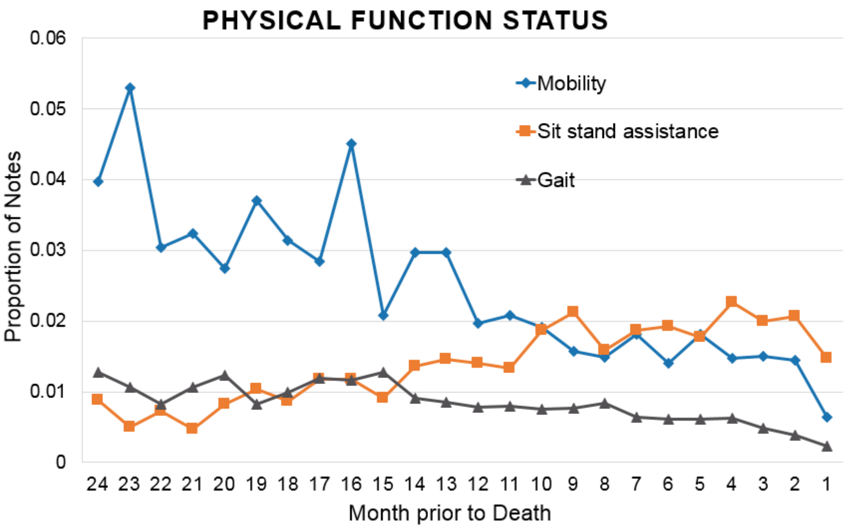
NLP in Biomedicine

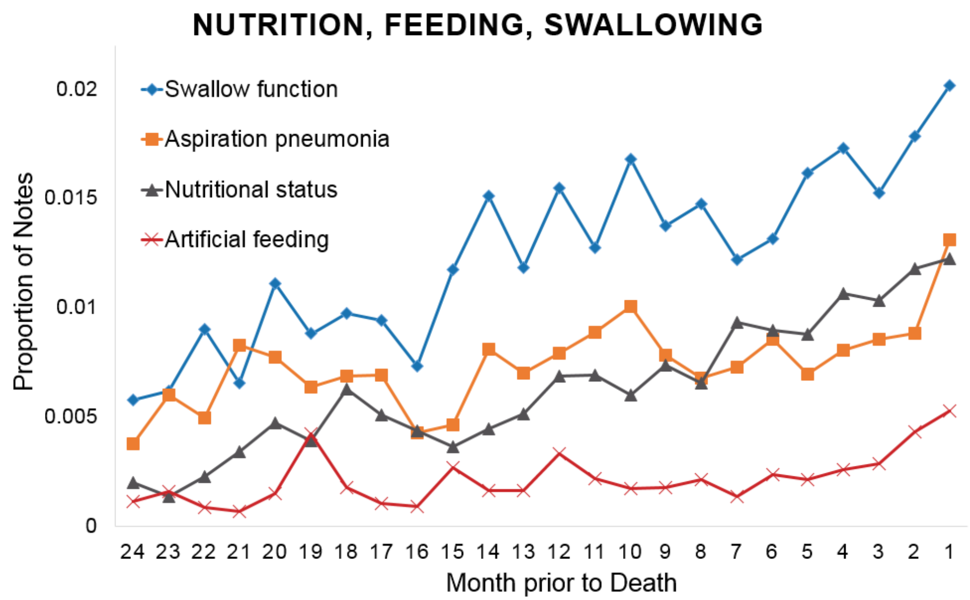


End of life care	**Family/hospice care**	care family hospice home dnr dementia palliative dni discussion goal intake daughter failure admission thrive
	Comfort care	comfort prn care cmo morphine family hospice measure transition pain comfortable palliative palliative dni dilaudid dnr
	Palliative care	care palliative pain prn continue family delirium time comfort review symptom well management agitation follow

Wang L, Lakin J, Riley C, Korach Z, Frain L, Zhou L. Disease Trajectories and End-of-Life Care for Dementias: Latent Topic Modeling and Trend Analysis Using Clinical Notes. AMIA Annu Symp Proc. 2018 (Distinguished Paper Award)

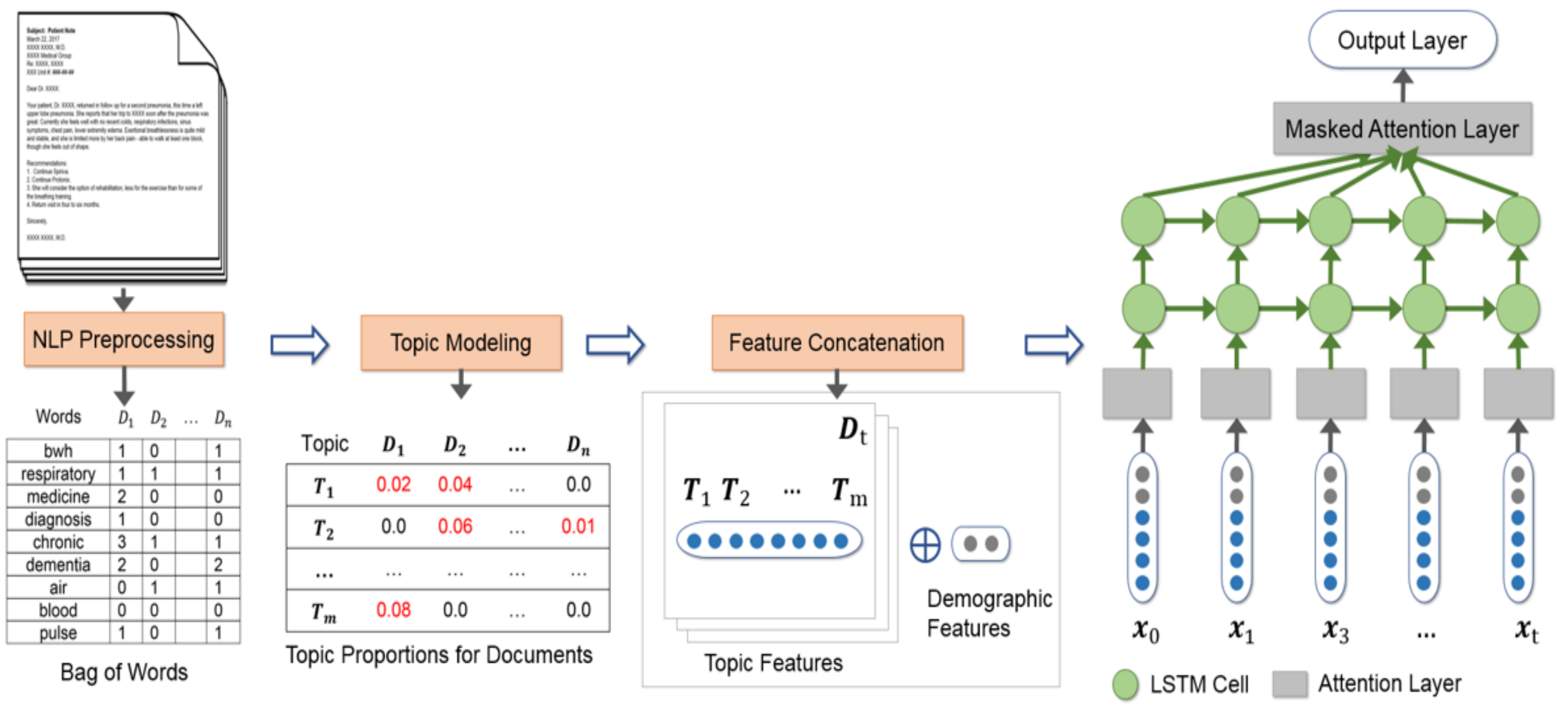
Topics in clinical notes during patients' last two years of life





Wang L, Lakin J, Riley C, Korach Z, Frain L, Zhou L. Disease Trajectories and End-of-Life Care for Dementias: Latent Topic Modeling and Trend Analysis Using Clinical Notes. *AMIA Annu Symp Proc.* 2018 Dec 5;2018:1056-1065 (Distinguished paper award).

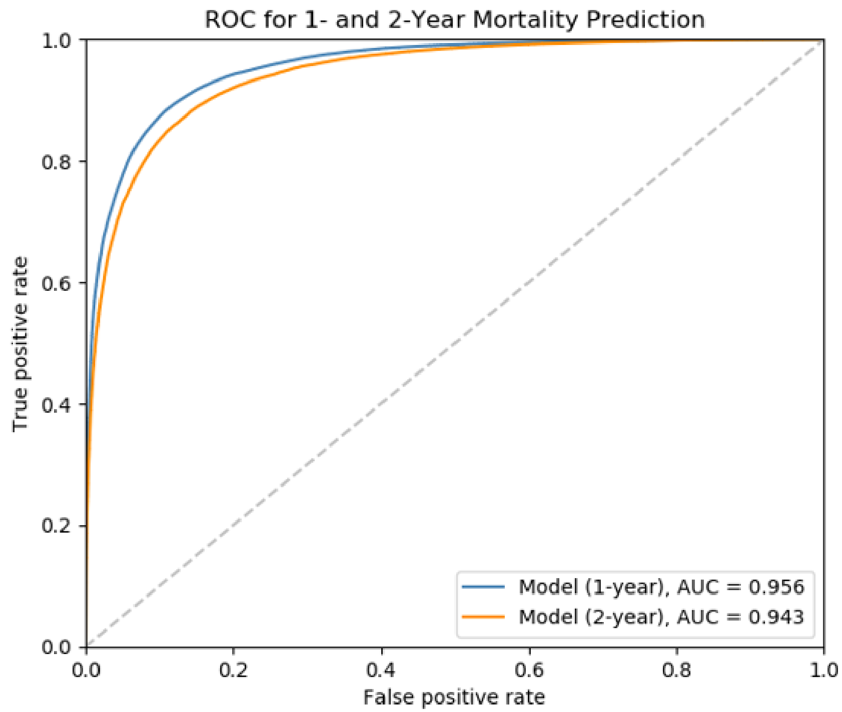
Deep learning for mortality prediction



Wang L, **Sha L**, Lakin JR, Bynum J, Bates DW, Hong P, Zhou L. Development and Validation of a Deep Learning Algorithm for Mortality Prediction in Selecting Patients With Dementia for Earlier Palliative Care Interventions. *JAMA Netw Open.* 2019.2(7):e196972.

Results

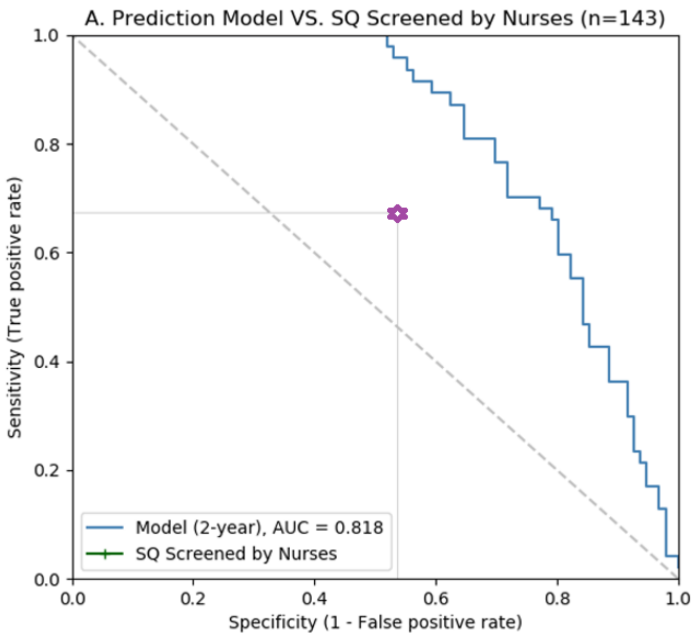
Our study shows promising results in patient stratification for clinical practice

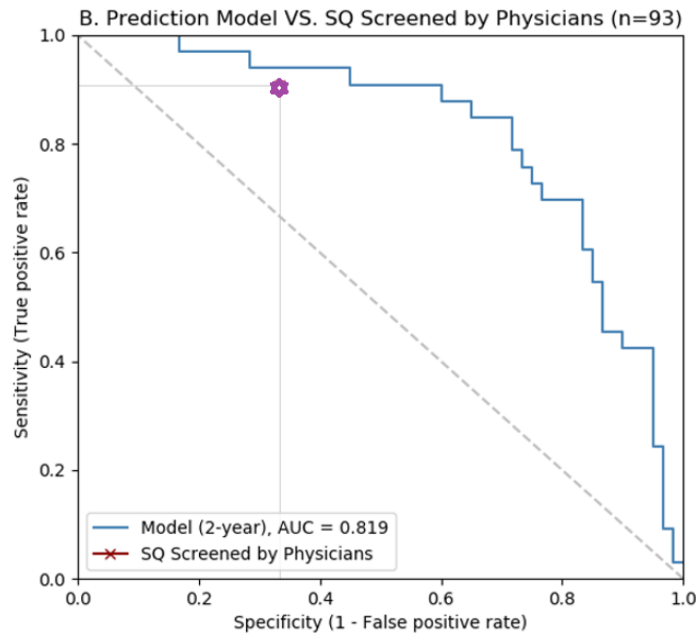


Wang L, Sha L, Lakin J, Bynum J, Bates DW, Hong P, Zhou L. JAMA Network Open, 2019.

Results

Our mortality prediction model performs better than clinician screening





Sensitivity: 0.674
Specificity: 0.792 (model) vs 0.536 (SQ)

Sensitivity: 0.909
Specificity: 0.525 (model) vs 0.330 (SQ)

MTERMS Applications

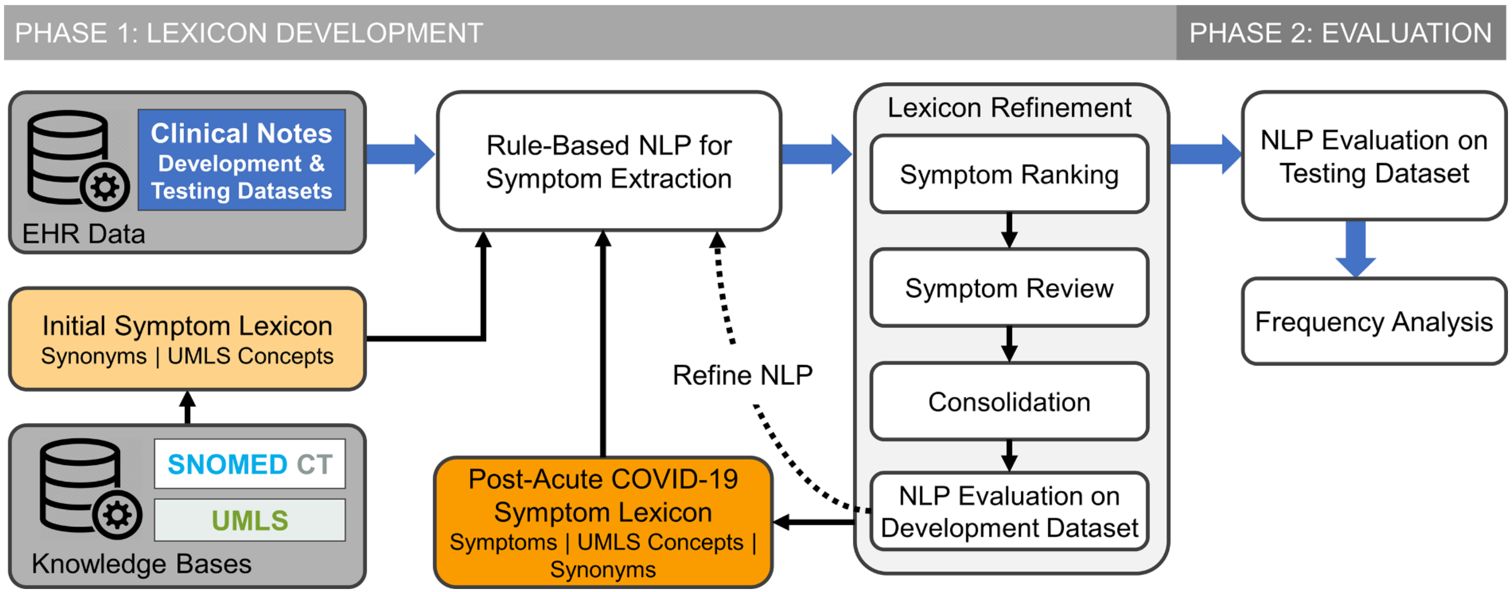
- Real-time pilots (integrated with Epic)
  - Allergy reconciliation module
    - Medication reconciliation module (in LMR)
  - Cancer Screening Follow-up (primary care)
  - Patient clinical deterioration based on nursing notes and EHR (inpatient)
- Near real-time
  - Patient mortality predication to improve palliative care intervention
- Research projects
  - Allergic and adverse reactions
  - Opioid use disorder patient identification
  - Gunshot intension classification
  - Malpractice cases (coding + similar cases)
  - Psychosis identification
  - Confounding factors for pharmacoepidemiology studies
  - Dementia/cognitive decline (bilingualism and cognitive reserve)
  - **PASCLex: Post-Acute Sequelae of COVID-19 (PASC) Symptom**
  - **Using Twitter data to understand public perceptions of approved vs. off-label user for COVID-19-related medications**
  - **Examination of Stigmatizing Language in EHRs**

PASCLex: Post-Acute Sequelae of COVID-19 (PASC) Symptom

- **PSAC syndrome or long COVID**: some patients have persistent symptoms and/or develop delayed or long-term complications after their recovery from acute COVID-19.
- Most early studies on PASC symptoms relied on patient survey data, manual chart, and in person follow-up.
  - Simple size, reporting bias
- Longitudinal EHR data serve as a rich data source for studying PASC symptoms
  - Structured data (lab results or diagnosis codes)
- NLP can automatically identify relevant symptoms and complications at different clinical stages from large volumes of longitudinal notes of a large patient cohort
- To capture wide variation in potential symptoms, a comprehensive lexicon encoded with a standard terminology is crucial for NLP tool development and utility and future EHR-based PASC analytics and research.

PASCLex: Methods

- Ontology-driven, EHR-guided and NLP-assisted approach
- PASC symptom lexicon was derived from 328,879 clinical notes of 26,177 COVID-19 patients documented between day 51-100 after their first positive COVID-19 test.



Wang L, Foer D, MacPhaul E, Lo YC, Bates DW, Zhou L. PASCLex: A Comprehensive Post-Acute Sequelae of COVID-19 (PASC) Symptom Lexicon Derived from Electronic Health Record Clinical Notes. Journal of Biomedical Informatics. 2021 Nov 13:103951. https://pubmed.ncbi.nlm.nih.gov/34785382/

PASCLex

PASCLex includes 355 symptoms (and 16,466 synonyms) consolidated from 1,520 Unified Medical Language System® (UMLS) concepts.

Selected examples of post-acute COVID-19 symptoms, consolidated Unified Medical Language System (UMLS) concepts, and synonyms from electronic health record clinical notes.

Symptoms	Consolidated UMLS concepts	Examples of Synonyms in Clinical Notes
Fatigue	C0015672:Fatigue, C0231218: Malaise, C0015674:Chronic Fatigue Syndrome, C0023380:Lethargy, C0392674: Exhaustion, C0024528:Malaise And Fatigue, C0424585:Tires Quickly, C0849970:Tired, C0439055:Tired All The Time, C2732413:Postexertional Fatigue, C3875100:Fatigue Due To Treatment, C4075947:Occasionally Tired	Fatigue, tiredness, malaise, tired, fatigued, lethargy, ill feeling, feeling unwell, feel tired, lethargic
Loss of appetite	C0003125:Anorexia Nervosa, C0232462:Decrease In Appetite, C0426587:Altered Appetite, C1971624:Loss Of Appetite, C0566582:Appetite Problem	Loss of appetite, decreased appetite poor appetite, appetite changes, change in appetite, decrease in appetite, appetite loss
Sleep apnea	C0037315:Sleep Apnea, C0003578:Apnea, C0020530:Hypersomnia With Sleep Apnea, C0751762:Primary Central Sleep Apnea, C1561861:Organic Sleep Apnea, C2732337:Sleep Hypoventilation	Sleep apnea, apneas, apnea, sleep disturbance, sleep disturbances, sleep problems, sleep disorder

The post-acute COVID-19 symptom lexicon can be accessed at: https://github.com/bylinn/Post_Acute_COVID19_Symptom_Lexicon.

Common PACS in clinical notes

50 most common post-acute COVID-19 patient symptoms in electronic health record clinical notes by symptom frequency, and corresponding precision of natural language processing (NLP) performance for unique symptom extraction.

Top 1–25 Symptoms	% frequency of symptoms	Precision	Top 26–50 Symptoms	% frequency of symptoms	Precision
Pain	43.1	0.94	Insomnia	11.2	0.94
Anxiety	25.8	0.98	Pain in extremities	10.7	1.0
Depression	24.0	0.90	Paresthesia	10.7	0.92
Fatigue	23.4	1.0	Peripheral edema	10.5	0.98
Joint pain	21.0	0.98	Palpitations	10.3	0.94
Shortness of breath	20.8	0.94	Diarrhea	10.3	0.92
Headache	20.0	0.92	Itching	9.4	0.92
Nausea and/or vomiting	19.9	1.0	Erythema	9.2	0.98
Myalgia	19.0	0.96	Lower urinary tract symptoms	8.7	0.98
Gastroesophageal reflux	18.6	0.94	Lymphadenopathy	8.3	0.96
Cough	17.5	0.92	Edema	7.9	0.88
Back pain	16.9	0.98	Weight gain	7.3	0.98
Stress	15.1	0.86	Sinonasal congestion	7.1	0.96
Fever	14.7	0.94	Pain in throat	6.4	0.98
Swelling	14.7	0.90	Abnormal gait	5.9	1.0
Bleeding	14.7	0.90	Respiratory distress	5.8	0.82
Weight loss*	14.2	0.98	Visual changes	5.8	0.92
Abdominal pain	14.1	0.98	Chills	5.6	0.86
Dizziness or vertigo	14.0	0.94	Urinary incontinence	5.6	0.96
Chest pain	12.5	0.90	Sleep apnea	5.4	0.94
Weakness	12.3	0.94	Confusion	5.4	0.98
Constipation	11.9	0.96	Hearing loss	5.2	1.0
Skin lesion	11.9	0.94	Problem with smell or taste	5.0	0.94
Wheezing	11.9	0.98	Difficulty swallowing	4.9	0.98
Rash	11.4	0.82	Loss of appetite	4.8	0.96

Journal of the American Medical Informatics Association, 29(10), 2022, 1668–1678
https://doi.org/10.1093/jamia/ocac114
Advance Access Publication Date: 18 July 2022
Research and Applications

Research and Applications

Using Twitter data to understand public perceptions of approved versus off-label use for COVID-19-related medications

Yining Hua[1,2], Hang Jiang[3], Shixu Lin[4], Jie Yang[4], Joseph M. Plasek[1,2], David W. Bates[1,2], and Li Zhou[1,2]

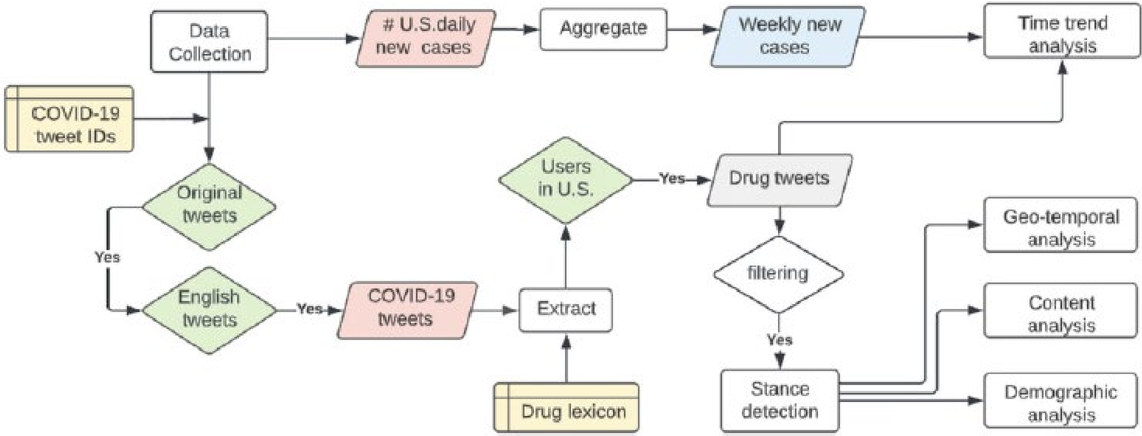


Figure 1. A comprehensive multimodal pipeline to study the public perception of drugs during the COVID-19 period.

ABSTRACT

Objective: Understanding public discourse on emergency use of unproven therapeutics is essential to monitor safe use and combat misinformation. We developed a natural language processing-based pipeline to understand public perceptions of and stances on coronavirus disease 2019 (COVID-19)-related drugs on Twitter across time.

Methods: This retrospective study included 609 189 US-based tweets between January 29, 2020 and November 30, 2021 on 4 drugs that gained wide public attention during the COVID-19 pandemic: (1) Hydroxychloroquine and Ivermectin, drug therapies with anecdotal evidence; and (2) Molnupiravir and Remdesivir, FDA-approved treatment options for eligible patients. Time-trend analysis was used to understand the popularity and related events. Content and demographic analyses were conducted to explore potential rationales of people's stances on each drug.

Results: Time-trend analysis revealed that Hydroxychloroquine and Ivermectin received much more discussion than Molnupiravir and Remdesivir, particularly during COVID-19 surges. Hydroxychloroquine and Ivermectin were highly politicized, related to conspiracy theories, hearsay, celebrity effects, etc. The distribution of stance between the 2 major US political parties was significantly different ($P < .001$); Republicans were much more likely to support Hydroxychloroquine (+55%) and Ivermectin (+30%) than Democrats. People with healthcare backgrounds tended to oppose Hydroxychloroquine (+7%) more than the general population; in contrast, the general population was more likely to support Ivermectin (+14%).

Conclusion: Our study found that social media users with have different perceptions and stances on off-label versus FDA-authorized drug use across different stages of COVID-19, indicating that health systems, regulatory agencies, and policymakers should design tailored strategies to monitor and reduce misinformation for promoting safe drug use. Our analysis pipeline and stance detection models are made public at https://github.com/ningkko/COVID-drug.

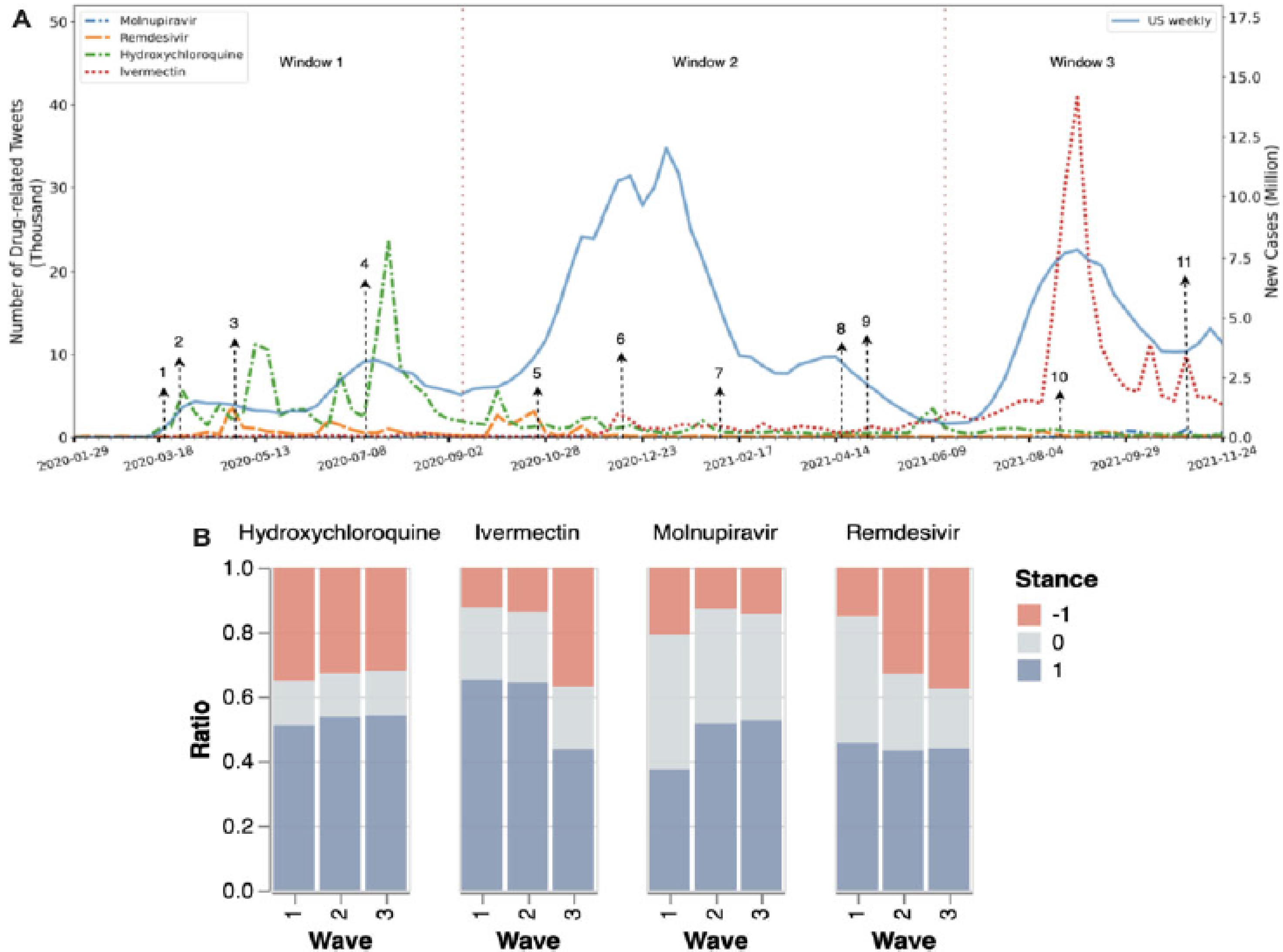


Figure 3. (A) The trends of (1) the number of tweets that mentioned COVID-19-related drugs: Hydroxychloroquine, Ivermectin, Molnupiravir, Remdesivir, and (2) weekly COVID-19 case counts (stepped line) in the United States. Wave boundaries are noted by dashed vertical lines. Major drug events are noted by numbers: (1) March 19, 2020: Trump declared Hydroxychloroquine a game-changer; (2) March 28, 2020: FDA approved a EUA to use Hydroxychloroquine for certain hospitalized patients; (3) May 1, 2020: FDA approved a EUA to use Remdesivir for severe patients; (4) July 15, 2020: FDA cautioned against the use of Hydroxychloroquine; (5) October 22, 2020: FDA approved Remdesivir for conditional use; (6) December 10, 2020: FDA cautioned against Ivermectin; (7) February 4, 2021: Merck cautioned against Ivermectin; (8) April 17, 2021: FDA clarified that Remdesivir was not approved; (9) May 1, 2021: FDA recalled a batch of Remdesivir vials, (10) August 21, 2021: FDA denounced Ivermectin as a COVID-19 treatment following an increase in overdoses; (11) November 4, 2021 Britain authorized Molnupiravir for COVID-19 treatment. (B) Distribution of percentage of tweets with positive (1, blue), neutral (0, gray), and negative (−1, red) stances for each drug. COVID-19: coronavirus disease 2019; EUA: Emergency Use Authorization FDA: US Food and Drug Administration.

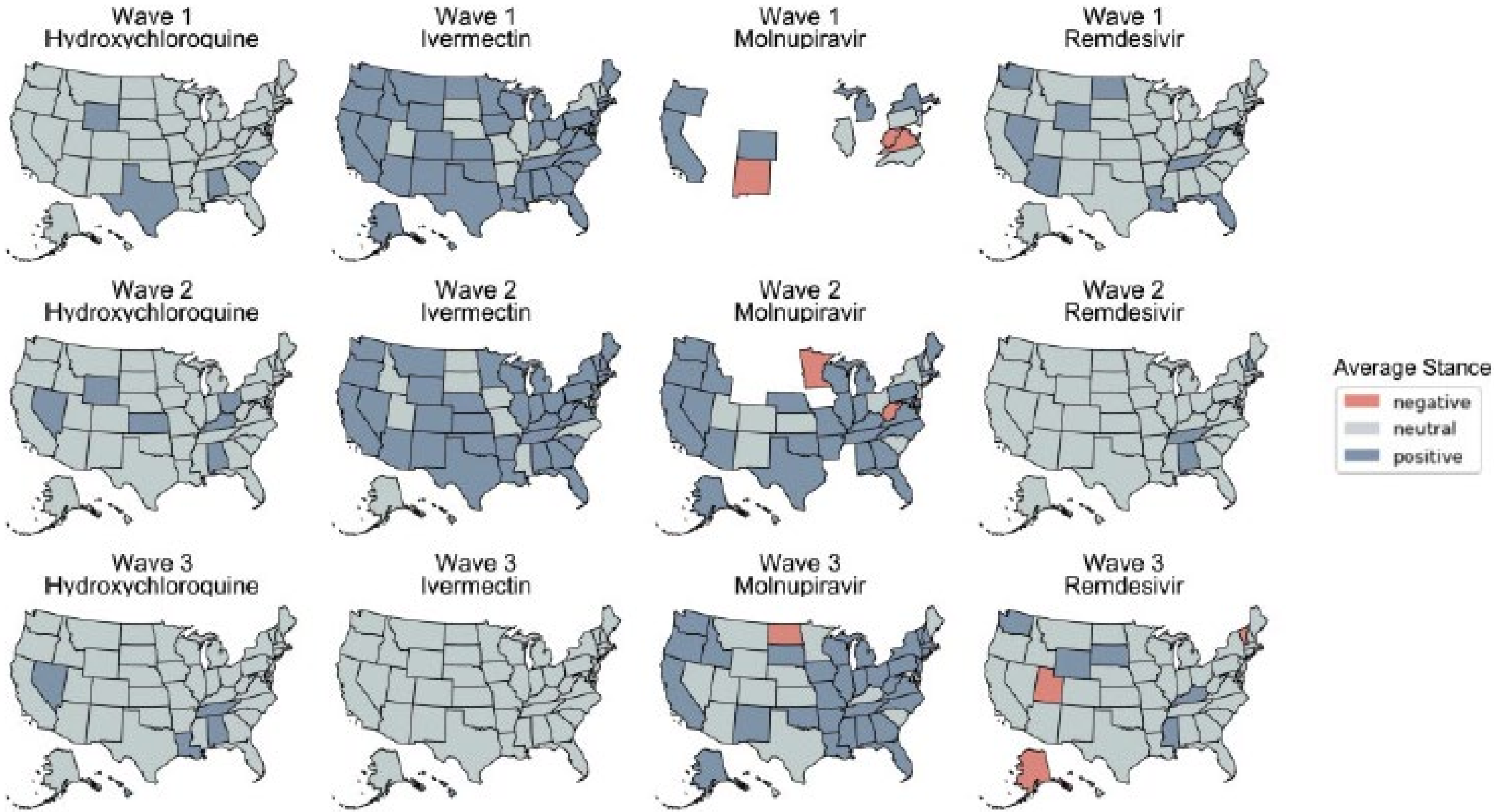


Figure 5. Longitudinal geo-temporal analysis of Tweeted sentiment of the 4 drugs by COVID-19 pandemic wave. The average sentiment of each state was classified into positive, neutral, and negative. COVID-19: coronavirus disease 2019.

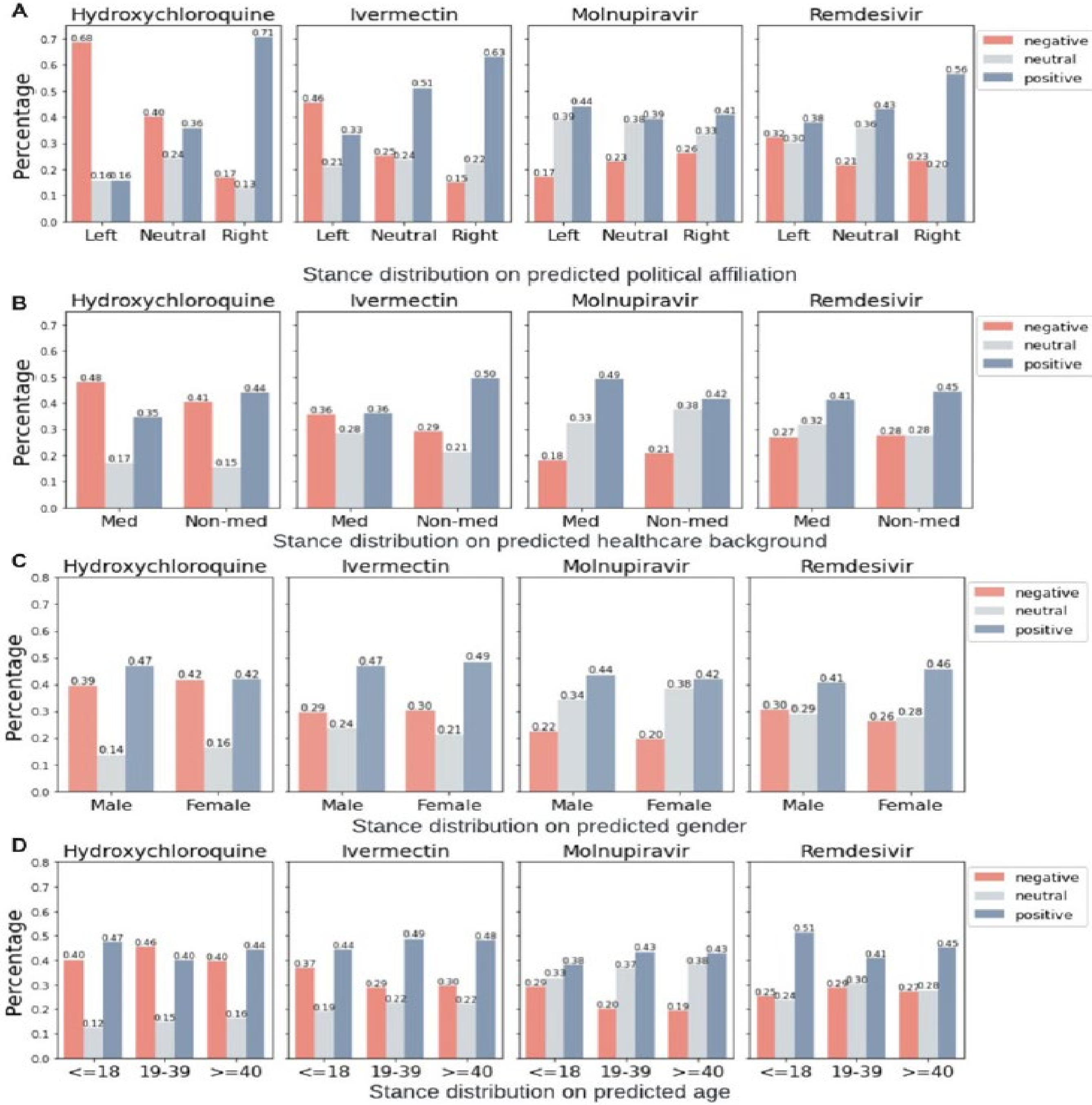


Figure 6. Stance distribution on predicted partisanship, age, and medical background for each drug. The exact numbers of tweets can be found in Supplementary Appendix SF.

Examination of Stigmatizing Language in EHR

Himmelstein G, Bates D, Zhou L. JAMA Netw Open. 2022 Jan PMID: 35084481. https://pubmed.ncbi.nlm.nih.gov/35084481/

- Stigmatizing language in the EHR may alter treatment plans, transmit biases between clinicians, and alienate patients.
  - However, neither the frequency of stigmatizing language, nor whether clinicians disproportionately use it in describing patients in particular demographic subgroups are known.
- **Question**: How frequently does stigmatizing language appear in the admission notes of patients who are hospitalized, and does the frequency vary by patients' medical conditions and race or ethnicity?
- Study Design:
  - Cross-sectional study of 48,651 admission notes about 29,783 unique patients by 1932 clinicians in 2018
  - Patients with diabetes, substance use disorder, or chronic pain
  - Patients' demographic characteristics: age, race and ethnicity, gender, and preferred language)
  - Clinicians' characteristics: gender, postgraduate year [PGY], and credential [physician vs advanced practice clinician]).

Examples of Stigmatizing Language

Table 2. Examples of Stigmatizing Language in Context, by Condition

Condition	Examples
Diabetes	Patient failed to show up to endocrine follow up
	Noncompliant with insulin regimen
	Patient refused diabetic diet
Substance use disorder	Started on opioids for pain control and admits to becoming addicted to them
	Avoid narcotics given history of abuse
	He is a habitual cocaine user
Chronic pain	Questionable if hyperalgesia or drug seeking behavior
	Patient has numerous psychiatric diagnoses including malingering
	Concern for secondary gain given narcotic seeking behavior

Examination of Stigmatizing Language in EHR

- Patient Characteristics
  - Age: 49 years; Female: 58.2%
  - 3.5% were non-Hispanic Asian, 8.4% non-Hispanic Black, 63.6% non-Hispanic White
  - 9.9% preferred a language other than English
- Of all admission notes, 2.5% contained stigmatizing language
- Stigmatizing language in hospital notes varied by medical condition and was more often used to describe non-Hispanic Black patients
  - Presented in 6.9% of notes for patients with diabetes, 3.4% for patients with substance use disorders, and 0.7% for patients with chronic pain.
  - Notes about non-Hispanic Black patients vs non-Hispanic White patients had a 0.67 (95%CI, 0.15 to 1.18) percentage points greater probability of containing stigmatizing language, with similar disparities in all 3 diagnosis-specific subgroups.
  - Greater diabetes severity and the physician-author being less advanced in their training was associated with more stigmatizing language.
- Training clinicians to minimize stigmatizing language in the EHR might improve patient-clinician relationships and reduce the transmission of bias between clinicians.

Agatha Christie

From Wikipedia, the free encyclopedia

For the video game series, see Agatha Christie (video game series).

Dame Agatha Mary Clarissa Christie, Lady Mallowan, DBE (*née* **Miller**; 15 September 1890 – 12 January 1976) was an English crime novelist, short story writer and playwright. She is best known for her 66 detective novels and 14 short story collections, particularly those revolving around her fictional detectives Hercule Poirot and Miss Marple. She also wrote the world's longest-running play, a murder mystery, *The Mousetrap*,[1] and six romances under the name **Mary Westmacott**. In 1971 she was made a Dame for her contribution to literature.[2]

Christie was born into a wealthy upper-middle-class family in Torquay, Devon. She served in a Devon hospital during the First World War, tending to troops coming back from the trenches, before marrying and starting a family in London. She was initially an unsuccessful writer with six rejections,[3] but this changed when *The Mysterious Affair at Styles*, featuring Hercule Poirot, was published in 1920.[4] During the Second World War she worked as a pharmacy assistant at University College Hospital, London, during the Blitz and acquired a good knowledge of poisons which featured in many of her novels.

Guinness World Records lists Christie as the best-selling novelist of all time. Her novels have sold roughly 2 billion copies, and her estate claims that her works come third in the rankings of the world's most-widely published books,[5] behind only Shakespeare's works and the Bible. According to Index Translationum, she remains the most-translated individual author – having been translated into at least 103 languages.[6] *And Then There Were None* is Christie's best-selling novel, with 100 million sales to date, making it the world's best-selling mystery ever, and one of the best-selling books of all time.[7] Christie's stage play *The Mousetrap* holds the world record for longest initial run. It opened at the

Dame Agatha Christie
Lady Mallowan
DBE

Born	Agatha Mary Clarissa Miller 15 September 1890 Torquay, Devon, England
Died	12 January 1976 (aged 85) Winterbrook, Oxfordshire, England

Christie is the best-selling novelist of all time. Her works come third in the rankings of the world's most-widely published books, behind only Shakespeare's works and the Bible.

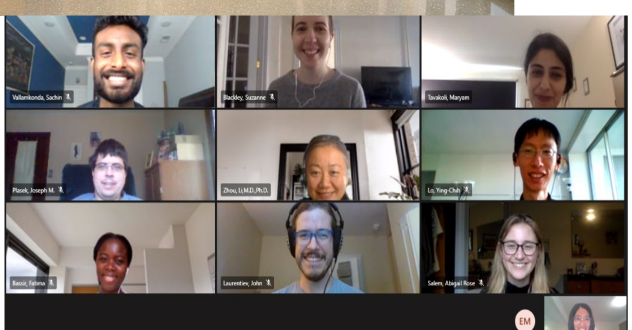
Text Analysis Claims Agatha Christie Had Alzheimer’s

- Compared a selection of Christie’s novels between the ages of 28 and 82, counting numbers of different words, indefinite nouns and phrases used in each
- Statistically significant drops in vocabulary and increases in repeated phrases and indefinite nouns
  - A book she wrote aged 81 showed 30% fewer word types than another book she wrote aged 63, 18% more repeated phrases, and almost three times as many indefinite words
- These language effects are recognized as symptoms of memory difficulties associated with Alzheimer's disease

What's Next?

- Clinical Decision Support
  - Retrieve relevant information, similar cases
  - Answer clinical/research questions
  - Identify high-risk, high-cost patients prospectively
- Enhancing EHR functions
  - Advanced search, spelling error correction, auto-fill, etc.
  - Improve clinical documentation and identify incomplete or inconsistent information
- Text Summarization and Generation
  - Summarize/generate a note, a specific condition, or the whole record…
- Speech Recognition
  - Further improve usability and integration with clinical workflow and the EHR
- Language and Diseases
- Multimodal data
- Others, e.g. computer-assisted coding

MTERMS Team



- <u>Best Paper Awards</u>
  - -Medinfo 2013
  - -Medinfo 2015 (nomination)
  - -AcademyHealth's Annual Research meeting 2017
  - - AMIA 2018, 2020
- <u>Competition Awards</u>
  - -IEEE Computer Society Student Global Challenge, finalist (2017)
  - -"Idea Lab with Microsoft: Machine Learning in Healthcare," BWH Innovation Hub, 1st Prize (2016)
  - -Health 2.0 Developer Challenge, Veterans Health Administration Care Coordination for Improved Outcomes Challenge, 2nd Place (2015)
- <u>Institutional Awards</u>
  - -Partners Excellence Award (in Leadership & Innovation, 2016,2018; in Team, Quality Treatment and Service, 2017)
  - -Innovation Discovery Award
- <u>Studies Reported by Media Outlet</u>

  The New York Times, The Wall Street Journal, CBS News, Fox News, Reuters, NPR, Physician's Weekly etc.

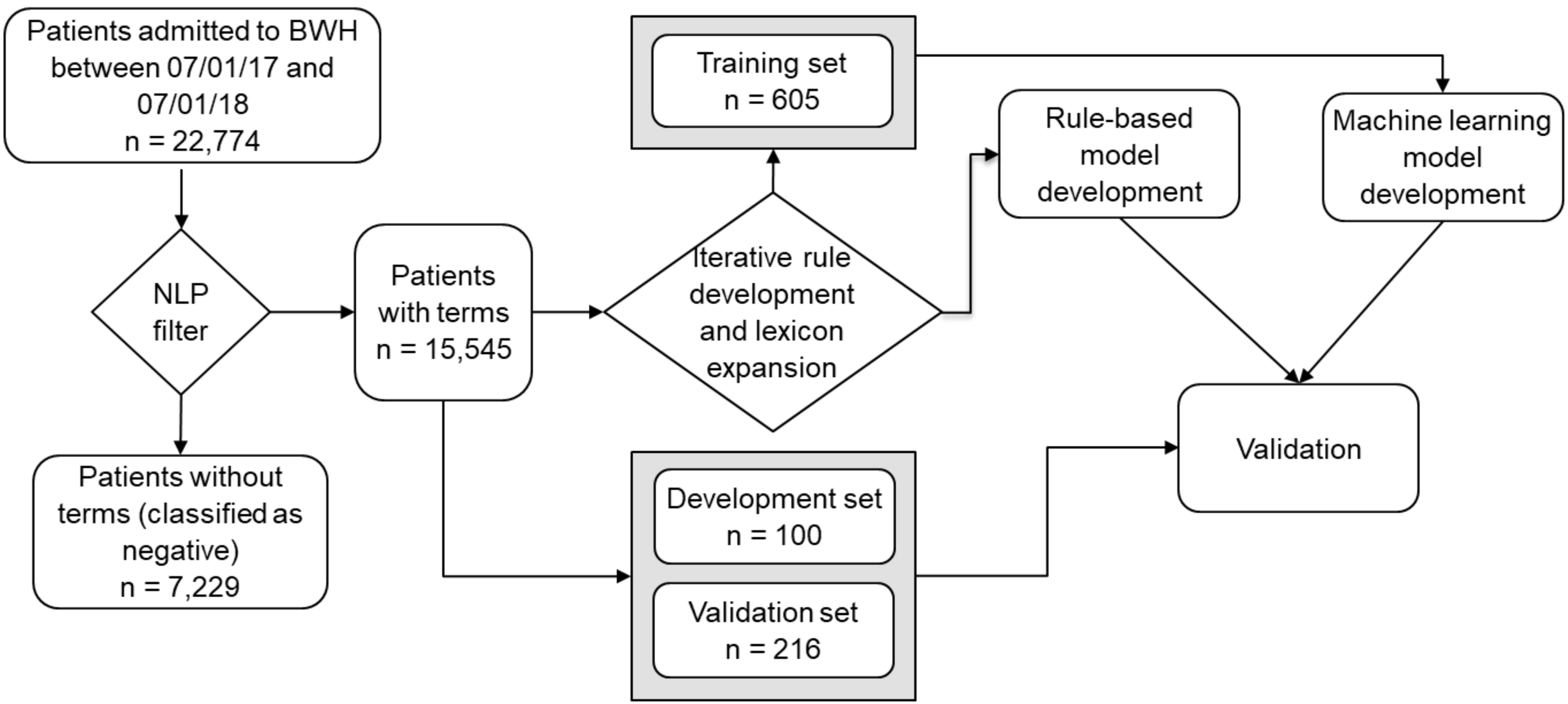
Additional Materials

NLP and Machine Learning to Identify Hospitalized Patients with OUD

- Misuse of prescription opioids is a major public health crisis
  - 0.3 million people in 2018
  - 130+ died every day from opioid overdoses
- 19% of patients who recognized a need for treatment did not receive help due to a lack of information on where to go
- Hospital encounters provide an opportunity for clinicians to offer effective interventions; however,
  - Existing hospital screening methods are resource-intensive
  - Many methods rely on structured EHR data (e.g. billing codes, prescription data, lab values), but these data may be incomplete
  - Substance use is often documented in free-text (e.g., ED notes, inpt progress notes)
- Automated methods are needed to identify potential patients and alert the provider and addiction services simultaneously, so a patient's care team can immediately present them with treatment options

Blackley SV, MacPhaul E, Martin B, Song W, Suzuki J, Zhou L. Using Natural Language Processing and Machine Learning to Identify Hospitalized Patients with Opioid Use Disorder. *AMIA Annu Symp Proc*. 2020. (Distinguished paper award)

Methods



Patients admitted to BWH between 07/01/17 and 07/01/18 n = 22,774
NLP filter
Patients without terms (classified as negative) n = 7,229
Patients with terms n = 15,545
Iterative rule development and lexicon expansion
Training set n = 605
Development set n = 100
Validation set n = 216
Rule-based model development
Machine learning model development
Validation

Results

		Rules	Machine learning models				DNN
			LR	SVM	KNN	RF	
Development set	Precision	0.9400	0.9601	0.9601	0.9033	0.9388	0.9167
	Recall	0.9420	0.9592	0.9592	0.8980	0.9388	0.8684
	F1 score	0.9399	0.9592	0.9592	0.8972	0.9388	0.8919
	Accuracy	0.9495	0.9592	0.9592	0.8980	0.9388	–
Validation set	Precision	0.9324	0.9654	**0.9730**	**0.9730**	**0.9730**	0.5000
	Recall	0.8052	0.9676	**0.9722**	**0.9722**	**0.9722**	0.6667
	F1 score	0.8563	0.9639	**0.9683**	**0.9683**	**0.9683**	0.5714
	Accuracy	0.9722	0.9676	0.9676	0.9676	**0.9722**	–

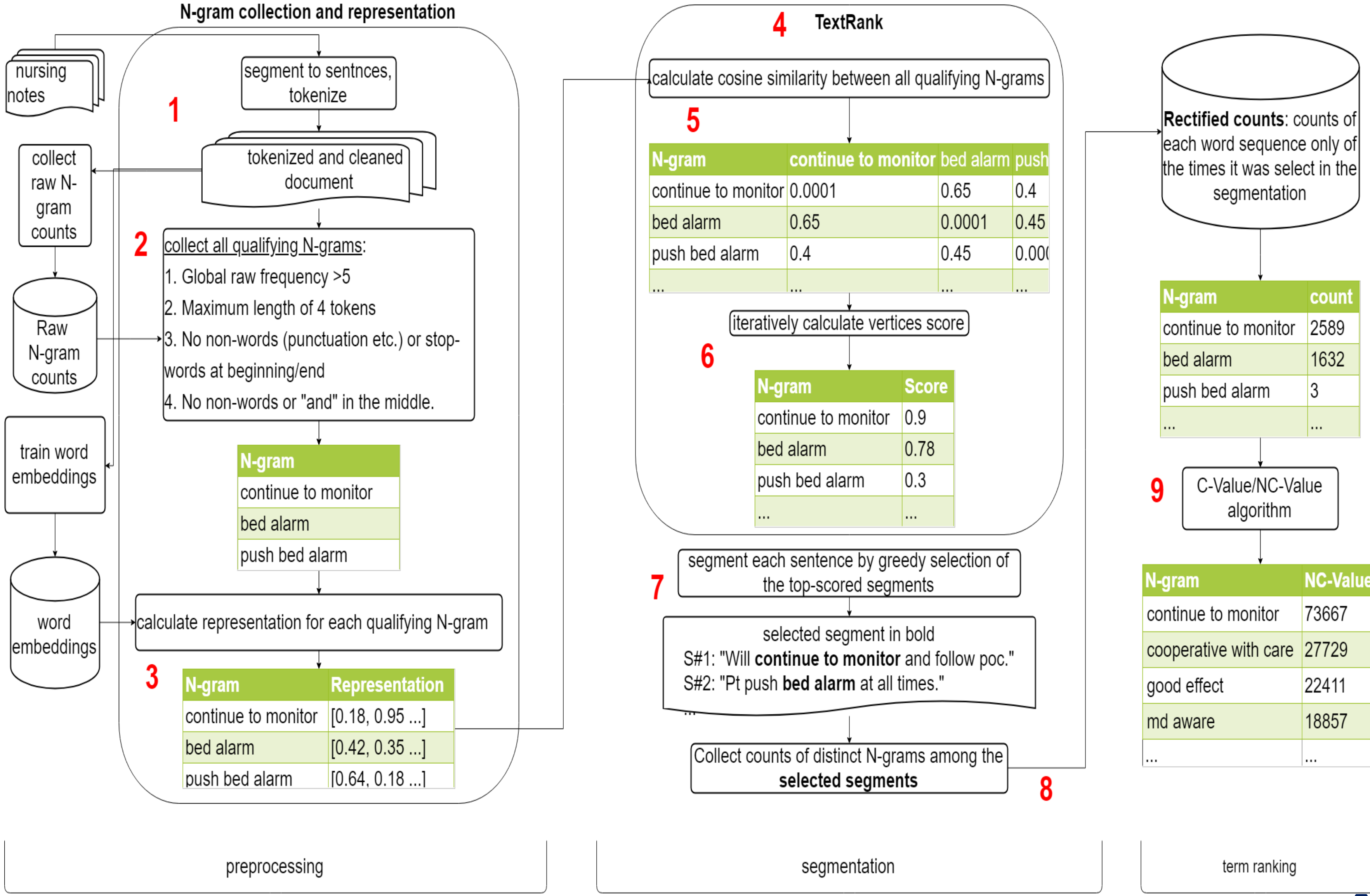
LR: logistic regression
SVM: support vector machine
KNN: k-nearest neighbors
RF: random forest
DNN: deep neural network

Mining nurses' "concerns" to predict patient deterioration

- Hospital rapid response (RR) teams respond to non–ICU patients at risk for rapid deterioration
  - Existing approaches to RR detection mainly focus on structured information
- RR teams often rely on nurses' clinical judgement typically documented narratively
  - "Staff member is worried about the patient"
- Develop a data-driven method to facilitate feature engineering of nursing notes for knowledge discovery and risk prediction
- Use Quality Phrase Mining to recognize nursing concern concept/phrases

Korach TZ, Zhou L. et al. Unsupervised machine-learning of topics documented by nurses about hospitalized patients prior to a rapid-response event. Appl Clin Inform. 2019 Oct;10(5):952-963.
Korach TZ, Zhou L. et al. Mining clinical phrases from nursing notes to discover risk factors of patient deterioration. Int J Med Inform. 2019 Dec 14;135:104053.

Phrase Mining Process



Predictive Value

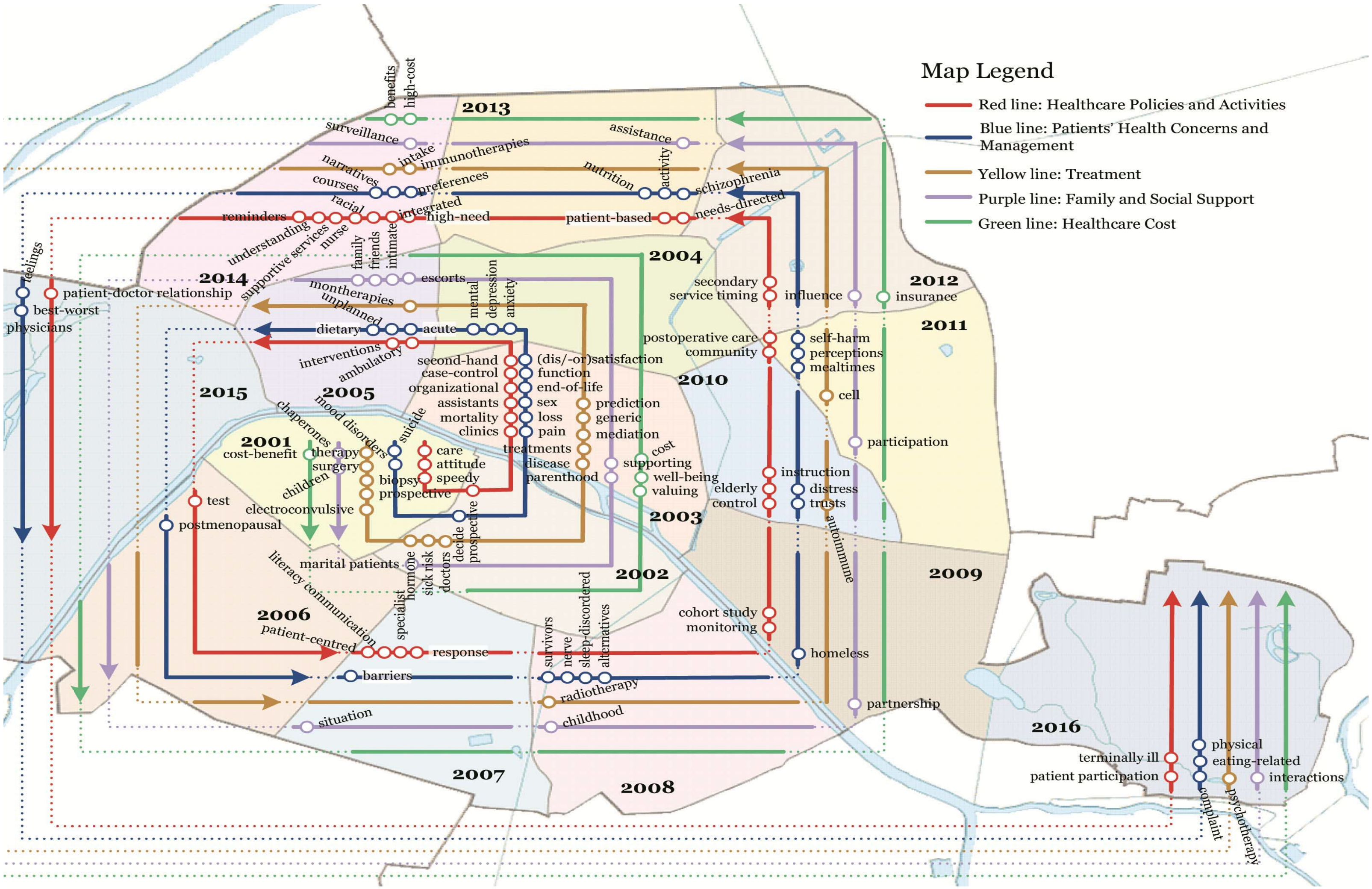
- Assessed the phrases' quality as features for predictive and explanatory modelling
- Time-dependent covariates Cox model using the phrases achieved a concordance index of 0.739.
- Clustering the phrases revealed clinical concepts significantly associated with Rapid Response (RR) event hazard.

Results

Phrase clusters and their association with the risk of rapid-response event. A positive coefficient increases the hazard for RR event and vice versa.

Effect	Cluster name	Hazard ratio (95 % confidence interval)	p-value	adjusted p-value	Example phrases
Increases hazard	patient decision**	1.68 (1.28–2.20)	< .001	< .001	pt aware, pt refused
	auxiliary tests**	1.63 (1.38–1.93)	< .001	< .001	am labs, ekg obtained
	cough**	1.47 (1.32–1.65)	< .001	< .001	non productive cough, productive cough
	emotional support**	1.44 (1.27–1.63)	< .001	< .001	emotional support given
	abdominal examination*	1.39 (1.04–1.85)	0.025	0.8	abd softly distended, non tender
	nurse's shift transfer*	1.38 (1.06–1.78)	0.015	0.49	shift assessment, shift summary
	patient verbal communication	1.24 (0.93–1.64)	0.14	> .99	patient reports, patient states
	MD awareness*	1.21 (1.02–1.43)	0.026	0.83	md made aware, md notified, md paged
	care management	1.21 (0.62–2.33)	0.58	> .99	review homecare, home care services
	fasting*	1.19 (1.01–1.39)	0.038	> .99	npo since midnight, remains npo, ivf infusion
	patient alarm*	1.15 (1.04–1.27)	< .001	0.27	bed alarm, call bell
	Intravenous medications	1.09 (0.91–1.29)	0.35	> .99	iv antibiotics, iv dilaudid
	aspiration	1.07 (0.93–1.23)	0.37	> .99	aspiration precautions maintained, infection afebrile
	Pre/post operation	1.05 (0.97–1.15)	0.22	> .99	post op, pre op
	dressing	1.03 (0.88–1.21)	0.72	> .99	primary dressing, wound vac
	assessment	1.03 (0.77–1.37)	0.86	> .99	team made aware, continue to monitor
	sex (female)	1.02 (0.91–1.16)	0.7	> .99	N/A
	age**	1.02 (1.02 to 1.02)	< .001	< .001	N/A
	skin barrier	1.01 (0.86–1.20)	0.87	> .99	barrier cream, open to air
No effect	Note's calendar hour	1 (0.99–1.01)	0.65	> .99	N/A
Decreases hazard	Intravenous catheter	0.95 (0.86–1.04)	0.28	> .99	blood return, picc line
	following commands	0.91 (0.82–1.01)	0.071	> .99	follows commands
	diet tolerance	0.86 (0.72–1.03)	0.1	> .99	tolerating house diet, takes pills whole
	completed tests*	0.86 (0.76 to 0.97)	0.014	0.43	cxr completed, echo completed
	vital signs*	0.85 (0.74 to 0.98)	0.024	0.76	hemodynamically stable, neuro exam stable
	risk criteria*	0.79 (0.65 to 0.97)	0.023	0.73	high risk criteria, initial assessment
	wound care	0.76 (0.56–1.03)	0.08	> .99	ace wrap, dressing cdi
	Pain*	0.75 (0.63 to 0.91)	< .001	0.091	abd pain, atc Tylenol, oxycodone prn
	Ambulation*	0.75 (0.59 to 0.95)	0.016	0.52	able to ambulate, oob to chair, stand by assist
	communication	0.73 (0.53–1.01)	0.054	> .99	family at bedside, responsible person
	plan of care	0.45 (0.18–1.18)	0.1	> .99	reassessment chart, pt consulted
	bowel movement**	0.29 (0.20 to 0.41)	< .001	< .001	bm overnight, bm this shift

A Map Tracking Patient Needs from 339 Medline Publications between 2001-2016



Tang C, Zhang H, Lai KH, She Y, Xiong Y, Zhou L. Developing a regional classifier to track patient needs in medical literature using spiral timelines on a geographical map. *In 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 2017 Nov 1 (pp. 874-879). IEEE.