

# *Information Retrieval for Medical Applications*

Byron C Wallace

b.wallace@northeastern.edu // byronwallace.com // @byron\_c\_wallace

# Some disclaimers

- In this talk I'll cover some of the broad areas (through the lens of data sources) in medical IR / NLP

# Some disclaimers

- In this talk I'll cover some of the broad areas (through the lens of data sources) in medical IR / NLP
- This is a *completely biased non-exhaustive* sampling, and like any good academic I'll mostly be talking about **my** work in detail (but I still hope it provides a sense of the space)

# Some disclaimers

- In this talk I'll cover some of the broad areas (through the lens of data sources) in medical IR / NLP
- This is a *completely biased non-exhaustive* sampling, and like any good academic I'll mostly be talking about **my** work in detail (but I still hope it provides a sense of the space)
- Forgive me if this seems disorganized or rambling; COVID has me a bit fuzzy (and kept me from being there in person!)

# Sources of data

PubMed



Biomedical literature



OhHeyGrrr!  
@ohheygrrr!

My #COVID19 symptoms update: still congested, sore throat and coughing. The only things I can handle ingesting are buttered white toast and orange juice. My son is doing regular 'welfare checks' on me by pounding on the door and bellowing "WELFARE CHECK!" 😂

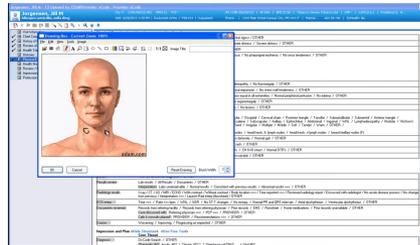
4:07 AM · Jul 11, 2022 · Twitter for iPhone

Social media

MEDITECH

Cerner

Epic



Electronic Health Records (EHRs)

# Sources of data

PubMed



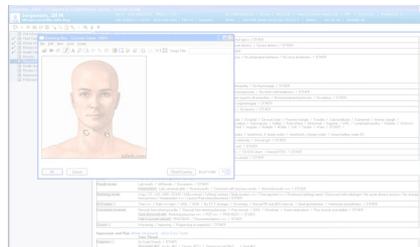
Biomedical literature



Social media

MEDITECH

Cerner  
Epic



Electronic Health Records (EHRs)

# PubMed

diabetes - Search Results - PubMed

pubmed.ncbi.nlm.nih.gov/?term=diabetes

NIH National Library of Medicine  
National Center for Biotechnology Information

byron.wallace@gmail...

PubMed.gov

diabetes

Search

Advanced Create alert Create RSS User Guide

Save Email Send to Sorted by: Best match Display options

MY NCBI FILTERS

RESULTS BY YEAR

1788 2023

TEXT AVAILABILITY

Abstract

Free full text

Full text

ARTICLE ATTRIBUTE

880,965 results

Page 1 of 88,097

**Diabetes insipidus and pregnancy.**

1 Chanson P, Salenave S.  
Cite Ann Endocrinol (Paris). 2016 Jun;77(2):135-8. doi: 10.1016/j.ando.2016.04.005. Epub 2016 May 9. PMID: 27172867 Review.  
Share **Diabetes insipidus** (DI) is a rare complication of pregnancy. It is usually transient, being due to increased placental production of vasopressinase that inactivates circulating vasopressin. ...  
Paperpile

**Gestational diabetes insipidus: a review of an underdiagnosed condition.**

2 Aleksandrov N, Audibert F, Bedard MJ, Mahone M, Goffinet F, Kadoch IJ.  
Cite J Obstet Gynaecol Can. 2010 Mar;32(3):225-31. doi: 10.1016/s1701-2163(16)34448-6. PMID: 20500966 Review.  
Share STUDY SELECTION: We reviewed 50 studies selected using the following key words: **diabetes insipidus**, pregnancy, arginine vasopressin, vasopressinase. CONCLUSION: Gestational **diabetes insipidus** is underdiagnosed because polyuria is often considered norma ...  
Paperpile

**History of Diabetes Insipidus.**

# PubMed

- ❑ Supports Boolean queries

# PubMed

- ❑ Supports Boolean queries

Medical Genetics	((Diagnosis AND genetics) OR (Differential Diagnosis[MeSH] OR Differential Diagnosis[Text Word] AND genetics) OR (Natural History OR Mortality OR Phenotype OR Prevalence OR Penetrance AND genetics) OR (therapy[Subheading] OR treatment[Text Word] OR treatment outcome OR investigational therapies AND genetics) OR (Genetic Counseling OR Inheritance pattern AND genetics) OR (Medical Genetics OR genotype OR genetics[Subheading] AND genetics) OR (DNA Mutational Analysis OR Laboratory techniques and procedures OR Genetic Markers OR diagnosis OR testing OR test OR screening OR mutagenicity tests OR genetic techniques OR molecular diagnostic techniques AND genetics))
------------------	--

# PubMed

- ❑ Supports Boolean queries
- ❑ Can search using controlled vocab (MeSH) terms
  - Normalizes for variants
  - MeSH terms are *manually* applied

# PubMed

- ❑ Supports Boolean queries
- ❑ Can search using controlled vocab (MeSH) terms
  - Normalizes for variants
  - MeSH terms are *manually* applied
- ❑ Provides “similar article” functionality

# PubMed

- ❑ Supports Boolean queries
- ❑ Can search using controlled vocab (MeSH) terms
  - Normalizes for variants
  - MeSH terms are *manually* applied
- ❑ Provides “similar article” functionality

## Computation of similar articles

The similarity between two documents is computed by adding up the weights (local wt1 × local wt2 × global wt) of all of the terms the two documents have in common. This provides an indication of how related two documents are. The resultant score is an example of a vector score. Vector scoring was originated by Gerard Salton and has a long history in text retrieval. **The interested reader is referred to Salton, Automatic Text Processing, Reading, MA: Addison-Wesley, 1989** for further

<https://pubmed.ncbi.nlm.nih.gov/help/#computation-of-similar-articles>

# Beyond PubMed



Database, Vol. 2011, Article ID baq036, doi:10.1093/database/baq036

## Review

# PubMed and beyond: a survey of web tools for searching biomedical literature

Zhiyong Lu\*

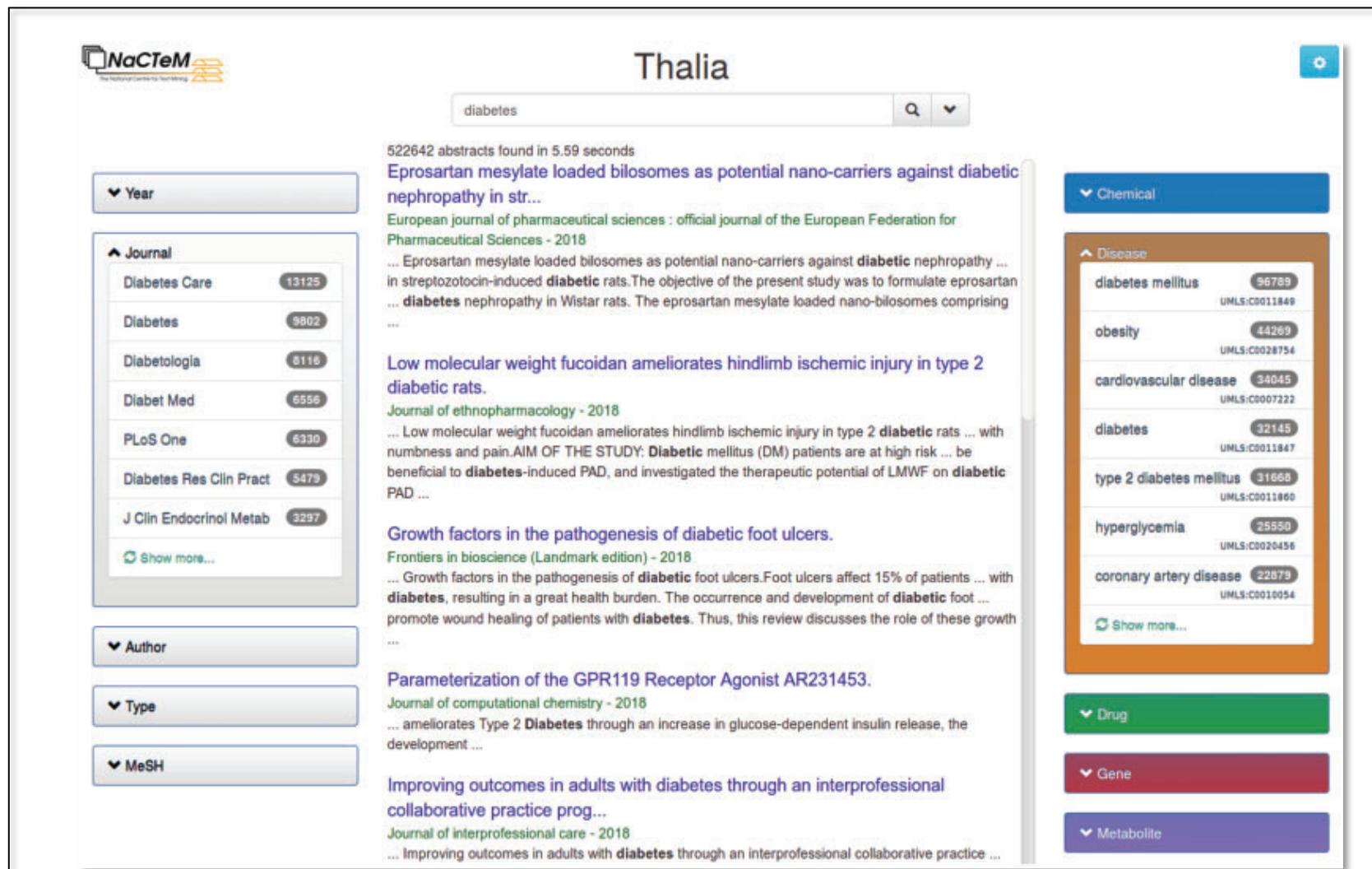
National Center for Biotechnology Information (NCBI), National Library of Medicine, Bethesda, MD 20894, USA

# Thalia: semantic search engine for biomedical abstracts

Axel J Soto, Piotr Przybyła, Sophia Ananiadou 

Bioinformatics, Volume 35, Issue 10, 15 May 2019, Pages 1799–1801,

<https://doi.org/10.1093/bioinformatics/bty871>



The screenshot displays the Thalia search engine interface. At the top left is the NaCTeM logo. The main header features the word "Thalia" and a search bar containing the query "diabetes". Below the search bar, it indicates "522642 abstracts found in 5.59 seconds".

On the left side, there are several filter panels:

- Year:** A dropdown menu.
- Journal:** A list of journals with their respective abstract counts:
  - Diabetes Care: 13125
  - Diabetes: 9802
  - Diabetologia: 8116
  - Diabet Med: 6556
  - PLoS One: 6330
  - Diabetes Res Clin Pract: 5479
  - J Clin Endocrinol Metab: 3297A "Show more..." link is also present.
- Author:** A dropdown menu.
- Type:** A dropdown menu.
- MeSH:** A dropdown menu.

The main content area displays three search results:

- Eprosartan mesylate loaded bilosomes as potential nano-carriers against diabetic nephropathy in str...**  
European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences - 2018  
... Eprosartan mesylate loaded bilosomes as potential nano-carriers against **diabetic** nephropathy ... in streptozotocin-induced **diabetic** rats. The objective of the present study was to formulate eprosartan ... **diabetes** nephropathy in Wistar rats. The eprosartan mesylate loaded nano-bilosomes comprising ...
- Low molecular weight fucoidan ameliorates hindlimb ischemic injury in type 2 diabetic rats.**  
Journal of ethnopharmacology - 2018  
... Low molecular weight fucoidan ameliorates hindlimb ischemic injury in type 2 **diabetic** rats ... with numbness and pain. AIM OF THE STUDY: **Diabetic** mellitus (DM) patients are at high risk ... be beneficial to **diabetes**-induced PAD, and investigated the therapeutic potential of LMWF on **diabetic** PAD ...
- Growth factors in the pathogenesis of diabetic foot ulcers.**  
Frontiers in bioscience (Landmark edition) - 2018  
... Growth factors in the pathogenesis of **diabetic** foot ulcers. Foot ulcers affect 15% of patients ... with **diabetes**, resulting in a great health burden. The occurrence and development of **diabetic** foot ... promote wound healing of patients with **diabetes**. Thus, this review discusses the role of these growth ...

On the right side, there are two filter panels:

- Chemical:** A dropdown menu.
- Disease:** A list of diseases with their respective abstract counts:
  - diabetes mellitus: 96789 (UMLS:C0011849)
  - obesity: 44269 (UMLS:C0028754)
  - cardiovascular disease: 34045 (UMLS:C0007222)
  - diabetes: 32145 (UMLS:C0011847)
  - type 2 diabetes mellitus: 31665 (UMLS:C0011860)
  - hyperglycemia: 25550 (UMLS:C0020456)
  - coronary artery disease: 22879 (UMLS:C0010054)A "Show more..." link is also present.

At the bottom right, there are three more filter panels:

- Drug:** A dropdown menu.
- Gene:** A dropdown menu.
- Metabolite:** A dropdown menu.

# Interactive Extractive Search over Biomedical Corpora

Hillel Taub Tabib, Micah Shlain, Shoval Sadde, Dan Lahav, Matan Eyal, Yaara Cohen, Yoav Goldberg



2020

<https://spike.apps.allenai.org/datasets/pubmed>

The screenshot displays the SPIKE Search over PubMed interface. The search query is "coronavirus lemma=treatment". The results are shown in a list of documents, with the first few entries highlighted. On the right side, there is an "Aggregations by Document" section showing a counts table for the "md:journal\_title" field. The table lists various journals and their corresponding document counts.

**Search Query:** coronavirus lemma=treatment

**Case Sensitivity:** Inensitive

**Sub-queries:** Sentence, Paragraph, Document, Section, Title, Journal\_title, Pmid, Year, Mesh\_list

**Matches:**

- Presentation of a case illustrating the benefits of traditional Chinese medicine ( TCM ) for **treatment** of **Coronavirus** disease 2019 ( COVID-19 ) in critically ill patients .
- To explore the pharmacologically active ingredients in Toujie Quwen granules ( TJQW ) for **treatment** of **coronavirus** disease 2019 ( COVID-19 ) in light of systemic pharmacology .
- SARS-CoV-2 ( Severe Acute Respiratory Syndrome **Coronavirus** 2)-related pneumonia , referred to as COVID-19 ( **Coronavirus** Disease 19 ) , is a public health emergency as it carries high morbidity , mortality , and has no approved specific pharmacological **treatments** .
- An overview of the preclinical discovery and development of bamlanivimab for the **treatment** of novel **coronavirus** infection ( COVID-19 ): reasons for limited clinical use and lessons for the future .
- Feline **coronavirus** ( FCoV ) causes the fatal disease feline infectious peritonitis , which is currently incurable by drug **treatment** , and no effective vaccines are available .

**Aggregations by Document**

Counts Table (2268 documents)

md:journal_title	Counts
REFERENCES	137
Medicine	34
Zhonghua wei zhong bing	31
Clinical infectious disease	28
Frontiers in immunology	20
PloS one	20
Scientific reports	20
Viruses	19
European review for medi	18
Journal of medical virolog	18
International journal of inf	17
Zhongguo Zhong yao za :	17
International journal of m	16

Showing 2485 matches. All Results have been downloaded

# Interactive Extractive Search over Biomedical Corpora

Hillel Taub Tabib, Micah Shlain, Shoval Sadde, Dan Lahav, Matan Eyal, Yaara Cohen, Yoav Goldberg

 ACL Anthology

2020

<https://spike.apps.allenai.org/datasets/pubmed>

**AI2** Allen Institute for AI



## SPIKE: Search over pubmed

### Example queries:

- Boolean: `coronavirus lemma=treatment` **RUN** (Sentences with specific terms.)
- Token: `incubation period ... from:* -|to to:* days` **RUN** (Extract incubation periods.)
- Structured: `<>v:virus $infection $causes a <>c:condition .` **RUN** (Things caused by kinds of infections.)

# Interactive Extractive Search over Biomedical Corpora

Hillel Taub Tabib, Micah Shlain, Shoval Sadde, Dan Lahav, Matan Eyal, Yaara Cohen, Yoav Goldberg



ACL Anthology

2020

<https://spike.apps.allenai.org/datasets/pubmed>

**AI2** Allen Institute for AI



## SPIKE: Search over pubmed

▼ There are many search systems over PubMed. What is special about this one?

- Ours is more complex and harder to use!

# Interactive Extractive Search over Biomedical Corpora

Hillel Taub Tabib, Micah Shlain, Shoval Sadde, Dan Lahav, Matan Eyal, Yaara Cohen, Yoav Goldberg



ACL Anthology

2020

<https://spike.apps.allenai.org/datasets/pubmed>

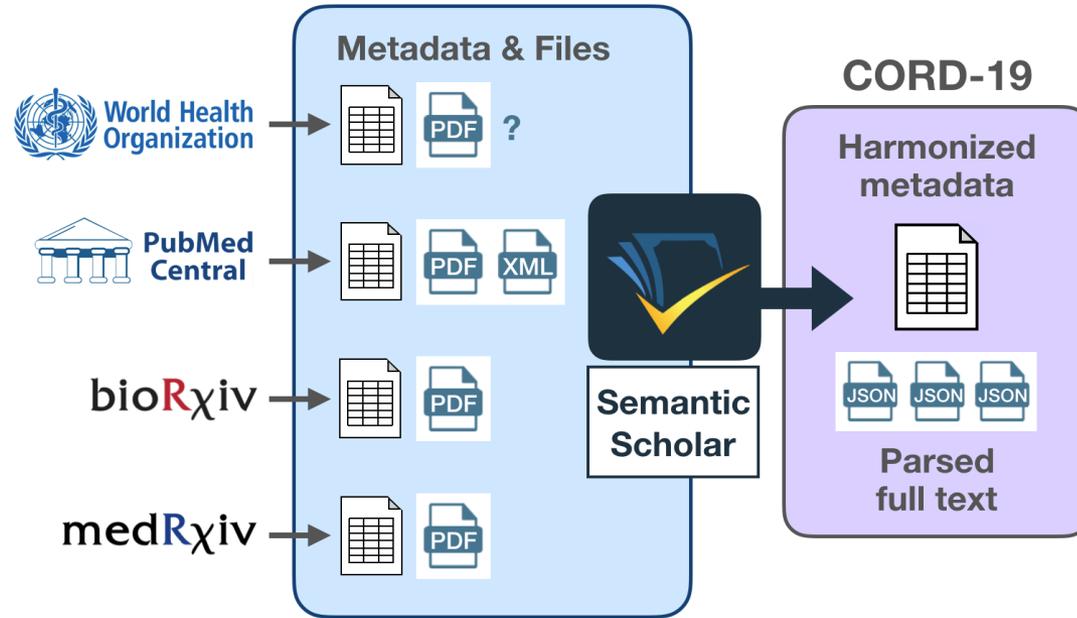
**AI2** Allen Institute for AI



## SPIKE: Search over pubmed

▼ **There are many search systems over PubMed. What is special about this one?**

- Ours is more complex and harder to use!
- However, it is also much more powerful.
- We are not competing with other systems – we aim to be complementary to them.



## CORD-19: The COVID-19 Open Research Dataset

Lucy Lu Wang<sup>1,\*</sup> Kyle Lo<sup>1,\*</sup> Yoganand Chandrasekhar<sup>1</sup> Russell Reas<sup>1</sup>

Jiangjiang Yang<sup>1</sup> Darrin Eide<sup>2</sup> Kathryn Funk<sup>3</sup> Rodney Kinney<sup>1</sup>

Ziyang Liu<sup>5</sup> William Merrill<sup>1</sup> Paul Mooney<sup>4</sup> Dewey Murdick<sup>6</sup>

Devvret Rishi<sup>4</sup> Jerry Sheehan<sup>3</sup> Zhihong Shen<sup>2</sup> Brandon Stilson<sup>1</sup>

Alex D. Wade<sup>5</sup> Kuansan Wang<sup>2</sup> Chris Wilhelm<sup>1</sup> Boya Xie<sup>2</sup>

Douglas Raymond<sup>1</sup> Daniel S. Weld<sup>1,7</sup> Oren Etzioni<sup>1</sup> Sebastian Kohlmeier<sup>1</sup>

<sup>1</sup>Allen Institute for AI <sup>2</sup>Microsoft Research <sup>3</sup>National Library of Medicine

<sup>4</sup>Kaggle <sup>5</sup>Chan Zuckerberg Initiative <sup>6</sup>Georgetown University <sup>7</sup>University of Washington

{lucyw, kylel}@allenai.org

Some tasks & corpora

# TREC biomedical tracks

Text REtrieval Conference (TREC)

...to encourage research in information retrieval from large text collections.



TREC Biomedical Tracks

trec-cds.org

## TREC Biomedical Tracks

[Home](#)  
[2022 CT Task](#)  
[2021 CT Task](#)  
[2020 PM Task](#)  
[2019 PM Task](#)  
[2018 PM Task](#)  
[2017 PM Task](#)  
[2016 CDS Task](#)  
[2015 CDS Task](#)  
[2014 CDS Task](#)  
[Mailing List](#)  
[TREC](#)

### Overview

Biomedical information retrieval is full of challenges, but has the potential to vastly improve the way biomedical information is accessed. This could improve the speed at which treatments are developed and disseminated into clinical practice. This site hosts the information for three of the five major medical track series that have run at the Text REtrieval Conference (TREC), with links to the other two major track series below. These tracks have sought to provide benchmark datasets and evaluate information retrieval systems focused on many of the most important information access problems in biomedicine.

1. **TREC Genomics** (2003-2007). This track focused on genomics researchers seeking relevant biomedical literature.
2. **TREC Medical Records** (2011-2012). This track focused on retrieving cohorts of patients from electronic health records (EHRs).
3. **Clinical Decision Support** (2014, 2015, 2016). This track focused on clinicians looking for evidence-based full-text literature to support diagnosis, treatment, and testing decisions.
4. **Precision Medicine** (2017, 2018, 2019, 2020). This track focused on oncologists looking for evidence-based treatment literature and clinical trials.
5. **Clinical Trials** (2021, 2022). This (ongoing) track focuses on matching patients to relevant clinical trials.

### 2021 Coordinators

**Kirk Roberts**, University of Texas Health Science Center at Houston (UTHealth)  
**William Hersh & Steven Bedrick**, Oregon Health and Science University (OHSU)  
**Dina Demner-Fushman**, U.S. National Library of Medicine (NLM)  
**Ellen Voorhees**, National Institute of Standards and Technology (NIST)

### Mailing List

<http://groups.google.com/d/forum/trec-cds>



A challenge on large-scale  
biomedical semantic indexing  
and question answering

*BioASQ organizes challenges on biomedical semantic indexing and  
question answering (QA)*

<http://bioasq.org/>



A challenge on large-scale  
biomedical semantic indexing  
and question answering

### **BioASQ Task 10b: Biomedical Semantic QA (Involves IR, QA, Summarization)**

... participants will have to respond to each test question with relevant concepts (from designated terminologies and ontologies), relevant articles (in English, from designated article repositories), relevant snippets (from the relevant articles), relevant RDF triples (from designated ontologies), exact answers (e.g., named entities in the case of factoid questions) and 'ideal' answers (English paragraph-sized summaries).



A challenge on large-scale  
biomedical semantic indexing  
and question answering

### **BioASQ Task 10b: Biomedical Semantic QA (Involves IR, QA, Summarization)**

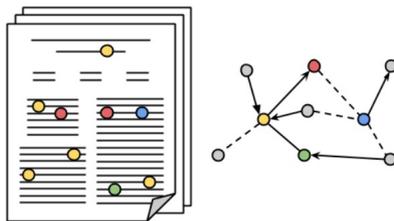
... participants will have to respond to each test question with relevant concepts (from designated terminologies and ontologies), relevant articles (in English, from designated article repositories), relevant snippets (from the relevant articles), relevant RDF triples (from designated ontologies), exact answers (e.g., named entities in the case of factoid questions) and 'ideal' answers (English paragraph-sized summaries).

... More than **4,200 training questions** (that were used as dry-run or test questions in previous year) are already available, along with their gold standard answers (relevant concepts, articles, snippets, exact answers, summaries). About **500 new test questions** will be used this year. All the questions are constructed by biomedical experts from around Europe.

# Multi-corpus benchmarks: Blurb

BLURB

[Leaderboard](#) [Paper](#) [Models](#) [Tasks](#) [Submit](#) [News](#)



## BLURB

**BLURB** is the **B**iomedical **L**anguage **U**nderstanding and **R**easoning **B**enchmark.

**BLURB** is a collection of resources for biomedical natural language processing. In general domains, such as newswire and the Web, comprehensive benchmarks and leaderboards such as **GLUE** have greatly accelerated progress in open-domain NLP. In biomedicine, however, such resources are ostensibly scarce. In the past, there have been a plethora of shared tasks in biomedical NLP, such as **BioCreative**, **BioNLP Shared Tasks**, **SemEval**, and **BioASQ**, to name just a few. These efforts have played a significant role in fueling interest and progress by the research community, but they typically focus on individual tasks. The advent of neural language models, such as **BERT** provides a unifying foundation to leverage transfer learning from unlabeled text to support a wide range of NLP applications. To accelerate progress in biomedical pretraining strategies and task-specific methods, it is thus imperative to create a broad-coverage benchmark encompassing diverse biomedical tasks.

Inspired by prior efforts toward this direction (e.g., **BLUE**), we have created BLURB (short for Biomedical Language Understanding and Reasoning Benchmark). BLURB comprises of a comprehensive benchmark for PubMed-based biomedical NLP applications, as well as a leaderboard for tracking progress by the community. BLURB includes thirteen publicly available datasets in six diverse tasks. To avoid placing undue emphasis on tasks with many available datasets, such as named entity recognition (NER), BLURB reports the macro average across all tasks as the main score. The BLURB leaderboard is model-agnostic. Any system capable of producing the test predictions using the same training and development data can participate. The main goal of BLURB is to lower the entry barrier in biomedical NLP and help accelerate progress in this vitally important field for positive societal and human impact.

# Multi-corpus benchmarks: Blurb

	BioBERT [34]	SciBERT [8]	BLUE [45]	BLURB
BC5-chem [35]	✓	✓	✓	✓
BC5-disease [35]	✓	✓	✓	✓
NCBI-disease [18]	✓	✓	-	✓
BC2GM [53]	✓	-	-	✓
JNLPBA [27]	✓	-	-	✓
EBM PICO [44]	-	✓	-	✓
ChemProt [31]	✓	✓	✓	✓
DDI [21]	✓	-	✓	✓
GAD [11]	✓	-	-	✓
BIOSSES [54]	-	-	✓	✓
HoC [20]	-	-	✓	✓
PubMedQA [25]	-	-	-	✓
BioASQ [42]	✓	-	-	✓

# Multi-corpus benchmarks: BigBIO

---

## **BIGBIO: A Framework for Data-Centric Biomedical Natural Language Processing**

---

**Jason Alan Fries<sup>1\*</sup> Leon Weber<sup>2,3\*</sup> Natasha Seelam<sup>4\*</sup> Gabriel Altay<sup>5\*</sup>**  
**Debajyoti Datta<sup>6†</sup> Ruisi Su<sup>7†</sup> Samuele Garda<sup>2†</sup> Sunny MS Kang<sup>8†</sup>**  
**Stella Biderman<sup>9,10†</sup> Matthias Samwald<sup>11†</sup> Stephen H. Bach<sup>12†</sup> Wojciech Kusa<sup>13†</sup>**  
**Samuel Cahyawijaya<sup>14†</sup> Fabio Barth<sup>2†</sup> Simon Ott<sup>11†</sup> Mario Sanger<sup>2†</sup> Bo Wang<sup>15</sup>**  
**Alison Callahan<sup>1</sup> Daniel Le3n Peria3n<sup>16</sup> Th3o Gigant<sup>7</sup> Patrick Haller<sup>2</sup>**  
**Jenny Chim<sup>17</sup> Jose Posada<sup>18</sup> John Giorgi<sup>19</sup> Karthik Rangasai Sivaraman<sup>20</sup>**  
**Marc Pamies<sup>21</sup> Marianna Nezhurina<sup>22</sup> Robert Martin<sup>2</sup> Moritz Freidank<sup>23</sup>**  
**Nathan Dahlberg<sup>7</sup> Shubhanshu Mishra<sup>24</sup> Shamik Bose<sup>7</sup> Nicholas Broad<sup>25</sup>**  
**Yanis Labrak<sup>26</sup> Shlok S Deshmukh<sup>27</sup> Sid Kiblawi<sup>28</sup> Ayush Singh<sup>7</sup> Minh Chien Vu<sup>29</sup>**  
**Trishala Neeraj<sup>30</sup> Jonas Golde<sup>2</sup> Albert Villanova del Moral<sup>25</sup> Benjamin Beilharz<sup>31</sup>**

<sup>1</sup>Stanford University <sup>2</sup>Humboldt-Universitat zu Berlin

<sup>3</sup>Max Delbruck Center for Molecular Medicine <sup>4</sup>Sherlock Biosciences <sup>5</sup>Tempus Labs Inc.

<sup>6</sup>University of Virginia <sup>7</sup>BigScience <sup>8</sup>Immuneering <sup>9</sup>EleutherAI <sup>10</sup>Booz Allen Hamilton

<sup>11</sup>Institute of Artificial Intelligence, Medical University of Vienna <sup>12</sup>Brown University

<sup>13</sup>TU Wien <sup>14</sup>The Hong Kong University of Science and Technology

<sup>15–25</sup>See Appendix A \*Equal Contribution †Equal Contribution

Corresponding Authors: jason-fries@stanford.edu leonweber@posteo.de  
nseelam1@gmail.com gabriel.altay@gmail.com

NAMED ENTITY RECOGNITION					
bc5cdr	ask a patient	anat em	an em	bionlp st 2011 epi	bionlp st 2011 ge
bioinfer	bionlp st 2013 gro	bionlp st 2019 bb	bionlp shared task 2009	bionlp st 2013 pc	bionlp st 2013 ge
bionlp st 2011 rel	bionlp st 2011 id	bionlp st 2013 cg	hprd50	genetag	genia ptm event corpus
jnlpba	genia term corpus	euadr	medmentions	linnaeus	meddocan
mantra gsc	gnormplus	ebm pico	chebi nactem	cellfinder	cantemist
cadec	bioscope	biorelex	biored	chemdner	chemprot
distemist	diann iber eval	ddi corpus	ctebmsp	cord ner	citation gia test collection
chia	nlm gene	ncbi disease	n2c2 2018 track2	n2c2 2014 deid	n2c2 2010
codiesp	n2c2 2009	pdr	osiris	pharmaconer	psytar
mlee	n2c2 2006 deid	progene	tmvar v3	thomas2011	swedish medical ner
mirna	mutation finder	pico extraction	tmvar v2	seth corpus	scai chemical
nlmchem	muchmore	pho ner	tmvar v1	scai disease	verspoor 2013

NAMED ENTITY DISAMBIGUATION				
ask a patient	medmentions	medal	mantra gsc	cantemist
bc5cdr	biorelex	msh wsd	mirna	nlmchem
bionlp st 2019 bb	cadec	nlm wsd	osiris	tmvar v3
gnormplus	codiesp	nlm gene	tmvar v2	spl adr 200db
linnaeus	citation gia test collection	ncbi disease	thomas2011	twadr

RELATION EXTRACTION					
bc5cdr	bionlp st 2013 ge	bionlp st 2011 rel	iepa	hprd50	
an em	genia relation corpus	biorelex	biored	ddi corpus	
bioinfer	euadr	chia	n2c2 2018 track2	n2c2 2010	
bionlp st 2013 gro	ill	chemprot	spl adr 200db	verspoor 2013	
bionlp st 2019 bb	chebi nactem	mlee	seth corpus		

TEXT CLASSIFICATION				
bioasq 2021 mesinesp	hallmarks of cancer	gad	meddialog	essai
bc7 litcovid	cas	codiesp	n2c2 2018 track1	n2c2 2014 risk factors
bioasq task c 2017	cantemist	n2c2 2008	ntcir 13 medweb	nlmchem
geokhoj v1	chemdner	n2c2 2006 smokers	pharmaconer	psytar
			psytar	scicite

EVENT EXTRACTION			
bionlp st 2011 ge	bionlp st 2011 epi	bionlp st 2013 gro	
bionlp st 2013 pc	bionlp st 2013 ge	bionlp st 2013 cg	genia ptm event corpus
bionlp shared task 2009	bionlp st 2011 id	mlee	pdr

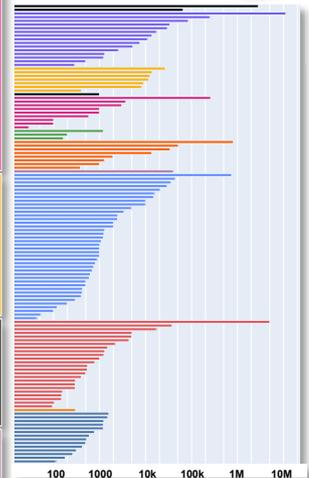
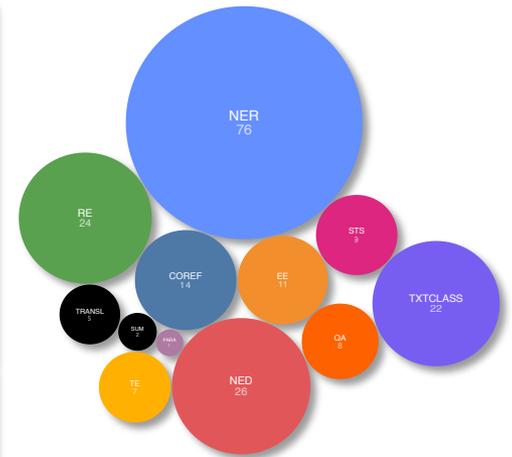
SEMANTIC SIMILARITY		
bio simlex	bio sim verb	mayosrs
ehr rel	minimayosrs	mqp
biosses	pmc patients	umnsrs

COREFERENCE RESOLUTION				
an em	bionlp st 2013 pc	bionlp st 2011 rel	genia ptm event corpus	biorelex
bionlp st 2011 ge	bionlp shared task 2009	bionlp st 2013 cg	mlee	pdr
bionlp st 2011 epi	bionlp st 2011 id	bionlp st 2013 ge	n2c2 2011	

QUESTION ANSWERING	
bioasq task b	biomrc
biology how why corpus	med qa
medqa qa	pubmed qa
medhop	sciq

TEXTUAL ENTAILMENT			
evidence inference	medqa rqe	medical data	sofact
mednli	medqa nli	scitail	

TRANSLATION		SUMMARIZATION	
diann iber eval	paramed	ntcir 13 medweb	meqsum
muchmore	scielo		multi xscience
			PARAPHRASING



- Task Categories**
- COREFERENCE RESOLUTION
  - EVENT EXTRACTION
  - NAMED ENTITY DISAMBIGUATION
  - NAMED ENTITY RECOGNITION
  - PARAPHRASING
  - QUESTION ANSWERING
  - RELATION EXTRACTION
  - SEMANTIC SIMILARITY
  - SUMMARIZATION
  - TEXTUAL ENTAILMENT
  - TEXT CLASSIFICATION
  - TRANSLATION

# Zero-shot performance

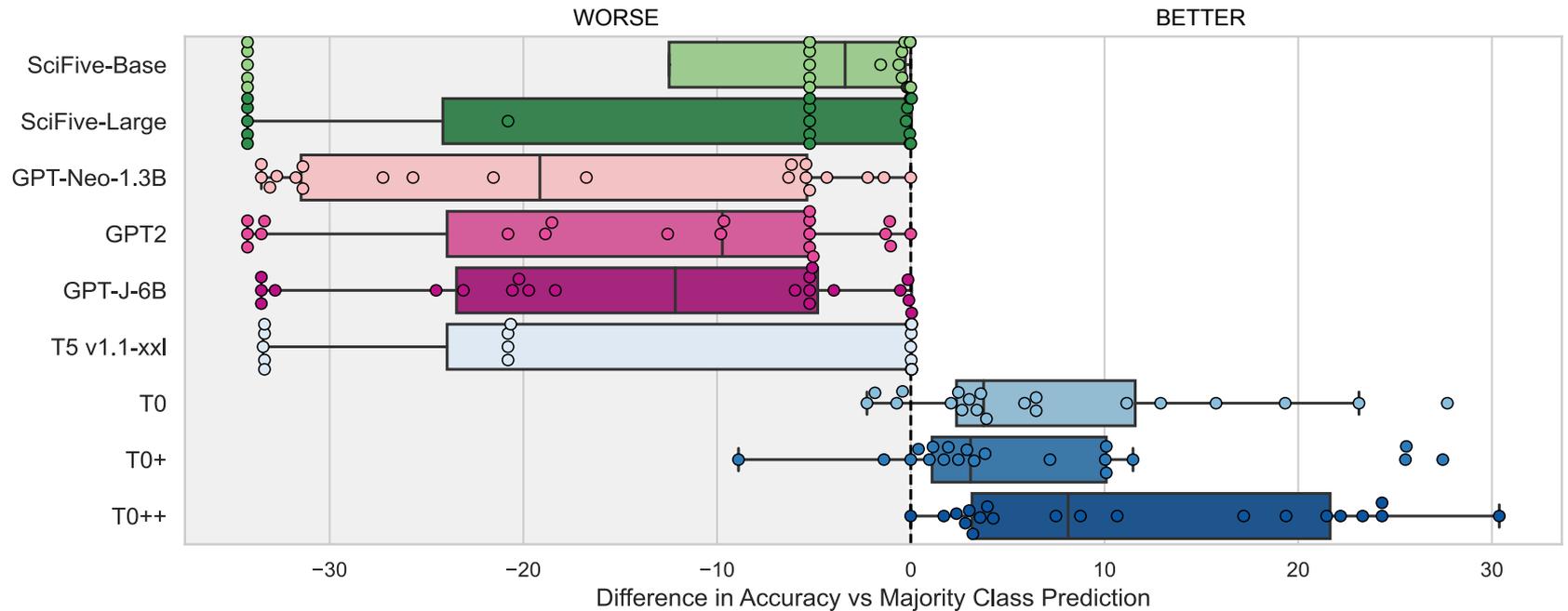


Figure 3: Zero-shot generalization to biomedical tasks. Box plots show pooled accuracy differences between a majority class baseline and zero-shot prediction for all datasets excluding BIOSSES. Points are per-prompt scores. T0 is the only language model class to outperform the majority baseline.

Retrieval and language technologies to support  
*evidence-based medicine (EBM)*

# Evidence-Based Medicine *n.*

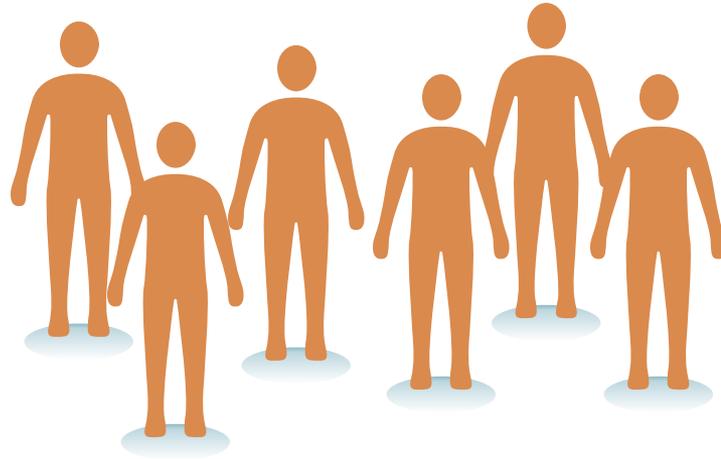
The conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients



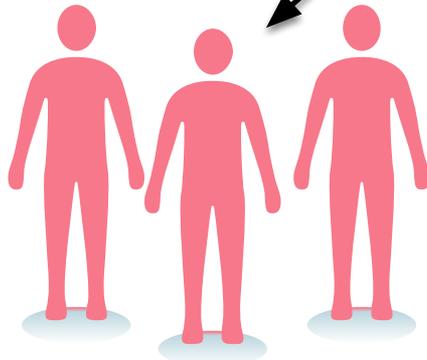
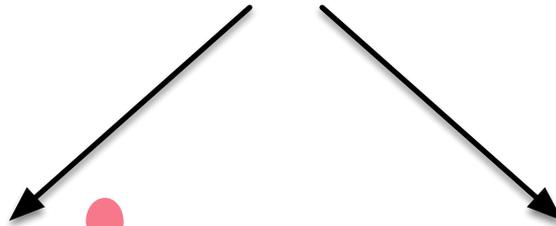
# Population, Intervention, Comparator, Outcome (PICO elements)

In **women with hypertension** does a program  
of **regular exercise** versus **no regular exercise**  
measurably reduce **hypertension**?

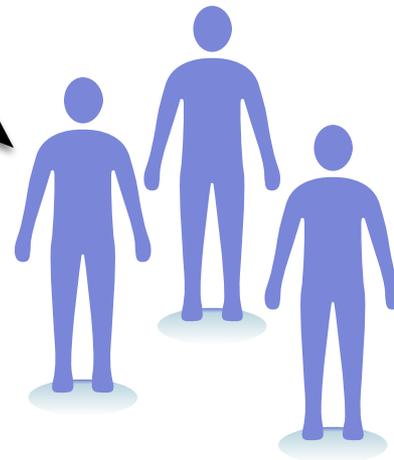
*Patients from a target population*



*Randomize*

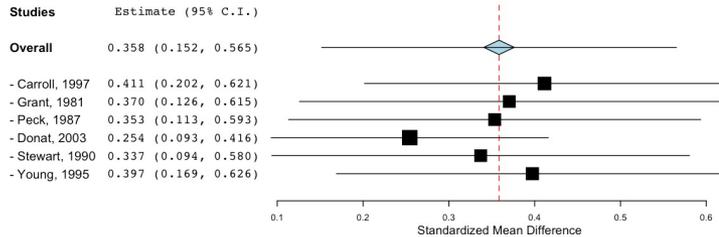
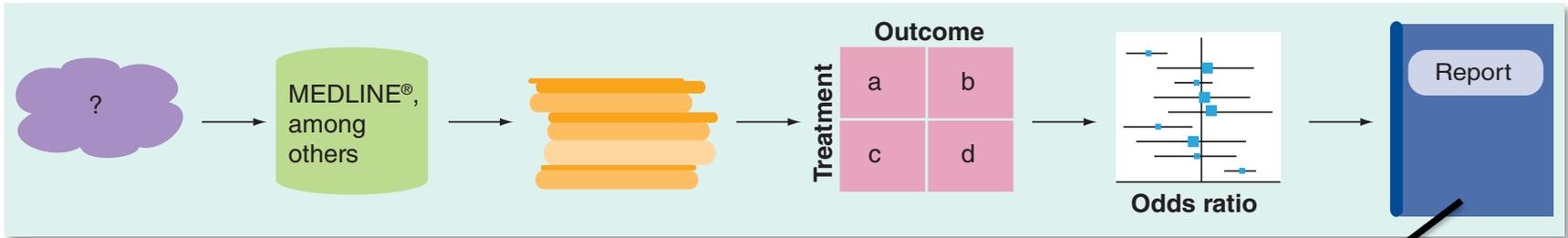


*Treatment A*

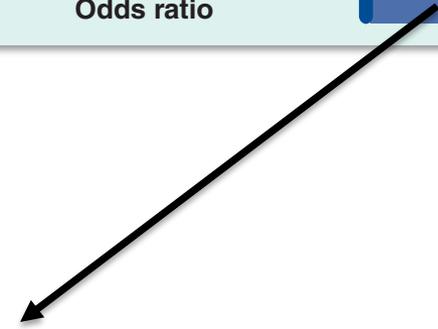


*Treatment B*





Meta-analysis



**Results.** Data on triptans were derived from 186 RCTs summarized in 9 systematic reviews (101,276 patients, most studied was sumatriptan, followed by zolmitriptan, eletriptan, naratriptan, almotriptan, rizatriptan, and frovatriptan). Compared with placebo, triptans resolved pain at 2 hours and 1 day, and increased the risk of mild and transient adverse events (high strength of the body of evidence [SOE]). Data on NSAIDs were derived from 3 systematic reviews (9 RCTs, 4,373 patients, most studied was ibuprofen, followed by diclofenac and ketorolac). Compared with placebo, NSAIDs probably resolved pain at 2 hours and 1 day, and increased the risk of mild and transient adverse events (moderate SOE). For other interventions, we included 134 RCTs and 6 comparative observational studies (33,623 patients). Compared with placebo, antiemetics (low SOE), dihydroergotamine (moderate to high SOE), ergotamine plus caffeine (moderate SOE) and acetaminophen (moderate SOE) reduced acute pain.

Narrative synopsis



## **Drug treatments for covid-19: living systematic review and network meta-analysis**

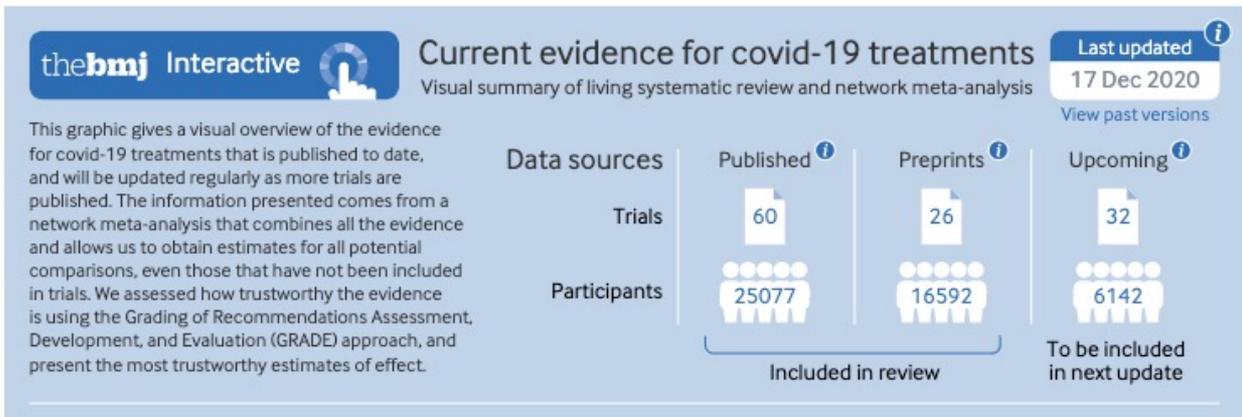
# Drug treatments for covid-19: living systematic review and network meta-analysis

## Research

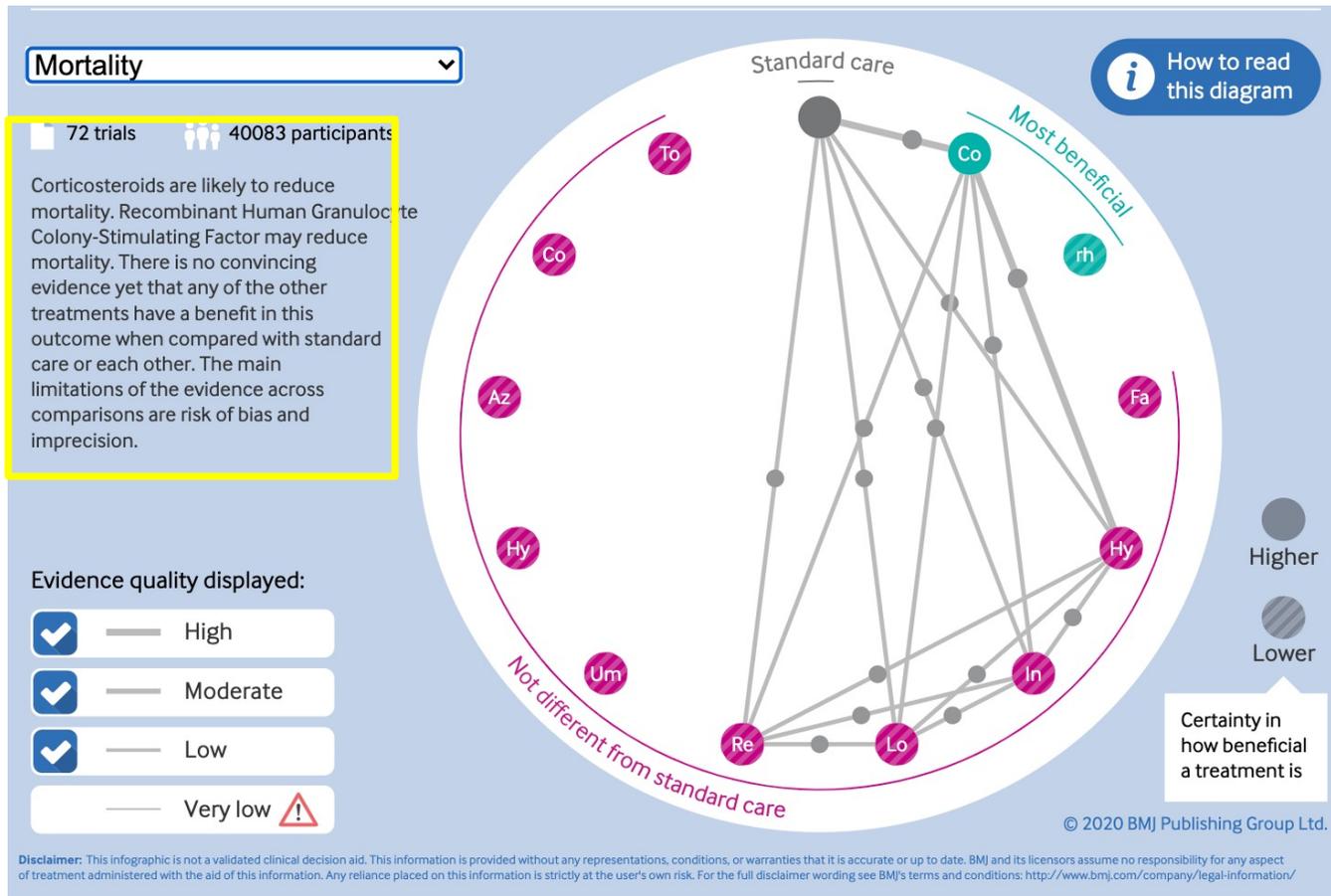
### Drug treatments for covid-19: living systematic review and network meta-analysis

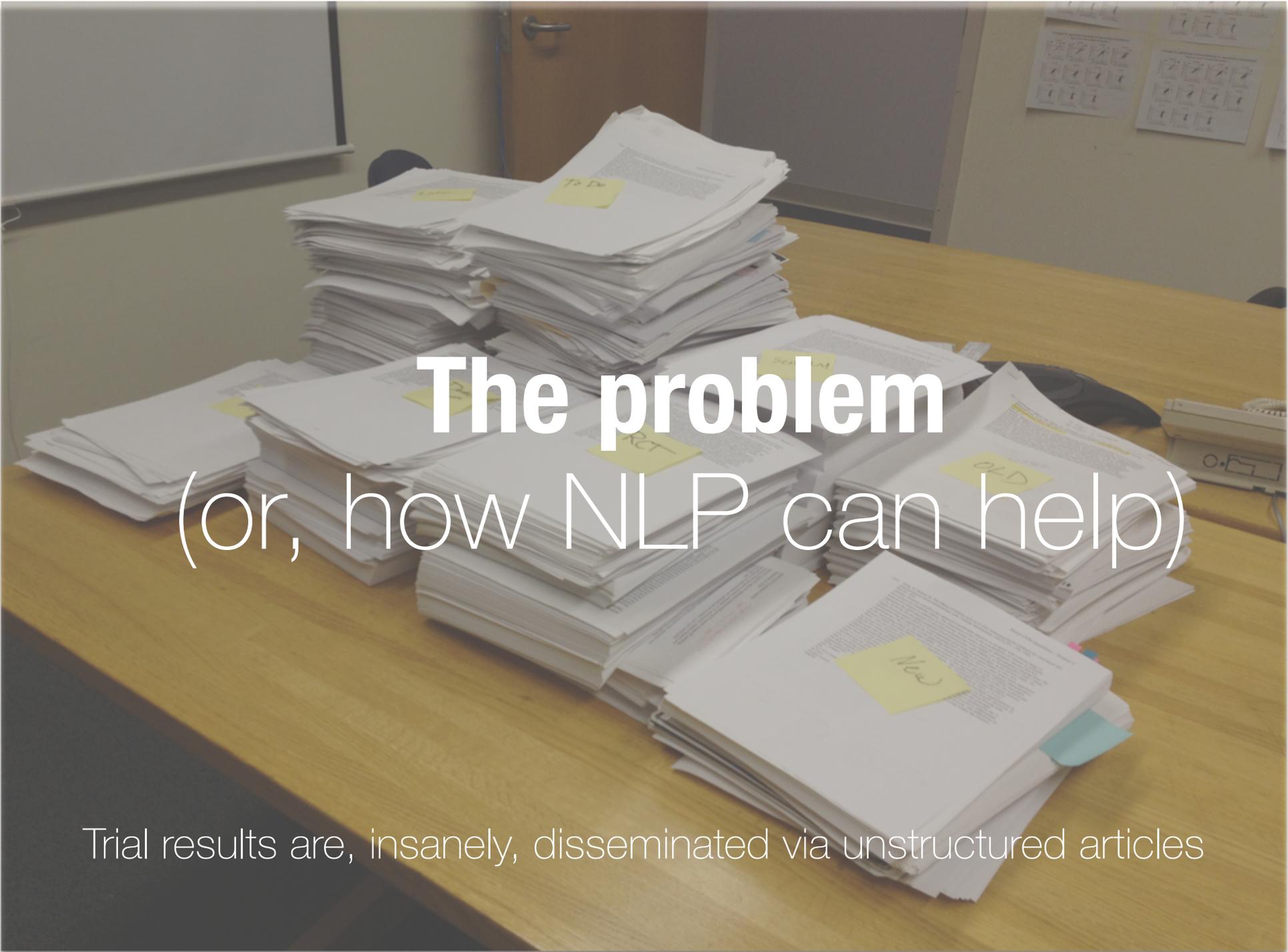
BMJ 2020 ; 370 doi: <https://doi.org/10.1136/bmj.m2980> (Published 30 July 2020)

Cite this as: *BMJ* 2020;370:m2980



# Drug treatments for covid-19: living systematic review and network meta-analysis



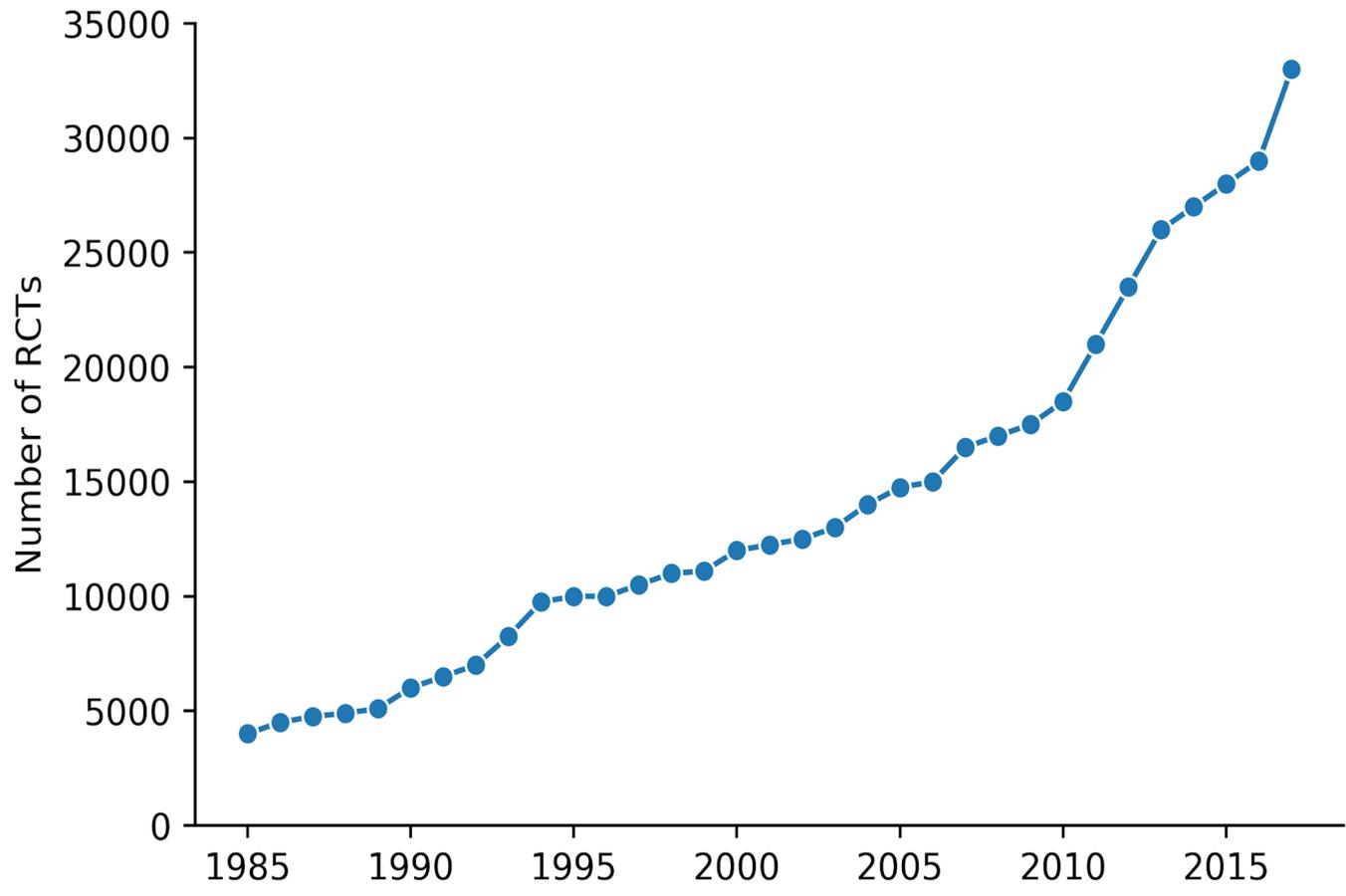


**The problem**  
(or, how NLP can help)

Trial results are, insanely, disseminated via unstructured articles

Estimated time to complete and publish a  
systematic review: **67.3 weeks**

*Borah et al, BMJ, 2017*

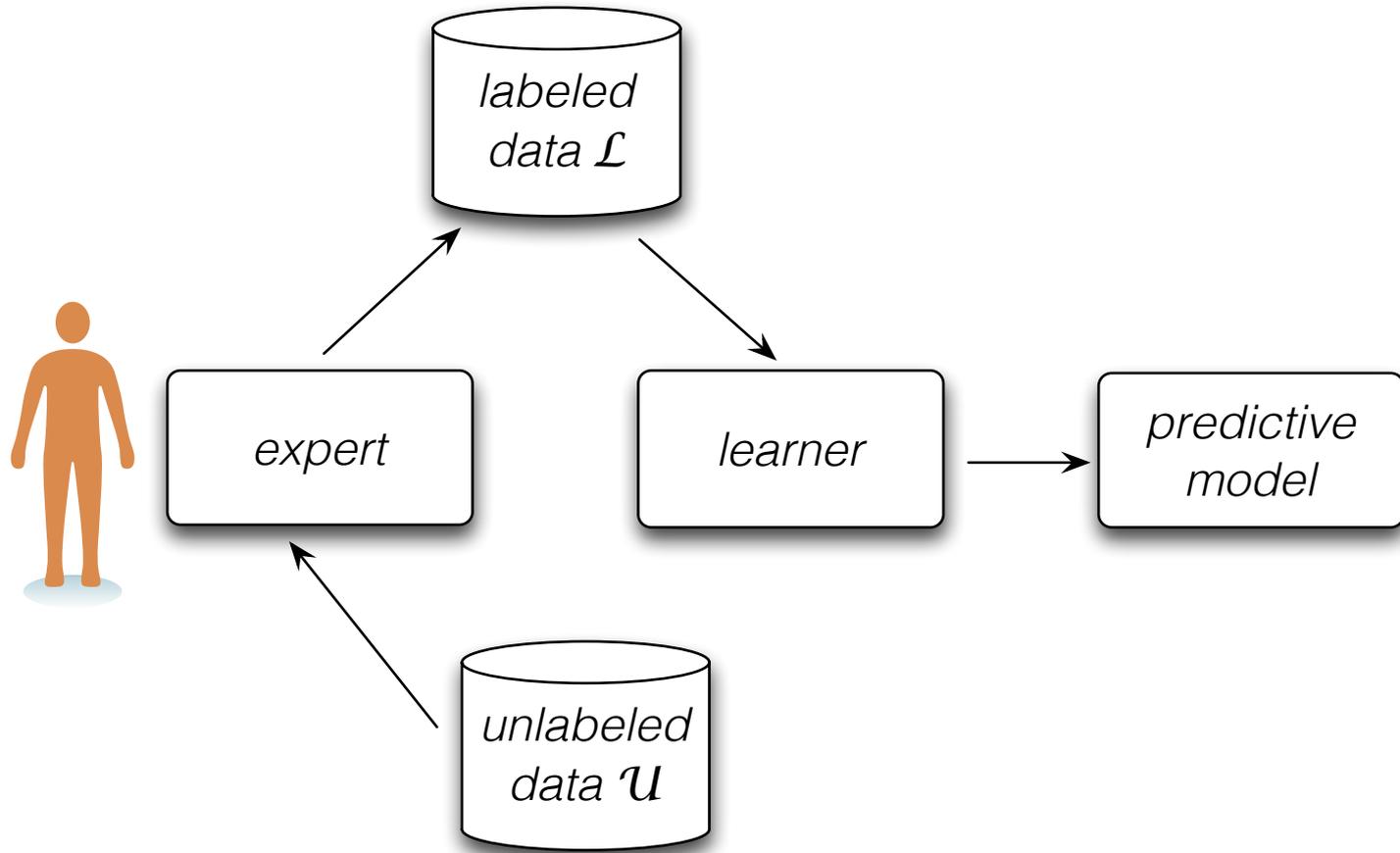


# Screening for reviews

Standard workflow: High-recall Boolean query (often quite complex) to PubMed, then manually *screen* for relevance



# Semi-automating screening



Research | [Open Access](#) | [Published: 15 June 2015](#)

## Faster title and abstract screening? Evaluating Abstrackr, a semi-automated online screening program for systematic reviewers

[John Rathbone](#) , [Tammy Hoffmann](#) & [Paul Glasziou](#)

[Systematic Reviews](#) **4**, Article number: 80 (2015) | [Cite this article](#)

**6891** Accesses | **69** Citations | **39** Altmetric | [Metrics](#)

Research | [Open Access](#) | [Published: 02 April 2020](#)

## Machine learning for screening systematic reviews: Abstrackr and E

[Amy Y. Tsou](#) 

### ORIGINAL ARTICLE

## A text-mining tool generated title-abstract screening workload savings: performance evaluation versus single-human screening

**Niamh Carey\***, Marie Harte, Laura Mc Cullagh

*National Centre for Pharmacoeconomics, Old Stone Building, Trinity Centre for Health Sciences, St James's Hospital, Dublin 8, Ireland*  
*Department of Pharmacology and Therapeutics, Trinity Centre for Health Sciences, St James's Hospital, Dublin 8, Ireland*

Accepted 24 May 2022; Published online 30 May 2022

ization in  
nance of

Research | [Open Access](#) | [Published: 15 June 2015](#)

## Faster title and abstract screening? Evaluating Abstrackr, a semi-automated online screening program for systematic reviewers

[John Rathbone](#) , [Tammy Hoffmann](#) & [Paul Glasziou](#)

*Systematic Reviews* 4, Article number: 80 (2015) | [Cite this article](#)

6891 Accesses | 69 Citations | 39 Altmetric | [Metrics](#)

Research | [Open Access](#) | [Published: 02 April 2020](#)

## Machine learning for screening systematic reviews: Abstrackr and E

[Amy Y. Tsou](#) 

### ORIGINAL ARTICLE

## A text-mining tool generated title-abstract screening workload savings: performance evaluation versus single-human screening

Niamh Carey\*, Marie Harte, Laura Mc Cullagh

National Centre for Pharmacoeconomics, Old Stone Building, Trinity Centre for Health Sciences, St James's Hospital, Dublin 8, Ireland  
Department of Pharmacology and Therapeutics, Trinity Centre for Health Sciences, St James's Hospital, Dublin 8, Ireland

Accepted 24 May 2022; Published online 30 May 2022

**Results:** Abstrackr reduced Stage 1 workload by 67% (5.4 days), when compared with Single-human screening. Sensitivity was high (91%). The false negative rate at Stage 1 was 9%; however, none of those citations were included following full-text screening. The high proportion of false positives ( $n = 2,001$ ) resulted in low specificity (72%) and precision (15.5%).

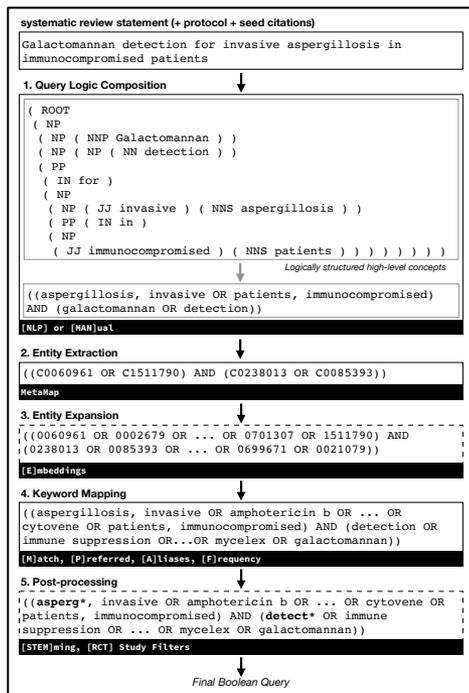
**Conclusion:** Abstrackr-assisted screening provided Stage 1 workload savings that did not come at the expense of omitting relevant citations. However, Abstrackr overestimated citation relevance, which may have negative workload implications at full-text

# Systematic Review Automation Tools for End-to-End Query Formulation

Hang Li  
University of Queensland  
Brisbane, Australia  
hang.li4@uq.net.au

Harrison Scells  
University of Queensland  
Brisbane, Australia  
h.scells@uq.net.au

Guido Zuccon  
University of Queensland  
Brisbane, Australia  
g.zuccon@uq.edu.au



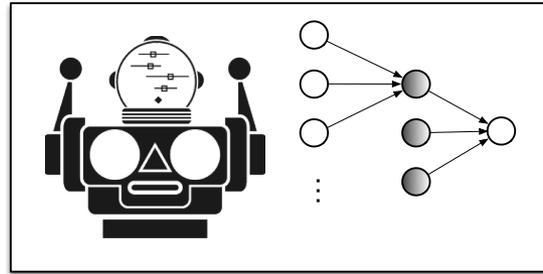
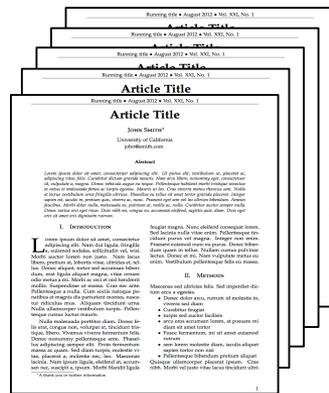
# Automatic Boolean Query Formulation for Systematic Review Literature Search

Harrison Scells  
University of Queensland  
Brisbane, Australia  
h.scells@uq.net.au

Bevan Koopman  
CSIRO  
Brisbane, Australia  
bevan.koopman@csiro.au

Guido Zuccon  
University of Queensland  
Brisbane, Australia  
g.zuccon@uq.edu.au

Justin Clark  
IEBH, Bond University  
Gold Coast, Australia  
jclark@bond.edu.au



RobotReviewer

Unstructured free-text articles  
describing clinical trials

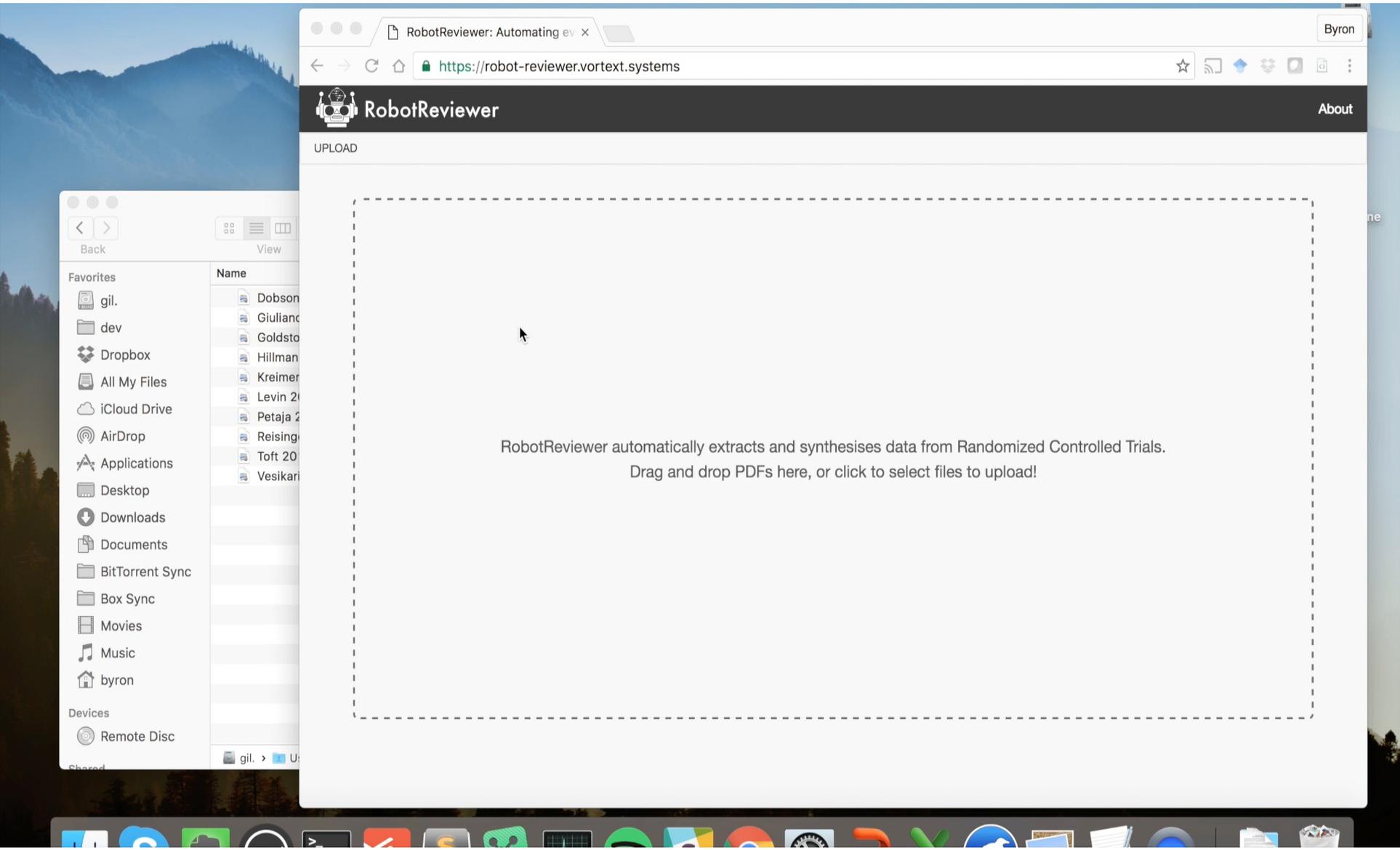
### Trial summaries

n	Participants	Interventions	Outcomes	punchline	finding
17	resistant hypertension in a sub-Saharan African population of T2DM patients from Cameroon, type 2 diabetes mellitus, 17 subjects presenting with resistant hypertension in specialized diabetes care units in Cameroon, sub-Saharan African population	alternative antihypertensive regimen (n = 8), on top of any ongoing regimen and prevailing lifestyle prescriptions, low-dose spironolactone, low-dose spironolactone, spironolactone	office systolic BP, diastolic BP, resistant hypertension, serum potassium, sodium, and creatinine levels, potassium levels, change in office and self-measured blood pressure (BP), sodium and creatinine levels	Their mean systolic and diastolic office BP decreased respectively from 158 ± 17 mmHg to 125 ± 11 mmHg (p = 0.009) and from 86 ± 11 to 72 ± 8 mmHg (p = 0.009).	↓ sig decrease
65	Patients, hypertensive patients, Sixty-five hypertensive patients not receiving medication, Patients with Hypertension	Intervention, electroacupuncture (EA, acupuncture intervention, Electroacupuncture	systolic blood pressure (SBP) and diastolic blood pressures (DBP, Blood Pressure, plasma concentration of norepinephrine, renin, long-lasting blood pressurelowering acupuncture effect, blood pressure, effectiveness of EA were peak and average SBP and DBP, blood pressures, peak and average SBP and DBP, Sympathetic and renin-aldosterone systems, underlying mechanisms of acupuncture with plasma norepinephrine, renin, and aldosterone, 24-hour ambulatory blood pressure monitoring	Hormone Responses EA did not affect plasma epinephrine (40 -6 to 38 -8 ng/mL; P > 0.05) in 25 patients.	— no diff



Iain Marshall  
King's College

Marshall, Iain J., Joël Kuiper, Edward Banner, and Byron C. Wallace. "Automating biomedical evidence synthesis: RobotReviewer." ACL (Demos) 2017. <http://www.robotreviewer.net/>

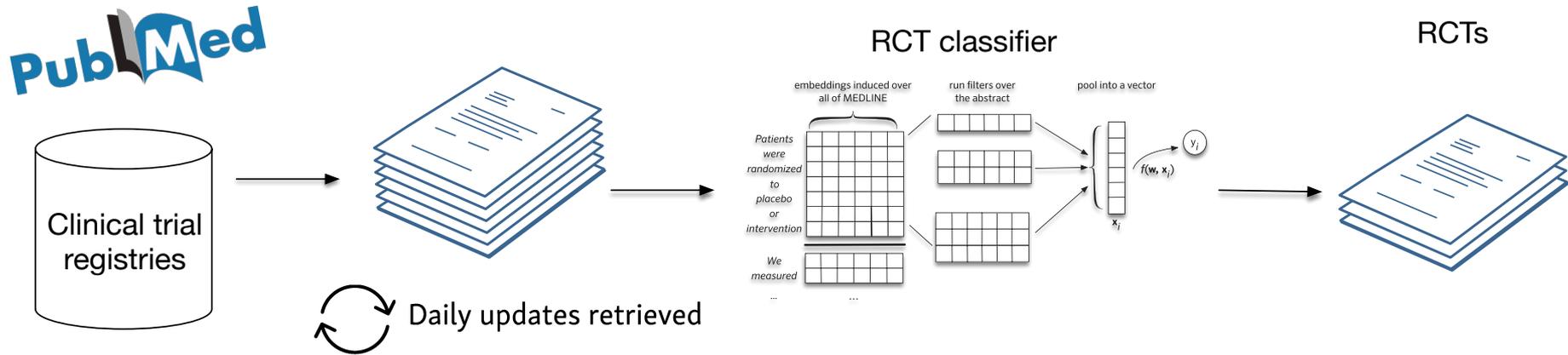


Marshall, Iain J., Joël Kuiper, Edward Banner, and Byron C. Wallace. "Automating biomedical evidence synthesis: RobotReviewer." ACL (Demos) 2017. <http://www.robotreviewer.net/>

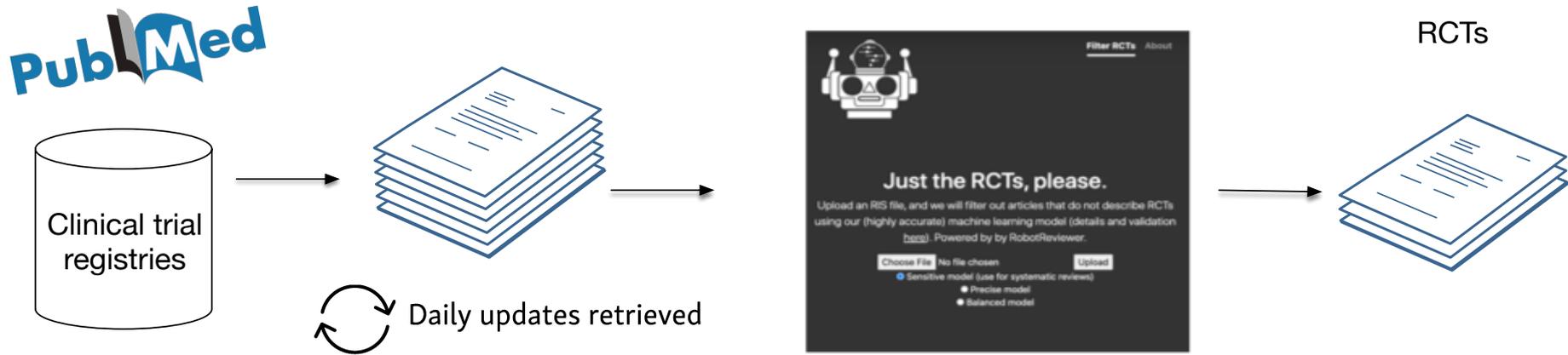
## Drug treatments for covid-19: living systematic review and network meta-analysis

the database and to capture eligible studies the day of or the day after publication. To identify randomised controlled trials, we filtered the results from the CDC's database through a validated and highly sensitive machine learning model.<sup>11</sup> We tracked preprints of

# Trialstreamer



# Trialstreamer



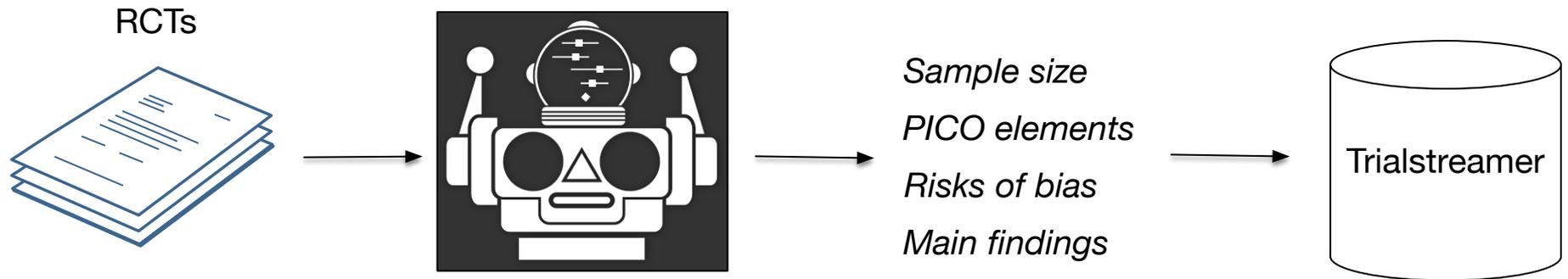
SPECIAL ISSUE PAPER | [Open Access](#) | |

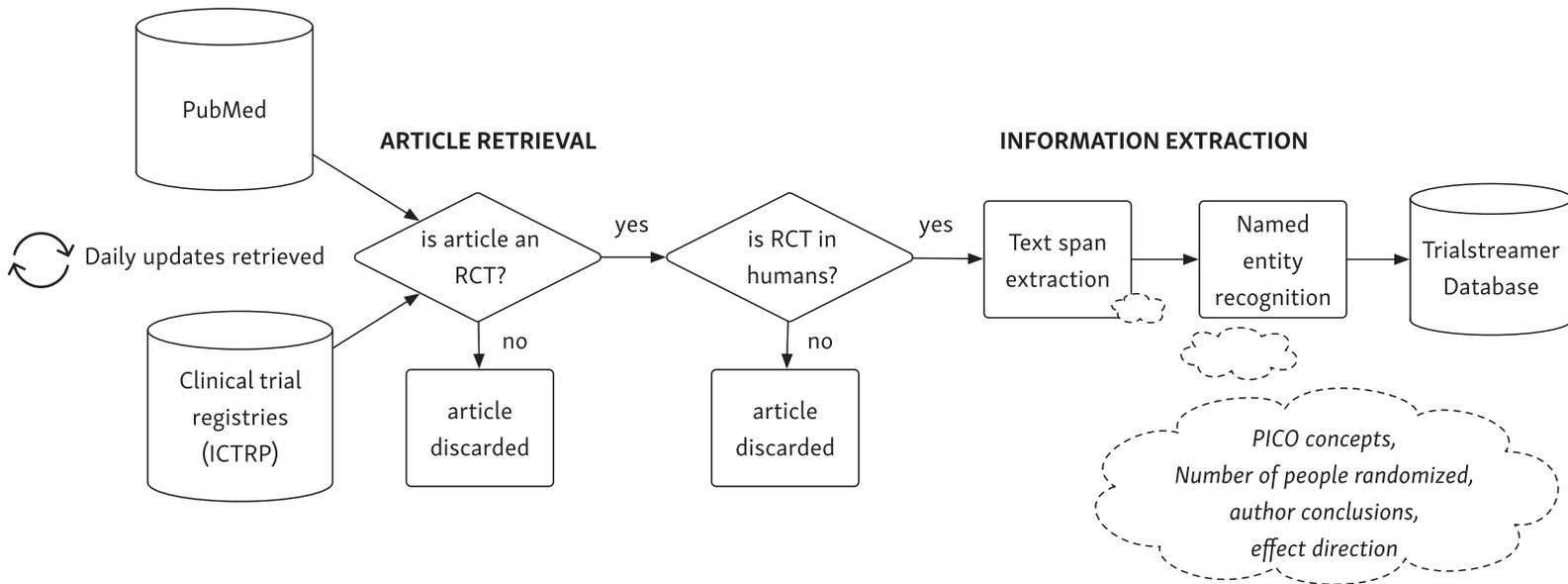
**Machine learning for identifying Randomized Controlled Trials: An evaluation and practitioner's guide**

Iain J. Marshall , Anna Noel-Storr, Joël Kuiper, James Thomas, Byron C. Wallace

First published 04 January 2018 | <https://doi.org/10.1002/jrsm.1287> | Citations: 28

# Trialstreamer





The screenshot shows a web browser window with the URL `trialstreamer.robotreviewer.net`. The page features the Trialstreamer logo, a navigation menu, and a search bar. Below the search bar, there are example search queries and information about open access and collaborations.

**TRIALSTREAMER** a living, annotated database of 689,627 randomized controlled trials. Last updated 21 minutes ago. [About](#)

Start typing a Population, Intervention, Comparator, or Outcome (PICO)

Try some of these examples:

- [What is the effect of remdesivir for treating COVID-19?](#)
- [What is the best ventilation strategy for acute respiratory distress syndrome \(ARDS\)?](#)
- [Do chloroquine or hydroxychloroquine improve outcomes in COVID-19 infection?](#)
- [What trials are underway on the use of Extracorporeal Membrane Oxygenation \(ECMO\) in COVID-19?](#)

Open access:

- Daily data updates freely available at [Zenodo](#) DOI: [10.5281/zenodo.3767068](https://doi.org/10.5281/zenodo.3767068)
- Source code available for [RobotReviewer](#), the [Trialstreamer database](#), and [this demonstration website](#).

A collaboration between:

<https://trialstreamer.ieai.robotreviewer.net/>

Nye et al. *Trialstreamer: Mapping and Browsing Evidence in Real-Time*. ACL (demos), 2020.

Marshall et al. *Trialstreamer: a living, automatically updated database of clinical trial reports* JAMIA, 2020.





**Welcome!** Trialstreamer is an artificial intelligence system, which finds and summarises new trial publications, registrations, and preprints in COVID-19. We monitor databases continuously, so you can find new RCT evidence the moment it is published, and surface the biggest and highest quality trials first. Please see the [About page](#) for information about how this all works.

COVID-19 [population] Remdesivir [interventions] Start typing a Population, Intervention, Comparator, or Outcome (PICO)

Showing 26 results

All (26) Published articles (4) Preprints (1) Registered trials (21)

Get large/high quality trials first Newest first

journal article

### Remdesivir for the Treatment of Covid-19 - Preliminary Report.

32445440 Beigel et al., The New England journal of medicine. 2020 10.1056/NEJMoa2007764

∞ 1063

CONCLUSIONS Remdesivir was superior to placebo in shortening the time to recovery in adults hospitalized with Covid-19 and evidence of lower respiratory tract infection.

**Population**

- 1059 patients (538 assigned to remdesivir and 521 to)
- adults hospitalized with Covid-19 and evidence of lower respiratory tract infection
- 1063 patients underwent randomization
- adults hospitalized with Covid-19 with evidence of lower respiratory tract involvement

**Interventions**

- remdesivir
- intravenous remdesivir
- placebo

**Outcomes**

- median recovery time
- time to recovery, defined by either discharge from the hospital or hospitalization for infection-control purposes only
- Serious adverse events

Probability of low risk of bias: 69%

journal article

### Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial.

32423584 Wang et al., Lancet (London, England). 2020 10.1016/S0140-6736(20)31022-9

∞ 237

Remdesivir use was not associated with a difference in time to clinical improvement (hazard ratio 1.23

**Welcome!** Trialstreamer is an artificial intelligence system, which finds and summarises new trial publications, registrations, and preprints in COVID-19. We monitor databases continuously, so you can find new RCT evidence the moment it is published, and surface the biggest and highest quality trials first. Please see the [About page](#) for information about how this all works.

**COVID-19 [population]** **Remdesivir [interventions]** Start typing a Population, Intervention, Comparator, or Outcome (PICO) Showing 26 results

All (26) Published articles (4) Preprints (1) Registered trials (21) Get large/high quality trials first **Newest first**

journal article

### Remdesivir for the Treatment of Covid-19 - Preliminary Report.

32445440 Beigel et al., The New England journal of medicine. 2020 10.1056/NEJMoa2007764

🔍 1063

**CONCLUSIONS** Remdesivir was superior to placebo in shortening the time to recovery in adults hospitalized with Covid-19 and evidence of lower respiratory tract infection.

#### Population

- 1059 patients (538 assigned to remdesivir and 521 to)
- adults hospitalized with Covid-19 and evidence of lower respiratory tract infection
- 1063 patients underwent randomization
- adults hospitalized with Covid-19 with evidence of lower respiratory tract involvement

#### Interventions

- remdesivir
- intravenous remdesivir
- placebo

#### Outcomes

- median recovery time
- time to recovery, defined by either discharge from the hospital or hospitalization for infection-control purposes only
- Serious adverse events

Probability of low risk of bias: 69%

journal article

### Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial.

32423584 Wang et al., Lancet (London, England). 2020 10.1016/S0140-6736(20)31022-9

🔍 237

Remdesivir use was not associated with a difference in time to clinical improvement (hazard ratio 1.23)

# RR-NLP

```
pip install rrnlp
```

```
import rrnlp

trial_reader = rrnlp.TrialReader()

ti_abs = {"ti": 'A Cluster-Randomized Trial of Hydroxychloroquine for Prevention of Covid-19',
          "ab": "'Background: Current strategies for preventing severe acute respiratory syndrome c

preds = trial_reader.read_trial(ti_abs)
```



<https://github.com/bwallace/RRnlp/>



# On demand summarization



COVID-19 [population] × Chloroquine [interventions] × Start typing a Population, Intervention, Comparator, or Outcome (PICO)

Showing 13 results

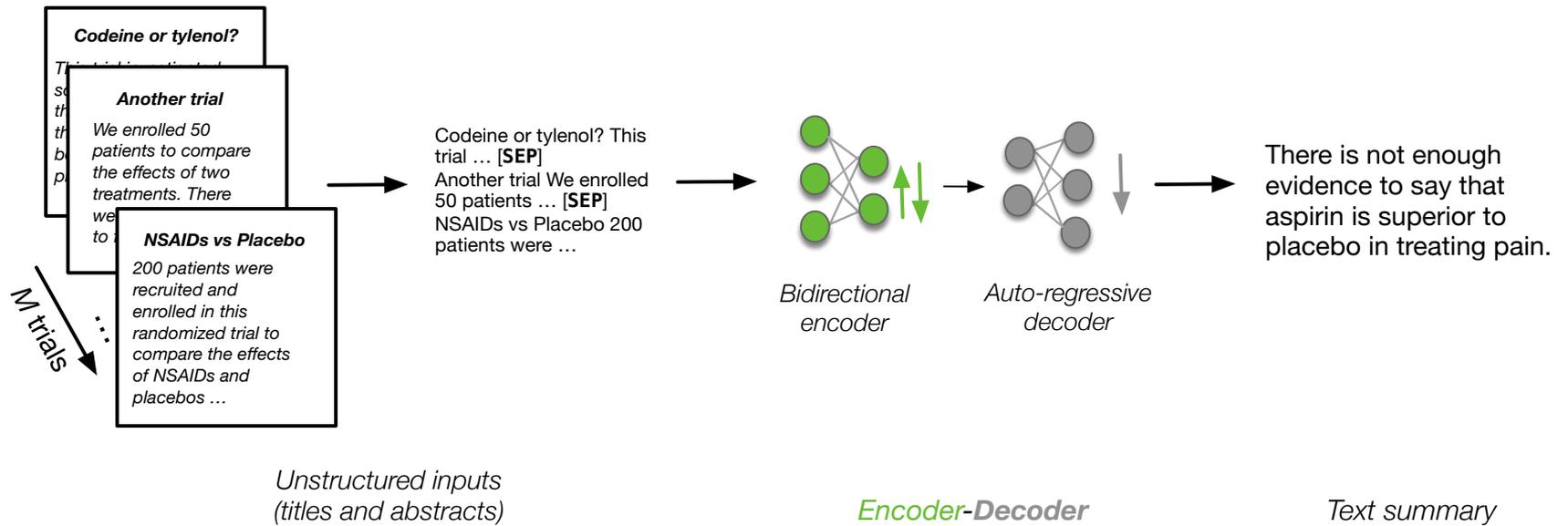
**Automatically generated summary (β!):** There is currently insufficient evidence to support the routine use of HCQ for the prophylaxis of SARS-CoV-2 infection in healthcare personnel. Further randomised controlled trials are needed to determine whether HCQ is of benefit to healthcare personnel and their carers, and to compare HCQ with other antiviral therapies.

All (168) Published articles (13) Preprints (6) Registered trials (149)

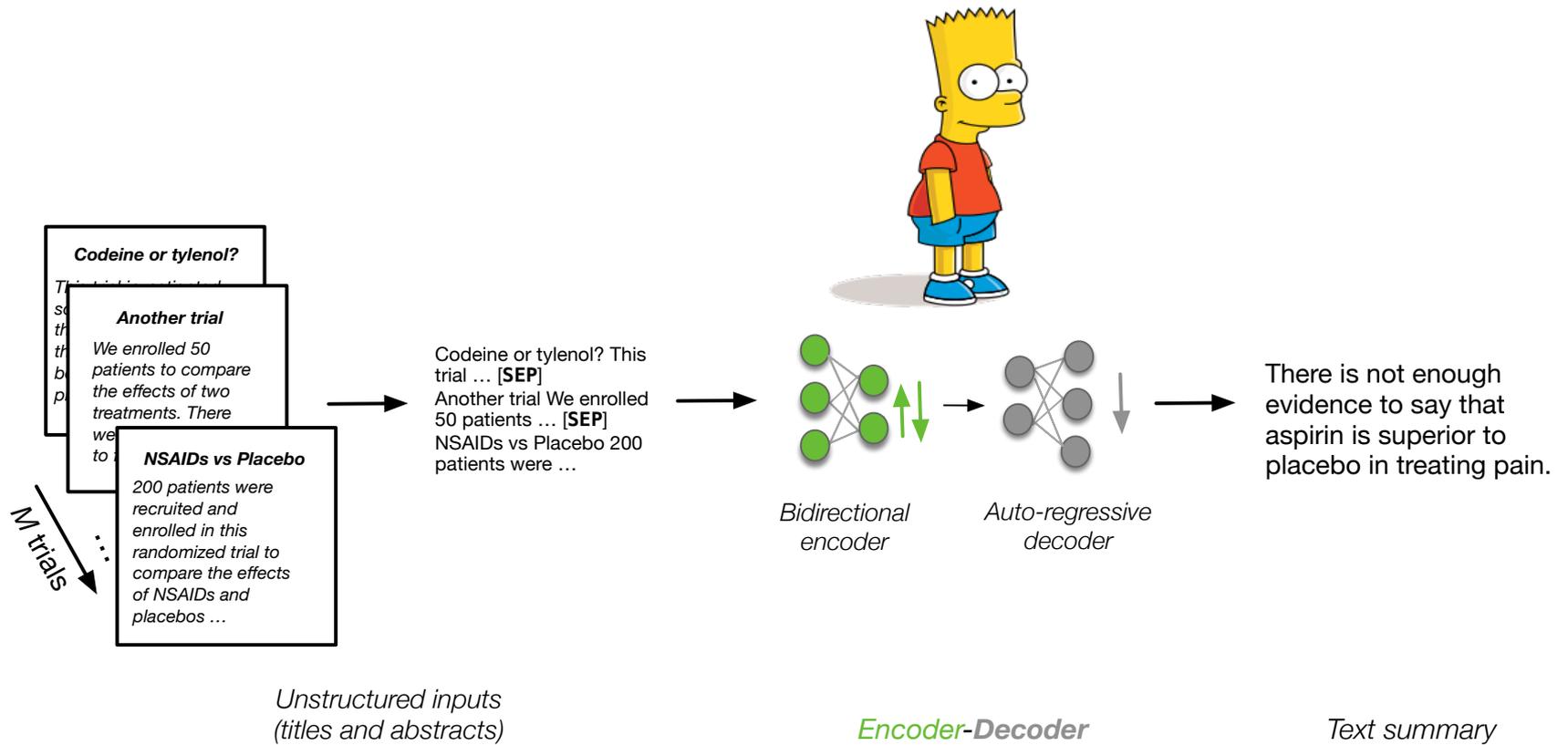
Get large/high quality trials first

Newest first



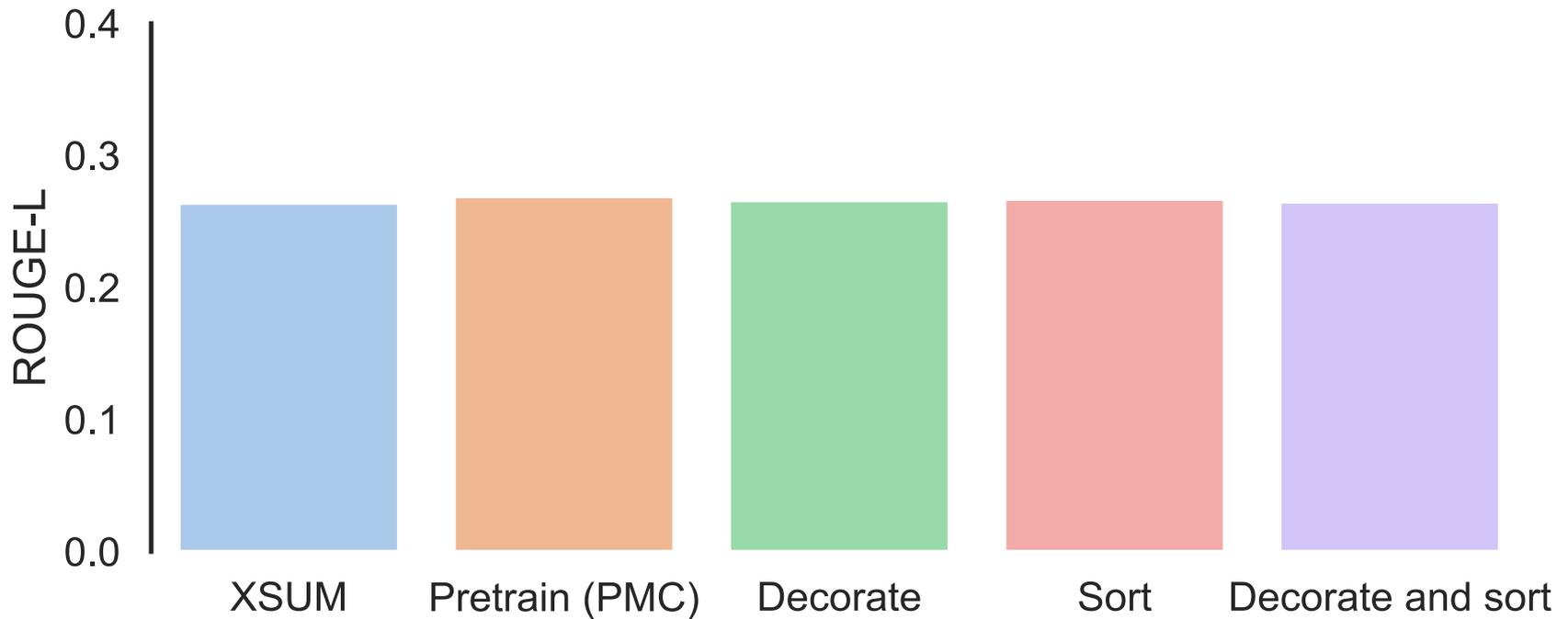


(Vanilla) multi-document summarization



(Vanilla) multi-document summarization

# ROUGE on different models: basically equivalent



*But this says nothing about how accurate these summaries are.*

Annotate x +

127.0.0.1:5000

## System Summary

Magnesium sulphate may reduce the incidence of eclampsia in women with mild to moderate preeclamatous hypertension. However, there is insufficient evidence to assess the effect of magnesium sulphate on other important outcomes, such as perinatal mortality and neurodevelopmental outcome. Further randomised controlled trials are needed to determine the optimal dose and route of administration, the optimal duration of prophylaxis, and the cost-effectiveness of this intervention.

## Reference

***Magnesium sulphate and other anticonvulsants for women with pre-eclampsia***

Magnesium sulphate more than halves the risk of eclampsia, and probably reduces maternal death. There is no clear effect on outcome after discharge from hospital. A quarter of women report side effects with magnesium sulphate.

---

**In the reference summary, characterize the authors' conclusion regarding the comparative effect of the main intervention(s) of interest, with respect to the outcome(s) summarized here. The summary concludes that the main intervention(s) ...**

Was harmful.

Had no effect.

Was beneficial.

Could not come to a conclusion due to a lack of relevant evidence.

**Now we ask that you assess the factuality of the system summary text, with reference to the assessment you made above regarding the key findings reported in the manually written summary. The system summary ...**

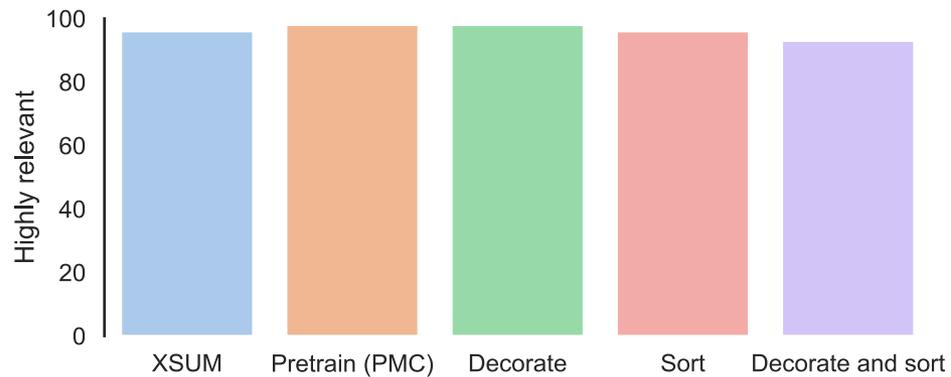
Strongly disagrees with or contradicts the reference summary.

Moderately disagrees with the reference summary.

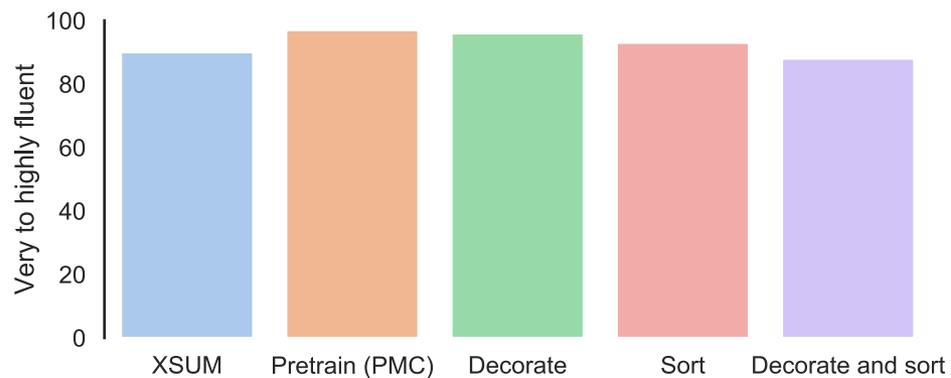
Partially agrees and partially disagrees with the reference summary.

Moderately agrees with the reference summary.

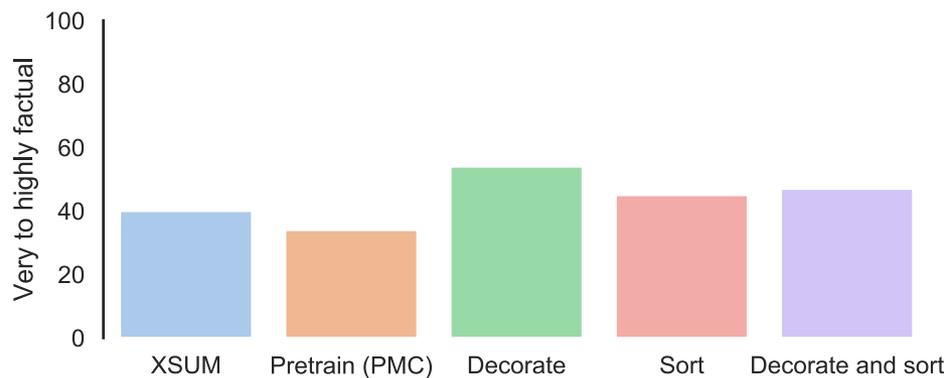
Strongly agrees with reference summary.



relevance



fluency



factuality

# An actual example

*Vaccines against SARS-CoV-2 cannot be recommended for routine clinical practice at this time. There is a need for well-designed RCTs with long-term follow-up to evaluate the efficacy and safety of vaccines against this disease in healthy adults ...*



# TL;DR: IR for biomedical literature

- Potential for real impact and “easy wins”
- Lots of readily accessible data (and tasks)
- Challenges are many: Many relate to how to build IR / language technology that is *actually useful*
- My advice: Work with domain experts to understand needs

# Sources of data

PubMed



Biomedical literature



OhHeyGrrr  
@ohheygrrr

My #COVID19 symptoms update: still congested, sore throat and coughing. The only things I can handle ingesting are buttered white toast and orange juice. My son is doing regular 'welfare checks' on me by pounding on the door and bellowing "WELFARE CHECK!" 😂

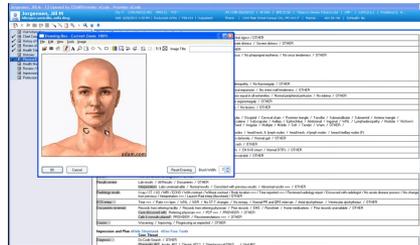
4:07 AM · Jul 11, 2022 · Twitter for iPhone

Social media

MEDITECH



Epic



Electronic Health Records (EHRs)

# Sources of data

PubMed



Biomedical literature



OhHeyGrrr  
@ohheygrrr

My #COVID19 symptoms update: still congested, sore throat and coughing. The only things I can handle ingesting are buttered white toast and orange juice. My son is doing regular 'welfare checks' on me by pounding on the door and bellowing "WELFARE CHECK!" 😂

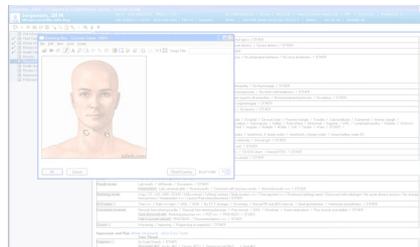
4:07 AM · Jul 11, 2022 · Twitter for iPhone

Social media

MEDITECH

Cerner

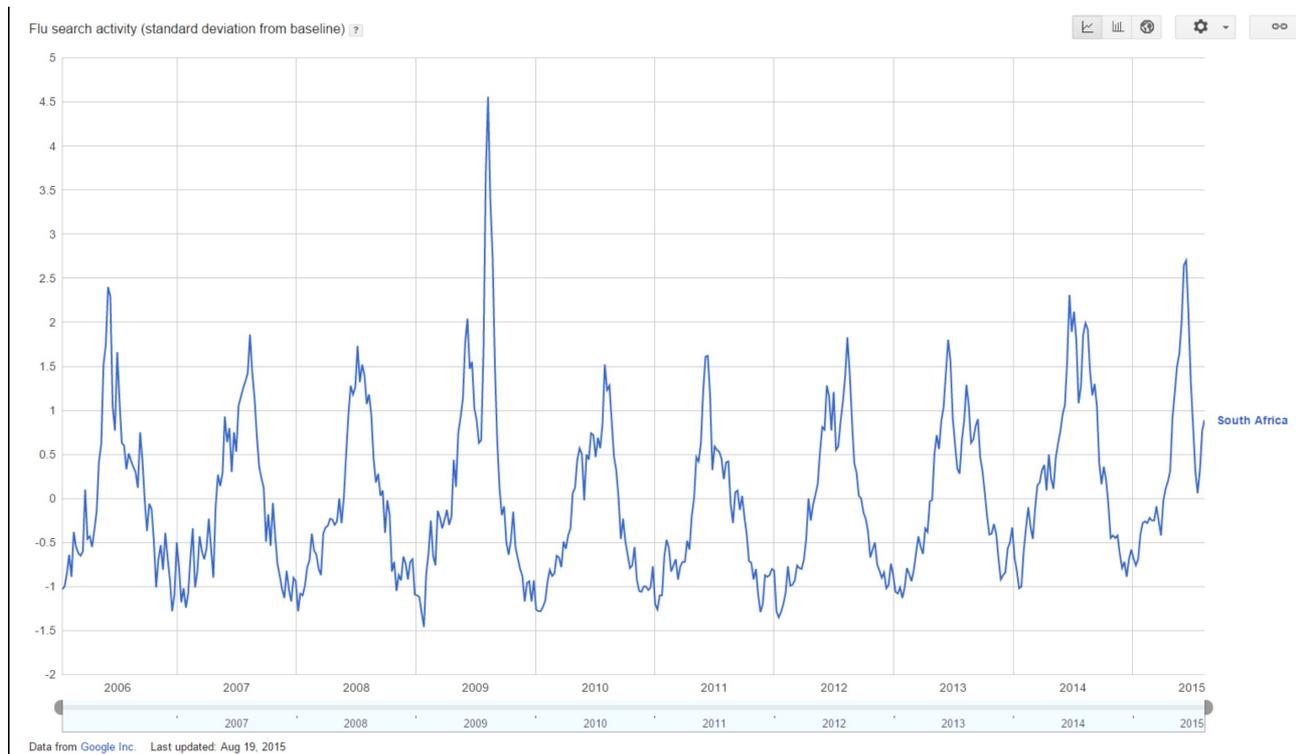
Epic



Electronic Health Records (EHRs)

# Social media + IR/NLP

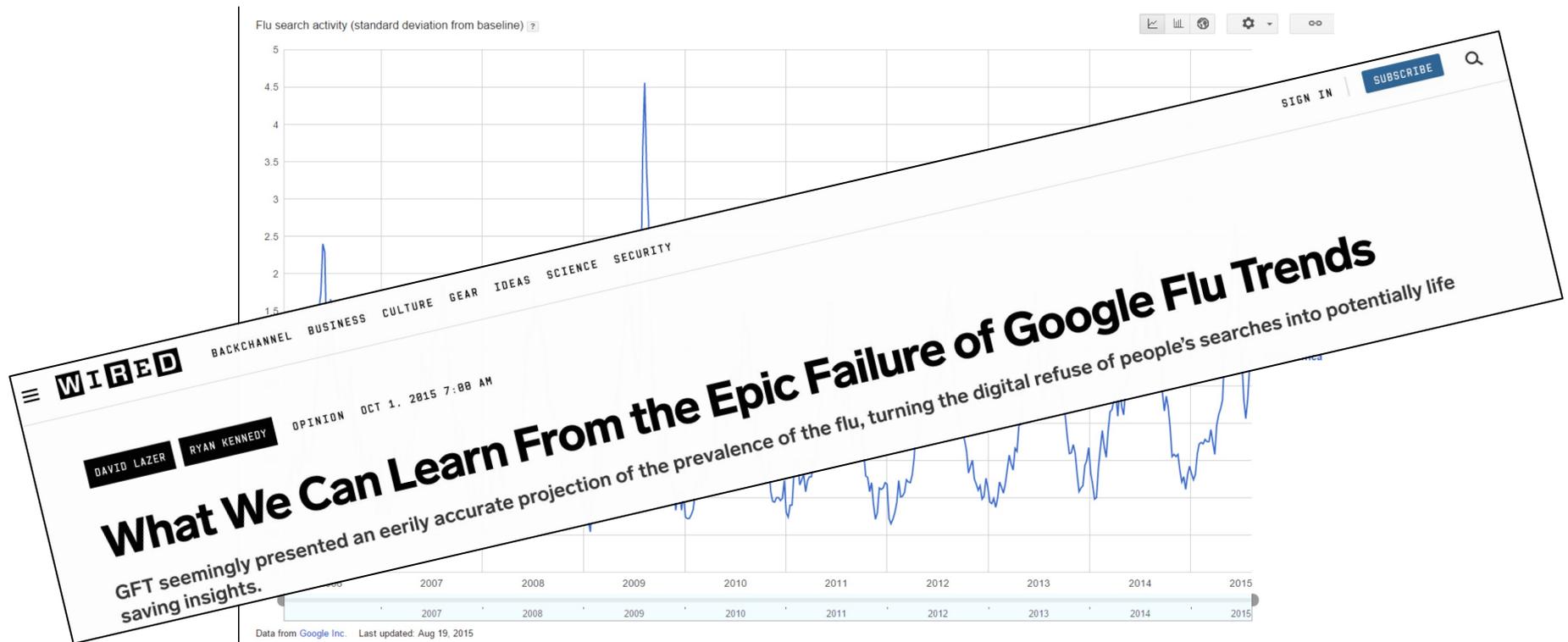
- Lots of work on mining Twitter for **surveillance** purposes (e.g., flu detection), and other epi-type questions



Google flu trends

# Social media + IR/NLP

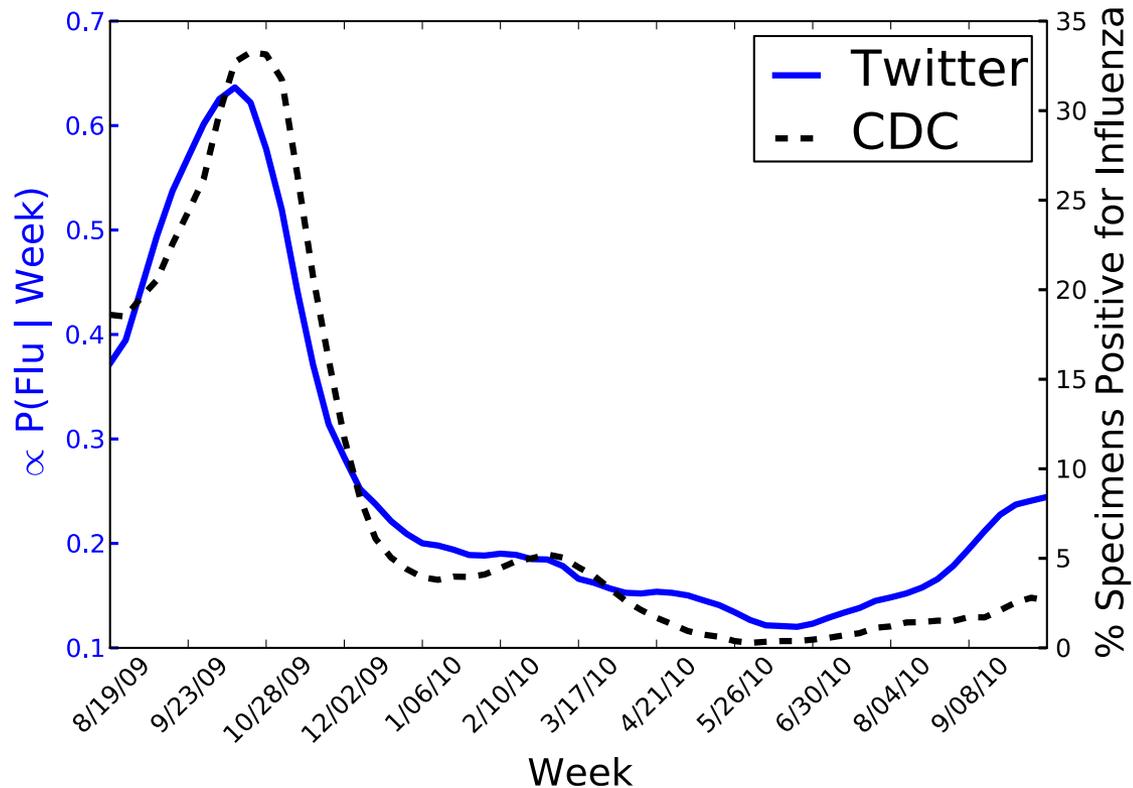
- Lots of work on mining Twitter for **surveillance** purposes (e.g., flu detection), and other epi-type questions



Google flu trends

# You Are What You Tweet: Analyzing Twitter for Public Health

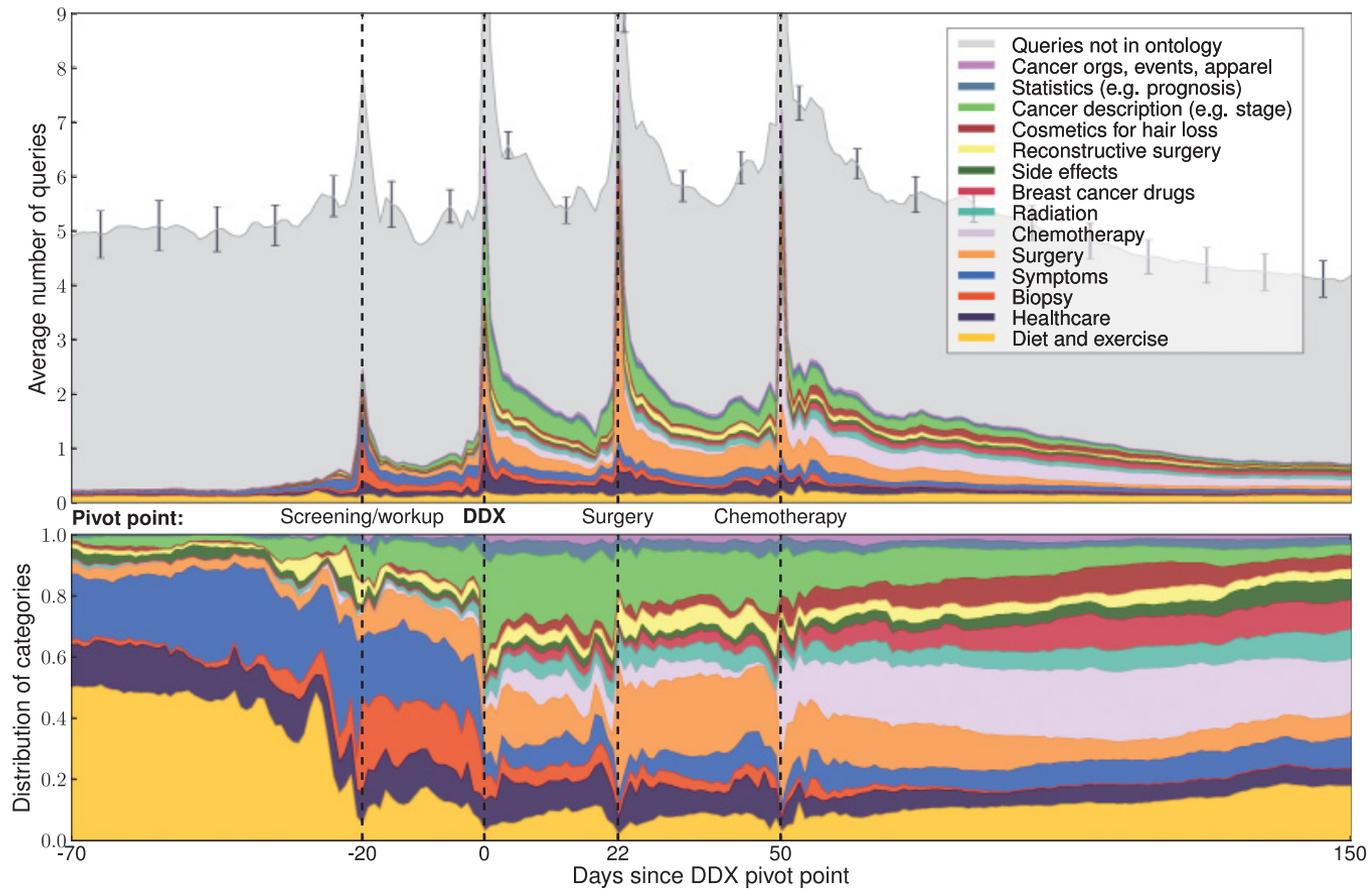
**Michael J. Paul and Mark Dredze**  
Human Language Technology Center of Excellence  
Center for Language and Speech Processing  
Johns Hopkins University  
Baltimore, MD 21218  
{mpaul, mdredze}@cs.jhu.edu



# Search and Breast Cancer: On Episodic Shifts of Attention over Life Histories of an Illness

MICHAEL J. PAUL, Johns Hopkins University

RYEN W. WHITE and ERIC HORVITZ, Microsoft Research



# COVID-19



ACL Anthology

[FAQ](#)

[Corrections](#)

[Submissions](#)

## Workshop on NLP for COVID-19 (NLP-COVID19)

2020

- [Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020](#) **19 papers**
- [Proceedings of the 1st Workshop on NLP for COVID-19 \(Part 2\) at EMNLP 2020](#) **38 papers**

# Fact-checking

## **COVIDLIES: Detecting COVID-19 Misinformation on Social Media**

**Tamanna Hossain<sup>\*,◇</sup> Robert L. Logan IV<sup>\*,◇</sup> Arjuna Ugarte<sup>\*,♠</sup> Yoshitomo Matsubara<sup>\*,◇</sup>  
Sean Young<sup>♠</sup> Sameer Singh<sup>◇</sup>**

<sup>◇</sup> Dept. of Computer Science, University of California, Irvine

<sup>♠</sup> Dept. of Emergency Medicine, University of California, Irvine

{[tthossai](mailto:tthossai@uci.edu), [rlogan](mailto:rlogan@uci.edu), [dugarte](mailto:dugarte@uci.edu), [yoshitom](mailto:yoshitom@uci.edu), [syoung5](mailto:syoung5@uci.edu), [sameer](mailto:sameer@uci.edu)}@uci.edu

# Fact-checking

---

**Tweet:** “Coronavirus CV19 was a top secret biological warfare experiment. That is why it is only affecting the poor.”

**Misconception:** “Coronavirus is genetically engineered.”

**Label:** Agree

---

**Tweet:** “It looks like we are all going to have to wait much longer for a #COVID19 vaccine.”

**Misconception:** “We’re very close to a vaccine.”

**Label:** Disagree

---

**Tweet:** “CDC: Coronavirus spreads rapidly in dense populations with public transit and regular social gatherings.”

**Misconception:** “Coronavirus cannot live in warm and tropical temperatures.”

**Label:** No Stance

---

Figure 1: **COVIDLIES Dataset.** Given a *tweet*, we annotate whether any of the known *misconceptions* are expressed in the tweet, in particular, if the tweet spreads the misconception (e.g., they **Agree**), combats the spread of the misconception (e.g., they **Disagree**), or takes No Stance towards the misconception.

# Health claims on reddit



## r/ibs

I just ordered **Metamucil** bc I read **psyllium may be better for IBS-D**. Or maybe the **fiber** is what is making me go more? **Definitely produces more gas.**

## r/Psychosis

Surprising I'm seeing research articles that **ketamine doesn't increase psychosis risk or induce psychosis** past the duration of the drug. I only took a brief look into it. Has anyone here had ketamine **induced psychosis**? What is r/psychosis experience with ketamine?

## r/Costochondritis

I've had **costo** for a while, usually comes and goes. Done all the heart / lung checks all clear. I've just recovered **covid** and what I'm left with is **chest pain / pressure**. I mean it could be a costo flare up which makes sense, **but also been reading about myocarditis after covid** and I'm worried.

# Annotating health-related reddit posts

Reddit post	Span labels	PIO elements from claims
<i>I've seen a bunch of posts on here from people who say that <b>glycopyrrolate</b> suddenly isn't working anymore for <b>hyperhidrosis</b>. I'm one of those person who has been facing this for a while now. Just wondering if anyone fixed it? Can't really ask my GP about it since he didn't even know the meds existed. He just prescribed them for me when I asked for it</i>	Claim: I've seen a bunch of posts on here from people who say that glycopyrrolate suddenly isn't working anymore for Hyperhidrosis Question: Just wondering if anyone fixed it?	<b>P</b> hyperhidrosis <b>I</b> glycopyrrolate
<i>so i recently read that <b>adderall</b> can trigger a psychotic break &amp; i was prescribed adderall years ago for my <b>adhd</b> but now i just have constant <b>hallucination episodes</b>. anyone else experience adderall induced psychosis?</i>	Claim: so i recently read that adderall can trigger a psychotic break Personal Experience: i was prescribed adderall years ago for my adhd but now i just have constant hallucination episodes Question: anyone else experience adderall induced psychosis?	<b>P</b> adhd <b>I</b> adderall <b>O</b> hallucinations
<i>I've had <b>costochondritis</b> for a while, usually comes and goes. Done all the heart/lung checks all clear. I've just recovered <b>covid</b> and what I'm left with is <b>chest pain</b>/pressure. I mean it could be a costo flare up which makes sense, but also <i>been reading about <b>myocarditis after covid</b></i> and I'm worried, how can I tell which is which?</i>	Claim: been reading about myocarditis after covid Personal Experience: I'm left with is chest pain/pressure Question: how can I tell which is which?	<b>P</b> costochondritis <b>I</b> covid <b>O</b> myocarditis, chest-pain

Table 1: Example annotations, which include: extracted spans (phase 1), and spans describing **Populations**, **Interventions**, and **Outcomes** — PIO elements — within them (phase 2). We collect the latter only for claims.

# Stats on annotations

Population type	# Posts	Average # per population			Average # per claim		
		Questions	Experiences	Claims	Populations	Interventions	Outcomes
<b>Very Common</b> (Dysthymia, Hypothyroidism, Gout, etc)	5467	1101.82	1654.00	114.83	0.82	2.66	3.57
<b>Common</b> (Chronic Fatigue Syndrome, Bulimia, Psychosis, etc)	9539	847.01	1141.72	74.27	1.05	2.95	3.22
<b>Rare</b> (Narcolepsy, Hyperhidrosis, Thyroid Cancer, etc)	7295	1028.50	1166.25	104.75	0.97	2.79	3.81

# Retrieving trustworthy evidence relevant to claims made on reddit

Surprising *I'm seeing research articles that ketamine doesn't increase psychosis risk or induce psychosis* past the duration of the drug. I only took a brief look into it. Has anyone here had ketamine induced psychosis? What is r/psychosis experience with ketamine?



r/psychosis

(A) Extract questions, experiences, and claims

## Questions

Has anyone here had ketamine induced psychosis?  
What is r/psychosis experience with ketamine?

## Personal experiences

None

## Claims

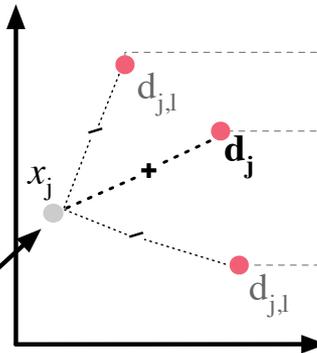
I'm seeing research articles that ketamine doesn't increase psychosis risk or induce psychosis.

(B) Extract PICO elements

**Population**  
*psychosis*

**Interventions**  
*ketamine*

**Outcomes**  
*psychosis*



## Ketamine and Psychosis History: Antidepressant Efficacy and Psychotomimetic Effects Postinfusion

**Abstract:** Because of a theoretical risk of exacerbating psychosis in predisposed patients, subjects with current psychotic symptoms or a past history of psychosis are typically excluded from ketamine trials.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure

(C) Retrieve relevant trustworthy evidence

# Retrieving reliable evidence

**Vitamin D** may prevent **autoimmune disease**

# Retrieving reliable evidence

Vitamin D may prevent autoimmune disease



# Retrieving reliable evidence

Vitamin D may prevent autoimmune disease



Vitamin D [interventions] x Autoimmune disease [outcomes] x Start typing a Population, Intervention, Comparator, or Outcome (PICO) Showing first 500 results only

All (500) Published articles (250) Preprints (0) Registered trials (250) Get large/high quality trials first Newest first

journal article

**Vitamin D and marine omega 3 fatty acid supplementation and incident autoimmune disease: VITAL randomized controlled trial.**

35082139 Hahn et al., BMJ (Clinical research ed.). 2022 10.1136/bmj-2021-066452

25871

For the vitamin D arm, 123 participants in the treatment group and 155 in the placebo group had a confirmed autoimmune disease (hazard ratio 0.78, 95% confidence interval 0.61 to 0.99, P=0.05).

# Similarity between claims and abstracts

$$\text{claim } \mathbf{x}_j = [\mathbf{p} \oplus \mathbf{c}_j \oplus \mathbf{pop}_j \oplus \mathbf{int}_j \oplus \mathbf{out}_j]$$

$$\phi(\mathbf{x}_j, d_l) = E_C(\mathbf{x}_j)^T E_D(d_l)$$

claim encoder

abstract encoder

abstract

# Similarity between claims and abstracts

$$\text{claim } \mathbf{x}_j = [\mathbf{p} \oplus \mathbf{c}_j \oplus \mathbf{pop}_j \oplus \mathbf{int}_j \oplus \mathbf{out}_j]$$

$$\phi(\mathbf{x}_j, d_l) = E_C(\mathbf{x}_j)^T E_D(d_l)$$

claim encoder                      abstract encoder

abstract



# Learning to rank evidence

$$\frac{\text{exp } \phi(x_j, d_j^+)}{\text{exp } \phi(x_j, d_j^+) + \sum_{l=1}^b \text{exp } \phi(x_j, d_{jl}^-)}$$

claim      relevant evidence

$b$  "in-batch" irrelevant abstracts      irrelevant abstract

# Supervision

We need (claim, evidence) supervision, but we don't have it

# Supervision

We need (claim, evidence) supervision, but we don't have it

## Alternative — **distant supervision**

- Start by pairing an annotated claim from Reddit with an arbitrary article in the Trialstreamer database (for which we have PICO elements)
- Replace PICO elements in the claim with those in Trialstreamer:  
Treat the resultant pair as a match

# Distant supervision

Surprising I'm seeing research articles that [INT] causes [OUT]  
past the duration of the drug

# Distant supervision

Surprising I'm seeing research articles that [INT] causes [OUT]  
past the duration of the drug



Surprising I'm seeing research articles that olanzapine causes discontinuation  
rate past the duration of the drug

# Results (assessed by MDs)

<b>Cumulative # of relevant abstracts @<math>k</math></b>				
<b>k</b>	<b>1</b>	<b>3</b>	<b>5</b>	<b>10</b>
<i>Pre-trained DPR (Karpukhin et al., 2020)</i>				
<b>Relevant</b>	6	16	29	58
<b>Somewhat relevant</b>	14	39	66	135
<b>Irrelevant</b>	80	245	405	807

# Results (assessed by MDs)

<b>Cumulative # of relevant abstracts @<math>k</math></b>				
<b>k</b>	<b>1</b>	<b>3</b>	<b>5</b>	<b>10</b>
<i>Pre-trained DPR (Karpukhin et al., 2020)</i>				
<b>Relevant</b>	6	16	29	58
<b>Somewhat relevant</b>	14	39	66	135
<b>Irrelevant</b>	80	245	405	807
<i><b>RedHOT-DER</b> trained on pseudo data</i>				
<b>Relevant</b>	18	62	101	201
<b>Somewhat relevant</b>	17	49	87	193
<b>Irrelevant</b>	65	189	312	606

# Sources of data

PubMed



Biomedical literature



OhHeyGrrr  
@ohheygrrr

My #COVID19 symptoms update: still congested, sore throat and coughing. The only things I can handle ingesting are buttered white toast and orange juice. My son is doing regular 'welfare checks' on me by pounding on the door and bellowing "WELFARE CHECK!" 😂

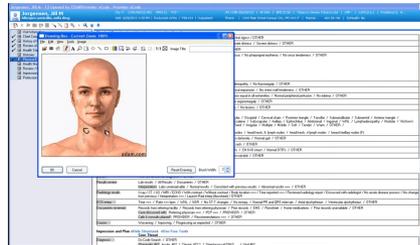
4:07 AM · Jul 11, 2022 · Twitter for iPhone

Social media

MEDITECH



Epic



Electronic Health Records (EHRs)

# Sources of data

PubMed



Biomedical literature



OhHeyGrrr  
@ohheygrrr

My #COVID19 symptoms update: still congested, sore throat and coughing. The only things I can handle ingesting are buttered white toast and orange juice. My son is doing regular 'welfare checks' on me by pounding on the door and bellowing "WELFARE CHECK!" 😬

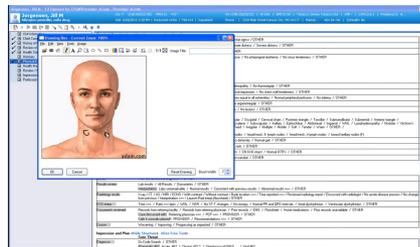
4:07 AM - Jul 11, 2022 - Twitter for iPhone

Social media

MEDITECH



Epic



Electronic Health Records (EHRs)

THE  
NEW YORKER

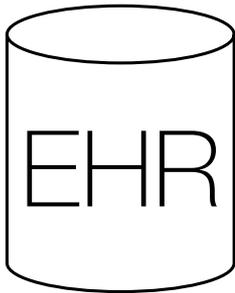
ANNALS OF MEDICINE NOVEMBER 12, 2018 ISSUE

WHY DOCTORS HATE  
THEIR COMPUTERS

*Digitization promises to make medical care easier and more efficient. But are screens coming between doctors and patients?*

By Atul Gawande

November 5, 2018



powered by

**Epic**

# IR + medical records

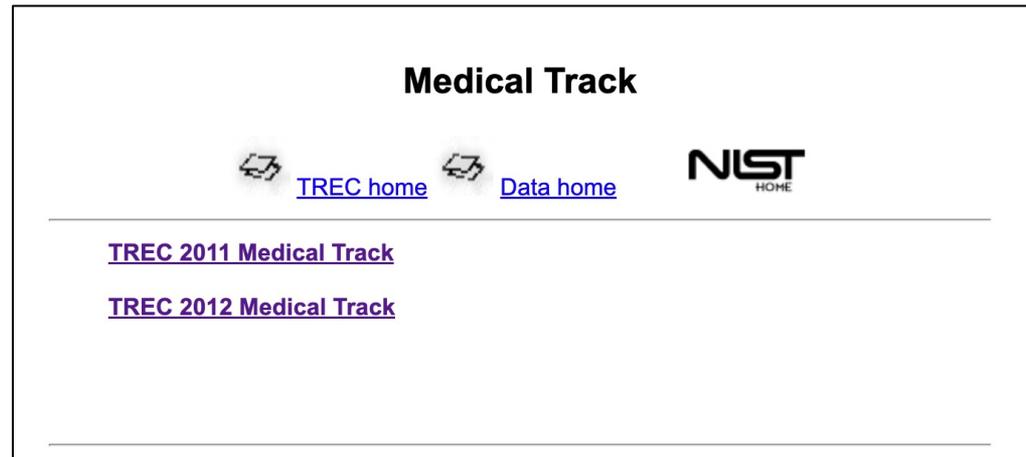
Possible (example!) tasks

- Retrieve snippets from notes within a given patients' records relevant to a given query (or image!)
- Retrieve patients eligible for *clinical studies*

# IR + medical records

## Possible (example!) tasks

- Retrieve snippets from notes within a given patients' records relevant to a given query (or image!)
- Retrieve patients eligible for *clinical studies*



# Example: Extractive query-focused summarization of EHR to aid diagnosis

*I'll present our work on this, but there is of course a long line of work on summarizing EHR; a nice (if slightly dated) survey:*

## Automated methods for the summarization of electronic health records

Rimma Pivovarov and Noémie Elhadad

RECEIVED 30 October 2014  
REVISED 15 February 2015  
ACCEPTED 15 March 2015  
PUBLISHED ONLINE FIRST 16 April 2015

**AMIA**  
INFORMATICS PROFESSIONALS. LEADING THE WAY.

**OXFORD**  
UNIVERSITY PRESS

### ABSTRACT

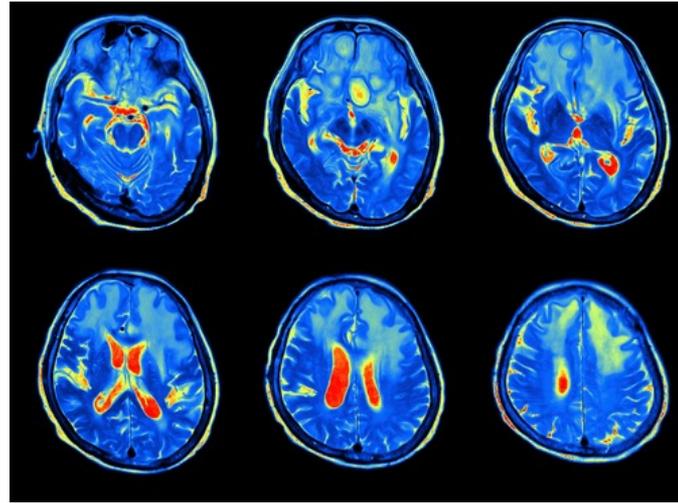
**Objectives** This review examines work on automated summarization of electronic health record (EHR) data and in particular, individual patient record summarization. We organize the published research and highlight methodological challenges in the area of EHR summarization implementation.

**Target audience** The target audience for this review includes researchers, designers, and informaticians who are concerned about the problem of information overload in the clinical setting as well as both users and developers of clinical summarization systems.

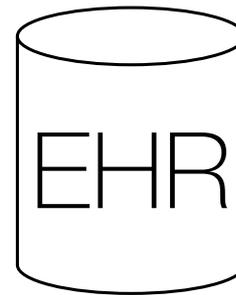
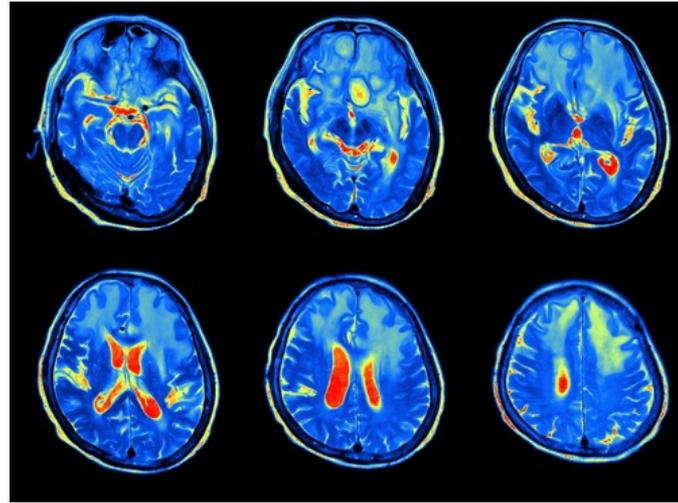
**Scope** Automated summarization has been a long-studied subject in the fields of natural language processing and human-computer interaction, but the translation of summarization and visualization methods to the complexity of the clinical workflow is slow moving. We assess work in aggregating and visualizing patient information with a particular focus on methods for detecting and removing redundancy, describing temporality, determining salience, accounting for missing data, and taking advantage of encoded clinical knowledge. We identify and discuss open challenges critical to the implementation and use of robust EHR summarization systems.

**Keywords:** clinical summarization, electronic health records, natural language processing, missing data, temporality, semantic similarity

# Motivation



# Motivation



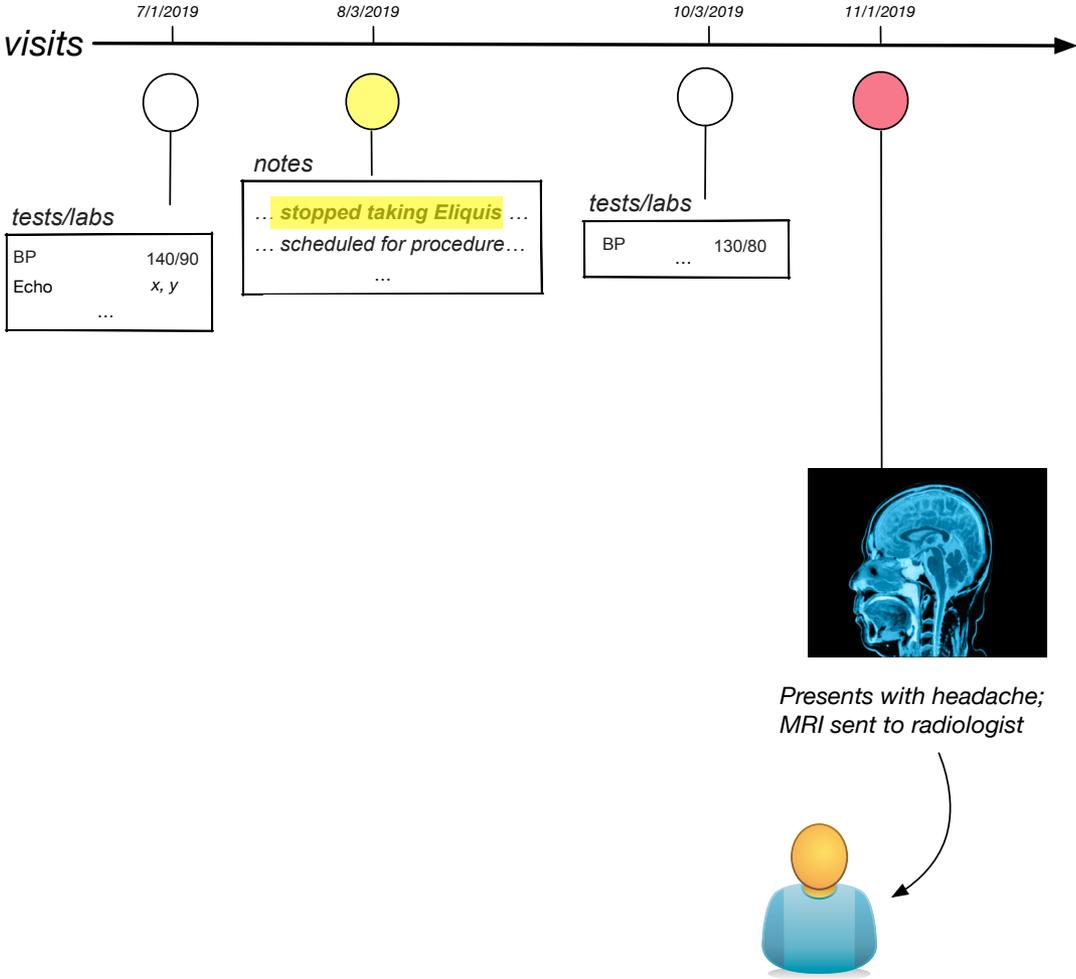
powered by

***Epic***

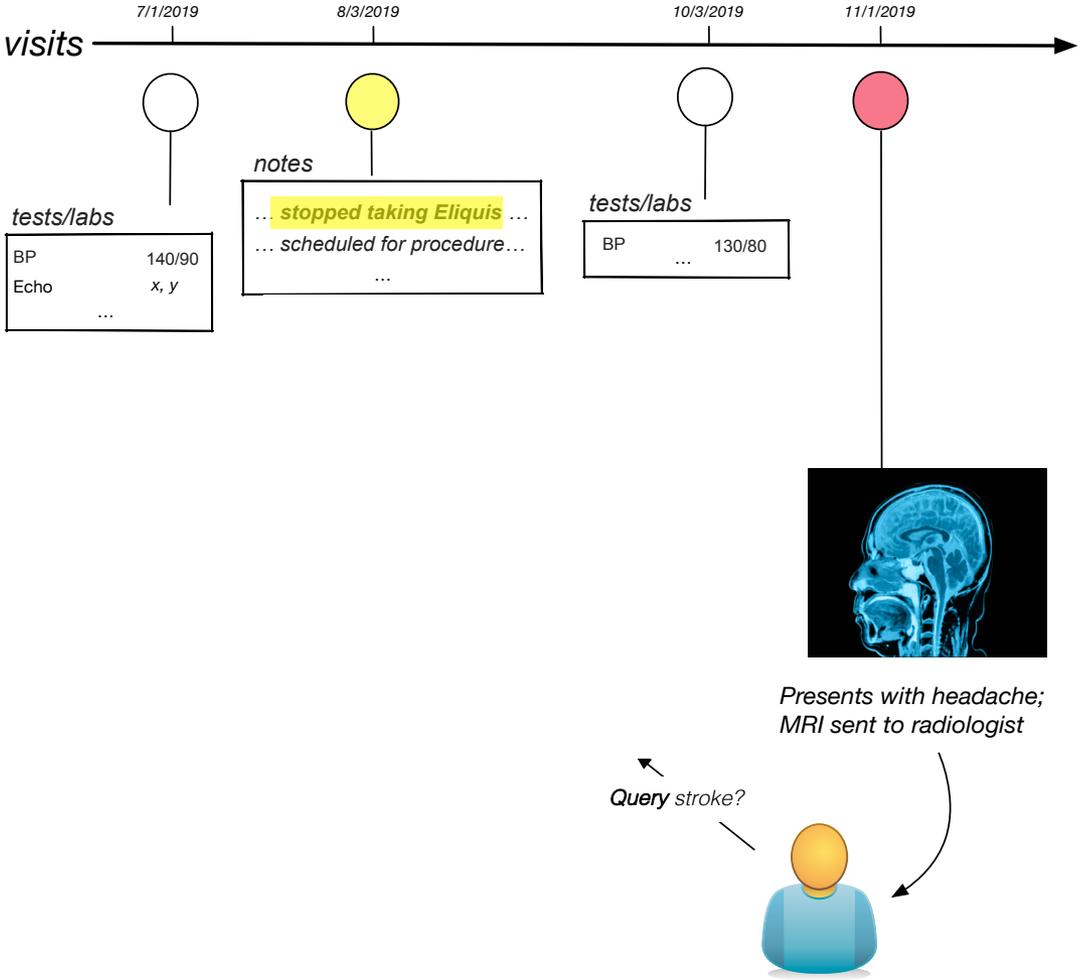
*Aim Train a model to provide extractive summaries relevant to a given query (potential diagnosis)*



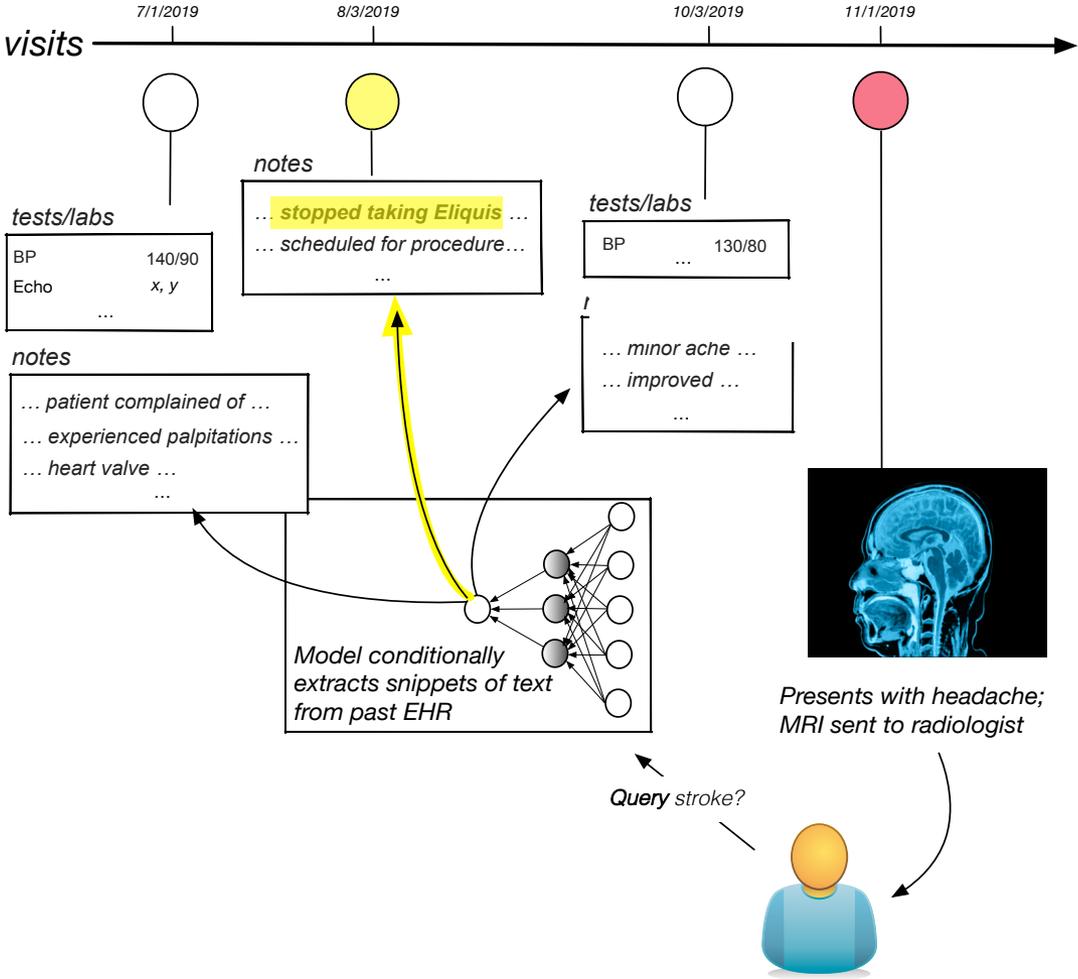
# Patient EHR



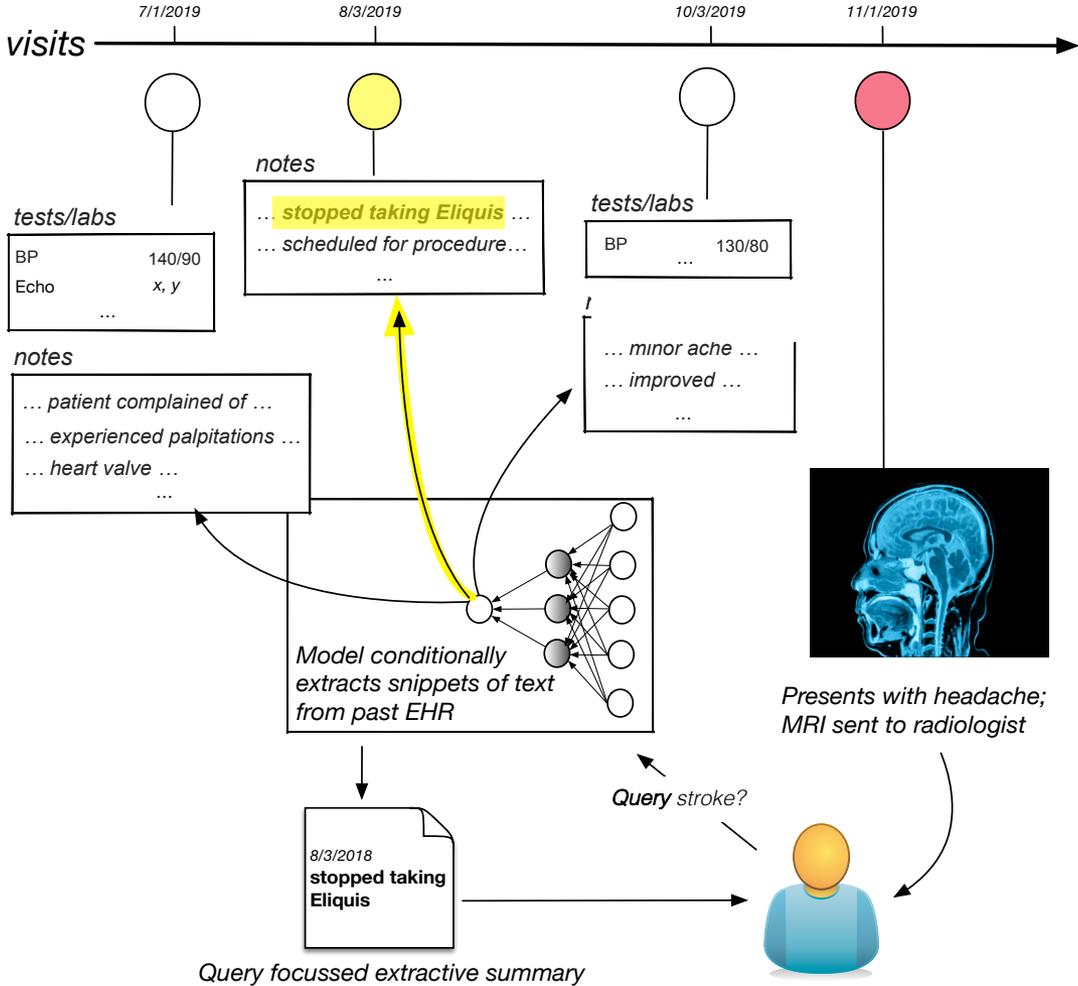
# Patient EHR



# Patient EHR



# Patient EHR



Query (potential diagnosis)



$(\mathbf{x}, q, \mathbf{a})$



Patient EHR



Relevance scores

# Unsupervised retrieval approach

$$B_{\text{mean}}(z) = \text{mean}(f_{\phi}^{\text{transformer}}(z))$$

$$\mathbf{a}_i = \text{Cosine}(B_{\text{mean}}(\mathbf{x}_i), B_{\text{mean}}(\mathbf{d}_q))$$

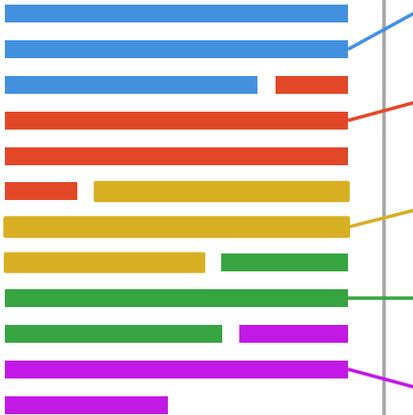
# Distant supervision (again!)

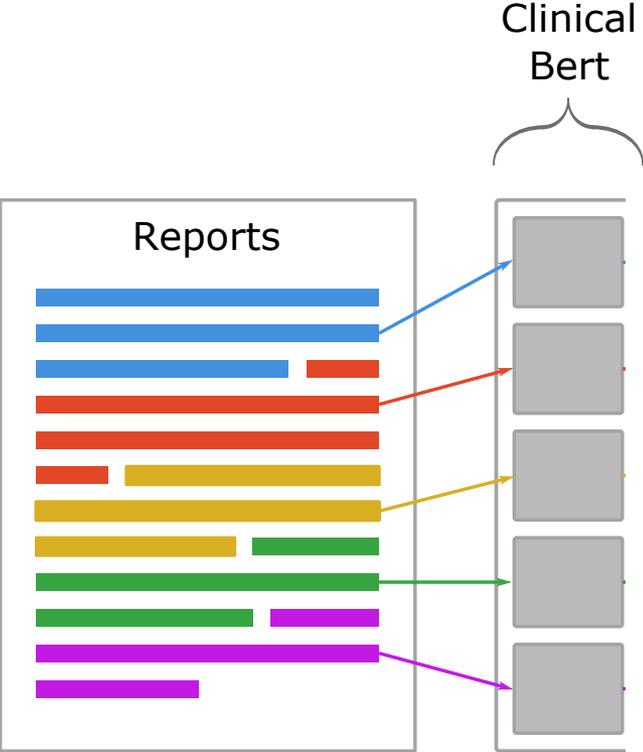
- Radiologists on our team enumerated sets of ICD codes that correspond to diagnoses of interest

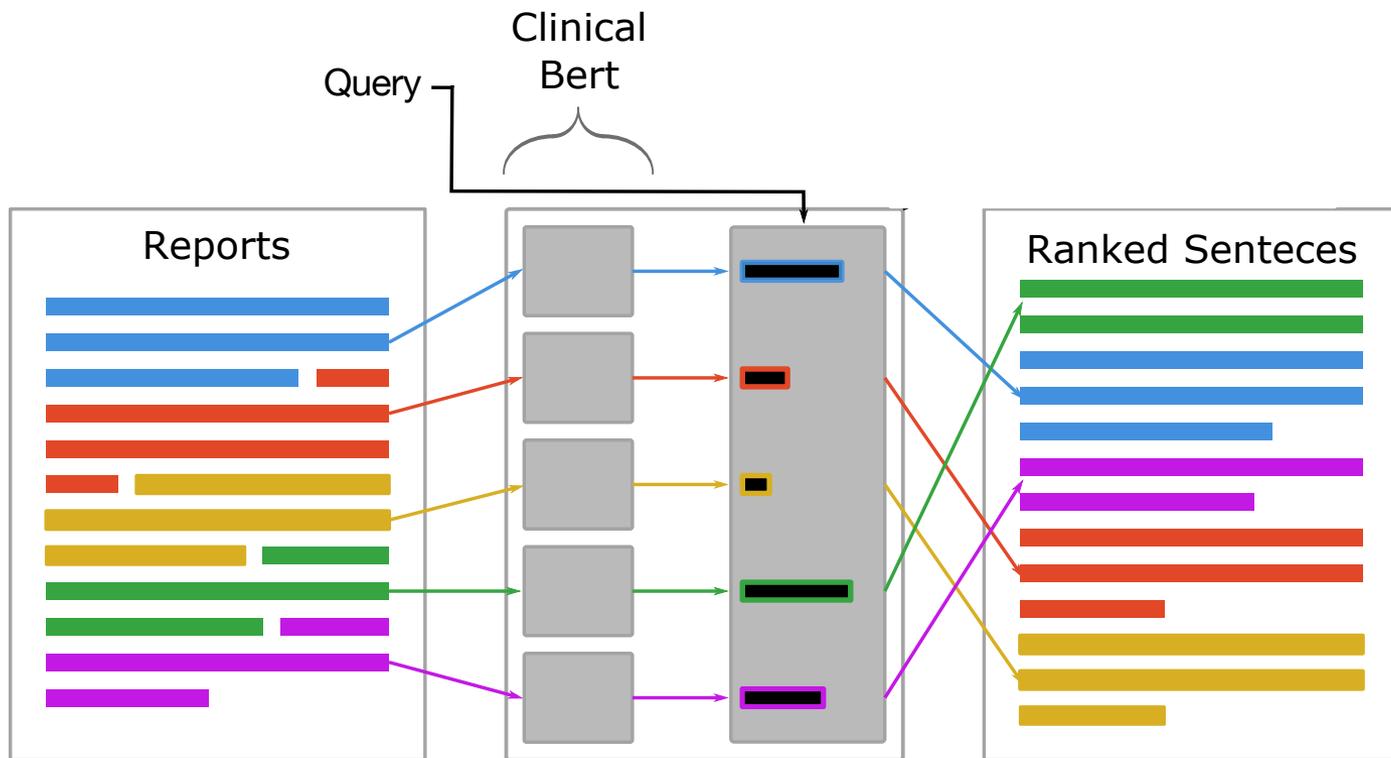
# Distant supervision (again!)

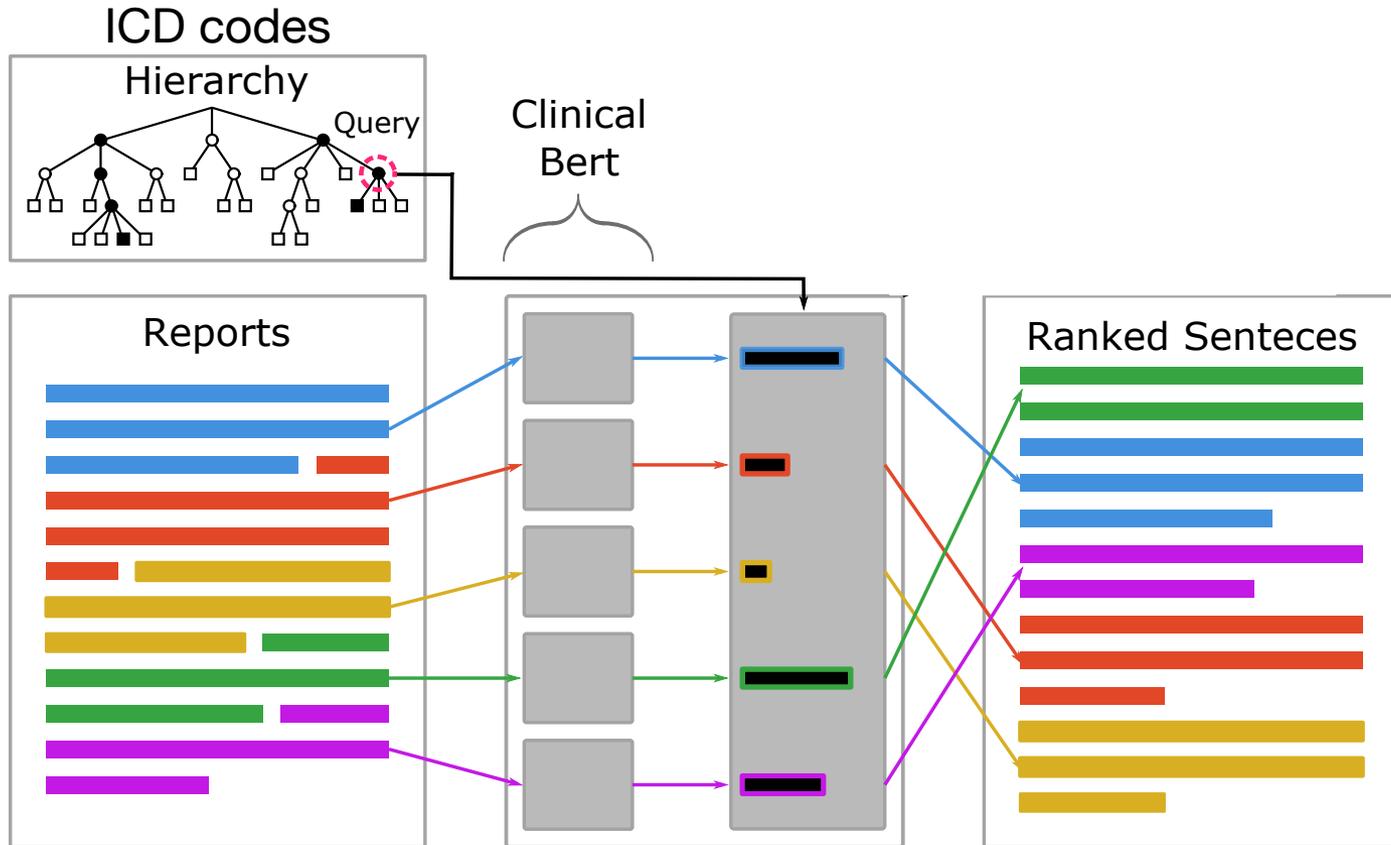
- Radiologists on our team enumerated sets of ICD codes that correspond to diagnoses of interest
- We train a model to predict *future* occurrences of these codes (but this is only a proxy task)

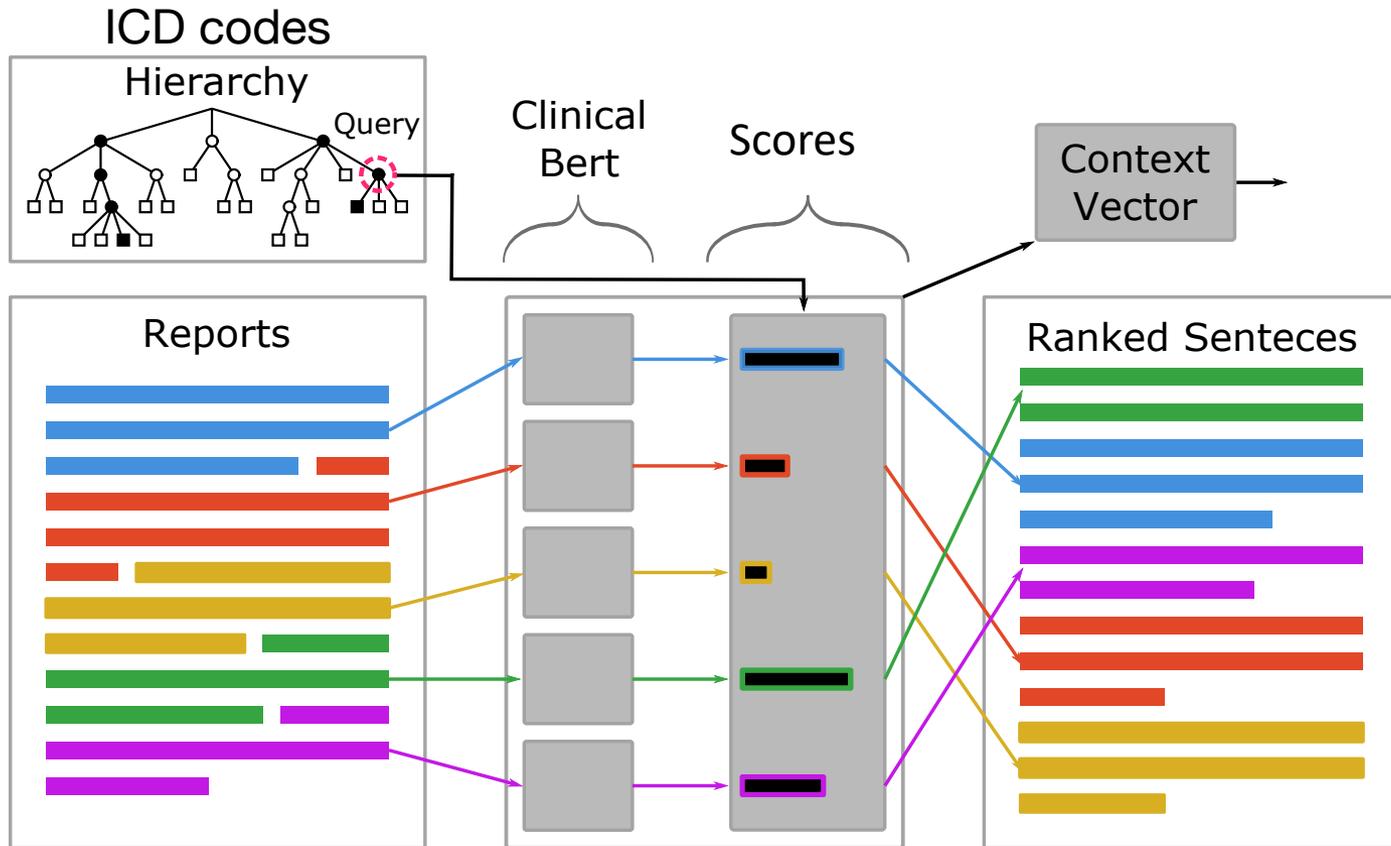
## Reports

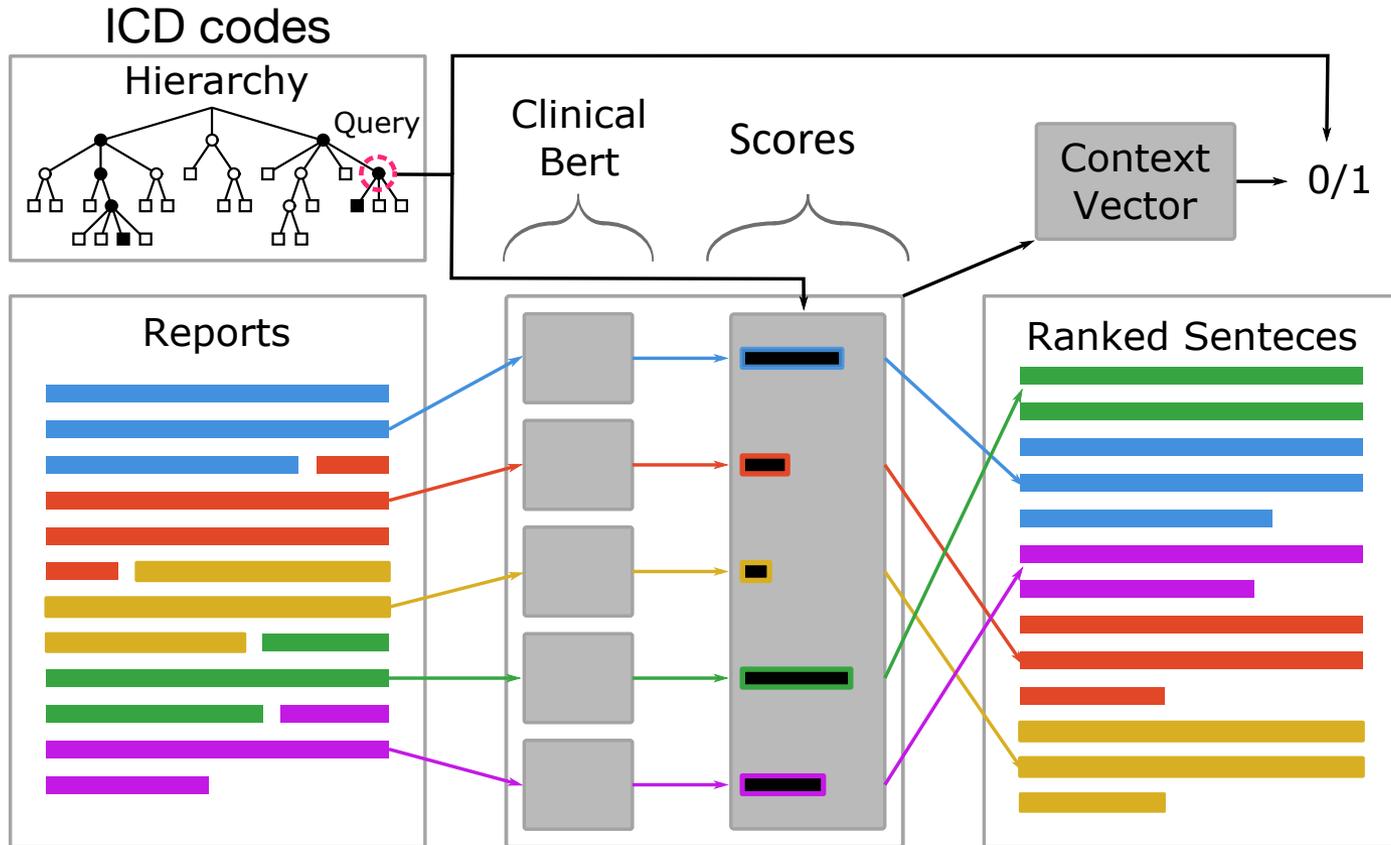












$$\mathbf{a}_i = \frac{\exp(B_{\text{cls}}(\mathbf{x}_i) \cdot \mathbf{e}_q)}{\sum_{i'} \exp(B_{\text{cls}}(\mathbf{x}_{i'}) \cdot \mathbf{e}_q)}$$

$$\mathbf{a}_i = \frac{\exp(B_{\text{cls}}(\mathbf{x}_i) \cdot \mathbf{e}_q)}{\sum_{i'} \exp(B_{\text{cls}}(\mathbf{x}_{i'}) \cdot \mathbf{e}_q)}$$

$$P(y = 1 | \mathbf{x}, q) = \sigma\left(\mathbf{U}_2 \text{ReLU}\left(\mathbf{U}_1 \left[\sum_i a_i B_{\text{cls}}(\mathbf{x}_i), \mathbf{e}_q\right] + \mathbf{b}_1\right) + \mathbf{b}_2\right)$$

# Embedding queries $\mathbf{e}_q$

Indicator one-hot

# Embedding queries $e_q$

Indicator one-hot

Description pass natural language description through BERT

# Embedding queries $e_q$

**Indicator** one-hot

**Description** pass natural language description through BERT

**Hierarchical** concatenate all descriptions of codes from the top of the ICD tree to the query code

# Embedding queries $\mathbf{e}_q$

Indicator one-hot

Description pass natural language description through BERT

Hierarchical concatenate all descriptions of codes from the top of the ICD tree to the query code

$$\mathbf{e}_q^{\text{hierarchy}} = B_{\text{cls}}([\text{CLS}], \mathbf{d}_{p_1^{(q)}}, [\text{SEP}], \mathbf{d}_{p_2^{(q)}}, \dots, [\text{SEP}], \mathbf{d}_{p_L^{(q)}}])$$

Evaluation

# Annotate

4 / 1123

Previous

Next

Patient File: instance\_335.pkl

Patient MRN: 15930

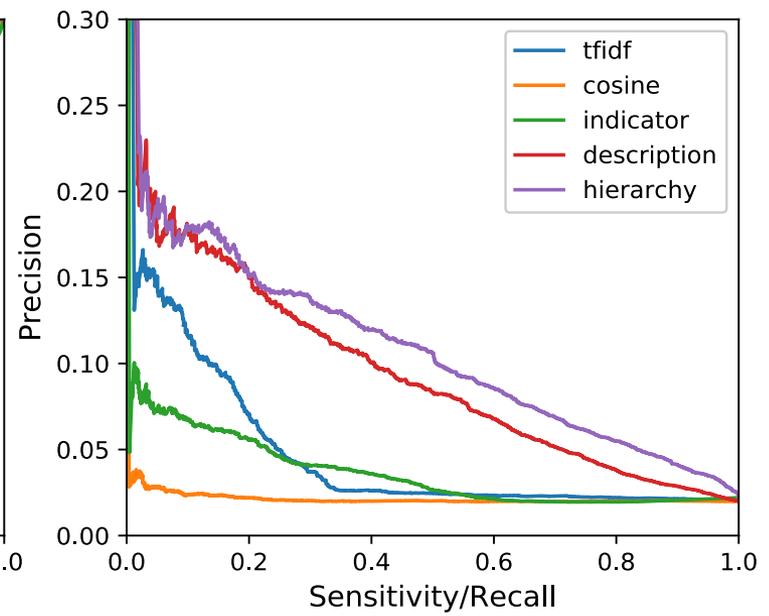
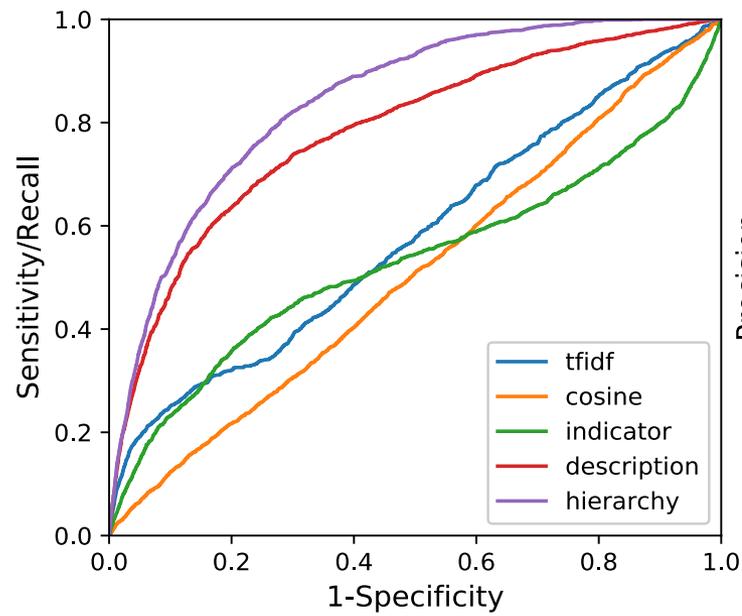
Future Reports

Past Reports

annotate the last 1000 sentences before the first mr in the past reports

Reports	Tags
<p>1/15</p> <p>Previous Radiology, Sun Apr 10 21:48 Next</p> <p>reason. Evaluate ischemic bowel, mv thrombosis and postal vein throm</p> <p>Admitting Diagnosis: MULTISYSTEM ORGAN FAILURE</p> <p>Field of view: 36 Contrast: OPTIRAY Amt: 150</p> <hr/> <p>[**Hospital 2**] <u>MEDICAL CONDITION:</u></p> <p><u>70 year old woman with MV/portal vein thrombosis w/ SB/colonic ischemia</u></p> <p><u>REASON FOR THIS EXAMINATION:</u></p> <p>Evalaute ischemic bowel, MV thrombosis and postal vein thrombosis</p>	<p>420-429: Other Forms O <span>Deselect</span></p> <p>Vascular <span>▲</span> 420-429: Other Forms O <span>▲</span></p> <p>Select a Tag <span>▲</span></p> <hr/> <p><b>420-429: Other Forms Of Heart Disease</b> <span>×</span></p> <hr/> <p>[CLS] ** ] medical condition : 70 year old woman with mv / portal vein thrombosis w / sb / colonic ischemia reason for this examination : [SEP] <span>×</span></p>

**Submit**



# Some take-aways

IR (and language technologies more generally) can realize considerable impact in health and medicine.

There are a rich set of resources (data and tasks) to get started in this space.

Supervision at scale often hard here: Often need to be creative for training.

Working with *domain experts* to frame problems in a useful way is critical, more so than using the deepest networks available.