



711th Human Performance Wing

NLP Covid 2020 Briefing

Dr. Ryan Kramer, Principal Investigator

Kyle Brown, Eric Schmidt, Kevin O'Neill, Tyler Whitlock, Wesley Holmes

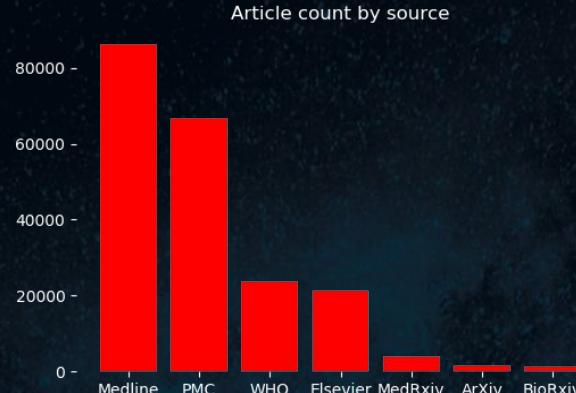
Summer of TDA 2020

August 7th, 2020

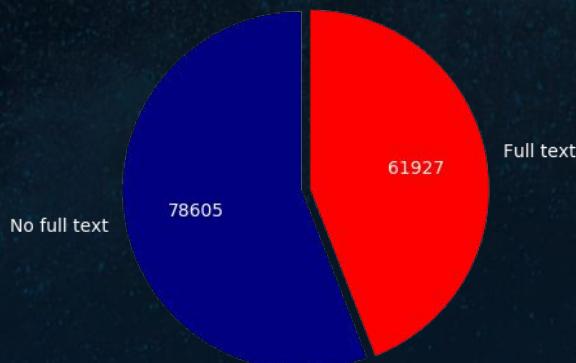
**Airman Systems Directorate
Battlespace Visualization Branch**

- Goal: Create dashboard(s) to enable researchers to quickly and efficiently search through the CORD-19 dataset to find relevant information
 - Documents, sentences, words
- CORD-19 dataset introduction
 - dataset statistics, paper counts etc
- Utility of each portion and their creation
 - Documents
 - Sentences
 - Words
- Use cases and conclusion from search
 - Remdesivir
 - Hydroxychloroquine

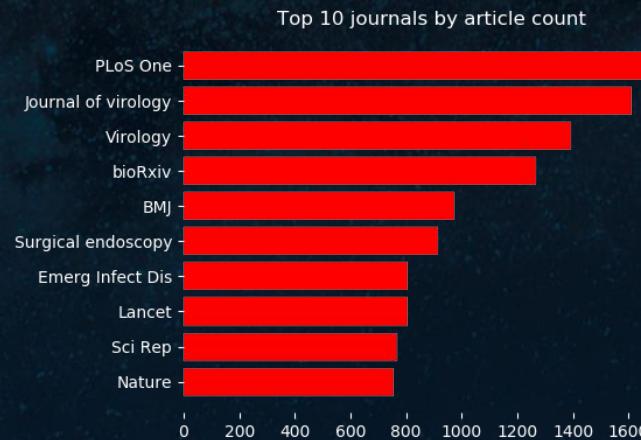
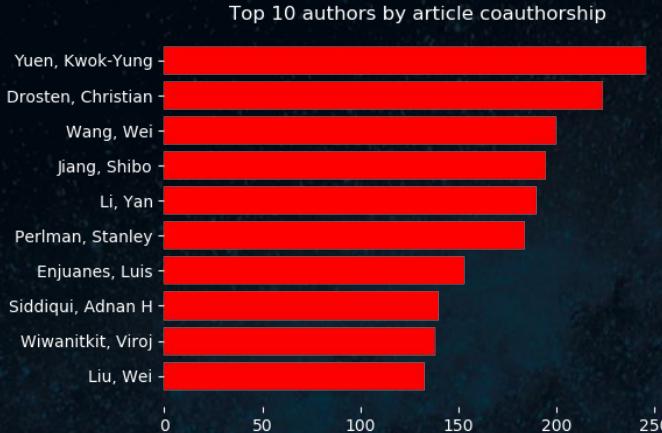
- The COVID-19 Open Research Dataset
 - over 140,000 scholarly articles published in peer-reviewed publications and archival services such as bioRxiv, medRxiv, and arXiv
- Papers covering research on COVID-19, as well as all coronaviruses such as SARS and MERS.
- Full text available for many articles
 - We restrict ourselves to articles with full text



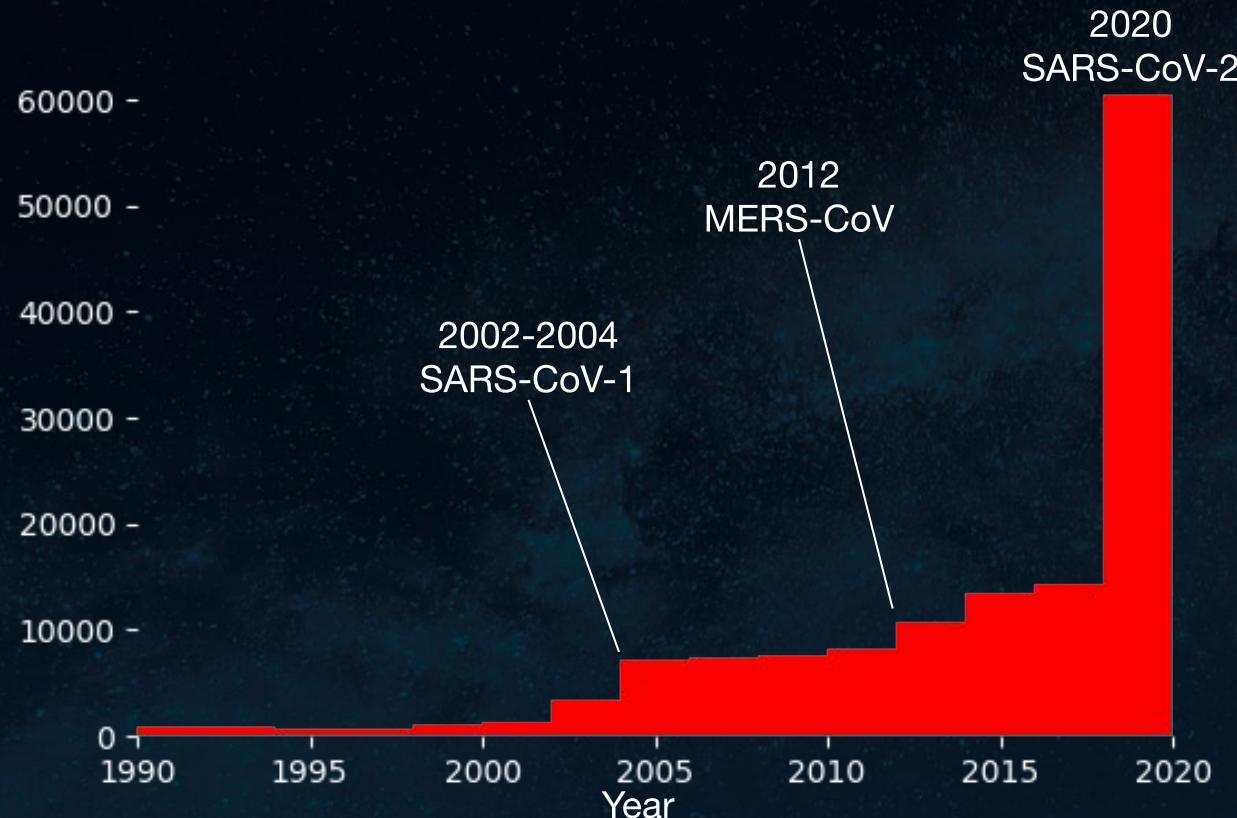
CORD19 corpus by full-text availability



- Additional metadata
 - Authors
 - 443,005 unique authors
 - Date of publication
 - Range: 1870-2020
 - Publication venue
 - 16,006 unique journals
- Updated daily with new publications



Articles in CORD-19 Dataset by Year

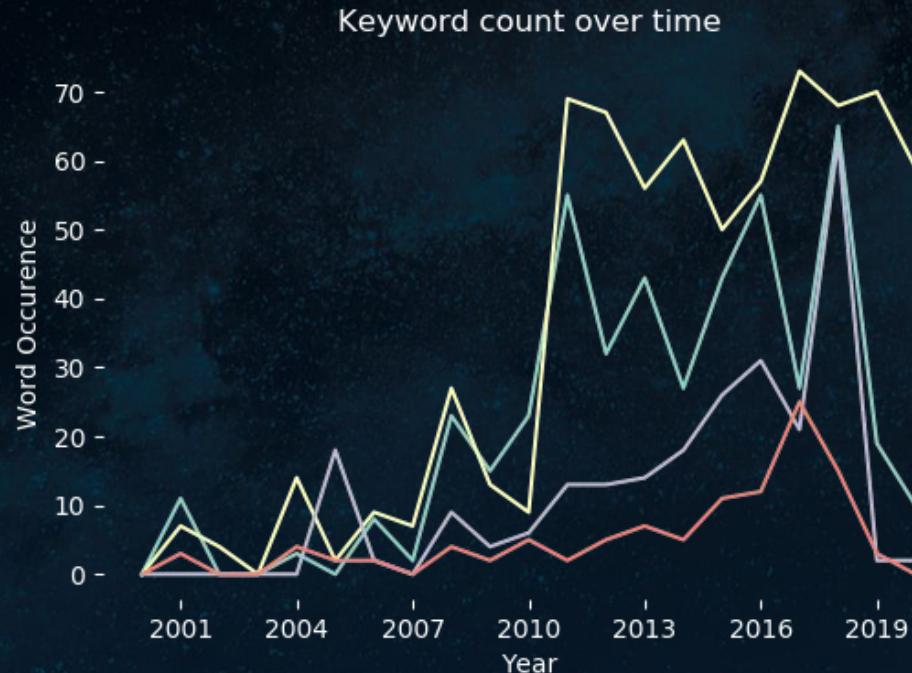


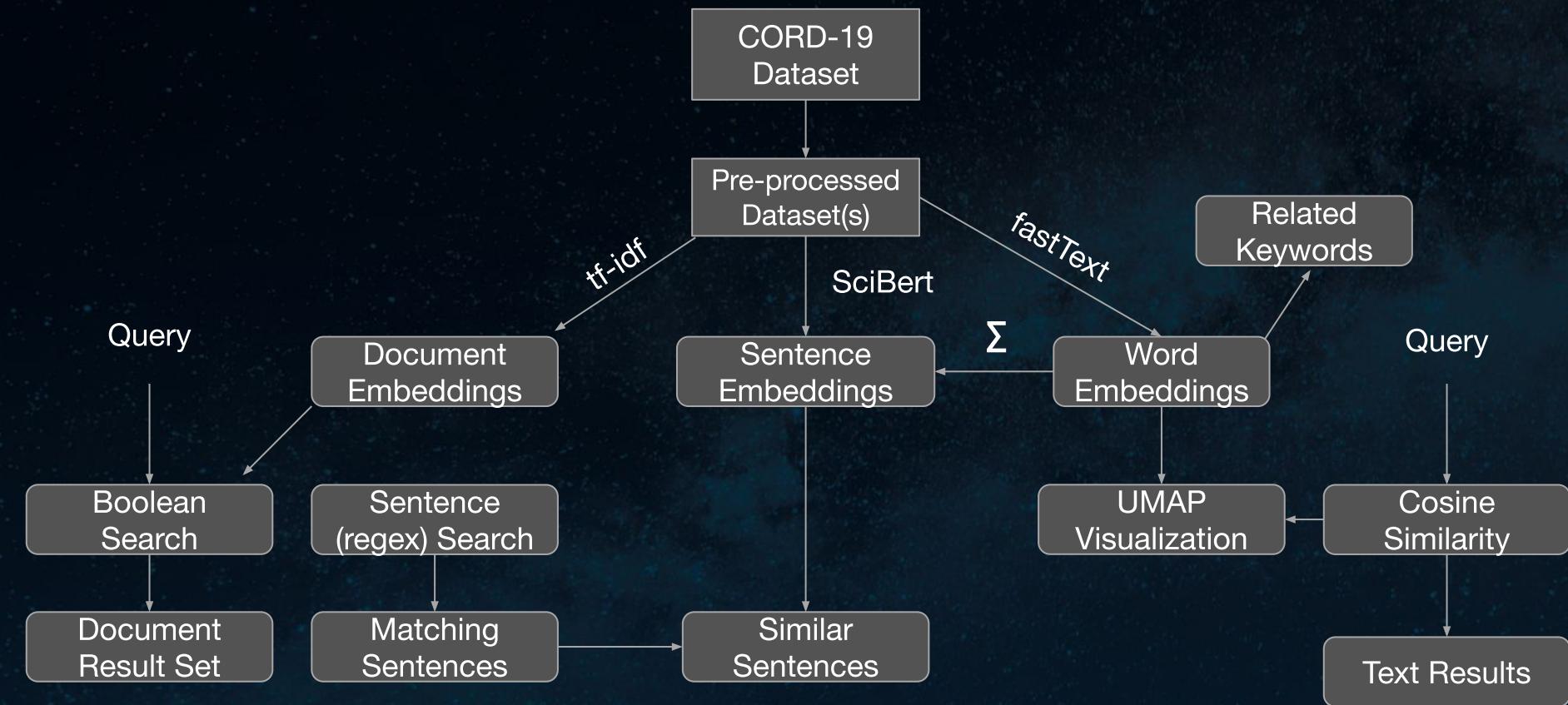
- Corpus can be used to analyze research trends over time
 - Algorithm to tabulate word usages across documents
 - Sparse matrix is memory-efficient, can calculate other NLP statistics relatively quickly

Example

Coronavirus genera
term usage over time

- Alphacoronavirus
 - Mammal virus, includes some varieties of common cold
- Betacoronavirus
 - Mammal virus, includes SARS, MERS, SARS-CoV2
- Gamma & Deltacoronavirus
 - Primarily avian viruses





- **Tokenization:** splitting the document into a sequence of *tokens*, which can be words, phrases, or sentences. We use word tokenization
- **Lemmatization:** removing inflections and other changes to words to get a base form
 - Examples: “cats” => “cat”; “broke” => “break”; “lingering” => “linger”
- **Stop-word removal:** exclude common words with little semantic content such as “the” and “with”
 - Produces smaller and more useful tf-idf embeddings
- Other processing steps
 - Remove overly short documents that may represent summaries, abstracts, or indexes
 - Convert to all lowercase and remove punctuation after tokenization
 - Remove selected words related to publishing, copyright, bibliography, etc.

- fastText
 - A library that uses neural network models for text classification and word representations
 - We use the word representations for finding similar keywords
 - Trained and ran entirely on CPU - able to use in real time
- SciBERT
 - BERT - Bidirectional Encoder Representations from Transformers
 - BERT is an unsupervised language representation model for NLP
 - SciBERT - a BERT model trained on a full text papers from semanticscholar.org
 - Used to build sentence embeddings which allow comparison of sentences even if they share no words

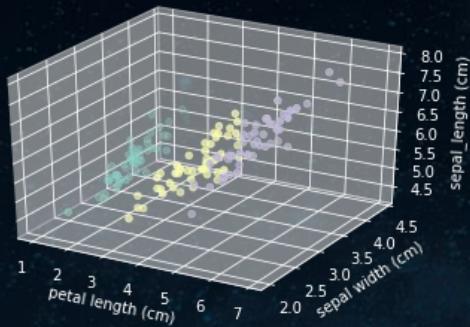
- Exploratory analysis with two dashboards: use of certain medications in treating COVID-19
 - What assertions are made about the efficacy of the medications?
 - How certain is the article about these assertions?
 - What are potential risks of the medications (side effects)?
- Look at two separate medications
 - Remdesivir
 - A general anti-viral medication
 - Hydroxychloroquine
 - Used in treatment of malaria and lupus

- The two dashboards utilize different approaches to find results
- Dashboard 1 (BERT) - Focuses on sentences and tailoring the results from a larger portion of the dataset
 - SciBERT embedding visualization and backbone
 - Built-in regex queries for results
 - Ideal use: Specific inquiries
- Dashboard 2 (tf-idf) - Focuses on narrowing down the relevant documents then returning sentences from that and further refinement
 - tf-idf document search
 - Keyword graph using fastText
 - Ideal use: Users with background knowledge in subject

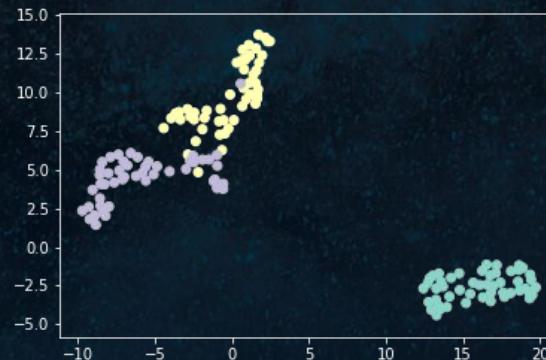
- BERT Dashboard Demo
 - Integrating some general data visualization with FastText and SciBERT sentence embeddings
 - Good for specific analysis of a research question
 - Visualization of sentence embeddings from SciBERT
 - Freely select from graph to examine potential clusters
- Use case: Remdesivir
- Exploratory investigation - could this be a valid treatment?
 - What does the literature say?
 - Pro's and Con's

- Pre-trained language representation model on **scientific** documents
 - Outperforms BERT (2)
 - Slightly outperforms BioBERT on 2 datasets and similar performance on 1 dataset (2)
- Create more **relevant** sentence embeddings through additional **context** provided by training on scientific documents
- Pytorch implementation for improved processing speed
- 768 dimensions for every sentence

- A general **non-linear** dimensionality reduction algorithm useful for visualization
- Better visualization results and faster processing time than alternatives such as t-SNE



*Plot of 3 of 4 dimensions
in Iris dataset*



*4 dimensions reduced to
2 with UMAP*

- Antiviral medication, sold under brand name Veklury
- Being tested as a treatment for COVID-19

Gilead says remdesivir coronavirus treatment reduces risk of death in severely sick patients

CNBC · Jul 10

- Remdesivir and coronavirus: More good news about antiviral COVID drug

USA TODAY · Jul 10

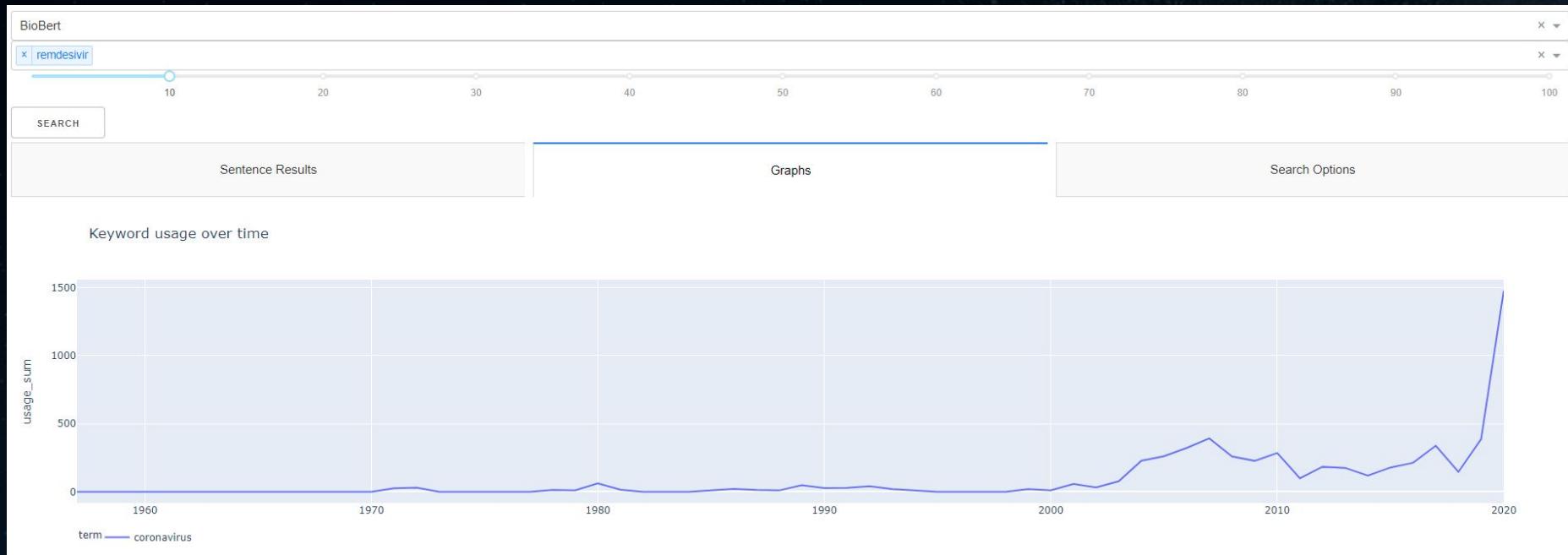
[View Full Coverage](#)

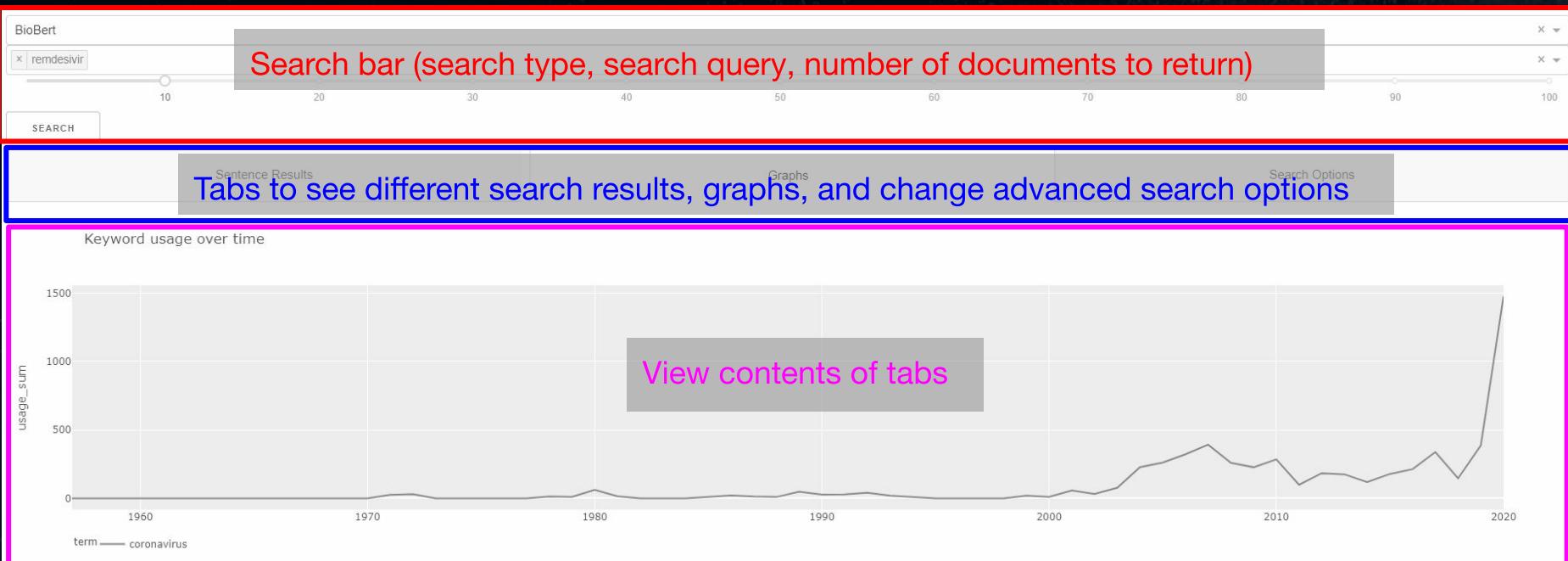


International Panel Issues Weak Recommendation for Remdesivir in Severe COVID-19

Drug Topics · 3 days ago







BioBert

remdesivir

10

20

30

40

50

60

70

80

90

X ▾

X ▾

100

SEARCH

Sentence Results

Graphs

Search Options

View sentences which are returned for a given query

Infected and mock-infected human lung epithelial cells (Calu 3) were treated with dilution series of lopinavir (50 μ M–0.05 μ M), ritonavir (50 μ M–0.05 μ M), interferon-beta (2800 IU/mL to 5.5 IU/mL), or remdesivir (10 μ M–0.02 μ M) and results were assayed 48 h post infection.

Title:
Advances in respiratory virus therapeutics – A meeting report from the 6th isiv Antiviral Group conference
doi: [10.1016/j.antiviral.2019.04.006](https://doi.org/10.1016/j.antiviral.2019.04.006)

Authors:
► Beigel, John H.; Nam, Hannah H.; Adams, Peter L.; Kraft, Amy; Ince, William L.; El-Kamary, Samer S....

Publication date:
2019-04-08

Journal:
Antiviral Res

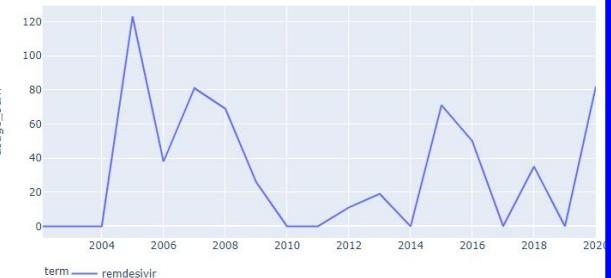
Abstract:
► The International Society for Influenza and other Respiratory Virus Diseases held its 6th Antiviral...

► The results of the predicted pharmacokinetics and pharmacodynamics properties of the azithromycin, chloroquine, lopinavir, oseltamivir, remdesivir, and ribavirin are presented in Tables 1–5.

► The results suggest that the high number of hydrogen bond formation could be responsible for the high binding score in lopinavir, remdesivir, and azithromycin (Elshah & Dandapani, 2012).

View graphs and visualizations of queries terms

Keyword usage over time



Advanced search options

Class Filter

sars mers covid19 none allow multiclass

Search Method

Containing Any Words

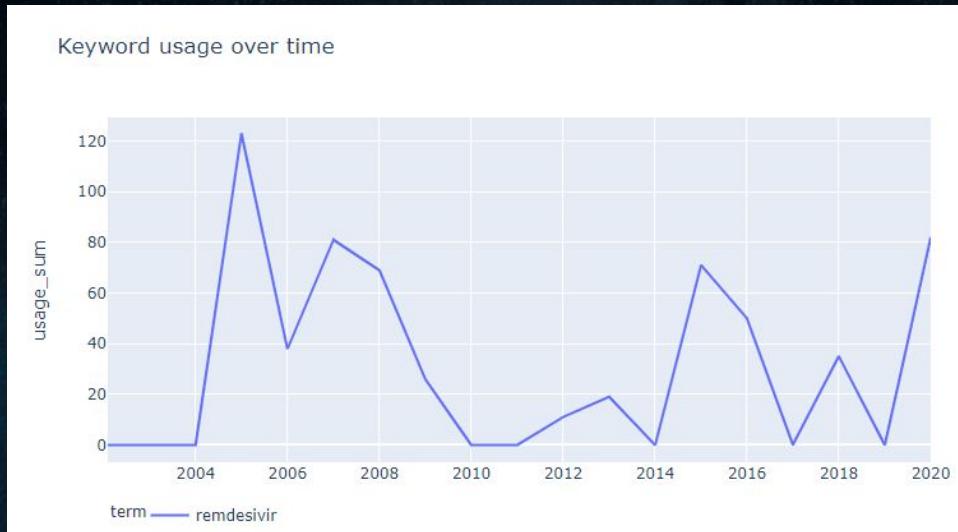
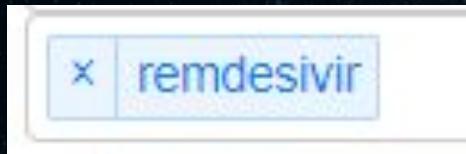
Filter By Author Name(s)

Filter By Journal Name(s)

Filter By Year Range

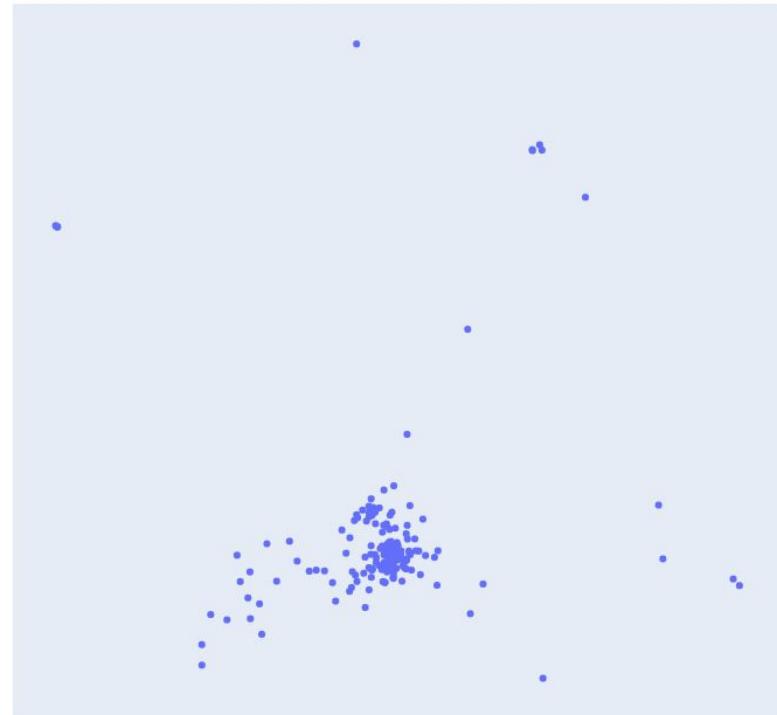
 Use authors Use journals Use year range Enable Strict Class Search

- Purpose: Examine usage trends by year to find patterns
- Keyword Search: Remdesivir
- Remdesivir has been looked at as a treatment for coronaviruses prior to COVID-19



Results

- ▶ Infected and mock-infected human lung epithelial cells (Calu 3) were treated with dilution series of lopinavir (50 μ M-0.05 μ M), ritonavir (50 μ M-0.05 μ M), interferon-beta (2800 IU/mL to 5.5 IU/mL), or **remdesivir** (10 μ M-0.02 μ M) and results were assayed 48 h post infection.
- ▶ The results of the predicted pharmacokinetics and pharmacodynamics properties of the azithromycin, chloroquine, lopinavir, oseltamivir, **remdesivir**, and ribavirin are presented in Tables 1-5.
- ▶ The results suggest that the high number of hydrogen bond formation could be responsible for the high binding score in lopinavir, **remdesivir**, and azithromycin (Elovely & Doerksen, 2013).
- ▶ [46](#) Several clinical trials are currently registered for the investigation of **remdesivir**, [12](#), [13](#) but no robust results are yet available.
- ▶ Pending results of several randomized ($n = 308$) clinical trials are expected to provide definitive insight into the efficacy of **remdesivir** as a therapeutic solution for the treatment of COVID-19.
- ▶ In addition to **remdesivir**, treatment with 17 compounds resulted in discernable dose-dependent antiviral activities, most of which could be segregated based on broad functional, structural, or target-based classes (Figure 3A).



Results

► Infected and mock-infected human lung epithelial cells (Calu 3) were treated with dilution series of lopinavir (50 μ M-0.05 μ M), ritonavir (50 μ M-0.05 μ M), interferon-beta (2800 IU/mL to 5.5 IU/mL), or **remdesivir** (10 μ M-0.02 μ M) and results were assayed 48 h post infection.

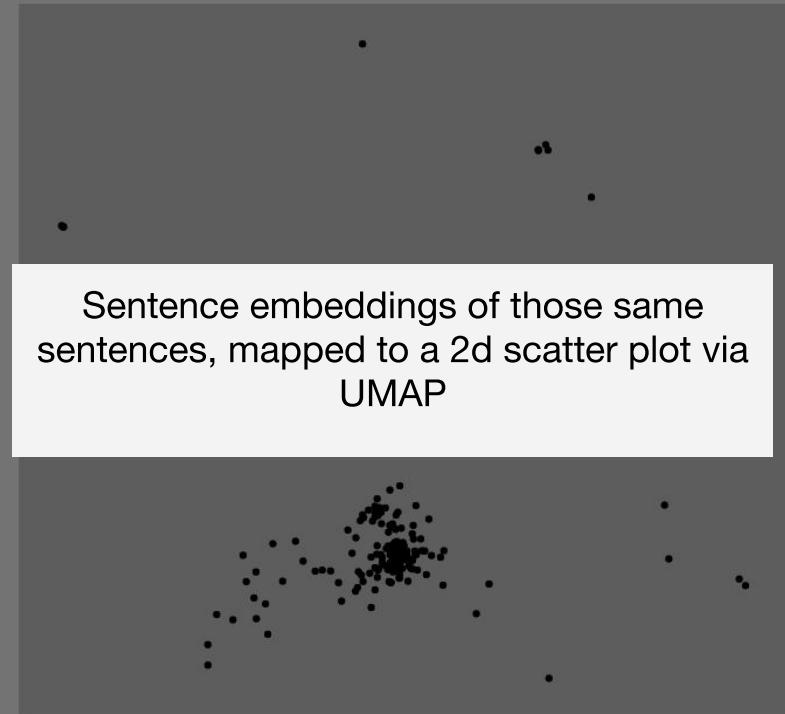
► The results of the predicted pharmacokinetics and pharmacodynamics properties of the azithromycin, chloroquine, lopinavir, oseltamivir, **remdesivir**, and ribavirin are presented in Tables 1-5.

► The results suggest that **remdesivir** may have therapeutic potential against COVID-19. The results show that **remdesivir** has potent antiviral activity against SARS-CoV-2 in vitro and in vivo. The results also show that **remdesivir** has a favorable pharmacokinetic profile and is well-tolerated in humans. The results suggest that **remdesivir** may be a promising therapeutic option for the treatment of COVID-19.

► Several clinical trials are currently registered for the investigation of **remdesivir**,^{12,13} but no robust results are yet available.

► Pending results of several randomized (n = 308) clinical trials are expected to provide definitive insight into the efficacy of **remdesivir** as a therapeutic solution for the treatment of COVID-19.

► In addition to **remdesivir**, treatment with 17 compounds resulted in discernable dose-dependent antiviral activities, most of which could be segregated based on broad functional, structural, or target-based classes (Figure 3A).



- **Purpose:** See returned sentences and most relevant information from them
- Each sentence is accompanied by metadata about the article it comes from
 - Can drill down and compare authors, journals, publication dates
 - Can view original paper to read more in-depth

▼
Infected and mock-infected human lung epithelial cells (Calu 3) were treated with dilution series of lopinavir (50 μ M-0.05 μ M), ritonavir (50 μ M-0.05 μ M), interferon-beta (2800 IU/mL to 5.5 IU/mL), or remdesivir (10 μ M-0.02 μ M) and results were assayed 48 h post infection.

Title:
Advances in respiratory virus therapeutics – A meeting report from the 6th isirv Antiviral Group conference

doi: [10.1016/j.antiviral.2019.04.006](https://doi.org/10.1016/j.antiviral.2019.04.006)

Authors:
► Beigel, John H.; Nam, Hannah H.; Adams, Peter L.; Krafft, Amy; Ince, William L.; El-Kamary, Samer S....

Publication date:
2019-04-08

Journal:
Antiviral Res

Abstract:
► *The International Society for Influenza and other Respiratory Virus Diseases held its 6th Antiviral ...*

►
The results of the predicted pharmacokinetics and pharmacodynamics properties of the azithromycin, chloroquine, lopinavir, oseltamivir, remdesivir, and ribavirin are presented in Tables 1-5.

►
The results suggest that the high number of hydrogen bond formation could be responsible for the high binding score in lopinavir, remdesivir, and azithromycin (Elokely & Doerksen, 2013).

- **Purpose: Examine clusters and similar sentences quickly**
- Each sentence can be found in the SciBERT embedding plot using UMAP
- Close or far distance in the embedding space represents semantic similarity or dissimilarity between the sentences

Results

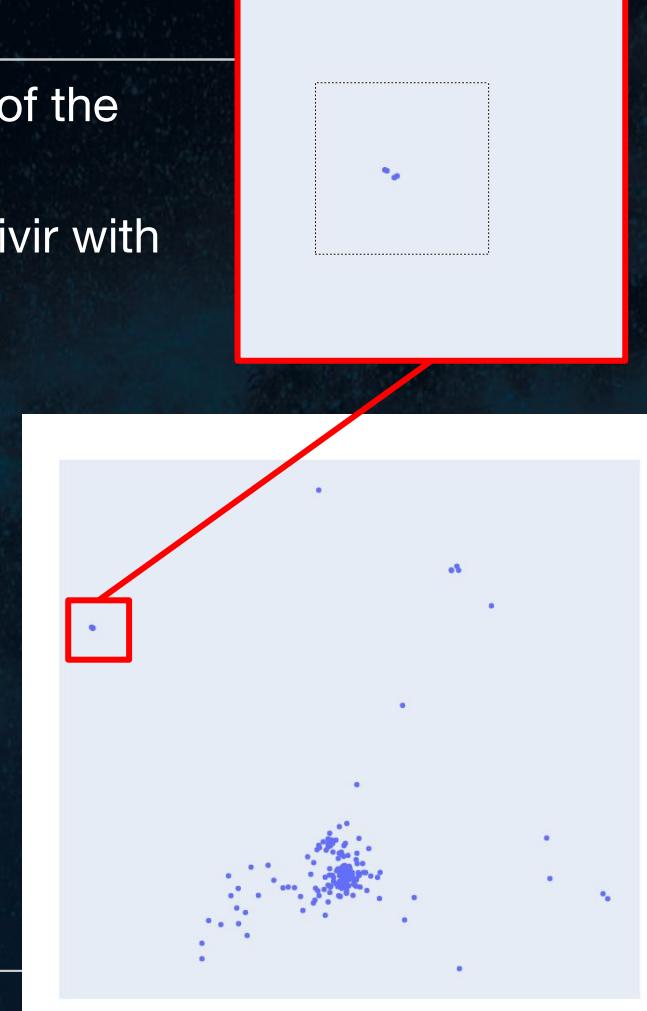


Infected and mock-infected human lung epithelial cells (Calu 3) were treated with dilution series of lopinavir (50 μ M-0.05 μ M), ritonavir (50 μ M-0.05 μ M), interferon-beta (2800 IU/mL to 5.5 IU/mL), or remdesivir (10 μ M-0.02 μ M) and results were assayed 48 h post infection.

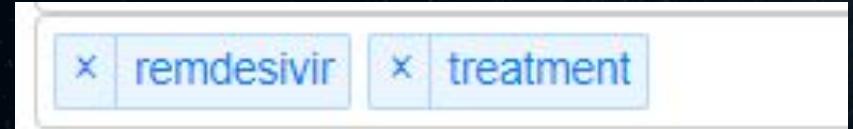


- Groups of points can be selected to inspect subsets of the data
- This cluster represents sentences comparing remdesivir with other antivirals like lopinavir and ritonavir

- ▶ Many of these trials will have results within months, and if **remdesivir** (produced by Gilead) and/or lopinavir plus ritonavir (produced by AbbVie as Kaletra and Aluvia, respectively) show effectiveness, they could potentially be used widely within a short time frame.
- ▶ Alternatively: 250 mg q 24 × 3 d (IV) ◊ Although lopinavir/ritonavir appears not to be effective, preliminary results with **Remdesivir** showed positive effect in 68% of cases [121](#).
- ▶ These include local standard of care in addition to **remdesivir**; chloroquine or hydroxychloroquine; ritonavir and lopinavir; and ritonavir, lopinavir, and interferon beta 1-alpha.
- ▶ Whilst our results showed that, saquinavir, ritonavir, and **remdesivir** can form irreversible interactions, which are considered an effective way for viral infections.



- Multiple keywords can be selected, searching for sentences containing all search terms
- Seems like the research on remdesivir shows promise for its use in treating COVID-19



- ▶ Pending results of several randomized ($n = 308$) clinical trials are expected to provide definitive insight into the efficacy of **remdesivir** as a therapeutic solution for the **treatment** of COVID-19.
- ▶ In addition to **remdesivir**, **treatment** with 17 compounds resulted in discernable dose-dependent antiviral activities, most of which could be segregated based on broad functional, structural, or target-based classes (Figure 3A).
- ▶ Findings: **Remdesivir treatment** results in a 33% significantly higher odds of discharge, a 29% significantly lower risk of death, and a 39% significantly lower risk for the combined endpoint of severe status and death.
- ▶ Beyond viral neutralization strategies aimed at preventing infection, other **treatment** approaches may decrease viral load and shorten disease duration, such as the antiretroviral drug **remdesivir** (preliminary results from trial NCT04292899).
- ▶ **Remdesivir**, a polymerase inhibitor developed as a **treatment** for Ebola virus 22 is showing very promising early results and will likely be confirmed in clinical trials 23 Figure 3 .
- ▶ Comparably, daily **treatment** with 20 μM **remdesivir** resulted in 7.3 log₁₀ and 7.9 log₁₀ reductions of intracellular SARS-CoV-2 viral titers at 48 hpi in nasal and bronchial HAE, respectively (Fig.



Preliminary clinical data for the investigational RNA antiviral agent, **remdesivir**, did not show a survival benefit in a study of 237 patients in China but did show a trend to earlier recovery [55](#), and the Adaptive COVID-19 **Treatment** Trial (ACTT) a study of 1,063 patients with lung involvement in the USA has not yet been peer reviewed but a press release from the National Institutes of Health (NIH) [56](#) reports a 31% faster recovery time ($p<0.001$) of 11 versus 15 days and possibly a trend to a survival benefit with 8% versus 11.6% mortality ($p=0.059$).



Findings: **Remdesivir treatment** results in a 33% significantly higher odds of discharge, a 29% significantly lower risk of death, and a 39% significantly lower risk for the combined endpoint of severe status and death.



Compared to vehicle-treated control animals, prophylactic **remdesivir treatment** resulted in significantly lower levels of MERS-CoV replication in the lungs, with lung viral loads 2.5 to 4 logs lower in each lung lobe (Fig.).



Comparably, daily **treatment** with 20 μ M **remdesivir** resulted in 7.3 log₁₀ and 7.9 log₁₀ reductions of intracellular SARS-CoV-2 viral titers at 48 hpi in nasal and bronchial HAE, respectively (Fig.).

- Preliminary clinical data for the investigational RNA antiviral agent, **remdesivir**, did not show a survival benefit of 237 patients, but did show a trend to earlier recovery
- Findings: **Remdesivir** treatment results in a 33% significantly higher odds of discharge, a 29% significantly lower risk of death, and a 39% significantly lower risk for the combined endpoint of severe status and death
- Compared to vehicle-treated control animals, prophylactic **remdesivir** treatment resulted in significantly lower levels of MERS-CoV replication in the lungs, with lung viral loads 2.5 to 4 logs lower in each lung lobe
- Comparably, daily treatment with 20 uM **remdesivir** resulted in 7.3 log₁₀ and 7.9 log₁₀ reductions of intracellular SARS-CoV2 viral titers at 48 hpi in nasal and bronchial HAE, respectively.

- Additionally, the BERT dashboard allows searching based on cosine similarity of sentence embeddings
- User can enter in a full sentence as a search term, and the dashboard will return sentences with semantic similarities, based on their embeddings

The screenshot shows a search interface for BERT. At the top, there is a search bar containing the text "remdesivir and other antivirals in shortening infection time". Below the search bar, the word "Results" is displayed. A list of ten bullet points follows, each representing a search result related to the query:

- ▶ SARS-CoV-2 infections may occur during boosted darunavir-based and/or on tenofovir-containing ART.
- ▶ Combined lopinavir and ritonavir is an effective regime in controlling HIV infection.
- ▶ Lopinavir and/or ritonavir has anti coronavirus activity in vitro.
- ▶ The HIV protease inhibitors lopinavir/ritonavir inhibit the major CoV protease 3CLpro.
- ▶ Lopinavir/Ritonavir have anti coronavirus activity in vitro.
- ▶ Chloroquine as prophylactic agent against COVID-19?
- ▶ Results of SARS-CoV-2 nucleic acid testing in newborn.
- ▶ As in HIV infection, ritonavir only acts boosting lopinavir plasma levels.

- This method is useful for finding research relating to a complex idea or hypothesis
- Can also show what types of ideas seem to co-occur
 - For example, much of the research into antivirals (like Remdesivir) has been driven by HIV research

37 Based on these studies, lopinavir/ritonavir appears to be a promising anti-SARScoronavirus agent.

Title:

Treatment and vaccines for severe acute respiratory syndrome

20 Lopinavir is a protease inhibitor with in-vitro activity against SARS-CoV and MERS-CoV.

Title:

Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study

SARS-CoV-2 infections may occur during boosted darunavir-based and/or on tenofovir-containing ART.

Title:

COVID-19 in people living with **human immunodeficiency virus**: a case series of 33 patients

The HIV protease inhibitors lopinavir/ritonavir inhibit the major CoV protease 3CLpro.

Title:

Potential new treatment strategies for COVID-19: is there a role for bromhexine as add-on therapy?

- Found evidence that Remdesivir may be a valid treatment for COVID-19 through an integrated dashboard that allows any end user to efficiently parse through a large version of the CORD-19 dataset
 - Higher rate of discharge
 - Lower rate of death
 - Lower levels of virus replication
- Found insights into remdesivir creation and development
 - HIV research

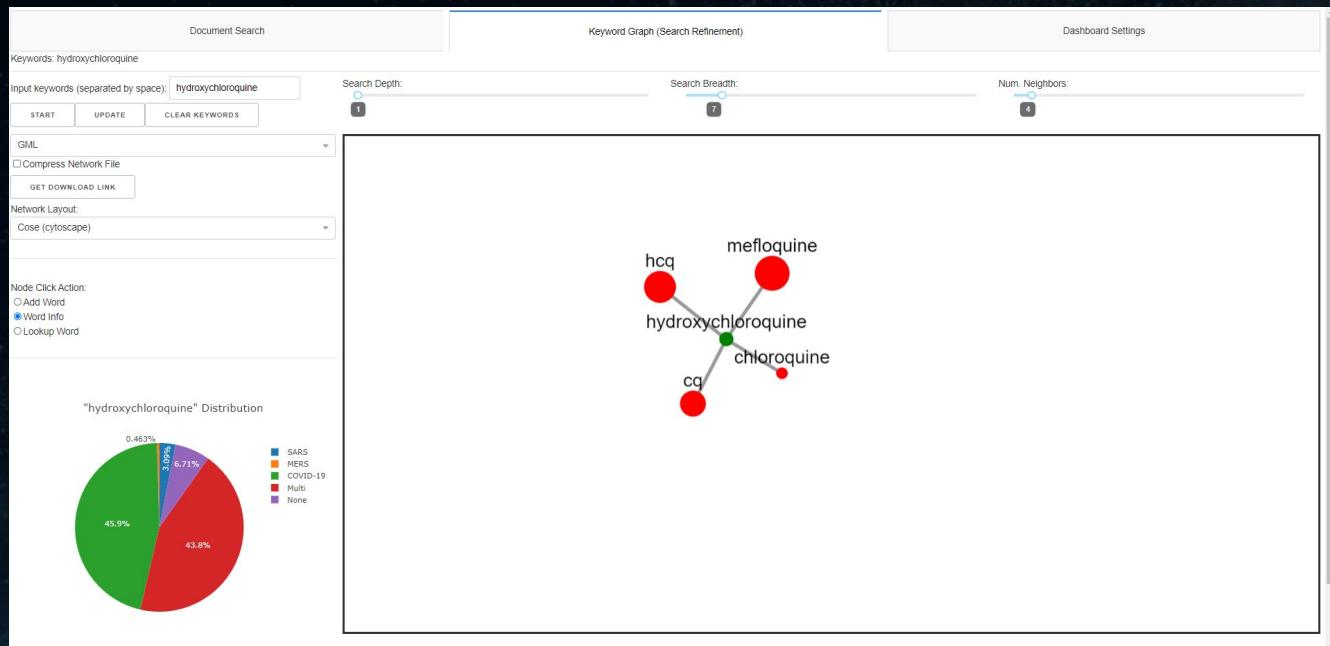
- Dashboard 2: tf-idf dashboard
 - Built on a backbone of tf-idf-based document search, with a great deal of added functionality and visualization implemented on top
- Unique features:
 - tf-idf document search
 - User-defined regex sentence search
 - Keyword graph (nearest neighbor search)
- Use case: Hydroxychloroquine for treatment of COVID-19

- Convert a document (sequence of word tokens, a.k.a. *terms*) to a numerical vector
 - One component for each word in the vocabulary
- **Term frequency (tf)**: The frequency of a term within the document
 - $tf(t,d) = (\text{number of occurrences of } t \text{ in } d) / (\text{number of terms in } d)$
- **Document frequency (df)**: Proportion of documents a term occurs in
 - $df(t) = (\text{number of documents containing } t) / (\text{number of documents})$
- **Inverse document frequency (idf)**: Gives more weight for terms that appear in less documents
 - $idf(t) = \log(1 / df(t))$
- tf-idf is the product of tf and idf; alternate ways of computing use the same approach

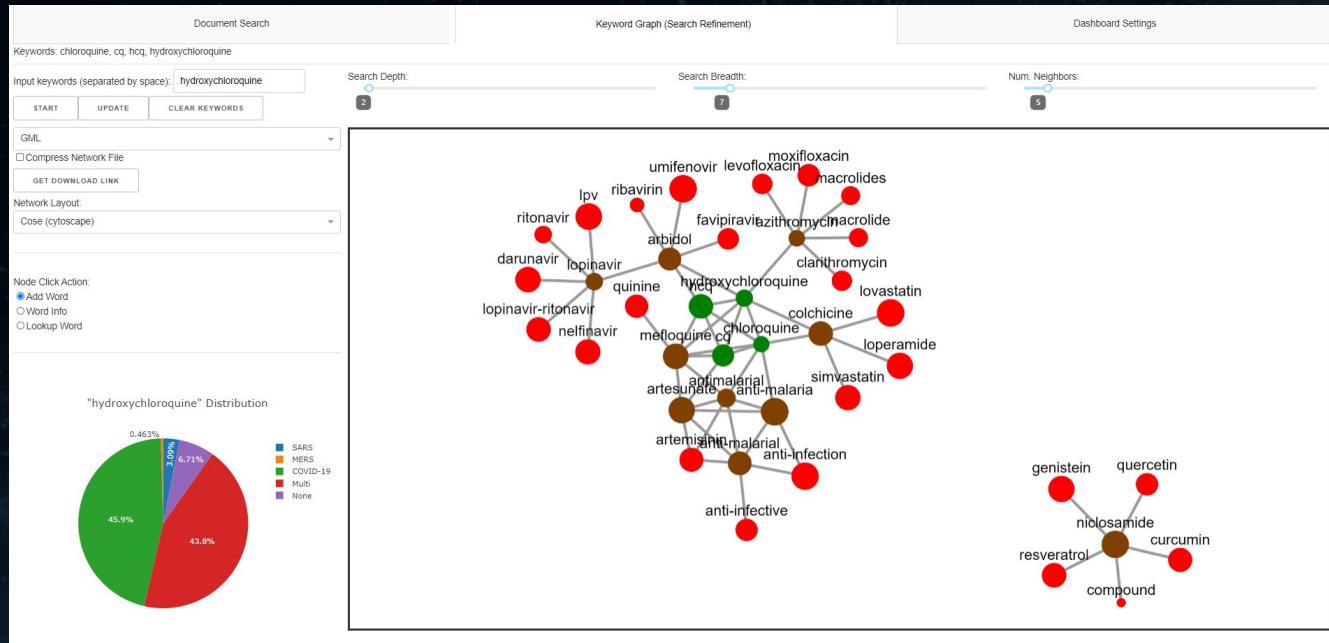
document	apple	approach	...	zoo
document 1	0.0	0.23821	...	0.0
document 2	0.0196833	0.0	...	0.42894257

- Efficient searching of documents using the tf-idf representation to compute a score
- Given a **query** q , treat it as a document and compute a tf-idf vector for it using the same vocabulary as the other documents
- For each document d , compute a **query score** $s(q, d)$ using **cosine similarity** (the normed inner product) between the document and query tf-idf vectors
- Normalize the query scores for all documents to be between 0 and 1 by dividing by the largest
 - The larger the query score, the more “relevant” the document is to the query
- **Boolean queries:** combine scores from two subqueries, grouped by parentheses
 - AND: minimum of two or more scores
 - OR: maximum of two or more scores
 - Example: (covid-19 OR 2019-ncov OR coronavirus) AND (spread OR outbreak OR quarantine)

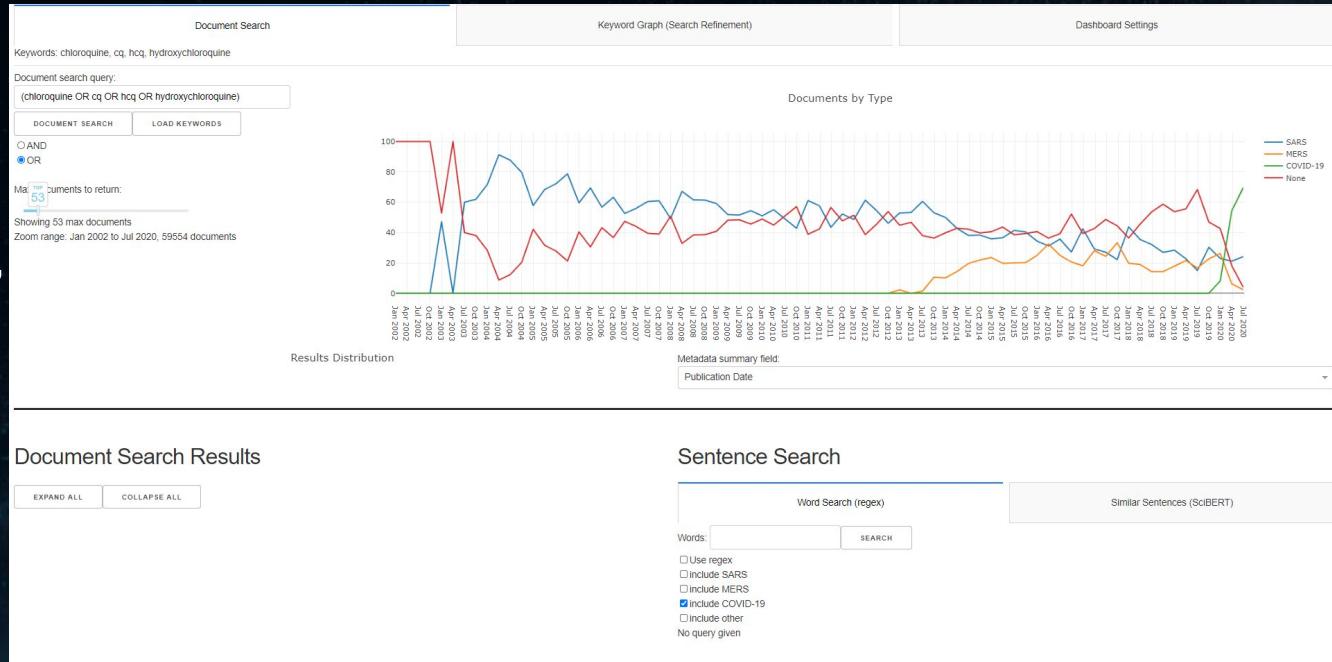
- Purpose: Find additional relevant words to query
- Graph of relationships between words (fastText)
- Search the word “hydroxychloroquine”
- Four nearest neighbors of the word that occur in the vocabulary
- Clicking node displays word information



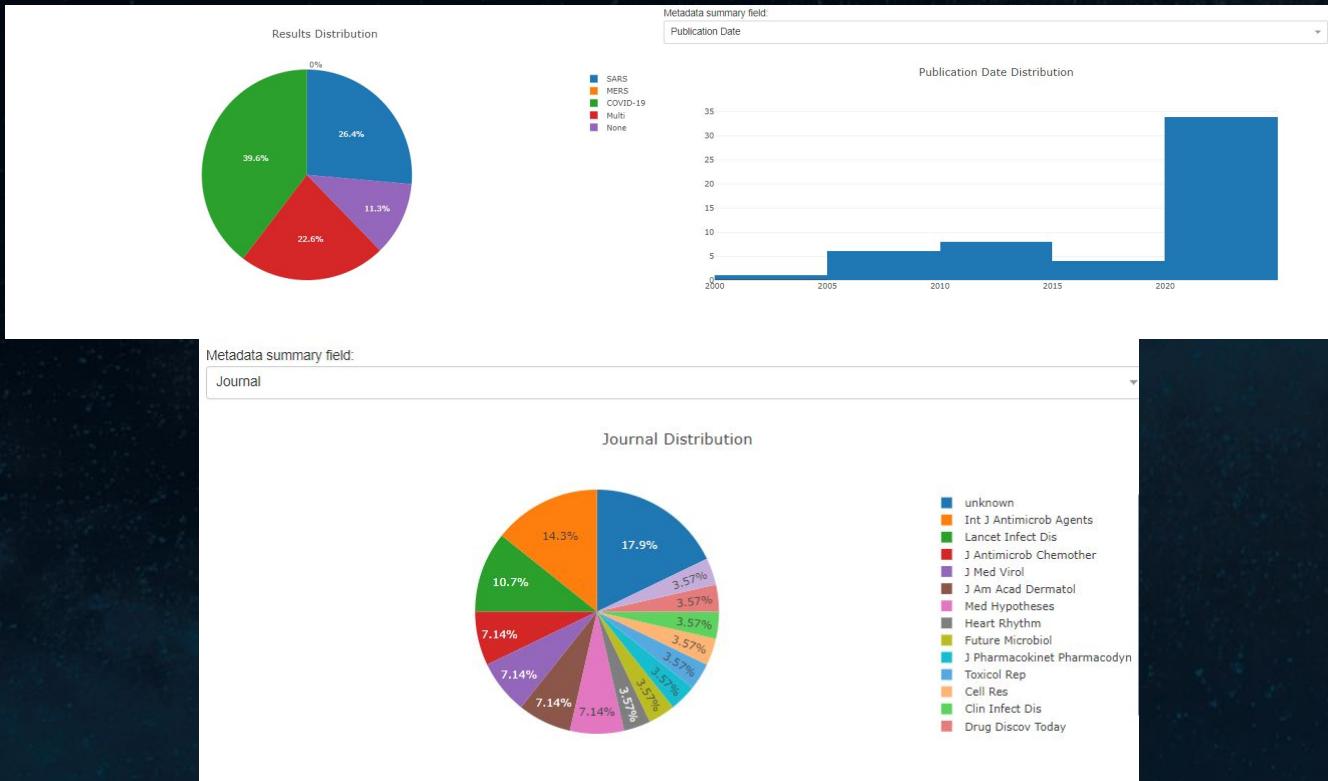
- Set node click to “add word” and click on desired words in the graph
- Starting word and all words selected are added to a list of keywords at the top
- Search depth and breadth to control number of nodes in network
- Num neighbors to control number of edges



- Purpose: Find documents relating to keyword set
- Click “Document Search” tab
- Click “Load Keywords” to build a document query from words selected in the keyword graph
- Click “Document Search” to run tf-idf-based search



- Purpose: Get a quick overview of results
- Distribution of topic of articles in result set indicated in pie chart
- Summaries of metadata:
 - Number of documents published each year in the results
 - Distribution of journals



- Document Results grouped by which human-coronavirus they describe
 - SARS, MERS, COVID-19, Multiple, None
- Metadata about the for the document, including abstract if it is present
- Full text link will use the local copy of the document from the CORD-19 dataset

Document Search Results

[EXPAND ALL](#) [COLLAPSE ALL](#)

► SARS Documents

▼ MERS Documents

▼ COVID-19 Documents

▼ An Updated Systematic Review of the Therapeutic Role of Hydroxychloroquine in Coronavirus Disease-19 (COVID-19) (2020-05-28)

Categories: COVID-19**Full text:** [ne8k1vez](#)**Authors:** Das, Salab, Bhownick, Subhrojyoti; Tiwari, Sayali; Sen, Sukanta**Journal:** Clin Drug Investig**DOI Link:** [10.1007/s40261-020-00927-1](https://doi.org/10.1007/s40261-020-00927-1)

▼ Abstract

BACKGROUND AND OBJECTIVE: The world is currently experiencing the Coronavirus Disease-19 (COVID-19) pandemic. There is no approved drug for the definitive treatment of the disease. Various drugs are being tried for the treatment of COVID-19, including hydroxychloroquine (HCQ). This study was performed to systematically review the therapeutic role of HCQ in COVID-19 from the available literature. **METHODS:** PubMed, Embase, ClinicalTrials.gov, ICTRP (WHO), Cochrane Library databases, and two pre-print servers (medRxiv.org and Research Square) were searched for clinical studies that evaluated the therapeutic role of HCQ on COVID-19 until 10 May 2020. The available studies were critically analyzed and the data were extracted. **RESULTS:** A total of 663 articles were screened and 12 clinical studies (seven peer-reviewed and published studies and five non-peer-reviewed studies from pre-print servers) with a total sample size of 3543 patients were included. Some of the clinical studies demonstrated good virological and clinical outcomes with HCQ alone or in combination with azithromycin in COVID-19 patients, although the studies had major methodological limitations. Some of the other studies showed negative results with HCQ therapy along with the risk of adverse reactions. **CONCLUSION:** The results of efficacy and safety of HCQ in COVID-19, as obtained from the clinical studies, are not satisfactory, although many of these studies had major methodological limitations. Stronger evidence from well-designed robust randomized clinical trials is required before conclusively determining the role of HCQ in the treatment of COVID-19. **Clinical prudence is required in advocating HCQ as a therapeutic armamentarium in COVID-19.** **ELECTRONIC SUPPLEMENTARY MATERIAL:** The online version of this article (10.1007/s40261-020-00927-1) contains supplementary material, which is available to authorized users.

▼ Further Information

Query Score: 1.0000**CORD-19 UID:** ne8k1vez**URL:** <https://doi.org/10.1007/s40261-020-00927-1> <https://www.ncbi.nlm.nih.gov/pubmed/32468425/>

► COVID-19 infection also occurs in patients taking hydroxychloroquine (2020-05-17)

► Is Coronavirus Disease 2019 (COVID-19) seen less in countries more exposed to Malaria? (2020-04-22)

► Aminoquinolines against coronavirus disease 2019 (COVID-19): chloroquine or hydroxychloroquine (2020-03-17)

► Chloroquine as prophylactic agent against COVID-19? (2020-04-12)

► Comment on: COVID-19: a recommendation to examine the effect of hydroxychloroquine in preventing infection and progression (2020-05-03)

► Chiral switches of chloroquine and hydroxychloroquine: potential drugs to treat COVID-19 (2020-05-01)

- Purpose: find sentences in result set relating to treatment
- regex “treat(ment|ing)” will match “treatment” or “treating”
 - Option to not use regex and search for a word or list of words instead
- Matches in sentences highlighted
- Can choose which categories to search in - here only COVID-19 documents
- Clicking on the ID in the title (blue link) will scroll to that document in the search results

Sentence Search

Word Search (regex)

Words: treat(ment|ing)

Use regex
 include SARS
 include MERS
 include COVID-19
 include other

Similar Sentences (SciBERT)

Results

▼ Fatal arrhythmias: Another reason why doctors remain cautious about chloroquine/hydroxychloroquine for treating COVID-19. ([_owpzsmoy](#))

- Background: Early during the current COVID-19 pandemic, hydroxychloroquine (HCQ) received a significant amount of attention as a potential antiviral **treatment**, such that it became one of the most commonly prescribed medications for COVID-19 patients. ([_find similar](#))
- Although CQ and HCQ are becoming the focus of **treating** COVID-19, they remain unendorsed by many physicians because of 1) limited clinical outcome data, 2) availability of other potentially more effective antiviral and interleukin inhibitors such as remdesivir and tocilizumab respectively, and 3) potential risk of malignant arrhythmia and sudden cardiac death (SCD) due to QT prolongation 7, 8. . ([_find similar](#))
- Consequently, many COVID-19 patients are already susceptible to arrhythmias and may be at even higher risk of deadly arrhythmogenic effects with COVID-19 drug combination **treatments**. ([_find similar](#))
- In both instances, fibrillation spontaneously occurred under the hypokalemic conditions (the heart was defibrillated when back in normal Tyrode's solution), highlighting a possible higher risk of HCQ in patients with hypokalemia and thus the importance of monitoring and **treating** low potassium levels as they can increase the probability of arrhythmic events 19. ([_find similar](#))
- A challenging aspect of COVID-19 **treatment** with respect to heart problems is the use of unusual drug combinations, such as HCQ with antibiotics like azithromycin, which prior to the COVID-19 pandemic were rarely used together. ([_find similar](#))

► New insights on the antiviral effects of chloroquine against coronavirus: what to expect for COVID-19? ([_da61tfr9](#))

► COVID-19 infection also occurs in patients taking hydroxychloroquine. ([_h6kiz5ui](#))

► Comment on: COVID-19: a recommendation to examine the effect of hydroxychloroquine in preventing infection and progression ([_ea4en1cc](#))

► Hydroxychloroquine and azithromycin as potential treatments for COVID-19; clinical status impacts the outcome ([_84ci8h6p](#))

- Purpose: find similar sentences that make an assertion about the efficacy of (hydroxy)chloroquine in treating COVID-19
- Sentences sorted by similarity (the first sentence here is the input sentence since it's in the results set)
 - Euclidean distance between SciBERT vectors
- Click on a sentence to expand and view document information

[Document Search](#)[Keyword Graph \(Search Refinement\)](#)[Dashboard Settings](#)

Keywords:

Sentence Search Results

Query: Although CQ and HCQ are becoming the focus of treating COVID-19, they remain unendorsed by many physicians because of 1) limited clinical outcome data, 2) availability of other potentially more effective antiviral and interleukin inhibitors such as remdesivir and tocilizumab respectively, and 3) potential risk of malignant arrhythmia and sudden cardiac death (SCD) due to QT prolongation 7, 8.

In: Fatal arrhythmias: Another reason why doctors remain cautious about chloroquine/hydroxychloroquine for treating COVID-19 ([Owpzsmoy](#))

Top Sentences

▼ 1. Although CQ and HCQ are becoming the focus of treating COVID-19, they remain unendorsed by many physicians because of 1) limited clinical outcome data, 2) availability of other potentially more effective antiviral and interleukin inhibitors such as remdesivir and tocilizumab respectively, and 3) potential risk of malignant arrhythmia and sudden cardiac death (SCD) due to QT prolongation 7, 8.

- ▶ Fatal arrhythmias: Another reason why doctors remain cautious about chloroquine/hydroxychloroquine for treating COVID-19 (2020-05-29)

▼ 2. Despite the dearth of evidence of efficacy, in light of the pressure that COVID-19 has posed on national health systems, several official guidelines have already incorporated hydroxychloroquine and chloroquine into the suggested treatment of patients with COVID-19 [28, 50, 51].

- ▶ Review: Hydroxychloroquine and Chloroquine for Treatment of SARS-CoV-2 (COVID-19) (2020-04-15)

▼ 3. Due to the aforementioned evidence, the negligible cost, its large worldwide use, and the known safety profile, CQ/ HCQ has been considered as a potentially useful drug in patients affected by SARS-CoV-2 [1, 5, 6] Despite in vitro activity in inhibiting the growth of several viruses, to date no acute virus infection has been successfully treated by Co/HCoq [7].

- ▶ Could Chloroquine /Hydroxychloroquine Be Harmful in Coronavirus Disease 2019 (COVID-19) Treatment? (2020-03-24)

Categories: SARS, COVID-19
Full text: [Zg1sudjg](#)
Authors: Guastalegname, Maurizio; Vallone, Alfredo
Journal: Clin Infect Dis
DOI Link: [10.1093/cid/ciaa321](#)
► Abstract
► Further Information

- ▶ There is cautious optimism that (hydroxy)chloroquine may have prophylactic and/or therapeutic effects against COVID-19, and understanding the mechanisms by which these drugs affect SARS-CoV-2 would be critical for optimizing and developing preventative and therapeutic strategies.
- ▶ 5. Although CQ and HCQ both have the potential to act against SARS-CoV-2, CQ, particularly at a higher dose, is associated with a higher risk of toxicity and should not be recommended for critically ill patients with COVID-19 [11].
- ▶ 6. While there is currently no general advice on the optimal dose for chloroquine in SARS-CoV-2, it appears that with dosages proposed in malaria studies or in the 5-day treatment schedules currently used for COVID-19, variations in plasma concentrations as a result of obesity, renal dysfunction, or underweight may be limited (Fig).
- ▶ 7. 236 Despite being researched for several decades and being supported by a large volume of in vitro and 238 animal study data for plausible mechanisms for antiviral effects, neither HCQ nor CQ are currently 239 recommended as antiviral agents for any of the infections they had been tested for in clinical trials 240 (with the exception of COVID-19 which is an evolving situation at the time of writing).
- ▶ 8. Although research concerning the use of CQ and HCQ in central COVID-19 is still preliminary, the potential use of these drugs is supported by

- *Fatal arrhythmias: Another reason why doctors remain cautious about chloroquine/hydroxychloroquine for treating COVID-19*
 - “Although CQ and HCQ are becoming the focus of treating COVID-19, they remain unendorsed by many physicians because of 1) limited clinical outcome data, 2) availability of other potentially more effective antiviral and interleukin inhibitors such as remdesivir and tocilizumab respectively, and 3) potential risk of malignant arrhythmia and sudden cardiac death (SCD) due to QT prolongation 7, 8.”
- *An Updated Systematic Review of the Therapeutic Role of Hydroxychloroquine in Coronavirus Disease-19*
 - “Although CQ and HCQ both have the potential to act against SARS-CoV-2, CQ, particularly at a higher dose, is associated with a higher risk of toxicity and should not be recommended for critically ill patients with COVID-19 [11] .”
 - “However, in this ongoing challenging scenario, considering the absence of any other definitive therapy in COVID-19, the mixed efficacy, and the safety profile of HCQ, we feel that clinicians should carefully weigh risks and benefits of HCQ alone or in combination with azithromycin.”

- Found evidence against the use of hydroxychloroquine as a treatment for COVID-19
 - Risk of arrhythmia and cardiac failure
- Integrated dashboard allows efficient navigation of the large CORD-19 dataset
- Keyword graph page allows the exploration of keywords to find related keywords to broaden a search

- Utilized two dashboards to determine the efficacy of 2 drugs for treatment of COVID-19
 - Remdesivir
 - Found evidence for the usage of Remdesivir as a treatment for COVID-19
 - Hydroxychloroquine
 - Found some evidence against the usage of hydroxychloroquine as a treatment for COVID-19
- Insights and results obtained in a fraction of the time (~3 minutes) versus traditional literature search methods

- Benefits of CORD-19 dashboards
 - Quick efficient search through the large dataset
 - Documents, Sentences
 - Input a sentence and find most similar sentences
 - Visualization tools to find insights into results that may not be apparent
 - Related keywords (tf-idf dashboard)
 - Related sentences in embedding space (BERT dashboard)

- Fine-tuning of Scibert Models
 - NLI dataset
 - AllenNLI
 - CORD19STS
 - Semantic Text Similarity version of Cord19
 - 32k sentence pairs evaluated for semantic similarity via users on mturk platform
 - Labels for improving
 - Similarity searches
 - Training

- Combination of both dashboards as a single unified dashboard
 - Requires server with higher memory and storage space
 - “Basic” and “Advanced” versions, similar to apps
 - Intelligently hide features
- Dashboard testing by clinicians to evaluate possible medications, uses, features
 - Feedback testing loop

- A detailed user manual has been included with the repository
 - Explain algorithms operating under-the-hood
 - Give future developers a “quick-start” to continue development
 - Describes a number of unfinished features which remain latent in the code, could be debugged and enabled

When a word or words is entered in the keyword search field, the algorithm will return sentences which contain that word, and also match several hard-coded regex queries which filter down the results to sentences which are likely to contain content relating to statements about results or conclusions. This filtering can be found in the `FilterSentencesByRegex` function.

Regex string	intent
<code>r\bresult[a-z]+'</code>	Sentences which contain the word "result", e.g. "the result was", "resulting in", "presented the following results" etc.
<code>r'p[=>][\s\d] \P[=>][\s\d]\balpha'</code>	Sentences which contain statistics, e.g. "p<.05", "alpha level of" etc

These strings can be added to or modified by changing the list of strings within that function.

The list of sentences generated by this function is then output to two output modalities.

The first modality is the most conventional: it presents the returned sentences as a scrollable result list of text. Metadata about the documents each sentence was drawn from is displayed as well, as content within a collapsible under each sentence.

The second modality is more visual: it presents as a scatter plot a 2d representation of the BERT embeddings of the same sentences as in the scrollable result list. This is accomplished by running UMAP on the full length vector embeddings to reduce them to a 2d vector space. Hovering over points on the scatter plot reveals the text of the full sentence it represents. Additionally, selecting a subset of the plotted points (plotly graphs include box and lasso select buttons by default) will also remove any sentences in the result list which are not present in the selected subset. This allows for closer inspection of any interesting clusters present within the embedding space.

embedding search

A third functionality of the dashboard is the embedding search function. This allows the user to type in a full sentence or phrase, and then returns the list of n sentences which most closely align with that phrase via cosine similarity.

These sentences are presented in a format identical to the scrollable results list returned by the keyword search function. Notably though, this is not accompanied by a plot of the relevant embeddings. This is because, since the returned sentences have been selected to have the smallest cosine similarity to an input sentence, they are likely to have fairly similar embeddings. Specifically, they should all roughly lie along the same line from the origin. This means that a

- Improvements to tf-idf dashboard
 - Usage of pre-computed SciBERT vectors to speed up similar sentence search
 - Fall back to using the SciBERT model if vectors for a document aren't present
 - Less overhead on first tf-idf document search
 - Don't use JIT compilation for vectorization of the query, as it has too much overhead
 - Extract most “relevant” words from a document set
 - Largest differences in mean tf-idf values between result set and entire dataset, scaled by normalized IDFs of each word

- Example of finding most relevant words with the search “(hydroxychloroquine OR hcq OR chloroquine OR cq)”
 - From a document set of ~150 documents
- Possible future integrations:
 - Word search for more details of a popular word
 - Select multiple words from the table and build a query

Metadata summary field: Popular Words			
Word	Word Score	Docu.	Freq.
hcq	1	219	
chloroquine	0.80112368679046631	1686	
cq	0.6621900796890259	399	
hydroxychloroquine	0.5285141603559854	1296	
covid-19	0.11229163184762081	15029	
drug	0.11616810371679306	20945	
azithromycin	0.11205510795116425	1569	
trial	0.05908700306852395	13986	
antimalarial	0.078541196880272476	909	
qtc	0.074148495720254318	263	
malaria	0.0724120410201462	3889	
mg	0.07102558016777039	15019	
dose	0.06515545379228141	13661	
sars-cov-2	0.059393537197351456	8804	
qt	0.058546971529722214	506	
prolongation	0.057964492589235306	893	

Metadata summary field: Popular Words			
Word	Word Score	Docu.	Freq.
prolongation	0.057964492589235306	893	
treatment	0.0568915121257380515	34980	
prophylaxis	0.050977181643247604	4208	
antiviral	0.047621291130781174	14496	
toxicity	0.047012973576784134	6431	
retinopathy	0.04655146598815918	467	
vitro	0.044717904180288315	16106	
efficacy	0.04230876639485359	12956	
inhibit	0.04007856547832489	14669	
dosage	0.03930031135678291	3432	
lupus	0.03785420798428986	1631	
adverse	0.03661166504025459	9120	
regimen	0.035551814121847153	4716	
patient	0.03514759615063667	37803	
concentration	0.034899838268756866	21848	

- McInnes, L, Healy, J, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, ArXiv e-prints 1802.03426, 2018
- Beltagy, Iz, Kyle Lo, and Arman Cohan. "SciBERT: A pretrained language model for scientific text." arXiv preprint arXiv:1903.10676 (2019).
- Wang, Lucy Lu, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Darrin Eide, Kathryn Funk, Rodney Michael Kinney, Ziyang Liu, William. Merrill, Paul Mooney, Dewey A. Murdick, Devvret Rishi, Jerry Sheehan, Zhihong Shen, Brandon Stilson, Alex D. Wade, Kuansan Wang, Christopher Wilhelm, Boya Xie, Douglas M. Raymond, Daniel S. Weld, Oren Etzioni and Sebastian Kohlmeier. "CORD-19: The Covid-19 Open Research Dataset." *ArXiv* abs/2004.10706 (2020).
- Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. "Enriching word vectors with subword information." *Transactions of the Association for Computational Linguistics* 5 (2017): 135-146.