

Brief Report: Research Publications Related to the Corona Virus from PUBMED search May 20, 2020

We have been downloading abstract record sets from the PubMed database relating to corona virus every 2 weeks, starting March 25. So we have files from March 25, April 8, April 22, May 5, and, now, May 20 (21), 2020.

For each of the previous 4 datasets, we have generated a basic report, summarizing simple analyses to introduce the data for potentially deeper probes. Those basic reports remain available; they treat:

- The Data & Background on the searching
- Looking at the data in *VantagePoint* desktop text analysis software [www.theVantagePoint.com]
- Research Trend
- Notes on players participating strongly in that research (countries, organizations, researchers)
- Topical emphases
- Exploratory analysis (trying one novel empirical effort each time)

Here, we shift to a more concise overview with a new exploration, but first need to address issues in searching and downloading the records.

Search Essentials

We apply the search [query](#), but apply it to the legacy version of the PubMed database [by shifting from “PMC” via pull-down to “PubMed”]

- "COVID-19" OR Coronavirus OR "Corona virus" OR "2019-nCoV" OR "SARS-CoV" OR "MERS-CoV" OR “Severe Acute Respiratory Syndrome” OR “Middle East Respiratory Syndrome”¹ = 34405 8:38am on May 21

We were not able to complete download on May 20. The interface no longer allows full download (instead limiting to the first 10,000 records) as XML, as previously, so we ran partial searches <10,000 to recombine in *VantagePoint* (VP) on desktop. Did so using separate searches, for select years. Download now by: 1) Save; Selection: all results; Format: PubMed; Create File.

- ("COVID-19" OR "Corona virus" OR "2019-nCoV" OR "SARS-CoV" OR "MERS-CoV" OR “Severe Acute Respiratory Syndrome” OR “Middle East Respiratory Syndrome”) – restricted to 2020 = 14990
[noted that the PubMed interface Year selection had some overlap – limiting the 34405 search to 2020 = 15981; taking all prior years = 18486; sum is 34467]
- ("COVID-19" OR "Corona virus" OR "2019-nCoV" OR "SARS-CoV" OR "MERS-CoV" OR “Severe Acute Respiratory Syndrome” OR “Middle East Respiratory Syndrome”) AND coronavirus – restricted to 2020 = 7969 -- SAVE All Results PubMed format
- ("COVID-19" OR "Corona virus" OR "2019-nCoV" OR "SARS-CoV" OR "MERS-CoV" OR “Severe Acute Respiratory Syndrome” OR “Middle East Respiratory Syndrome”) NOT coronavirus – restricted to 2020 = 7021 -- SAVE All Results PubMed format
- Sum of AND & NOT = 14990 (for 2020)
- Combining the 14990 with our prior search result [May 6 = 28264], and removing duplicates on PubMed ID = 33667 [so short of 34405 by a lot]
- Grab recent years other than 2020 to see if fill in pretty well

¹ In MEDLINE detail: "COVID-19"[All Fields] OR ("coronavirus"[MeSH Terms] OR "coronavirus"[All Fields]) OR "Corona virus"[All Fields] OR "2019-nCoV"[All Fields] OR "SARS-CoV"[All Fields] OR "MERS-CoV"[All Fields] OR "Severe Acute Respiratory Syndrome"[All Fields] OR "Middle East Respiratory Syndrome"[All Fields].

- 2006-19 full search query -- SAVE All Results PubMed format = 9990 [when combine, get 33976, still quite short], so
- Ran 1949-2005 full search query -- SAVE All Results PubMed format = 8567
 - Combined datasets & Removed dups on PubMed ID = 34181 [vs. 34405 on May 21 vs. 34258 on May 20 – so within 0.65%]
- Note that PubMed search results are dynamic, updating within a day, based on this experience. Also note that updates appear to gain earlier year publications too.

We are investigating use of a PubMed API to facilitate future downloads.

Results

On May 21, search yields **34,181** PubMed abstract records, compared to prior:

- March 25, 2020: Initial dataset - 19,538 records.
- April 8, 2020: Update yielded 21,314 records.
- April 22, 2020: Update yielded 24,479 records.
- May 6, 2020: Update yielded 28,264 records.

As mentioned, we elect not to redo the rudimentary research profiling presented in the prior biweekly reports. Those remain available to peruse research participation, topical emphases, and research trend. For those of you wanting to answer “**who, what, where, when?**” questions, we make the data available in *VantagePoint* format, along with free trial use of the software to explore your interests.

This week we explore one question brought up in recent reflections on COVID-19 – how are topical emphases changing within 2020? Previously we explored shifts from pre-2020 to 2020 publications (May 6 search report).

We matched the current 15752 publications in 2020 from the May 21 34181 dataset against the prior 10346 for 2020 from the May 6 dataset.

- 5406 have been added since our May 6 download (i.e., ~50%)
- Checking publication dates of the 5406 “new”:
 - 192 dated 2020 only
 - 31 dated in Jan or Feb
 - 58 dated in March
 - 4468 dated in May
 So, these are, by and large, really recent publications.
- In the 15752 total 2020 papers file in *VantagePoint*, made a 2-value field of the 5406 Recent additions and the 10346 prior 2020 publications.
- Selective topical contrasts are of interest. MeSH terms are available in 13% or fewer of the records (13% for Descriptors; 3% for Primary or Top Level). Abstract NLP phrases are present for 49% of the 15752 2020 publications. Title NLP phrases, in contrast, appear in 99% of the records. Of those, 637 appear in 10 or more records. A matrix of those
- Table 1 shows the title phrases with highest ratios of appearances in the newer (since May 6 download) versus earlier (2020 publications in the May 6 download). It is limited to those with more records in the new download².
 - One might peruse the values sorted from highest ratio of new/earlier to search for potentially interesting articles.
 - A few terms that pop out: urology, child, recovery,

² “available” was extracted from “not available” – of no substantive interest & removed.

Table 1. Title Phrases appearing more often in the most recent 2020 publications

	# Records	10346	5406	
# Records	Title (NLP) (Phrases) >=10	in May 6 download	since May 6 download	Ratio
11	Urology	2	9	450%
15	child	4	11	275%
11	epicenter	3	8	267%
18	recovery	5	13	260%
10	innovation	3	7	233%
10	symptom	3	7	233%
13	expression	4	9	225%
13	mechanical ventilation	4	9	225%
16	follow-up	5	11	220%
16	multiple sclerosis	5	11	220%
12	COVID-19 Testing	4	8	200%
12	healthcare	4	8	200%
12	heart failure	4	8	200%
18	literature	6	12	200%
24	anxiety	8	16	200%
20	Canada	7	13	186%
17	Impacts	6	11	183%
11	neurosurgeons	4	7	175%
11	psychological distress	4	7	175%
11	SARS-CoV-2 RNA	4	7	175%
11	State	4	7	175%
22	women	8	14	175%
16	Intubation	6	10	167%
16	severe Covid-19 pneumonia	6	10	167%
13	Cardiovascular diseases	5	8	160%
13	corticosteroids	5	8	160%
13	vitamin D	5	8	160%
18	adolescents	7	11	157%
10	COVID	4	6	150%
10	depression	4	6	150%
10	Influenza	4	6	150%
10	men	4	6	150%
10	neurological manifestations	4	6	150%
10	Place	4	6	150%
10	psychiatry	4	6	150%

10	Smoking	4	6	150%
15	radiotherapy	6	9	150%
27	results	11	16	145%
34	Lockdown	14	20	143%
12	emergency	5	7	140%
12	Parkinson's Disease	5	7	140%
12	severe COVID-19 patients	5	7	140%
12	sex	5	7	140%
12	Smell	5	7	140%
12	surge	5	7	140%
12	validation	5	7	140%
31	Opportunities	13	18	138%
19	Pulmonary Embolism	8	11	138%
26	Covid-19 lockdown	11	15	136%
14	acute Kidney Injury	6	8	133%
14	Nepal	6	8	133%
102	COVID-19 Era	44	58	132%
23	ethics	10	13	130%
27	caring	12	15	125%
36	COVID-19 patient	16	20	125%
20	New York City	9	11	122%
11	COVID-19 quarantine	5	6	120%
11	persons	5	6	120%
11	Susceptibility	5	6	120%
11	TMPRSS2	5	6	120%
22	fear	10	12	120%
13	asthma	6	7	117%
13	quality	6	7	117%
26	efficacy	12	14	117%
15	lungs	7	8	114%
15	SARS-CoV-2 pneumonia	7	8	114%
15	scoping review	7	8	114%
32	covid-19 response	15	17	113%
17	Vaccines	8	9	113%
19	inflammatory bowel disease	9	10	111%
29	health	14	15	107%