

Brief Report: Research Publications Related to the Corona Virus from PUBMED search July 1, 2020

We have been downloading abstract record sets from the PubMed database relating to corona virus every 2 weeks, starting March 25. See: <http://www.techminingforglobalgood.org/open-covid-19-research-for-analysis/>.

For the first several datasets, we generated basic reports, summarizing rudimentary analyses to introduce the data for potentially deeper probes. Those reports remain available; they variously address:

- The data & background on the searches
- 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)
- Changing topical emphases
- Exploratory analysis (trying one novel empirical effort each time) – e.g., for May 21, we compared topic emphases through May 6, 2020 to emphases from then forward.

Here, we offer a more concise overview with a new exploration, but first address issues in searching and downloading the records. Starting with May 20, we encountered limitations in PubMed downloading.

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”]¹

- "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”²

The interface now limits downloads to the first 10,000 records, so we ran a combination of partial searches to recombine in *VantagePoint* on desktop. We use the core query: (search query as just above), plus additional constraints or terms to divide into datasets <=10000.

- We divided year sets (the PubMed interface gives a slider to choose years) into 1949-2005; 2006-2019; 2020-2021. Then downloaded the first two datasets separately.
- Further restricted 2020-21, as it had 29165 records [2 of those dated 2021]
 - (search query as above) with added terms to restrict further:
 - AND coronavirus AND clinical = 9713 saved
 - AND coronavirus NOT clinical (= 10876 – so need to reduce further):
 - AND viral: 6732 saved
 - NOT viral: 4144 saved
 - [search query as above] NOT coronavirus = 11556
 - AND clinical = 2823 saved
 - NOT clinical = 8733 saved
- Downloaded by choosing: Save; Selection: all results; Format: PubMed; Create File

¹ We use the legacy, peer-reviewed journal papers for consistency. PMC includes preprints, many lacking metadata indexing and some full text that does not combine well with analyses of abstract records.

² 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].

- [downloads to desktop computer]
- Imported using the PubMed configuration file; removed duplicates; and thereby consolidated these in one *VantagePoint* file = 47607 records.

****Comparisons of prior year datasets indicate high stability (i.e., only 5 new records on July 1 dated pre-2020). So, in future searches, one could just add for 2020 on and consolidate with prior year results (removing duplicates).**

Results

Series of Bi-weekly searches:

- March 25, 2020: Initial dataset - 19538 records.
- April 8, 2020: 21314
- April 22, 2020: 24479
- May 6, 2020: 28264
- May 21, 2020: 34181
- June 3, 2020: 38183
- June 17, 2020: 43254
- July 1, 2020: 47607

Explosive growth in COVID-19 research publication this year continues!

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

We have spotlighted different special analyses in the biweekly reports, to stimulate consideration of other analytical inquiries that might generate research intelligence of value in countering COVID-19. Let us know if you have a particular interest. Prior analyses include:

- March 25 and several following biweekly reports gave screenshots of the dataset in *VantagePoint* and showed the explosive growth trend for this literature. They also tallied statistics on research participation, publications, and topical emphases.
- March 25 additionally mapped **collaboration among countries**.
- April 8 broke out **top MeSH (Medical Subject Headings) Descriptors by recent month**
- April 22 **clustered topics**
- May 6 compared MeSH Qualifiers pre-2020 vs. 2020 to date.
- May 21 spotlighted **topics (title phrases) growing most strongly in the prior 2 weeks**.
- June 3 identified leading author organizations publishing in this literature in 2020 and MeSH Qualifier Topics escalating in coverage from Jan.-Apr, 2020 to May, 2020 forward.
- June 17 and July 1 explore **prevalence of terms associated with different approaches** to addressing the pandemic.

This week we again explore topical coverage as it relates to four approaches to addressing COVID-19, as distinguished by Ron Kostoff and colleagues^{3 4}. We proceed as follows:

- Use *VantagePoint* to extract noun phrases in titles, and in abstracts, using its Natural Language Processing (NLP)
- Run the thesauri and list cleanup routines of “Refine NLP” on the abstract (NLP) phrases; and on the title (NLP) phrases; then combine those with MeSH Descriptors.
- Extract “My Keywords” terms/phrases associated with each of Kostoff’s 4 approaches from the 431398 combo NLP terms to analyze [without removing those appearing in a single record]; also FIND the 3 contingent terms in the Abstract field to yield 125 keywords.

The extracted keywords reflect different factors possibly bearing upon the COVID-19 pandemic. We reviewed results of June 17 to inspect how the terms found in the core PubMed (legacy) COVID-19 abstract records were used. Some terms can play various roles vis-à-vis COVID-19. For instance, depression was included by Kostoff as a possible factor damaging immune system resistance to such a viral attack. However, in the COVID-19 literature, we found depression addressed as an effect of social isolation.

Based on such examination, and recognizing that no categorization will be definitive, we revise the 129 keywords by removing 5, re-stemming a couple, and re-categorizing some 5 others. Furthermore, we added contingency conditions to 3:

- chronic stress AND immun
- alcohol AND lifestyle
- air pollution AND (suscept or incidence or factor)

So in analyzing the July 1 data, we search for 122 keywords, and then add the 3 with contingencies just noted.

Of those 125 terms/phrases for which we searched in the 47607 abstract records, we found 86 of these target keywords. Here are total record counts (with how many terms each):

- 1) Vaccines to prevent COVID-19 infection 3290 (3)
[ranging from 2 for virus interference to 3249 for vaccine]
- 2) Restrict exposure to COVID-19 virus 2120 (9)
[ranging from 6 for hand-washing to 1104 for quarantine]
- 3) Tactical – reduce effect of COVID-19 virus 1183 (32)
[ranging from 1 for avigan & 2 others to 286 for lopinavir]
- 4) Strategic means to bolster Immune system
 - Biotoxins 1028 (6)
[ranging from 1 for T-2 toxin to 759 for respiratory syncytial virus]
 - Iatrogenic 992 (11)
[ranging from 1 for gamma radiation and another to 507 for immunosuppr]
 - Lifestyle 801 (6)
[ranging from 2 for high-fat diet to 621 for obes (obesity)]
 - Occup/env 159 (16)
[ranging from 1 for WiFi and another to 89 for nanoparticle]
 - Psych/soc/ec 6 (3)
[ranging from 1 for childhood adversity to 4 for (chronic stress AND immune)]

³ Kostoff, R.N., Briggs, M.B., Porter, A.L., Aschner, M., Spandidos, D.A., and Tsatsakis, A. (2020), COVID-19: Post-lockdown guidelines, *International Journal of Molecular Medicine* 46, 463-466; DOI: 10.3892/ijmm.2020.4640.

⁴ Kostoff, R.N., Briggs, M.B., and Porter A.L., (in prep), COVID-19: Preventing future pandemics, Georgia Institute of Technology. 2020. PDF. <https://smartech.gatech.edu/handle/1853/62907>.

Given the vagaries of terminology and limitations of coverage, we would not make much of detailed differences. Also, the terms tap only a minority of these COVID-19 abstracts – some 8814 of 47607. Some 35% of the records lack abstracts; 44% lack MeSH descriptors. And then the particular terms that we relate to the categories of interest certainly cannot cover all aspects of research inquiries.

That said, we note the minimal research directed to psycho-social factors in contributing to the pandemic, and, conversely, the maximal attention to vaccine development. Observers might see value in increasing attention to the strategic factors potentially affecting human immune system strength, especially in moving beyond COVID-19 to reduce future pandemic risks.

Appendix 1 lists the individual terms together with the factors that they reflect.

We inquired to what degree those terms were addressed during three periods of research pertaining to COVID-19. For this analysis, we constructed the periods as

- 1) publications dated prior to 2020;
- 2) 2020-1: those dated either 2020 without further detail or January-April; and
- 3) 2020-2: those dated May, 2020, or later. Notice the explosive growth in research – the number of articles published this year exceeds all prior.

Table 1 breaks out each Factor in terms of the number of records in the given period. Roughly, early 2020 has about half as many PubMed-indexed papers as do each of the other two periods. So, one could roughly normalize by considering half the early-2020 counts.

Table 1. Shifting Factor Emphases over Time

	# Records	18451	7751	21405
# Records	8 Factors \ Date - 3 periods	2019 (1949-thru)	2020-1 (Jan-Apr)	2020-2 (May or later)
3290	Vaccination	2415	276	599
2120	Restrict Exposure	351	467	1302
1183	Tactical	211	233	739
1028	Biotoxins	935	45	48
992	Iatrogenic	368	140	484
801	Lifestyle	256	104	441
159	Occup/Env	80	14	65
6	Psych/Soc/Econ	3	0	3

Surprisingly to ourselves, as non-expert observers, research on **vaccines** appears far less emphasized this year. Conversely, aspects of **restricting exposure** (“RE”) to the virus show greatly heightened attention, as the pandemic unfolds in 2020 – e.g., quarantine. Likewise, **tactical** means to counter the virus, such as hydrochloroquine, are receiving increased research attention.

Recent attention to the strategic factors (the last five categories) is mixed. Possible contributing effects of **biotoxins** are receiving less attention than in research prior to this year – possibly reflecting their perceived less-immediate influences apt to counter the pandemic. On the other hand, **lifestyle** considerations are up to a degree, most recently.

Kostoff and colleagues make the point that the strategic factors warrant strong attention, to help counter the current corona virus threat and, especially, to diminish future viral pandemic threats. Were we able to bolster our immune systems, we would be better off in multiple respects.

Appendix 1 Terms Associated with Factors (July 1)

		heavy metal	Occup/Env
aflatoxin	Biotoxins	herbicide	Occup/Env
anatoxin-A	Biotoxins	insecticide	Occup/Env
biotoxin	Biotoxins	microplastics	Occup/Env
dietary toxic cyanobacteria	Biotoxins	nanoparticle	Occup/Env
microcystin-LR	Biotoxins	PAHs	Occup/Env
mycotoxin	Biotoxins	PCBs	Occup/Env
ochratoxin	Biotoxins	perfluorooctanesulfonate	Occup/Env
pseudomonas aeruginosa	Biotoxins	perfluorooctanoic acid	Occup/Env
respiratory syncytial virus	Biotoxins	pesticide	Occup/Env
rhinovirus	Biotoxins	PFOA	Occup/Env
streptomyces californicus	Biotoxins	PFOs	Occup/Env
T-2 toxin	Biotoxins	polyaromatic hydrocarbon	Occup/Env
yessotoxin	Biotoxins	polychlorinated biphenyls	Occup/Env
acetaminophen	Iatrogenic	sodium fluoride	Occup/Env
adjuvanted vaccine	Iatrogenic	ultraviolet radiation	Occup/Env
anaesthetic	Iatrogenic	UV radiation	Occup/Env
anesthetic	Iatrogenic	WiFi	Occup/Env
antibiotic	Iatrogenic	wireless radiation	Occup/Env
gamma radiation	Iatrogenic	childhood adversity	Psych/Soc/Ec
iatrogenic	Iatrogenic	restraint stress	Psych/Soc/Ec
immunosuppr	Iatrogenic	stressful life events	Psych/Soc/Ec
non-steroidal antiinflammatory	Iatrogenic	facemask	RE
NSAID	Iatrogenic	hand wash	RE
serotonin reuptake inhibitor	Iatrogenic	hand-washing	RE
surgical stress	Iatrogenic	large gathering	RE
chronic sleep restriction	Lifestyle	lockdown	RE
high-cholesterol diet	Lifestyle	quarantine	RE
high-fat diet	Lifestyle	sanitize	RE
obes	Lifestyle	social distanc	RE
smoking	Lifestyle	social isolation	RE
substance abuse	Lifestyle	acalabrutinib	Tactical
western-style diet	Lifestyle	actemra	Tactical
aromatic halogenated		anakinra	Tactical
disinfection	Occup/Env	antiretroviral	Tactical
benzo(a)pyrene	Occup/Env	apremilast	Tactical
cell tower	Occup/Env	arbidol	Tactical
corexit	Occup/Env	avastin	Tactical
crude oil	Occup/Env	avigan	Tactical
endocrine disrupt	Occup/Env	azithromycin	Tactical
fine particulate	Occup/Env		

baricitinib	Tactical	Alcohol AND lifestyle	Lifestyle
canakinumab	Tactical	air pollution AND	Occup/Env
chloroquine	Tactical	(suscept or incidence or	
colchicine	Tactical	factor)	
colcrys	Tactical		
convalescent plasma	Tactical		
EIDD-2801	Tactical		
favipiravir	Tactical		
fingolimod	Tactical		
galidesivir	Tactical		
gilenya	Tactical		
hydrochloroquine	Tactical		
ilaris	Tactical		
ivermectin	Tactical		
jakafi	Tactical		
kaletra	Tactical		
keczara	Tactical		
kineret	Tactical		
leronlimab	Tactical		
lopinavir	Tactical		
macrilimumab	Tactical		
methylprednisolone	Tactical		
nanomedic	Tactical		
olumiant	Tactical		
oseltamivir	Tactical		
otezla	Tactical		
remdesivir	Tactical		
rionavir	Tactical		
ruxolitinib	Tactical		
sarilumab	Tactical		
scorpion venom	Tactical		
tamiflu	Tactical		
tocilizumab	Tactical		
tofacitinib	Tactical		
umifenovir	Tactical		
xeljanz	Tactical		
antibody-dependent	Vaccine		
Fc receptor	Vaccine		
imprinting	Vaccine		
I-REV	Vaccine		
vaccin	Vaccine		
virus interference	Vaccine		
chronic stress AND immun	Psych/Soc/Ec		