

# Research Publications Related to the Corona Virus from PUBMED search May 6, 2020

## The Data

Here, we take a search query [available at “query” link below]<sup>1</sup> and applied it to the legacy<sup>2</sup> version of the PubMed database:

- "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”

At the PubMed search interface, entering the above search query, we

- “Send to” – pull down menu and select “File.”
- Pull down menu for “Format” – we choose XML
- For “Sort by” we leave as “Most Recent”
- Click “Create File” and Download

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.

We will keep the simple research profiles from prior searches available, should you want to revisit. The ones based on the prior searches contain similar content as this one, with some variations.

We make this latest dataset available as a “vpt” format file for further text analyses using *VantagePoint* software [www.theVantagePoint.com]. We also make that software available free on a 60-day trial basis for researchers and analysts who want to investigate particular treatments or biomarkers, explore associations, etc.

## Perspective on the Covid-19-related Data

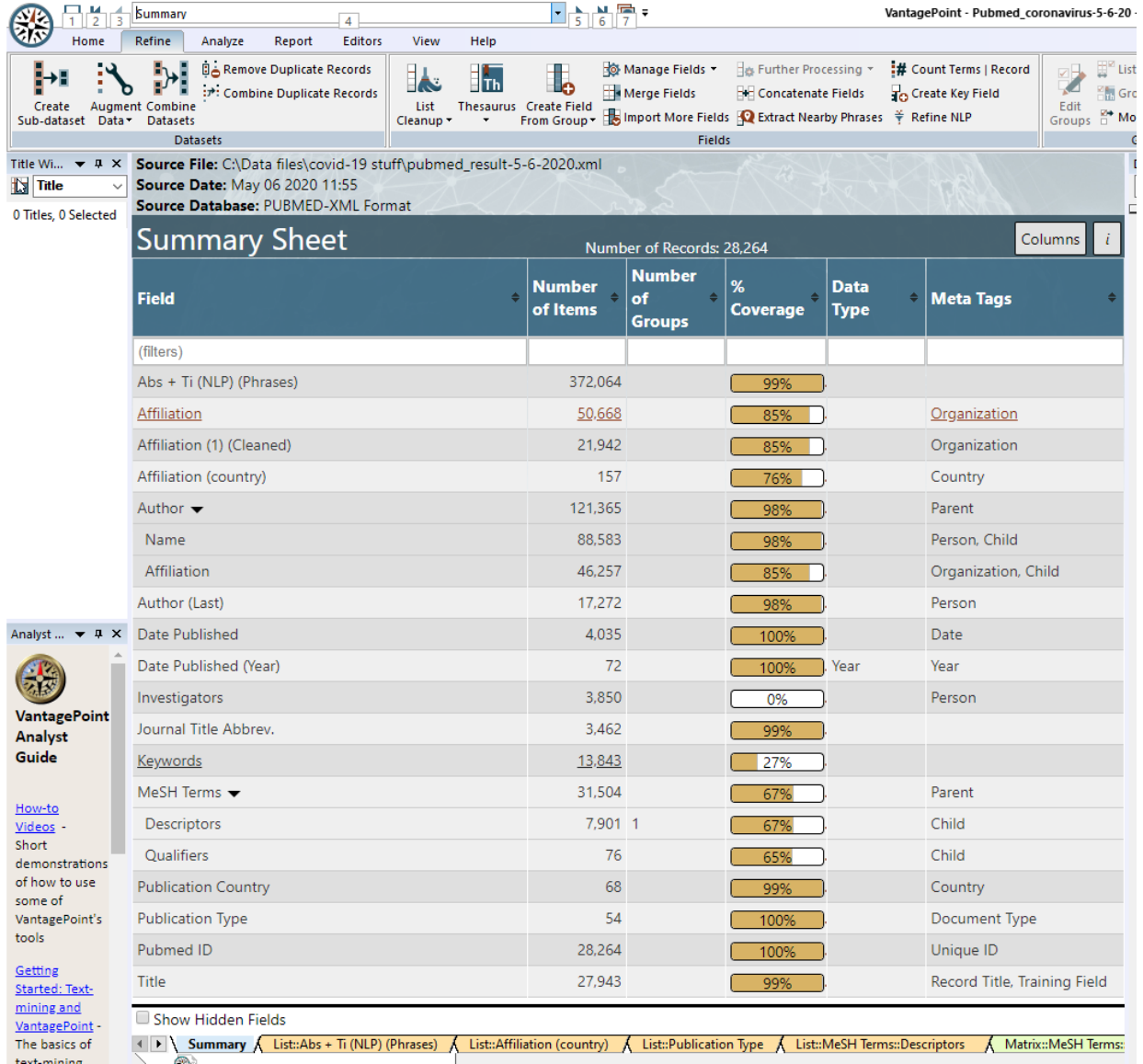
Figure 1 offers a screenshot of the file summary sheet in *VantagePoint*. We present limited data fields here to keep down clutter; Appendix 1 presents a screenshot of the window used to import additional PubMed (MEDLINE) fields of interest. This file reflects some essential data cleaning – using Natural Language Processing (NLP) routines to extract noun phrases from titles and abstracts for topical analyses. A quick tour of the dataset provides some perspective on the biomedical research activity pertaining to COVID-19, as indexed in MEDLINE.

---

<sup>1</sup> 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].

<sup>2</sup> We prefer using the legacy interface because it provides MEDLINE indexed abstract records of published research appearing in scholarly journals. This offers a good degree of comparability, whereas mining other of the compilations demands considerable attention to what is being mixed together. The PubMed PMC open access interface includes preprints, some records lack metadata, and some records are full text – so it is potentially hard to gauge how complete or representative such a dataset is.

**Figure 1. VantagePoint Summary View of the Dataset**



- **Breadth of Research Participation** is extensive:
  - Authors or co-authors of the papers reside in 156 countries or regions (such as Hong Kong, Taiwan).
  - 31 countries host researchers publishing 100 or more of these papers
  - Of these 28264 publications, the only 2 countries authoring more than 1605 of them are the US (5715) and China (5309)<sup>3</sup> – between them, 39% of the total.
  - Surprising to an outsider – e.g., Saudi Arabia shows as #16 with 476 papers.
- **Publication Types:** 23364 of these 28264 publications are journal articles (including 2373 reviews). For some analyses, one might want to focus on those.
  - Only 56 are identified as clinical trials.

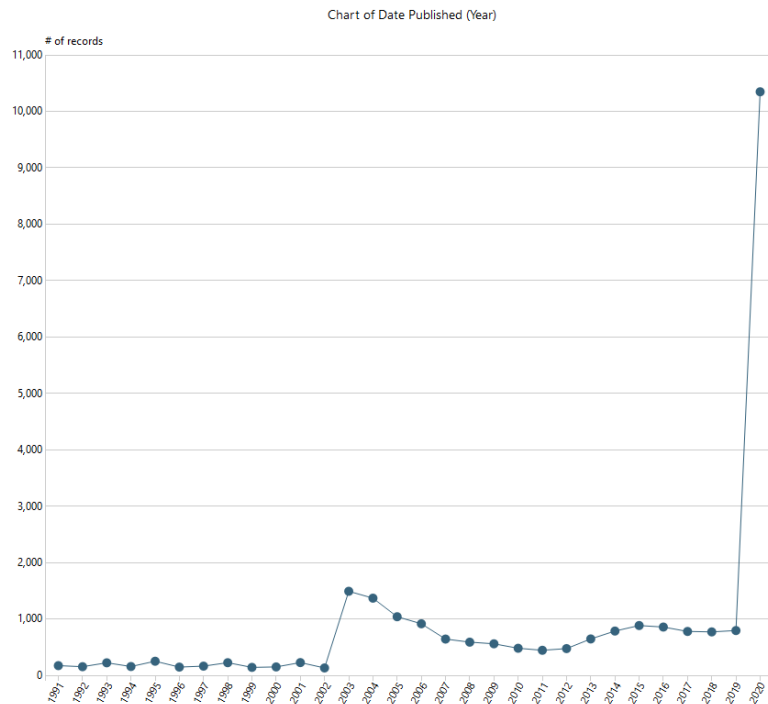
<sup>3</sup> 280 records indicate “PR”, but inspection indicates these are China (People’s Republic).

- **Topical Content:**
  - Medical Subject Heading Terms (MeSH) appear for some 67% of the records – a few tallies of MeSH descriptors:
    - About half of the papers address human subjects (10800) and half, animals (9869), but Chinese are more apt to study humans (1159) than animals (405), based on MeSH classifications
    - Animals studied are most prominently: mice (2810 of the papers), swine (1521), chickens (1086), and cattle (678)
    - Studies emphasize adults (1972) over children (813)
  - Analyses of intersections represented can help focus on particular research thrusts – a matrix of top MeSH descriptors (those appearing in >200 records) vs. top MeSH descriptors shows the 6400 papers addressing “coronavirus infections”
    - Roughly evenly split between human and animal studies, with heavy concentrations in mice, swine, and chickens, but relatively less in cattle.
    - Over 2000 paper abstracts concern both coronavirus infections and viral pneumonia
  - Analyses of MeSH qualifiers could also serve to discern pockets of research of special interest to a researcher – e.g., many studies address genetics (6221) and immunology (4649), prevention & control (3306), transmission (1967), drug effects (1218), and so forth.
  - Further analyses of the topical content (e.g., from titles and abstract noun phrases) holds promise to enable focus on particular attributes and their intersections in subsets of studies. *VantagePoint* facilitates such analyses.
  
- Quick scans of the **top researchers**<sup>4</sup> show
  - 679 organizations affiliated with 10 or more of the 28264 papers
  - 18 organizations affiliated with 100 or more papers; of which the top two are University of Hong Kong and Chinese Academy of Sciences.
  
- The **Research Trend** is informative:
  - From the first paper in 1949, we see fewer than 10 papers/year until 1962.
  - We first see over 100 papers in a year in 1981 (40 years ago!).
  - Then, over 1000 papers per year first in 2003.
  - Figure 2 shows the publication trend over the past 30 years – note the explosion of attention in 2003 and, again, now (more so) in 2020 (and that’s just in the beginning of this year).
  - Messages herein include that research attention is “exploding” nowadays, but corona-related viruses have been studied for decades.

---

<sup>4</sup> PubMed data on Author Affiliations tallied here are partly cleaned in *VantagePoint* using tailored fuzzy match, list cleanup routines tuned to authors and their affiliations. Full, detailed tuning by humans was not done.

**Figure 2. Trend in Corona Virus Related Publications from PUBMED**  
[as of May 6, 2020 PubMed COVID-19 search]



Exploring research interests; this dataset can help answer basic “**who, what, where, when?**” questions. Juxtaposing the available fields of data can point to research activity of particular concern:

- For instance, a matrix of **Authors (or Organizations) X Topics** could point to hotbeds of research on particular corona virus subjects.
- A matrix of **Topics X Years** can help gauge the emergence of various topics. Our profile based on the April 8 search results presented such a matrix. It allows one to scan topical frequency over time.
- Or, one might inspect a matrix of **Topics X Topics** to pursue intersecting factors (e.g., viral antibodies together with assays; viral RNA with cell lines).
- Another matrix could counterpose active researchers in particular countries who have recently published on a topic of special interest (e.g., American researchers engaging viral RNA re: coronaviruses).

➤ **Contrasts in Topical Emphases between 2020 and Earlier Research Literature as Indexed in PubMed**

- Note that COVID-19-related research published so far in 2020 (10,346 papers indexed in PubMed) represents 58% as much as generated in all previous years (17,918).
- Not surprisingly, emphases shift. Some **MeSH Qualifiers** showing relatively **heavy attention in 2020**<sup>5</sup>:
  - Supply & distribution (50 in 2020, up from 26 prior)
  - Diagnostic imaging (131 vs. 211 prior)
  - Therapy (319 vs. 700 prior)
  - Complications (292 vs. 665)

<sup>5</sup> We have not investigated whether National Library of Medicine has altered indexing.

- Some MeSH Qualifiers showing **diminished attention** in 2020:
    - Microbiology (2 vs. 1418)
    - Biosynthesis (7 vs. 1147)
    - Veterinary (34 vs. 3230)
    - Parasitology (0 vs. 119)
    - Radiation effects (0 vs. 53)
    - Cytology (5 vs. 363)
    - Metabolism (106 vs. 3578)
    - Immunology (136 vs. 4513)
  - Looking at the top 580 **MeSH Descriptors** (those in 50 or more records), some drawing relatively **heavy attention**:
    - Betacoronavirus (1805 in 2020 vs. 20 prior)
    - Pandemics (1890 vs. 16)
    - Pneumonia, Viral (2223 vs. 10)
    - Telemedicine (47 vs. 30)
  - Some showing **diminished attention** in 2020:
    - Molecular Sequence Data (0 vs. 1776)
    - Infectious bronchitis virus (2 vs. 1505)
    - Specific Pathogen-Free Organisms (0 vs. 276)
    - Gene Expression Regulation, Viral (0 vs. 227)
    - Encephalitis, Viral (0 vs. 54)
    - Adenoviridae Infections (0 vs. 54)
    - DNA Mutational Analysis (0 vs. 55)
- We also consolidated 25,539 terms appearing in title and abstract NLP phrases appearing in >3 records (less a few very frequent terms). Then reduced to 417  $\geq$  100 records after removing 51 less interesting terms. We applied *VantagePoint's* PCA (Principal Components Analysis) to group those terms into 19 topics, and compared their frequency in 2020 to previous.
- Most striking increase in attention is to Wuhan (2384 in 2020 vs. 5 prior), not surprisingly. Relatively high emphasis occurs for acute respiratory distress syndrome (ARDS) (556 vs. 531 prior).
  - A number of viral topics show far less attention in 2020, including: MHV, RSV, IBV, PEDV, TGEV, FIP, FCoV, BSV, and genomic RNA.
- Appendix 2 lists the full sets of MeSH Qualifiers, MeSH Descriptors, and PCA Factors with their frequencies in the 17,918 records published prior to 2020 compared to the 10,346 records published in 2020.

## Appendix 1. Additional Fields Available for Handy Import for Analyses



## Appendix 2. Frequencies of various Topics in the COVID-19 Research Literature prior to 2020 compared to in 2020 papers (based on search of PubMed on May 6, 2020).

	MeSH Terms::Qualifiers	1	2	
	# Records	17918	10346	57.7%
# Records		pre-2020	2020	
76	supply & distribution	26	50	192.3%
342	diagnostic imaging	211	131	62.1%
281	trends	179	102	57.0%
1019	therapy	700	319	45.6%
957	complications	665	292	43.9%
117	legislation & jurisprudence	82	35	42.7%
54	surgery	38	16	42.1%
174	economics	123	51	41.5%
11	agonists	8	3	37.5%
82	ethics	60	22	36.7%
429	psychology	314	115	36.6%

518	standards	382	136	35.6%
1020	drug therapy	765	255	33.3%
4	radiotherapy	3	1	33.3%
1967	transmission	1516	451	29.7%
917	therapeutic use	719	198	27.5%
434	organization & administration	342	92	26.9%
24	rehabilitation	19	5	26.3%
5308	epidemiology	4251	1057	24.9%
153	analogs & derivatives	124	29	23.4%
2528	diagnosis	2055	473	23.0%
313	adverse effects	256	57	22.3%
73	education	60	13	21.7%
3306	prevention & control	2734	572	20.9%
110	nursing	92	18	19.6%
716	statistics & numerical data	609	107	17.6%
588	mortality	504	84	16.7%
959	etiology	824	135	16.4%
26	pharmacokinetics	23	3	13.0%
55	chemically induced	49	6	12.2%
46	ethnology	41	5	12.2%
118	history	106	12	11.3%
2377	methods	2146	231	10.8%
1990	pathogenicity	1827	163	8.9%
565	physiopathology	520	45	8.7%
406	antagonists & inhibitors	377	29	7.7%
1218	drug effects	1133	85	7.5%
118	instrumentation	110	8	7.3%
16	urine	15	1	6.7%
6534	virology	6139	395	6.4%
665	administration & dosage	625	40	6.4%
1728	pathology	1627	101	6.2%
4112	isolation & purification	3886	226	5.8%
1412	pharmacology	1362	50	3.7%
2709	physiology	2617	92	3.5%
1421	blood	1373	48	3.5%
6221	genetics	6012	209	3.5%
1458	classification	1411	47	3.3%
663	enzymology	642	21	3.3%
2177	chemistry	2113	64	3.0%
4649	immunology	4513	136	3.0%
3684	metabolism	3578	106	3.0%

1803	analysis	1759	44	2.5%
525	ultrastructure	516	9	1.7%
61	toxicity	60	1	1.7%
631	growth & development	621	10	1.6%
368	cytology	363	5	1.4%
160	chemical synthesis	158	2	1.3%
86	deficiency	85	1	1.2%
3264	veterinary	3230	34	1.1%
1154	biosynthesis	1147	7	0.6%
1420	microbiology	1418	2	0.1%
119	parasitology	119		0.0%
53	radiation effects	53		0.0%
40	transplantation	40		0.0%
32	embryology	32		0.0%
21	cerebrospinal fluid	21		0.0%
16	blood supply	16		0.0%
12	anatomy & histology	12		0.0%
6	abnormalities	6		0.0%
6	injuries	6		0.0%
3	diet therapy	3		0.0%
2	poisoning	2		0.0%
2	secondary	2		0.0%
1	congenital	1		0.0%
1	innervation	1		0.0%

	<b>MeSH Terms::Descriptors</b>	1	2	
	# Records	17918	10346	57.7%
# Records		pre-2020	2020	
10800	Humans	8700	2100	24.1%
1825	Betacoronavirus	20	1805	9025.0%
1890	Pandemics	109	1781	1633.9%
2223	Pneumonia, Viral	130	2093	1610.0%
50	Telemedicine	3	47	1566.7%
260	Clinical Laboratory Techniques	40	220	550.0%
55	Personal Protective Equipment	9	46	511.1%
54	Mental Health	10	44	440.0%
62	Civil Defense	16	46	287.5%
52	Chloroquine	17	35	205.9%
87	Cough	29	58	200.0%
213	Tomography, X-Ray Computed	83	130	156.6%



53	Critical Illness	21	32	152.4%
94	Clinical Trials as Topic	38	56	147.4%
99	Delivery of Health Care	49	50	102.0%
159	Epidemics	79	80	101.3%
130	Practice Guidelines as Topic	65	65	100.0%
68	Neoplasms	35	33	94.3%
149	Italy	78	71	91.0%
101	Comorbidity	55	46	83.6%
103	Critical Care	58	45	77.6%
67	Stress, Psychological	38	29	76.3%
176	Fever	101	75	74.3%
83	Communication	49	34	69.4%
62	Forecasting	37	25	67.6%
405	Public Health	244	161	66.0%
90	Information Dissemination	55	35	63.6%
70	Medicine, Chinese Traditional	44	26	59.1%
161	United Kingdom	102	59	57.8%
93	Biomedical Research	59	34	57.6%
253	Quarantine	162	91	56.2%
1410	China	904	506	56.0%
106	Disease Progression	68	38	55.9%
6400	Coronavirus Infections	4242	2158	50.9%
66	Internationality	44	22	50.0%
111	Health Policy	75	36	48.0%
54	Emergency Medical Services	37	17	45.9%
121	Radiography, Thoracic	84	37	44.0%
107	Europe	75	32	42.7%
90	Mass Screening	64	26	40.6%
248	Peptidyl-Dipeptidase A	177	71	40.1%
85	Pregnancy Complications, Infectious	61	24	39.3%
96	Models, Theoretical	69	27	39.1%
85	Epidemiological Monitoring	62	23	37.1%
56	Sex Factors	41	15	36.6%
102	Intensive Care Units	75	27	36.0%
50	Respiratory Insufficiency	37	13	35.1%
97	Immunization, Passive	72	25	34.7%
304	Health Personnel	226	78	34.5%
83	Internet	62	21	33.9%
53	India	40	13	32.5%
188	Contact Tracing	142	46	32.4%
74	Drugs, Chinese Herbal	56	18	32.1%

80	Risk	61	19	31.1%
533	Global Health	407	126	31.0%
51	Fear	39	12	30.8%
201	Severity of Illness Index	154	47	30.5%
446	Travel	343	103	30.0%
123	Respiration, Artificial	95	28	29.5%
169	Patient Isolation	131	38	29.0%
95	Brazil	74	21	28.4%
105	Models, Statistical	82	23	28.0%
148	Prognosis	116	32	27.6%
65	Germany	51	14	27.5%
71	Interleukin-6	56	15	26.8%
199	Risk Assessment	157	42	26.8%
478	Young Adult	378	100	26.5%
345	World Health Organization	273	72	26.4%
87	Inflammation	69	18	26.1%
136	Pneumonia	108	28	25.9%
117	Respiratory Distress Syndrome, Adult	93	24	25.8%
588	Infection Control	469	119	25.4%
89	Masks	71	18	25.4%
50	Protein Domains	40	10	25.0%
61	Disease Notification	49	12	24.5%
61	Immunocompromised Host	49	12	24.5%
504	United States	407	97	23.8%
63	Age Distribution	51	12	23.5%
137	Health Knowledge, Attitudes, Practice	111	26	23.4%
53	Drug Discovery	43	10	23.3%
81	Research	66	15	22.7%
532	Risk Factors	434	98	22.6%
186	Hospitalization	153	33	21.6%
62	History, 20th Century	51	11	21.6%
142	Disaster Planning	117	25	21.4%
57	Anti-Inflammatory Agents	47	10	21.3%
82	Public Health Practice	68	14	20.6%
2462	Disease Outbreaks	2054	408	19.9%
86	Occupational Exposure	72	14	19.4%
487	Retrospective Studies	408	79	19.4%
74	Nucleic Acid Amplification Techniques	62	12	19.4%
253	Republic of Korea	212	41	19.3%
247	Pregnancy	207	40	19.3%
56	Protective Clothing	47	9	19.1%

602	Lung	506	96	19.0%
63	Data Collection	53	10	18.9%
95	Australia	80	15	18.8%
297	Infectious Disease Transmission, Patient-to-Professional	251	46	18.3%
66	Genomics	56	10	17.9%
2136	Coronavirus	1813	323	17.8%
179	Disease Transmission, Infectious	152	27	17.8%
813	Child	692	121	17.5%
55	Adaptation, Psychological	47	8	17.0%
1108	Aged	947	161	17.0%
230	Age Factors	197	33	16.8%
84	Centers for Disease Control and Prevention, U.S.	72	12	16.7%
337	Zoonoses	289	48	16.6%
50	Specimen Handling	43	7	16.3%
239	Surveys and Questionnaires	206	33	16.0%
283	Population Surveillance	244	39	16.0%
117	Emergency Service, Hospital	101	16	15.8%
83	Biomarkers	72	11	15.3%
333	Chiroptera	289	44	15.2%
1645	Middle Aged	1428	217	15.2%
69	Adrenal Cortex Hormones	60	9	15.0%
292	Infant, Newborn	254	38	15.0%
117	Computer Simulation	102	15	14.7%
79	France	69	10	14.5%
332	Real-Time Polymerase Chain Reaction	290	42	14.5%
105	History, 21st Century	92	13	14.1%
65	Sentinel Surveillance	57	8	14.0%
132	International Cooperation	116	16	13.8%
50	Lung Diseases	44	6	13.6%
50	Vaccines	44	6	13.6%
479	Aged, 80 and over	422	57	13.5%
68	Aerosols	60	8	13.3%
60	Serine Endopeptidases	53	7	13.2%
269	Cytokines	238	31	13.0%
929	Antiviral Agents	823	106	12.9%
178	Cross-Sectional Studies	158	20	12.7%
197	Treatment Outcome	175	22	12.6%
387	Communicable Disease Control	344	43	12.5%
1972	Adult	1753	219	12.5%
118	Hospitals	105	13	12.4%
174	Virus Shedding	155	19	12.3%

110	Algorithms	98	12	12.2%
66	Biological Evolution	59	7	11.9%
86	Chronic Disease	77	9	11.7%
241	Incidence	216	25	11.6%
80	Drug Evaluation, Preclinical	72	8	11.1%
70	Predictive Value of Tests	63	7	11.1%
50	Occupational Health	45	5	11.1%
61	Glucocorticoids	55	6	10.9%
217	Viral Load	196	21	10.7%
53	Asia	48	5	10.4%
215	Disease Reservoirs	195	20	10.3%
625	Time Factors	567	58	10.2%
883	Adolescent	805	78	9.7%
102	Serologic Tests	93	9	9.7%
126	Japan	115	11	9.6%
154	Disease Susceptibility	141	13	9.2%
3358	Male	3078	280	9.1%
72	Protein Multimerization	66	6	9.1%
86	Personnel, Hospital	79	7	8.9%
261	Virus Internalization	240	21	8.8%
87	Gene Expression Profiling	80	7	8.8%
50	Software	46	4	8.7%
3753	Female	3459	294	8.5%
212	Diagnosis, Differential	196	16	8.2%
122	Cohort Studies	113	9	8.0%
841	Genome, Viral	779	62	8.0%
96	Anti-Bacterial Agents	89	7	7.9%
166	Reproducibility of Results	154	12	7.8%
669	Child, Preschool	621	48	7.7%
448	Receptors, Virus	416	32	7.7%
154	Nasopharynx	143	11	7.7%
70	Fatal Outcome	65	5	7.7%
635	Infant	591	44	7.4%
94	Hemorrhagic Fever, Ebola	88	6	6.8%
259	Antibodies, Neutralizing	243	16	6.6%
429	Cross Infection	403	26	6.5%
99	Community-Acquired Infections	93	6	6.5%
99	Immunoglobulin M	93	6	6.5%
100	Animals, Wild	94	6	6.4%
50	Respiratory Protective Devices	47	3	6.4%
153	Protease Inhibitors	144	9	6.3%

51	Nasal Cavity	48	3	6.3%
51	Organ Specificity	48	3	6.3%
137	Drug Design	129	8	6.2%
52	Hot Temperature	49	3	6.1%
52	Up-Regulation	49	3	6.1%
1111	Viral Vaccines	1047	64	6.1%
87	Molecular Diagnostic Techniques	82	5	6.1%
71	Animal Husbandry	67	4	6.0%
71	Multiple Sclerosis	67	4	6.0%
71	Public Health Administration	67	4	6.0%
107	Respiratory System	101	6	5.9%
74	Trypsin	70	4	5.7%
130	HEK293 Cells	123	7	5.7%
56	Down-Regulation	53	3	5.7%
56	Leukocyte Count	53	3	5.7%
412	Models, Molecular	390	22	5.6%
94	Dipeptidyl Peptidase 4	89	5	5.6%
593	Influenza, Human	562	31	5.5%
136	Coinfection	129	7	5.4%
216	Singapore	205	11	5.4%
642	Communicable Diseases, Emerging	610	32	5.2%
81	Occupational Diseases	77	4	5.2%
61	Respiratory Tract Diseases	58	3	5.2%
143	Computational Biology	136	7	5.1%
1024	Middle East Respiratory Syndrome Coronavirus	974	50	5.1%
82	Interferon-beta	78	4	5.1%
82	Models, Genetic	78	4	5.1%
151	Influenza A Virus, H1N1 Subtype	144	7	4.9%
65	Methylprednisolone	62	3	4.8%
87	Peptide Hydrolases	83	4	4.8%
1117	Phylogeny	1066	51	4.8%
1355	Spike Glycoprotein, Coronavirus	1295	60	4.6%
226	Evolution, Molecular	216	10	4.6%
3003	SARS Virus	2871	132	4.6%
253	Protein Conformation	242	11	4.5%
347	Sequence Alignment	332	15	4.5%
95	Virology	91	4	4.4%
95	Virus Attachment	91	4	4.4%
528	Sensitivity and Specificity	506	22	4.3%
74	Catalytic Domain	71	3	4.2%
401	Prevalence	385	16	4.2%

76	Epitopes, T-Lymphocyte	73	3	4.1%
890	Reverse Transcriptase Polymerase Chain Reaction	855	35	4.1%
51	Air Microbiology	49	2	4.1%
51	Animal Feed	49	2	4.1%
77	Viral Tropism	74	3	4.1%
237	Seasons	228	9	3.9%
79	Survival Rate	76	3	3.9%
53	Immunoglobulins	51	2	3.9%
186	Cluster Analysis	179	7	3.9%
80	Drug Therapy, Combination	77	3	3.9%
243	Canada	234	9	3.8%
54	HIV-1	52	2	3.8%
110	Cell Line, Tumor	106	4	3.8%
140	Vaccines, Synthetic	135	5	3.7%
84	Logistic Models	81	3	3.7%
199	Crystallography, X-Ray	192	7	3.6%
114	Molecular Epidemiology	110	4	3.6%
57	Vietnam	55	2	3.6%
118	Membrane Fusion	114	4	3.5%
296	Host-Pathogen Interactions	286	10	3.5%
90	Cell Fusion	87	3	3.4%
61	Lymphocytes	59	2	3.4%
94	Peptide Fragments	91	3	3.3%
94	Tumor Necrosis Factor-alpha	91	3	3.3%
597	Polymerase Chain Reaction	578	19	3.3%
189	Models, Biological	183	6	3.3%
127	Immunoglobulin A	123	4	3.3%
546	Diarrhea	529	17	3.2%
259	Taiwan	251	8	3.2%
99	Adjuvants, Immunologic	96	3	3.1%
266	Immunity, Innate	258	8	3.1%
134	Interferons	130	4	3.1%
305	Viral Nonstructural Proteins	296	9	3.0%
102	Mice, Transgenic	99	3	3.0%
103	Vaccines, DNA	100	3	3.0%
71	Aging	69	2	2.9%
72	Macaca mulatta	70	2	2.9%
328	Protein Structure, Tertiary	319	9	2.8%
221	T-Lymphocytes	215	6	2.8%
186	Seroepidemiologic Studies	181	5	2.8%
75	Attitude of Health Personnel	73	2	2.7%

75	Immune Sera	73	2	2.7%
113	Dose-Response Relationship, Drug	110	3	2.7%
305	Binding Sites	297	8	2.7%
115	Follow-Up Studies	112	3	2.7%
393	Immunoglobulin G	383	10	2.6%
674	Feces	657	17	2.6%
162	Cross Reactions	158	4	2.5%
81	Administration, Intranasal	79	2	2.5%
205	RNA Replicase	200	5	2.5%
534	Antibodies, Monoclonal	521	13	2.5%
4557	Severe Acute Respiratory Syndrome	4447	110	2.5%
83	HIV Infections	81	2	2.5%
296	Cysteine Endopeptidases	289	7	2.4%
296	Epitopes	289	7	2.4%
85	In Vitro Techniques	83	2	2.4%
1751	RNA, Viral	1710	41	2.4%
86	B-Lymphocytes	84	2	2.4%
87	RNA	85	2	2.4%
87	Sequence Analysis, RNA	85	2	2.4%
131	Ribavirin	128	3	2.3%
1340	Virus Replication	1310	30	2.3%
180	CD8-Positive T-Lymphocytes	176	4	2.3%
92	Central Nervous System	90	2	2.2%
92	Multiplex Polymerase Chain Reaction	90	2	2.2%
235	Prospective Studies	230	5	2.2%
94	Radiography	92	2	2.2%
424	Disease Models, Animal	415	9	2.2%
9869	Animals	9661	208	2.2%
96	Molecular Structure	94	2	2.1%
2029	Antibodies, Viral	1987	42	2.1%
97	Intestine, Small	95	2	2.1%
342	Genetic Variation	335	7	2.1%
98	Immunoenzyme Techniques	96	2	2.1%
98	RNA, Small Interfering	96	2	2.1%
304	Recombination, Genetic	298	6	2.0%
203	Dog Diseases	199	4	2.0%
106	Nucleocapsid	104	2	1.9%
429	Protein Binding	421	8	1.9%
108	Survival Analysis	106	2	1.9%
54	3' Untranslated Regions	53	1	1.9%
54	Nursing Staff, Hospital	53	1	1.9%

54	Syndrome	53	1	1.9%
757	Swine Diseases	743	14	1.9%
163	Signal Transduction	160	3	1.9%
55	Chemokines	54	1	1.9%
55	NF-kappa B	54	1	1.9%
167	Structure-Activity Relationship	164	3	1.8%
56	RNA Interference	55	1	1.8%
226	Camelus	222	4	1.8%
113	Protein Structure, Secondary	111	2	1.8%
57	Multivariate Analysis	56	1	1.8%
59	Leukocytes, Mononuclear	58	1	1.7%
119	Immunity, Cellular	117	2	1.7%
477	Mutation	469	8	1.7%
180	Epithelial Cells	177	3	1.7%
180	Immunization	177	3	1.7%
60	Carcinoembryonic Antigen	59	1	1.7%
60	DNA-Directed RNA Polymerases	59	1	1.7%
60	Random Allocation	59	1	1.7%
60	Testis	59	1	1.7%
486	Hong Kong	478	8	1.7%
61	Cell Nucleus	60	1	1.7%
1229	Amino Acid Sequence	1209	20	1.7%
187	Cat Diseases	184	3	1.6%
63	Ferrets	62	1	1.6%
255	Gene Expression	251	4	1.6%
64	Antibodies	63	1	1.6%
64	Coronavirus NL63, Human	63	1	1.6%
65	Papain	64	1	1.6%
67	Enzyme Inhibitors	66	1	1.5%
135	CD4-Positive T-Lymphocytes	133	2	1.5%
68	Microbial Sensitivity Tests	67	1	1.5%
414	Vaccination	408	6	1.5%
351	Dogs	346	5	1.4%
214	Peptides	211	3	1.4%
72	Cell Proliferation	71	1	1.4%
72	Epitope Mapping	71	1	1.4%
72	Phenotype	71	1	1.4%
72	Point Mutation	71	1	1.4%
293	Brain	289	4	1.4%
74	Oligonucleotide Array Sequence Analysis	73	1	1.4%
817	Chlorocebus aethiops	806	11	1.4%



743	Vero Cells	733	10	1.4%
223	Recombinant Fusion Proteins	220	3	1.4%
300	Saudi Arabia	296	4	1.4%
907	Respiratory Tract Infections	895	12	1.3%
152	Coronavirus OC43, Human	150	2	1.3%
1300	Viral Proteins	1283	17	1.3%
155	Gene Expression Regulation	153	2	1.3%
233	Virion	230	3	1.3%
634	Virus Diseases	626	8	1.3%
636	Porcine epidemic diarrhea virus	628	8	1.3%
81	Picornaviridae Infections	80	1	1.3%
490	Recombinant Proteins	484	6	1.2%
82	Dendritic Cells	81	1	1.2%
493	Nucleocapsid Proteins	487	6	1.2%
85	Endopeptidases	84	1	1.2%
260	Kidney	257	3	1.2%
88	Vaccines, Inactivated	87	1	1.1%
1521	Swine	1504	17	1.1%
90	CD13 Antigens	89	1	1.1%
182	Apoptosis	180	2	1.1%
376	DNA, Viral	372	4	1.1%
94	Rotavirus Infections	93	1	1.1%
95	Intestinal Mucosa	94	1	1.1%
96	Conserved Sequence	95	1	1.1%
96	Cytoplasm	95	1	1.1%
96	Glycosylation	95	1	1.1%
292	Viral Matrix Proteins	289	3	1.0%
297	Liver	294	3	1.0%
101	Sequence Homology	100	1	1.0%
203	Ontario	201	2	1.0%
103	Respiratory Syncytial Virus Infections	102	1	1.0%
104	Hydrogen-Ion Concentration	103	1	1.0%
104	Paramyxoviridae Infections	103	1	1.0%
104	Substrate Specificity	103	1	1.0%
1670	Cell Line	1654	16	1.0%
209	Influenza A virus	207	2	1.0%
317	Genotype	314	3	1.0%
107	Interferon-alpha	106	1	0.9%
219	Acute Disease	217	2	0.9%
225	Glycoproteins	223	2	0.9%
804	Antigens, Viral	797	7	0.9%

2810	Mice	2786	24	0.9%
1173	Viral Envelope Proteins	1163	10	0.9%
119	Interferon Type I	118	1	0.8%
119	Orthomyxoviridae	118	1	0.8%
597	Neutralization Tests	592	5	0.8%
121	Lymphocyte Activation	120	1	0.8%
243	Sequence Homology, Amino Acid	241	2	0.8%
123	Virus Assembly	122	1	0.8%
247	Kinetics	245	2	0.8%
124	Adenoviridae	123	1	0.8%
753	Enzyme-Linked Immunosorbent Assay	747	6	0.8%
126	Metapneumovirus	125	1	0.8%
126	Sequence Deletion	125	1	0.8%
127	Mice, Knockout	126	1	0.8%
255	Escherichia coli	253	2	0.8%
255	Macrophages	253	2	0.8%
260	Animals, Newborn	258	2	0.8%
135	Sequence Homology, Nucleic Acid	134	1	0.7%
141	Amino Acid Substitution	140	1	0.7%
143	Flow Cytometry	142	1	0.7%
440	Virulence	437	3	0.7%
147	Turkeys	146	1	0.7%
151	Antibody Formation	150	1	0.7%
454	RNA, Messenger	451	3	0.7%
303	Rats	301	2	0.7%
609	Cats	605	4	0.7%
155	Coronavirus, Canine	154	1	0.6%
157	Rhinovirus	156	1	0.6%
159	DNA, Complementary	158	1	0.6%
163	RNA Viruses	162	1	0.6%
168	Temperature	167	1	0.6%
343	Open Reading Frames	341	2	0.6%
173	Communicable Diseases	172	1	0.6%
176	Plasmids	175	1	0.6%
535	Sequence Analysis, DNA	532	3	0.6%
185	Rabbits	184	1	0.5%
190	Virus Cultivation	189	1	0.5%
194	Viral Structural Proteins	193	1	0.5%
204	Coronavirus 229E, Human	203	1	0.5%
205	Spleen	204	1	0.5%
225	Genetic Vectors	224	1	0.4%

709	Mice, Inbred BALB C	706	3	0.4%
487	Mice, Inbred C57BL	485	2	0.4%
291	Chick Embryo	290	1	0.3%
293	Fluorescent Antibody Technique	292	1	0.3%
299	Transcription, Genetic	298	1	0.3%
327	DNA Primers	326	1	0.3%
329	Species Specificity	328	1	0.3%
329	Viruses	328	1	0.3%
1332	Coronaviridae	1328	4	0.3%
364	Cloning, Molecular	363	1	0.3%
391	Coronavirus, Feline	390	1	0.3%
912	Poultry Diseases	910	2	0.2%
495	Genes, Viral	494	1	0.2%
1086	Chickens	1084	2	0.2%
1122	Base Sequence	1120	2	0.2%
615	Cells, Cultured	614	1	0.2%
678	Cattle	677	1	0.1%
1507	Infectious bronchitis virus	1505	2	0.1%
777	Transmissible gastroenteritis virus	776	1	0.1%
921	Membrane Glycoproteins	920	1	0.1%
2101	Murine hepatitis virus	2100	1	0.0%
1776	Molecular Sequence Data	1776		0.0%
668	Coronaviridae Infections	668		0.0%
479	Hepatitis, Viral, Animal	479		0.0%
385	Cattle Diseases	385		0.0%
362	Gastroenteritis, Transmissible, of Swine	362		0.0%
315	Microscopy, Electron	315		0.0%
276	Specific Pathogen-Free Organisms	276		0.0%
274	Feline Infectious Peritonitis	274		0.0%
262	Coronavirus, Bovine	262		0.0%
252	Cricetinae	252		0.0%
239	Blotting, Western	239		0.0%
230	Cytopathogenic Effect, Viral	230		0.0%
227	Gene Expression Regulation, Viral	227		0.0%
219	Interferon-gamma	219		0.0%
218	Demyelinating Diseases	218		0.0%
216	Vaccines, Attenuated	216		0.0%
201	Immunohistochemistry	201		0.0%
193	Transfection	193		0.0%
189	Trachea	189		0.0%
183	Intestines	183		0.0%

168	Molecular Weight	168		0.0%
166	Cell Membrane	166		0.0%
159	Protein Biosynthesis	159		0.0%
156	HeLa Cells	156		0.0%
156	Middle East	156		0.0%
156	Rotavirus	156		0.0%
153	Newcastle disease virus	153		0.0%
152	Case-Control Studies	152		0.0%
148	Mutagenesis, Site-Directed	148		0.0%
146	Capsid	146		0.0%
145	Mice, Inbred Strains	145		0.0%
141	Viral Plaque Assay	141		0.0%
140	Electrophoresis, Polyacrylamide Gel	140		0.0%
139	Viral Fusion Proteins	139		0.0%
136	Protein Processing, Post-Translational	136		0.0%
132	Birds	132		0.0%
129	Hemagglutination Inhibition Tests	129		0.0%
124	Encephalomyelitis	124		0.0%
124	Hemagglutinins, Viral	124		0.0%
122	Antibody Specificity	122		0.0%
120	Nucleic Acid Conformation	120		0.0%
118	Fluorescent Antibody Technique, Indirect	118		0.0%
118	Spinal Cord	118		0.0%
118	Tumor Cells, Cultured	118		0.0%
116	Viral Core Proteins	116		0.0%
114	Coronavirus, Turkey	114		0.0%
114	DNA	114		0.0%
110	Rodent Diseases	110		0.0%
106	Cell Adhesion Molecules	106		0.0%
103	Respiratory Syncytial Viruses	103		0.0%
101	Influenza A Virus, H5N1 Subtype	101		0.0%
101	Influenza in Birds	101		0.0%
99	Membrane Proteins	99		0.0%
96	Hepatitis, Animal	96		0.0%
93	Serotyping	93		0.0%
90	Bacterial Infections	90		0.0%
90	Enteritis	90		0.0%
89	Vaccinia virus	89		0.0%
87	Defective Viruses	87		0.0%
86	Gastroenteritis	86		0.0%
86	Golgi Apparatus	86		0.0%

85	Antigens, CD	85		0.0%
85	Enteritis, Transmissible, of Turkeys	85		0.0%
84	Endoplasmic Reticulum	84		0.0%
84	T-Lymphocytes, Cytotoxic	84		0.0%
83	Milk	83		0.0%
82	Poultry	82		0.0%
82	Promoter Regions, Genetic	82		0.0%
81	Administration, Oral	81		0.0%
80	Horses	80		0.0%
79	Newcastle Disease	79		0.0%
77	L Cells	77		0.0%
76	Hepatitis A	76		0.0%
75	Antibodies, Bacterial	75		0.0%
75	Escherichia coli Infections	75		0.0%
74	Bronchitis	74		0.0%
74	Green Fluorescent Proteins	74		0.0%
74	Lymph Nodes	74		0.0%
73	Neurons	73		0.0%
73	T-Lymphocyte Subsets	73		0.0%
72	Fibroblasts	72		0.0%
72	Immunity, Maternally-Acquired	72		0.0%
71	Common Cold	71		0.0%
71	Microscopy, Fluorescence	71		0.0%
71	Nasal Mucosa	71		0.0%
70	Phosphorylation	70		0.0%
69	Amino Acid Motifs	69		0.0%
69	Mice, Inbred C3H	69		0.0%
69	Nucleic Acid Hybridization	69		0.0%
69	RNA-Binding Proteins	69		0.0%
68	Bird Diseases	68		0.0%
67	Bioterrorism	67		0.0%
67	Colostrum	67		0.0%
67	Dimerization	67		0.0%
67	Mice, Nude	67		0.0%
66	Cell Survival	66		0.0%
66	Mice, Inbred A	66		0.0%
65	Gene Deletion	65		0.0%
65	Parvoviridae Infections	65		0.0%
63	Bacteria	63		0.0%
63	Monocytes	63		0.0%
63	Oligodendroglia	63		0.0%

63	Orthomyxoviridae Infections	63		0.0%
63	Reoviridae	63		0.0%
62	Hemagglutination Tests	62		0.0%
61	Animals, Suckling	61		0.0%
61	Astrocytes	61		0.0%
60	Genetic Predisposition to Disease	60		0.0%
60	Immunity, Humoral	60		0.0%
60	Mutagenesis	60		0.0%
58	In Situ Hybridization	58		0.0%
58	Restriction Mapping	58		0.0%
57	5' Untranslated Regions	57		0.0%
57	Genes, Reporter	57		0.0%
57	Microscopy, Confocal	57		0.0%
56	COS Cells	56		0.0%
56	Hybridomas	56		0.0%
56	Infectious bursal disease virus	56		0.0%
56	Protein Transport	56		0.0%
56	Serial Passage	56		0.0%
55	Baculoviridae	55		0.0%
55	DNA Mutational Analysis	55		0.0%
55	Peptide Library	55		0.0%
54	Adenoviridae Infections	54		0.0%
54	Encephalitis, Viral	54		0.0%
53	Cytotoxicity, Immunologic	53		0.0%
53	Immunity, Mucosal	53		0.0%
52	Blotting, Northern	52		0.0%
52	Mycoplasma Infections	52		0.0%
52	Organ Culture Techniques	52		0.0%
52	Precipitin Tests	52		0.0%
51	Cell Differentiation	51		0.0%
51	Enzyme Activation	51		0.0%
51	Germ-Free Life	51		0.0%
50	Body Weight	50		0.0%
50	Culture Techniques	50		0.0%
50	Jejunum	50		0.0%

	<b>PCA groups of Title &amp; Abstract NLP Phrases</b>	<b>1</b>	2	
	# Records	17918	10346	57.7%
# Records		pre-2020	2020	
2389	Map: Wuhan	5	2384	47680.0%

2222	Map: SARS-CoV	1891	331	17.5%
2063	Map: MERS-CoV	1724	339	19.7%
1893	Map: mouse hepatitis virus (MHV)	1846	47	2.5%
1457	Map: antibodies	1384	73	5.3%
1327	Map: respiratory syncytial virus (RSV)	1201	126	10.5%
1323	Map: infectious bronchitis virus (IBV)	1301	22	1.7%
1166	Map: lungs	972	194	20.0%
1087	Map: acute respiratory distress syndrome (ARDS)	531	556	104.7%
1075	Map: calves	976	99	10.1%
847	Map: porcine epidemic diarrhea virus (PEDV)	823	24	2.9%
665	Map: TGEV	660	5	0.8%
639	Map: S protein	607	32	5.3%
607	Map: HCoV	554	53	9.6%
550	Map: feline infectious peritonitis (FIP)	538	12	2.2%
388	Map: FCoV	377	11	2.9%
364	Map: genomic RNA	359	5	1.4%
156	Map: bovine corona virus (BCV)	154	2	1.3%
141	Map: b>Conclusion	25	116	464.0%

## Background

The U.S. Government is helping to provide access to research related to Covid-19. A number of resources are being made available, including:

- <https://www.whitehouse.gov/briefings-statements/call-action-tech-community-new-machine-readable-covid-19-dataset/>
- Follow the link to here: <https://pages.semanticscholar.org/coronavirus-research>
- Follow the link to “PubMed's PMC open access corpus using this [query](#) (COVID-19 and coronavirus research)”
- Also, among many resources: *NIH Extramural Nexus* listserv: <https://list.nih.gov/cgi-bin/wa.exe?SUBED1=extramuralnexus&A=1>