

1 **Utilizing artificial intelligence to manage COVID-19 scientific evidence torrent**  
2 **with Risklick AI: a critical tool for pharmacology and therapies development.**

3

4 **Short title: Risklick AI-based management of COVID-19 scientific evidences**

5

6 **Preliminary author list:**

7 Quentin Haas<sup>1,3</sup>, David Vicente Alvarez<sup>2</sup>, Nikolay Borissov<sup>1,3</sup>, Sohrab Ferdowski<sup>2</sup>,  
8 Leonhard von Meyenn<sup>1,3</sup>, Sven Trelle<sup>3</sup>, Douglas Teodoro<sup>2</sup> and Poorya Amini<sup>1,3</sup>

9

10 <sup>1</sup> Risklick AG, Spin-off University of Bern, Bern, Switzerland.

11 <sup>2</sup> HES-SO University of Applied Sciences and Arts Western Switzerland, Geneva,  
12 Switzerland.

13 <sup>3</sup> Clinical Trial Unit Bern, University of Bern, Bern, Switzerland.

14

15 **Corresponding author:**

16 Poorya Amini

17 Risklick AG, Spin-off

18 University of Bern

19 Spitalackerstrasse 26, 3013 Bern, Switzerland

20 Tel: +41 78 781 41 35

21 Email: poorya.amini@risklick.ch

22 Website: www.risklick.ch

23

24 **Authorship note:** Quentin Haas and David Vicente Alvarez contributed equally to this  
25 work.

26

27 **Number of Tables:** 1

28 **Number of Figures:** 5

29 **Word count:** 2760

30 **Keywords:** Artificial intelligence, COVID-19, search platform, Risklick.

31

32 **ABSTRACT**

33

34 **Introduction:** The SARS-CoV-2 pandemic has led to one of the most critical and  
35 boundless waves of publications in the history of modern science. The necessity to  
36 find and pursue relevant information and quantify its quality is broadly acknowledged.  
37 Modern information retrieval techniques combined with artificial intelligence (AI)  
38 appears as one of the key strategies for COVID-19 living evidence management.  
39 Nevertheless, most AI projects that retrieve COVID-19 literature still require manual  
40 tasks. **Methods:** In this context, we present a novel, automated search platform, called  
41 Risklick AI, which aims to automatically gather COVID-19 scientific evidence and  
42 enable scientists, policy makers and healthcare professionals to find the most relevant  
43 information tailored to their question of interest in real time. **Results:** Here, we compare  
44 the capacity of Risklick AI to find COVID-19-related clinical trials and scientific  
45 publications in comparison to clinicaltrials.gov and Pubmed in the field of  
46 pharmacology and clinical intervention. **Discussion:** The results demonstrate that  
47 Risklick AI is able to find COVID-19 evidences more effectively, both in terms of  
48 precision and recall, compared to the baseline platforms. Hence, Risklick AI could  
49 become a useful alternative assistant to scientists fighting the COVID-19 pandemic.

50

## 51 INTRODUCTION

52 The SARS-CoV-2 pandemic resulted in one of the largest waves of publications and  
53 clinical trials in the history of modern science, with the number of articles doubling  
54 every 20 days and unprecedented clinical trial rate [1–3] . In this context, it has become  
55 virtually impossible for scientists, policy makers and healthcare workers to keep up  
56 with the speed at which data are generated. Moreover, this situation limited the  
57 possibilities offered to professionals involved in the pandemic to read entire articles  
58 thoroughly, as well as to properly evaluate the limitations of the data. In addition, this  
59 outburst of publications also impacted the average quality of research papers [4,5].

60 The necessity to effectively gather scientific evidence that encompasses only relevant  
61 information with acceptable quality has been one of the most important modern  
62 challenges in science. This issue has become strikingly evident throughout the COVID-  
63 19 crisis. In this context, artificial intelligence (AI) appears to be the best strategy to  
64 seek the most relevant scientific evidence in a minimum amount of time [6,7]. AI-based  
65 strategies are now required to diminish time of research, increase performance, and  
66 reduce errors and oversights in the research of references performed by scientists and  
67 health professionals. The proliferation of AI-based initiatives to address the COVID-19  
68 pandemic has resulted in the creation of numerous technologies, such as LitCovid and  
69 the COVID-NMA Project, among others [8,9]. However, most of the developed COVID-  
70 19 tools, such as the cited examples, require manual steps in the analytic process.  
71 Hence, a fully automated, AI-based efficient tool is still missing in the context of the  
72 COVID-19 pandemic in order to optimize the access and the management of specific  
73 knowledge and research results.

74 In this context, we developed a novel, automated scientific evidence management  
75 platform called Risklick AI. The tool aims to gather and manage COVID-19-related

76 literature using natural language processing (NLP), a technology allowing computers  
77 to process and analyze large amounts of data expressed in natural language [10–12].  
78 The tool combines classic statistical word frequency methods, so called bag-of-words,  
79 with state-of-the-art masked language models [13–15]. Hence, using artificial  
80 intelligence, Risklick AI allows computers to analyze human language with more  
81 meaning than with the usual processed and programmed responses. In this study, we  
82 compare the capacities of Risklick AI to find COVID-19-related clinical trials compared  
83 to clinicaltrials.gov [16] and scientific publications in comparison with Pubmed. We  
84 compared query outcomes of Risklick AI to clinicaltrials.gov and Pubmed on COVID-  
85 19 pharmacologically relevant treatments, as considered by the authorities [17]. Here,  
86 we demonstrate that Risklick AI represents the more effective technology with the  
87 potentiality to assist scientists in finding and pursuing relevant COVID-19-related  
88 scientific evidences.

89

## 90 **METHODS**

### 91 Data collection

92 On a daily basis, Risklick AI collects and updates clinical trials data on from wide  
93 sources such as clinical trials registries and datasets from World Health Organization  
94 (WHO) [18]. Moreover, publications' metadata like titles, abstracts, journal names,  
95 publication date, digital object identifier number and others are collected and updated  
96 from sources like PubMed, Embase, BioRxiv and MedRxiv from "Living Evidence on  
97 COVID-19," and "CORD-19" datasets (19) .

### 98 Technology

99 All the data for clinical trials and publications is preprocessed to align to a predefined  
100 data format and added to Elasticsearch, which serves as a full text search and analytics  
101 engine for clinical trials and publications. The indexed data and queries are normalized  
102 using a pipeline of text preprocessing techniques like tokenization, lowercasing, stop  
103 words removal, and reducing words to their root form. The indices are maintained in a  
104 Elasticsearch cluster. The index model parameters are tuned using a set of manually  
105 annotated queries. The similarity measure was computed using the divergence from  
106 randomness model (DFR) with the term frequency normalization set to 20.0 [19]. A  
107 detailed description of the pipeline is provided by [Ferdowsi et al](#) [15].

108

109 To increase the recall of relevant documents to user query, we apply query expansion  
110 techniques using a COVID-specific ontology of standardized medical terms, their  
111 synonyms, classes, and sub-classes engineered by clinical trial domain experts [20].  
112 For instance, once the user search for heparin, the query automatically expands to all  
113 three major of heparin: unfractionated heparin (UFH), low molecular weight heparin  
114 (LMWH), ultra-low-molecular weight heparin (ULMWH), and their trade names based  
115 on COVID-specific ontology (e.g. Nadroparin, Fraxiparin, Fraxodi, Calciparine,

116 Bemiparin, Zibor, Ivor, Enoxaparin, Clexane, Iovenox, Fragmine, Dalteparin,  
117 Dociparstat).

118

## 119 Experimental setup

120 At the time of analysis, more than 1800 interventional studies linked to COVID-19 were  
121 available on clinicaltrials.gov. In addition, more than 48'500 COVID-19-related  
122 publications were available on Pubmed. In order to compare Risklick AI's performance  
123 with other COVID-related search platforms, we defined and used a common set of  
124 search queries, which were executed on a specific day for all platforms. To assess our  
125 clinical trial search engine, we compared Risklick AI with the most advanced and  
126 biggest clinical trial registry— clinicaltrials.gov. In addition to COVID-19 cases,  
127 clinicaltrials.gov covers all of COVID-19 clinical trials from other registries like  
128 clinicaltrialsregister.eu and chichtr.org, as specified by clinicaltrials.gov  
129 ([https://clinicaltrials.gov/ct2/who\\_table](https://clinicaltrials.gov/ct2/who_table)). Hence, clinicaltrials.gov appears as the  
130 adapted gold-standard to allow comparison with Risklick AI.

131 The day the queries are run, the platform retrieves the latest dataset from  
132 clinicaltrials.gov and it is indexed in the Risklick AI platform. The comparison comprises  
133 of only interventional clinical trials having unique clinicaltrials.gov identifier (NCT-  
134 number). To compare the clinical trials found by the different types of queries, data  
135 from Risklick AI, clinicaltrials.gov, and corona-trials.org are collected for categories like  
136 antibiotic, anticoagulant, and antiviral, as well as more fine granular queries for specific  
137 drugs like Remdesivir, Tocilizumab, Azithromycin, Hydroxychloroquine, and Heparin  
138 (suppl. Table 1).

139 Risklick AI and PubMed were then compared regarding their publications search  
140 performance. Before running the query, the latest COVID-19-related publications are  
141 retrieved from PubMed using the predefined queries in the Institute of Social and

142 Preventive Medicine (ISPM) Bern and added to a new index in Risklick AI [21]. This  
143 way we ensure that the queries executed on a specific day on PubMed and Risklick AI  
144 retrieve publications based on the same data distribution for the specific day on both  
145 platforms. To compare the scientific publications found by the different queries, data  
146 from Risklick AI and Pubmed were collected for antithrombotic, dexamethasone and  
147 Favipiravir (suppl. Table 2).

148 All the drug categories used in this study (antibiotic, antithrombotic, antiviral, and  
149 anticoagulant) are resumed in suppl. Table 3.

150

#### 151 Validation

152 Verification and validation procedures were performed by two separate and  
153 independent immunologists. All clinical trials and scientific publications were analyzed  
154 and verified manually. To optimize the result comparison between the different search  
155 tools, *recall* (the number of positive class predictions made out of all positive examples  
156 in the dataset), *precision* (the number of positive class predictions that actually belong  
157 to the positive class), and *F1-score* (single score that balances both the concerns of  
158 precision and recall in one number) were calculated [22].

159

#### 160 Data analysis

161 Retrieved publications were individually and manually scored as true-positive or false-  
162 positive. Graphs were created using Prism 8.0.

163

## 164 RESULTS

### 165 Comparison search performance for clinical trials

166 The capacity of Risklick AI to retrieve COVID-19-related clinical trials was analyzed.  
167 When compared to clinicaltrials.gov and covid-trials.org, regarding its capacity to find  
168 COVID-19-related clinical trials, Risklick AI found more raw clinical trials than other  
169 tools for different categories of treatments such as antibiotic anticoagulant and antiviral  
170 (Figure 1A). In average, Risklick found 1.9-times more clinical trials than  
171 clinicaltrials.gov for these 3 treatments, and 8.2-times more than covid-trials.org for  
172 these same 3 treatments. When investigating key molecules of each category, such  
173 as Hydroxychloroquine, Remdesivir, Azithromycin, Tocilizumab or Heparin (Figure  
174 1B), Risklick AI presented more raw output compared to the 2 other research tools. No  
175 clinical trial connected to COVID-19 was found (n.d.) on covid-trials.org for the Heparin  
176 query.

177 In order to compare the search capacity of Risklick AI in comparison with  
178 clinicaltrials.gov, COVID-19-related search was restricted to the same database, using  
179 only clinical trials registered on clinicaltrials.gov. This strategy was applied for both  
180 drug classes and specific drugs. Using Hydroxychloroquine data for illustration, we first  
181 segregated publications found only by Risklick AI or clinicaltrials.gov (“Unique”) from  
182 publications found by both tools (“Common”) (Figure 1C). Then, unique publications  
183 were analyzed and separated between true-positive (“True”) and false-positive  
184 (“False”) results (figure 1D). Ultimately, we calculated the total number of true positives  
185 of the publications by adding the categories common and unique along with true-  
186 positive (Figure 1E).

187 We further analyzed accuracy of both tools for drug classes. There, Risklick AI showed  
188 a higher number of relevant clinical trials for antibiotic (8.9%) (Figure 2A-C),



189 anticoagulant (29.4%) (Figure 2D-F), and antiviral drugs (47.2%) (Figure 2G-I)  
190 associated with COVID-19 in comparison to clinicaltrials.gov on the same reference  
191 database. Recall, precision and F1 score measures for the 3 drugs categories were  
192 systematically higher for Risklick AI compared to clinicaltrials.gov (suppl Table 3). The  
193 detailed analysis reveals that the higher score of Risklick AI is due to a higher number  
194 of true-positives (that is, higher recall), and a lower number of false-positives (that is,  
195 higher precision) in the unique findings cohort relative to clinicaltrials.gov (Figure  
196 2B,E,H). The analysis was then extended to specific drugs. There, Hydroxychloroquine  
197 (Figure 1 C-E), Tocilizumab (Figure 3A-C), and Heparin (Figure 3D-F) all presented a  
198 higher number of relevant clinical trials associated with COVID-19 compared to  
199 clinicaltrials.gov. Again, the higher score of Risklick AI is due to a higher number of  
200 true-positives, and a lower amount of false-positives unique findings in comparison to  
201 clinicaltrials.gov for these three drugs (Figure 1D and Figure 2 B,E). Regarding  
202 Azithromycin, the same number of relevant clinical trials was found in both search tools  
203 (Figure 3G-I). However, in opposition to clinicaltrials.gov, Risklick AI uncovered no  
204 false-positive outcomes (Figure 3H). Ultimately, no difference was observed between  
205 Risklick AI and clinicaltrials.gov regarding Remdesivir (Suppl. Fig.1). When taken  
206 together, Risklick AI presented an average recall of 99.25% compared to 86.61% for  
207 clinicaltrials.gov. By extension, Risklick AI also presented a F1 score of 97.59%, while  
208 clinicaltrials.gov had 88.57% (Table 1).

209

210 Risklick AI search performance regarding COVID-19-related publications

211

212 The data retrieval was extended to COVID-19-related scientific publications by  
213 comparing Risklick to Pubmed search capacities. We restricted the search to the  
214 Pubmed database using Boolean search tool. There, we investigated the number of

215 relevant publications restricted to COVID-19 for antithrombotic (+61.4%) (Figure 4A-  
216 C), Dexamethasone (+114.3%) (Figure 4 D-F) and Favipiravir (+38.3%) (Figure 4 G-  
217 I). As for the comparison with clinicaltrials.gov, the superiority of Risklick AI compared  
218 to Pubmed is due to a more important number of true-positives, and a lower amount of  
219 false-positive unique findings (Figure 4 B,E,H). Taken together, the Risklick search  
220 presented an average recall of 86.66% compared to 61.26% for Pubmed. In addition,  
221 the average F1 score for Risklick reached 90.28% compared to 71.68% for Pubmed  
222 (Table 1).

223

#### 224 Evaluation of Risklick AI's publication search tool

225

226 Risklick AI offers the possibility to find COVID-19-related publications using Boolean-  
227 based search or NLP-based search methods and further combining the results of both  
228 methods. Here, we compare the capacity of each technology to find COVID-19-related  
229 publications. Hence, Boolean-based search ("Risklick bool"), NLP-based search  
230 ("Risklick NLP"), and NLP-based search supplemented with pre-print publications  
231 ("Risklick NLP+PP") were compared for antithrombotic (Figure 5 A-C),  
232 Dexamethasone (Figure 5 D-F), and Favipiravir (Figure 5 G-I). The searches in Risklick  
233 AI and clinicaltrials.gov are run based on same dataset for the specific day and based  
234 on same queries. Overall, Risklick AI NLP and Risklick NLP+PP offer more  
235 publications than Boolean-based search (+23.7% and +118.3%, respectively),  
236 although each search strategy presents various rates of false-positive outcomes  
237 (Figure 5 B,E,H). Used synchronously, both search methods offer a more complete,  
238 pertinent overview of currently available literature on the given treatments linked to  
239 COVID-19. Regarding clinical trials, clinicaltrials.gov uses Medical Subject Headings  
240 (MeSH) terms for query expansion, but does not match misspelled or differently spelled

241 words for a disease or intervention. Risklick AI combines query expansion technology  
242 based on ontology defined by experts together with NLP techniques. The NLP  
243 techniques allow us to better deal with misspelled and similarly spelled words, which  
244 improved the quality of the search.

## 245 **DISCUSSION**

246 The COVID-19 outbreak has resulted in one of the biggest waves of publications in the  
247 history of modern science [2,23]. In these conditions, it has become clear that COVID-  
248 19 data retrieval and monitoring would be one of the main challenges of the current  
249 and future pandemics [24]. To address this dilemma, we automatically gathered and  
250 centralized all COVID-19 scientific information from scattered sources on a daily basis.  
251 Several intelligent algorithms and models were then developed to retrieve query  
252 relevant scientific evidences from a centralized database. Both Boolean and NLP-  
253 based search methods have been used to find query relevant scientific evidences.

254 In this study, the search performance of our methodology was compared to  
255 clinicaltrials.gov when screening the same database of clinical trials. Several  
256 molecules were selected to this purpose based on their connection to COVID-19 trials  
257 currently performed worldwide, as well as their important number of citations in the  
258 scientific literature. Overall, the abilities of the Risklick AI method to find relevant clinical  
259 trials against specific intervention queries were higher than the reference search tools,  
260 both for drug classes as for single treatments. Interestingly, the Risklick AI performance  
261 was largely due to a higher true-positive and lower false-positive outcome in  
262 comparison to clinicaltrials.gov. We believe this is due to the power of the full text  
263 search engine combined with the Boolean model plus the improved semantics brought  
264 by the COVID ontology.

265 When extended to COVID-19-related publications, Risklick AI also confirmed a  
266 superior search capability compared to the medical reference tool Pubmed, using the  
267 Boolean search engine. By extension, we compared the capacities of Risklick AI to  
268 find the scientific COVID-19 literature for pharmacological keywords using the Boolean  
269 and NLP approach. Molecules and categories selected for this analysis were chosen

270 based on their relevance to COVID19. These molecules were not engaged into  
271 numerous clinical trials as for molecules chosen in the clinicaltrials.gov comparison.  
272 There, we observed that both strategies offered a broad overview of key search articles  
273 with a high proportion of unique outcomes. In addition, we also confirmed the capacity  
274 of Risklick AI to find preprint (PP) literature database with a high true-positive outcome,  
275 allowing for broad search perspectives in a context of permanent novelty not covered  
276 by Pubmed.

277 On the one hand, Boolean search is still used in recent platforms like PubMed,  
278 Embase, and others. On the other hand, recent advancements in NLP and full-text  
279 searches enable better gathering of queries, sentences, and documents. These  
280 developments reduce the need for preprocessing and normalization steps and they  
281 improve the quality of context-based searches.

282 Our methodology offers two search interfaces to find documents on the same datasets:  
283 one for Boolean search and one for NLP context-based search. This way users can  
284 arbitrarily combine the results of both approaches and thus improve precision and  
285 recall of their results. By extension, the evaluation results demonstrate the potential of  
286 the proposed method to help scientists and decision makers to triage key information  
287 out of the torrent of scientific papers from the COVID-19 pandemic. Consequently,  
288 Risklick AI could play a key role in the development of novel drugs and strategies  
289 targeting COVID-19, and could therefore become an important ally in fields such as  
290 pharmacology and epidemiology to organize the medical response against the SARS-  
291 CoV-2 virus. Moreover, in perspective of the current situation, Risklick AI could play an  
292 primordial role in the monitoring of all COVID-19 vaccines effectiveness, particularly in  
293 perspective of the numerous variants and associated serotypes of SARS-CoV-2. By  
294 extension, Risklick AI could offer significant advantages in the data management of

295 other diseases and pathologies for clinicians and fundamental researchers. Since the  
296 underlying technology is generic, the framework can be used in other diseases and  
297 areas to manage relevant scientific evidences.

298

299 **Statement of ethic:** This study did not involve human or animal material or data. Ethics  
300 approval was not required.

301 **Conflict of interests:** The authors QH, NB, LvM, and PA are working for Risklick.

302 **Funding source:** This work was supported by Innosuisse project funding number  
303 41013.1 IP-ICP.

304 **Author contributions:** QH, PA and NB designed the study. QA, PA and NB wrote the  
305 manuscript. DVA, SF and DT designed and implemented the clinical trial and  
306 publication retrieval technologies. QH and NB performed the experimental work. All  
307 authors had full access to the data, helped draft the report or critically revised the draft,  
308 contributed to data interpretation, reviewed and approved the final version of the report.

309    **REFERENCES**

- 310    1     Thorlund K, Dron L, Park J, Hsu G, Forrest JI, Mills EJ. A real-time dashboard  
311           of clinical trials for COVID-19. *Lancet Digit Heal*. 2020 Jun;2(6):e286–7.
- 312    2     Palayew A, Norgaard O, Safreed-Harmon K, Andersen TH, Rasmussen LN,  
313           Lazarus J V. Pandemic publishing poses a new COVID-19 challenge. *Nat Hum*  
314           *Behav*. 2020 Jul;4(7):666–9.
- 315    3     Brainard J. Scientists are drowning in COVID-19 papers. Can new tools keep  
316           them afloat? *Science*. 2020 May
- 317    4     Teixeira da Silva JA, Tsigaris P, Erfanmanesh M. Publishing volumes in major  
318           databases related to Covid-19. *Scientometrics*. 2020 Aug;1–12.
- 319    5     Besançon Id L, Peiffer-Smadja Id N, Segalas Id C, Jiang H, Id PM, Smout Id C,  
320           et al. Open Science Saves Lives: Lessons from the COVID-19 Pandemic.  
321           *bioRxiv*. 2020 Aug;2020.08.13.249847.
- 322    6     Hutson M. Artificial-intelligence tools aim to tame the coronavirus literature.  
323           *Nature*. 2020 Jun
- 324    7     Extance A. How AI technology can tame the scientific literature. *Nature*. 2018  
325           Sep;561(7722):273–4.
- 326    8     Chen Q, Allot A, Lu Z. Keep up with the latest coronavirus research. *Nature*.  
327           2020 Mar;579(7798):193.
- 328    9     Boutron I, Chaimani A, Meerpohl JJ, Hróbjartsson A, Devane D, Rada G, et al.  
329           The COVID-NMA Project: Building an Evidence Ecosystem for the COVID-19  
330           Pandemic. *Ann Intern Med*. 2020 Sep
- 331    10    Jurafsky Daniel, Martin James. *Speech and Language Processing*, 2nd Edition



332 . 2008.

333 11 Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al.  
334 HuggingFace's Transformers: State-of-the-art Natural Language Processing.  
335 arXiv. 2019 Oct.

336 12 Teodoro D, Mottin L, Gobeill J, Gaudinat A, Vachon T, Ruch P. Improving  
337 average ranking precision in user searches for biomedical research datasets.  
338 Database (Oxford). 2017 Jan;2017:83.

339 13 Harris ZS. Distributional Structure. WORD. 1954 Aug;10(2-3):146-62.

340 14 Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep  
341 Bidirectional Transformers for Language Understanding. arXiv. 2018 Oct  
342 ;1:4171-86.

343 15 Ferdowsi S, Borissov N, Kashani E, Alvarez DV, Copara J, Gouareb R, et al.  
344 Information retrieval in an infodemic: the case of COVID-19 publications.  
345 bioRxiv. 2021 Jan;2021.01.29.428847.

346 16 Tse T, Williams RJ, Zarin DA. Update on registration of clinical trials in  
347 ClinicalTrials.gov. Chest. 2009 Jul;136(1):304-5.

348 17 Ahuja AS, Reddy VP, Marques O. Artificial intelligence and COVID-19: A  
349 multidisciplinary approach. Integr Med Res. 2020 Sep;9(3):100434.

350 18 WHO Clinical trials primary registries platform. [https://www.who.int/clinical-](https://www.who.int/clinical-trials-registry-platform/network/primary-registries)  
351 [trials-registry-platform/network/primary-registries](https://www.who.int/clinical-trials-registry-platform/network/primary-registries).

352 19 Amati G, Van Rijsbergen CJ. Probabilistic models of information retrieval based  
353 on measuring the divergence from randomness. ACM Trans Inf Syst. 2002  
354 Oct;20(4):357-89.

- 355 20 Diao L, Yan H, Li F, Song S, Lei G, Wang F. The research of query expansion  
356 based on medical terms reweighting in medical information retrieval. *Eurasip J*  
357 *Wirel Commun Netw.* 2018 Dec;2018(1):105.
- 358 21 COVID-19 projects of ISPM. Available from: [https://ispmbern.github.io/covid-](https://ispmbern.github.io/covid-19/living-review/collectingdata.html)  
359 [19/living-review/collectingdata.html](https://ispmbern.github.io/covid-19/living-review/collectingdata.html)
- 360 22 Hamel C, Kelly SE, Thavorn K, Rice DB, Wells GA, Hutton B. An evaluation of  
361 DistillerSR's machine learning-based prioritization tool for title/abstract  
362 screening – impact on reviewer-relevant outcomes. *BMC Med Res Methodol.*  
363 2020 Dec;20(1):256.
- 364 23 Callaway E. Will the pandemic permanently alter scientific publishing? *Nature.*  
365 2020 Jun;582(7811):167–8.
- 366 24 Muhammad LJ, Islam MM, Usman SS, Ayon SI. Predictive Data Mining Models  
367 for Novel Coronavirus (COVID-19) Infected Patients' Recovery. *SN Comput*  
368 *Sci.* 2020 Jul;1(4):206.

369

370

371 **FIGURE LEGENDS**

372 **Figure 1. Risklick AI clinical trials outcome for COVID-19 compared to other web-**  
373 **based resource registries.** (A, B) Total raw number of clinical trials found by Risklick  
374 AI compared to other registries for drug classes (A) and specific treatments (B) used  
375 against COVID-19. (C-E) Analysis of search capacity of registered clinical trials by  
376 Risklick AI and clinicaltrials.gov based on the same dataset for hydroxychloroquine.  
377 Clinical trials were separated between common and unique outcomes (C). Unique  
378 outcomes were validated and separated between true-positive (True) and false-  
379 positive (False) results (D). Final total number of true positive clinical trials is comprised  
380 of the addition of common findings and unique, true-positive findings (E). n.d, no data.

381

382 **Figure 2. Risklick AI clinical trials search capacity for drug classes connected to**  
383 **COVID-19 compared to clinicaltrials.gov based on the same dataset. (A-C)**  
384 Analysis of search capacity of registered clinical trials by Risklick AI and  
385 clinicaltrials.gov on the same database for antibiotic drugs. Clinical trials were  
386 separated between common and unique outcomes (A). Unique outcomes were  
387 validated and separated between true-positives (true) and false-positives (False)  
388 results (B). The final total number of true positive clinical trials is the addition of  
389 common findings and unique, true-positive findings (C). The same procedure was  
390 performed for anticoagulant (D-F) and antiviral (G-I) drugs.

391

392 **Figure 3. Risklick AI clinical trials search capacity for specific treatments**  
393 **associated with COVID-19 in comparison with clinicaltrials.gov on the same**  
394 **dataset.** (A-C) Analysis of search capacity of registered clinical trials by Risklick AI and  
395 clinicaltrials.gov on the same dataset for Tocilizumab. Clinical trials were separated  
396 between common and unique outcomes (A). Unique outcomes were validated and  
397 separated between true-positive (true) and false-positive (wrong) results (B). Total final  
398 number of true positive clinical trials is the addition of common findings and unique,  
399 true-positive findings (C). The same procedure was performed for Heparin (D-F) and  
400 Azithromycin (G-I). n.s, no data.

401

402 **Figure 4. Risklick AI publication search capacity for specific treatments**  
403 **associated with COVID-19 compared to Pubmed on the same publication**  
404 **dataset.**(A-C) Analysis of search capacity of COVID-19-related publications by  
405 Risklick AI and Pubmed on the same publication dataset for antithrombotic.  
406 Publications were separated between common and unique outcomes (A). Unique  
407 outcomes were validated and separated between true-positive (true) and false-positive  
408 (wrong) results (B). Total final number of true positive publications is the addition of  
409 common findings and unique, true-positive findings (C). Same procedure was  
410 performed for Dexamethasone (D-F) and Favipiravir (G-I).

411

412 **Table 1. Risklick AI, clinicaltrials.gov and Pubmed average recall, precision and**  
413 **F1 score for all the different molecules and treatments groups searched.**

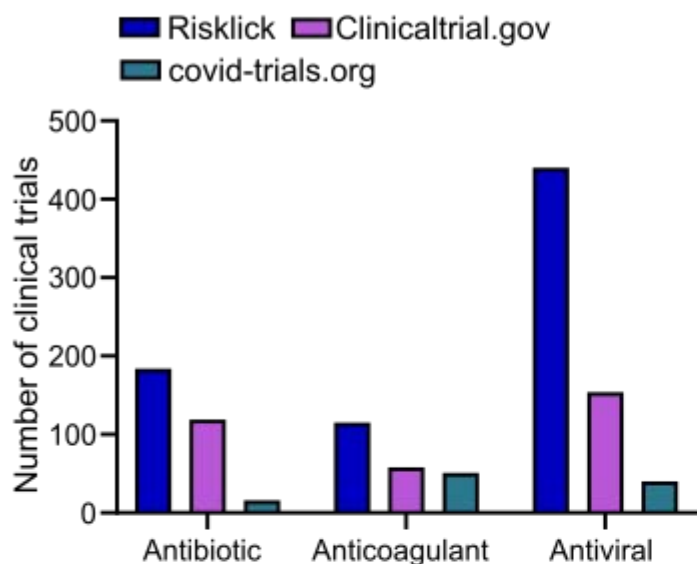
414

415 **Figure 5. Risklick AI publication search capacity for specific treatments**  
416 **associated with COVID-19 using Boolean or Natural-language processing (NLP)**  
417 **search methods.** (A-C) Analysis of search capacity of COVID-19-related publications  
418 by Risklick AI using Boolean search tool (bool), Natural-language processing (NLP)  
419 research tool, and NLP with the database extended to pre-print (PP) publications for  
420 antithrombotic drugs. Publications were separated between common and unique  
421 outcomes (A). Unique outcomes were validated and separated between true-positive  
422 (true) and false-positive (wrong) results (B). Total final number of true positive  
423 publications is the addition of common findings and unique, true-positive findings (C).  
424 The same procedure was performed for Dexamethasone (D-F) and Favipiravir (G-I).  
425

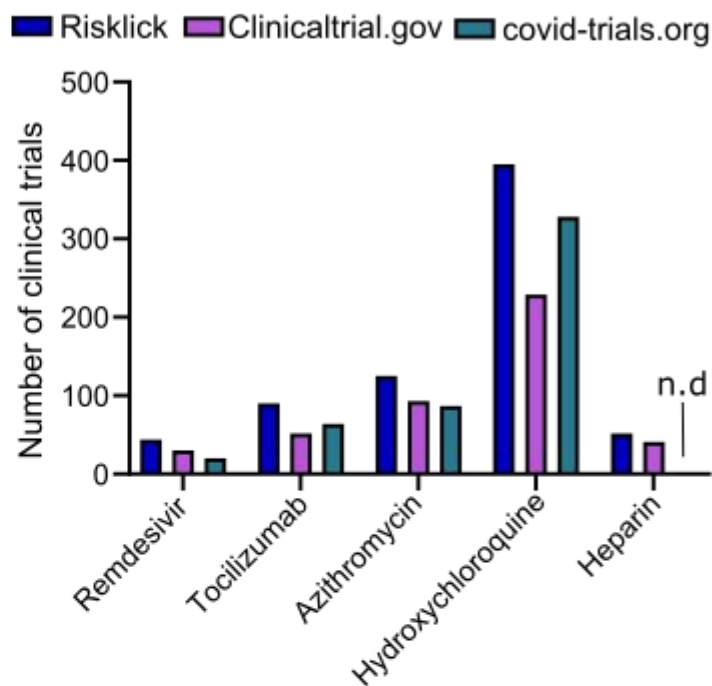


# Figure 1

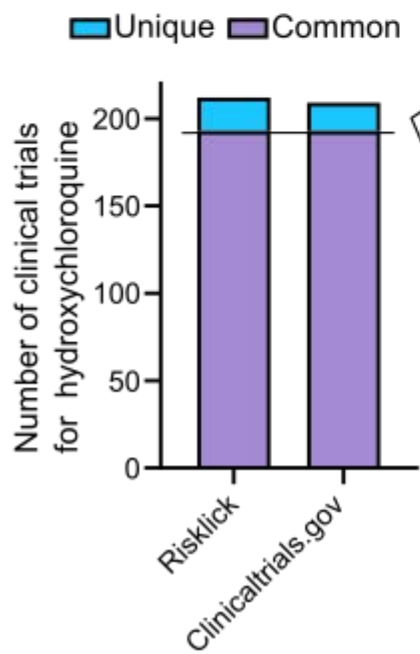
## A



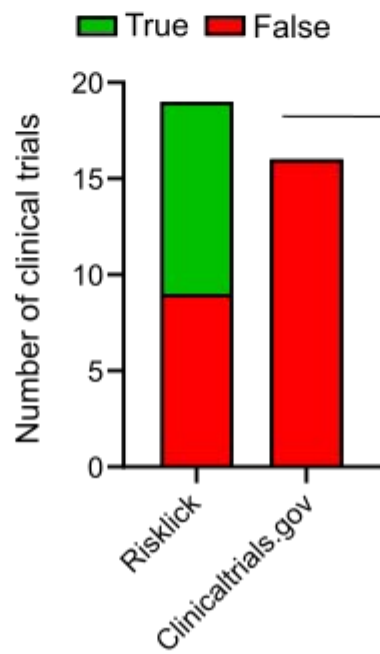
## B



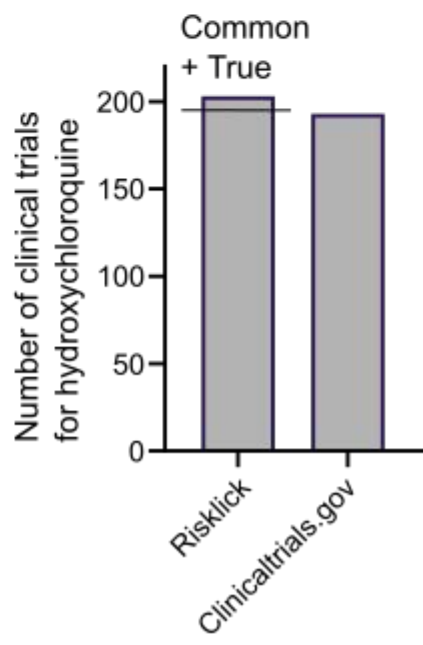
## C

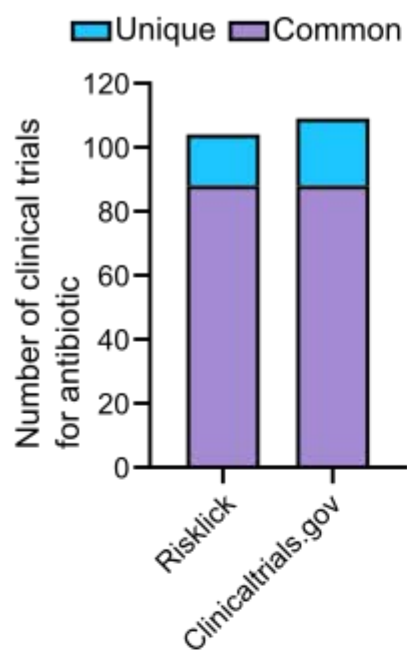
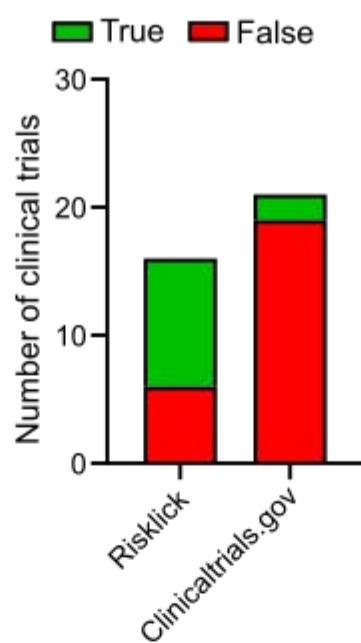
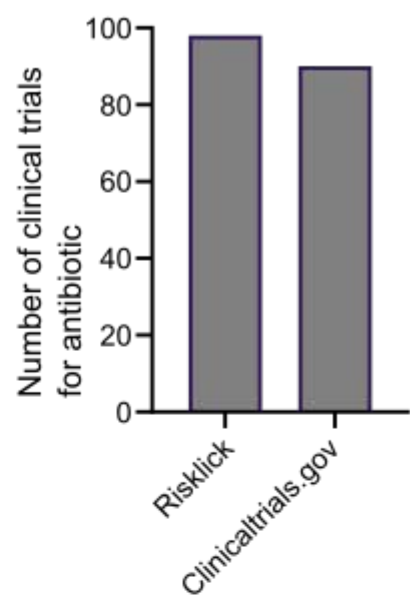
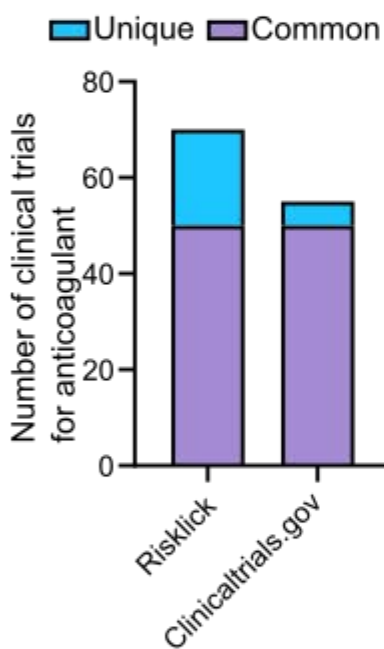
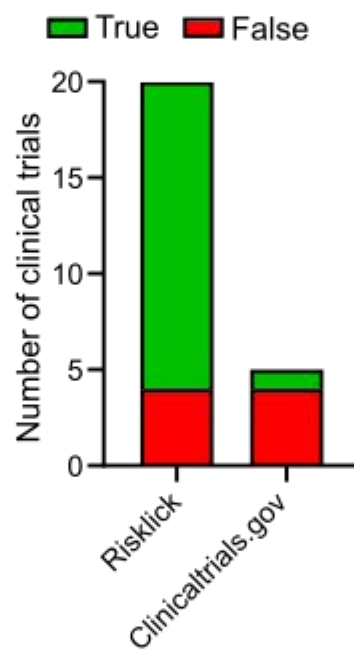
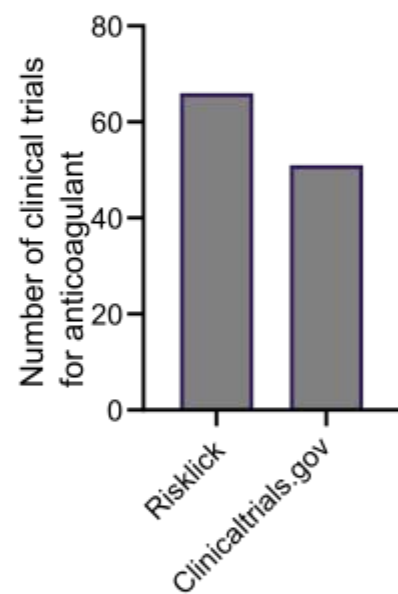
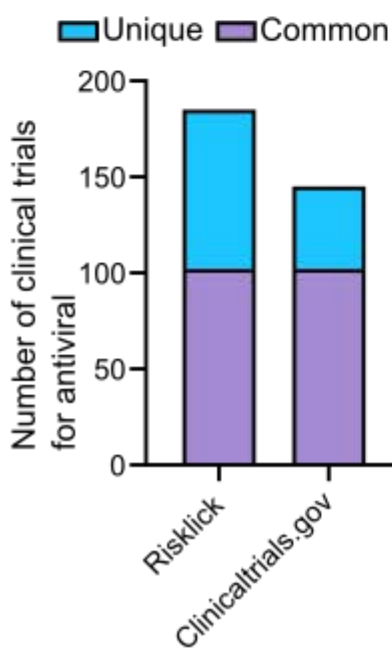
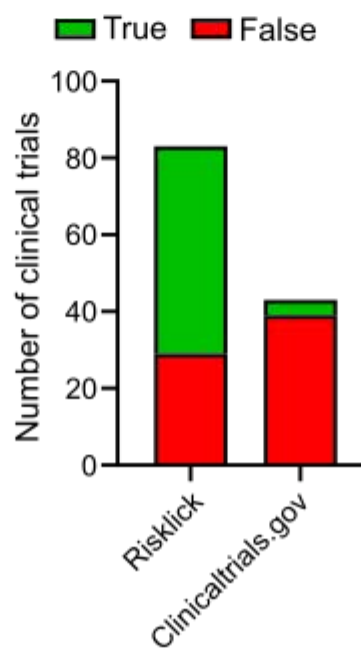
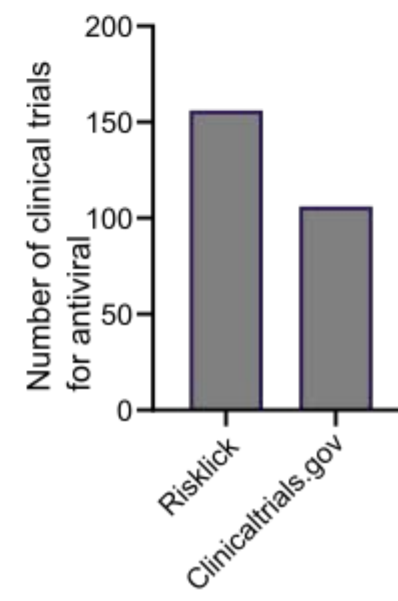


## D

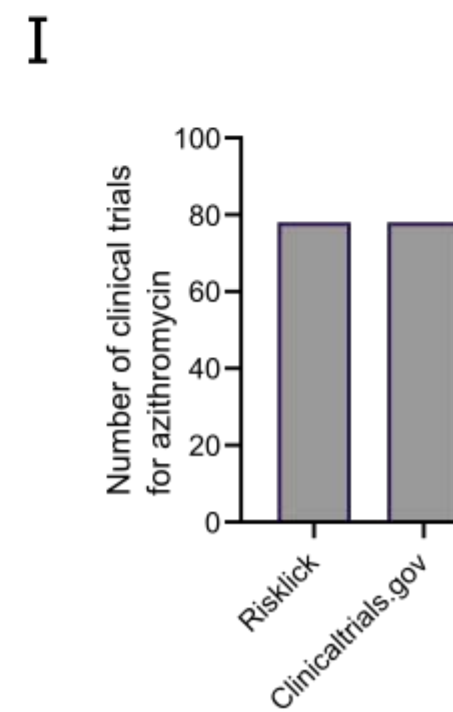
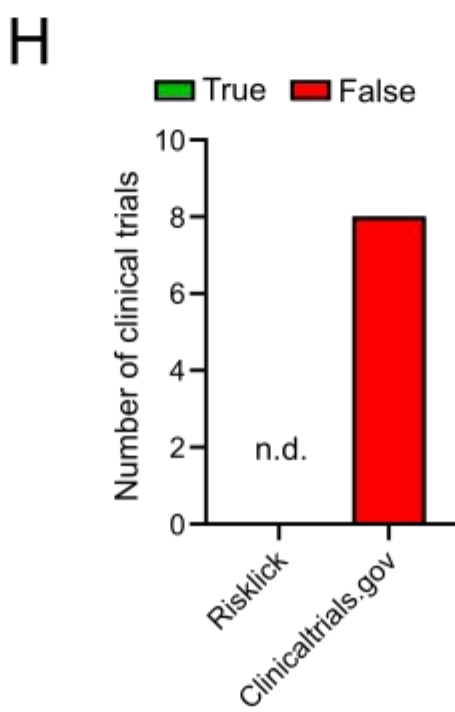
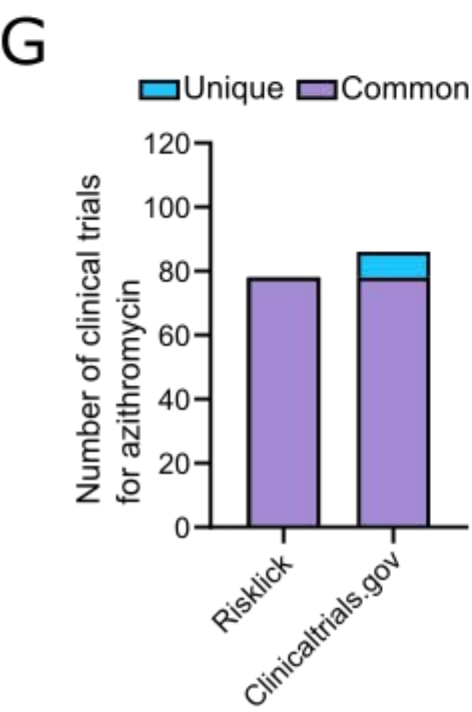
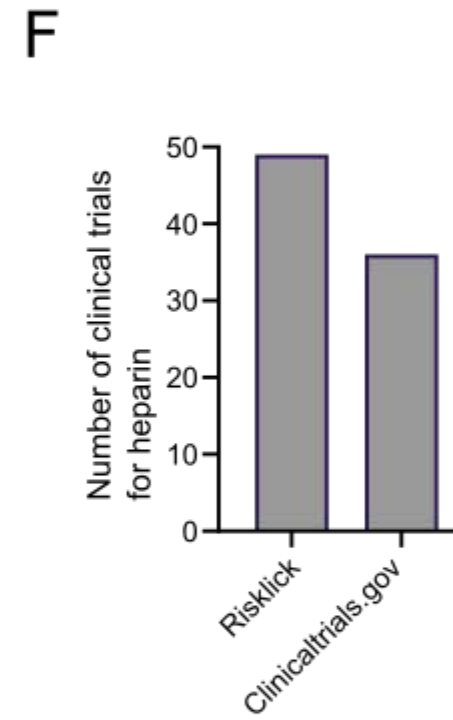
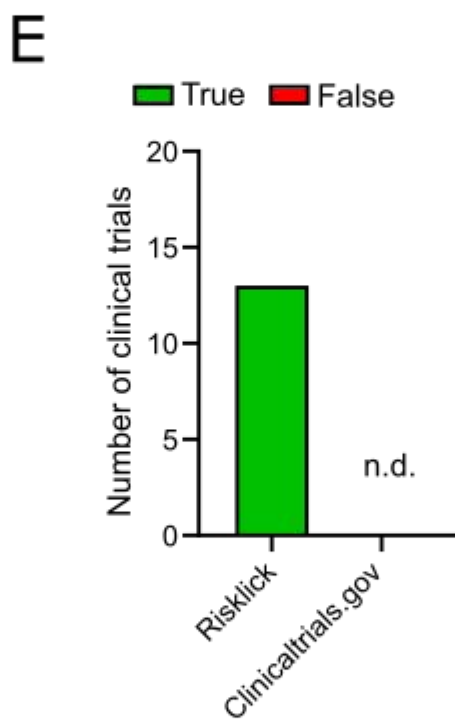
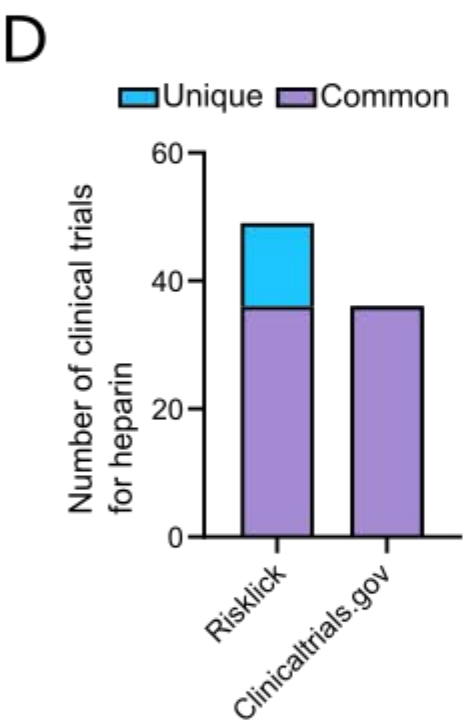
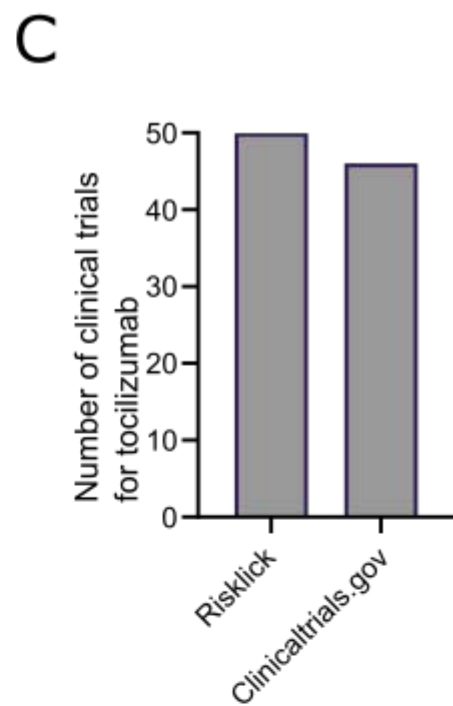
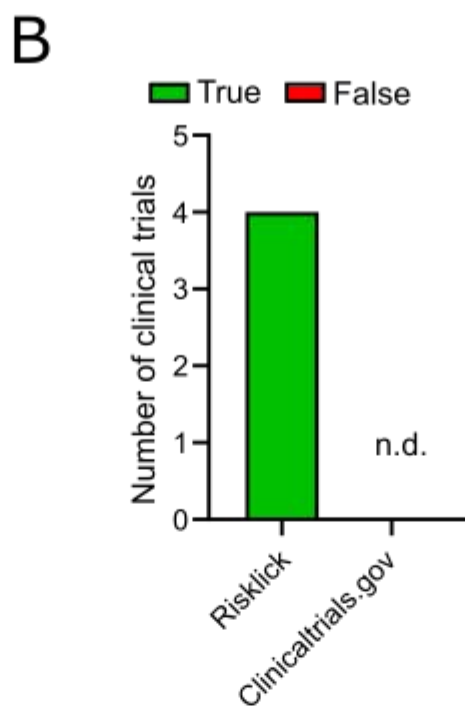
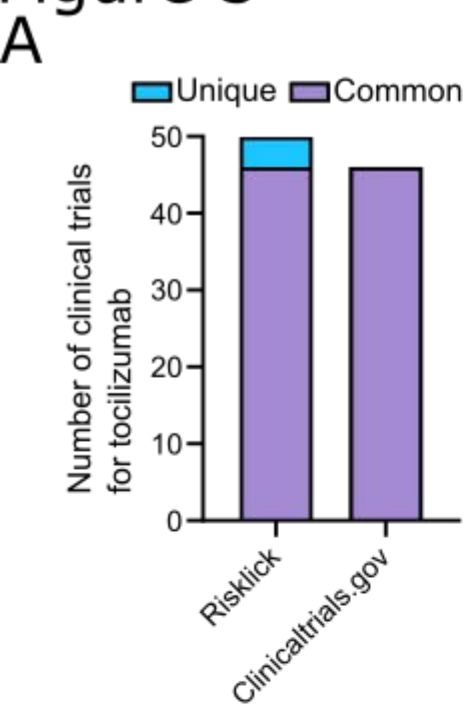


## E



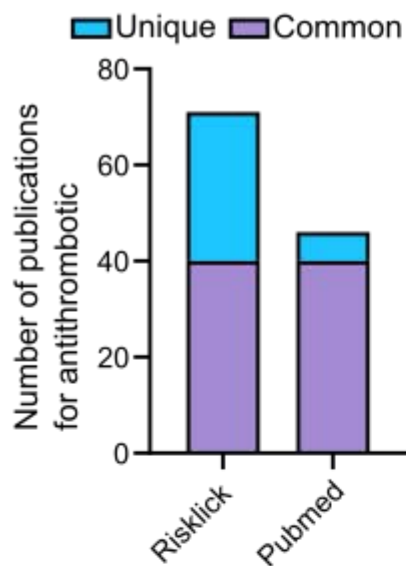
**Figure 2****A****B****C****D****E****F****G****H****I**

# Figure 3

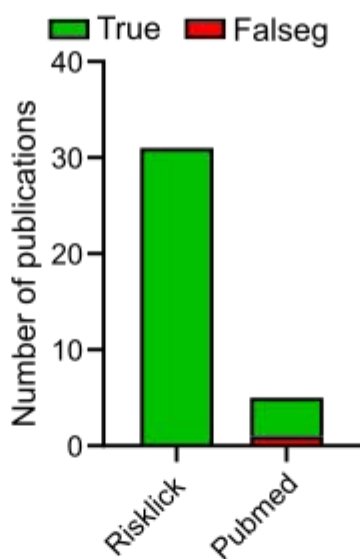


# Figure 4

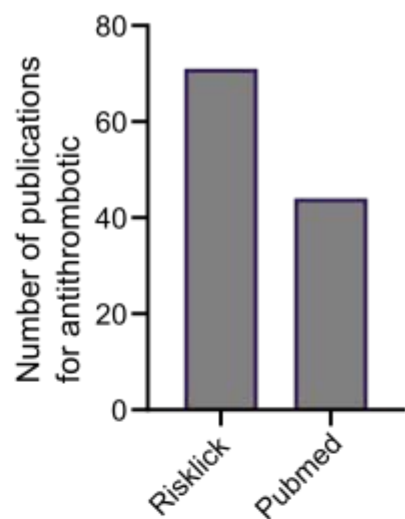
## A



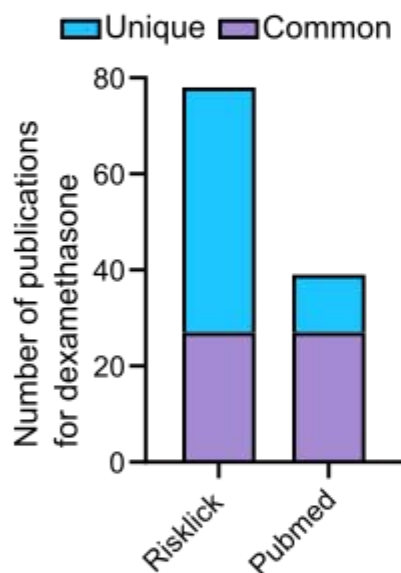
## B



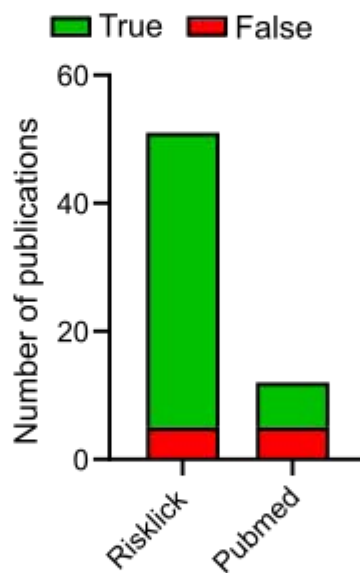
## C



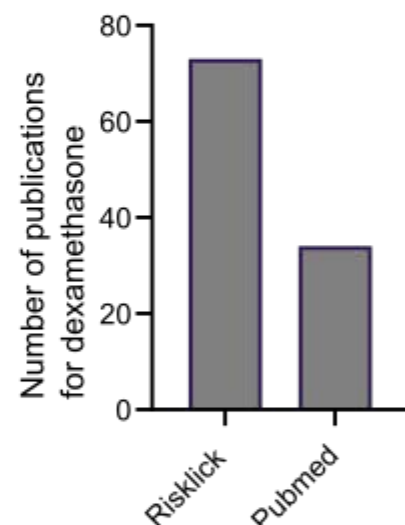
## D



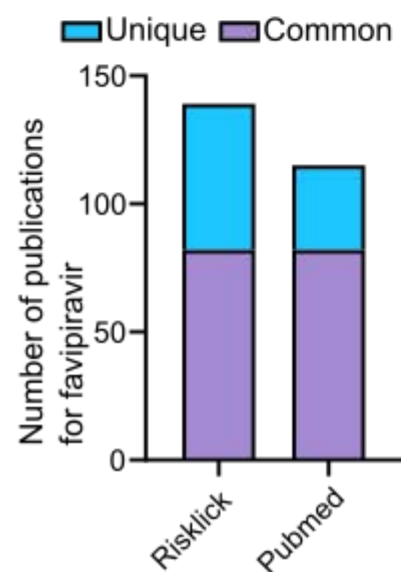
## E



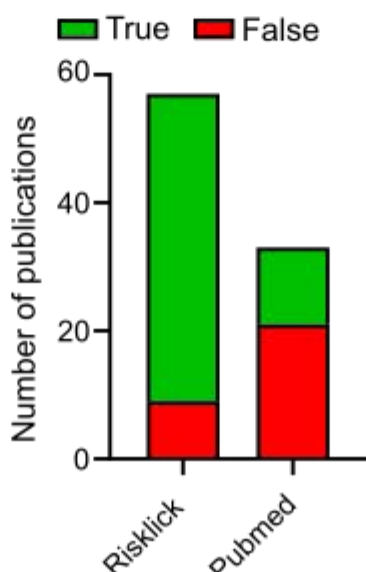
## F



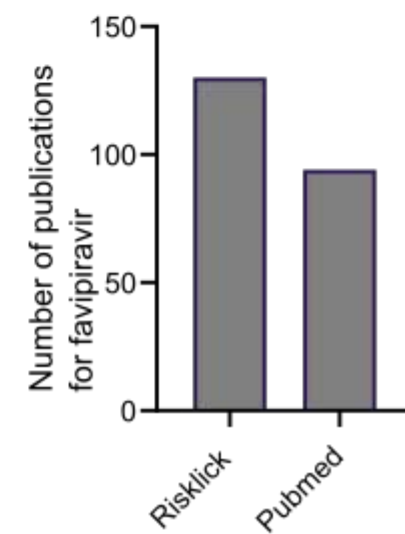
## G

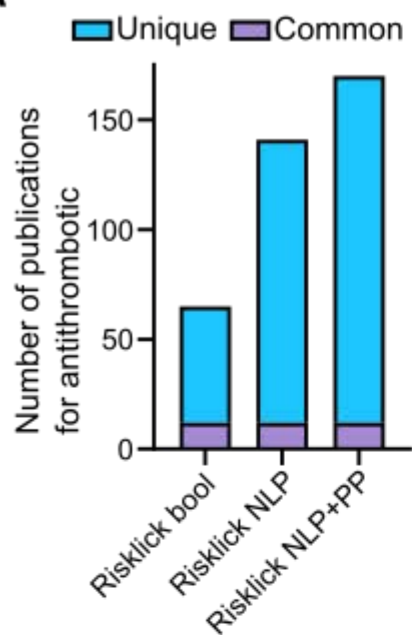
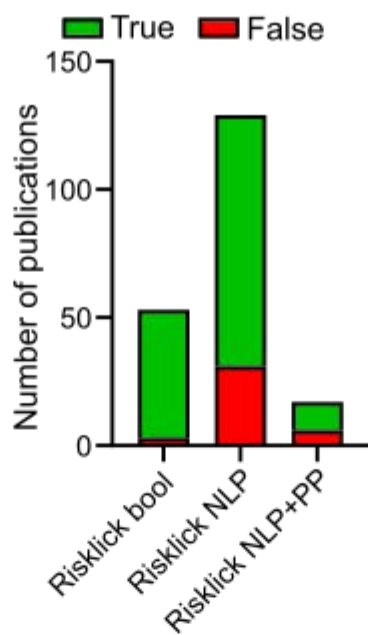
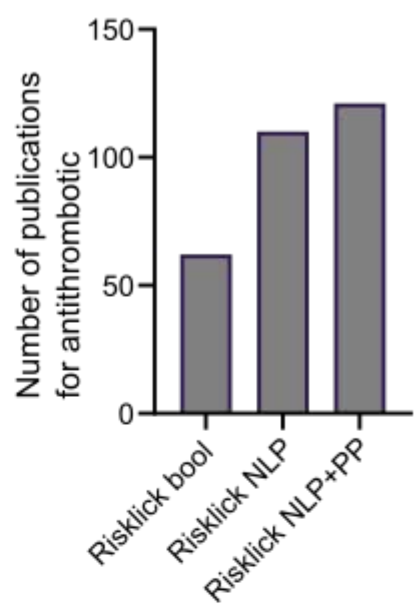
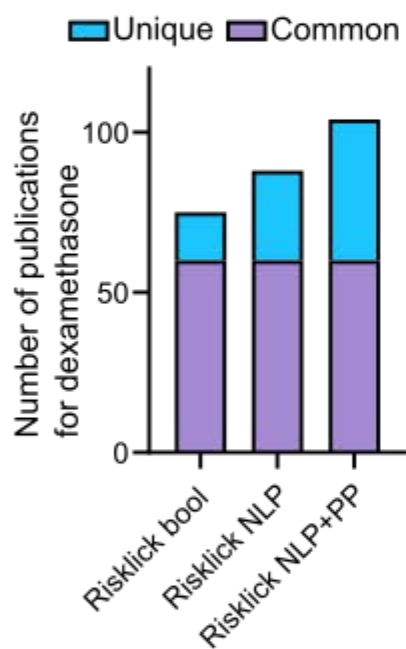
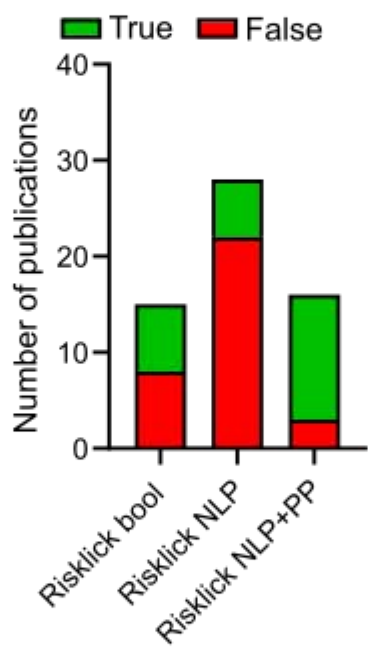
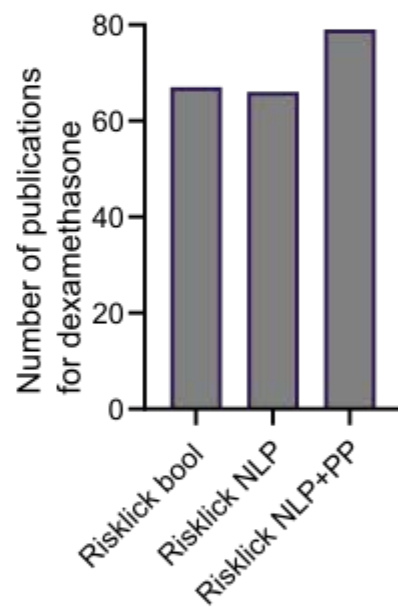
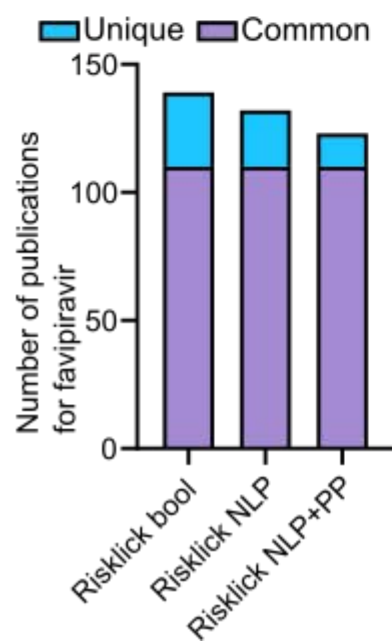
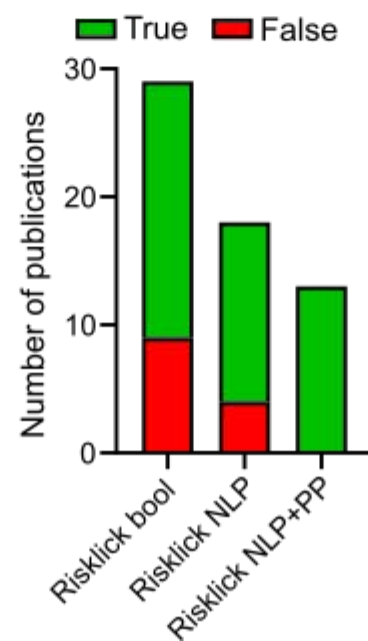
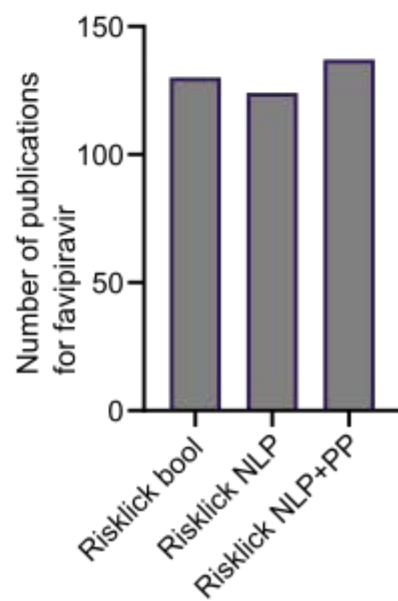


## H

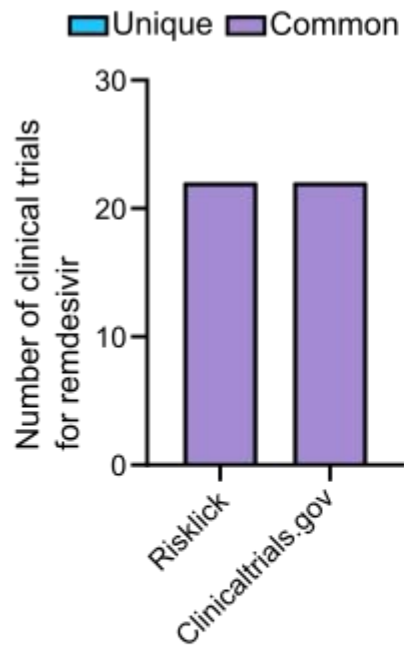


## I

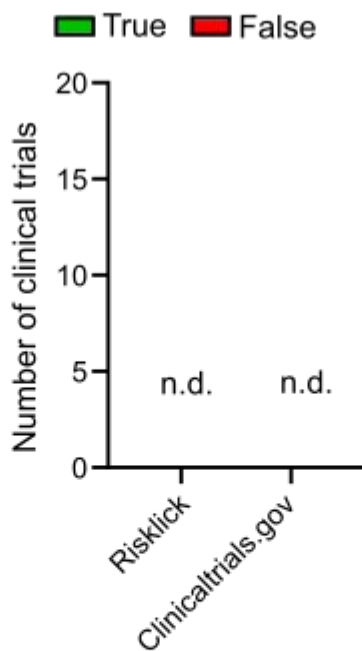


**Figure 5****A****B****C****D****E****F****G****H****I**

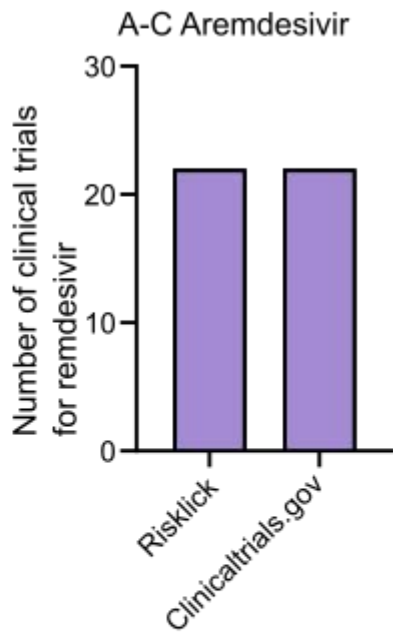
A



B



C



Supp. Figure 2: Risklick Covid-19-related clinical trial search capacities for Remdesivir in comparison to clinicaltrial.gov on the same publication registry. Clinical trials were separated between common and unique outcomes (A). Unique outcomes were analyzed and separated between true-positive (True) and false-positive (False) results (B). Total final number of clinical trials outcomes is the addition of the common findings and unique, true-positive findings (C). n.d., no data.

## Clinical Trials Queries

<b>Query</b>	<b>Date of retrieval</b>
Antibiotic	21.08.2020
anticoagulant	21.08.2020
antiviral	21.08.2020
Remdesivir	26.08.2020
Tocilizumab	26.08.2020
Azithromycin	26.08.2020
Hydroxychloroquine	26.08.2020
Heparin	26.08.2020

## Publications Queries

<b>Query</b>	<b>Date of retrieval</b>
Antithrombotic	10.09.2020
Dexamethasone	10.09.2020
Favipiravir	18.09.2020

Antibiotic		Anticoagulant	Antiviral		Antithrombotic	
Co-trimoxazole	Moxifloxacin	Fondaparinux	Eidd-2801	Darunavir	Anbinex	Prasugrel
Metronidazole	Piperacillin	Argatroban	Ledipasvir	Favicovir	Bemiparin	Tirofiban
Meropenem	Roxithromycin	Warfarin	Merimepodib	Favipiravir	Alteplase	Cangrelor
Vancomycin	Sulfamethoxazole	Idraparinux	Nelfinavir	Galidesivir	Anticoagulant	Clopidogrel
Rifampicin	Tazobactam	Dabigatran	Raltegravir	Inosine pranobex	Fondaparinux	Ticagrelor
Rifabutin	Teicoplanin	Antithrombin	rbt-9	Ixekizumab	Argatroban	Dipyridamole
Triamcinolone	Tetracycline	Apixaban	regn10933	lopinavir/ritonavir	Warfarin	
Rifapentine	Toraymyxin	Bivalirudin	zsp1273	Nitazoxanide	Idraparinux	
Bleomycin	Trimethoprim	Edoxaban	d961h	Novaferon	Dabigatran	
Amoxicillin	Zithromax	Enoxaparin	Pimodivir	Oseltamivir	Antithrombin	
Azithromycin		Heparin	Molnupiravir	Peramivir	Apixaban	
Carriomycin		Bemiparin	Lingdancao	Polyinosinic	Bivalirudin	
Cefdinir		Enoxaparin	Ingavirin	Remdesivir	Edoxaban	
Cefepime		Nafamostat Mesylate	prep-001	Ribavirin	Enoxaparin	
Cefotaxime		Rivaroxaban	Isenolic	Sofosbuvir	Heparin	
Ceftaroline fosamil		Tinzaparin	al-794	Tenofovir	Bemiparin	
Ceftazidime			Leibovitz	Triazavirin	enoxaparin	
Ceftriaxone			Codavax	Umifenovir	Nafamostat Mesylate	
Cephalosporin			bcx-1813	Veldona	rivaroxaban	
Citrox			Aciclovir	vmip	tinzaparin	
Clarithromycin			Aceinavir	Zanamivir	Antiparasitic	
Clavulanate			Amantadine	btl-tml-covid	Atovaquorine	
Clindamycin			Azivudine	c-10276	Levamisole	
Doxycycline			Azvudine	Koronavir	Niclosamide	
Erythromycin			Baloxavir	nnrti	Piperaquine	
Fosfomycin			Clevudine	pf-07304814	Tizoxanide	
Kanamycin			Cobicistat	Polycytidylic acid	dwrx2003	
Levofloxacin			Combivir		Antiplatelet	
Linagliptin			Danoprevir		Dantonic	

Suppl. Table 2



Treatment Pubmed vs clinicaltrials.gov	Research tool	Total raw results	Unique results	Unique, right	Total, right	Total relevant (Risklick+ Pubmed)	TP	TN	FP	FN	Recall	Precision	F1 Score
<b>Remdesivir</b>	Risklick	104	16	10	98	100	98	0	6	2	98.0%	94.2%	96.1%
	clinicaltrials.gov	109	21	2	90		90	0	19	10	90.0%	82.6%	86.1%
<b>Anticoagulant</b>	Risklick	70	20	16	66	67	66	0	4	1	98.5%	94.3%	96.4%
	clinicaltrials.gov	55	5	1	51		51	0	4	16	76.1%	92.7%	83.6%
<b>Antiviral</b>	Risklick	185	83	54	156	160	156	0	29	4	97.5%	84.3%	90.4%
	clinicaltrials.gov	145	43	4	106		106	0	39	54	66.3%	73.1%	69.5%
<b>Azithromycin</b>	Risklick	78	0	0	78	78	78	0	0	0	100.0%	100.0%	100.0%
	clinicaltrials.gov	86	8	0	78		78	0	8	0	100.0%	90.7%	95.1%
<b>Heparin</b>	Risklick	49	13	13	49	49	49	0	0	0	100.0%	100.0%	100.0%
	clinicaltrials.gov	36	0	0	36		36	0	0	13	73.5%	100.0%	84.7%
<b>Remdesivir</b>	Risklick	22	0	0	22	22	22	0	0	0	100.0%	100.0%	100.0%
	clinicaltrials.gov	22	0	0	22		22	0	0	0	100.0%	100.0%	100.0%
<b>Tocilizumab</b>	Risklick	50	4	4	50	50	50	0	0	0	100.0%	100.0%	100.0%
	clinicaltrials.gov	46	0	0	46		46	0	0	4	92.0%	100.0%	95.8%
<b>Hydroxy- chloroquine</b>	Risklick	212	19	10	203	203	203	0	9	0	100.0%	95.8%	97.8%
	clinicaltrials.gov	209	16	0	193		193	0	16	10	95.1%	92.3%	93.7%

Treatment Risklick  
vs Pubmed

<b>Antithrombotic</b>	Risklick	55	22	22	55	64	55	0	0	9	85.9%	100.0%	92.4%
	Pubmed	43	10	9	42		42	0	1	22	65.6%	97.7%	78.5%
<b>Dexamethasone</b>	Risklick	71	44	38	65	72	65	0	6	7	90.3%	91.5%	90.9%
	Pubmed	39	12	7	34		34	0	5	38	47.2%	87.2%	61.3%
<b>Favipiravir</b>	Risklick	107	43	34	98	117	98	0	9	19	83.8%	91.6%	87.5%
	pubmed	104	40	19	83		83	0	21	34	70.9%	79.8%	75.1%