Identifying the Quality Nuggets Amid the Explosion of COVID-19-Related Scientific Communication: An Insurmountable Challenge?

Lavi Oud

To the Editor

The recent reports by Bose and colleagues [1] and by Yanai [2] in this journal review the contemporary data on the management of patients infected with the severe acute respiratory syndrome coronavirus 2, which causes coronavirus disease 2019 (COVID-19), and provide a meta-analysis of the prevalence of diabetes and hypertension among COVID-19 patients with severe vs. non-severe disease, respectively. Both studies provide timely contributions on key clinical and epidemiological issues crucial to clinicians and scientists in the face of the evolving COVID-19 pandemic, as clinicians have been accustomed to instantly accessible high-quality information to support evidence-based decisions.

However, while the narrative review by Bose et al would not be expected to specifically address all published data on the covered topics, it is possible that the authors may not have provided readers with the full spectrum of the data reported by the time of review. Correspondingly, Yanai did not provide the data search strategy used to identify the studies included in the meta-analysis and may not have captured all the relevant publications to adequately estimate the odds of the examined comorbidities in COVID-19 patients with severe vs. non-severe illness.

As discussed below, the abovementioned concerns are, however, not specific to these two studies. Rather, these concerns underscore broader challenges facing clinicians and scientists alike in data discovery that may be unique to the current pandemic and that may slow the growth in better understating of COVID-19, while risking making published synthesizing studies dated at a much accelerated pace.

An increasingly evident corollary of the evolving COVID-19 pandemic has been an unprecedented pace of growth in scientific communication on its clinical, basic science, translational, and societal implications, likely reflecting a corresponding unparalleled sense of emergency (as compared to prior global health crises) among clinicians and scientists to investigate and disseminate their findings. While much of this communication takes place through "non-traditional" venues, such as social media and "pre-prints" (that is, scientific manuscripts posted prior to peer-review and formal acceptance by scientific publications), a tremendous number of publications are added daily through scientific, peer-reviewed journals. The abovementioned heightened sense of emergency in need to for timely dissemination of cutting edge COVID-19-related science would be expected understandably to affect scientific journal editors. However, the need to balance rigorous peer-review, often with its relatively slow pace, against the perceived time pressure to inform clinicians and scientists about the evolving data on all COVID-19-related facets essential to the efforts to mitigate and contain the pandemic, has placed journal editors in an unusual position that may have affected at times the quality and volume of published manuscripts, despite the best intentions [3, 4]. Notably, while the publicity provided to select COVID-19-related reports make them doubtlessly more visible to health care professionals, recent experience has demonstrated that even publication in high-impact scientific journals does not assure the veracity of reported data [5, 6].

The abovementioned state of publication volume and velocity is unlikely to abate in the foreseeable future, creating a substantial challenge for clinicians and scientists in teasing out all relevant published information to inform clinical practice at the point-of-care and planned and ongoing research efforts. At the same time, the identified publications are expected to pass a screening process for a measure of scientific rigor by the publishing journals prior to being indexed in bibliographical datasets. Multiple online resources are presently available for citation retrieval (termed "search engines" hereafter) of scientific publications from bibliographical datasets. However, it is unknown whether the present exponential pace of growth in the volume of published COVID-19-related manuscripts allows adequate identification of relevant citations for specific queries through use of search engines and, as importantly, whether the number of the retrieved citations is manageable for review by clinicians and scientists.

Among the commonly used publicly accessible search engines, PubMed, which serves as an interface to MEDLINE and is managed by the Unites States National Library of Medicine, is considered among the most authoritative. In order to illuminate the challenges facing health care professionals searching across

Table 1. Retrieved COVID-19-Related Citations Obtained Through a Search of the PubMed Search Engine[a]

Citation category[b]	Citation numbers
All	38,485
Clinical studies[c]	546
Letters	7,704
Case reports	1,515
Editorials	3,130
Reviews[d]	3,867
Meta-analyses	88
Comments	2,400
Other[e]	19,235

[a]Search results for citations published from January 1, 2020 through July 20, 2020. The search strategy is detailed in the manuscript. [b]The citation categories included in the table are based on the filter options included in the PubMed search engine and reflect the indexing classification of each citation by the National Library of Medicine. [c]The "Clinical studies" category in the table includes all the following filter options provided in the PubMed search engine: Clinical Study; Clinical Trial Phase I; Clinical Trial Phase II; Clinical Trial Phase III; Clinical Trial Phase IV; Multicenter Study; Observational Study; Pragmatic Clinical Trial; Clinical Trial; Comparative Study; Validation Study; Controlled Clinical Trial; and Randomized Controlled Trial. [d]The "Review" category includes 475 systematic reviews. [e]The "Other" category includes all the citations not classifiable by the filter options included in the table. These include predominantly citations indexed as "Journal Article".

scientific publications on COVID-19 for guidance on clinical practice and as background for research, the PubMed dataset was searched to quantify the COVID-19-related publications indexed to date. The identified citations were then categorized using PubMed's filtering tools, thus providing an overall and category-specific quantification of the citation volume. The search included citations from January 1 through July 20, 2020, using the following search approach: "Coronavirus Disease 2019" OR "COVID-19" OR "SARS-CoV-2" OR "coronavirus" OR "corona virus" OR "coronaviridae" OR "coronavirinae" OR "betacoronavirus" OR "covid19" OR "covid 19" OR "nCoV" OR "CoV2" OR "sarscov2" OR "2019nCoV" OR "novel CoV" OR "wuhan virus". No language restrictions were applied. Although searchers may be focused on more narrow aspects of published data on COVID-19 (e.g., specific molecular-level processes, clinical manifestations involving a specific system), the current stage of the pandemic is at such an early phase that more general data are expected to be a common focus of data queries. Thus, the search strategy was geared to allow broad and comprehensive citation retrieval, to provide a general perspective on the data volume faced by searchers at the present phase of the pandemic.

The search findings are summarized in the Table 1. Over 38,000 COVID-19-related citations were identified, of which 96.7% were in English. For perspective, the H1N1 influenza pandemic, which has started in the spring of 2009, had a global case estimate of nearly 61 million and possibly over 500,000 deaths during the first 12 months [7]. However, there were "only" about 2,000 PubMed citations on this virus in 2009, with the total number of PubMed citations (that is, including the subsequent 11 years) of about 21,000.

Only 546 (1.4%) of the COVID-19-related citations were identified by PubMed as clinical studies of any type. This number of clinical studies, while considerable, may be potentially manageable for screening and narrowing down by investigators, though not by busy clinicians. However, a substantial number of clinical study findings (mostly observational) were and continue to be communicated as letters to the editor [8-10], thus remaining opaque to the current filtering systems of PubMed and invisible as such to searchers. Indeed, there were over 7,700 citations identified as letters, comprising 20% of all identified COVID-19-related citations. This latter number of citations is no longer readily manageable for further narrowing to pertinent ones, with the challenge compounded by the lack of abstracts for this type of citations. It appears that letters have become a preferential tool for COVID-19-related scientific communication, possibly due to faster time to publication.

Given the tremendous volume of retrievable original reports, publications synthesizing the available data on COVID-19-related topics such as that reported by Bose et al [1] can be indispensable starting points for clinicians and scientists. However, another striking finding of the citation search was that despite the brief duration the COVID-19 pandemic, spanning barely over 6 months, there was a remarkably large volume of review articles, with over 3,800 review citations, making it all but impracticable for clinicians and scientists to make an informed selection among the retrieved citations. Moreover, because the majority of the citations in this latter group appear to be narrative reviews, the capacity of these publications to fully peruse, appraise, and synthesize preceding COVID-19 data, which can be of low quality due to risk of bias and serious imprecision [11], is uncertain. Using the abovementioned example of the H1N1 influenza pandemic as a comparator, the number of all review citations on this virus as of this writing was 1,268 (with 128 citations in 2009). Last, the rapid pace of growth in reported COVID-19-related data risk making even the most meticulously searched reviews dated within the span of weeks or months.

In contrast to the voluminous number of published reviews, the number of citations identified as meta-analyses (88) on COVID-19 related topics appears more manageable for screening and review. However, it appears that not all such studies are truly identifiable through data search. Thus, the meta-analysis reported by Yanai [2] is invisible to searches focusing on this type of publication. This lack of visibility to targeted search is because the meta-analysis was published in a letter form, is identified in PubMed as such, and is not retrieved when a search on COVID-19-related meta-analyses is performed. Thus, the actual total number of published meta-analyses on COVID-19 and the number of those published in a letter form (as well as the frequency of the latter publication category among all meta-analyses, as of this writing) cannot be credibly determined.

PubMed provides searchers with multiple functions to enhance the relevance of retrieved citations, including among others Medical Subject Heading terms and Boolean operators. However, only a small minority of queries were found in a previous study to include either [12], and it is uncertain whether the use of these or other functions could narrow sufficiently the number of retrieved COVID-19-related citations, while maintaining their relevance. Nor can the use of these functions over-

come the limitations inherent in PubMed indexing procedures.

In conclusion, the abovementioned findings on the current state of published scientific communication on COVID-19 indicate that the volume of retrievable citations to be perused to identify all those pertinent to a general search and possibly to that of more focused ones is over an order of magnitude greater than that reported during the early phase of the most recent global pandemic [7]. While it appears that the present approaches to bibliographical indexing may not adequately meet the current data search needs, workable, timely solutions are not readily apparent. Temporizing approaches, such as consistent inclusion of the study type in manuscript titles, especially when reported in a letter form [13, 14], may offer better data transparency, though this will not reduce the volume of potentially relevant retrieved citations. However, the practicality of even such limited-scope approaches remains to be determined. At the same time many citations relevant to the topics of interest may not be readily identifiable. Thus, although elaborate search strategies may usually help health care professionals identify a workable number of relevant citations for specific queries, clinicians with limited data retrieval skills and experienced investigators alike will likely continue to face considerable challenges when seeking to identify the rapidly evolving evidence to inform the care of COVID-19 patients and to address the broader implications of the evolving pandemic.

Acknowledgments

None to declare.

Financial Disclosure

None to declare.

Conflict of Interest

None to declare.

Informed Consent

Not applicable

Author Contributions

Lavi Oud has performed the design of work, acquisition of data, analysis and interpretation , composition, drafting, revising, editing, and final approval.

Data Availability

The data supporting the findings of this study are available via PubMed.gov and can be accessed from https://pubmed.ncbi.nlm.nih.gov/.

References

1. Bose S, Adapa S, Aeddula NR, Roy S, Nandikanti D, Vupadhyayula PM, Naramala S, et al. Medical management of COVID-19: evidence and experience. J Clin Med Res. 2020;12(6):329-343.
2. Yanai H. A Significance of High Prevalence of Diabetes and Hypertension in Severe COVID-19 Patients. J Clin Med Res. 2020;12(6):389-392.
3. Chotirmall SH, Martinez FJ, Schumacker PT, Cooke CR, Seam N, Brochard L, Tighe RM, et al. Life at the editorial "COVID Frontline". The American Thoracic Society Journal Family. Am J Respir Crit Care Med. 2020;201(12):1457-1459.
4. Kaplan LJ, Bleck TP, Buchman TG, Dellinger RP, Deutschman CS, Marshall JC, Maslove DM, et al. Pandemic-related submissions: the challenge of discerning signal amidst noise. Crit Care Med. 2020;48(8):1099-1102.
5. Mehra MR, Ruschitzka F, Patel AN. Retraction-Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. Lancet. 2020;395(10240):1820.
6. Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Retraction: cardiovascular disease, drug therapy, and mortality in COVID-19. N Engl J Med. 2020;382(26):2582.
7. Pandemic (H1N1pdm09 virus). Centers for Disease Control and Prevention. https://www.cdc.gov/flu/pandemic-resources/2009-h1n1-pandemic.html.
8. Zhang H, Zhou P, Wei Y, Yue H, Wang Y, Hu M, Zhang S, et al. Histopathologic changes and SARS-CoV-2 immunostaining in the lung of a patient with COVID-19. Ann Intern Med. 2020;172(9):629-632.
9. Haberman R, Axelrad J, Chen A, Castillo R, Yan D, Izmirly P, Neimann A, et al. COVID-19 in immune-mediated inflammatory diseases - case series from New York. N Engl J Med. 2020;383(1):85-88.
10. Huang Y, Chen S, Yang Z, Guan W, Liu D, Lin Z, Zhang Y, et al. SARS-CoV-2 viral load in clinical samples from critically ill patients. Am J Respir Crit Care Med. 2020;201(11):1435-1438.
11. Siemieniuk RA, Bartoszko JJ, Ge L, Zeraatkar D, Izcovich A, Kum E, Pardo-Hernandez H, et al. Drug treatments for covid-19: living systematic review and network meta-analysis. BMJ. 2020;370:m2980.
12. Herskovic JR, Tanaka LY, Hersh W, Bernstam EV. A day in the life of PubMed: analysis of a typical day's query log. J Am Med Inform Assoc. 2007;14(2):212-220.
13. Pinto LC, Bertoluci MC. Type 2 diabetes as a major risk factor for COVID-19 severity: a meta-analysis. Arch Endocrinol Metab. 2020;64(3):199-200.
14. Kunutsor SK, Laukkanen JA. Cardiovascular complications in COVID-19: A systematic review and meta-analysis. J Infect. 2020;81(2):e139-e141.