Triaging diabetes research in perplexed COVID-19 resources: Avoiding research waste through a corpus-based approach

Dear Editor,

After the emerging of COVID-19, concerns raised regarding research waste due to poor quality studies published without effective evidence-based response.[1] A torrent of coronavirus-related articles has been released within a few months after the pandemic. Although COVID-19 is a hot topic for researchers, careful steps should be taken toward shedding light on the overall approach to fund, conduct, and publish research papers to avoid research waste.

The burden of coronavirus has been globally increasing in diabetic patients.[2] Since diabetes along with other comorbidities are significant predictors of morbidity and mortality in patients with COVID-19, future research is urgently needed on COVID-19 and diabetes, and its clinical management.[3]

On the other hand, attempts should be made to systematize ongoing data by reviewing the rapidly growing body of recent research in order to present a bibliometric analysis of published papers.[4] Applying "linguistic corpora" is a very useful tool. Corpora yield trends and patterns based on the semantic relationship within texts. The value of the data from a corpus lies in the fact that researchers can compare the results of other studies or texts.[5] Thus, it is required to critically analyze the information about COVID-19 and diabetes.

The data-processing capability of empirical data helps review research trends via applying systematic citation analysis procedures to summarize and identify salient research themes which significantly contribute to the dynamic interaction of research in the literature. Table 1 shows common terms, along with their frequency of occurrence, collocates, patterns, and comparisons in COVID-19 literature.

Certain diabetes corpora have been developed which give us the insight to consider them in coronavirus-related diabetes research. The advantages of using corpora instead of systematic reviews include being remarkably faster, being much easier to perform, finding quick replies to questions with a few clicks, and access to multiple relevant sources with a few clicks.

To sum up, given the potential of linguistic corpora as a tool in trendspotting research, there is compelling evidence which drives us to triage research needs during and after the coronavirus pandemic to avoid research waste using a corpus-based approach. Due to data-processing capability of a bulk of empirical linguistic data and interface with academic disciplines, using linguistic corpora, as a solution, can remarkably alleviate concerns about the waste in research by improving quality and effectiveness, that is, making corpus-based decisions and prioritizing research projects to monitor and guarantee an optimum conduction of projects and publishing papers on diabetes-COVID-19. Developing new specialized diabetes corpora, in addition to using available corpora, is strongly recommended to extract patterns and trends through natural language processing and text mining techniques by identifying semantic relations between concepts to extract patterns and trends from texts.

Table 1: Results of searching the key terms related to COVID-19 in English-Corpora.org

Frequent key terms	Collocates	Patterns	Compare	Key terms related to diabetes (frequency of occurrence)
Coronavirus	Virus	Stay-at-home	Social*	Diabetes (6147)
COVID-19	Ban	Social*	Economic*	Diabetic (712)
Social distancing	Stockpile	Social care	Ban	Diabetics (218)
Flatten the curve	Disinfect*	Social cohesion	Obey	COVID-19+diabetes (168)
Work* home	Remotely	Economic	Social media	Mortality+diabetes (10)
Zoom	Reopen	Hoard*	Social care	Infection+diabetes (27)
Wuhan	Dangerous	Ban (noun)	Social distancing	Type 2 diabetes (389)
Hoard*	Spray	Ban (verb)	Social distance	Obesity (1350)
Toilet paper	Masks	Lockdown	Social contact	Glucose (497)
Curbside	Surfaces	Contact	Economic impacts	Glucose control (12)
Pandemic	Hands	Quarantine	Economic consequence	Metabolic* (450)
Reopen	Clean	Contact	Economic costs	Glycaemic* (17)
Defy	Sanitizer	Restriction	Ban travel	Insulin (914)

[a]Frequent terms: Terms related to COVID-19 found with the highest frequency in the literature, [b]Collocates: Terms which are mostly collocated with COVID-19 in the literature, [c]Patterns: The patterns in which words related to COVID-19 occur in the literature, [d]Compare: Comparing how common phrases related to COVID-19 have changed over time in the literature, [e]Key terms related to diabetes (frequency): Diabetes-related terms with their frequency of occurrence in COVID-19 literature

Conflicts of interest

There are no conflicts of interest.

Davood Khalili[1,2], Mohsen Varzandeh[3]

[1]Prevention of Metabolic Disorders Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran, [2]Department of Biostatistics and Epidemiology, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran, [3]Health System Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran

Address for correspondence: Dr. Mohsen Varzandeh, Shiraz University of Medical Sciences, Shiraz, Iran.
E-mail: mvar61@gmail.com, mvarzandeh@hotmail.com

REFERENCES

1. Glasziou PP, Sanders S, Hoffmann T. Waste in covid-19 research. BMJ 2020;369:m1847.
2. Singh AK, Gupta R, Ghosh A, Misra A. Diabetes in COVID-19: Prevalence, pathophysiology, prognosis and practical considerations. Diabetes Metab Syndr 2020;14:303-10.
3. Hussain A, Bhowmik B, do Vale Moreira NC. COVID 19 and diabetes: Knowledge in progress. Diabetes Res Clin Pract 2020;162:108142.
4. Latif S, Usman M, Manzoor S, Iqbal W, Qadir J, Tyson G, *et al.* Leveraging data science to combat COVID 19: A comprehensive review. IEEE Transactions on Artificial Intelligence; 2020;1.
5. Kwary DA. A corpus and a concordance of academic journal articles. Data Brief 2018;16:94-100.

Access this article online	
Quick Response Code:	**Website:** www.jmsjournal.net
	DOI: 10.4103/jrms.JRMS_920_20