



## Is the content of scientific studies written on COVID-19 biostatistically satisfying?

Selcen Yüksel 

Department of Biostatistics, Ankara Yıldırım Beyazıt University Faculty of Medicine, Ankara, Turkey

Received: April 22, 2020 Accepted: April 27, 2020 Published online: May 07, 2020

Coronavirus disease 2019 or COVID-19 has rapidly emerged as a global pandemic. For the past three months, researchers have been competing to test their hypotheses on COVID-19. To search the clinical characteristic prognostic factors of patients with COVID-19 and to evaluate the diagnosis and treatment strategies on this pandemic, many researchers have immediately carried out clinical studies on COVID-19. This was such a speed that 4,831 articles from ((((((“COVID-19”[Title] OR “2019-ncov”[Title] OR “cov-19”[Title] OR “coronavirus-19”[Title] OR “SARS-CoV-2”[Title] OR “Severe acute respiratory syndrome coronavirus 2”[Title] OR “Wuhan virus”[Title] OR “Wuhan novel coronavirus”[Title] search filter were published in PubMed until 21/04/2020. Although these efforts are well-regarded due to oriented to get an urgent cure, they may sometimes need judgment as to scientific grounds.

Biostatistics is the most important discipline to provide tools and methods to find the structure in and to provide a deeper insight into data, analyze it, and quantify uncertainty. In the American Statistical Association’s ethical guidelines, inclusion of detailed statistical points for any clinical study is stated.<sup>[1]</sup> The main quality checklist elements of any clinical study are as follows: calculation of the required sample size and defining statistical methodology deeply.<sup>[2]</sup> If a study has satisfied these two points, the results can be counted on to be unbiased. An important criterion of any prestigious scientific journal for a clinical article to be candidate to publish is that it must contain statistical method section. This section mainly

must include used statistical analyses and software information. Apart from this, sample size calculation, the statistical power of the study, should be considered in this section or anywhere in the article.

In this context, a question has aroused, if the published papers on COVID-19 provide the statistical criteria which are critically checked before by editors. Or are these criteria eased off to win COVID-19 clinical hypotheses testing competition? There are two biostatistical questions for COVID-19 studies: how many of the published articles mentioned the sample size calculation in their paper in this limited time? How many of them included statistical method or statistical analysis section separately in their studies? The answers are so important that reliability of the conclusions reached in any article solely depend on sufficiency of statistical methodology and sample size. Electronic searches were performed in electronic libraries of two universities (Ankara University and Ankara Yıldırım Beyazıt University). Two main filters were used without date restriction. The first filter included all studies in online databases without any restriction. The second filter included the studies published at academic journals and conference materials. The last search was carried out on the date of 21.04.2020. The search results are summarized in Table 1.

It is clear that, in both databases, about 4% of all documents with the title on COVID-19 contain some basic statistical terms. In the first database, only 137 of 10,942 academic studies (1.25%) about COVID-19 mentioned the sample size or statistical power or type-II error in their context. In this database, totally 244 studies included these keywords and 56% of them

**Corresponding author:** Selcen Yüksel, MD. Ankara Yıldırım Beyazıt Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, 06800 Çankaya, Ankara, Türkiye.

e-mail: selcenpehlivan@yahoo.com

Cite this article as:

Yüksel S. Is the content of scientific studies written on COVID-19 biostatistically satisfying? Turk J Phys Med Rehab 2020;66(2):223-224.

**Table 1.** Search results from two databases-Search date: 21.04.2020

	No restriction	Academic journals and conference materials	No restriction	Academic journals and conference materials
	1 <sup>st</sup> electronic library		2 <sup>nd</sup> electronic library	
Just in Title	174.819	10.942	92.582	14.873
Title AND TX ("data analysis")	440	144	265	156
Title AND TX ("statistical method")	7	3	5	4
Title AND TX ("statistical analysis")	166	121	144	122
Title AND TX ("statistical power")	9	4	5	5
Title AND TX ("Type-II error")	1	0	1	0
Title AND TX ("sample size")	234	133	187	134
Title AND TX (software)	5.965	240	2.469	262
Title AND TX ("statistical software")	21	17	26	17
Title: TI ("COVID-19" OR "2019-ncov" OR "cov-19" OR "coronavirus-19" OR "SARS-CoV-2" OR "Severe acute respiratory syndrome coronavirus 2" OR "Wuhan virus" OR "Wuhan novel coronavirus")				

were academic studies. Using the same keywords in the second database, 139 of 14,873 academic studies (0.93%) about COVID-19 had these keywords anywhere in their context. Therefore, it is clear that the articles published under the title "COVID-19" do not satisfy the readers, according to the sample size data.

The same disappointment is relevant for the section that any clinical study must include: statistical method. In the first and second databases, only 268 of 10,942 (2.44%) and 282 of 14,873 (1.89%) academic studies about COVID-19 mentioned the data analysis or statistical method or statistical analysis in their context, respectively. In the second database, totally 613 studies (0.35% of all results) included these keywords and 43.7% of them were academic studies.

For the evaluation of the number of statistical software used in articles, within 268 articles using statistical analysis in their all text, 257 (95%) in the first database and 257 (98%) in the second database referred statistical software. Although these ratios were high within the papers including statistical procedure, in all academic studies, the number of studies including statistical software or software keywords was 257 (2.34%) at the first database.

In conclusion, although the world presently needs fast solutions for COVID-19, reliability and accuracy of the clinical research results must be also satisfactory. It is, therefore, a prerequisite to provide an adequate sample size obtained from statistical calculations and a proper statistical methodology. Using different keywords at different time points, the search results can be changed; however, the reality will be the same. Journal editors are also required to act selectively for the candidate papers on COVID-19 to keep the garbage of biased information away from the literature.

#### **Declaration of conflicting interests**

The author declared no conflicts of interest with respect to the authorship and/or publication of this article.

#### **Funding**

The author received no financial support for the research and/or authorship of this article.

#### **REFERENCES**

1. American Statistical Association, Ethical Guidelines for Statistical Practice, April 14, 2018.
2. Weihs C, Ickstadt K. Data Science: the impact of statistics. *Int J Data Sci Anal* 2018;6:189-94.