Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Letter to the Editor

Questions about Tosepu et al. (2020) “Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia”

Keywords:
COVID-19
Indonesia
Temperature
Weather
Correlation

1. Introduction

At the start of global pandemic in early 2020, numerous peer-reviewed journal articles examined weather data and COVID-19 cases to see if the stability of the virus was affected by weather conditions. For example, Tan and Schultz (2021) documented 158 articles published before 31 March 2021 that contained quantitative statistical correlations on the weather–COVID-19 relationship. One of the earliest articles to be published on this topic was Tosepu et al. (2020a) “Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia”. This article was received by Science of the Total Environment on 2 April 2020, accepted on the same day, and published online two days later. Tosepu et al. (2020a) is also one of the most highly cited articles on the weather–COVID-19 relationship (Zhoud and Zhoud, 2021) with 623 citations on Google Scholar and 283 in Web of Science, as of 9 January 2022. In this Letter to the Editor, I raise five questions about the article. My questions pertain to the origins of the data in their figure, how the analysis was conducted, how their results are interpreted, and how they are presented in the article.

2. Discussion: questions about the article

Tosepu et al. (2020a) obtained daily COVID-19 data for Jakarta from the Ministry of Health of the Republic of Indonesia and compared it to daily weather data (i.e., maximum, minimum, and average temperatures; relative humidity; rainfall amount) for February 2020 from the Meteorological Department of the Republic of Indonesia. The principal data for the study are shown in six time-series plots in their Fig. 1 (part also replicated in the graphical abstract to this present Letter to the Editor). Tosepu et al. (2020a) then used the Spearman rank correlation test to calculate the correlation between these time series. These correlation coefficients were presented in their Table 1, showing that the only statistically significant correlation was between the average temperature and number of COVID-19 cases.

With this background, we can interrogate Tosepu et al. (2020a) and ask questions about the data. The first panel of Fig. 1 “Cases of the Covid-19” ranges from 1 to 11 days. However, the remaining five panels range from 1 to 29 days, which seems to be referred to in the text as “weather data for the period of February 2020” (i.e., there were 29 days in February 2020). The discrepancies between the caption and the x axes in the panels means it is unclear exactly how many days were used in the analysis performed in this study, nor does the article clearly indicate what specific dates were involved in the analysis. Neither the text nor the figure caption provided a clear statement as to which 11 and 29 dates were associated with values along the x axes in these panels.

Also, the text offers different interpretations of this information. The text reads, “Fig. 1 shows that covid-19 [sic.] that occurred in Jakarta experienced a rapid increase. The first finding numbered 177 cases, the next report found 40 cases, and on 29 March 2020 the total number of covid-19 [sic.] cases numbered 678.” These numbers are inconsistent with a similar statement in the highlights where the spread is referred to as “very fast”. In fact, the first panel shows a decrease from day 1 to day 11, with a slight rise afterward. Also, it is unclear what days are associated with “the first finding” and “the next report”. Are these labeled days 1 and 2 in Fig. 1? If so, what dates were these associated with? The caption says that the plots contain data over the range “from January to March 29, 2020”, but no starting day is indicated. Moreover, this time period is greater than the number of data points in each of the panels.

Understanding the exact number of days and the specific dates of these data are critical. These data in their figure go into the calculation of the Spearman correlation coefficients presented in their Table 1. Thus, the discrepancy between the number of days in the COVID-19 cases and weather panels (and the data being unavailable to readers) prevents the reader from making their own assessment of the relationships purported by the authors. An analysis with only 11 data points would have a different interpretation than one with 29 data points.

• Question 1: How many days (11 or 29) and which specific dates were used in the analysis that produced the Spearman correlation coefficients? Why is there a discrepancy between the number of days in the panels in Fig. 1?

• Question 2: How does the COVID-19 data in Fig. 1 illustrate a “rapid increase” and a “very fast” spread?

On the temperature minimum and maximum panels of their Fig. 1, some days have temperatures of 0 °C. Yet, the average temperatures all appear to be above 26 °C, which is consistent with the text: “the lowest average temperature of 26.1°C and the highest temperature of 28.6°C.” The reported 0 °C for a maximum temperature would be less than the minimum temperature on days 7 and 11, which makes no sense, and is inconsistent with the text which states the ranges in maximum and minimum temperatures. Also, a temperature of 0 °C would be a record low temperature for some days have temperatures of 0 °C. Yet, the average temperatures all appear to be above 26 °C, which is consistent with the text:

http://dx.doi.org/10.1016/j.scitotenv.2022.154078
0048-9697/© 2022 Elsevier B.V. All rights reserved.
• Question 3: What is the meaning of the zero-temperature values, and were they included in the calculations for the Spearman correlation coefficients?

Nowhere in the manuscript is the time lag between the weather data and the COVID-19 cases discussed. If the authors are testing the hypothesis that the weather affects the spread of COVID-19, then some time lag must be applied between the weather experienced by an individual when infected by the virus and when that individual's positive case is recorded in the official database. The time lag between these two events is the sum of the time for the viral infection to become manifest as symptoms of COVID-19, the time for the symptoms to become severe enough for the patient to seek medical attention, the time for the test to come back positive, and the time for the case to be recorded in the database. Other published studies on the weather–COVID-19 relationship typically assume time lags of 2–14 days, although 51% of 158 such studies do not state what time lag they applied (Tan and Schultz, 2021).

A follow-up paper, Tosepu et al. (2020b) in Public Health of Indonesia, also appears to have similar issues with its data, figure, and analyses. Specifically, Tosepu et al. (2020b) write that “Another research direction is to lag the effects of meteorological factors on COVID-19 in Indonesia,” suggesting that they did not perform a lag analysis in either study. Indeed, Kerr et al. (2021) also identify the lack of a time lag in Tosepu et al. (2020b) to be problematic. If the authors of Tosepu et al. (2020a) had not assumed a time lag in their calculations, they would be testing the relationship between COVID-19 and the weather on the date the case was recorded in the database rather than on the approximate day that the infection occurred. If so, then this result could not be used to support the statement that weather was a factor in the spread of COVID-19 as the weather would happen after the transmission had already occurred. As their article does not record the fact that a time lag was included, nor is it listed as being a limitation of this study (Tosepu et al., 2020a, p. 3), the question arises as to whether the time lag was included.

• Question 4: Was a time lag applied to the data in the Tosepu et al. (2020a) analysis? If so, what was it? If not, then how would this change the interpretation of their results?

Despite my critical questions above, let's take the authors' results at face value. They write, “The weather is an important factor in determining the incidence rate of covid-19 [sic.] in Jakarta. Temperature average was significantly correlated with covid-19. Our findings can be used as an input in suppressing covid-19 disease in Indonesia.”

A correlation of 0.392 between the average temperature and the number of COVID-19 cases only explains 15% of the variance. Even if the result is statistically significant, this does not imply practical significance or importance, as the recent debates about the value of calculating statistical significance in various journals and disciplines have discussed (e.g., Nicholls, 2000; Cumming, 2014; Nuzzo, 2014; Ponsky and Oswald, 2014). The question should be whether a relationship of this magnitude between average temperature and the number of COVID-19 cases is meaningful and actionable. How low this percentage of variance explained can be considered an “important factor” is unclear. Moreover, there appears to be a change in wording from the main body of the text (“weather is an important factor”) to the highlights (“one of the factors that triggered the spread of Covid-19”). Highlights in Science of the Total Environment should “capture the novel results of your research” (https://www.elsevier.com/journals/science-of-the-total-environment/0048-9697/guide-for-authors). Yet, there is a big difference in interpretation between writing that temperature “triggered the spread of COVID-19”, which is what the authors wrote, and writing that temperature is a factor affecting the spread, which is what they should have written [again, if their results are taken at face value]. A factor that initiates the pandemic versus influences the pandemic are two different things.

• Question 5: Given that the average daily temperature was the only variable having a statistically significant relationship (and only a factor that apparently explained 15% of the variance), how do the authors explain that the “weather” can be said to be an “important factor” and one that “triggered” the spread of COVID-19 during the onset of a pandemic?

3. Conclusions

This review has raised some serious questions about the data, analysis, and interpretation of the results of Tosepu et al.'s (2020a) study relating COVID-19 and the weather in Jakarta, Indonesia. In particular, I identified the following five questions:

• the number of days and dates of the data that was analyzed,
• how the authors claimed a rapid increase in COVID-19 from the data presented,
• why apparently unrealistic temperatures of 0 °C were included in the analysis,
• the problem with the lack of a time lag between the weather on the day of infection and the date the case was recorded, and
• the interpretation that the weather is an “important factor” that “triggered” the spread of COVID-19 in Jakarta.

For most readers encountering Tosepu et al. (2020a) for the first time, a quick glance at the inconsistencies in the x axes in their Fig. 1 would have been the first sign to indicate potential problems with this article. In fact, other authors have reported on the general low-quality of the body of literature addressing the weather–COVID-19 relationship during the earliest phase of the pandemic (e.g., Zaitechik et al., 2020; Kerr et al., 2021; Quintana et al., 2021; Tan and Schultz, 2021; WMO, 2021). Although scientists were heavily invested in learning everything about the pandemic, especially early on, and studies were conducted and published quickly, one of the hallmarks of science is the ability to re-examine previously published results and re-evaluate the process that was conducted at the time.

The scientific community should not shy away from critical reflection about how it does science and how it communicates science. I believe that the scientific community would benefit from reading answers to the five questions from the authors of Tosepu et al. (2020a, 2020b). Their responses would bring clarity to the questions that I had when reading, allowing other readers in the future to understand this article better. I look forward to the engagement from the authors of Tosepu et al. (2020a).

CRediT authorship contribution statement

David M. Schultz: Conceptualization, Method, Formal analysis, Investigation, and Writing.

Declaration of competing interest

I declare no competing financial interests.

Acknowledgments

Partial funding was provided by the Natural Environment Research Council grant NE/N003918/1 to the University of Manchester. I thank the two anonymous reviewers and the Editor Prof. Barceló for their comments that improved this manuscript.

Data availability statement

No data was created or analyzed in this study.

References

Cumming, G., 2014. The new statistics: why and how. Psychol. Sci. 25, 7–29. https://doi.org/10.1177/0956797613504966.
Kerr, G.H., Badr, H.S., Gardner, L.M., Perez-Saex, J., Zaitechik, B.F., 2021. Associations between meteorology and COVID-19 in early studies: inconsistencies, uncertainties, and recommendations. One Health 12, 100225. https://doi.org/10.1016/j.ongehlt.2021.100225.
Nicholls, N., 2000. The insignificance of significance testing. Bull. Am. Meteorol. Soc. 81, 981–986. https://doi.org/10.1175/1520-0477(2000)081<0981:TSCT>2.3.CO;2.

Nuzzo, R., 2014. Statistical errors: P values, the ‘gold standard’ of statistical validity, are not as reliable as many scientists assume. Nature 506, 150–152.

Ploisky, L., Oswald, F.L., 2014. How big is “big”? Interpreting effect sizes in L2 research. Lang. Learn. 64, 878–912. https://doi.org/10.1111/lang.12079.

Quintana, A.V., Clemons, M., Hoeve, K., Liu, A., Balbes, J., 2021. A descriptive analysis of the scientific literature on meteorological and air quality factors and COVID-19. GeosHealth 5. https://doi.org/10.1029/2020GH000367.

Tan, L., Schultz, D.M., 2021. How is COVID-19 affected by weather? Meta-regression of 158 studies and recommendations for best practices in future research. Weather Clim. Soc. https://doi.org/10.1175/WCAS-D-21-0132.1 in press.

Tosepu, R., Gunawan, J., Effendy, D.S., Ahmad, L.O.A.I., Lestari, H., Bahar, H., Asfian, P., 2020a. Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. Sci. Total Environ. 725, 138436. https://doi.org/10.1016/j.scitotenv.2020.138436.

Tosepu, R., Effendy, D.S., Ahmad, L.O.A.I., Lestari, H., Bahar, H., Asfian, P., Sakka, A., 2020. Vulnerability of Weather on COVID-19 Pandemic in West Java, Indonesia. Public Health of Indonesia 6 (4), 123–128. https://doi.org/10.36685/pbi.6(4).357.

Zaitchik, B.F., et al., 2020. A framework for research linking weather, climate and COVID-19. Nat. Commun. 11, 5730. https://doi.org/10.1038/s41467-020-19546-7.