medRxiv preprint doi: https://doi.org/10.1101/2020.04.23.20077008; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

The closer to the Europe Union headquarters, the higher risk of COVID-19? Cautions regarding ecological studies of COVID-19

Shuai Li[1,2*], Xinyang Hua[3]

[1] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Victoria, Australia

[2] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom

[3] The National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

*Corresponding author

Address: Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, 207 Bouverie Street, Parkville, Victoria 3051, Australia

Email: shuai.li@unimelb.edu.au

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.23.20077008; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

Several ecological studies of the coronavirus disease 2019 (COVID-19) have reported correlations between group-level aggregated exposures and COVID-19 outcomes. While some studies might be helpful in generating new hypotheses related to COVID-19, results of such type of studies should be interpreted with cautions. To illustrate how ecological studies and results could be biased, we conducted an ecological study of COVID-19 outcomes and the distance to Brussels using European country-level data. We found that, the distance was negatively correlated with COVID-19 outcomes; every 100 km away from Brussels was associated with approximately 6% to 17% reductions (all $P<0.01$) in COVID-19 cases and deaths in Europe. Without cautions, such results could be interpreted as the closer to the Europe Union headquarters, the higher risk of COVID-19 in Europe. However, these results are more likely to reflect the differences in the timing of and the responding to the outbreak, *etc*. between European countries, rather than the 'effect' of the distance to Brussels itself. Associations observed at the group level have limitations to reflect individual-level associations – the so-called ecological fallacy. Given the public concern over COVID-19, ecological studies should be conducted and interpreted with great cautions, in case the results would be mistakenly understood.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.23.20077008; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Introduction

As of 22 April 2020, the coronavirus disease 2019 (COVID-19) pandemic had been reported to cause more than 2.4 million cases and 169,006 deaths worldwide[1]. Many studies have been conducted to understand this 'unknown' new disease. Some studies have found that regional aggregated exposures (i.e., Bacillus Calmette-Guerin vaccination, air quality index) were correlated with regional aggregated COVID-19 outcomes (i.e., numbers of cases and deaths)[2-4]. Such type of studies using group-level data, rather than individual-level data, is typically called ecological studies[5]. While some ecological studies might be helpful in generating new hypotheses related to COVID-19, their results should be interpreted with cautions, as ecological studies by nature are sometimes more vulnerable to bias than studies using individual-level data.

To illustrate how ecological studies and relevant results could be biased, we presented an ecological study of investigating the relationships between COVID-19 outcomes and the distance to Brussels, where the Europe Union (EU) headquarters located, in Europe.

Methods

We obtained daily numbers of COVID-19 cases and deaths for European countries from the European Centre for Disease Prevention and Control (https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide). To minimise the results being impacted by small numbers of cases, we included countries with at least 2,000 cumulative cases (up to and including 22 April 2020) only. A total of 26 countries were included (Figure 1). We studied three COVID-19 outcomes: confirmed number of cases per million people, confirmed number of deaths per million people, and case fatality rate defined as the proportion of deaths in confirmed cases. For each country, the distance to Brussels was defined as the direct distance (as the crow flies) from its capital to Brussels, and was measured using an online tool (https://www.freemaptools.com/how-far-is-it-between.htm).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.23.20077008; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

We investigated the correlations between the distance and each of the three outcomes (log-transformed) using Pearson correlation coefficient. We estimated the 'effect' of the distance on COVID-19 outcomes using a log-linear regression model.

Results

All the three COVID-19 outcomes were negatively correlated with the distance to Brussels (Figure 1); the correlation coefficient ranged from -0.5 to -0.7, and all the P-values were less than 0.01.

From the log-linear regression, every 100km away from Brussels was associated with 10.3% (95% confidence interval [CI], 5.4% to 15.1%) reduction in confirmed COVID-19 cases per million people, with 16.7% (95% CI, 10.0% to 23.3%) reduction in confirmed COVID-19 deaths per million people, and 6.4% (95% CI, 1.8% to 11.0%) reduction in COVID-19 case fatality rate.

Discussion

From our ecological study, we found that for European countries their distance to Brussels was negatively correlated with their COVID-19 outcomes, and all the correlations were statistically significant. If no cautions were given to our study, such results could be interpreted as the closer to the EU headquarters, the higher risk of COVID-19 in Europe – a conclusion with limited plausibility.

Ecological bias, or ecological fallacy, is the major limitation of ecological studies in making causal inference. This bias is usually interpreted as the failure of the observed association at the group level to reflect the biological effect at the individual level[5]. While a strong individual-level effect of an exposure on an outcome could result an effect at the group level, i.e., a population with more smokers tend to have more lung cancers compared with a (comparable) population with fewer smokers, the reverse is not always held.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.23.20077008; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

One source for ecological bias is that the group-level measures do not necessarily reflect the measures at the individual level, as the latter are not measured; after all, health conditions occur at the individual level.

Even the purpose of an ecological study is to make inference at the group level, rather than at the individual level, cautions are still needed. Confounders between groups and effect modification by group could also introduce biased ecological associations. In terms of our example, there are considerable differences in the timing of the outbreak, responding to the outbreak, test capacity, healthcare system, population structure, *etc.* between European countries that could confound the association at the country level. Our observed ecological association is likely to reflect the impacts of these differences on the different COVID-19 outcomes between the studied European countries, rather than the 'effect' of the distance to Brussels itself.

Nevertheless, ecological studies indeed have some values, especially in generating hypotheses. Some discoveries about the causes of cancer could be attributed to the hypotheses generated by internationally comparing cancer incidences[6]. The current ecological studies of COVID-19 might inspire some further studies to research in depth to provide more evidence for supporting, or falsifying, relevant hypotheses.

Understanding the differences in health conditions between populations is the driving force of the development of epidemiology; however, modern epidemiology tends to more focus on the individual level. Comparing group-level data could bring the public-health orientation back to epidemiology, but simply conducting ecological studies would have little help[5,6].

Given the public concern over COVID-19, any 'novel/striking finding' from ecological studies would have the potential of becoming eye-catching headlines and attract considerable media coverage, and be easily interpreted problematically. Researchers should take cautions when conducting ecological studies and discussing relevant results.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.23.20077008; this version posted April 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Conflict of interest

The authors declare no conflict of interest.

Author's contributions

SL conceived the study and analysed the data. SL and XH wrote the manuscript.

References

1. Coronavirus disease 2019 (COVID-19) Situation Report – 93. (Accessed 23 Apr, 2020, at https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200422-sitrep-93-covid-19.pdf?sfvrsn=35cf80d7_4.)

2. Miller A, Reandelar MJ, Fasciglione K, Roumenova V, Li Y, Otazu GH. Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: an epidemiological study. medRxiv 2020:2020.03.24.20042937.

3. Wu X, Nethery RC, Sabath BM, Braun D, Dominici F. Exposure to air pollution and COVID-19 mortality in the United States. medRxiv 2020:2020.04.05.20054502.

4. Travaglio M, Popovic R, Yu Y, Leal N, Martins LM. Links between air pollution and COVID-19 in England. medRxiv 2020:2020.04.16.20067405.

5. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd. Philadelphia: Lippincott Williams & Wilkins; 2008.

6. Pearce N. The ecological fallacy strikes back. Journal of Epidemiology and Community Health 2000;54:326-7.

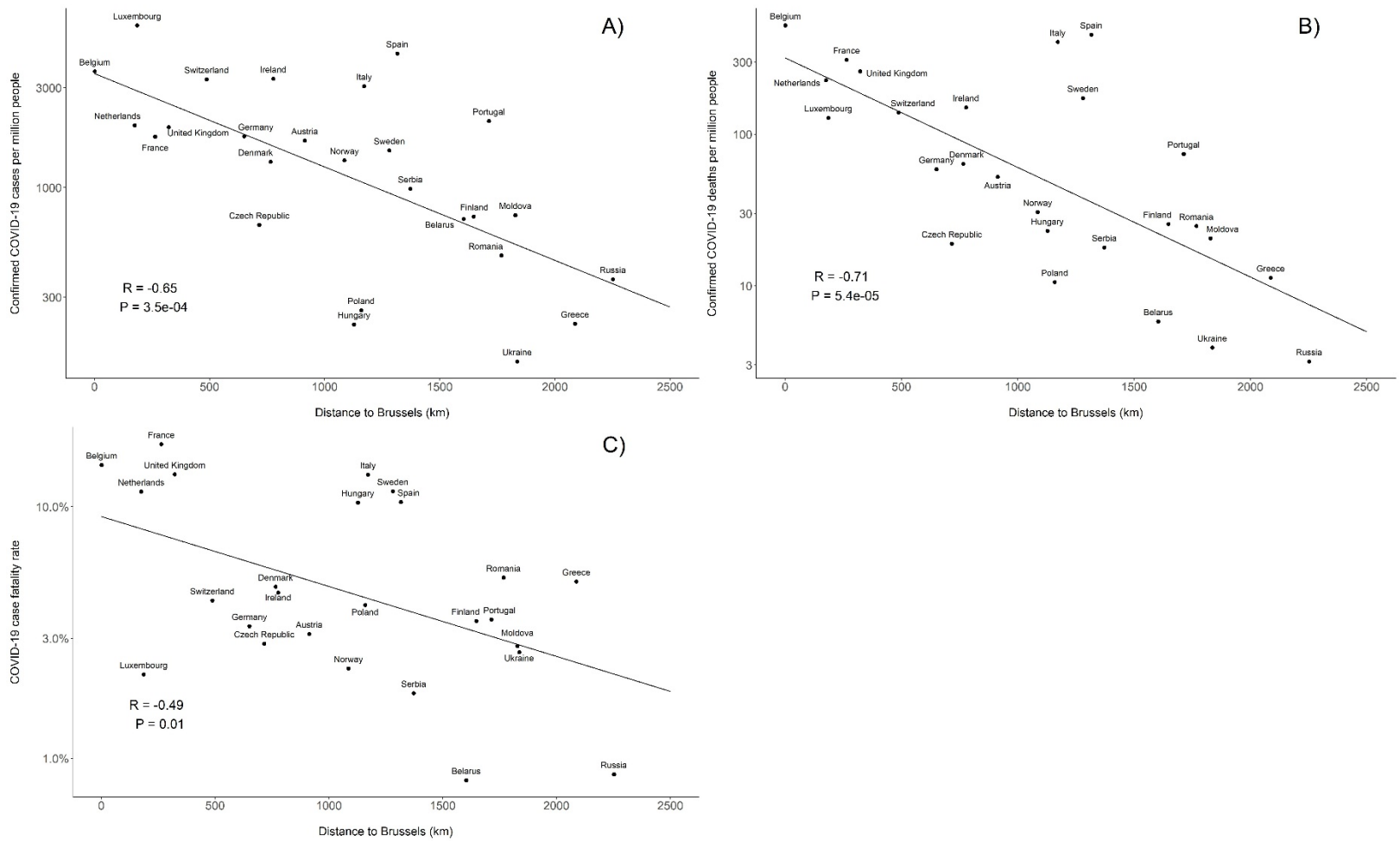

Figure 1 Correlations between COVID-19 outcomes and the distance to Brussels for European countries

The y-axis is in log scale. R is the Pearson correlation coefficient and P is the corresponding P-value. A) Correlation for the number of confirmed COVID-19 cases per million people; B) Correlation for the number of confirmed COVID-19 deaths per million people; C) Correlation for the COVID-19 case fatality rate