CEG
Centro de Estudos Geográficos

THE IMPORTANCE OF SPATIAL ANALYSIS OF COVID-19 PANDEMIC FOR HEALTH GEOGRAPHY:
CHALLENGES AND PERSPECTIVES

Ana Isabel Ribeiro[1]

Cláudia Jardim Santos[2]

ABSTRACT – The COVID-19 pandemic brought an unparalleled opportunity for spatial analysis. More than ever people are creating maps to document the space-time diffusion of the COVID-19 pandemic. However, despite recent technical and computational improvements, spatial analysis of epidemiological surveillance data is still affected by a number of challenges, especially when dealing with near-real-time data of an emergent disease. This paper summarizes the key challenges for the spatial analysis of the COVID-19 pandemic and possible solutions.

Keywords: Spatial models; SARS-CoV-2; public health; health geography.

RESUMO – IMPORTÂNCIA DA ANÁLISE ESPACIAL DA PANDEMIA DE COVID-19 PARA A GEOGRAFIA DA SAÚDE: DESAFIOS E PERSPETIVAS. A pandemia de COVID-19 trouxe uma oportunidade incomparável para a utilização da análise espacial. Mais do que nunca, as pessoas estão a criar mapas para documentar a difusão espaço-temporal da pandemia de COVID-19. Contudo, apesar dos progressos técnicos e computacionais, a análise espacial de dados de vigilância epidemiológica continua a ser afetada por vários desafios, principalmente quando se trata de lidar com dados em quase tempo real de uma doença emergente. Este artigo resume os principais desafios para a análise espacial da pandemia de COVID-19 e possíveis soluções.

Palavras-chave: Modelação espacial; SARS-CoV-2; saúde pública; geografia da saúde.

[1] Investigadora e Docente, Unidade de Investigação em Epidemiologia, Instituto de Saúde Pública (EPIUnit), Universidade do Porto, Departamento de Ciências da Saúde Pública e Forenses e Educação Médica, Faculdade de Medicina, Universidade do Porto, Alameda Prof. Hernâni Monteiro, 4200-319, Porto, Portugal. E-mail: ana.isabel.ribeiro@ispup.up.pt

[2] Investigadora, Unidade de Investigação em Epidemiologia, Instituto de Saúde Pública (EPIUnit), Universidade do Porto, Porto, Portugal. E-mail: claudiasantos@ispup.up.pt

I.   INTRODUCTION

The coronavirus disease (COVID-19) pandemic brought an unparalleled opportunity for spatial analysis. More than ever people are creating maps to document the space-time diffusion of COVID-19 pandemic.

Since the 17th century, disease mapping has been considered as a vital tool in tracking and combating disease diffusion. When computerised geographic information systems were born, the possibilities for analysing, visualising and detecting patterns of disease dramatically increased (Kamel Boulos & Geraghty, 2020). And, now, we have seen a revolution in health geography through Web-based tools, which further expanded these technical capacities (Kamel Boulos & Geraghty, 2020).

However, despite the recent technical advances in Geographic Information Systems (GIS) and spatial statistics, spatial analysis of epidemiological surveillance data is affected by a number of challenges, especially when dealing with near-real-time data of an emergent disease. This paper aims to summarize the key challenges for spatial analysis of COVID-19 pandemic and to discuss possible solutions, based on the evidence and practices available in the first months of COVID-19 pandemic (from December 2019 until June 2020). Note that, although these challenges are interconnected, for convenience reasons, this paper treats them as separate issues.

II.   CHALLENGES AND PERSPECTIVES OF SPATIAL ANALYSIS: THE CASE OF COVID-19 PANDEMIC

Challenge nº 1: Protection of geoprivacy

Safeguarding patient privacy while preserving the spatial resolution required for spatial analysis and cluster detection is a major challenge in disease mapping. Location and health data from patients are considered identifiable and personal information and, therefore, are subject to the General Data Protection Regulation 2016/679 (GDPR), a legislation that aims to provide control to individuals over their personal information. Similarly, mobile phone geolocation data – increasingly used to track human movement and access the efficacy of lockdown measures – is under the same rules and poses even more data protection questions.

In response, a variety of methods have been proposed to mask patients' geolocation (Chen *et al.*, 2017). For example, the most common geomasking method is to spatially aggregate data, as done by the Directorate-General of Health (Direção-Geral da Saúde, DGS), which therefore only discloses data at the municipal level. Moreover, the DGS applies an additional restriction by omitting municipalities with less than three cases, which makes spatial analysis challenging and introduces a new layer of uncertainty. While disaggregating data is nearly impossible, dealing with the second problem could be achieved by using, for instance, imputation methods for masked count data.

Challenge nº 2: Low spatial resolution and high geographical uncertainty

In Portugal, so far, individual-level data on COVID-19 cases are publicly disclosed by sub-regions (Nomenclature of territorial units for statistics – NUT, 3) and aggregated data (counts) is accessible by municipality. Using NUT 3, and even municipalities, as spatial units of analysis may conceal local outbreaks and important socioeconomic and biophysical variation.

Therefore, COVID-19 spatial analysis derived from aggregated data may be affected by the Modifiable Areal Unit Problem (MAUP) (Openshaw & Taylor, 1979), which happens when the number of spatial units (the scale) used to define the same area affects the study conclusions, namely geographical patterns and the magnitude of the associations. If the geographical units are large, is more likely that associations found at the aggregate level will diverge from the same associations found at individual level leading to the so-called ecological fallacy (Aikins & Ribeiro, 2020). Choosing the ideal spatial resolution for a particular investigation is difficult. Thus, analysing the same data using multiple geographical scales is a way of assessing the potential impact of MAUP.

A second issue is the Uncertain Geographic Context Problem (UGCoP). Case data is available according to the patients' municipality of occurrence, but focusing only on occurrence location can introduce substantial uncertainty in research results because people may spend a considerable amount of time in other municipalities and may acquire the disease in these locations (e.g. work, transportation, etc.) (Ribeiro, 2018).

Finally, a common problem in spatial analyses of rare events (e.g. diseases) is the well-known Problem of Small Numbers that is related to statistical instability when calculating rates in areas with low population and few cases, leading to random fluctuation and unreliable rates (Pina, Alves, Ribeiro, & Olhero, 2010). Spatial smoothing methods and the calculation of uncertainty intervals are widely used solutions for the problem.

Challenge nº 3: Lack of completeness and representativeness of patient and covariate data

Case data only includes confirmed cases. Confirmed case counts are not enough to comprehend the true magnitude of the COVID-19 pandemic. Although the true number of undetected cases is still to be ascertained, in Europe, the ratio of the total estimated cases to the observed cases was found to around 2.3 (Böhning, Rocchetti, Maruotti, & Holling, 2020). Compiling datasets that include suspected, probable, and negative test counts could substantially improve our understanding of COVID-19 space-time dynamics (Desjardins, Hohl, & Delmelle, 2020). Data on deaths may be subject to the same issues, as we may be missing deaths among persons infected with SARS-CoV-2, but who were not diagnosed with COVID-19.

For a better understanding of the space-time dynamics, data on population mobility and social networks are increasingly used. In the absence of universal, full-coverage datasets (at least in Portugal and other European nations), population mobility and social

networks are being tracked using anonymized phone location data. Yet, these datasets may be prone to data completeness and representativeness limitations too. Previous research demonstrated that mobile phone users and social media users are disproportionally distributed according to age, gender, and geography (Wesolowski, Eagle, Noor, Snow, & Buckee, 2012) .

Finally, the past couple of months we assisted to an exponential increase in web--based surveys to extract data on COVID-19 pandemic. Web-based voluntary recruitment introduces important selection bias, firstly by excluding people not on the internet and secondly by introducing self-selection bias. For instance, the overrepresentation of women and highly educated individuals are common problems in this type of study recruitment strategy (Rossi *et al.*, 2020). In these cases, weighting adjustments can reduce bias due to lack of population representativeness.

Challenge nº 4: Geographical comparisons may be affected by different sources of bias

Differences in the availability and practice of SARS-CoV-2 testing may contribute to spatial disparities in COVID-19 incidence across territories. In addition, if screening practices change through time, we can observe sudden incidence increases in certain regions simply due to increased screening. This means that estimates of incidence, case--fatality rates, and trends in incidence at country, regional and municipality level might not be directly comparable across jurisdictions.

For instance, as advanced by a recent study on COVID-19 spatiotemporal diffusion in the US (Desjardins *et al.*, 2020), the state of New York has a testing rate of 4.9 tests per 1000 population, three times higher than the national average. This high level of testing contributed to a better ascertainment of cases and it partially explains why this state presents a high cumulative incidence of COVID-19.

Similarly, differences in the numbers of deaths might reflect geographical inequalities in testing and disease coding practices, but they can reflect differences in population age--structure. Jurisdictions with older populations will necessarily have a higher number of deaths, which is neither novel nor unexpected. Thus, crude case-fatality and mortality rates cannot be directly compared. To avoid misleading conclusions, age-specific rates should be used instead, or one should calculate age-standardized rates, which denote the number of events that would have been expected if the jurisdictions being compared had similar age distribution.

An ecological analysis is defined as the assessment of the associations between disease incidence and variables of interest (e.g. social or environmental covariates) and it is usually the goal of many spatial analyses. Due to the presence of spatial autocorrelation (i.e. higher similarity of closer units) and due to the general lack of aggregated data on health determinants, these studies are particularly prone to bias. Spatial models that account for the spatial structure of the data are therefore required to correctly estimate the effects of these socioeconomic and environmental correlates (Pina *et al.*, 2010).

Challenge nº 5: True interdisciplinarity is still missing

Research around COVID-19 is not limited to the health and biological sciences, but it has attracted scientists from various fields including geographers. Nonetheless, very few research projects integrating the health (e.g. public health) and social and earth sciences (e.g. geography) have been conducted.

Interdisciplinary work is widely recognized as a breeding ground for innovation and it is, possibly, the only way of understanding complex problems, such as the COVID-19 pandemic. Health researchers, who hold vast knowledge on the biological mechanisms of disease transmission and health surveillance systems, need to team up with geographers (and other scientists) who are widely known for bridging social sciences and natural sciences, for their proficiency in GIS that could be used to track contagion, and for their understanding of human-environment relationships. But the same applies to geographers, who should get familiar with medical and biological terms, epidemiological research methods, causality frameworks, epidemic models, and so on.

As *two heads are better than one, not because either is infallible, but because they are unlikely to go wrong in the same direction* (C. S. Lewis, 1898-1963), the current COVID-19 should be used to leverage true interdisciplinarity research.

III. CONCLUSION

Spatial analysis tools help monitor and manage public health. When conducting these analyses, it is crucial to ensure patients' privacy by masking their geolocation while also choosing the ideal spatial resolution to reduce stigma and geographic uncertainty. In the beginning of an outbreak or when dealing with rare cases of a disease, spatial smoothing methods and the calculation of uncertainty intervals should be considered to avoid random fluctuation and unreliable rates. Health geography can only mirror reality when the quality of the data is assured, therefore, the completeness and representativeness of data are essential to understand the space-time dynamics of a disease. Geographical comparisons can be useful to assess the evolution of the disease between and within areas; however, these comparisons can only be viable if possible sources of bias are taken into account. Thus, on the verge of a public health crisis the collaboration of biomedical, social, and natural sciences experts is essential for making faster and careful informed decisions.

ACKNOWLEDGMENTS

REFERENCES

Aikins, E., & Ribeiro, A. I. (2020). *Elements of Health and Medical Geography.* Dubuque: Kendall Hunt Publishing Company.

Böhning, D., Rocchetti, I., Maruotti, A., & Holling, H. (2020). Estimating the undetected infections in the Covid-19 outbreak by harnessing capture-recapture methods. *International Journal of Infectious Diseases, 97*, 197-201. Doi: 10.1016/j.ijid.2020.06.009

Chen, C. C., Chuang, J. H., Wang, D. W., Wang, C. M., Lin, B. C., & Chan, T. C. (2017). Balancing geo-privacy and spatial patterns in epidemiological studies. *Geospatial Health, 12*(2), 294-299. Doi: 10.4081/gh.2017.573

Desjardins, M. R., Hohl, A., & Delmelle, E. M. (2020). Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Applied Geography, 118*, 102202. Doi: 10.1016/j.apgeog.2020.102202

Kamel Boulos, M. N., & Geraghty, E. M. (2020). Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. *International Journal of Health Geographics, 19*, 8. Doi: 10.1186/s12942-020-00202-8

Openshaw, S., & Taylor, P. J. (1979). A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem. In N. Wrigley (Ed.), *Statistical applications in the spatial sciences* (pp. 127-144). London: Pion.

Pina, M. F., Alves, S., Ribeiro, A. I., & Olhero, A. (2010). Spatial Epidemiology: New Approaches to Old Questions. *Universitas Odontológica, 29*(63), 47-65.

Ribeiro, A. I. (2018). Public health: why study neighborhoods?. *Porto biomedical journal, 3*(1), e16-e16. Doi: 10.1016/j.pbj.0000000000000016

Rossi, R., Socci, V., Talevi, D., Mensi, S., Niolu, C., Pacitti, F., (…), & Di Lorenzo, G. (2020). COVID-19 pandemic and lockdown measures impact on mental health among the general population in Italy. An N=18147 web-based survey. *medRxiv*, pre-print, 1-12. Doi: 10.1101/2020.04.09.20057802

Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W., & Buckee, C. O. (2012). Heterogeneous mobile phone ownership and usage patterns in Kenya. *PLoS One, 7*(4), e35319. Doi: 10.1371/journal.pone.0035319