# Spatial data analysis and the use of maps in scientific health articles

Luciana Bertoldi Nucci[1]*, Patrick Theodore Souccar[2], Silvia Diez Castilho[3]

[1]PhD in Epidemiology – Professor at Health Science Postgraduate Program, Pontifícia Universidade Católica de Campinas (PUC-Campinas), Campinas, SP, Brazil
[2]Medical degree from PUC-Campinas. Medical Resident at Hospital Sírio-Libanês, São Paulo, SP, Brazil
[3]PhD in Child and Adolescent Health – Professor at Medical School, PUC-Campinas, Campinas, SP, Brazil

## Summary

**Introduction:** Despite the growing number of studies with a characteristic element of spatial analysis, the application of the techniques is not always clear and its continuity in epidemiological studies requires careful evaluation.

**Objective:** To verify the spread and use of those processes in national and international scientific papers.

**Method:** An assessment was made of periodicals according to the impact index. Among 8,281 journals surveyed, four national and four international were selected, of which 1,274 articles were analyzed regarding the presence or absence of spatial analysis techniques.

**Results:** Just over 10% of articles published in 2011 in high impact journals, both national and international, showed some element of geographical location.

**Conclusion:** Although these percentages vary greatly from one journal to another, denoting different publication profiles, we consider this percentage as an indication that location variables have become an important factor in studies of health.

**Keywords:** epidemiological studies, spatial analysis, geographic mapping, statistics data interpretation.

## Introduction

The diversity of statistical analysis techniques applied to epidemiological data in scientific articles is growing, confirming the need for interaction between different areas of knowledge, such as statistics, geography, and the health sciences. Increasingly, analyses already known in the exact sciences are being applied to epidemiological data. As an example, Generalized Linear Models with their specificities according to the type of distribution of data include logistic regression analysis, that is one of the techniques applied the most to epidemiological data in recent decades.[1,2]

The interaction between statistics and health sciences dates from the first half of the 20th century, with the institutionalization process of epidemiology as a scientific discipline, continuing with the creation of professorships at the main schools of public health such as the London School of Hygiene and Johns Hopkins, resulting in the incorporation of techniques and principles from Biostatistics into the study of epidemiological problems.[1]

The first records about the relationship between disease and the environment have been attributed to Hippocrates (480 BC),[3] while spatial data analysis was introduced by John Snow in the middle of the 19th century, through studies on the cholera epidemic in London, showing the relationship between infection and its spatial distribution which, in turn, led to the discovery of the association between the disease and the city's water supply.[4] The major change that has developed since then was that epidemiology no longer uses merely geographical elements, but went on to establish a dialogue with this and other fields of knowledge.[5,6]

Similarly, the interaction between Epidemiology, Geography and Statistics became imperative. Although such analyses do not intend to establish causal associations at the individual level, they can become a valuable tool in evaluating the impact of social processes and structures in the determination of health events.[7] Finding out about the spatial structure and dynamics enables the characterization of the place and conditions in which

these health events occur, in addition to tracking the spread of morbidities, and identifying priority sites of intervention and the result after these interventions.[6] Geoprocessing of environmental and health information also enables the identification of variables that reveal the social, economic and environmental structure, where the health risks are present, although not determining their cause, which is extremely useful for understanding patterns of diseases.[3,8]

Despite the growing number of studies with characteristic elements of spatial analysis, the application of analytical techniques suitable for epidemiological studies requires careful consideration. Therefore, the objective of this study is to verify the dissemination and use of these techniques in scientific articles published in national and international journals with significant impact in the area of health.

## METHOD

An assessment was made in accordance with the impact index of the 8,281 periodicals found in the database of the Web of Knowledge in August 2012, according to the JCR (Journal Citation Report – the impact factor of Thomson Reuters ISI – Institute for Scientific Information) and Qualis Criterion – CAPES (Higher Education Personnel Improvement Coordination).

Among these 8,000+ journals, 20 titles with the highest impact indexes were eligible for review, and a further 20 among the 93 selected when restricting the search criteria to Brazil. Four national and four international publications were chosen out of these 40 (20 national and 20 international). The selection of these eight journals was carried out at random by the authors with the aim of including general approaches from all areas of health-related knowledge (Chart 1).

Articles related to investigations or research (research articles) from all the numbers of the year of 2011 for each of the periodicals in question were assessed for the presence or absence of maps and spatial analysis techniques. Letter type articles, editorials, book reviews, and others not characterized as scientific research or investigations were excluded.

Once the articles were selected, they were divided into categories[9] in ascending order of complexity:

1. Illustrative map – used to describe the location of the sample or population studied. This category also included those that are the result of the reproduction of other works;
2. Thematic map – displaying the research results in maps. This depends on the calculation of indicators made with the data collected, followed by georeferencing. Concepts of thematic cartography are used in this case, but also simple geoprocessing software;
3. Analytical map – this group includes those in which the space is one of the central objectives of the work. It is worth noting that it requires formulation of hypotheses, knowledge of the area and the etiology of health event studied, mastery of statistical techniques and GIS (Geographic Information Systems) and the use of specialized software.

The statistical analysis was carried out for the comparison of proportions using the chi-squared test with a sig-

**CHART 1** Periodicals selected among those with the highest impact factors according to the Journal Citation Report (JCR), Institute for Scientific Information (ISI) and CAPES Qualis Criterion and their respective quality indexes. Study conducted in August 2012.

| Classification according to the JCR | Periodical | Impact factor (ISI) | Impact factor in the last 5 years | Qualis/Capes criteria |
|---|---|---|---|---|
| 2 | *The New England Journal of Medicine*[10] | 53.298 | 50.075 | A1 |
| 9 | *Nature*[11] | 36.280 | 36.235 | A1 |
| 15 | *The Lancet*[12] | 38.278 | 33.797 | A1 |
| 18 | *Science*[13] | 31.201 | 32.452 | A1 |
| 1 | *Memórias do Instituto Oswaldo Cruz*[14] | 2.147 | 1.975 | B1*/B1** |
| 4 | *Revista de Saúde Pública*[15] | 1.328 | 1.436 | B2*/A2** |
| 7 | *Anais da Academia Brasileira de Ciências*[16] | 1.094 | 1.208 | B2*/B2** |
| 14 | *Revista da Sociedade Brasileira de Medicina Tropical*[17] | 0.681 | 0.889 | B3*/B2** |

* Medicine 1; **Public Health.

nificance level of 5%. The open source software adopted was Mendeley[18,19] for organization of bibliographic references, and the OpenEpi website for the statistical tests.[20]

## RESULTS

From all of the articles published in 2011 in the eight periodicals selected, we identified those that had maps, since this is the characteristic element of spatial data analysis, and they were subsequently stored in a database of bibliographic references. We obtained 127 articles from a total of 1,274 published in the period of the journals in question (Table 1).

**TABLE 1** Total number and percentage of articles with a characteristic element of spatial analysis (map), published in 2011 in the periodicals selected among those with higher impact factors, according to the Journal Citation Report, Institute for Scientific Information (ISI) and CAPES Qualis Criterion. Study conducted in August 2012.

| Periodical | Number of articles in 2011 | |
| --- | --- | --- |
| | Total | With map |
| *The New England Journal of Medicine* | 249 | 2.4% (6) |
| *Nature* | 152 | 3.9% (6) |
| *The Lancet* | 273 | 9.2% (25) |
| *Science* | 73 | 2.7% (2) |
| **Total, International** | **747** | **5.2% (39)** |
| *Memórias do Instituto Oswaldo Cruz* | 169 | 9.5% (16) |
| *Revista de Saúde Pública* | 120 | 6.7% (8) |
| *Anais da Academia Brasileira de Ciências* | 112 | 38.4% (43) |
| *Revista da Sociedade Brasileira de Medicina Tropical* | 126 | 16.7% (21) |
| **Total, National** | **527** | **16.7% (88)** |
| **Total** | **1,274** | **10.0% (127)** |

In general, the international journals presented different percentages for the use of spatial analysis, with proportionally higher use in *The Lancet* (9.2%), followed by *Nature* (3.9%), *Science* (2.7%), and *The New England Journal of Medicine* (2.4%) with the smallest numbers.

The national journals present a lower total number of articles, due to the small number of editions during the year compared with international journals. There was, however, a high percentage of use of geographical location in the articles. The largest proportion occurred in the *Anais da Academia Brasileira de Ciências* (38.4%), followed by the *Revista da Sociedade Brasileira de Medicina Tropical* (16.7%), the *Memórias do Instituto Oswaldo Cruz* (10.6%), and the *Revista de Saúde Pública* (6.7%) appearing in last place.

The maps identified in the international journals were mostly global (43.6%), followed by India (10.3%), and the United States (7.7%). Brazil appears in fourth place with 2 of the 39 articles found (5.1%) in international journals.

Regarding the maps identified in national journals, the majority (76%) represented the Brazilian territory, and the others referred to countries from various continents, among which there was a greater expression of Latin America, with Argentina and Colombia accounting for 4.5% each, followed by Bolivia, French Guiana, Panama and Mexico with 1.1% each. In addition, around 90% of the national articles that presented maps from Brazil studied a specific region of the country, with a clear predominance of the Southeast (36%), half of which are concentrated in the state of São Paulo, followed by the Northeast (24%), North (22%), South (10%) and Midwest (8%) in last place.

Fifty nine of the 127 articles selected had illustrative type maps (46.5%), 58 articles had thematic maps (45.7%), and 10 were classified as analytical maps (7.9%). The 39 international articles had a predominance of the thematic type (61.5%), followed by illustrative (30.8%), and analytical in third place (7.9%). In the case of national periodicals, the largest proportion was illustrative maps (53.4%), with thematic maps in second place (38.6%) and analytical (8.0%) in last place. However, a difference was noted in categories of the maps presented in national and international journals, with the majority of the illustrative maps in national journals and the majority of the articles with thematic maps in the international journals (p=0.048) (Table 2).

**TABLE 2** Level of complexity of the characteristic elements of spatial analysis (maps) of the 127 articles published in 2011 in the journals selected from those with the highest impact factors, according to the Journal Citation Report, Institute for Scientific Information (ISI) and CAPES Qualis Criterion. Study conducted in August 2012.

| | Illustrative % (n) | Thematic % (n) | Analytical % (n) |
| --- | --- | --- | --- |
| International | 30.8% (12) | 61.5% (24) | 7.7% (3) |
| National | 53.4% (47) | 38.6% (34) | 8.0% (7) |
| **Total** | **46.5% (59)** | **45.7% (58)** | **7.9% (10)** |

Chi-squared test, p=0.048.

## DISCUSSION

The frequent changes undergone in the contemporary world, in various areas such as the economic, political and cultural spheres, create a need to develop and use new methodological, theoretical and conceptual tools in order to understand it. In this process, distinct sciences

are implemented to visualize knowledge in a more comprehensive manner through the convergence of concepts to respond to new necessities.

With regard to space, there are many ways to understand and study it, including maps. They are very practical tools for this purpose, since they enable the representation of multiple characteristics of a particular location. However, certain techniques must be observed to do so adequately and clearly. Such information can be very important if, for example, it allows greater insight into the manifestation of health and disease in a particular region, their coverage, and integration between data about the environment, health and society, as well as planning and management of services.

Geoprocessing is known as a group of techniques for collecting, handling and presenting information on a determined space. Among these techniques, it is worth highlighting: remote mapping, data digitization, automation of mapping work, the Global Positioning System (GPS) and Geographic Information Systems (GIS).[6,21-23] In recent decades, with increased access to computers, the storage and processing of data has been streamlined, facilitating the dissemination and realization of works with spatial reference or georeferenced information. Even so, in the case of Brazil, only municipalities that have invested heavily in the construction of GISs have managed to develop them in a satisfactory and detailed manner. The alternative strategy that also enables satisfactory results is to use, for example, socioeconomic and health references obtained from censuses and surveys of the Brazilian Institute of Geography and Statistics (IBGE) and Datasus.[24]

In the area of prediction and prevention, geoprocessing is a useful tool for planning intervention measures with populations exposed to risks, in order to improve actions, teams and costs in the tracking of cases, such as dengue fever, for example. In this case, planning leads to more streamlined and specific actions for vector control, according to the locations, severity, prevalence and seasonal variation.[8] However, the use of this tool is not only limited to infectious diseases, or the identification and control of vectors as it was at the beginning of its development. Other examples would be the identification of toxic substances capable of generating health hazards to the population in a determined space, or verification of occupational diseases in a geographic area.[25]

In our results we found a predominance of maps representing Brazil (76%) in national journals, with no incidence in Anglo-Saxon American countries. In the international periodicals, the majority of the selected articles represented the global geographic division (43.6%), with India in second place (10.3%), surpassing the United States by a good margin, with the latter in third place with little more than 7.7% of images relating to its territory. Based on the locations of the maps, we can perceive, both in the national and international journals, that the highest incidence is associated with large research centers. With regards to the main focus of this study, which was to quantify and differentiate the use of elements of different types of geographical analysis, we selected 127 articles, corresponding to about 10% of the articles searched, with higher incidence in national journals (16.7%) than international ones (5.2%).

Another important piece of information that we managed to observe is the comparison between the national and international impact factors. While the index of the journal classified as the best in our search of international items was over 50 points, the highest among the national journals was just over 2 points, which shows that the average number of citations of published scientific articles in national journals is still low and do not reach the world average. It is worth noting that, according to Teixeira et al., the number of Brazilian medical journals cited in national publications is 11.74%, which is considered a good percentage.[26] The valuation of national research through citation in national journals could contribute to increasing these indexes of the impact rating of Brazilian journals.

Despite the results of the literature search indicating that just over 10% of the articles published in the year 2011 in high-impact journals, at both the national and international level, presented some kind of element related to geographic location, we found no quantitative data on the use of maps in scientific articles prior to our research. We consider the percentage obtained as an indication that location variables have become an important factor in health studies. However, note that the variation in these percentages from one periodical to another was high, denoting different publishing profiles, and it was possible to identify periodicals with a more epidemiological focus and others directed more toward genetics and/or new technologies in health.

With regard to the different degrees of complexity analyzed, a clear difference was noted between the journals, given that among national ones there is a predominance of the illustrative category, while thematic maps predominate in international periodicals. This disparity can be explained by the need to conceptualize Brazil's position or the country addressed in the article, as well as its regions. In addition, the low frequency of the analytical category (8%) is evident. One of the possible explana-

tions is that, in order to conceive the study, a greater knowledge of spatial analysis techniques is essential, as well as greater use of specific software and the use of predictive models explaining the health-disease process.

As for the elaboration of the maps, it was noted that many of them were developed solely using image-editing software, which could reflect, once again, the lack of interest in or even mastery of geoprocessing, cartography methods and use of spatial analysis software. Another fact that was evident and noted during the screening of the articles, is that the authors often prefer to use tables or descriptions to represent the results of their work, instead of using maps. In the same line of thought, it is important to stress the need for methodological articles for dissemination of spatial analysis techniques so that their incorporation into studies becomes more automated. This would facilitate and popularize their use, since in addition to being a useful tool in the presentation of research results, they could also be used in health management, enabling the evaluation of risk factors and demonstrating points in which certain resources can be better allocated. Furthermore, there was a lack of correlation between the cartographic elements and the text in various articles, given that in these cases the image was only arranged to illustrate something, without a proper explanation in the text about its production processes, or even a specific analysis of its findings. Therefore, the full potential of this tool was not used.

## Conclusion

We can see that the multidisciplinary integration process remains grueling. Although recent, the introduction of Geography, adding spatio-temporal components to epidemiological analyses, requires greater dissemination of its importance. The publication of theoretical studies about possible analysis and study methods with practical applications in the health area, and the inclusion of spatial analysis techniques, contributes greatly to the understanding of epidemiological processes of health-disease. It is believed, therefore, that the percentage of articles published in high-impact scientific journals that include spatial analysis of data will soon increase, especially those focused on epidemiology, reaching much higher levels, as seen with the introduction of statistics in the area of health.

## Resumo

Análise de dados espaciais e o uso de mapas em artigos científicos na área da saúde

**Introdução:** apesar do crescente número de estudos com elemento característico de análise espacial, a aplicação de técnicas de análise nem sempre é clara e sua continuidade em estudos epidemiológicos requer uma avaliação cuidadosa.

**Objetivo:** verificar a disseminação e utilização de técnicas de análise de dados espaciais em artigos científicos nacionais e internacionais.

**Método:** foi feita uma avaliação de periódicos de acordo com o índice de impacto. Dentre 8.281 periódicos pesquisados, foram selecionados quatro nacionais e quatro internacionais, dos quais foram analisados 1.274 artigos quanto à presença de técnicas de análise espacial.

**Resultados:** pouco mais de 10% dos artigos publicados no ano de 2011 em revistas de alto impacto, tanto nacionais quanto internacionais, apresentaram algum elemento de localização geográfica.

**Conclusão:** embora exista uma variação de percentuais de um periódico para outro, denotando perfis de publicação diferentes, consideramos esse percentual um indicativo de que variáveis de localização têm se tornado um fator importante nos estudos da área da saúde.

**Palavras-chave:** estudos epidemiológicos, análise espacial, mapeamento geográfico, interpretação estatística de dados.

## References

1. Barata RB. Epidemiologia e saber científico. Rev Bras Epidemiol. 1998; 1(1):14-27.
2. Barreto ML. Crescimento e tendência da produção científica em epidemiologia no Brasil. Rev Saúde Pública. 2006; 40(N Esp):79-85.
3. Musa GJ, Chiang PH, Sylk T, Bavley R, Keating W, Lakew B, et al. Use of GIS mapping as a public health tool-from cholera to cancer. Health Serv Insights. 2013; 6:111-6.
4. Snow J. Sobre a transmissão do cólera. 2.ed. São Paulo: Hucitec, 1999. 249 p.
5. Costa MCN, Teixeira MGLC. A concepção de "espaço" na investigação epidemiológica. Cad Saúde Pública. 1999; 15(2):271-9.
6. Hino P, Villa TCS, Sassaki CM, Nogueira JDA, dos Santos CB. Geoprocessing in health area. Rev Lat Am Enfermagem. 2006; 14(6):939-43.
7. Marshall RJ. A review of methods for the statistical analysis of spatial patterns of diseases. J R Statist Soc. 1991; 154(3):421-41.
8. Barcellos C, Bastos FI. [Are geoprocessing, environment, and health a possible combination?] Cad Saúde Pública. 1996; 12(3):389-97.
9. Rojas LI, Barcellos C, Peiter P. Utilização de mapas no campo da Epidemiologia no Brasil: Reflexões sobre trabalhos apresentados no IV Congresso Brasileiro de Epidemiologia. Inf Epidemiol SUS. 1999; 8(2):27-35.
10. The New England Journal of Medicine [Internet]. Waltham (MA): Massachusetts Medical Society. Vol. 364, No. 1, Jan. 2011 – Vol. 365, No. 26, Dez. 2011. Available from: http://www.nejm.org/.
11. Nature [Internet]. London (UK): Nature Publishing Group. Vol. 469, No. 7328, Jan. 2011 – Vol. 480, No. 7378, Dez. 2011. Available from: http://www.nature.com/.
12. The Lancet [Internet]. London (UK): Elsevier. Vol. 377, No. 9759, Jan. 2011 – Vol. 378, No. 9809, Dez. 2011. Available from: http://www.thelancet.com/journals/lancet/issue/current.
13. Science [Internet]. Washington DC: American Association for the Advancement of Science. Vol. 331, No. 6013, Jan. 2011 – Vol. 334, No. 6063, Dez. 2011. Available from: http://www.sciencemag.org/.

14. Memórias do Instituto Oswaldo Cruz [Internet]. Rio de Janeiro (RJ): Instituto Oswaldo Cruz. Vol. 106, No. 1, Fev. 2011 – Vol. 106, No. 8, Dez. 2011. Available from: http://memorias.ioc.fiocruz.br/.

15. Revista de Saúde Pública [Internet]. São Paulo (SP): Faculdade de Saúde Pública da Universidade de São Paulo. Vol. 45, No. 1, Fev. 2011 – Vol. 45, No. 6, Dez. 2011. Available from: http://www.rsp.fsp.usp.br/.

16. Anais da Academia Brasileira de Ciências [Internet]. Rio de Janeiro (RJ): Academia Brasileira de Ciências. Vol. 83, No. 1, Mar. 2011 – Vol. 83, No. 4, Dez. 2011. Available from: http://www.abc.org.br/rubrique.php3?id_rubrique=52.

17. Revista da Sociedade Brasileira de Medicina Tropical [Internet]. Uberaba (MG): Sociedade Brasileira de Medicina Tropical. Vol. 44, No. 1, Jan./Fev. 2011 – Vol. 44, No. 6, Nov./Dez. 2011. Available from: http://www.sbmt.org.br/site/revistas.

18. Mendeley Desktop [Internet]. Version 1.10.3. Mendeley Ltd. 2008-2014. Available from: http://www.mendeley.com/.

19. The Mendeley Support Team. Getting started with Mendeley. Mendeley Desktop; 2011:1-16. Available from: http://www.mendeley.com.

20. Dean AG, Sullivan KM, Mir R. OpenEpi – Estatísticas epidemiológicas de código aberto para a Saúde Pública. OpenEpi Proj. 2014. Available from: http://www.openepi.com.

21. Kolifarhood G, Khorasani-Zavareh D, Salarilak S, Shoghli A, Khosravi N. Spatial and non-spatial determinants of successful tuberculosis treatment outcomes: an implication of Geographical Information Systems in health policy-making in a developing country. J Epidemiol Glob Health. 2015; 5(3):221-30.

22. López-Abente G, Aragonés N, García-Pérez J, Fernández-Navarro P. Disease mapping and spatio-temporal analysis: importance of expected-case computation criteria. Geospat Health. 2014; 9(1):27-35.

23. Curtis AB, Kothari C, Paul R, Connors E. Using GIS and secondary data to target diabetes-related public health efforts. Public Health Rep. 2013; 128(3):212-20.

24. Skaba DA, Carvalho MS, Barcellos C, Martins PC, Terron SL. [Geoprocessing of health data: treatment of information on addresses]. Cad Saúde Pública. 2004; 20(6):1753-6.

25. Zheng RZ, Zhang QH, He YX, Zhang Q, Yang LS, Zhang ZH, et al. Historical long-term exposure to pentachlorophenol causing risk of cancer – A community study. Asian Pac J Cancer Prev. 2013; 14(2):811-6.

26. Teixeira RKC, Botelho NM, Petroianu A. [References from Brazilian medical journals in national publications]. Rev Assoc Med Bras. 2013; 59(6):571-5.