

# VETERINARIA RIVISTA DI SANITÀ PUBBLICA VETERINARIA **ITALIANA**

**Special Issue GeoVet2023**



# Communication Breakdown – Of Disease Clusters, a Trillium and One Health

Olaf Berke<sup>1\*</sup>

<sup>1</sup>Department of Population Medicine, University of Guelph, 50 Stone Rd E, Guelph, Ontario, Canada N1G 2W1 - CA

\*Corresponding author at: Department of Population Medicine, University of Guelph, 50 Stone Rd E, Guelph, Ontario, Canada N1G 2W1 - CA

E-mail: oberke@uoguelph.ca

---

*Veterinaria Italiana*, Vol. 60 No. 4 (2024): Special Issue GeoVet23 DOI: 10.12834/VetItt.3483.27128.3

---

## Abstract

One health is based on an interdisciplinary collaboration across professions using a common language. Geographic epidemiology is the study of spatial patterns of population health in a study area. Such spatial patterns (trend, cluster and clustering) require clear definition to be meaningful in science communication. However, the term “disease cluster” has been defined in the literature in various and rather different ways. When geographic epidemiology is unable to make sense of its own concepts it is questionable how respective research results can benefit one health. The goal of this study was to clarify the disease cluster concept.

Examples of disease cluster definitions from the literature were used for illustration. The epidemiological triangle of causation (agent, host and environment) was used to conceptualize geographic epidemiological data analysis. The term disease cluster was distinguished from related concepts (clustering, high-risk area, hot spot and outbreak) additionally the semantics and statistical meaning of expectation and prediction were reviewed to further identify the cluster concept as a statistical outlier.

The new paradigm of the geographic epidemiological trillium is proposed here and embedded within the spatial generalized linear mixed model to clarify concepts of spatial patterns and guide epidemiological research and teaching.

## Keywords

cluster, communication, epidemiological triangle, geographic epidemiology, One Health

---

## Introduction

The COVID-19 pandemic highlighted the importance of proper scientific communication specifically in the public health arena. When terms and concepts, such as a disease cluster, are loosely defined and casually used any communication effort and health promotion campaign is unlikely to succeed.

Clear definition of terms and concepts is equally key for “one health.” As a concept, one health has a long history dating back at least 150 years to Rudolf Virchow, who promoted the one medicine concept by recognizing that some diseases affect humans and animals alike. Later, Calvin Schwabe added an environmental component to arrive at the one health concept, whereby the health of human and animal populations implies a shared healthy environment (Berke et al., 2022). Interdisciplinary collaboration to study and improve human, animal and environmental health is therefore a posit for the one health approach. Collaboration across disciplines requires a common language and careful use of terms and concepts.

Geographic epidemiology is an interdisciplinary field of study, which can support the one health approach. Geographic epidemiology is here distinguished from spatial epidemiology. Spatial epidemiology is the study of spatial patterns in health outcome data in relation to their location in a study area, whereas geographic epidemiology is an extension based on the notion of “place” rather than “location.” A location is defined by its coordinates. In contrast, place is a relational concept, which further describes a location by environmental determinants of the individual or population studied in that place (Cromley and McLafferty, 2012, p. 12). Thus geographic epidemiology is the study of population health with respect to location, distance between locations, neighbourhood structures as well as determinants of the chemical-, physical-, social-, economical-, political-, legal- and ecological environment, and therefore naturally

supports the one health approach. For example, Cortes-Ramirez et al. (2021) link the spatial risk patterns of rates of hospitalizations due to zoonoses (incl. leptospirosis and Q-fever) to social and environmental risk factors (incl. rainfall, social disadvantage, proportion of high-risk occupations, sex ratio).

Distinguished spatial patterns of disease occurrence are: trend, clustering and clusters (Berke and Waller, 2010). Waller and Gotway (2004, p. 1) remind of the importance of an identification of patterns in disease occurrence as a “first step towards increased understanding and possibly control of that particular disease.” Of the three spatial data patterns “disease cluster” appear to be of special importance. Haining and Maheswaran (2016) explain that disease cluster (real or apparent) may trigger “public anxiety and media interest” and thus must be addressed by public health in a timely fashion; and this implies typically high costs which further underlines the importance of cluster identification. John Snow’s detection of a cholera cluster around the Broadstreet water pump in London, 1854, is considered a pioneering and successful case study in public health (Berke, 2015). However, Kleinhues (1996, p. vii) claims, that it is not agreeable among experts how to define “what a disease cluster actually is”. Such a lack of common language prohibits effective health communication and ends in a communication breakdown.

The goal of this study was to clarify the disease cluster concept within geographic epidemiology. Further objectives were (i) using examples from the literature around cluster and clustering to demonstrate the ambiguity around these terms, (ii) embed the spatial data patterns within the theory of the epidemiological triangle, (iii) relate the disease cluster concept to the spatial generalized linear mixed model (GLMM), and (iv) propose a definition for a spatial disease cluster and distinguish cluster from outbreaks and hot spot concepts.

## Material and Methods

A few examples of disease cluster definitions from the literature were compiled and compared. This was not an exhaustive or even systematic review of proposed definitions, but presents exemplars from the literature to demonstrate the conceptual variations.

The epidemiological triangle of causation (agent, host and environment) was proposed to embed the spatial patterns of disease occurrence in a new epidemiological paradigm – the geographic epidemiological trillium – as a concept for geographic epidemiological data analysis. Key to such a data analysis is the spatial GLMM.

The trillium paradigm and spatial GLMM together provide a segue to further the understanding of spatial patterns of disease occurrence, and especially that of a disease cluster. The distinction between the spatial data patterns of trend, cluster and clustering required a closer look at the semantics and statistical meanings of the terms expectation and prediction.

Finally, the term disease cluster was distinguished from related concepts: high-risk area, hot spot and outbreak.

## Results

### Presentation of spatial patterns in the literature

Three spatial patterns of the geographic variation in health outcomes are distinguished: trend, clustering and clusters. The following presents some insight into how these patterns are conceptualized in the literature. First consider the “trend” pattern.

According to John Tukey (1977) data consists of two components:

$$\text{data} = \text{smooth} + \text{rough}$$

This can equally be expressed as “data = fit + residual,” where the “fit” is with respect to a statistical model and represents the expected value. In other words:

$$\text{data} = \text{trend} + \text{residual}.$$

The “trend” component is typically a model based on certain explanatory (or predictor) variables. Therefore the trend pattern in spatial data represents variability that can be explained (or predicted) by variation in explanatory variables.

Further distinguish between a “spatial” trend in any coordinate direction from a “geographic” trend, which might be directional but may also be related to any explanatory variable or predictor, e.g. land-use patterns or variations in

population density.

The trend pattern is less controversial as long as a model for the trend can be agreed upon. However, conceptualizations of the cluster and clustering patterns seem more ambiguous. Clarification regarding the distinction between cluster and clustering is often presented with reference to Besag and Newell (1991). These authors distinguish “cluster” as a local and “clustering” as a global pattern in spatial data. The importance is that these are distinct spatial patterns, which can be studied using “tests for clustering” and “tests for the detection of clusters,” respectively. Unfortunately, a reference to methods, i.e. tests, can not replace a definition for these concepts. Help comes in the form of the Dictionary of Epidemiology (Porta *et al.*, 2014), which defines a disease “cluster” as follows:

1) “Aggregations of relatively uncommon events or diseases in space and/or time in amounts that are believed or perceived to be greater than could be expected by chance.”

In other words, definition (1) implies that a cluster exists where more cases are observed than expected and that expectation is with reference to a certain model. Not all authors agree with this definition and thus provide their own (working) definitions, which are exemplified by the following quotes from frequently referenced textbooks as well as a review article:

2) “Cluster ... any area within the study region of significant elevated risk ... This definition is often referred to as hot spot clustering” (Lawson and Williams, 2001, p. 92)

3) “cluster ... an area with an unusual elevated disease incidence rate ... local clusters and global clustering” (Wheeler, 2007)

4) “we call a set of aggregated cases a cluster of cases... Clustering of cases... many small clusters of cases throughout the area.” (Tango, 2010, p.11/12)

5) “Cluster ... a collection of cases inconsistent with our null hypothesis of no clustering, whereas... clustering (is) the overall propensity of cases to cluster together” (Waller and Gotway, 2004, p.161)

To summarize, definitions (2) and (3) refer to a cluster as an area, whereas in (4) and (5) clusters are cases or sub-populations. Furthermore, these definitions seem not to follow the proposal by Besag and Newell (1991) to distinguish between clusters and clustering as distinct patterns. Sometimes it appears that cluster is used as a noun and clustering is the related verb, but both words are meant as reference to the same concept. This is most clearly presented in (4) where a single cluster is contrasted to clustering as a situation of multiple clusters. This must be confusing as (1) and (3) certainly attach conceptually different attributes to clusters as opposed to clustering: local versus global features. In (2) reference is made to a cluster as hot spot clustering, which implies that there are further definitions for clusters. Indeed, Andrew Lawson repeatedly reports (e.g. Lawson, 2018, p. 133) that three conflicting definitions or conceptualizations of a cluster exist: hot spot cluster, residual cluster and grouping cluster. And this had led Kleinhues previously to the conclusion, that “there is no universally agreed definition of exactly what a cluster is” (Kleinhues, 1996, p. vii). Similar, Wakefield *et al.* (2000, p. 128) state “much of the controversy surrounding ‘clusters’ stems from difficulty in giving a definition of a ‘cluster’.”

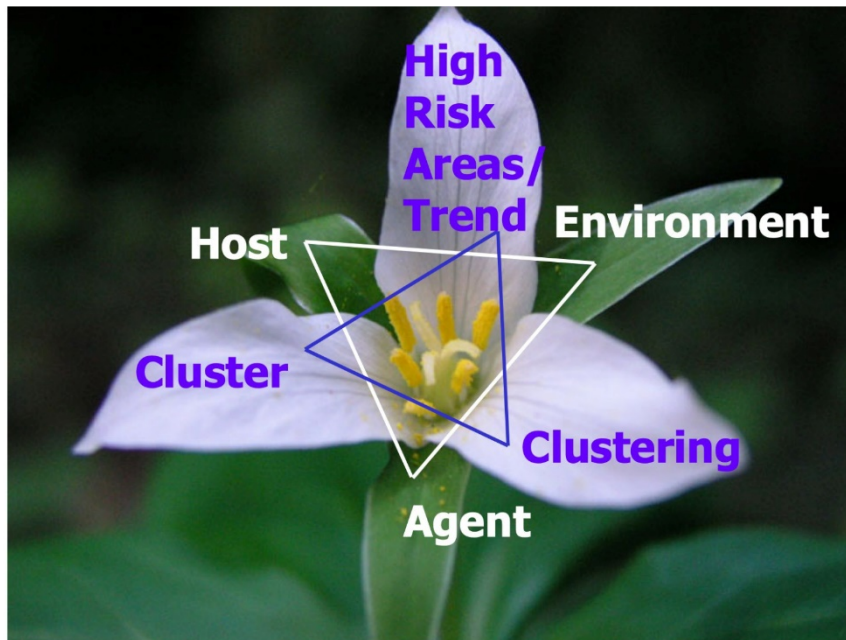
In summary, the above definitions of disease cluster, (2) to (5), are not coherent and therefore contribute to a communication breakdown. Furthermore, rather than being rooted in theory, cluster definitions in the literature often make reference to statistical methods (tests, expected values) to facilitate the statistical detection of clusters.

## Explaining spatial data patterns and the epidemiological triangle

Much epidemiological thinking and teaching is based on a theory, whereby health and disease in a population depend on the environment the population lives in (Bhopal, 2008, p. 128). One model for this relation is the epidemiological triangle, which connects an agent of disease with a host population and their environment (Bhopal, 2008, p. 130). The model applies to various health conditions including chronic or infectious diseases. The purpose of the triangle model is to analyze causal interactions to inform one health strategies.

Quotes from the literature can be used to associate spatial data patterns (trend, cluster and clustering) with the agent, host and environment corners of the epidemiological triangle as a theoretical foundation for geographic epidemiological research.

In figure 1 the geographic epidemiological trillium is presented to summarize and visualize the resulting relations between the spatial data patterns and the epidemiological triangle. The white trillium (*Trillium grandiflorum*) is the flower of Ontario and has long been used for medicinal purposes by native peoples. The three green sepals represent the corners of the epidemiological triangle of agent, host and environment; and the three white petals represent the corresponding spatial patterns. The connections are: (i) agent of disease with disease clustering, (ii) environment with trend and high-risk areas, and (iii) host population with disease cluster; as is described in the following subsections.



**Figure 1.** The geographic epidemiological trillium, representing with its 3 white petals and 3 green petals the link between the epidemiological triangle and the spatial patterns of disease occurrence.

### Cluster ~ host population

Following definitions (3) and (4) a cluster is a set of cases or better a sub-population of cases and surrounding non-cases of the study population. Thus the cluster pattern relates to the “host.” In other words, part of the host population can form a cluster and similarly several sub-populations can result in multiple clusters. Because clusters occur in certain areas they are local features of spatial data. However, a cluster is a sub-population but is not a place nor an area as suggested by definitions (2) and (3).

### Clustering ~ agent of disease

Next a connection between the “clustering” pattern and the “agent” of a disease was identified. This relation is evident from the following quotes:

6) “spatial clustering... is a description of the underlying disease process” (Diggle, 2003, p. 130)

7) “A disease exhibits spatial clustering if there is residual spatial variation in risk” (Wakefield et al., 2000, p. 128/129)

This aligns well with Besag and Newell (1991) who consider clustering a global (as opposed to local) feature of spatial data. Therefore clustering is a characteristic of the agent or the disease under study.

### Trend ~ environment

Finally, the “environment” component of the epidemiological triangle was studied and an association with the spatial “trend” pattern uncovered. A quote from the literature for this link is provided by Diggle (2003, p. 133):

8) “The function  $r(x)$  is called the risk surface. In contrast to spatial clustering, spatial variation in risk is a description of the study region.”

Considering “study region” a synonym for the environment in the epidemiological triangle reveals the link with the trend pattern. According to quote (7), clustering occurs due to residual spatial variation in risk, which implies the presence of a model fit or trend, which is typically the expected value of a regression model. Therefore the trend is predominantly expressed by explanatory and predictor variables describing the environment of places where population health data has been recorded. The trend function might also include information about the host and agent insofar these are known (e.g. vaccination rate) but such information is for places at respective population levels and thus mostly describe the environment and lesser the host or agent.

In summary, the trend pattern is typically a function of location and the environment, i.e. a function of space and place. Therefore the distinction between spatial and geographic epidemiology could also carryover to a distinction between a spatial trend function in spatial coordinates as opposed to high- and low-risk areas explained by environmental covariates.

## The geographic epidemiological trillium and spatial regression models

A theoretical concept for disease cluster is incomplete without a methodological approach for its analysis. The previously reviewed definitions from the literature make reference to tests, but a model-based approach for data analysis is more favourable. The spatial GLMM model allows to accommodate clusters and clustering via a combination of two random effect terms, which are independent of each other.

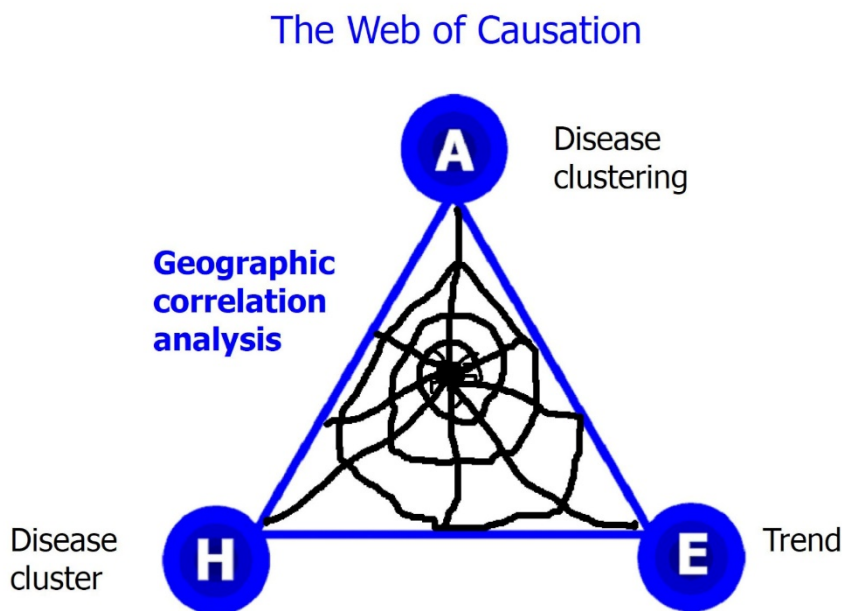
As an example of a spatial GLMM consider the Poisson rate model for observed counts  $Y_i$  of a rare disease among a population-at-risk of size  $N_i$  with intensity  $\lambda_i = N_i \theta_i$  in regions or places  $i = 1, \dots, k$ :

$$Y_i \sim \text{Poisson}(\lambda_i)$$
$$\log(\theta_i) = X_i \beta + v_i + u_i + \text{offset},$$

where:  $X_i \beta$  is a linear trend,  $v_i$  are spatially correlated random effects, and  $u_i$  are uncorrelated random effects representing extra Poisson variation, and  $\log(N_i)$  being the offset. The random effects are distributed independent of each other and with an expectation of zero. Briefly, the  $v_i$  model residual spatial dependence whereas the  $u_i$  model noise and sporadic outliers. Alternate models and modeling approaches depend on the spatial nature of the data, i.e. whether places are spatially continuous or contiguous leading to geostatistical and conditional autoregressive models. Furthermore, the models require a specification of the distribution of their parameters and respective estimation methods resulting in a variety of frequentist and Bayesian modeling approaches.

Nancy Krieger associated the causal web analysis in epidemiology with a spiders web (Krieger, 1994). This web is visualized in figure 2: A spider web connects the three corners of the epidemiological triangle, while the triangle is in balance, the disease status of the population is under control, i.e. endemic. If a disease gets out of control, i.e. become epidemic, studying population health via cluster detection, assessment of clustering and trend estimation might provide information about (geographic) patterns in disease occurrence and inform control initiatives at agent, host and population levels. Key to such analysis might be the spatial GLMM. However, regression modeling in geographic epidemiology (a.k.a. geographic correlation analysis) is an ecological study of observational data and thus not a definitive causal analysis.

Nancy Krieger used the spider web analogy as a connection to a popular book: "Charlotte's Web." In that story, the spider Charlotte weaves secret messages into its spider web to save a life. The interrelations between cluster, clustering and trend are represented by the web in figure 2 and may hold similar messages for epidemiologist, if only it is possible to discover the message from the observed data using a spatial GLMM.



**Figure 2.** The geographic epidemiological trillium combined with Charlotte's web of causation (after Krieger, 1994) as an attempt to visualize the relation between disease cluster, clustering and trend detection via geographic correlation analysis to uncover (causal?) relations among agent, host and environment characteristics.

## Of disease clusters and outliers

Using the notation of the spatial GLMM it is now possible to portray the reason for the confusion around the terms cluster, clustering and high-risk area; and to separate out a disease cluster as an outlier in spatial population health data. In paraphrasing definition (1) a disease cluster is defined as a case aggregation, which exceeds what can be expected by chance. Indeed, cases alone can not be identified rather these are part of their source populations in certain places. Thus a cluster is a sub-population of cases and non-cases in a place, where the case frequency exceeds the expected value under the assumption of equal risk.

According to the spatial GLMM presented above, the expected value is related to the trend  $X_i\beta$ . Now the question is, what is the meaning of the word “expectation”? If expectation refers to the expected value in a statistical sense, then a disease cluster is defined via an excess value of the combined random effects  $u_i + v_i$ . But that notion would muddle the cluster and clustering concepts, because clustering is already represented via the spatially correlated random effects  $v_i$ ; and according to the spatial GLMM and the trillium paradigm cluster and clustering are separate or independent features in spatial data. In other words, global clustering does not detect a local cluster. The definition of disease clusters could alternatively be based on the notion of “what could be predicted under a chosen model” rather than what could be expected. The predicted value in a statistical sense is based on the trend plus the correlated random effect:  $X_i\beta + v_i$ . This leaves the uncorrelated random effect  $u_i$  as the sole basis for cluster detection, which is in line with the above trillium paradigm.

While prediction and expectation are two different statistical concepts, these are not so strictly separated in colloquial language. As an example consider the semantics of tomorrow’s weather forecast. The weather forecast considers the most recent weather observations and their deviation from the historical average weather for tomorrow. Now assume, that the current weather is a heat wave, which results in higher temperatures than normal, i.e. the historically expected temperature. Therefore a good weather forecast will be a higher temperature, higher than the normal historically expected temperature. The normal temperature is the statistically expected value, whereas the forecast is the statistically predicted value. Yet, what is in common language the “expected” temperature for tomorrow is the statistically “predicted” higher temperature of the forecast, rather than the expected lower temperature from the historical average.

In summary, a disease cluster is detected as an “outlier” or group of neighbouring outliers in the distribution of the uncorrelated random effects of a respective spatial GLMM. While a plethora of methods for outlier detection exists, the key idea is, that a cluster is always an outlier in relation to a model. Including all relevant risk factors and their inter-relations in a model, should predict any otherwise unusual observations or cluster in the raw observed data. However, missing relevant factors in a model fit to the data at hand, i.e. confounders, can result in outliers or disease clusters.

## Discussion and Conclusion

This study revealed that the term disease cluster is defined by researchers in a variety of ways alluding to it being either an area or population. Some authors use the terms cluster and clustering as noun and verb for the same concept, while others consider clusters and clustering as separate and distinct spatial data pattern. This hinders proper communication between stakeholders and posts a barrier for one health programming.

Using the epidemiological triangle theory of agent, host and environment and connecting these to the spatial patterns of cluster, clustering and trend (or high-risk areas) resulted in the geographic epidemiological trillium paradigm. This paradigm also presents an approach to teaching the concepts of cluster as an outlier in population health data and clustering as a characteristic of the health issue under study. This theoretical footing further connects logically to the spatial GLMM and provides an approach to analyze the spatial patterns using observational data.

The following definition for the term spatial disease cluster is proposed here:

*A spatial disease cluster is a sub-population in a defined study area, where the disease frequency exceeds what can be predicted under a certain statistical model for given observations. A spatial disease cluster is an outlier in epidemiological data and conditional on the model used to analyze these data.*

Note, the term “disease” is used in this definition as a generic term for any population health issue. Similar definitions can also be used in the context of temporal and spatio-temporal situations. Further note, the definition makes reference to the data itself; this relates to the way data are analyzed and especially the spatial aggregation level and support area the data represent. The modifiable areal unit problem (MAUP) is well known and implies here that a disease cluster can emerge or disappear or shift its location, when data are aggregated to different spatial support areas (Waller and Gotway, 2004).

While the definition provides a distinction of the cluster pattern ( $u_i$ ) from clustering ( $v_i$ ) as well as trend and high-risk

areas ( $X_i\beta$ ), it also allows to distinguish clusters from outbreaks and hot spots. The terms cluster, “outbreak” and “hot spot” have been reviewed and discussed by Lessler et al. (2017), who point out the ambiguity in the use of the terms and the resulting confusion. The authors recommend using more precise terminology (e.g. emergence vs. transmission hot spot) to ensure more effective policy for disease control. While this is a good point, the essay is only advocating for more precise terminology without providing clear language examples.

Therefore it is proposed to distinguish outbreaks, cluster and hot spots as follows. An “outbreak” can be a single case of a newly emerging disease (e.g. ebola in Canada), but more often describes an event in which several linked cases occur in a short period. The epidemiological link does not imply geographic proximity of cases at exposure. As an example contaminated food could be shipped and consumed across Canada. On the other hand a cluster consists of more than one case and typically is a sub-population of several cases and non-cases, which are observed in close proximity. Thus, an outbreak can generate a cluster and often the epidemiological link between cases is approximated by geographic closeness, due to insufficient knowledge about the exposure and agent. Furthermore, a “hot spot” is an area, where a large number of cases is observed, i.e. where the caseload is extreme. This concept of a hot spot is based on the raw observed number of cases without adjustment for risk factors, i.e. is different from the notion of a trend or high-risk area.

A limitation of this study is the fact, that it is not based on a more systematic (e.g. scoping) review of the literature to explore the full variety of the use of the terms disease cluster and clustering as well as the implied conceptual differences. The examples from the literature are thus incomplete but considered here as sufficient to display how confusing the communication around these concepts can be.

In conclusion, a spatial disease cluster is an outlier in spatial data and therefore relative to an underlying model on which the analysis is based. The concept of a cluster is distinct from those of clustering, high-risk area, hot spot and outbreak, and such distinction will help future one health communication. Cluster are related to the host population in the epidemiological trillium and can be detected using a spatial GLMM. And finally a word of caution: a cluster identified as an outlier in the data using a statistical model alone, should only be considered a “candidate cluster” that calls for further public health considerations regarding the context of the data and disease.

## Competing Interest Statement

There are no financial or non-financial competing interests to be declared.

---

---

## References

- Berke, O. (2015, p. 627-633). London Cholera Epidemic and Epidemiology. Trefil, J. (editor) Discoveries in Modern Science: Exploration, Invention, Technology. MacMillan Reference USA.
- Berke, O., Mallare, J., Jeyabalan, T. & Clow, K. (2022). Rudolf Virchow—The epic. *Environmental Health Review* 65(2): 37-39. <https://doi.org/10.5864/d2022-008>
- Berke, O. & Waller, L. (2010). On the effect of diagnostic misclassification bias on the observed spatial pattern in regional count data - a case study using West Nile virus mortality data from Ontario, 2005. *Spatial and Spatio-temporal Epidemiology*, 1(2-3), 117–122. <https://doi.org/10.1016/j.sste.2010.03.004>
- Besag, J., E. & Newell, J. (1991). The detection of clusters in rare diseases. *JRSS A* 154: 143-155
- Besag, J., York, J. & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 43, 1–20. <https://doi.org/10.1007/BF00116466>
- Bhopal, R. (2008). *Concepts of Epidemiology*. (2nd edn.). Oxford University Press, Oxford.
- Cortes-Ramirez, J., Vilcins, D., Jagals, P., & Soares Magalhaes, R. J. (2020). Environmental and sociodemographic risk factors associated with environmentally transmitted zoonoses hospitalisations in Queensland, Australia. *One Health*, 12, 100206. <https://doi.org/10.1016/j.onehlt.2020.100206>



- Cromley, E.K. & McLafferty, S.L. (2012). *GIS and Public Health* (2nd edn.). Guilford Press, New York.
- Diggle, P. (2003). *Statistical Analysis of Spatial Point Patterns* (2nd edn.). Arnold, London.
- Haining, R.P. & Maheswaran, R. (2016). *Geographic Information Systems in Spatial Epidemiology and Public Health*. In: Lawson, A.B., Banerjee, S., Haining, R.P. & Ugarte, M.D. (Eds.). *Handbook of Spatial Epidemiology*. CRC Press / Chapman & Hall, Boca Raton.
- Kleinhues, P. (1996). Preface. In: Alexander F.E. & Boyle P. (Eds.). *Methods for Investigating Localized Clustering of Disease*. IARC Scientific Publication No. 135.
- Krieger, N. (1994) *Epidemiology and the web of causation: has anyone seen the spider?* *Social Science & Medicine*. 39:887-903. [https://doi.org/10.1016/0277-9536\(94\)90202-x](https://doi.org/10.1016/0277-9536(94)90202-x)
- Lawson, A. & Williams, F. (2001). *An Introductory Guide to Disease Mapping*. Wiley, New York.
- Lawson, A. (2018). *Bayesian Disease Mapping* (3rd Edn.). CRC Press / Chapman & Hall, Boca Raton.
- Lessler, J., Azman, A. S., McKay, H. S., & Moore, S. M. (2017). *What is a Hotspot Anyway?* *The American Journal of Tropical Medicine and Hygiene*, 96(6), 1270–1273. <https://doi.org/10.4269/ajtmh.16-0427>
- Waller, L. & Gotway, C.A. (2004). *Applied Spatial Statistics for Public Health Data*. Wiley, New York.
- Wheeler, D.C. (2007). *A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996 – 2003*. *Int J Health Geogr* 6, 13. <https://doi.org/10.1186/1476-072X-6-13>
- Porta, M.S., Greenland, S., Hernán, M. dos Santos Silva, I. & Last, J.M. (2014). *Dictionary of Epidemiology* (6th edn.). Oxford University Press, Oxford.
- Tango, T. (2010). *Statistical Methods for Disease Clustering*. Springer, New York.
- Tukey, J. (1977). *Exploratory Data Analysis*. Addison Wesley, Reading MA.
- Wakefield, J.C., Kellsall, J.E., Morris, S.E. (2000, p. 128/9) *Clustering, cluster detection and spatial variation in risk*. In: Elliott, P., Wakefield, J.C., Best, N.G., & Briggs, D.J. (eds.) *Spatial Epidemiology: Methods and Applications*. Oxford University Press, Oxford.