

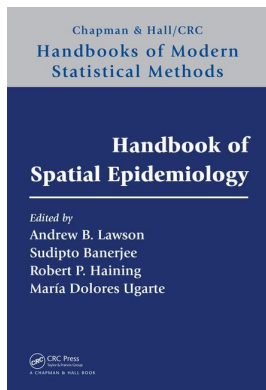
This article was downloaded by: 10.3.98.104

On: 05 Aug 2020

Access details: *subscription number*

Publisher: *CRC Press*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



Handbook of Spatial Epidemiology

Andrew B. Lawson, Sudipto Banerjee, Robert P. Haining, María Dolores Ugarte

Interpreting Clusters of Health Events

Publication details

<https://www.routledgehandbooks.com/doi/10.1201/b19470-5>

Geoffrey Jacquez, Jared Aldstadt

Published online on: 04 Apr 2016

How to cite :- Geoffrey Jacquez, Jared Aldstadt. 04 Apr 2016, *Interpreting Clusters of Health Events* from: Handbook of Spatial Epidemiology CRC Press

Accessed on: 05 Aug 2020

<https://www.routledgehandbooks.com/doi/10.1201/b19470-5>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

3

Interpreting Clusters of Health Events

Geoffrey Jacquez

*Department of Geography
State University of New York at Buffalo
Buffalo, New York*

Jared Aldstadt

*Department of Geography
State University of New York at Buffalo
Buffalo, New York*

CONTENTS

3.1	Introduction	57
3.2	How Are Cluster Studies Used and What Is Their Function?	58
3.2.1	Public health response	58
3.2.2	Confirmatory studies (association)	59
3.2.3	Hypothesis testing (advancing knowledge)	59
3.2.4	Are cluster studies science (role of cluster studies in scientific inquiry)?	59
3.3	Relevant Issues	61
3.4	Case Study	62
3.5	Conclusions	63
	References	63

3.1 Introduction

There has been some debate over the years regarding the utility of disease cluster analysis. While it is widely recognized that cluster analysis may be part of a public health response to an outbreak of disease, or to reports of a possible cluster by concerned citizens, the contribution and role of cluster analysis in spatial epidemiology is less clear. Applications frequently cited as successful exemplars include Snow's investigation of cholera (Snow 1855), the discovery of the link between fluoride in drinking water and dental caries (Dean 1938), and the identification of mercury poisoning as the cause of Minamata's disease (Harada 1995). Other etiological connections triggered by clustering of adverse health events include discovery of acquired immune deficiency syndrome (Friedman-Kien et al. 1981) and the association between a food allergy and tick bites (Commins et al. 2011). For chronic diseases such as cancer, the link to a putative cause is often unclear, since disease latency is often long, actual exposures are not measured, and the number of cases is often small, making a finding of a statistically significant excess difficult. Chronic disease clusters successfully linked to specific exposures include the Libby Montana cluster of lung cancer from asbestos exposures related to vermiculite mining (McDonald et al. 2004); the clustering of leukemia, lymphoma, and

adverse birth outcomes in Camp Lejeune, North Carolina, due to exposures to trichloroethylene, benzene, and other carcinogens from drinking water (ATSDR 2010); and a possible cluster of brain and central nervous systems cancers in Toms River, New Jersey, possibly caused by exposures to styrene, acrylonitrile, and styrene-acrylonitrile (SAN) trimer (ATSDR 1997). Although there have been some positive cluster studies, the majority of them do not find a significant excess, leading Neutra and others to view cluster analysis as an expensive endeavor that yields little insights into the existence of clusters, let alone their underlying causes (Neutra 1990). What then are the plausible underlying causes of disease clusters? And how have cluster analyses advanced our understanding of disease processes?

This chapter provides an overview of issues central to the interpretation of disease clusters. Here, we concern ourselves not with methods (refer to [Chapter 8](#) in this volume for statistical tests for clustering and surveillance); rather, we discuss when cluster studies have been used, their role in scientific inquiry, and relevant issues such as case ascertainment, the small numbers problem, and aspects of uncertainty. We seek to identify what can be known, and the pitfalls encountered along the way. A case study is included to demonstrate that several of the pitfalls can be avoided.

This chapter is organized around three questions or topics:

1. How are cluster studies used, and what is their function?
2. Are cluster studies science, and what is the role of cluster analysis in scientific inquiry?
3. What are the issues to consider when interpreting cluster study results?

3.2 How Are Cluster Studies Used and What Is Their Function?

Cluster studies are commonly used in three capacities: (1) as part of a public health response to a cluster allegation brought forward by concerned citizens, (2) in confirmatory studies that search for an association between a putative environmental exposure and a health outcome, and (3) in a hypothesis testing framework to systematically advance scientific knowledge. We consider each of these in turn.

3.2.1 Public health response

Kingsley et al. (2007) and others provide a useful review of cluster studies at the Centers for Disease Control and Prevention (CDC) and how they may be used within the context of a public health response to a clustering of health events identified and reported by concerned citizens (Buehler et al. 2004; CDC 1990; Kingsley et al. 2007). Attendant issues within this framework include preselection bias, known colloquially as the “Texas sharpshooter problem,” which may increase the type I error (false positives). The rationale is that by constantly surveying their environment, citizens act to “preselect” apparent local disease excesses and thereby report chance clusters with higher probability. In practice, the majority of cluster studies have negative results—that is, they do not find a statistically significant excess of disease cases (Schulte et al. 1987; Warner and Aldrich 1988). This failure to result in positive findings is one of the reasons why Neutra (1990) suggested cluster studies are a wasteful enterprise for public health agencies, especially when those resources might reliably be used in vaccination, screening, and other activities that reduce disease burden. The application of clustering methods within the public health response framework may be called

pre-epidemiology since a designed sample may be lacking and an epidemiologically sound study design is imposed only after a statistically significant cluster has been demonstrated.

3.2.2 Confirmatory studies (association)

Once geographically statistically significant clusters of health events have been found, it is tempting to seek possible causes by searching for associations with environmental exposures and other factors. While such an activity may suggest causal hypotheses, it does not have the power to reject hypotheses or test predictions (Jacquez 2004). From that perspective, confirmatory studies that search for association have little appeal for spatial epidemiology.

3.2.3 Hypothesis testing (advancing knowledge)

What, then, is the role of cluster studies in spatial epidemiology? We can think of spatial epidemiology as the documentation and analysis of spatial disease patterns to better understand the causes and correlates of disease. As noted above, cluster studies in the public health response framework often lack a sound sampling design, as they are by definition a post hoc response to alleged clusters. Confirmatory studies seek to identify associations between clusters and spatial patterns of environmental variables, and thus do not employ an inferential framework consistent with the scientific method put forward by Popper and others (Platt 1964; Popper 1968). The coupling of epidemiological study designs with cluster analysis techniques supports a more robust inference structure (Meliker et al. 2009a, 2009b) and may be referred to as post-epidemiology. Here, cluster analysis methods are applied to the residual risk that is not explained by the risk factors identified in a parent epidemiological study. Any clusters so identified then are above and beyond the causal factors identified in the parent epidemiological study. This approach has been successfully employed in case-control studies to identify potential clustering of testicular cancers (Sloan et al. 2013), childhood leukemia and diabetes (Schmiedel et al. 2011), and breast cancers (Jacquez et al. 2013), among others. When good-quality disease registries are available that document most, if not all, incident cases, it is possible to use cluster studies to identify statistically significant local excesses of disease, provided known explanatory (e.g., nuisance) factors have been taken into account. Spatial patterns in known explanatory factors may be accounted for using “neutral models” (Goovaerts and Jacquez 2004, 2005). This makes it possible to use disease cluster analysis to test hypotheses regarding the spatial patterns of disease, thereby increasing the utility of clustering methods in spatial epidemiology.

3.2.4 Are cluster studies science (role of cluster studies in scientific inquiry)?

The above discussion identified two components necessary in order for cluster studies to advance spatial epidemiological knowledge: (1) a sound sampling design, usually from an epidemiological study (e.g., case-control or cohort) or from a complete census of incident cases such as may be provided by a disease registry and (2) null hypotheses that support the inclusion of known factors that might explain observed clusters (e.g., neutral models). Several questions and issues must be addressed when interpreting clusters in spatial epidemiology:

1. *Is there an excess?* Most clustering methods fall under the rubric of global, local, or focused tests. Here, global clustering refers to the existence of excess risk somewhere in the study area; local clustering methods report where those clusters are found, usually with their spatial extent, population size affected, and an estimate

of the excess risk; and focused clustering assesses whether there is elevated risk in the immediate vicinity of a suspected location or focus (Besag and Newell 1991; Lawson 1989; Waller et al. 1995). Statistical clustering methods are used to identify whether there is a statistically significant excess, and where that excess may be found.

2. *Is it real?* The identification of a statistically significant excess does not mean that the disease cluster is real, in the sense of reflecting a true, underlying excess in disease risk. Spurious findings of an excess of health events are expected by chance. But other factors can play a role as well, giving rise to false cluster findings. The CDC makes a distinction between false and true clusters (CDC 1990). False clusters lack a plausible biological explanation and may be comprised of cases with symptoms or illnesses with unrelated causes. True clusters are comprised of a statistically significant excess of cases with a common, plausible biological explanation. Hence, biological plausibility, as well as statistical significance, must be considered when assessing whether a cluster is real.
3. *What does it mean?* When a real excess has been identified, the challenge of interpretation begins in earnest. This may involve the identification of the set of possible causes that may underlie the observed excess. Once these have been enumerated, these specific alternative explanations may be excluded in a systematic fashion by analysis of additional data or information. The set of explanations may include chance, case-attractor hypotheses, spatial pattern in covariates, and the action of causative exposures and disease processes. Case-attractor hypotheses describe processes that bring cases together in the absence of elevated disease risk. Examples include patients moving to be near treatment facilities, and modifications in behaviors that cause either cases or noncases to cluster. An example of the latter is the healthy worker effect, where in order to be employed, workers must be in comparatively sound health (Fornalski and Dobrzyński 2010).
4. *What's the hypothesis?* Almost all tests for clustering employ null and alternative hypotheses. The null hypothesis is a statement regarding the spatial pattern expected in the absence of a cluster process, while the alternative hypothesis is embodied in specification of the spatial weights (Griffiths 1995). Often, the meaning of the null and alternative hypotheses in terms of the underlying disease process is given little thought. One example is the use of complete spatial randomness (CSR) as the null hypothesis in many statistical tests of spatial disease clustering. CSR usually implies that the underlying disease risk, in the absence of a cluster process, will be uniform across the study area. In practice, this seldom is true, since disease risk may be associated with covariates such as age, income, and ethnicity. When working with area-based data, one solution is to use covariate-adjusted rates. Another solution is to employ randomization or Monte Carlo methods with randomization procedures that account for the covariate structure. One may also model the expected risk while incorporating known risk factors into the model, and then inspect the residuals (observed minus expected risk) for spatial pattern. But unless spatial structure in the covariates is fully accounted for, CSR may not be a reasonable null hypothesis.

For the alternative hypothesis, the specification of spatial weights should be accomplished in a manner that reflects the underlying cluster process. For example, when diffusion is suspected, it may make sense to link adjacent locations to one another, for example, specify a spatial weight of 1 for nearby locations. But if transport on a network is suspected, one might choose to link locations using the

actual road network. In practice, one thus may wish to explore a suite of alternative hypotheses, each corresponding to its own spatial weight set. This requires careful enumeration of what the set of alternative hypotheses might be, coupled with construction of the spatial weights for each alternative hypothesis. In certain instances, knowledge of the scale of the clustering may be absent or anecdotal, in which case spatial weight sets corresponding to different spatial scales may be employed, in a sensitivity analysis of the effect of spatial scale on the results of the cluster analysis.

5. *What is the sampling frame and design?* Interpretation of a cluster finding must be premised on the sampling frame and sampling design. Especially within the public health framework mentioned above, the data may be encountered, for example, cases that have been reported by concerned citizens. Here, a designed sample that is representative of an underlying study population is absent, and any inferences that may be drawn thus would apply only to the sample. When registry data are analyzed, all cases conceivably could be included in the study, and inferences would then apply to the entire population covered by the registry. When a postepidemiological framework is used, the scope of inference will apply to the study population from which the sample in the parent epidemiological study was drawn.
6. *What inferences can be drawn regarding underlying disease processes?* When the data consist of cases reported by concerned citizens, it is difficult to make inferences regarding underlying disease processes, since there usually will be several alternative explanations for an observed disease cluster. But when the data are sampled in a systematic fashion, and tests for spatial, temporal, and space–time disease clustering are used, it may be possible to construct inferences regarding an underlying disease process (see, e.g., Jacquez et al. 2013, table 1). But this is a difficult prospect, especially in the absence of additional information, since different diseases in different situations can give rise to similar space–time patterns of case occurrence.

3.3 Relevant Issues

There are several other issues that can impact the interpretation of cluster findings. These include case ascertainment, incomplete reporting, the small numbers problem, and geocoding location error.

Accurate *case ascertainment* is critical. Are cases correctly diagnosed, and to what extent are there misdiagnoses? Considerable effort has gone into describing and understanding the spatial patterns of West Nile virus (WNV) transmission in North America since its introduction in 1999 (Nash et al. 2001; Ruiz et al. 2010; Hayes et al. 2005). Most human WNV infections are asymptomatic, and many result in very common clinical features that are likely to be misdiagnosed (Davis et al. 2006). In this case, clusters of human WNV illnesses may be a result of variability in diagnoses, host characteristics, or strain virulence, as well as an indicator of infection risk. Additionally, comparative studies have shown that spatial analysis and clustering results vary considerably, depending on whether data are obtained through a notifiable disease reporting system or administrative health records (Jones et al. 2012; Yiannakoulias and Svenson 2009).

Incomplete reporting can reduce the power to detect a true cluster. And when the extent of incomplete reporting is spatially structured, it can give rise to the finding of clusters when

they are, in fact, absent. One example of this is maps where the disease rate varies dramatically across administrative boundaries. While this may reflect a true difference in disease burden, it is often explained by reporting differences across administrative units. Passive disease surveillance systems are particularly susceptible to spatially heterogeneous intensity of reporting that confound cluster analysis. Economic and social barriers to obtaining proper healthcare and associated differences in treatment-seeking behavior may lie at the root of this problem. For example, incomplete reporting has been suggested as a cause of geographical clustering of autism incidence in California, and the apparent associations between parental education levels and autism incidence (Van Meter et al. 2010).

The *small numbers problem* refers to the increase in variance in disease rates (calculated as the number of incident cases, e.g., divided by the size of the at-risk population) as the size of the at-risk population in the denominator decreases. Many cluster analysis methods account for differences in population size, but not all of them do. The local Moran statistic, for example, is often applied to raw disease rates, does not account for population size, and hence can yield spurious cluster findings (false positives). Some practitioners recommend that disease rates be smoothed using an empirical Bayesian smoother prior to local Moran analysis. Such smoothing introduces spurious spatial autocorrelation into the resulting smoothed rates. Since this spurious autocorrelation is not accounted for in the null hypothesis (which in most studies is CSR), the cluster analysis results are unreliable.

Geocoding is frequently used in cluster analysis to identify the spatial coordinates of health events (Abe and Stinchcomb 2008; Goldberg 2008). That geocodes (the geographic coordinates that are the result of geocoding an address) have an associated location uncertainty is well known (Krieger et al. 2001; Bonner et al. 2003; Oliver et al. 2005; Whitsel et al. 2006). Location uncertainty impacts disease cluster analysis by decreasing statistical power (Rushton et al. 2006; Zandbergen et al. 2012) and introducing bias into exposure models that use geocoded locations (Mazumdar et al. 2008). However, at the time of this writing, geocoding location error is routinely ignored in the interpretation of cluster results (Jacquez 2012; Whitsel et al. 2006).

3.4 Case Study

Sloan et al. (2015) examine testicular cancer cases to determine whether there are spatial or spatial-temporal clusters of risk. Excess risk for testicular cancer is largely unexplained, and this study seeks to generate hypotheses by detecting locations of shared exposure among testicular cancer patients. Case-control study design, detailed residential history data, and adjustment for personal risk factors are employed to rigorously test the null hypothesis of uniform risk with nearest-neighbor-based Q-statistics (Jacquez et al. 2005).

The study is based on reliable case data and a rigorous case-control study design. The case data are taken from the Danish Cancer Registry and include 3297 cases diagnosed between 1991 and 2003. Two sets of birth-date-matched controls were obtained from the Danish Civil Registration System. The second set of controls was obtained to address the possibility that clustering results are due to choice of control group (Nordsborg et al. 2013). Residential histories were also obtained from the registration system. Inclusion of residential histories accounts for population mobility and also allows for the temporal alignment of cases by age at diagnosis or time prior to diagnosis. The temporal alignment of cases allows for an examination of important ages of exposure and accounts for the latency period between exposure and diagnosis.

Risk factors were examined first in a conditional logistic regression model that included individual-level characteristics and community-level socioeconomic status. Family history of testicular cancer was the only significant predictor of increased risk and was incorporated into the cluster detection analysis. The problem of multiple testing was addressed by performing simulation studies and checking for correspondence with the SaTScan procedure. Sloan et al. (2012) employed residential history data from the same registration system in a simulation study to provide a threshold for significant clusters in the face of multiple testing. Local space–time clusters detected with the Q-statistics were reexamined for spatial clustering using SaTScan, which is robust in the face of multiple testing.

This study did not find convincing evidence of clusters of risk for testicular cancer. This “negative” result is itself instructive and hypothesis generating. It may be that environmental exposures do not play a role in testicular cancer risk. The result may also indicate that important environmental exposures do not vary at a scale that can be detected by this method. Exposures within households or an environmental factor that is practically uniform throughout Denmark are an example. In either case, these robust results will be important when aggregating studies from different regions or in the design of future studies examining the behavioral and environmental risks for testicular cancer.

3.5 Conclusions

Disease cluster analysis consists of a spectrum of techniques, from inferential pattern analysis (e.g., tests for disease clustering) to modeling approaches, including regression, geostatistical, and Bayesian models. This chapter has concentrated primarily on inferential pattern analysis and the interpretation of clusters of health events.

While the 1980s and 1990s might be correctly typified as an era of pre-epidemiology in cluster analysis, advances in recent years, especially the use of rigorous sampling designs, have greatly strengthened the inferential structure of cluster analysis. The included case study demonstrates that available data and methodologies have moved cluster studies into the realm of post-epidemiology. Increasingly, health surveillance systems are using unstructured data streams, such as those from Twitter and Google search engines. These kinds of data tend to be heterogeneous, and the sampling method is only partially known, without a formal sampling design. An important area of future research in cluster analysis is how to impose representative sampling frames and designs on such data streams.

References

- Abe, T. and D. Stinchcomb. (2008). Geocoding practices in cancer registries. In *Geocoding Health Data*, ed. G. Rushton, M. P. Armstrong, J. Gittler, B. R. Greene, C. E. Pavlik, M. M. West, and D. Zimmerman, 111–125. Boca Raton, FL: CRC Press.
- ATSDR (Agency for Toxic Substances and Disease Registry). (1997). Childhood cancer incidence health consultation: A review and analysis of cancer registry data, 1979–1995 for Dover Township (Ocean County), New Jersey. Atlanta, GA: U.S. Department of Health and Human Services, ATSDR, and New Jersey Department of Health and Senior Services, Hazardous Site Health Evaluation Program, Division of Epidemiology, Environmental and Occupational Health.

- ATSDR (Agency for Toxic Substances and Disease Registry). (2010). Camp Lejeune, North Carolina: Health study activities frequently asked questions (FAQs). Atlanta, GA: ATSDR, April 2.
- Besag, J. and J. Newell. (1991). The detection of clusters in rare diseases. *Journal of the Royal Statistical Society Series A*, 154, 143–155.
- Bonner, M. R., D. Han, J. Nie, P. Rogerson, J. E. Vena, and J. L. Freudenheim. (2003). Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology*, 14, 408–412.
- Buehler, J. W., R. S. Hopkins, J. M. Overhage, D. M. Sosin, and V. Tong. (2004). Framework for evaluating public health surveillance systems for early detection of outbreaks: Recommendations from the CDC Working Group. *MMWR Recommendations and Reports*, 53, 1–11.
- CDC (Centers for Disease Control and Prevention). (1990). Guidelines for investigating clusters of health events. *Mortality and Morbidity Weekly Report*, 39, 1–16.
- Commins, S. P., H. R. James, L. A. Kelly, S. L. Pochan, L. J. Workman, M. S. Perzanowski, K. M. Kocan, J. V. Fahy, L. W. Nganga, and E. Ronmark. (2011). The relevance of tick bites to the production of IgE antibodies to the mammalian oligosaccharide galactose- α -1,3-galactose. *Journal of Allergy and Clinical Immunology*, 127, 1286–1293.
- Davis, L. E., R. DeBiasi, D. E. Goade, K. Y. Haaland, J. A. Harrington, J. B. Harnar, S. A. Pergam, M. K. King, B. DeMasters, and K. L. Tyler. (2006). West Nile virus neuroinvasive disease. *Annals of Neurology*, 60, 286–300.
- Dean, H. T. (1938). Endemic fluorosis and its relation to dental caries. *Public Health Report*, 53, 1443–1452.
- Fornalski, K. W. and L. Dobrzyński. (2010). The healthy worker effect and nuclear industry workers. *Dose-Response*, 8, 125–147.
- Friedman-Kien, A., L. Laubenstein, M. Marmor, K. Hymes, J. Green, A. Ragaz, J. Gottlieb, F. Muggia, R. Demopoulos, and M. Weintraub. (1981). Kaposi sarcoma and *Pneumocystis pneumonia* among homosexual men—New York City and California. *MMWR Morbidity and Mortality Weekly Report*, 30, 305–308.
- Goldberg, D. 2008. *A Geocoding Best Practices Guide*. Springfield, IL: North American Association of Central Cancer Registries.
- Goovaerts, P. and G. M. Jacquez. (2004). Accounting for regional background and population size in the detection of spatial clusters and outliers using geostatistical filtering and spatial neutral models: The case of lung cancer in Long Island, New York. *International Journal of Health Geographics*, 3, 14.
- Goovaerts, P. and G. M. Jacquez. (2005). Detection of temporal changes in the spatial distribution of cancer rates using local Moran's I and geostatistically simulated spatial neutral models. *Journal of Geographical Systems*, 7, 137–159.
- Griffiths, D. 1995. Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. In *Practical Handbook of Spatial Statistics*, ed. Arlinghaus, S. L., D. A. Griffith, W. C. Arlinghaus, W. D. Drake, and J. D. Nystuen, 65–82. Boca Raton, FL: CRC Press.

- Harada, M. (1995). Minamata disease: Methylmercury poisoning in Japan caused by environmental pollution. *CRC Critical Reviews in Toxicology*, 25, 1–24.
- Hayes, E. B., N. Komar, R. S. Nasci, S. P. Montgomery, D. R. O’Leary, and G. L. Campbell. (2005). Epidemiology and transmission dynamics of West Nile virus disease. *Emerging Infectious Disease*, 11, 1167–1173.
- Jacquez, G., J. Barlow, R. Rommel, A. Kaufmann, M. Rienti, G. AvRuskin, and J. Rasul. (2013). Residential mobility and breast cancer in Marin County, California, USA. *International Journal of Environmental Research and Public Health*, 11, 271–295.
- Jacquez, G. M. (2004). Current practices in the spatial analysis of cancer: Flies in the ointment. *International Journal of Health Geographics*, 3, 22.
- Jacquez, G. M. (2012). A research agenda: Does geocoding positional error matter in health GIS studies? *Spatial and Spatio-Temporal Epidemiology*, 3, 7–16.
- Jacquez, G. M., A. Kaufmann, J. Meliker, P. Goovaerts, G. AvRuskin, and J. Nriagu. (2005). Global, local and focused geographic clustering for case-control data with residential histories. *Environmental Health*, 4, 4.
- Jones, S. G., W. Conner, B. Song, D. Gordon, and A. Jayakaran. (2012). Comparing spatio-temporal clusters of arthropod-borne infections using administrative medical claims and state reported surveillance data. *Spatial and Spatio-Temporal Epidemiology*, 3, 205–213.
- Kingsley, B. S., K. L. Schmeichel, and C. H. Rubin. (2007). An update on cancer cluster activities at the Centers for Disease Control and Prevention. *Environmental Health Perspectives*, 115, 165–171.
- Krieger, N., P. Waterman, K. Lemieux, S. Zierler, and J. Hogan. (2001). On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. *American Journal of Public Health*, 91, 1114–1116.
- Lawson, A. B. 1989. Score tests for detection of spatial trend in morbidity data. Dundee, Scotland: Dundee Institute of Technology.
- Mazumdar, S., G. Rushton, B. J. Smith, D. L. Zimmerman, and K. J. Donham. (2008). Geocoding accuracy and the recovery of relationships between environmental exposures and health. *International Journal of Health Geographics*, 7, 13.
- McDonald, J., J. Harris, and B. Armstrong. (2004). Mortality in a cohort of vermiculite miners exposed to fibrous amphibole in Libby, Montana. *Occupational and Environmental Medicine*, 61, 363–366.
- Meliker, J. R., P. Goovaerts, G. M. Jacquez, G. A. Avruskin, and G. Copeland. (2009a). Breast and prostate cancer survival in Michigan: Can geographic analyses assist in understanding racial disparities? *Cancer*, 115, 2212–2221.
- Meliker, J. R., G. M. Jacquez, P. Goovaerts, G. Copeland, and M. Yassine. (2009b). Spatial cluster analysis of early stage breast cancer: A method for public health practice using cancer registry data. *Cancer Causes Control*, 20, 1061–1069.
- Nash, D., F. Mostashari, A. Fine, J. Miller, D. O’Leary, K. Murray, A. Huang, A. Rosenberg, A. Greenberg, and M. Sherman. (2001). The outbreak of West Nile virus infection in the New York City area in 1999. *New England Journal of Medicine*, 344, 1807–1814.

- Neutra, R. R. (1990). Counterpoint from a cluster buster. *American Journal of Epidemiology*, 132, 1–8.
- Nordsborg, R. B., J. R. Meliker, A. K. Ersbøll, G. M. Jacquez, and O. Raaschou-Nielsen. (2013). Space-time clustering of non-Hodgkin lymphoma using residential histories in a Danish case-control study. *PLoS One*, 8, e60800.
- Oliver, W. N., K. Matthews, M. Siadat, F. Hauck, and L. Pickle. (2005). Geographic bias relating to geocoding error in epidemiologic studies. *International Journal of Health Geographics*, 4, 29.
- Platt, J. (1964). Strong inference. *Science*, 146, 347–353.
- Popper, K. 1968. *The Logic of Scientific Discovery*. New York: Harper and Rowe.
- Ruiz, M. O., L. F. Chaves, G. L. Hamer, T. Sun, W. M. Brown, E. D. Walker, L. Haramis, T. L. Goldberg, and U. D. Kitron. (2010). Local impact of temperature and precipitation on West Nile virus infection in *Culex* species mosquitoes in northeast Illinois, USA. *Parasites & Vectors*, 3, 19.
- Rushton, G., M. Armstrong, J. Gittler, B. Greene, C. Pavlik, M. West, and D. Zimmerman. (2006). Geocoding in cancer research—a review. *American Journal of Preventive Medicine*, 30, S16–S24.
- Schmiedel, S., G. Jacquez, M. Blettner, and J. Schüz. (2011). Spatial clustering of leukemia and type 1 diabetes in children in Denmark. *Cancer Causes and Control*, 22, 849–857.
- Schulte, P. A., R. L. Ehrenberg, and M. Singal. (1987). Investigation of occupational cancer clusters: Theory and practice. *American Journal of Public Health*, 77, 52–56.
- Sloan, C. D., R. Bastrup-Norstrom, G. M. Jacquez, P. J. Landrigan, O. Raaschou-Nielsen, and J. R. Meliker. (2013). Space-time analysis of testicular cancer clusters using residential histories: A case-control study in Denmark. *PLoS One*, 10, e0120285.
- Sloan, C. D., G. M. Jacquez, C. M. Gallagher, M. H. Ward, O. Raaschou-Nielsen, R. B. Nordsborg, and J. R. Meliker. (2012). Performance of cancer cluster Q-statistics for case-control residential histories. *Spatial and Spatio-Temporal Epidemiology*, 3, 297–310.
- Sloan, C. D., R. B. Nordsborg, G. M. Jacquez, O. Raaschou-Nielsen, and J. R. Meliker. (2015). Space-time analysis of testicular cancer clusters using residential histories: A case-control study in Denmark. *PLoS One*, 10, e0120285.
- Snow, J. (1855). *On the Mode of Communication of Cholera*. London: John Churchill, New Burlington Street.
- Van Meter, K. C., L. E. Christiansen, L. D. Delwiche, R. Azari, T. E. Carpenter, and I. Hertz-Picciotto. (2010). Geographic distribution of autism in California: A retrospective birth cohort analysis. *Autism Research*, 3, 19–29.
- Waller, L. A., B. W. Turnbull, G. Gustafsson, U. Hjalmar, and B. Andersson. (1995). Detection and assessment of clusters of disease: An application to nuclear power plant facilities and childhood leukaemia in Sweden. *Statistics in Medicine*, 14, 3–16.
- Warner, S. C. and T. E. Aldrich. (1988). The status of cancer cluster investigations undertaken by state health departments. *American Journal of Public Health*, 78, 306–307.

- Whitsel, E., P. Quilbrera, R. Smith, D. Catellier, D. Liao, A. Henley, and G. Heiss. (2006). Accuracy of commercial geocoding: Assessment and implications. *Epidemiologic Perspectives and Innovations*, 3, 8.
- Yiannakoulis, N. and L. Svenson. (2009). Differences between notifiable and administrative health information in the spatial-temporal surveillance of enteric infections. *International Journal of Medical Informatics*, 78, 417–424.
- Zandbergen, P. A., T. C. Hart, K. E. Lenzer, and M. E. Camponovo. (2012). Error propagation models to examine the effects of geocoding quality on spatial analysis of individual-level datasets. *Spatial and Spatio-Temporal Epidemiology*, 3, 69–82.