pagepress

Geospatial Health
Geospatial Health 2026; 21:1544
doi:10.4081/gh.2026.1544



# Ten common pitfalls in spatial epidemiology and how to avoid them

Behzad Kiani,[1] Nima Kianfar,[2] Munazza Fatima,[3] Robert Bergquist[4]

[1]The Frazer Institute, Faculty of Health, Medicine, and Behavioural Sciences, The University of Queensland, Brisbane, Australia; [2]Department of Geospatial Information Systems, Faculty of Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran; [3]Department of Geography and Geoinformatics, Faculty of Physical and Mathematical Sciences, The Islamia University of Bahawalpur, Bahawalpur, Pakistan; [4]Geospatial Health, Ingerod, Brastad, Sweden

## Abstract

Spatial epidemiology provides powerful tools for understanding geographic patterns in health and disease, but methodological and interpretative pitfalls can undermine the validity of spatial analyses. This editorial highlights ten common pitfalls spanning spatial dependence, scale, ecological inference, small-area estimation, spatial confounding, hotspot interpretation, model validation, measurement and statistical uncertainty, and causal interpretation. For each, we provide practical guidance to support more rigorous and reliable spatial epidemiological research. As geospatial data, artificial intelligence, and analytical methods continue to advance, careful spatial reasoning remains essential to ensure that methodological sophistication translates into valid, interpretable, and meaningful public health evidence.

**Key words:** spatial epidemiology, spatial dependence, model validation, exposure assessment, uncertainty analysis.

*Correspondence: Behzad Kiani, The Frazer Institute, Faculty of Health, Medicine, and Behavioural Sciences, The University of Queensland, Brisbane, Australia. E-mail: B.kiani@uq.edu.au*

Spatial epidemiology has become a cornerstone of modern public health, providing the framework to understand how diseases vary across space and time and why populations experience different risks. Advances in geospatial data, spatial statistics, and Artificial Intelligence (AI) have greatly expanded its analytical capabilities, enabling increasingly sophisticated disease mapping, risk modelling, and forecasting. However, these advances also introduce new methodological challenges: increasingly sophisticated methods do not guarantee valid inference. Spatial analyses can still produce misleading conclusions when issues such as spatial dependence, scale, uncertainty, inadequate validation, or causal interpretation are overlooked. As the field continues to evolve, methodological progress requires not only increasingly advanced analytical tools but also stronger spatial reasoning. Here, we outline ten common pitfalls that can undermine spatial epidemiological inference and provide practical guidance for avoiding them (Table 1).

## 1. Ignoring spatial autocorrelation

Most conventional statistical methods assume that observations are independent, an assumption that is rarely valid in spatial epidemiology because nearby locations often share common underlying environmental, demographic, behavioural, or healthcare-related characteristics. Failure to account for spatial autocorrelation can bias parameter estimates, distort statistical inference, and lead to misleading conclusions. Spatial autocorrelation should therefore be routinely assessed and, where appropriate, incorporated into the analytical framework.

## 2. Ignoring spatial scale and the modifiable areal unit problem

Spatial patterns are inherently scale-dependent. Associations observed at one geographic scale may weaken, disappear, or even reverse when examined at another because of the Modifiable Areal Unit Problem (MAUP) (Tuson *et al.*, 2019). The choice of spatial unit should therefore be justified, alternative spatial scales explored where feasible, and findings interpreted in the context of the selected spatial resolution.

## 3. Committing the ecological fallacy

Ecological analyses are essential for understanding population health, but they cannot establish individual-level associations. Confusing population-level (ecological) associations with individual-level associations remains one of the most common interpretational errors in spatial epidemiology (Shih *et al.*, 2023). To avoid unsupported causal or individual-level interpretations, researchers should clearly state whether their findings apply to populations or individuals and interpret them accordingly.

## 4. Using unstable small-area rates

Disease maps may create a false impression of precision when based on unstable rates. Small populations and rare outcomes may produce extreme estimates driven by random variation, whereas excessive smoothing can obscure meaningful local heterogeneity (Quick & Song, 2024). Both uncertainty and population denominators should therefore be considered when interpreting small-area

estimates, and appropriate rate-stabilisation techniques applied where necessary.

## 5. Failing to account for spatial confounding

Spatial confounding arises when exposures, outcomes, and covariates share similar geographic patterns, making observed associations difficult to interpret (Marques *et al.*, 2022). Spatial random effects may account for residual spatial structure but do not, by themselves, resolve spatial confounding. Careful causal reasoning, transparent model specification, and thoughtful consideration of plausible spatial confounders remain essential for valid inference.

## 6. Overinterpreting hotspot analyses

Hotspot and cluster analyses identify where disease is concentrated, not why. Although statistically significant clusters may indicate elevated disease risk, they can also reflect population structure, surveillance intensity, data quality, or analytical choices rather than underlying disease processes (Tango, 2021). Cluster detection should therefore be viewed as a hypothesis-generating tool that requires epidemiological interpretation and independent validation.

## 7. Introducing spatial data leakage

Predictive performance is meaningful only if models generalise beyond the data on which they were developed. Spatial data leakage occurs when geographically proximate observations are included in both the training and test datasets, allowing information to be unintentionally shared between them. This can lead to overly optimistic estimates of model performance because nearby observations are often spatially correlated (Roberts *et al*., 2017). Validation strategies should therefore preserve spatial independence through approaches such as spatial blocking, geographically independent testing, temporal validation, or external validation.

## 8. Overlooking measurement uncertainty

Spatial analyses are only as reliable as the data on which they are based. Measurement uncertainty introduced through geocoding errors, residential mobility, exposure misclassification, and spatial or temporal misalignment between exposures and health outcomes can compromise the validity of spatial epidemiological findings (Zhang *et al.*, 2016). Transparent reporting of data quality, spatial and temporal resolution, exposure assignment procedures, and sensitivity analyses is therefore essential.

## 9. Ignoring statistical uncertainty

Every spatial estimate carries statistical uncertainty, yet maps and model outputs often convey a greater degree of precision than warranted. Uncertainty may arise from sampling variability, model specification, parameter estimation, spatial prediction, and extrapolation beyond observed locations (Delmelle *et al.*, 2022). Confidence or credible intervals, prediction intervals, uncertainty maps, and appropriate validation should therefore be routine components of spatial epidemiological reporting rather than optional supplements.

## 10. Equating spatial association with causation

Spatial epidemiology is exceptionally powerful for identifying geographic patterns and generating hypotheses, but spatial association alone does not establish causation. Causal inference requires appropriate study design, temporal ordering, control of confounding, and explicit causal assumptions. Interpretations should remain consistent with the study objectives and clearly distinguish between descriptive, predictive, explanatory, and causal analyses.

## Looking ahead

Spatial epidemiology is entering an era of unprecedented opportunity. Advances in geospatial data, artificial intelligence, and computational methods are transforming our ability to understand, predict, and prevent disease. However, methodological innovation must be matched by methodological rigour. Sophisticated analytical tools cannot compensate for inadequate study design, poor data quality, inappropriate validation, or overinterpretation of results. Ultimately, the future of spatial epidemiology will depend not only on developing more advanced methods, but on applying them wisely.

**Table 1.** Practical guidance for avoiding common pitfalls in spatial epidemiology.

| No. | Pitfall | Why it matters | Good practice |
|---|---|---|---|
| 1 | Spatial autocorrelation | Biases statistical inference | Assess spatial dependence and use spatial models when appropriate |
| 2 | Spatial scale and MAUP | Findings may vary across spatial units | Justify spatial units and assess sensitivity to scale |
| 3 | The ecological fallacy | Area-level findings may not apply to individuals | Match conclusions to the level of inference |
| 4 | Unstable small-area rates | Random variation may produce misleading maps | Quantify uncertainty and consider rate stabilisation |
| 5 | Confounding factors | Shared spatial structure may bias associations | Identify spatial confounders and use causal reasoning |
| 6 | Hotspot interpretation | Clusters do not explain disease mechanisms | Treat clusters as hypothesis-generating and validate findings |
| 7 | Data leakage | Predictive performance may be overestimated | Use spatially independent or external validation |
| 8 | Measurement uncertainty | Data errors can bias findings | Report data quality, spatial resolution, and sensitivity analyses |
| 9 | Uncertain data | Results may appear more precise than they are | Report and communicate uncertainty |
| 10 | Uncertain associations | Association does not imply causation | Align interpretation with the study design |

## References

Delmelle EM, Desjardins MR, Jung P, Owusu C, Lan Y, Hohl A, Dony, C. 2022. Uncertainty in geospatial health: Challenges and opportunities ahead. Ann Epidemiol 65;15–30.

Marques I, Kneib T, Klein N. 2022. Mitigating spatial confounding by explicitly correlating Gaussian random fields. Environmetrics 33:e2727.

Quick H, Song G. 2024. Reliable event rates for disease mapping. J Official Statistics 40:333–47.

Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W, Warton DI, Wintle BA, Hartig F, Dormann CF. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40:913–29.

Shih Y-CT, Bradley C, Yabroff KR. 2023. Ecological and individualistic fallacies in health disparities research. J Nat Cancer Inst 115:488–491.

Tango T. 2021. Spatial scan statistics can be dangerous. Stat Methods Med Res 30:75–86.

Tuson M, Yap M, Kok MR, Murray K, Turlach B, Whyatt D. 2019. Incorporating geography into a new generalized theoretical and statistical framework addressing the modifiable areal unit problem. Int J Health Geogr 18:6.

Zhang Z, Manjourides J, Cohen T, Hu Y, Jiang Q. 2016. Spatial measurement errors in the field of spatial epidemiology. Int J Health Geogr 15:21.

Received: 6 July 2026; Accepted: 21 July 2026.

Contributions: Behzad Kiani conceptualised the article and wrote the original draft. Nima Kianfar contributed to the conceptualisation and drafting of the manuscript. Munazza Fatima and Robert Bergquist reviewed and edited the manuscript. All authors reviewed and approved the final version.

Conflict of interest: the authors declare no potential conflict of interest, and all authors confirm accuracy.

Ethics approval: not applicable.

Availability of data and materials: all data generated or analyzed during this study are included in this published article.