KI REPORTS
KIReports.org

REVIEW

Geospatial Modeling Methods in Epidemiological Kidney Research: An Overview and Practical Example

Check for updates

R. Blake Buchalter[1,2], Sumit Mohan[3,4] and Jesse D. Schold[5,6]

[1]Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA; [2]Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA; [3]Division of Nephrology, Department of Medicine, Columbia University Vagelos College of Physicians and Surgeons, New York, New York, USA; [4]Department of Epidemiology, Columbia University Mailman School of Public Health, New York, New York, USA; [5]Department of Surgery, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA; and [6]Department of Epidemiology, Colorado School of Public Health, Aurora, Colorado, USA

Geospatial modeling methods in population-level kidney research have not been used to full potential because few studies have completed associative spatial analyses between risk factors and exposures and kidney conditions and outcomes. Spatial modeling has several advantages over traditional modeling, including improved estimation of statistical variation and more accurate and unbiased estimation of coefficient effect direction or magnitudes by accounting for spatial data structure. Because most population-level kidney research data are geographically referenced, there is a need for better understanding of geospatial modeling for evaluating associations of individual geolocation with processes of care and clinical outcomes. In this review, we describe common spatial models, provide details to execute these analyses, and perform a case-study to display how results differ when integrating geographic structure. In our case-study, we used U.S. nationwide 2019 chronic kidney disease (CKD) data from Centers for Disease Control and Prevention's Kidney Disease Surveillance System and 2006 to 2010 U.S. Environmental Protection Agency environmental quality index (EQI) data and fit a nonspatial count model along with global spatial models (spatially lagged model [SLM]/pseudo-spatial error model [PSEM]) and a local spatial model (geographically weighted quasi-Poisson regression [GWQPR]). We found the SLM, PSEM, and GWQPR improved model fit in comparison to the nonspatial regression, and the PSEM model decreased the positive relationship between EQI and CKD prevalence. The GWQPR also revealed spatial heterogeneity in the EQI-CKD relationship. To summarize, spatial modeling has promise as a clinical and public health translational tool, and our case-study example is an exhibition of how these analyses may be performed to improve the accuracy and utility of findings.

Kidney Int Rep (2024) **9,** 807–816; https://doi.org/10.1016/j.ekir.2024.01.017
KEYWORDS: chronic kidney disease; geographic information science; GIS; kidney transplantation; spatial epidemiology; spatial modeling
© 2024 International Society of Nephrology. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Geographic information science is a relatively new theoretical and methodological approach to evaluating epidemiological policy and research aims, only becoming widely used since the broader application of geographic information systems (GIS; i.e., geographical computer programs) in the mid-1990s.[1] Since then, GIS and spatial analysis have been utilized extensively to address research in many health care contexts, including chronic conditions (namely, cancer, obesity, diabetes, heart disease, among others)[2–5]; however, there is a paucity of research literature incorporating geospatial epidemiological methods in the context of kidney disease, including CKD and kidney transplantation. In cancer research specifically, GIS methods have revolutionized disease surveillance or study of novel exposures and have had a transformational impact on targeting real-world interventions and policy.[2,6–10] Most existing epidemiological spatial studies of kidney disease, transplantation, and outcomes use descriptive spatial methods, such as disease mapping and cluster analyses with only a handful of studies applying associative methods.[11–21] Many previous reviews have described in depth the importance of geography and strengths of spatial analysis in epidemiology, and the central concept is that health

Correspondence: R. Blake Buchalter, Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, JJN-446, 9211 Euclid Avenue, Cleveland, Ohio 44106, USA. E-mail: buchalr@ccf.org

Received 25 October 2023; revised 19 December 2023; accepted 8 January 2024; published online 23 January 2024

Table 1. Brief summary of comparative advantages, disadvantages, and limitations of various types of modeling in population health research

	Type of model			
		Global spatial models		
Comparative advantage/disadvantage characteristics	Traditional regressions	Spatially lagged models	Spatial error models	Local spatial models (e.g., GWR)
Easier implementation and interpretation	X			
Stronger methodological development	X			
Stronger literature base	X			
Accounting for spatial structure of data		X	X	X
Improved reduction of type I and type II error		X	X	X
Require larger sample size of areal units				X
Require higher geographic continuity of areal units				X
Directly measure local geographic differences in modeled relationships				X

GWR, geographically weighted regression.

conditions and outcomes are heavily influenced by where individuals live and/or work.[22-24] Without adjusting for the inherent spatial distribution of the population and risk factors linked to geographic location, traditional (nonspatial) associative modeling becomes biased due to violation of model assumptions, potentially resulting in bias via misrepresentations of coefficient direction and magnitude of effects and underestimation of standard errors.[25] These issues can lead to flawed results and interpretation, which may be particularly problematic for those risk factors and exposures with more complicated relationships with kidney outcomes. A summarized description of comparative advantages, disadvantages, and limitations of traditional and spatial modeling in epidemiology can be found in Table 1.

The primary goal of this overview is not to discuss strengths and weaknesses of previous spatial population-level kidney research, but instead to display the potential importance of accounting for geography in modeling through application of spatial models and comparisons with traditional statistical modeling. We describe common and useful types of spatial models in epidemiology, show how to execute these analyses, and provide examples of how results in epidemiological analyses of population kidney data shift when integrating geographic structure to models. We also discuss how wider application of spatial methods for the study of kidney disease and outcomes may better inform translational public health and clinical and translational science in the field.

Types of Geospatial Regression

The aim of geospatial regression in epidemiology is to integrate and/or account for the geographic distribution of population characteristics into associative modeling with health conditions,[22] which is achieved by directly or indirectly targeting spatial autocorrelation (also known as spatial dependence or clustering) in spatial health data (i.e., outcome and/or predictor values that vary across space).[25] There are many types of associative spatial models, but they can be categorized into 2 broader groups, namely "global" spatial models and "local" spatial models.[26,27]

Global spatial models are named as such because they account for spatial autocorrelation across the entire study area via spatial parameters, but resulting model coefficients do not vary across space (i.e., global results represent predictor-outcome relationships representing the entire geographic extent of the study area).[26] Global spatial models commonly used in epidemiological research include classical spatial autoregressive models, such as SLMs and spatial error models (SEMs), among others.[28] Spatial autoregressive models account for geographic space via inclusion of spatial model terms, which reduce or remove spatial autocorrelation.[29] SLMs incorporate spatial structure (a lag or weighted average of neighboring values) via the outcome variable, whereas SEMs incorporate spatial structure into model residuals (error terms).[25] SLMs have the components of the traditional regression and include a spatial representation of the outcome variable on the right-hand side of the model equation. The SEM applies the basic linear regression formula but accounts for clustering in the residuals with a different model term. It is generally best practice to perform both SLM and SEM models to aid in identifying whether spatial relationships are better adjusted for by accounting for the influence neighboring outcome values (SLM) or by accounting for clustering in model residuals (SEM).

Conversely, local spatial models, also known as spatially varying coefficients models, do not use spatial model terms and instead construct separate model equations for specified spatial zones (such as for a county and its 10 closest neighbors). Local spatial models, including variations of the commonly used geographically weighted regression (GWR) model, produce local coefficients showing local variation in predictor-outcome relationships. GWR models are useful for examining how predictor-outcome associations vary

across geographic areas while adjusting for related variables. GWR models have been used in chronic disease epidemiology extensively, including for cancer, diabetes, and heart disease[30–32]; however, they are comparatively rare in epidemiological research of kidney conditions and outcomes.[13,19] GWR has been identified as a promising translational public health tool due to its unique ability, along with other spatially varying coefficients models, of identifying areas of disparate risk factor or exposure-outcome relationships that are adjusted for related variables.[5,33–35] The inferential ability of GWR is unclear,[36] but it remains a useful tool for exploration of spatial heterogeneity.

Expanded explanation and model equations of each of these types of models (including nonspatial count regressions) can be found in the Supplementary Material.

Pitfalls of Geospatial Data

Comparative advantages and disadvantages of traditional and geospatial regression can be found in Table 1; however, there are broader limitations of epidemiological spatial data that should be considered in the context of formulating models. The ecological fallacy is a consistent issue in epidemiological analysis of area-level (e.g., county-level) data,[37] where statistical relationships between areal variables may simply be coincidental. The fallacy extends to analyses of epidemiological spatial data, where spatial relationships between variables without plausible explanations for identified relationships can also be coincidental.[38] Therefore, it is important to consider from a theoretical perspective why variables may be associated both statistically and spatially. For example, socioeconomic status, race or ethnicity distributions, and disease burden of U.S. communities do not occur at random across geographic space, and communities near one another may be more similar to one another compared with those further away due to a constellation of factors such as historical racism (e.g., redlining and segregation), availability of amenities and services (e.g., hospitals and groceries), and built environment factors (e.g., locations of highways), among others.[39,40] Another routinely noted pitfall in analysis of spatial data is the modifiable areal unit problem, which is a long-recognized issue in spatial epidemiology where units with arbitrary boundaries, such as counties, are selected to compare health-related variables between areas.[41,42] In short, the modifiable areal unit problem is both a geographic shape and an aggregation problem, where values of disease or risk factor values are aggregated for counties or other units that have arbitrary shapes and sizes. The modifiable areal unit problem is not easily rectifiable, and researchers have observed that it is not likely to be completely resolved ever,[43] yet it remains an important limitation and consideration when pursuing spatial analyses.

Case-Study

Though the population burden and patient-level progression of CKD due to certain factors, such as hypertension and diabetes, are well-established,[44,45] only a limited number of studies have examined relationships with environmental variables, such as ambient air pollution or water quality.[46–49] To examine how integration of spatial structure to modeling affects resulting coefficients, we performed a spatial modeling case-study exploring associations between county-aggregated CKD prevalence and a measure of overall county-level environmental quality, the EQI. EQI has been utilized extensively as a predictor variable in epidemiological literature[50–55] but not in relation to CKD prevalence. We utilized county-level diabetes and hypertension data as adjustment variables and have purposely not interpreted results for these variables because there is strong existing patient-level evidence of relationships with CKD.

Methods

Data

We obtained county-level CKD, diabetes, and hypertension prevalence for Centers for Medicare and Medicaid Services beneficiaries aged 65 and older from the Centers for Disease Control and Prevention's Kidney Disease Surveillance System for 3083 U.S. counties, which is based on claims data from the 5% Medicare random sample from the Centers for Medicare and Medicaid Services. This sample includes beneficiaries who had at least 1 inpatient or outpatient visit during the calendar year, estimated at over 1.8 million annually. Further description of the dataset is available in the existing literature.[56,57] In our analysis, the percentage of the Centers for Medicare and Medicaid Services beneficiary population with CKD was considered the outcome variable, whereas percentage of the Centers for Medicare and Medicaid Services beneficiary population with (i) diabetes and (ii) hypertension serve as adjustment variables. We obtained county-level data for the main predictor variable, EQI, from the most recent U.S. Environmental Protection Agency data release, representing years 2006 to 2010.[58] Briefly, the EQI is a composite indicator of the overall quality of the ambient environment in a given county, including the air, water, land, built, and sociodemographic domains.[58] Though we focused on environmental quality and certain important adjustment variables, theoretically, any county-level population characteristics could be used for adjustment in our models.

Statistical Analysis

We performed a traditional type of count modeling, quasi-Poisson (QP) regression, along with 2 types of global spatial regression (SLM and PSEM) and a local spatial regression, GWQPR. All analyses were completed in R version 4.3 or GeoDa version 1.2, and maps were produced in both GeoDa version 1.2 and ArcGIS version 2.4.[59-61] Both R and GeoDa are free publicly available software suites with GIS capability, whereas ArcGIS requires licensure. Full R code and case-study datasets can be found at the following Github link: https://github.com/spatialepidemiology/spatial_kidney_review.

Count Modeling

We fit a county-level QP general additive model (GAM) with CKD counts as the outcome, county population 65 years and older as the offset, EQI as the main predictor, and diabetes and hypertension percentage as adjustment variables. The overdispersion-resistant QP GAM count model was selected over a negative binomial model to allow for comparison with the GWQPR. Geographically weighted negative binomial regression is not yet developed for open-source statistical software, such as R. We did not pursue zero-inflated versions of the count modeling, because only 0.6% of counties had counts of zero. For comparison purposes, we computed goodness-of-fit measures, including deviance R^2 and 2 indicators of spatial autocorrelation in model residuals: global and local Moran's I. Global Moran's I values were computed in R, whereas local Moran's I clustering and outlier maps were computed in GeoDa software. We produced maps of the local Moran's I values of deviance residuals for comparison with the spatial models. Further details, explanations, and model specifications for the nonspatial QP model can be found in the Supplementary Material.

Global Spatial Modeling

For the 2 global geospatial models (SLM and PSEM), a necessary first step was constructing the spatial structure of the dataset. For the SLM, we selected a k-nearest neighbors spatial weights matrix to form the dataset spatial structure. For the PSEM, we decomposed the spatial geometry into spatial coordinates for inclusion in a tensor product smooth of thin plate regression splines. Details, explanations, and specifications for creating the spatial data structure can be found in the Supplementary Material. We fit a spatially lagged auto-covariate QP GAM (SLM) and a tensor product smooth spatial QP GAM (PSEM) using the same base formula as the nonspatial QP model. SLMs and SEMs produce spatial coefficient parameters (lag coefficient for SLM and spatial spline coefficients for PSEM), which reflect spatial dependence in the data. In the SLM, when the lag coefficient is greater than zero, it indicates that counties are expected to have higher predictor variable values if, on average, their neighbors have higher values. We computed the same goodness-of-fit measures for the global spatial models as the nonspatial model. Local Moran's I values of residuals were mapped for the 2 global spatial models. Further details, explanations, and model specifications for SLM and PSEM can be found in the Supplementary Material.

Local Spatial Modeling

We implemented a generalized GWQPR using the same base model formula as for the QP GAM. Details, explanations, and specifications of the GWQPR can be found in the Supplementary Material. We computed the same goodness-of-fit measures for the GWR as for the QP GAM and global spatial models. Local Moran's I values of residuals and multiple testing-adjusted local coefficients were then mapped. We also produced a forest plot directly comparing resulting coefficients for EQI from each of the models (nonspatial QP GAM, SLM, PSEM, and GWQPR).

Results

Count Modeling Results

Results from the nonspatial QP GAM showed that worsening EQI was associated with higher county-level CKD prevalence (prevalence ratio [PR]: 1.013; 95% confidence interval: 1.010–1.016). In other words, for every 1 unit increase (worsening) of the EQI, county-aggregated CKD prevalence among the Medicare population significantly increased by 1.3% (Table 2). Global Moran's I results showed significant clustering in the residuals (Moran's I: 0.304; $P < 0.0001$). Clustering in the residuals is also clear in local Moran's I value map in Figure 1a. Clustering in the residuals indicates that an underlying spatial process or pattern exists in the data that must be accounted for via spatial modeling. Goodness-of-fit metrics showed a deviance R^2 value of 0.88 for the nonspatial QP GAM. Increasing deviance R^2 indicates improving model fit, where R^2 is bounded in a range from zero to one.

Global Spatial Modeling Results

Results from the SLM showed that increasing EQI is associated with higher CKD prevalence in U.S. counties (PR: 1.014; 95% confidence interval: 1.011–1.016) (Table 2). In comparison to the nonspatial count model, the PR for the EQI-CKD in counties was slightly larger (1.014 vs. 1.013), representing a roughly 0.1% increase compared with the nonspatial model (Figure 2). Global Moran's I results for the SLM showed that significant clustering in the residuals was decreased by nearly half in comparison to the linear regression; however, it was still significant (Moran's I: 0.188; $P < 0.0001$), which is

Table 2. County-level quasi-Poisson regression, spatial lag model, pseudo-spatial error model, and geographically weighted quasi-Poisson regression results for associations between environmental quality, adjustment variables, and chronic kidney disease prevalence among Medicare beneficiaries aged 65 and older ($N = 3083$ counties)

	Quasi-Poisson regression	Spatially lagged model	Pseudo-spatial error model	Geographically weighted quasi-Poisson regression		
Parameter	Exponentiated estimate (95% CI)	Exponentiated estimate (95% CI)	Exponentiated estimate (95% CI)	Minimum exponentiated estimate	Median exponentiated estimate	Maximum exponentiated estimate
Intercept	0.215 (0.214–0.216)[a]	0.178 (0.175–0.181)[a]	0.214 (0.213–0.215)[a]	0.213	0.215	0.216
Spatial parameter[b]	–	2.297 (2.126– 2.482)[a]	See note[b]	–	–	–
Parameter	Prevalence ratio (95% CI)	Prevalence ratio (95% CI)	Prevalence ratio (95% CI)	Minimum prevalence ratio	Median prevalence ratio	Maximum prevalence ratio
Environmental Quality Index	1.013 (1.010–1.016)[a]	1.014 (1.011–1.016)[a]	1.007 (1.004–1.010)[a]	1.011	1.012	1.015
Diabetes	0.997 (0.990–1.004)	0.997 (0.990–1.003)	1.013 (1.007–1.019)[a]	0.997	0.998	0.999
Hypertension	1.332 (1.324–1.340)[a]	1.305 (1.298–1.313)[a]	1.304 (1.297–1.311)[a]	1.326	1.328	1.336
Diagnostics						
Deviance R^2	0.88	0.89	0.93		0.88	
Residual global Moran's I value	0.304 ($P < 0.0001$)[a]	0.188 ($P < 0.0001$)[a]	0.026 ($P = 0.0009$)[a]		0.286 ($P = 0.0001$)[a]	

CI, confidence interval; CKD, chronic kidney disease; PSEM, pseudo-spatial error model; SLM, spatially lagged model.
[a]Significant at $P < 0.05$.
[b]Spatial parameter for SLM is a coefficient for a spatially lagged auto-covariate of CKD counts by county offset by the population count of those 65 years or older. Spatial parameter for the PSEM is a tensor product smooth term with 24 spline knots and therefore cannot be represented by a single coefficient. See tensor product smooth contour map (Figure S1) for representation of spatial parameter.

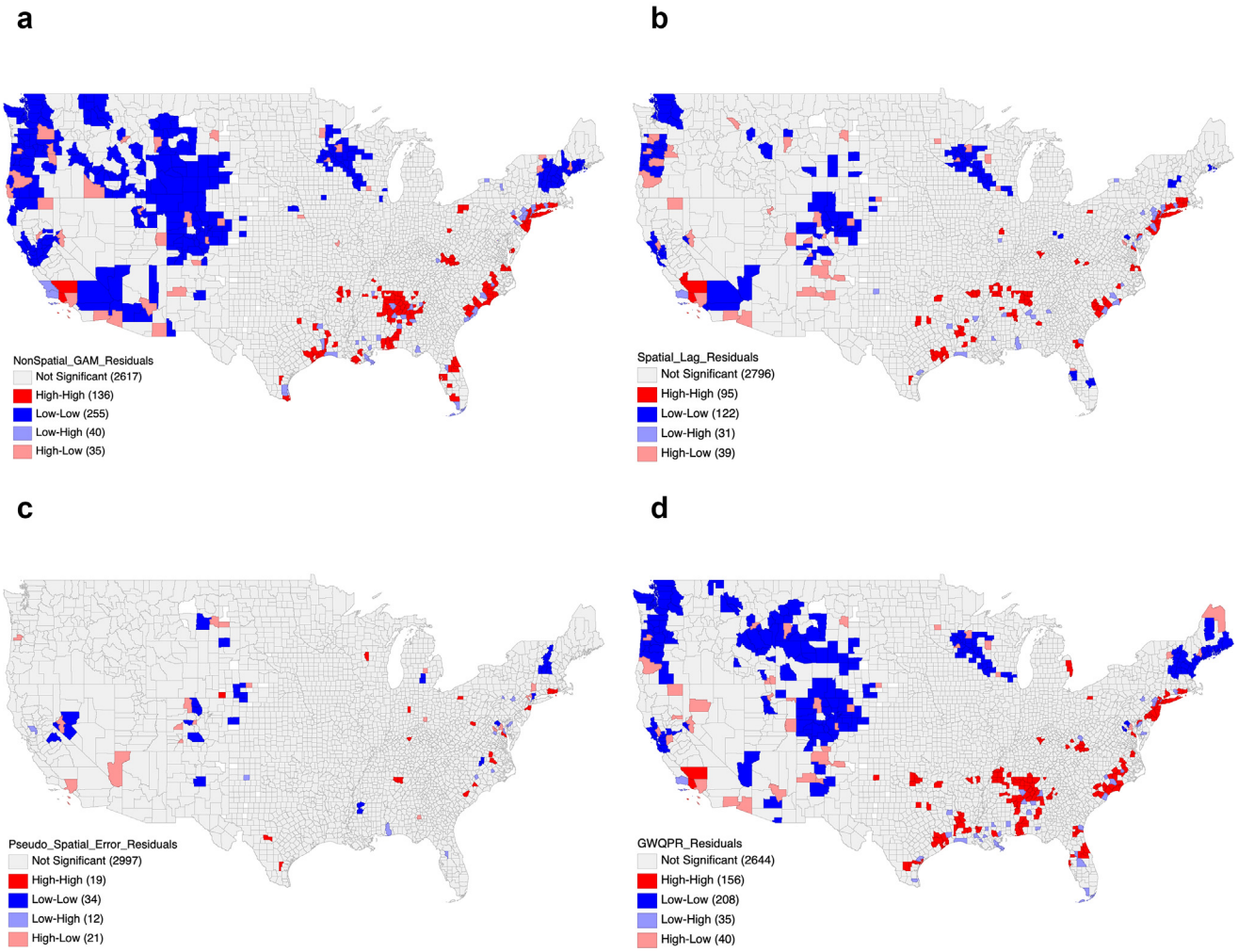

Figure 1. Local Moran's I clustering and outlier results for residuals of (a) nonspatial quasi-Poisson general additive model, (b) spatially lagged model with auto-covariate, (c) pseudo-spatial error model, and (d) geographically weighted quasi-Poisson regression at alpha <0.01. High-High (clustering): significantly higher values than expected and surrounded by other high values; Low-Low (clustering): significantly lower values than expected and surrounded by other low values; Low-High (outliers): significantly lower values than expected and surrounded by high values; High-Low (outliers): significantly higher values than expected and surrounded by low values.

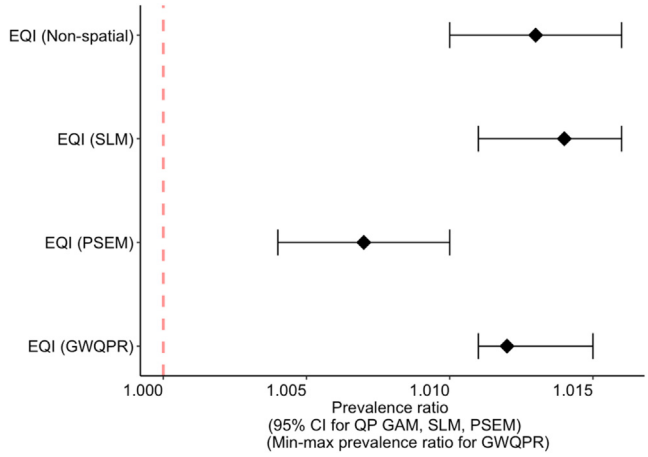

Figure 2. Forest plot comparing EQI-CKD prevalence ratios for statistical and spatial models with 95% confidence interval for nonspatial quasi-Poisson regression, SLM, and PSEM; and minimum-maximum prevalence ratio for the GWQPR. Estimate for GWQPR represents median coefficient estimate. GWQPR prevalence ratio should not be interpreted in the context of statistical significance. CKD, chronic kidney disease; EQI, environmental quality index; GWQPR, geographically weighted quasi-Poisson regression; PSEM, pseudo-spatial error model; SLM, spatially lagged model.

clear in the local Moran's I clustering and outlier map (Figure 1b). A significant exponentiated lag coefficient of 2.297 indicated that counties near to one other are more likely to have higher predictor variable values, on average, if neighboring counties have higher predictor values. Goodness-of-fit metrics for the SLM showed a strong deviance R^2 value of 0.89.

The PSEM also showed that increasing EQI is significantly associated with higher CKD prevalence in U.S. counties (PR: 1.007; 95% confidence interval: 1.002–1.008) (Table 2). The PR for EQI was roughly half in the PSEM in comparison to the nonspatial model. Comparisons of PSEM coefficients with nonspatial QP and SLM coefficients for EQI can be seen graphically in Figure 2. Global Moran's I results for the PSEM showed significant clustering in the residuals less extreme than the nonspatial QP GAM and the SLM (Moran's I: 0.090; $P < 0.0001$), which is again clear in the local Moran's I clustering and outlier map (Figure 1c). The highly significant smooth term of the PSEM also indicated significant spatial dependence in the residuals, which is clear in the contour map of exponentiated partial effects of spatial coordinates on CKD prevalence (Supplementary Figure S1). Goodness-of-fit metrics for the PSEM again had a strong deviance R^2 value of 0.93. Among the nonspatial QP and global spatial models, the PSEM had the best fit based on deviance R^2, whereas the nonspatial model had the worst fit.

Local Spatial Modeling Results

Unlike the nonspatial count model and global spatial models, the GWQPR does not produce a single coefficient for each predictor variable, and instead coefficients for each county can be aggregated in the form of minimum, median, and maximum coefficients. The median PR for the relationship between EQI and CKD prevalence was 1.012, whereas the minimum coefficient was 1.011 and the maximum coefficient was 1.015 (Table 2). Mapping results of the relationships between EQI and CKD burden show spatial heterogeneity. PRs for the EQI-CKD relationship were increasingly positive moving from eastern to western counties (Figure 3). Global Moran's I results for the GWR showed significant clustering in the residuals (Moran's I: 0.286; $P = 0.0001$), yet were still improved over the nonspatial model. With a deviance R^2 of 0.88, the GWQPR explained approximately the same percent of deviance as the nonspatial model. Local Moran's I results indicated clustering and outliers were visually slightly reduced in comparison to the nonspatial model (Figure 1d).

Taken together, all spatial models improved upon the base QP regression, as measured by statistical reductions in global Moran's I values and visual shrinkage of clustering in local Moran's I values. The GWQPR model also provided the ability to produce local coefficient maps to inform geographic differences in predictor-outcome relationships. Comparing the EQI model coefficients in Figure 2, one can see the effect shifts caused by the introduction of spatial modeling. Methodologically, our case-study results show the importance of testing a variety of spatial models.

Discussion

As both consumers and investigators of research with geospatial exposures and population health aims, we should strive to integrate rigorous spatial methods to improve our understanding of complex data. This is particularly important with the increasing recognition of ecological and environmental factors that contribute to health outcomes, as well as the opportunity to have more targeted interventions that are cognizant of structural barriers that can differ across the country. Spatial methods in regression modeling can be highly effective tools for effectively describing relationships between predictor variables and outcomes in geographically referenced population-level epidemiological research of kidney conditions and outcomes. Our case-study is not only an exhibition of how to flexibly perform these types of analyses for predictors of CKD prevalence, but also a display of the potential strengths of spatial modeling. In comparison with nonspatial regression, our best fitting global spatial model (PSEM) decreased the PR in the relationship between EQI and CKD burden by nearly half, indicating that accounting for spatial dependency in CKD burden across counties substantially changed resulting

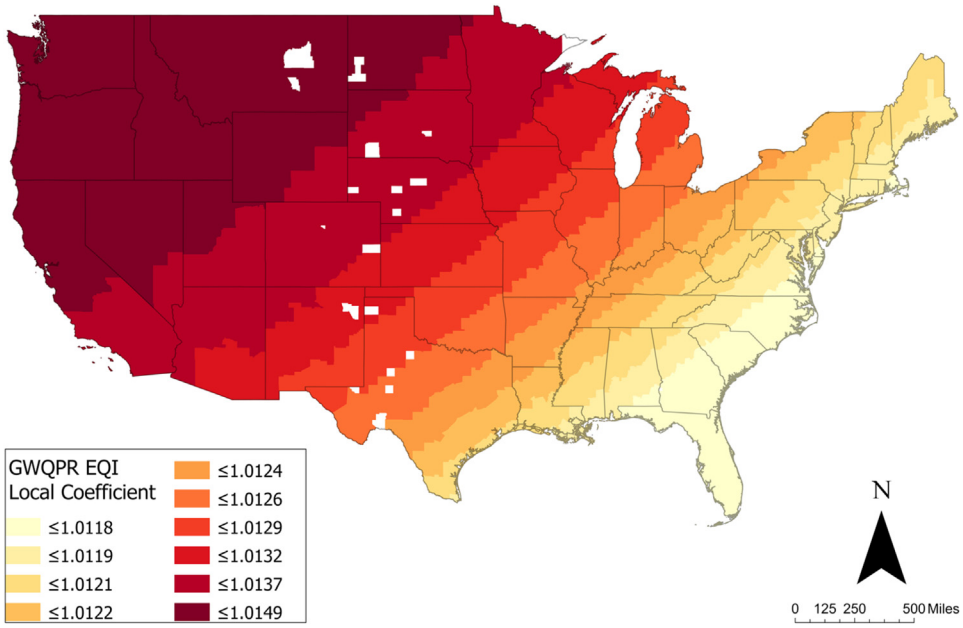

Figure 3. Quantile map of local prevalence ratios for county-level environmental quality index-CKD prevalence association. White space signifies missingness in Medicare data. CKD, chronic kidney disease.

model coefficients in our case-study. In the geographically weighted model, the PR for the EQI-CKD prevalence association is clearly spatially heterogeneous across U.S. counties. In other words, without applying spatial modeling to account for the geographic patterns present in our case-study data, predictor-outcome relationships were mis-estimated and did not integrate spatial variations into resulting coefficients. We have provided a visual representation of model differences via a forest plot in Figure 2, which displays the shifts in PRs for EQI by model type. Though effect shifts of this size may or may not be important translationally, the results are important from a methodological perspective, indicating a need for translating spatial considerations into patient-level analyses, such as through hierarchical spatial models. Though beyond the scope of an introduction to spatial regression in the context of population-level kidney research, Bayesian spatial modeling is gaining in popularity in the epidemiological and clinical realms with the release and expansion of the "INLA" and "rstan" R packages over the past 10 to 15 years.[62-64] Bayesian spatial regressions are powerful tools that support broad flexibility in modeling by allowing borrowed strength across geographic and temporal domains.[64]

The effect of environmental exposures on CKD prevalence is beginning to be understood,[46-49] and those environmental factors with less clearly understood relationships with CKD, such as heat exposure, air pollution, water quality, and other environmental toxins or exposures, have potential to be clarified more accurately via spatial models. Environmental exposures are not only difficult to measure with reasonable certainty but also can vary strongly temporally (e.g., by season or exposures due to one-time disasters) and across geography.[65-69] Though not presented in this review, epidemiological spatiotemporal models[70,71] would also be useful to examine relationships between geographically and temporally referenced environmental exposures and patient-level kidney disease outcomes. Integration of spatial and temporal structure into analysis of patient data is typically completed via hierarchal modeling,[72-74] which are methods that have been used previously in the epidemiological kidney literature but remain rare.[20,21] Spatial autoregressive (i.e., SLM and SEM models) and GWR models can both be extended to include hierarchical structure for patient data accomodation.[75,76] Hierarchical spatial modeling in epidemiological kidney research is a clear next step to begin providing clinical and translational science–oriented evidence that accounts for patient residence and other geographically referenced variables.

GWQPR modeling and local coefficient mapping contributed novel findings and improved the predictive accuracy of our case-study results, displaying that modeled relationships between environmental quality and CKD burden vary across U.S. counties. These local models have seldom been applied in epidemiological kidney research[13,19]; however, they have potential to serve as a powerful translational public health tool to identify subregional areas where further study can be targeted to reduce disease burden. We found that the association between EQI and CKD prevalence increases moving from eastern to western U.S. counties; however, this result should be interpreted only from an exploratory perspective due to the fact that

significance testing was not possible. Considering that these results are model-adjusted, they are more robust than simple mapped comparisons, further suggesting the utility of local spatial modeling. Going forward, these methods could also be applied to improve understanding of factors driving the identified geographic variations in kidney disease outcomes, including outcomes after transplantation outcomes across space.[12]

Our case-study had several limitations, including its cross-sectional ecological nature (lack of patient-level data) and lack of adjustment for demographic factors. Considering these limitations, care should be taken in directly translating our results in a causal framework. Instead, our results should be used as an exhibition of how global and local spatial modeling can be utilized to approach population-level kidney research in new ways. Broad integration of spatial or spatiotemporal data structures into modeling of kidney conditions or outcomes could have strong impacts on translational public health as well as clinical and translational science in the field by clarifying predictor-outcome relationships with improved accuracy and informing local targeting of funds, programs, and interventions. As both consumers and investigators of research with geospatial exposures and population health research aims, we should carefully evaluate studies with potential opportunities for methodological improvement and strive to integrate rigorous spatial methods to improve our understanding of complex data that is meaningful to prospective patient care and health care policy.

DISCLOSURE

RBB is supported in part by National Institutes of Health Award Number T32CA094186. SM receives grant funding from Kidney Transplant Collaborative and the National Institutes of Health; and personal fees from Kidney International Reports and Health Services Advisory Group outside of the submitted work. SM also reports consultancy for Sanofi, HSAG; patents or royalties at Columbia University; advisory or leadership role as Deputy Editor of Kidney International Reports (ISN), Chair of UNOS Data advisory committee, Member of ASN Quality committee, and National Faculty Chair of ETCLCJDS. JDS declared no competing interests.

SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Types of Spatial Regressions.

Case Study Supplementary Methods.

Figure S1. Contour map of the exponentiated partial effects of spatial coordinates (x, y) on CKD prevalence via thin plate spline tensor product smooth.

Supplementary Reference.

REFERENCES

1. Clarke KC, McLafferty SL, Tempalski BJ. On epidemiology and geographic information systems: a review and discussion of future directions. *Emerg Infect Dis*. 1996;2:85–92. https://doi.org/10.3201/eid0202.960202
2. Sahar L, Foster SL, Sherman RL, et al. GIScience and cancer: state of the art and trends for cancer surveillance and epidemiology. *Cancer*. 2019;125:2544–2560. https://doi.org/10.1002/cncr.32052
3. Jia P, Cheng X, Xue H, Wang Y. Applications of geographic information systems (GIS) data and methods in obesity-related research. *Obes Rev*. 2017;18:400–411. https://doi.org/10.1111/obr.12495
4. Cuadros DF, Li J, Musuka G, Awad SF. Spatial epidemiology of diabetes: methods and insights. *World J Diabetes*. 2021;12:1042–1056. https://doi.org/10.4239/wjd.v12.i7.1042
5. Tabb LP, Ortiz A, Judd S, Cushman M, McClure LA. Exploring the spatial patterning in racial differences in cardiovascular health between blacks and whites across the United States: the REGARDS study. *J Am Heart Assoc*. 2020;9:e016556. https://doi.org/10.1161/JAHA.120.016556
6. Openshaw S, Charlton M, Craft AW, Birch JM. Investigation of leukaemia clusters by use of a geographical analysis machine. *Lancet*. 1988;331:272–273. https://doi.org/10.1016/S0140-6736(88)90352-2
7. Brody JG, Aschengrau A, McKelvey W, Rudel RA, Swartz CH, Kennedy T. Breast cancer risk and historical exposure to pesticides from wide-area applications assessed with GIS. *Environ Health Perspect*. 2004;112:889–897. https://doi.org/10.1289/ehp.6845
8. Turnbull BW, Iwano EJ, Burnett WS, Howe HL, Clark LC. Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *Am J Epidemiol*. 1990;132(suppl 1):136–143. https://doi.org/10.1093/oxfordjournals.aje.a115775
9. Lian M, Struthers J, Schootman M. *Comparing GIS-Based Measures in Access to Mammography and Their Validity in Predicting Neighborhood Risk of Late-Stage Breast Cancer. Published Online 2012*.
10. Poulstrup A, Hansen HL. Use of GIS and exposure modeling as tools in a study of cancer incidence in a population exposed to airborne dioxin. *Environ Health Perspect*. 2004;112:1032–1036. https://doi.org/10.1289/ehp.6739
11. Bowe B, Xie Y, Xian H, Lian M, Al-Aly Z. Geographic variation and US county characteristics associated with rapid kidney function decline. *Kidney Int Rep*. 2017;2:5–17. https://doi.org/10.1016/j.ekir.2016.08.016
12. Buchalter RB, Huml AM, Poggio ED, Schold JD. Geographic hot spots of kidney transplant candidates wait-listed post-dialysis. *Clin Transpl*. Published online September 14, 2022. 2022;36:e14821. https://doi.org/10.1111/ctr.14821
13. VanDervort DR, López DL, Orantes CM, Rodríguez DS. Spatial distribution of unspecified chronic kidney disease in el Salvador by crop area cultivated and ambient temperature. *MEDICC Rev*. 2014;16:31–38. https://doi.org/10.37757/MR2014.V16.N2.6
14. Hansson E, Mansourian A, Farnaghi M, Petzold M, Jakobsson K. An ecological study of chronic kidney disease in five Mesoamerican countries: associations with crop and heat. *BMC Public Health*. 2021;21:840. https://doi.org/10.1186/s12889-021-10822-9

15. Cao Y, Stewart K, Kalil R. Geographic patterns of end-stage renal disease and kidney transplants in the Midwestern United States. *Appl Geogr*. 2016;71:133–143. https://doi.org/10.1016/j.apgeog.2016.05.001
16. Mohan S, Mutell R, Patzer RE, Holt J, Cohen D, McClellan W. Kidney transplantation and the intensity of poverty in the contiguous United States. *Transplantation*. 2014;98:640–645. https://doi.org/10.1097/TP.0000000000000125
17. Soret S, McCleary KJ, Wiafe SA, Rivers PA, Montgomery SB. Access to kidney transplantation in California: a geographic information systems (GIS) assessment. Accessed January 8, 2024. https://proceedings.esri.com/library/userconf/health01/papers/hc01_p04a/hc01_p04a.html
18. Cannon RM, Anderson DJ, MacLennan P, et al. Perpetuating disparity: failure of the kidney transplant system to provide the most kidney transplants to communities with the greatest need. *Ann Surg*. 2022;276:597–604. https://doi.org/10.1097/SLA.0000000000005585
19. Chan TC, Fan IC, Liu MSY, Su MD, Chiang PH. Addressing health disparities in chronic kidney disease. *Int J Environ Res Public Health*. 2014;11:12848–12865. https://doi.org/10.3390/ijerph111212848
20. Kürüm E, Nguyen DV, Banerjee S, Li Y, Rhee CM, Şentürk D. A Bayesian multilevel time-varying framework for joint modeling of hospitalization and survival in patients on dialysis. *Stat Med*. 2022;41:5597–5611. https://doi.org/10.1002/sim.9582
21. Li Y, Nguyen DV, Banerjee S, et al. Multilevel modeling of spatially nested functional data: spatiotemporal patterns of hospitalization rates in the US dialysis population. *Stat Med*. 2021;40:3937–3952. https://doi.org/10.1002/sim.9007
22. Kirby RS, Delmelle E, Eberth JM. Advances in spatial epidemiology and geographic information systems. *Ann Epidemiol*. 2017;27:1–9. https://doi.org/10.1016/j.annepidem.2016.12.001
23. Ostfeld RS, Glass GE, Keesing F. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends Ecol Evol*. 2005;20:328–336. https://doi.org/10.1016/j.tree.2005.03.009
24. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environ Health Perspect*. 2004;112:998–1006. https://doi.org/10.1289/ehp.6735
25. Anselin L. Spatial regression. The SAGE Handbook of Spatial Analysis. SAGE Publications Ltd; 2009:255–276.
26. Fotheringham AS, Brunsdon C. Local forms of spatial analysis. *Geogr Anal*. 1999;31:340–358. https://doi.org/10.1111/j.1538-4632.1999.tb00989.x
27. Anselin L. Under the hood issues in the specification and interpretation of spatial regression models. *Agric Econ*. 2002;27:247–267. https://doi.org/10.1111/j.1574-0862.2002.tb00120.x
28. Osei F. Current statistical methods for spatial epidemiology: a review. *Austin Biom Biostat*. 2014;1:7.
29. Dray S, Legendre P, Peres-Neto PR. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol Modell*. 2006;196:483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015
30. Gebreab SY, Roux AVD. Exploring racial disparities in CHD mortality between blacks and whites across the United States: a geographically weighted regression approach. *Health Place*. 2012;18:1006–1014. https://doi.org/10.1016/j.healthplace.2012.06.006
31. Siordia C, Saenz J, Tom SE. An introduction to macro- level spatial nonstationarity: a geographically weighted regression analysis of diabetes and poverty. *Hum Geogr*. 2012;6:5–13. https://doi.org/10.5719/hgeo.2012.62.5
32. Goovaerts P, Xiao H, Adunlin G, et al. Geographically weighted regression analysis of percentage of late-stage prostate cancer diagnosis in Florida. *Appl Geogr*. 2015;62:191–200. https://doi.org/10.1016/j.apgeog.2015.04.018
33. Sharma A. Exploratory spatial analysis of food insecurity and diabetes: an application of multiscale geographically weighted regression. *Ann GIS*. 2023;29:485–498. https://doi.org/10.1080/19475683.2023.2208199
34. Oluyomi AO, Gunter SM, Leining LM, Murray KO, Amos C. COVID-19 community incidence and associated neighborhood-level characteristics in Houston, Texas, USA. *Int J Environ Res Public Health*. 2021;18:1495. https://doi.org/10.3390/ijerph18041495
35. Balamurugan A, Delongchamp R, Bates JH, Mehta JL. The neighborhood where you live is a risk factor for stroke. *Circ Cardiovasc Qual Outcomes*. 2013;6:668–673. https://doi.org/10.1161/CIRCOUTCOMES.113.000265
36. Wheeler DC. Geographically weighted regression. In: Fischer MM, Nijkamp P, eds. *Handbook of Regional Science*. Springer; 2021:1895–1921.
37. Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *Am J Public Health*. 1994;84:819–824. https://doi.org/10.2105/ajph.84.5.819
38. Spatial aggregation and the ecological fallacy. *Chapman Hall CRC Handb Mod Stat Methods*. 2010;2010:541–558. https://doi.org/10.1201/9781420072884-c30
39. Tung EL, Cagney KA, Peek ME, Chin MH. Spatial context and health inequity: reconfiguring race, place, and poverty. *J Urban Health*. 2017;94:757–763. https://doi.org/10.1007/s11524-017-0210-x
40. Xiao WQ. The road to racial justice: resolving the disproportionate health burden placed on communities of color by highway pollution. *Colum Hum Rts LRev*. 2020;52:911.
41. King PE. Problems of spatial analysis in geographical epidemiology. *Soc Sci Med D Med Geogr*. 1979;13:249–252. https://doi.org/10.1016/0160-8002(79)90046-7
42. Gehlke CE, Biehl K. Certain effects of grouping upon the size of the correlation coefficient in census tract material. *J Am Stat Assoc*. 1934;29:169–170. https://doi.org/10.1080/01621459.1934.10506247
43. Manley D. Scale, aggregation, and the modifiable areal unit problem. In: Fischer MM, Nijkamp P, eds. *Handbook of Regional Science*. Springer Verlag; 2014:1157–1171.
44. Ku E, Lee BJ, Wei J, Weir MR. Hypertension in CKD: core curriculum 2019. *Am J Kidney Dis*. 2019;74:120–131. https://doi.org/10.1053/j.ajkd.2018.12.044
45. Pyram R, Kansara A, Banerji MA, Loney-Hutchinson L. Chronic kidney disease and diabetes. *Maturitas*. 2012;71:94–103. https://doi.org/10.1016/j.maturitas.2011.11.009
46. Obrador GT, Schultheiss UT, Kretzler M, et al. Genetic and environmental risk factors for chronic kidney disease. *Kidney Int Suppl*. 2017;7:88–106. https://doi.org/10.1016/j.kisu.2017.07.004

47. Soderland P, Lovekar S, Weiner DE, Brooks DR, Kaufman JS. Chronic kidney disease associated with environmental toxins and exposures. *Adv Chronic Kidney Dis*. 2010;17:254–264. https://doi.org/10.1053/j.ackd.2010.03.011
48. Glaser J, Lemery J, Rajagopalan B, et al. Climate change and the emergent epidemic of CKD from heat stress in rural communities: the case for heat stress nephropathy. *Clin J Am Soc Nephrol*. 2016;11:1472–1483. https://doi.org/10.2215/CJN.13841215
49. Wu MY, Lo WC, Chao CT, Wu MS, Chiang CK. Association between air pollutants and development of chronic kidney disease: a systematic review and meta-analysis. *Sci Total Environ*. 2020;706:135522. https://doi.org/10.1016/j.scitotenv.2019.135522
50. Jagai JS, Messer LC, Rappazzo KM, Gray CL, Grabich SC, Lobdell DT. County-level cumulative environmental quality associated with cancer incidence. *Cancer*. 2017;123:2901–2908. https://doi.org/10.1002/cncr.30709
51. Gearhart-Serna LM, Hoffman K, Devi GR. Environmental quality and invasive breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2020;29:1920–1928. https://doi.org/10.1158/1055-9965.EPI-19-1497
52. Nance D, Rappazzo KM, Jensen ET, et al. Increased risk of eosinophilic esophagitis with poor environmental quality as measured by the environmental quality index. *Dis Esophagus*. 2021;34:doab041. https://doi.org/10.1093/dote/doab041
53. Jian Y, Messer LC, Jagai JS, et al. Associations between environmental quality and mortality in the contiguous United States, 2000-2005. *Environ Health Perspect*. 2017;125:355–362. https://doi.org/10.1289/EHP119
54. Gray CL, Lobdell DT, Rappazzo KM, et al. Associations between environmental quality and adult asthma prevalence in medical claims data. *Environ Res*. 2018;166:529–536. https://doi.org/10.1016/j.envres.2018.06.020
55. Patel AP, Jagai JS, Messer LC, et al. Associations between environmental quality and infant mortality in the United States, 2000-2005. *Arch Public Health*. 2018;76:1–11. https://doi.org/10.1186/s13690-018-0306-0
56. Vijayasarathi A, Loehfelm T, Duszak R, Hawkins CM. JOURNAL CLUB: Radiologists' online identities: what patients find when they search radiologists by name. *AJR Am J Roentgenol*. 2016;207:952–958. https://doi.org/10.2214/AJR.16.16437
57. Centers for Medicare and Medicaid Services. Doctors and clinicians National downloadable file. Accessed October 21, 2023. https://data.cms.gov/provider-data/dataset/mj5m-pzi6
58. United States Environmental Protection Agency. Environmental quality index-technical report (2006-2010) (final, 2020). 2020. Accessed November 6, 2023. https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=350545
59. ESRI. *ArcGIS Pro 2.4.0. Redlands CA*. Environmental Systems Research Institute; 2019.
60. Anselin L, Syabri I, Kho Y. GeoDa: An introduction to spatial data analysis. *Geograph Analysis*. 2006;38(1):5–22.
61. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2023. Accessed February 12, 2024. https://www.R-project.org
62. Lawson AB. Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. 3rd ed. Routledge & CRC Press; 2018.
63. Martino S, Rue H. Implementing approximate Bayesian inference using Integrated Nested Laplace Approximation: a manual for the inla program. Norwegian University of Science and Technology. 2009. Accessed November 20, 2023. https://www.math.ntnu.no/preprint/statistics/2008/S2-2008.pdf
64. Moraga P. *Geospatial Health Data: Modeling and Visualization With R-INLA and Shiny*. Chapman and Hall/CRC Press; 2019.
65. Aly NA, Casillas G, Luo YS, et al. Environmental impacts of Hurricane Florence flooding in eastern North Carolina: temporal analysis of contaminant distribution and potential human health risks. *J Expo Sci Environ Epidemiol*. 2021;31:810–822. https://doi.org/10.1038/s41370-021-00325-5
66. Committee on Human and Environmental Exposure Science in the 21st Century; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Research Council. *Exposure Science in the 21st Century: A Vision and a Strategy*. National Academies Press; 2012.
67. Bell ML, Dominici F, Ebisu K, Zeger SL, Samet JM. Spatial and temporal variation in PM2. 5 chemical composition in the United States for health effects studies. *Environ Health Perspect*. 2007;115:989–995. https://doi.org/10.1289/ehp.9621
68. Bell ML, Ebisu K, Peng RD. Community-level spatial heterogeneity of chemical constituent levels of fine particulates and implications for epidemiological research. *J Expo Sci Environ Epidemiol*. 2011;21:372–384. https://doi.org/10.1038/jes.2010.24
69. Fletcher BA, Lin S, Fitzgerald EF, Hwang SA. Association of summer temperatures with hospital admissions for renal diseases in New York State: a case-crossover study. *Am J Epidemiol*. 2012;175:907–916. https://doi.org/10.1093/aje/kwr417
70. Meliker JR, Sloan CD. Spatio-temporal epidemiology: principles and opportunities. *Spat Spatio-Temporal Epidemiol*. 2011;2:1–9. https://doi.org/10.1016/j.sste.2010.10.001
71. Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with R-INLA. *Spat Spatio-Temporal Epidemiol*. 2013;7:39–55. https://doi.org/10.1016/j.sste.2013.07.003
72. Mobley LR, Kuo TM, Urato M, Subramanian S. Community contextual predictors of endoscopic colorectal cancer screening in the USA: spatial multilevel regression analysis. *Int J Health Geogr*. 2010;9:1–11. https://doi.org/10.1186/1476-072X-9-44
73. MacNab YC. Hierarchical Bayesian modeling of spatially correlated health service outcome and utilization rates. *Biometrics*. 2003;59:305–315. https://doi.org/10.1111/1541-0420.00037
74. Carlin BP, Banerjee S. Hierarchical multivariate CAR models for spatio-temporally correlated survival data. *Bayesian Stat*. 2003;7:45–63.
75. Dong G, Harris R. Spatial autoregressive models for geographically hierarchical data structures. *Geogr Anal*. 2015;2:173–191. https://doi.org/10.1111/gean.12049
76. Chen DR, Truong K. Using multilevel modeling and geographically weighted regression to identify spatial variations in the relationship between place-level disadvantages and obesity in Taiwan. *Appl Geogr*. 2012;32:737–745. https://doi.org/10.1016/j.apgeog.2011.07.018