Published in final edited form as:

Spat Spatiotemporal Epidemiol. 2011 June ; 2(2): 91–101. doi:10.1016/j.sste.2011.03.002.

Effects of varying temporal scale on spatial models of mortality patterns attributed to pediatric diarrhea

S Leyk^{1,2,*}, BJJ McCormick¹, and JR Nuckols^{1,3}

BJJ McCormick: ben.mccormick@nih.gov; JR Nuckols: john.nuckols@colostate.edu

¹ Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, DHHS, Bethesda, Maryland, USA

² Department of Geography, University of Colorado, Boulder, Colorado, USA

³ Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, Colorado, USA

Abstract

Public health data is often highly aggregated in time and space. The consequences of temporal aggregation for modeling in support of policy decisions have largely been overlooked. We examine the effects of changing temporal scale on spatial regression models of pediatric diarrhea mortality patterns, mortality rates and mortality peak timing, in Mexico. We compare annual and decadal level univariate models that incorporate known risk factors. Based on normalized sums of squared differences we compare between annual and decadal coefficients for variables that were significant in decadal models. We observed that spurious relationships might be created through aggregating time scales; eliminating inter-annual variation and resulting in inflated model diagnostics. In fact, variable selection and coefficient values can vary with changing temporal aggregation. Some variables that were significant at the decadal level were not significant at the annual level. Implications of such aggregation should be part of risk communication to policy makers.

Keywords

Temporal aggregation; spatial models; diarrhea; uncertainty; disease modeling

© 2011 Elsevier Ltd. All rights reserved.

*Corresponding author: Stefan Leyk, Department of Geography, University of Colorado at Boulder, UCB 260, Boulder, CO, 80305, USA (Tel: (303) 492-4695. FAX: (303) 492-7501), SL: stefan.leyk@colorado.edu.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

1st author conceptualized the study, designed the methodology, created the models, and conducted the statistical evaluation; he wrote and revised this paper. 2nd author was involved in designing the conceptual idea and statistical evaluation, co-developed the methodological steps and revised the paper. 3rd author was involved in the conceptual design and strategic planning of the research, as well as in the methodological development of the modeling approaches; he revised the paper. All authors read and approved the final manuscript.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 Introduction

Research using exploratory analyses (Strina et al., 2003; Hamer et al., 1998), time series analyses (Kale et al., 2004) and spatial analyses (Balk et al., 2003; Pande et al., 2008) has greatly improved our understanding of the outbreak etiology of diarrheal disease. Spatial and spatio-temporal patterns, spatial dependence, clusters and hotspots of diarrheal disease have been investigated and observed (Jepsen et al., 2009; Chaikaew et al., 2009; Leyk et al., in press). Purely temporal variations in spatial disease patterns and their association with risk factors have also been described (Kelly-Hope et al., 2008; Anyamba et al., 2006; Alonso et al., 2006). However, when disease outcome (registry) data are compiled in the context of health policy, they are typically aggregated over time in order to derive measures at the same time interval as reported causal risk factors (Abellan et al., 2008). The period of aggregation can be on the order of every 5 or 10 years, for example when census data are released. The integration of the time dimension into mapping models has been proposed to improve the interpretation of risk patterns (Waller et al., 1997; Abellan et al., 2008). In this paper we explore the effect of changing the temporal scale on spatial models of mortality patterns of pediatric diarrhea (age < 5 years) in Mexico over the period 1979 to 1988. We define mortality patterns as the timing of peak mortality and the magnitude of annual mortality rates (cases per 100,000 population in this age group). We compare models fitted using data temporally aggregated over 10 years, “decadal model”, with models using data on an annual basis. The aim is to assess differences in spatial distribution of mortality patterns and their associations with known or expected risk factors due to temporal aggregation of observed disease data. The implications of using results from such models in setting international and national global health policy are discussed.

2 Data and data preparation

Mortality data and mortality patterns

We obtained Mexican mortality data from 1979 to 1988 for each state ($N=32$) in Mexico from the Instituto Nacional de Estadística Geografía e Informática (INEGI). The data were compiled from death certificates and cause classified using the International Classification of Diseases (ICD; <http://www.who.int/classifications/icd>). We used all deaths of children younger than 5 years of age and classified ICD-9 codes 001–009 (intestinal infectious disease) in our study. We calculated monthly mortality rates of diarrhea per 100,000 children in this age group based on state-specific age-classified population estimates, which were projected to each year using linear interpolation between census statistics from 1980 and 1990 (INEGI, 2009). We calculated the annual mortality rates MR_a for each state, and then averaged them over the ten year study period to produce an average annual decadal mortality rate MR_{dec} . We determined the month with the highest number of deaths from the raw data for each year (annual mortality peak timing; MPT_a) and averaged a scalar value ranging from 1.0 (May 1st) to 12.99 (April 30th) representing the peak timing over the study period (decadal mortality peak timing; MPT_{dec}). We selected May as starting point of this digital scale because it represented the earliest peak month we observed in some years during the study period.

Climate Data

We obtained monthly climatic data for Mexico and the same period as our mortality data set from interpolated climate maps (Mitchell and Jones, 2005) with pixel size 0.5° (latitude) by 0.5° (longitude), or approximately $2,500 \text{ km}^2$. The climatic variables used were: number of wet days, precipitation in mm, averages of daily mean, maximum, and minimum temperatures ($^\circ\text{C}$), diurnal temperature range and vapor pressure in hectopascal. We derived climate variable values for each state using the average of the pixel intersecting the location

of its capital city and the adjoining pixels which intersect the state area. In most cases this allowed us to link the averaged climate variable to the area of the most populated regions in each state. For each climate variable, we derived peak timing, the month when the value has its maximum, mean values and maximum (peak) values for each year, and used these values to create an average variable over the decadal modeling period (1979–1988).

Elevation

We obtained 1km spatial resolution elevation data (*dem*) from the Shuttle Radar Topography Mission (SRTM) digital elevation model (<http://www2.jpl.nasa.gov/srtm/>), and calculated average elevation within the aforementioned climate grid cells used in our study.

Population and population density

As mentioned above, state population by age were obtained from the census statistics of 1980 and 1990 (INEGI, 2009). Based on the annual state-specific projections and decadal averages of total population we calculated population density ($popd_a$, $popd_{dec}$; people per km²). We used spatial gridded databases from the Gridded Population of the World (GPW) and the Global-Rural Mapping Project (GRUMP) (<http://sedac.ciesin.columbia.edu/gpw/index.jsp>) (CIESIN, 2005) of the year 1990 and with a spatial resolution of 30'' (approximately 1 km²), to create demographic variables, including proportion of urban area (*propur*), or percentage of urban population (*popurb*) for each state. Because the data were only available for 1990, we assumed them to be constant over our study period. We also derived a variable (*popslp*) representing urban population growth rate for the largest Mexican localities within each state using decennial population counts (1930–1980) (INEGI, 2009), assuming the rate of growth prior to our study period could influence known risk factors for enteric disease.

Socio-economic variables

We obtained the proportion of population (households) in each state without residential access to drinking water (*nowater*), sewage treatment facilities (*nosew*) and electricity (*noelectr*) from the 1980 and 1990 census data (INEGI, 2009), and projected these statistics to each year of the study using a linear function. We also obtained and projected the proportion of the population older than 15 years that is illiterate (*illit15*), using the same method. We then calculated a decadal (averaged annual) value for each of these variables for each state.

3 Methods

We used a three-step approach to identify, develop and evaluate models for annual (MR_a) and decadal mortality rates (MR_{dec}), and corresponding mortality peak timing (MPT_a , MPT_{dec}) for the period 1979–1988: (i) a preliminary analysis to determine whether spatial or non-spatial model procedures should be used, (ii) development of spatial models (annual, decadal) for MR, and MPT, and (iii) an analysis of model coefficients; testing for consistency of relationships between mortality patterns and predictive risk factors in annual models, and for the deviation of these annual relationships from the corresponding ones in a decadal model.

3.1 Preliminary analysis: Spatial autocorrelation, spatial clustering and non-spatial baseline models

We calculated standard deviation in annual mortality patterns, MPT_a and MR_a , in order to quantify their variability during the study period. We visually evaluated the spatio-temporal distribution of mortality patterns across Mexican states for the period of our decadal model (1979–1988) by creating maps of MPT_{dec} and MR_{dec} (Fig. 1). The mapped data suggest a

spatio-temporal pattern, with the peaks in mortality earliest in the states around Federal District (state 9) (Alonso et al., in press) and increasingly time-shifted with increasing distance to the North-West and South-East from Federal District. We then tested for spatial autocorrelation in mortality patterns and its persistence over time by calculating the Moran's I statistic (Moran, 1950) on the decadal and annual time scale. As shown in Table 1, we found highly significant ($p < 0.01$) spatial autocorrelation for the decadal mortality patterns as well as for most of the annual patterns indicating that the outcome variables are not independent. Based on methods to calculate scan statistics reported by Kulldorff and others (Kulldorff, 1997; Kulldorff et al., 1998 and 2009) we used SaTScan software (Kulldorff, 2010) to test for purely spatial clusters in the temporally aggregated mortality patterns, MR_{dec} and MPT_{dec} , as well as for space-time clusters in the annual patterns, MR_a and MPT_a . We employed circular scanning windows of varying sizes up to a maximum defined by capturing 40% of the study population (children, age < 5 years) to evaluate observed and expected mortality patterns, across the modeling units (centroids of state polygons) in our study area. For the spatio-temporal cluster analysis, we added a cylinder of the same radius, with varying height corresponding to time (years) to calculate a space-time scan statistic. Based on likelihood ratio tests we tested for statistically significant ($p < 0.05$) and purely spatial clusters of low and high values in decadal mortality patterns as well as space-time clusters in annual distributions of mortality patterns.

In a final determination as to whether we should employ a spatial or non-spatial model we fitted univariate non-spatial baseline models for each risk factor using Ordinary Least Square (OLS) estimation and tested for spatial autocorrelation in the model residuals. We observed spatial dependency in the model residuals i.e., significant spatial autocorrelation in the model residuals based on Moran's I indicating non-random error terms, and thus the need to use a spatial model (Anselin and Griffith, 2005).

3.2 Spatial models of decadal and annual mortality patterns and evaluation of model coefficients

Based on the results of our preliminary analyses, we selected spatial regression models as our modeling procedure on both time scales. We fitted spatial lag models, which use a spatially lagged variable, and spatial error models that account for spatial patterns in the error structure based on Maximum Likelihood (ML) estimations (Anselin, 1988a, 2002). To compute spatial lag and spatial error variables, and to estimate the autoregressive spatial coefficients for the model, we created spatial weight matrices based on first-order rook contiguity, which takes into account the presence of commonly shared borders of neighboring spatial modeling units (LeSage and Pace, 2009), which in our study were polygons delineating political boundaries of the states in Mexico.

To determine the appropriate adjustment, spatial lag or spatial error, in our models we used a sequence of simple and robust Lagrange Multiplier (LM) diagnostics for a missing spatially lagged dependent variable (LM(lag)) and for error dependency (LM(error)) (Anselin, 1988b). If the results of both simple LM tests were found to be statistically significant and the results of the robust LM tests were not, we selected the final model based on the most significant test result of the simple LM tests. If the results of the robust LM tests were also found to be significant, the final model was selected based on the most significant robust test value (Anselin, 1988b).

Based on the LM tests described above we found spatial lags to be more appropriate than spatial error adjustment for all the spatial regression models described below. We therefore fitted univariate spatial lag models for MPT_{dec} and MR_{dec} by treating each risk factor as an independent variable. We then fitted univariate annual models for MPT_a and MR_a using the same risk factors. We compared statistically significant ($p < 0.05$) risk factors derived from

the decadal models for MPT_{dec} and MR_{dec} in terms of the percent of annual models for which the same risk factors were determined to be statistically significant. We tested all spatial regression models for residual normality (Jarque-Bera test) and heteroscedasticity (Breush-Pagan and Koenker-Bassett tests), as well as for multicollinearity (Belsley et al., 1980). We evaluated the model performance based on the coefficient of determination (R^2), a Log-Likelihood measure, and the Akaike Information Criterion (AIC) (Akaike, 1974).

To assess the effect of changing temporal scales on models of MPT and MR, we examined the deviation (in per cent) of annual coefficients from decadal coefficients in these models ($NSSD_{dec}$) using the equation:

$$NSSD_{dec} = 100 \times \left(\frac{\sum (b_{dec} - b_a)^2}{\sum (b_{dec})^2} \right) = 100 \times \left(\frac{\sum (b_{dec} - b_a)^2}{(10 \times (b_{dec})^2)} \right) \quad (1)$$

where,

$NSSD_{dec}$ = normalized sum of squared differences in per cent of b_a from b_{dec}

b_{dec} = decadal model coefficient

b_a = annual model coefficients

$\sum (b_{dec} - b_a)$ = The summed differences (SD) over the individual years within the decadal period, in our case 1979–1988.

We calculated the SD measure to determine if annual coefficient values systematically under- or over-estimated the decadal coefficient values. To assess variability in annual model coefficients across years, we calculated the normalized sum of squared differences between the coefficient values b_a and the best-fit line of these coefficient values $y(b_a)$ ($NSSD_a$). These measures provide a quantifiable basis for estimating the variation of annual model coefficients and thus the consistency of the identified annual relationships for models of both MPT and MR. We graphed the coefficient values of the annual models and corresponding standard errors in estimating the coefficients in order to compare them graphically over the time period as well as to the values of decadal model coefficients.

4 Results

The standard deviations (Table 2) indicate where the variations in both MPT_a (e.g., States 15, 19, 28, 29) and MR_a (e.g., States 1, 3, 29) values are highest across the years. The annual spatial distributions of observed MR_a and MPT_a are illustrated in Figures 2 and 3, respectively.

We observed statistically highly significant spatial autocorrelation ($p < 0.01$) in the observed decadal and annual mortality patterns in all years, except 1988 for MPT_a (Table 1). When aggregated to a decade, Moran's I for MPT_{dec} and MR_{dec} was = 0.70 and 0.43, respectively. When considered an independent annual data, the range in Moran's I was 0.27 – 0.71 for MPT_a , and 0.29 – 0.50 for MR_a .

Maps resultant from our analyses for clusters of high and low observed mortality rate (MR) and peak timing (MPT) based on scan statistics, indicate high consistency in both spatial and spatio-temporal scenarios at the two temporal scales used in our study. The most likely spatial clusters in high and low decadal mortality rate (1979–1988; Figures 4C, D) were spatially consistent with space-time clusters derived using annual mortality rates covering up to 8 of the 10 years (Figures 4A, B). Similarly, the location of most likely clusters in observed low (Figure 5C,D) and high (Figure 5A,B) values of MPT were generally consistent in both the spatial and spatio-temporal analyses, but for shorter time periods than for MR. The exception was the state of Tamaulipas, in Northeast Mexico, which was

identified in clusters of high values consistently each year for the period 1979–1984, but was not in the spatial cluster of high decadal MPT values. For both variables, MR and MPT, some variation could be detected for secondary clusters e.g., different locations or only single years in space-time clusters.

The results of our comparison of statistically significant risk factors derived by decadal versus annual univariate spatial lag models of mortality peak timing (MPT) and mortality rate (MR) are presented in Table 3. For decadal models the regression coefficient and the level of statistical significance are reported; for the annual models the percent years of the decadal modeling period for which each of these risk factors was significant at the $p < 0.05$ level is reported. The statistically significant risk factors meeting the $p < 0.05$ criteria in the decadal model and in the highest percent of annual models for MPT were maximum temperature peak value (90%), elevation (60%), and mean annual vapor pressure (40%). The statistically significant risk factors meeting this criteria in the decadal model and in the highest percent of annual models for MR were percent illiteracy in people older than 15 years (100%), percent households with no electricity (90%), percent urban population (70%), and maximum temperature peak value (70%).

Table 4 shows for each decadal model for outcomes MPT_{dec} and MR_{dec} , two univariate models for which the spatial lag variable and the risk factor were significant ($p < 0.05$), the model diagnostics indicated high model performance (based on R^2 , Log Likelihood and AIC), and for which no spatial autocorrelation was found in the model residuals. We present model results based on the independent variables population density (*popd*; $R^2 = 0.72$) and elevation (*dem*; $R^2 = 0.71$) for modeling MPT_{dec} and percentage of households without access to electricity (*noelectr*; $R^2 = 0.44$) and peak value of maximum monthly temperature (*tmaxPV*; $R^2 = 0.44$) for modeling MR_{dec} . The spatial lag variables (lag) were highly significant ($p < 0.01$) in all four decadal models (Table 4). All other temperature-related variables behaved similar to, but were less significant than, *tmaxPV* and *tmaxPT*, respectively.

When testing the models shown in Table 4 in more detail at the annual level we found the accuracy measures (R^2) for the MPT_a models ranged from 0.17 in 1988 to 0.78 in 1981 for *popd* and from 0.16 in 1988 to 0.78 in 1981 for *dem*, with a mean R^2 over the decade of 0.34 for both independent variables. Similarly for the MR_a models R^2 ranged from 0.26 in 1980 to 0.53 in 1987 for *tmaxPV* and from 0.30 in 1980 to 0.53 in 1987 for *noelectr*, with a mean R^2 of 0.41 for both variables. Annual and decadal model coefficients using the same risk factors are graphically illustrated in Figures 6 and 7 for comparison. The annual coefficients of *noelectr* and the lag variable in the annual MR_a models showed high similarity to the decadal model coefficients and limited variation across the years; the coefficients of *tmaxPV* and the corresponding lag variable in contrast show an oscillating trend around the decadal coefficient (Fig. 6) which could be an indication for a gradually changing yet mostly significant relationship across the years. Interestingly the models, which incorporated *illit15* as independent variable, showed the highest model performance in most years but their lag variables were tested not significant and did not contribute to the reduction of variance in estimating MR_a . The coefficients of *popd* and *dem* and corresponding lag variables show higher variation and higher standard errors in the annual models for MPT_a than the ones for MR_a resulting in systematic deviations from the decadal MPT_{dec} model coefficients (Fig. 7).

We found higher variability in the coefficients of the variables in the annual MPT_a models (NSSD_a between 7% and 18%) than in those of the annual MR_a models (NSSD_a between 1 and 7%; Table 4) when quantified by the NSSD_a measure to the best-fit line (Table 4),

We found distinct differences between annual and decadal coefficients for variables in the MPT models (all $NSSD_{dec} \geq 16\%$, up to 149%), with annual coefficients of the lag variable (see section 3.2) and independent variables mostly lower in value than the decadal values indicating underestimation (Table 4, Figure 7). The coefficients in the annual MR models resulted in much less deviation (most $NSSD_{dec}$ values $\leq 11\%$; Table 4, Figure 6).

5 Discussion and Concluding Remarks

The collection of public health data, in particular in developing countries, is often intended for policy making on a regional or national scale, but rarely done with a view to elucidating mechanistic relationships with contemporary socio-economic or environmental correlates. As a consequence researchers in disease mapping often use summary disease data that is aggregated in the temporal or spatial domain (Abellan et al., 2008) and not originally intended for modeling disease outcomes. Statistically, aggregation of data can yield spurious associations and patterns based on small samples, but superior data is rarely available and tests of the effects of data aggregation are rarely shown (Beale et al., 2008).

In this paper we examine the impact of changing temporal scale of disease data on the variation in correlative relationships, and thus the historical patterns of mortality from pediatric diarrhea in Mexico in a spatial model framework. Our intention was to provide important insights into the dynamics of relationships between environmental risk factors and pediatric diarrhea detectable on the given spatial scale (states) whose neglect can lead to the derivation of spurious relationships (Abellan et al., 2008) or to effects of “ecological fallacy” introduced during seasonality analysis (Fisman, 2007). As has been reported by other studies (Pande et al., 2008; Jepsen et al., 2009; Chaikaew et al., 2009; Alonso et al., in press) our preliminary analysis showed distinct spatial patterns and spatial autocorrelation of the disease data as well as in non-spatial model residuals. However moving beyond the spatial patterns (Tiao et al., 1990; Weatherhead et al., 1998), we found that temporal aggregation of data affected the spatial regression models on the annual and decadal scale differently. The computation of mean values to derive MPT_{dec} in this paper was found appropriate due to low degrees of dispersion of the annual values. However, if the values would be more dispersed more robust methods would be needed to reflect their circularity. Statistically, temporal aggregation of the disease data for the decadal model resulted in higher model performance, suggesting averaging the disease burden and thereby eliminating variability over time, improves the fit of the independent variables. However, annual models over the same time period exhibited considerable inter-annual variability in the regression coefficients of the independent and spatial lag variables (Figs. 6 and 7). This finding demonstrates that the use of statistical tests for goodness of fit based on temporally aggregated data can mask underlying and significant variation in the consistency of correlative relationships between risk factors and the disease outcome of interest. The finding could have important consequences in terms of interpretation of risk factor models for policy development due to the ambiguity generated by different time scales concerning relative importance of “significant” variables in a model of aggregated data. For example, we found several risk factors that exhibited significant associations with mortality patterns from pediatric diarrhea in the decadal spatial regression models, but not in most annual regression models (Table 3). These risk factors include *tmaxPT*, *popd*, and *popdurb* for annual mortality peak timing in annual models (MPT_a); and *wetPV* in the models for annual mortality rate (MR_a). Other risk factors were significant in both the decadal and most annual models (e.g. for MPT: *tmaxPV*; and for MR: *noelectr*, *illit15* or *popurb*).

Our approach allowed quantification of the effect of replacing inter-annual variability of outcomes and conditions by a decadal average on the model structure. For example, in the model for mortality peak timing (MPT) the annual coefficients systematically deviated from

the decadal coefficients due to a high variation. As a result, decadal model coefficients substantially over- or underestimated their counterparts in the annual models (Fig. 7). This finding is further verified by our observation that the annual spatial distributions of clusters of MPT_a , as derived from space-time scan statistics, change considerably between years (Fig. 5), while the independent variables, *dem* and *popd*, did not, or changed only to a limited degree. Without elucidation of these differences, model diagnostics typically used in policy analysis would have indicated adequate decadal models, which could lead to misinterpretations if used for policy decisions on resource allocation. In this specific case risk factors that have no significant relationship to mortality peak timing on the annual time scale would have been used to make this policy decision.

Less discrepancy was found between the results produced by a series of annual models and a decadal model of mortality rate (MR) when using the risk factor *noelectr*, which was statistically significant in the decadal model and in 90% of the annual models for MR_a (Table 3). With the exception of the model intercepts, which show a gradual decrease during the considered period, the annual coefficients in this model exhibited relatively low inter-annual variation, and unremarkable deviation from their counterpart coefficients in the decadal model (Fig. 6). Based on these findings, one could surmise that the decadal model was appropriate for policy development and resource allocation. Similar interpretations could be derived for models that incorporate *tmaxPV* as independent variable (Table 3). However, upon closer scrutiny, we found the coefficients in the annual models for this risk factor exhibited a distinct oscillating pattern around the counterpart coefficient in the decadal model, resulting in a sinusoidal pattern (Fig. 6). Thus, the model structure of the decadal model varies substantially from those of the annual models which nevertheless were significant in 70% of the years.

In summary, modeling can be used to identify risk factors that correlate with a disease outcome such as pediatric mortality due to diarrhea. The use of aggregated data is often a necessity to pursue immediate public health questions, in particular in policy decisions regarding developing countries. However, it is essential to appropriately communicate implications and limitations with regard to the treatment of the data used to non-expert audiences. In this case study of policy-relevant disease modeling, the consequence of temporal aggregation is not necessarily a model with poor explanatory power; indeed our decadal models were statistically superior. The more insidious consequence is the difference in the mechanistic interpretation of the decadal model, a result that would remain unnoticed if there was an over-reliance on the statistics with little regard for the interpretation. In our case we have identified that different temporal scales may involve different risk factors or substantial changes in the relationships between the disease outcome and risk factors. From a policy perspective, the interpretation of such models may yield different advice, which emphasizes the necessity to incorporate the considered temporal scale into risk communication. In this study, we have presented tools that could prove useful in scrutiny of aggregated data if the non-aggregated information can be ascertained.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The Etiology, Risk Factors and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development Study (MAL-ED) is carried out as a collaborative study supported by the Bill & Melinda Gates Foundation, the Foundation for the NIH and the National Institutes of Health/ Fogarty International Center. SL and JRN were also funded by interagency personnel agreements between the Fogarty International Center (NIH)

and the University of Colorado at Boulder and Colorado State University, respectively. The authors thank two anonymous reviewers for their constructive critics.

References

- Abellan JJ, Richardson S, Best N. Use of Space–Time Models to Investigate the Stability of Patterns of Disease. *Environ Health Perspect.* 2008; 116:1111–1119. [PubMed: 18709143]
- Akaike H. A new look at the statistical model identification. *IEEE Trans Automatic Control.* 1974; 19(6):716–723.
- Alonso WJ, Acuña-Soto R, Giglio R, Nuckols JR, Leyk S, Schuck-Paim C, Viboud C, Miller MA, McCormick BJJ. Spatio-temporal Patterns of diarrhoeal mortality in México. *Epidemiol Infect.* in press.
- Alonso, WJ.; Kelly-Hope, L.; Viboud, C.; Hirano, EW.; Miller, MA. Changes in the spatial dynamics of seasonal diarrhea in Mexico in 1979–2001. 55th ASTM Annual Meeting; 2006; Atlanta. 2006. 55th ASTM Annual Meeting Abstract Book
- Anselin L. Lagrange Multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geogr Anal.* 1988b; 20:1–17.
- Anselin, L. *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers; Dordrecht, The Netherlands: 1988a.
- Anselin L. Under the hood. Issues in the specification and interpretation of spatial regression models. *Agric Econ.* 2002; 27:247–67.
- Anselin L, Griffith DA. Do Spatial Effects Really Matter in Regression Analysis? *Papers in Regional Science.* 1988; 65(1):11–34.
- Anyamba A, Chretien JP, Small J, Tucker CJ, Linthicum KL. Developing global climate anomalies suggest potential disease risks for 2006–2007. *Int J Health Geogr.* 2006; 5:60. [PubMed: 17194307]
- Balk D, Pullum T, Storeygard A, Greenwell F, Neuman M. Spatial analysis of childhood mortality in West Africa. *Population Space and Place.* 2003; 10(33):175–216.
- Beale L, Abellan JJ, Hodgson S, Jarup L. Methodologic Issues and Approaches to Spatial Epidemiology. *Environ Health Perspect.* 2008; 116:1105–1110. [PubMed: 18709139]
- Belsley, DA.; Kuh, E.; Welson, ER. *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity.* New York: Wiley; 1980.
- Center for International Earth Science Information Network (CIESIN), Columbia University. Small Area Estimates of Poverty and Inequality (SAEPI) database. Palisades, NY: CIESIN, Columbia University; 2005. Available at <http://sedac.ciesin.columbia.edu/povmap>
- Chaikaew N, Tripathi NK, Souris M. Exploring spatial patterns and hotspots of diarrhea in Chiang Mai, Thailand. *Int J Health Geogr.* 2009; 8:36. [PubMed: 19552822]
- Fisman DN. Seasonality of Infectious Diseases. *Annu Rev Public Health.* 2007; 28:127–43. [PubMed: 17222079]
- Hamer D, Simon J, Thea D, Keusch G. Childhood diarrhea in Sub-Saharan Africa. *Child Health Research Project Special Report.* 1998; 2:32.
- INEGI - Instituto Nacional de Estadística Geografía e Informática (INEGI). 2009
- Jepsen MR, Simonsen J, Ethelberg JS. Spatio-temporal cluster analysis of the incidence of *Campylobacter* cases and patients with general diarrhea in a Danish county, 1995–2004. *Int J Health Geogr.* 2009; 8:11. [PubMed: 19228436]
- Kale PL, Andreozzi VL, Nobre FF. Time Series Analysis of Deaths Due to Diarrhoea in Children in Rio de Janeiro, Brazil, 1980–1998. *J Health Popul Nutr.* 2004; 22(1):27–33. [PubMed: 15190809]
- Kelly-Hope LA, Alonso WA, Thiem VD, Canh DG, Anh DD, Lee H, Miller MA. Temporal Trends and Climatic Factors Associated with Bacterial Enteric Diseases in Vietnam, 1991–2001. *Environ Health Perspect.* 2008; 116:7–12. [PubMed: 18197292]
- Kulldorff M. A spatial scan statistic. *Communications in Statistics: Theory and Methods.* 1997; 26:1481–1496.
- Kulldorff, M. *SaTScan v9.0: Software for the spatial and space-time scan statistics.* Silver Spring, MD: Information Management Services; 2010.

- Kulldorff M, Athas W, Feuer E, Miller B, Key C. Evaluating cluster alarms: A space-time scan statistic and brain cancer in Los Alamos. *Am J Public Health*. 1998; 88:1377–1380. [PubMed: 9736881]
- Kulldorff M, Huang L, Konty K. A scan statistic for continuous data based on the normal probability model. *Int J Health Geogr*. 2009; 8:58. [PubMed: 19843331]
- LeSage, JP.; Pace, K. *Introduction to Spatial Econometrics*. Chapman & Hall/CRC; 2009.
- Leyk S, Phillips TP, Smith JM, Nuckols JR. Spatio-temporal Trends of Diarrheal Mortality of Children in Association with Hydrographic Regions of Brazil. *Cartogr Geogr Inf Sci*. in press.
- Mitchell TD, Jones PD. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int J Climat*. 2005; 25:693–712.
- Moran PAP. Notes on Continuous Stochastic Phenomena. *Biometrika*. 1950; 37:17–33. [PubMed: 15420245]
- Pande S, Keyzer MA, Arouna A, Sonneveld BG. Addressing diarrhea prevalence in the West African Middle Belt: social and geographic dimensions in a case study for Benin. *Int J Health Geogr*. 2008; 7:17. [PubMed: 18433488]
- Strina A, Cairncross S, Barreto M, Larrea C, Prado M. Childhood diarrhea and observed hygiene behavior in Salvador, Brazil. *Am J Epidemiol*. 2003; 157(11):1032–1038. [PubMed: 12777367]
- Tiao GC, Reinsel GC, Xu D, Pedrick JH, Zhu X, Miller AJ, DeLuisi JJ, Mateer CL, Wuebbles DJ. Effects of autocorrelation and temporal sampling schemes on estimates of trend and spatial correlation. *J Geophys Res*. 1990; 95(D12):20507–20517.
- Waller LA, Carlin BP, Xia H, Gelfand AM. Hierarchical spatio-temporal mapping of disease rates. *J Am Stat Assoc*. 1997; 92:607–617.
- Weatherhead EC, Reinsel GC, Tiao GC, Meng X-L, Choi D, Cheang W-K, Keller T, DeLuisi J, Wuebbles DJ, Kerr JB, Miller AJ, Oltmans SJ, Frederick JE. Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *J Geophys Res*. 1998; 103(D14):17149–17162.

Figure 1. Observed spatial distribution of decadal mortality patterns across Mexican states for the study period (1979–1988), children under 5 years of age
 (A) Mean mortality peak timing (MPT_{dec}); (B) Mean mortality rate (MR_{dec}); (C) Index of Mexican states used in our study.

Figure 2.
Observed annual mortality rates in Mexican states due to diarrhea, children under 5 years of age (1979–1988)

Figure 3.
Observed peak timing (month) of mortality in Mexican states due to diarrhea, children under 5 years of age (1979–1988)

Figure 4. Spatial and spatio-temporal clusters: Observed mortality rates in Mexican states due to diarrhea, children under 5 years of age (1979–1988)

Likelihood of space-time clusters of low (A) and high (B) annual mortality rates (MR_a); and spatial clusters of low (C) and high (D) mean mortality rates across the study period.

Figure 5. Spatial and spatio-temporal clusters: observed mortality peak timing in Mexican states due to diarrhea, children under 5 years of age (1979–1988)

Likelihood of space-time clusters of low (A) and high (B) values of annual mortality peak timing (MPT_a); and spatial clusters of low (C) and high (D) values of mean mortality peak timing across the study period (MPT_{dec}).

Figure 6. Coefficients of annual spatial lag models for mortality rates due to diarrhea in Mexican states, children under 5 years of age (1979–1988)

Coefficient and intercept values of univariate spatial lag models using percent population without access to electricity, *noelectr*, (left) and peak value of maximum monthly temperature, *tmaxPV*, (right) as independent variables. Dashed horizontal lines indicate decadal model coefficients; circles and dots indicate non-significant and significant coefficients ($p < 0.05$), respectively; vertical lines indicate standard errors of coefficients.

Figure 7. Coefficients of annual spatial lag models for mortality peak timing due to diarrhea in Mexican states, children under 5 years of age (1979–1988)
Coefficient and intercept values of univariate spatial lag models using population density, *popd*, (left) and elevation, *dem*, (right) as independent variables. Dashed horizontal lines indicate decadal model coefficients; circles and dots indicate non-significant and significant coefficients ($p < 0.05$), respectively; vertical lines indicate standard errors of coefficients.

Table 1Spatial autocorrelation^a of mortality patterns for the period 1979–88

Modeling Interval and Time Period	Moran's I for	
	Mortality Peak Timing ^b	Mortality Rate ^b
Decadal		
1979–88	0.70***	0.43***
Annual		
1979	0.27**	0.34**
1980	0.43***	0.29**
1981	0.71***	0.36***
1982	0.50***	0.40***
1983	0.38***	0.42***
1984	0.36***	0.44***
1985	0.38***	0.48***
1986	0.40***	0.44***
1987	0.37***	0.50***
1988	0.10	0.38***

^aStatistical significance: 0.1% level(***), on the 1% level (**) and on the 5% level (*).

^bMPT_{dec}: average annual mortality peak timing (1979–88); MR_{dec}: average annual mortality rate; MPT_a: annual mortality peak timing; MR_a: annual mortality rate.

Mean Mortality Peak Timing and Mortality Rate due to Diarrheal Disease in Mexico over the Period 1979–1988, Children under 5 years of Age

Table 2

State ID	State name	Mean (decadal) Mortality Peak Timing				Mean Annual Mortality Rate (deaths/100,000 ^b)			
		Mean month ^a	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
1	Aguascalientes	8.0	0.94	24.31	10.20				
2	Baja California Norte	10.4	1.07	10.58	5.05				
3	Baja California Sur	10.2	1.62	15.22	10.06				
4	Campeche	7.7	1.25	17.33	4.41				
5	Chiapas	7.1	0.32	30.93	6.43				
6	Chihuahua	8.9	0.88	14.06	8.01				
7	Coahuila De Zaragoza	7.3	1.06	13.27	6.21				
8	Colima	8.4	1.07	21.16	6.72				
9	Distrito Federal	5.7	0.67	16.13	4.74				
10	Durango	8.1	1.45	5.78	2.74				
11	Guanajuato	7.5	1.18	33.74	7.85				
12	Guerrero	7.9	0.57	17.77	6.64				
13	Hidalgo	6.3	0.48	22.59	6.87				
14	Jalisco	7.8	0.63	16.62	6.27				
15	Mexico	6.2	1.81	23.12	6.59				
16	Michoacan de Ocampo	7.0	0.94	19.65	6.44				
17	Morelos	7.4	0.52	14.79	3.28				
18	Nayarit	8.1	0.88	9.28	3.26				
19	Nuevo Leon	7.9	1.97	7.58	2.30				
20	Oaxaca	7.6	0.52	45.04	8.60				
21	Puebla	6.4	0.84	30.14	7.22				
22	Queretaro de Arteaga	7.4	1.35	27.46	6.96				
23	Quintana Roo	8.1	1.66	11.05	4.74				
24	San Luis Potosi	8.3	1.62	17.32	3.64				
25	Sinaloa	8.4	1.71	6.98	3.38				
26	Sonora	9.5	1.43	15.33	5.11				
27	Tabasco	7.6	0.70	29.40	8.67				

Mean (decadal) Mortality Peak Timing					
Mean Annual Mortality Rate (deaths/100,000 ^b)					
State ID	State name	Mean month ^a	Std Dev	Mean	Std Dev
28	Tamaulipas	7.1	2.92	10.06	4.01
29	Tlaxcala	6.7	2.00	30.55	9.41
30	Veraacruz-Llave	7.2	0.63	17.08	4.63
31	Yucatan	7.5	0.71	18.22	5.68
32	Zacatecas	7.2	1.03	15.15	5.49

^aThe numbers shown represent digital months. In the analysis we transformed them to scalar values such that May 1st (5.0 in the table)= 1.0, June 1st (6.0 in the table) = 2.0, etc. May represented the earliest peak month in most years and was thus selected as starting point of the digital scale of peak timing.

^bDenominator is per 100,000 children under age 5

Table 3

Comparison of statistically significant risk factors in decadal versus annual univariate spatial lag models of peak timing (MPT) and rate (MR) in mortality due to diarrhea, Mexican children under the age of 5 years (1979–88)

Risk factor	Definition	y=MPT _{dec}		y=MPT _a		y=MR _{dec}		y=MR _a	
		decadal model coefficient ^a	% annual models with p<0.05	decadal model coefficient ^a	% annual models with p<0.05	decadal model coefficient ^a	% annual models with p<0.05	decadal model coefficient ^a	% annual models with p<0.05
illitl5	% illiterate population > 15 y.		20		100	0.515***			
noelectr	% household w/ no electricity		0	0.281*	90				
nowater	No access to drinking water		0		20				
nosew	No access to sewage facilities		10		10				
popslp	Max. urban growth 1930–80	0.0218*	10		0				
popd	Population density	-0.00023**	30		0				
popurb	% urban population		10	-0.158*	70				
popdurb	Urban population density	-0.000175**	20		0				
precPT	Precipitation peak timing		10	-4.693*	30				
precPV	Precipitation peak value		10		0				
tmaxPT	Max. temperature peak timing	0.293**	30	-3.462**	60				
tmaxPV	Max. temperature peak value	0.113***	90	-0.648*	70				
vapAV	Vapor pressure annual mean	-0.122**	40		0				
wetPV	# wet days peak value	-0.050*	30	-0.345*	40				
dem	Elevation	-0.00033**	60		10				

^aStatistical significance < 0.1% level (***), <1% level (**) or and <5% level (*) or and <0.05 level (hashed).

Table 4

Performance assessment of selected^a univariate spatial lag models for decadal annual mortality peak timing (MPT_{dec}) and mortality mean rate (MR_{dec})

	SE	Coefficient analysis ^c		Model Diagnostics	
		Model coefficient ^b	NSSD _{dec} %	NSSD _g %	
MPT_{dec} I	0.540				
lag	0.758***		16 (-) ^d	8 ^e	R ² 0.72
constant	1.960*				AIC 64.63 (92.8) ^f
popd	-0.00023**		26 (-)	18	Log Likelihood -29.31
MPT_{dec} II	0.550				
lag	0.668***		23 (-)	13	R ² 0.71
constant	2.990**		85 (+)	7	AIC 66.08 (83.46)
dem	-0.00033**		53 (-)	18	Log Likelihood -29.04
MR_{dec} I	6.600				
lag	0.524***		3 (-)	2	R ² 0.44
constant	4.280		44 (+)	5	AIC 222.35 (227.01)
noelectr	0.2807*		3 (+)	1	Log Likelihood -107.17
MR_{dec} II	6.560				
lag	0.518***		3 (-)	1	R ² 0.44
constant	30.090**		11 (+)	4	AIC 221.9 (226.45)
tmaxPV	-0.648*		11 (-)	7	Log Likelihood -106.96

^aDecadal models for which the spatial lag variable and the independent variable were significant (p<0.05)

^bStatistical significance at < 0.1% level(***), <1% level (**) or <5% level (*)

^cNSSD: Normalized sum of squared differences

^dNSSD_{dec}: NSSD of the annual coefficient values from the horizontal line defined by the decadal coefficients ((-) and (+) indicate negative or positive bias, respectively, given by the summed differences (SD))

^eNSSD_g: NSSD between the coefficient values b_g and the best-fit line of these coefficient values y(b_g)

^fThe numbers in parentheses after AIC represents the AIC measure of the Ordinary Least Squares (OLS) baseline model to denote the improvement in model diagnostics by using a spatial lag variable.