Biophysical, infrastructural and social heterogeneities explain spatial distribution of waterborne gastrointestinal disease burden in Mexico City

Andrés Baeza-Castro[1], Alejandra Estrada-Barón[2], Luis A. Bojórquez[2], Hallie Eakin[1], Ana E. Escalante[2]*

[1.] School of Sustainability, Arizona State University, Tempe AZ 85287 USA.

[2.] Laboratorio Nacional de Ciencias de la Sostenibilidad (LANCIS), Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico.

Keywords: flooding, neighborhood effect, giardiasis, spatial statistics, water management.

*Corresponding author: aescalante@iecologia.unam.mx

ABSTRACT

Megacities of the developing world are particularly prone to hydrological hazards and infectious diseases from waterborne pathogens, which disproportionally affect the most vulnerable populations. In this work we aim to elucidate to what extent physical and socio-economic factors related to hydrological hazards (flooding and scarcity) are associated with disease risk in a Megacity. We conducted spatial statistics and multivariate regressions for data sets of two periods (2007-2009 and 2010-2014) to investigate the distribution of gastrointestinal disease incidence in Mexico City, a megacity with more than 8 million people and high inequality in socio-economic and risk to hazards. Results indicate a significant association between the number of sewage-storm water flooding events and higher disease incidences at the center (lowlands) of the city. We also found that at the periphery (highlands), higher incidence is associated with insufficient sewage infrastructure and deficiencies of infrastructure inside houses. Our findings suggest different routes of exposure to waterborne pathogens through the city, associated with different biophysical and socio-economic factors (i.e. deficiencies in the management of the drainage system *vs.* lack of water infrastructure) and their spatial heterogeneity. The significant association between gastrointestinal disease incidence and the heterogeneity in public infrastructure provision supports the need of dialogue between public works and public health sector. This dialogue will be critical to develop mutually supportive managerial policies and spatially tailored interventions aimed to manage the socio-hydrological vulnerability of fast changing megacity.

1 INTRODUCTION

To date, more than 50% of the human population lives in cities, and this percentage is predicted to rise to 70% by the year 2050 (United Nations, 2016). Urbanization has had great impact on improving quality of life, yet poses great challenges for planning and management of the built environment, urban services and economic policy (Dye, 2008; UN-HABITAT, 2017). Urbanization in the developing world is often unregulated and unplanned and occurs most rapidly in the periphery of cities, which are often highly vulnerable to different hazards, such as extreme rainfall, flooding, landslides and water scarcity (Pick and Butler, 1997; IPCC, 2012). Unplanned growth also results in significant spatial heterogeneity in terms of culture, economy, and infrastructure, creating inequalities in the provision of urban services exposure of hazard and responses from managers and infrastructure providers (Jacobi *et al.*, 2010; Ahern, 2013; Pickett *et al.*, 2017; Zhou, Pickett and Cadenasso, 2017). In the vulnerability literature, it has long been understood that risk (i.e., the probability that a potentially damaging event will occur) is a product of both environmental conditions and the social processes that determine vulnerability (Turner *et al.*, 2003). In cities, these social and political processes define the actions of governments and decision-makers that in turn influence the creation of the built environment, the socio-economic heterogeneity and the concomitance vulnerability to hydrological hazards (Eakin *et al.*, 2017).

Despite its relevance for urban planning and health policy decisions, we still lack a full understanding of the role of the built environment, socio-economic attributes and hydrological processes in the production of epidemiological vulnerability in large urban areas (Vlahov *et al.*, 2005, 2007; Santos-Vega, Martinez and Pascual, 2016). In other words, there is a need to investigate the association between socio-hydrological vulnerability and epidemiological outcomes. In particular, there has been a growing interest in understanding how socio-economic heterogeneities found in large cities influence the spread and persistence of infectious diseases, taking into account for example, the spatial distribution of pathogens (Ahern *et al.*, 2005). Waterborne pathogens are of particular interest in the epidemiology literature, as they can be in contact with human populations in many different ways, and cause a variety of infectious diseases (Hunter, 1997). It is therefore of epidemiological importance to understand the role that socio-hydrological processes and water management have on epidemiological outcomes in urban areas.

In this work, we investigated the spatial distribution of risk of gastrointestinal diseases caused by waterborne pathogens, and the influence of socio-hydrological risk factors in Mexico City (MC) on this distribution. This work thus contributes to "socio-hydrology", a growing body of research "aimed at understanding the dynamics and co-evolution of coupled human-water systems" (Sivapalan, Savenije and Blöschl, 2012). Under this framework, socio-hydrological risk is defined as the potential for harm resulting from the combination of both hydrological processes (e.g., precipitation, runoff, subsidence) and social processes (e.g., the built environment, infrastructural conditions, human behavior and decision-making).

We hypothesize that in MC the spatial distribution of gastrointestinal disease incidence (GDI) reflects the spatial heterogeneity of the factors that shape the vulnerability of the city to hydrological hazards that determine the way people are exposed to water-borne pathogens. To test this hypothesis, we conducted a set of spatial statistical analyses on data derived from public records of GDI over the period between 2007 and 2014. The results suggest strong differentiation in the type of hydrological risk, and the factors underlying the association of water and health risk at the center and periphery of the city. We conclude that the city providers and institutions of health and water must foster dialogue and initiatives to consider how they can tackle together the challenges of waterborne-related health problems and water infrastructure management.

2 DATA AND METHODS

2.1 Data collection and handling

All data were georeferenced at the scale of census block; this required data processing prior to the analyses, which is described next:

2.1.1 Gastrointestinal Disease Incidence (GDI) records

The outcome variable of interest was incidence rate of gastrointestinal disease (GDI). Incidence rate is defined as new cases per unit of time divided by the population at risk. We requested annual reports of gastrointestinal disease cases recorded from each of the 279 public clinics across MC, for the years 2007 to 2014, from the Secretary of Health (Secretaría de Salud) to be used as a proxy for incidence in the population. There are also many private hospitals and clinics in the city; we did not have access to data from these sources. The data are categorized by the Secretariat of Health according to the diagnosed cause of disease: “parasitic” (Group 1), “bacterial” (Group 2), and “not defined” (Group 3) (Table S1).

2.1.2 Flooding records

We obtained records of flood events for the period 2007 to 2014, reported from citizens and corroborated by SACMEX, the city water authority. This dataset includes the specific neighborhood where the events occurred and the year.

2.1.3 Residential services and socioeconomic variables

We compiled data on the number of homes per census block with access to piped water, access to the sewage system, construction materials, and toilet facilities, as well as the level of education of the inhabitants and the percentage of people under 18 from the Population Censuses conducted in 2005 and 2010 (INEGI, 2011). We complemented this information with the purchasing power index, as a proxy for income (CONEVAL, 2013). Finally, we obtained data on the location of informal food stands and the location of sewers in the city using data from Mexico City Statistical Department.

2.1.4 Gastrointestinal Disease Incidence (GDI)

Patients with symptoms of gastrointestinal infection can access health centers and hospitals with different levels of specialization, starting from local health clinics (first level), up to specialized hospitals (third level) (PAHO, 2002). Thus, to avoid double counting specific cases, we only considered records from first level health centers, which are the clinics that only admit patients from neighborhoods that are geographically close to the clinic. To represent the gastrointestinal disease data at the scale of the census block, we distributed the cases recorded in each clinic among all the census blocks that fell within the area of influence of the clinic, assuming that the proportion of cases recorded in the clinic were proportional to the population of the census block. For this, we used the information about the proportion of the population in each census block associated with the type of health providers (INEGI, 2011). Finally, to calculate gastrointestinal disease incidence, we divided the number of cases by the total population of the census block. This procedure was conducted in ArcGIS, using a spatial proximity method by the spatial join function (ESRI, 2011).

2.1.5 Socio-economic and hydrological data

The records of flood events were provided at the neighborhood level; neighborhoods typically encompass all or part of several census blocks. To convert the data to census block unit, we assumed that occurrence of flood events associated with each census block was proportional to the area of neighborhood intersecting with each census block, such that a census block that was only partially within a neighborhood would only be assigned partial flooding events. For instance, two census blocks of the same size within a neighborhood that was flooded 10 times in a year would thus be considered to have experienced 5 events, each. To find the area proportion for each neighborhood in a census block, we constructed a joint shape-file layer, combining census block and neighborhood (polygons) within the software ARC-GIS (ESRI, 2011).

Finally, the socio-economic variables that reported number of people per census block were normalized by the population size to use proportion instead of absolute values.

2.2 Data Analyses

Since geo-referenced data face a type of effect known as spatial dependence (Anselin, 1988, 2001), conventional statistical tools are not appropriate to analyze the spatial association of GDI risk with socio-hydrological vulnerability at the scale of the city. The concept of spatial dependence or autocorrelation comes from the likelihood that geographic units that are closer in space have a tendency to influence each other and consequently have similar attributes. In other words, the magnitude of the variable of interest may be determined by the values of the same or others variables in locations nearby. Therefore, a proper analysis of spatial data that takes into account these spatial dependencies is needed for this case.

We rely on the use of a statistical modeling approach that considers spatial dependency when defining univariate and multivariate correlations and regression analysis (Anselin, 1988). We first conducted univariate statistics to evaluate the spatial dependency of the GDI records, as well as the socio-hydrological factors. We then conducted bivariate statistical tests to evaluate the global dependency of GDI to each factor. We then evaluated the local influence of each factor on the GDI in each census block using spatial correlation methods. Finally, we conducted regression analyses to formally evaluate the relative contribution of each factor to explain GDI, including the spatial association of the data. All the analyses were conducted over the accumulated incidence recorded for two different periods of time 2007-09 and 2010-14.

2.2.1 Spatial autocorrelation of GDI

To evaluate the spatial dependency of GDI and the socio-hydrological risk factors independently, a univariate Moran Index (Moran´s I) was calculated (Anselin, 1988, 1996). The Moran's I consists of identifying the global degree of spatial association of data. In other words, how much the magnitude of an indicator in one location is influenced by the magnitude of the indicator in the area close to it (Anselin, 1996, 2001). To calculate the Moran's I, a contiguity weight matrix needs to be specified (appendix A). This matrix was also used for the subsequent statistical analyses.

2.2.2 Global spatial association of GDI and socio-hydrological risk factors

In order to determine the existence of a spatial relationship between GDI and the analyzed factors, we conducted a bivariate statistical test using the bivariate Moran's Index (Wartenberg, 1985). This index captures the influence of an independent variable (e.g. flooding events) on the magnitude of a dependent variable (i.e. GDI) including not just the effect of the dependent variable in the focal point, but also including the effect of the neighbors, which were defined by the contiguity weight matrix. This indicator reports in a single value the global spatial association between GDI and each socio-hydrological factor.

2.2.3 Local Indicators of Spatial Association (LISA) of GDI and socio-hydrological risk factors

The next step was to identify areas of the city with different levels of disease burden (GDI) that responded similarly to a particular socio-hydrological factor. This was done by calculating the Local Indicators of Spatial Association (LISA) (Anselin, 1995). These indicators were calculated to evaluate the spatial dependency between the level of GDI in each census block, and the magnitude of the socio-hydrological risk in the nearest neighborhoods defined, again, by the contiguity matrix. These indicators signify localize differences in the association between the dependent (GDI) and independent variables (analyzed factors). Depending on the sign of the indicator (positive or negative), these local associations can express positive-positive associations, positive-negative, negative-positive or negative-negative. We reported in the main text only the type of association that implies an enhancement in GDI associated to an increase in the analyzed factor (i.e. positive-positive response). All other associations are reported in the supplementary material (figures S1(a), S1(b), S2(a) and S2(b)).

2.2.4 Spatial Regression Model

To evaluate the relative contribution of each socio-hydrological risk factor in explaining the variance of GDI in space, we constructed a multivariate regression model that considered altogether the socio-hydrological factors (i.e. socio-economic, demographic and hydrological) in a single equation. We first defined a model that includes all the variables that were previously analyzed independently. We then conducted an Akaike information criteria test in order to obtain a model that considers only the variables that explain most of the variance in the data, without over-fitting. To evaluate if a spatial parameter had to be included in the final model, a Moran´s test was conducted on the residuals of the regression. If this test reported a significative value it means that the regression residuals are spatially auto-correlated, thus a Lagrange Multiplier test was applied to determine the type of spatial dependency to include in the model, which could be as a residual parameter or as a spatial lag parameter (Anselin, 2001, 2003; Elhorst, 2014) (appendix A).

3 RESULTS

3.1 Spatial distribution of GDI

The spatial distribution of GDI for each group of pathogens over the entire city reveals that the incidence records are concentrated in the center of the city, in the lowlands, and at the periphery of the city, in the highlands (figure 1). This pattern is consistent for the different categories of pathogens (parasitic (Group 1), bacterial (Group 2), and non-well defined categories (Group 3)) and for the two analyzed periods (2007-2009 and 2010-2014). We also observed an overall reduction in cases from period 1 to period 2 (table S1). Infections caused by protozoans (Group 1) are more frequently observed in the center of the city. Nonetheless, a significant number of cases is also observed at the periphery (figure 1b). Incidence of bacteria related infections (Group 2) are less commonly observed, and mostly concentrated at the center of the city in the lowest part of the watershed.

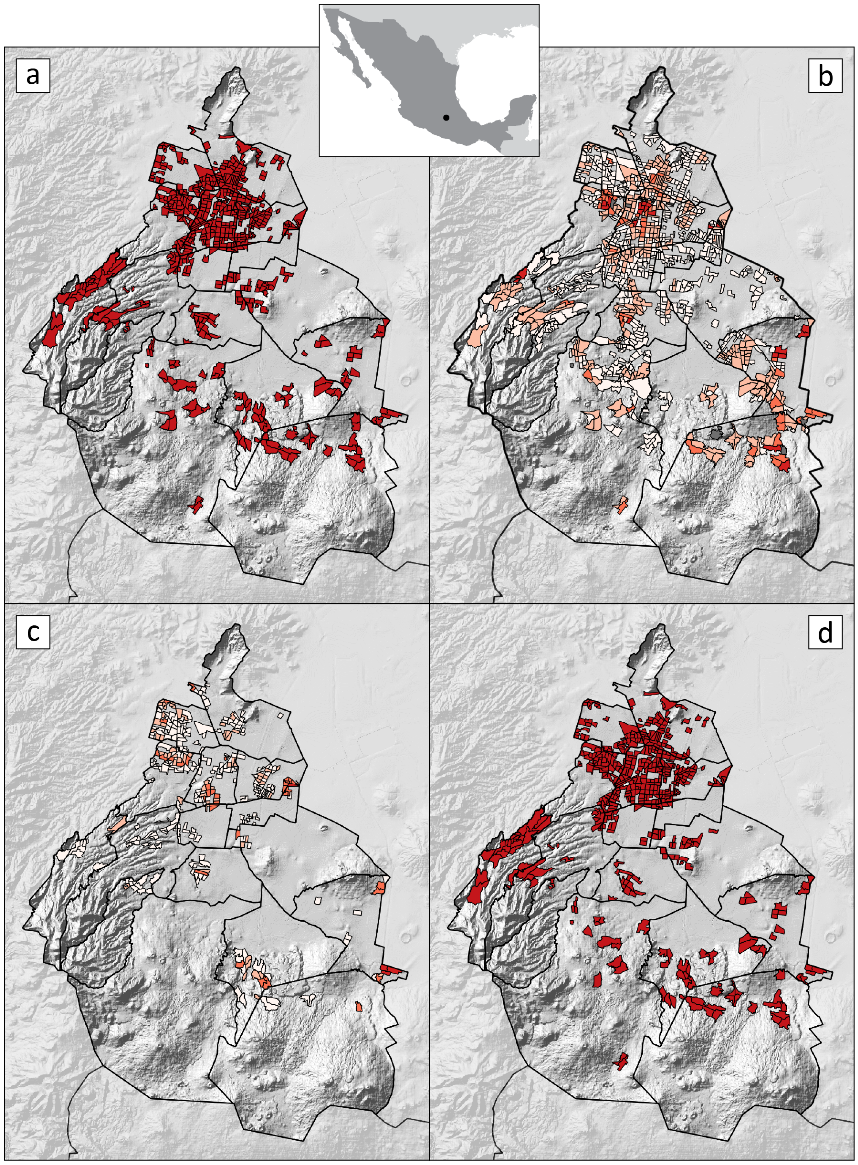

Figure 1. Distribution of Gastrointestinal Disease Incidences (GDI) at census block scale in Mexico City for the period 2010-14. (a) Total GDI; (b) GDI attributable to Group 1, parasitic (protozoans) microorganisms; (c) GDI attributable to Group 2, bacterial microorganisms; (d) GDI attributable to Group 3, not well-defined causal agent. For (a) and (d), red census blocks correspond to those with incidence rate > 1 [case/1,000 inhabitants per year]. For (b) and (c), tones of reds indicate low (light) or high incidence rates (dark). Municipality's limits are represented by solid black lines. Grey background depicts the elevation differences.

3.2 *Spatial autocorrelation of variables and GDI*

For all the groups of pathogens there is a strong spatial autocorrelation for the entire city (table 1). The concentration of incidences reported at the center of the city as well as in the periphery contributes to this high autocorrelation. Similarly, socio-hydrological risk factors are also significantly associated with themselves. Especially, the spatial autocorrelation of the flood events is high and significant in the Moran's I. This is due to the fact that most of the flooding events are concentrated at the center of the city (table S2).

Results from the bivariate analyses indicate a spatial association of most of the socio-hydrological factors analyzed with GDI. Of the eleven factors analyzed using the bivariate Moran Index, eight presented a significant spatial autocorrelation with GDI (table 1). These include three factors positively correlated with hydrological risk: number of flooding events, number of sewer drains, and homes without toilet facilities. The other five variables correspond to socio-economic factors that are not directly associated with hydrological risk. Two of these exhibited a positive correlation with GDI: number of street food stands and level of education. The rest of the variables that were negatively correlated with hydrological risk are: % of people <18 years old, income index, and homes with dirt floor.

Table 1. Bivariate spatial autocorrelation index between GDI and analyzed factors. The table shows the autocorrelation coefficient obtained from the Bivariate Moran´s Index for each period analyzed (2007-09; 2010-14). The asterisks show the level of significance: ***, 99%; **, 95%; *, 90%.

	2007-09				2010-14			
Factor	**Group of pathogen**				**Group of pathogen**			
	All	**Group 1**	**Group 2**	**Group 3**	**All**	**Group 1**	**Group 2**	**Group 3**
Flooding	0.128***	0.033**	0.076***	0.144***	0.211**	0.103**	0.081**	0.241**
Homes with dirt floor	-0.031***	-0.009	-0.052***	-0.035***	-0.048***	-0.013*	-0.020***	-0.057**
Homes without water supply	-0.007	0.008	-0.026***	-0.01	0.000	0.010	-0.005	-0.004
Homes without toilet facilities	0.023*	0.023*	0.017*	0.022*	0.058**	0.025*	0.009	0.067**
Homes without sewage	0.007	0.012	-0.007	0.006	0.001	0.008	-0.001	-0.001
% people < 18 years old	-0.041***	-0.011	-0.055***	-0.046***	-0.013*	0.008	-0.019	-0.019**
Level of education	0.055***	0.007	0.087***	0.063***	0.029**	0.003	0.012	0.036**
Income Index	-0.006	-0.019*	0.039***	-0.003	-0.011	-0.017*	0.001	-0.010
Sewer drains	0.115***	0.029**	0.090***	0.123***	0.107**	0.038**	0.237*	0.133**
Street food stands	0.11***	0.052**	0.028**	0.125***	0.109***	0.278**	0.034*	0.128**

3.3 At the local scale, the city differs in the factors associated with GDI

Results from the local indicators of spatial association test (LISA) show that the factors spatially associated with high GDI differ significantly between the center (lowlands) of the city and periphery (highlands) (figure 2). Specifically, at the center, incidences have a positive correlation (positive/positive) with flooding reports, the % of homes without toilet facilities and food stands. At the periphery, a negative association exists between GDI and income in the closest neighborhoods. Homes without toilet facilities also present a significant positive association with GDI in the periphery of the city. A full display of the analysis, including all factors and correlations is presented in the supporting material (figures S1(a), S1(b), S2(a) and S2(b)).

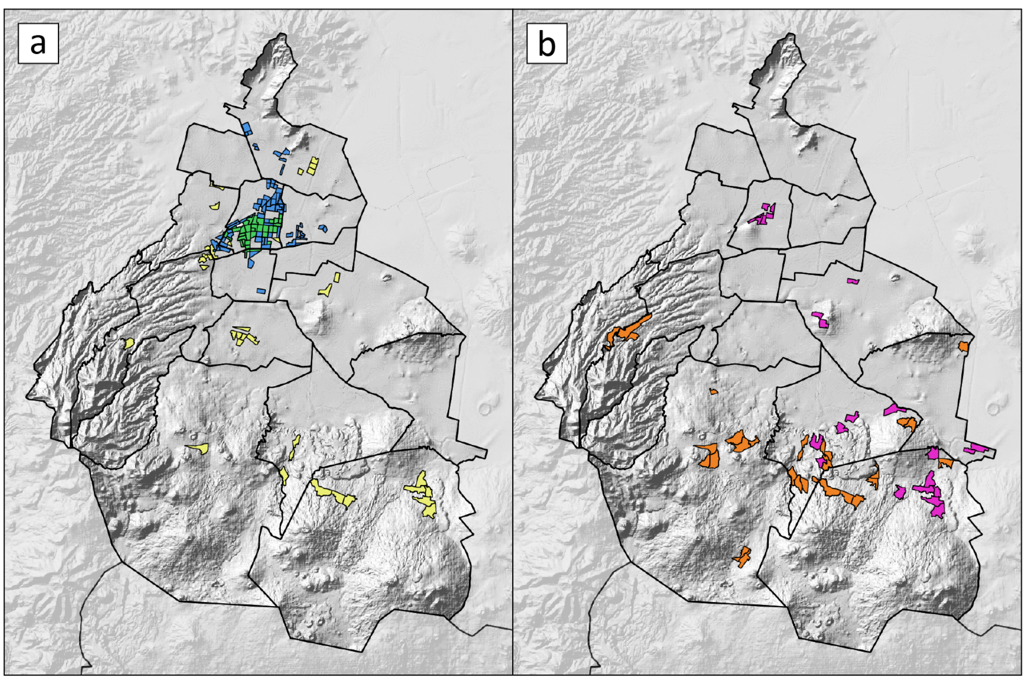

Figure 2. Clusters of bivariate local indicators of spatial associations (Bivariate LISA) between GDI and the analyzed factors for the period 2010-14. (a) census blocks with significant positive/positive association of GDI and flooding (navy blue) and homes without toilet facilities (yellow); the association overlap of both variables is represented in green**;** (b) census blocks with significant associations of GDI and infrastructure, particularly with homes without water supply (orange) and homes without sewage (purple).

3.4 The influence of the spatial component in the center of the city explains a 50% increase in the risk of GDI

The spatial pattern observed by mapping the distribution of cases in the city, and the differences reported in the LISA test, are suggestive of strong differences between the center and periphery in terms of both the manifestation of GDI and factors that contribute to this risk. Thus, we decided to conduct a separate regression analysis for the center and for the periphery. The criteria to define these two regions were based on the hydrological and topographic characteristics of MC drainage system: the area defined as highlands corresponds to the region where water from rainfall drains naturally on the surface of the watershed toward the center. The area defined as lowlands represents the region in which the water drains through man-made infrastructure, such as pipes and canalized rivers.

Results from the regression model conducted for lowlands show that flooding reports is the only factor significantly associated with GDI (table 2). The model also evidences the presence of spatial dependence in the GDI, this spatial dependence is captured in the regression as the spatial lag parameter (rho) for the dependent variable (GDI), which is a weighted average of GDI at neighboring locations. Similar results were obtained when the regression was applied to each type of pathogen independently.

For the regression model conducted for highlands, different factors were significantly associated with GDI, and these factors changed depending of the pathogen group analyzed. Nonetheless, two factors were significant in all regression cases: % of homes with dirt floor and % of homes without toilet facilities. For highlands, the regression did not include a residual or lag spatial parameter indicating that the spatial association is not as strong in the highlands.

Table 2. Regression model. Table of the contribution of each statistically significant factor in explaining the variance of GDI per analyzed period (2007-09; 2010-14) and city region (center and periphery). The asterisks show the level of significance: ***, 99%; **, 95%; *, 90%.

Factor	Period 2007-09							Period 2010-14					
	Center-Lowlands				Periphery-Highlands			Center-Lowlands			Periphery-Highlands		
	Group of pathogen				Group of pathogen			Group of pathogen			Group of pathogen		
	All	Group 1	Group 2	Group 3	All	Group 1	Group 3	All	Group 2	Group 3	All	Group 1	Group 3
Intercept	0.374***	0.0473***	0.002***	0.439	-0.280**	0.052***	-0.595**	0.373***	0.002***	0.522	0.827**	0.016***	0.741***
Flooding	0.014***	0.002**	0.000**	0.010***				0.020***	0.000**	0.001**			
Homes with dirt floor					1.294**	-0.050**	0.020**				1.302	-0.048	1.158
Homes without water supply							-0.138						
Homes without toilet facilities					-0.100**		0.447***				0.635***	0.041***	0.447***
Homes without sewage												0.045***	
% people < 18 years old													-2.573***
Level of education						-0.005***	0.072				-2.764***		
Income Index					1.262***	0.033***	0.802***				0.411**		0.425***
rho	0.451***	0.206***	0.173***	0.550***				0.588***	0.534***	0.541***			

4 DISCUSSION

The spatial statistical analyses conducted on GDI records in the highly populated and flood-prone area of MC revealed a significant spatial association between incidence and the hydrological risk to flooding and scarcity. While an overall reduction in cases was observed during the investigated periods, the positive association between gastrointestinal disease incidences and socio-hydrological risk remains strong through time.

The set of univariate and bivariate global spatial autocorrelation analyses conducted in this work indicates a high diversity in the factors that help explain the heterogeneous spatial distribution of risk between the center of the city and its periphery. At the center of the city, in the lowlands, incidences correlate strongly with areas highly vulnerable to flooding, while at the periphery, more incidences were associated with lack of sanitation, infrastructure deficiencies and low income. While the observed relationship between GDI and flooding in the center of Mexico City is strong and lasting, the mechanisms behind this pattern are not well understood. The area of high positive association has been historically the most prone to flooding. Yet, it has been historically recognized as one of the cores of the city´s wealth (Romero Lankao, 2010a), with only few areas with poor residents. Moreover, since pre-Columbian times, elaborate infrastructure and engineering investments have been made in the city center as the principal means of reducing the risk of large scale floods (Ezcurra *et al.*, 1999). However, aging of infrastructure together with the growth of the city and its population over the last 50 years have increased the demand on the system to not only eliminate storm water but also cope with increased waste water discharge. The problems of capacity continue to plague its effectiveness, and the city is constantly struggling to ensure discard of all its water (Romero Lankao, 2010a).

The peripheral areas correspond to neighborhoods with high levels of economic informality (Romero Lankao, 2010b; Gilbert and De Jong, 2015), underscoring the linkages between fast, irregular and illegal urban settlements and health (Khan, 2012). In these areas with fast growth and high informality, the lack of services incentivize people to find strategies to locally manage the distribution, storage and use of water (Eakin *et al.*, 2016). Household management of water introduces many opportunities for contamination, taxing households not only with the costs of water storage and delivery, but also the burden of disease (Cohen *et al.*, 2008).

The results of the LISA test and the regression analyses indicate that, in the center of MC, the risk of being exposed to waterborne pathogens in one location increases significantly when the location is surrounded by high incidence neighborhoods. This is known as the "neighborhood effect"(Witten, Witten and Karen, 2017). According to Ellen et al., (Ellen and Turner, 1997; Ellen, Mijanovich and Dillman, 2001) neighborhood conditions can influence health outcomes in the short term by influencing the behavior and attitude of the residents in the surrounding areas and, in the long term, by the reinforced and accumulated effect of the external conditions that shape neighborhood heterogeneity in terms of poverty, education and other adverse environmental conditions, such as infrastructure. Our results are suggesting that it is the latter factors that are influencing higher incidence in the center. Two pieces of evidence support this conclusion: first, the strong clustering of flood risk as an external forcing, and second, the final regression model did include other socio-economic factors such as income and education in the urban periphery where scarcity is the main factor. However, the effect of behavior cannot be underestimated, as we found also a strong correlation between GDI and the number of street food stands, which can be considered an indicator of a common daily consumption behavior of MC residents. The movement of people to areas of the city with flood risk coupled with the practice of eating in street venues increases risk and chance of exposure.

The spatial heterogeneity in GDI in MC is similar to patterns observed in other rapidly growing urban areas of the developing world. For instance, Reiner and collaborators found that average incidences of cholera at the center (core) of the city of Dhaka were two times higher than at the periphery (Reiner *et al.*, 2012). Similar patterns were observed for the same city but for rotavirus infections (Martinez *et al.*, 2016). In Ahmedabad, India, Santos and collaborators found strong heterogeneity in malaria incidence related to socio-economic differences, with higher risk levels at the center of the city compared to the borrows far from the historic center (Santos-Vega *et al.*, 2016). While the mechanisms of transmission of these diseases differ, all the studies point to the management of socio-hydrological risk, and/or differences in socio-economic conditions and densification patterns of core urban areas as explanatory factors.

The significance of flooding to GDI in the city center and lack of water infrastructure to GDI in the urban fringe underscores the need to consider the dynamic interaction between water management at the regional scale and the control of waterborne diseases in public health planning. Cities are likely to be more vulnerable to hydrological hazards under climate change as they simultaneously grow in area and population (Cutter *et al.*, 2012). Population growth at the periphery is increasing the demand for drinking water, and in turn, is diminishing the capacity of the sewage system. Moreover, an increase in extreme rainfall over time has been associated with the process of urbanization and the drainage of MC's lakes (Benson-Lira *et al.*, 2016), and as these processes continue in the city and surrounding regions, ~~–~~it is expected the quantity of runoff into the sewage system will increase, overwhelming a system already operating over its capacity, and hampered by age and subsidence (Romero Lankao, 2010b). If policy and health programs are to be improved under such scenarios, there will be an increasing need for dynamic disease transmission models that couple the social factors and ecological determinants of disease along with traditional epidemiological models.

Our results underscore the need for a more integrated governance of water resource infrastructure and public health. Today little anticipatory dialogue exists between sector agencies in relation to proactive risk management, yet they operate in similar spatial domains and consider similar socio-economic factors to assess risk and vulnerability. The agencies that govern and operate water resources and infrastructure rarely consider health factors when defining actions and plans. Public health institutions, on the other side, often overlook the significance of biophysical and infrastructural conditions that influence the disease burden in urban areas (Parkes and Horwitz, 2009). While it is unrealistic to expect any sector agency to expand its mandate to address the issues and demands of other agencies, coordinated action could produce synergistic outcomes of benefit to multiple sector objectives. Our research illustrates the close relationship between the built environment, exposure to hazards and the burden of disease. While these relationships are often subject to speculation, our analysis of available data demonstrates that the spatial correlations exist, and that the factors contributing to disease vary with both exposure to different forms of hydrological stress and the varied infrastructural and socioeconomic conditions of the city. As we move to confront the challenges of the Anthropocene, improved governance will be needed to address the complex interactions between sector dynamics and hazards in the megacity.

ACKNOWLEDGEMENTS

We thank The National Institute of Statistics and Geography (INEGI), The Secretary of Health of Mexico City and The Mexico City water Operator (SACMEX) for data. Fidel Serrano Candela, Bertha Hernanadez, and Sergio Bourguet for technical assistance. This research was funded by the National Science Foundation [Grant No. 1414052] and the Inter-American Institute for Global Change Research [Project number: CRN3108].

APPENDIX A

A.1 Spatial dependence

The concept of spatial dependence, or spatial autocorrelation, has its roots on the Tobler's First Law of Geography, which states that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1979). In this sense, geographic units that are closer in space must have the propensity to influence each other, and consequently have similar attributes. In other words, spatial dependence implies that the values of a certain variable is determined by the values of the same variable in other locations nearby. This relationship can be formally expressed as (Anselin, 1988):

$$cov(x_i, x_j) = E(x_i, x_j) - E(x_i) \cdot E(x_j) \neq 0, \text{ where } i \neq j \qquad \text{(A-1)}$$

Where $i = 1, \ldots, N$ and $j = 1, \ldots, N$, represent the two locations, while x_i and x_j are the values of the variable of interest location i and j respectively. $cov(x_i, x_j)$ can take positive and negative values. Positive values indicate that high values of a variable in i are associated with high values in j. Finally, when the values of a particular variable in two locations are distributed randomly it is consider a non-dependency.

A.2 Spatial autocorrelation index

To calculate the global spatial dependency we use the Moran's Index (*I*), which is a non-parametric index of the spatial dependency across the entire landscape using local the covariance and weighed by the elements of a contingency matrix, w_{ij}. The mathematical definition of Moran's I is found in the following equation (Anselin, 1995, 1996):

$$I = \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum(x_i - \bar{x})^2} \tag{A-2}$$

Where n is the number of spatial locations, w_{ij} are the elements of the contiguity matrix (see *A.2.1*), and $(x_i - \bar{x})$ and $(x_j - \bar{x})$ are the normalized values of variable x in location *i* and *j*.

The index is calculated under the null hypothesis of no spatial dependency, and assuming normal distribution, but it is important to remark that this is just an approximation, since the empirical distribution of this statistic is unknown.

A.2.1 Contiguity Matrix

The contiguity matrix, W, also known as the spatial weight index, is a binary matrix that identifies the area of influence of each spatial unit. Thus, when two spatial units, i and j, are considered neighbors, the value of cell $w_{ij} = 1$, and zero in other case (Anselin, 1996, 2001). The way in which the spatial weights are assigned depends on the type of contiguity that is used. For this work we use a Queen Contiguity, in which its units are considered neighbors if they share boundaries and corners.

A.3 Bivariate Moran spatial autocorrelation statistic

Bivariate Moran quantifies the global spatial dependency between two variables, and it is calculated based on the multivariable spatial correlation (Wartenberg, 1985):

$$I_{X,Y} = \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \tag{A-3}$$

Where $(y_i - y)$ are the normalized values of variable y in location *i.*

The values of $I_{X,Y}$ must be interpreted similarly to those obtained by the univariate Moran Index. That is, positive values reflect a positive relationship between variables *x* and *y*.

A.4 Spatial autocorrelation indicators

The local indicator for spatial autocorrelation (LISA) is a measure of spatial autocorrelation in a single location. Their mathematical representation is shown below(Anselin, 1995):

$$I_i = (x_i - \bar{x}) \sum_j w_{ij}\left(x_j - \bar{x}\right) \tag{A-4}$$

A.5 Spatial Regression Analysis

A regression model of the form

$$Y_i = \rho W Y_i + \boldsymbol{\beta} X + \varepsilon_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(A-5)}$$

was used to incorporate the full set of predictors and the spatial dependency observed in the data. $\boldsymbol{Y}_i$ is an $N \times 1$ vector of observations of the dependent variable, with one observation for every unit in the sample, X is a $N \times K$ matrix of exogenous explanatory variables , $\boldsymbol{\beta}$ is a $K \times 1$ vector with unknown parameters to be estimated, ε_i is a $N \times 1$ vector of disturbance terms, where ε_i is assume to be independently and identically distributed for all i, with zero mean and variance σ^2. In order to capture the spatial dependency observed in the incidence data, the model incorporates an additional regressor in the form of a spatially lagged variable, WY_i (Anselin, 2001). This variable captures cross-section dependencies, in which exist a covariance structure in different locations derives from the geographic space (Anselin 1998, Anselin, 2001). The term ρ is the unknown spatial lag coefficient, and W is the $N \times N$ contiguity matrix.

REFERENCES

Ahern, J. (2013) 'Urban landscape sustainability and resilience: the promise and challenges of integrating ecology with urban planning and design', *Landscape Ecology*. Springer Netherlands, 28(6), pp. 1203–1212. doi: 10.1007/s10980-012-9799-z.

Ahern, M., Kovats, R. S., Wilkinson, P., Few, R. and Matthies, F. (2005) 'Global health impacts of floods: epidemiologic evidence.', *Epidemiologic reviews*. Oxford University Press, 27(1), pp. 36–46. Available at: http://epirev.oxfordjournals.org/content/27/1/36.full (Accessed: 12 April 2016).

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, *Operational Regional Science Series*. Springer Science+Business Media Dordrecht. doi: 10.1007/978-94-015-7799-1.

Anselin, L. (1995) 'Local indicators of spatial association — LISA.', *Geographical Analysis*, 27(2), pp. 93–115. doi: 10.1111/j.1538-4632.1995.tb00338.x.

Anselin, L. (1996) 'The Moran Scatterplot as an ESDA tool to assess local instability in spatial association', *Spatial analytical perspectives on GIS*, pp. 111–125.

Anselin, L. (2001) 'Spatial econometrics', *A companion to theoretical econometrics*, pp. 310–330. doi: 10.1002/9780470996249.

Anselin, L. (2003) 'An Introduction to Spatial Regression Analysis in R', *University of Illinois, Urbana-Champaign*, p. 25. Available at:

http://labs.bio.unc.edu/Buckley/documents/AnselinIntroSpatRegres.pdf.

Benson-Lira, V., Georgescu, M., Kaplan, S. and Vivoni, E. R. (2016) ‘Loss of a lake system in a megacity: The impact of urban expansion on seasonal meteorology in Mexico City’, *Journal of Geophysical Research: Atmospheres*, 121(7), pp. 3079–3099. doi: 10.1002/2015JD024102.

Cohen, S. A., Egorov, A. I., Jagai, J. S., Matyas, B. T., DeMaria, A., Chui, K. K. H., Griffiths, J. K. and Naumova, E. N. (2008) ‘The SEEDs of two gastrointestinal diseases: Socioeconomic, environmental, and demographic factors related to cryptosporidiosis and giardiasis in Massachusetts’, *Environmental Research*, 108(2), pp. 185–191. doi: 10.1016/j.envres.2008.06.009.

CONEVAL, (Consejo Nacional de Evaluación de la Política de Desarrollo Social) (2013) ‘Nota técnica del Rezago Social en las AGEB urbanas en México’, pp. 1–7. Available at: http://www.coneval.gob.mx/Medicion/IRS/Paginas/Rezago_social_AGEB_2010.aspx.

Cutter, S., Osman-Elasha, B., Campbell, J., Cheong, S.-M., McCormick, S., Pulwarty, R., Supratid, S., Ziervogel, G., Calvo, E., Mutabazi, K. D., Arnall, A., Arnold, M., Bayer, J. L., Bohle, H.-G., Emrich, C., Hallegatte, S., Koelle, B., Oettle, N., Polack, E., Ranger, N., Rist, S., Suarez, P. and Wilches-Chaux, G. (2012) ‘Managing the Risks from Climate Extremes at the Local Level’, in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. Cambridge: Cambridge University Press, pp. 291–338. doi: 10.1017/CBO9781139177245.008.

Dye, C. (2008) ‘Health and Urban Living’, *Science*, 319(5864).

Eakin, H., Bojórquez-Tapia, L. A., Janssen, M. A., Georgescu, M., Manuel-Navarrete, D., Vivoni, E. R., Escalante, A. E., Baeza-Castro, A., Mazari-Hiriart, M. and Lerner, A. M. (2017) ‘Opinion: Urban resilience efforts must consider social and political forces.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 114(2), pp. 186–189. doi: 10.1073/pnas.1620081114.

Eakin, H., Lerner, A. M., Manuel-Navarrete, D., Hernández Aguilar, B., Martínez-Canedo, A., Tellman, B., Charli-Joseph, L., Fernández Álvarez, R. and Bojórquez-Tapia, L. (2016) ‘Adapting to risk and perpetuating poverty: Household’s strategies for managing flood risk and water scarcity in Mexico City’, *Environmental Science & Policy*, 66, pp. 324–333. doi: 10.1016/j.envsci.2016.06.006.

Elhorst, J. P. (2014) *Spatial Econometrics From Cross-Sectional Data to Spatial Panels*. doi: 10.1007/978-3-642-40340-8.

Ellen, I. G., Mijanovich, T. and Dillman, K.-N. (2001) ‘Neighborhood Effects on Health: Exploring the Links and Assessing the Evidence’, *Journal of Urban Affairs*, 23(3&4), pp. 391–408. doi: 10.1111/0735-2166.00096.

Ellen, I. G. and Turner, M. A. (1997) ‘Does neighborhood matter? Assessing recent evidence’, *Housing Policy Debate*, 8(4), pp. 833–866. doi: 10.1080/10511482.1997.9521280.

ESRI (2011) ‘ESRI ArcGIS Desktop: Release 10. Redlands, CA Institute, Environmental Systems Research’.

Ezcurra, E., Mazari-Hiriart, M., Pisanty, I. and Aguilar, A. G. (1999) ‘The basin of Mexico : critical environmental issues and sustainability’. UNU Press.

Gilbert, L. and De Jong, F. (2015) ‘Entanglements of Periphery and Informality in Mexico City’, *International Journal of Urban and Regional Research*, 39(3), pp. 518–532.

doi: 10.1111/1468-2427.12249.

Hunter, P. R. (1997) *Waterborne disease: epidemiology and ecology.*, *Waterborne disease: epidemiology and ecology.* John Wiley & Sons Ltd.

INEGI (2011) *Censo General de Poblacion y Vivienda 2010*. Mexico City.

IPCC (2012) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*. Edited by C. B. Field, V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, S. K. Allen, M. Tignor, and P. M. Midgley. Cambridge, UK: Cambridge University Press.

Jacobi, P., Kjellen, M., McGranahan, G., Songsore, J. and Surjadi, C. (2010) *The citizens at risk: from urban sanitation to sustainable cities*. Routledge.

Khan, M. H. (2012) ‘Urban health in megacities of developing countries’, *Public Health Forum*, 20(2), p. 29.e1-29.e3. Available at: http://www.sciencedirect.com/science/article/pii/S0944558712000157 (Accessed: 12 April 2016).

Martinez, P. P., King, A. A., Yunus, M., Faruque, A. S. G. and Pascual, M. (2016) ‘Differential and enhanced response to climate forcing in diarrheal disease due to rotavirus across a megacity of the developing world.’, *Proceedings of the National Academy of Sciences of the United States of America*, 113(15), p. 1518977113-. doi: 10.1073/pnas.1518977113.

PAHO (2002) ‘Perfil del sistema de servicios de salud méxico’, *Pan American Health Organization*.

Parkes, M. W. and Horwitz, P. (2009) ‘Water, ecology and health: ecosystems as settings for promoting health and sustainability.', *Health promotion international*, 24(1), pp. 94–102. doi: 10.1093/heapro/dan044.

Pick, J. and Butler, E. (1997) *Mexico megacity*. Boulder Colorado: Westview Press.

Pickett, S. T. A., Cadenasso, M. L., Rosi-Marshall, E. J., Belt, K. T., Groffman, P. M., Grove, J. M., Irwin, E. G., Kaushal, S. S., LaDeau, S. L., Nilon, C. H., Swan, C. M. and Warren, P. S. (2017) 'Dynamic heterogeneity: a framework to promote ecological integration and hypothesis generation in urban systems', *Urban Ecosystems*. Springer US, 20(1), pp. 1–14. doi: 10.1007/s11252-016-0574-9.

Reiner, R. C., King, A. A., Emch, M., Yunus, M., Faruque, A. S. G. and Pascual, M. (2012) 'Highly localized sensitivity to climate forcing drives endemic cholera in a megacity.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 109(6), pp. 2033–6. doi: 10.1073/pnas.1108438109.

Romero Lankao, P. (2010a) 'Water in Mexico City: what will climate change bring to its history of water-related hazards and vulnerabilities?', *Environment and Urbanization*. SAGE PublicationsSage UK: London, England, 22(1), pp. 157–178. doi: 10.1177/0956247809362636.

Romero Lankao, P. (2010b) 'Water in Mexico City: what will climate change bring to its history of water-related hazards and vulnerabilities?', *Environment and Urbanization*, 22(1), pp. 157–178. doi: 10.1177/0956247809362636.

Santos-Vega, M., Bouma, M. J., Kohli, V. and Pascual, M. (2016) 'Population Density, Climate Variables and Poverty Synergistically Structure Spatial Risk in Urban Malaria in India', *PLOS Neglected Tropical Diseases*. Edited by M. Maheu-Giroux. Public Library of Science, 10(12), p. e0005155. doi: 10.1371/journal.pntd.0005155.

Santos-Vega, M., Martinez, P. P. and Pascual, M. (2016) ‘Climate forcing and infectious disease transmission in urban landscapes: integrating demographic and socioeconomic heterogeneity’, *Annals of the New York Academy of Sciences*, 1382(1), pp. 44–55. doi: 10.1111/nyas.13229.

Sivapalan, M., Savenije, H. H. G. and Blöschl, G. (2012) ‘Socio-hydrology: A new science of people and water’, *Hydrological Processes*. John Wiley & Sons, Ltd, 26(8), pp. 1270–1276. doi: 10.1002/hyp.8426.

Tobler, W. R. (1979) ‘Cellular Geography’, *Philosophy in geography*, 20, pp. 379–386. doi: 10.1007/978-94-009-9394-5.

Turner, B. L., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., Eckley, N., Hovelsrud-Broda, G. K., Kasperson, J. X., Kasperson, R. E., Luers, A., Martello, M. L., Mathiesen, S., Naylor, R., Polsky, C., Pulsipher, A., Schiller, A., Selin, H. and Tyler, N. (2003) ‘Illustrating the coupled human-environment system for vulnerability analysis: three case studies.’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 100(14), pp. 8080–5. doi: 10.1073/pnas.1231334100.

UN-HABITAT (2017) *New urban agenda*.

United Nations (2016) ‘The World’s Cities in 2016: Data Booklet.’, *Economic and social affair*, p. 29. Available at: http://www.un.org/en/development/desa/population/publications/pdf/urbanization/the_worlds_cities_in_2016_data_booklet.pdf.

Vlahov, D., Freudenberg, N., Proietti, F., Ompad, D., Quinn, A., Nandi, V. and Galea, S. (2007) ‘Urban as a determinant of health.’, *Journal of urban health : bulletin of the New York Academy of Medicine*, 84(3 Suppl), pp. i16-26. doi: 10.1007/s11524-007-

9169-3.

Vlahov, D., Galea, S., Gibble, E. and Freudenberg, N. (2005) ‘Perspectives on urban conditions and population health’, *Cadernos de Saúde Pública*. Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz, 21(3), pp. 949–957. doi: 10.1590/S0102-311X2005000300031.

Wartenberg, D. (1985) ‘Multivariate Spatial Correlation: A Method for Exploratory Geographical Analysis’, 17(4).

Witten, K., Witten and Karen (2017) ‘Neighborhoods and Health’, in *International Encyclopedia of Geography: People, the Earth, Environment and Technology*. Oxford, UK: John Wiley & Sons, Ltd, pp. 1–4. doi: 10.1002/9781118786352.wbieg0151.

Zhou, W., Pickett, S. T. A. and Cadenasso, M. L. (2017) ‘Shifting concepts of urban spatial heterogeneity and their implications for sustainability’, *Landscape Ecology*. Springer Netherlands, 32(1), pp. 15–30. doi: 10.1007/s10980-016-0432-4.

SUPPLEMENTARY MATERIAL

Table S1. Average number of gastrointestinal disease cases (incidences) in Mexico City per year for the period 2007-09 and 2010-14. Cases are divided into three categories or groups based on the associated causal pathogen, as reported by the Health Secretariat.

Group	Gastrointestinal Diseases (GD)	Pathogen	Average number of cases per year (2007 - 09)	Average number of cases per year (2010 - 14)
Group 1	· Amebiasis	*Entamoeba histolytica*	21,966	12,333
	· Giardiasis	*Giardia lamblia*		
	· Other protozoan intestinal infections			
Group 2	· Bacterial food-borne illness	*Salmonella spp.*	15,307	8,536
	·Other illness from *Salmonella*	*S. enterica*		
	·Paratyphoid fever,	*S. typhi*		
	·Typhoid fever	*Shigella spp.*		
	·Shigellosis			
Group 3	·Intestinal infections from other organisms	Non well defined	72,528	42,222

Table S2. Spatial autocorrelation index for GDI and analyzed factors. The table shows the autocorrelation coefficient obtained from the Moran´s Index for each variable and period analyzed (2007-09; 2010-14). The asterisks show the level of significance: ***, 99%; **, 95%; *, 90%.

		Period 2007-09	**Period 2010-14**
Causal agent	All	0.356***	0.241***
	Group 1	0.226***	0.098***
	Group 2	0.269***	0.066***
	Group 3	0.379***	0.379***
Variables	Flooding	0.704***	0.764***
	Homes with dirt floor	0.5000***	0.287***
	Homes without water supply	0.519***	0.566***
	Homes without toilet facilities	0.093***	0.080**
	Homes without sewage	0.118***	0.138***
	% people < 18 years old	0.403***	0.331***
	Level of education	0.536***	0.483***
	Income Index	0.384***	0.389***
	Sewer drains	0.487***	0.487***
	Street food stands	0.224***	0.202***

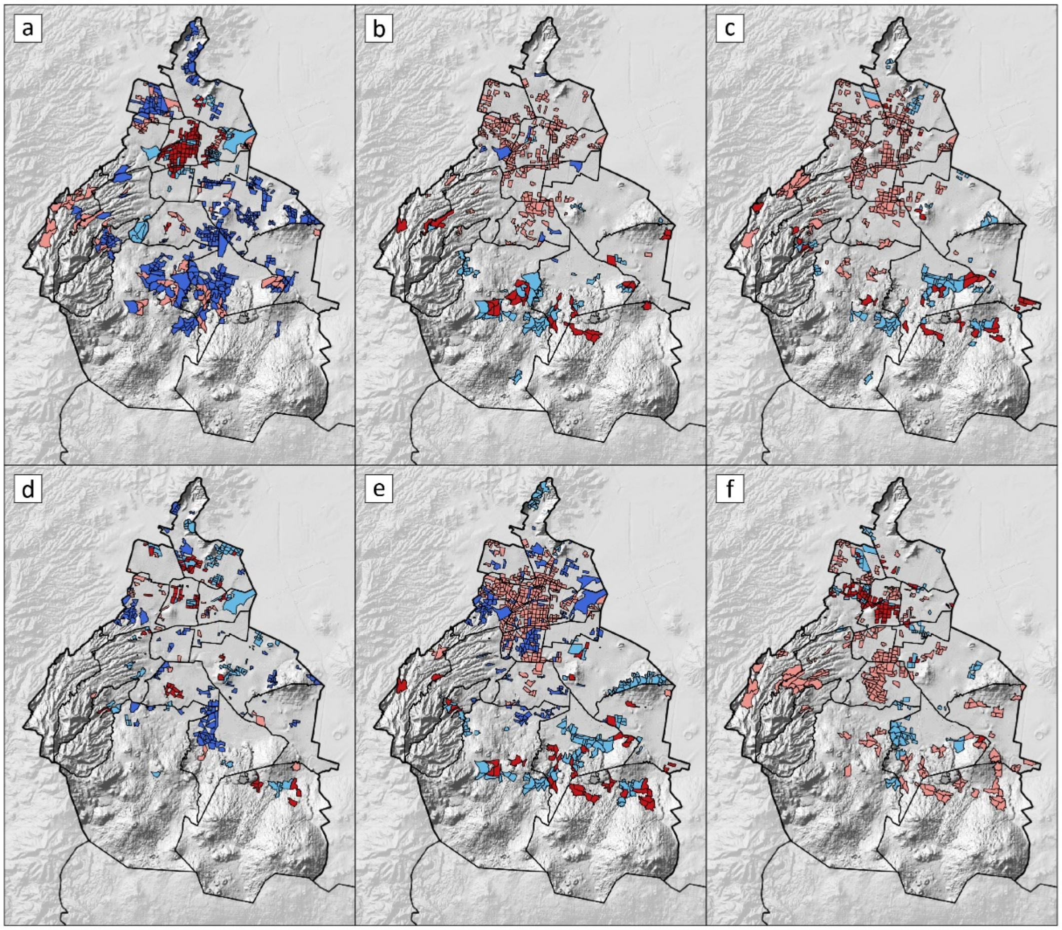

Figure S1(a). Bivariate LISA between Gastrointestinal Disease Incidences (GDI) and the analyzed factors for the period 2007-09. Each map shows the census blocks with significant LISA of GDI and ***(a) flooding***, ***(b) homes without water supply, (c) homes without sewer, (d) homes without toilet facilities, (e) homes with dirt floor, and (f) street food stands***. Each color represents one of the four types of possible association: (i) dark blue corresponds with ***low-low association*** which implies that low levels of GDI is significantly correlated with low values of the analyzed factor of the neighboring areas, (ii) light blue symbolizes ***low-high association*** which implies that low levels of GDI is significantly correlated with high values of the analyzed factor, (iii) pink symbolizes ***high-low association*** which implies that high levels of GDI in is significantly correlated with low values of an analyzed factor, and (iv) red symbolizes ***high-high association*** which implies that high levels of GDI in is significantly correlated with high values of an analyzed factor.

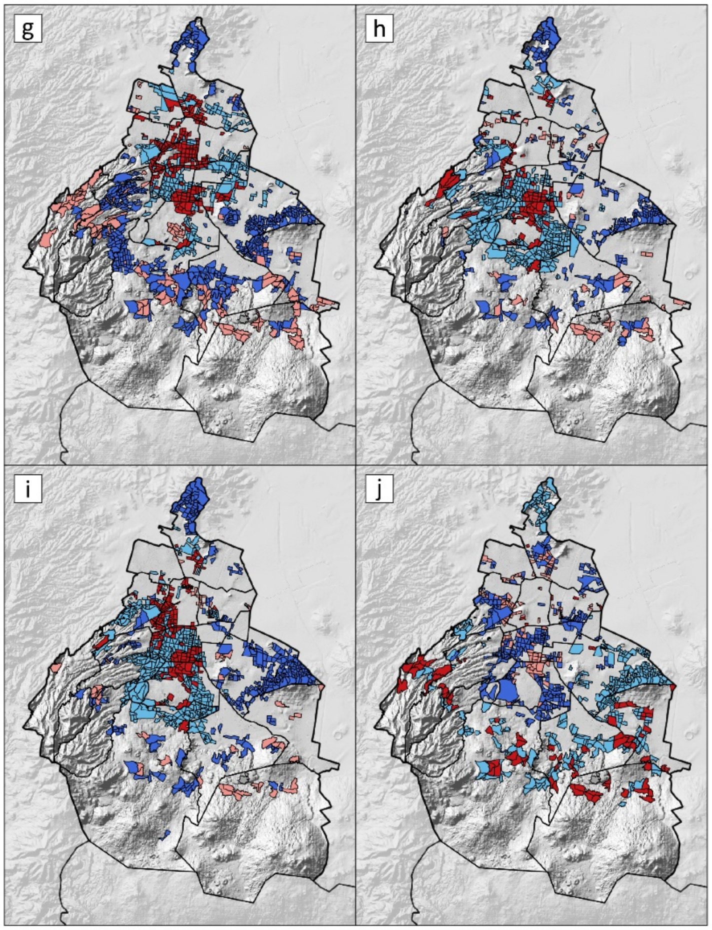

Figure S1(b). Bivariate LISA between Gastrointestinal Disease Incidences (GDI) and the analyzed factors for the period 2007-09. Each map shows the census blocks with significant LISA of GDI and ***(g) sewer drains, (h) income index, (i) level of education, and (j) % people < 18 years old***. Each color represents one of the four types of possible association: (i) dark blue corresponds with ***low-low association*** which implies that low levels of GDI is significantly correlated with low values of the analyzed factor of the neighboring areas, (ii) light blue symbolizes ***low-high association*** which implies that low levels of GDI is significantly correlated with high values of the analyzed factor, (iii) pink symbolizes ***high-low association*** which implies that high levels of GDI in is significantly correlated with low values of an analyzed factor, and (iv) red symbolizes ***high-high association*** which implies that high levels of GDI in is significantly correlated with high values of an analyzed factor.

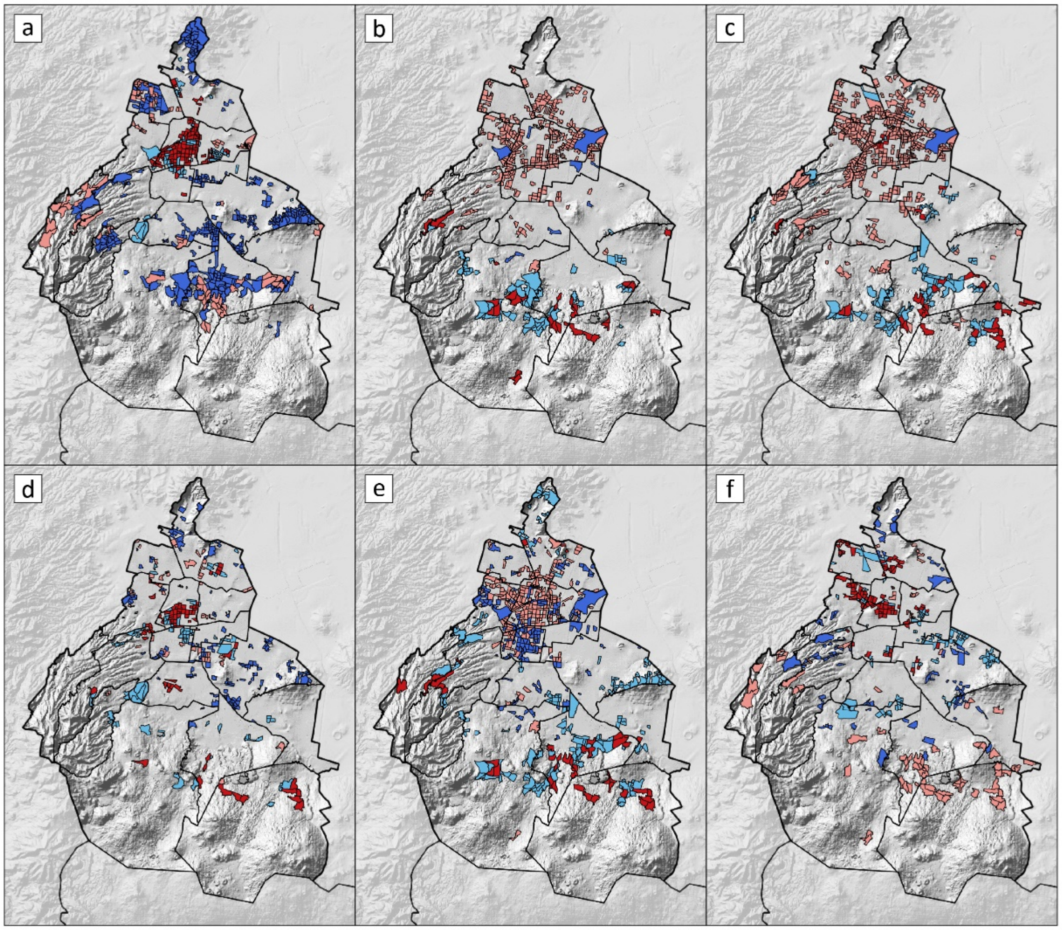

Figure S2(a). Bivariate LISA between Gastrointestinal Disease Incidences (GDI) and the analyzed factors for the period 2010-14. Each map shows the census blocks with significant LISA of GDI and ***(a) flooding***, ***(b) homes without water supply, (c) homes without sewer, (d) homes without toilet facilities, (e) homes with dirt floor, and (f) street food stands***. Each color represents one of the four types of possible association: (i) dark blue corresponds with ***low-low association*** which implies that low levels of GDI is significantly correlated with low values of the analyzed factor of the neighboring areas, (ii) light blue symbolizes ***low-high association*** which implies that low levels of GDI is significantly correlated with high values of the analyzed factor, (iii) pink symbolizes ***high-low association*** which implies that high levels of GDI in is significantly correlated with low values of an analyzed factor, and (iv) red symbolizes ***high-high association*** which implies that high levels of GDI in is significantly correlated with high values of an analyzed factor.

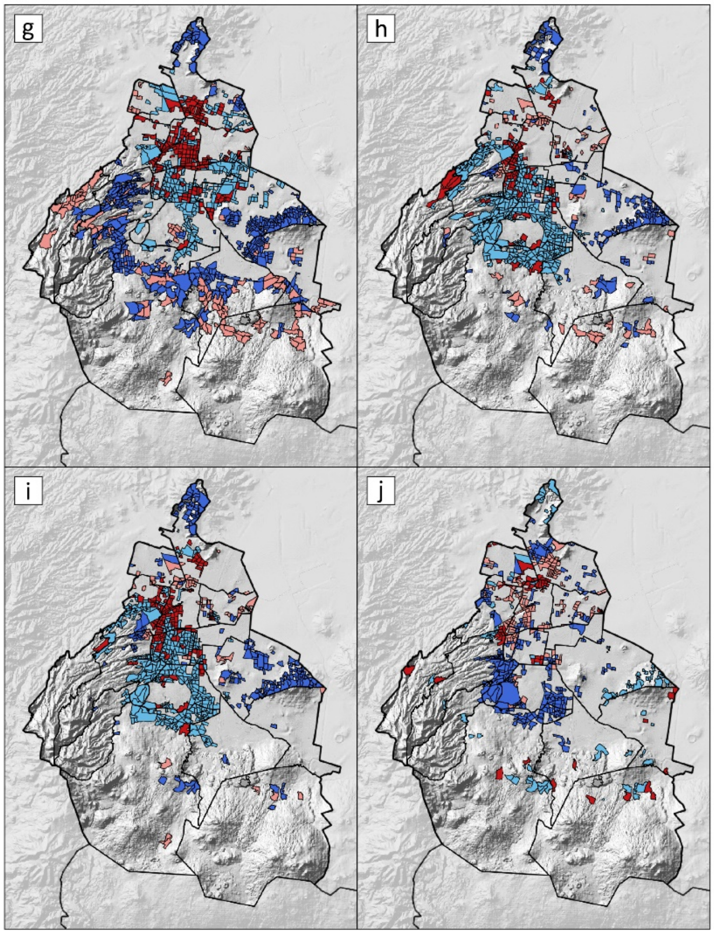

Figure S2(b). Bivariate LISA between Gastrointestinal Disease Incidences (GDI) and the analyzed factors for the period 2010-14. Each map shows the census blocks with significant LISA of GDI and ***(g) sewer drains, (h) income index, (i) level of education, and (j) % people < 18 years old***. Each color represents one of the four types of possible association: (i) dark blue corresponds with ***low-low association*** which implies that low levels of GDI is significantly correlated with low values of the analyzed factor of the neighboring areas, (ii) light blue symbolizes ***low-high association*** which implies that low levels of GDI is significantly correlated with high values of the analyzed factor, (iii) pink symbolizes ***high-low association*** which implies that high levels of GDI in is significantly correlated with low values of an analyzed factor, and (iv) red symbolizes ***high-high association*** which implies that high levels of GDI in is significantly correlated with high values of an analyzed factor.