Spatiotemporal Analysis of Arboviruses and Health Inequalities in Latin America.

Mabel Carabali

Department of Epidemiology, Biostatistics and Occupational Health

Faculty of Medicine

McGill University, Montreal, Canada

June 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy (Ph.D)

© Mabel Carabali, 2020

Table of Contents

Abstract

Dengue, chikungunya, and Zika are three different arboviruses transmitted to humans by the same vector, *Aedes* mosquitoes. These diseases present similar symptoms, are illnesses for which specific curative treatments do not exist, and for which sufficiently safe and effective vaccines are not yet introduced. Over the last 20 years, the burden of notified arboviruses has increased 15-fold worldwide. Particularly in Latin America, where Colombia and Brazil are experiencing up-to 60% of the overall arboviral burden. The heterogeneous distribution of these arboviruses across neighborhoods, socioeconomic strata (SES) and ethnic groups, suggests that social determinants of health (SDH) are playing a role in their presence and expansion. However, information on the effect and role of these SDH on the observed inequalities for arboviral diseases is still limited.

This manuscript-based thesis focuses on the assessment and quantification of social inequalities in the burden of dengue, Zika, and chikungunya in Brazil and Colombia from 2007 to 2017. This dissertation integrates approaches from social epidemiology and spatial statistics; using Bayesian spatiotemporal analysis with individual and aggregated data via hierarchical mixed models. The included set of analyses aimed at 1) identification of high-risk disease areas; 2) estimation and decomposition of socioeconomic disparities, and 3) the estimation of between and within ethnic disparities.

The first manuscript presents the application of a joint spatial marked point process model for non-severe and severe dengue in Colombia. This method, which has not been used before in this context, analyses the spatial location of cases using individual and area-level data simultaneously. This method allowed the identification of key sociodemographic factors (age, SES, and distance between cases) and clustering, accounting for spatial autocorrelation and uncertainty in surveillance data. The second and third manuscripts include the assessment and decomposition of SES-inequalities on arboviruses in Brazil and Colombia. Using the Relative and Absolute Concentration Index of inequality, the second manuscript assesses the SES-inequality and documents their temporal trend, describing the presence of a non-monotonic distribution of cases across the SES distribution and changes in the magnitude of inequalities during outbreaks. The

third manuscript shows the results of the decomposition analysis, indicating that year of notification, age, presence of healthcare facilities, and sanitation are key contributors to the overall SES-inequality on arboviral diseases in all study sites.

The last manuscript examines the overall and severe dengue burden across and within different ethnic groups. Despite that African Ancestry is considered "protective" for severe dengue, several studies in Latin America showed increased severity and mortality among self-identified Black or African descendants. To study this phenomenon in the Colombian context, I used spatiotemporal hierarchical models correcting for underreporting and misclassification. The results showed a small increase of severity among Afro-Colombians that was not observed when correcting for underreporting and misclassification. It is possible to consider then, that the paradoxical findings are likely related to the differential reporting among Afro-Colombians and intersectionality, linked to differential effects of SES and access to health care across and within ethnic groups.

This thesis contributes to the body of evidence about health inequalities on arboviruses by providing robust estimates about the socioeconomic, ethnic, and spatial distribution of arboviral cases in Latin America. My approach, which is particularly quantitative, has the capacity to expand and improve upon the current body of evidence about arboviruses and other infectious diseases. The methods and findings presented in this thesis could be used in other endemic epidemic settings with similar sociodemographic characteristics for policy making. Specifically, to identify areas of constant presence of arboviruses and for targeting strategies to decrease disparities at the local level.

Résumé

La dengue, le chikungunya et le Zika sont trois arbovirus différents transmis aux humains par le même vecteur, les moustiques *Aedes*. Ces maladies présentent des symptômes similaires, et il n'existe pas de traitements curatifs spécifiques ni de vaccins avec une efficacité éprouvée. Durant les 20 dernières années, le nombre des cas notifiés d'arbovirus a augmenté de 15 fois dans le monde. Cette augmentation est plus marquée en Amérique latine, où la Colombie et le Brésil représentent jusqu'à 60% du fardeau global des arboviroses. La distribution inégale des cas d'arbovirus à travers les quartiers, les strates socioéconomiques (SSE) et les groupes ethniques suggère que les déterminants sociaux de la santé ont un rôle important dans leur propagation. Cependant, les informations disponibles sur l'effet et le rôle de ces SDH sur les inégalités dans la distribution des arboviroses sont incomplètes.

Cette thèse évalue et quantifie les inégalités sociales du fardeau de la dengue, du chikungunya et du Zika au Brésil et en Colombie (2007-2017). Utilisant des données individuelles et agrégées de surveillance épidémiologique, des approches tirées de l'épidémiologie sociale et des analyses spatio-temporelles Bayésiennes ont été mises à profit pour 1) l'identification des zones à haut risque, 2) l'estimation et la décomposition des disparités socio-économiques, et 3) l'estimation des disparités entre diffèrent groupes ethniques.

Le 1er manuscrit montre l'application d'un modèle spatial des processus ponctuels permettant l'analyse de la localisation des cas en utilisant simultanément les données individuelles et agrégées. Cette approche novatrice a permis d'identifier les principaux facteurs sociodémographiques et les grappes spatiales, en surmontant les limites méthodologiques associées à l'unique utilisation de données agrégées. Les 2e et 3e manuscrits consistent à évaluer et décomposer des inégalités sociales des arboviroses au Brésil et en Colombie. Le 2e manuscrit en utilisant les indices de concentration relative et absolue, évalue les inégalités sociales et leurs tendances temporelles. Il décrit la présence d'une distribution non monotone des cas à travers la SSE et des changements dans l'ampleur des inégalités pendant les épidémies. Le 3e manuscrit utilisant des analyses de décomposition, indique que l'année du report, l'âge, et la présence

d’établissements de santé sont les principaux contributeurs de l'inégalité socioéconomique des arboviroses.

Le dernier manuscrit examine le fardeau de la dengue à travers et à l’intérieur des groupes ethniques. Avoir ascendance Africaine est considéré comme « protectif » pour la dengue sévère, mais dans l’Amérique Latine plusieurs études ont montré une sévérité accrue parmi les personnes auto-identifies noires ou d’ascendance Africaine. Pour étudier ce phénomène dans le contexte colombien, j'ai utilisé des modèles spatio-temporels mixtes, corrigeant la sous-déclaration et la classification erronée des cas. Les résultats indiquent que le risque élevé de la sévérité chez les Afro-Colombiens disparaissent après la correction pour la sous-déclaration. Cependant, cette tendance paradoxale, serait probablement liée aux différences dans le processus de notifications des cas parmi les Afro-Colombiens et à l’intersectionnalité entre SSE, recours aux soins de santé et l’ethnicité.

Cette thèse contribue à l’étude des inégalités sur les arboviroses en fournissant des estimations robustes sur leurs distributions socioéconomique, ethnique et spatiale en Amérique latine. L’approche, notamment quantitative de cette thèse peut permettre d'élargir et d'améliorer les connaissances actuelles des arboviroses et autres maladies infectieuses. Les résultats et surtout les méthodes présentées dans cette thèse pourraient être utilisés dans d’autres contextes endémiques ayant des caractéristiques sociodémographiques similaires, pour l’élaboration des politiques publiques. Plus précisément, ces résultats et méthodes pourraient être utilisez pour l’identification des zones de présence constante d'arboviroses et élaboration des stratégies ciblées, visant à réduire les disparités.

List of Abbreviations and Acronyms

ACI: Absolute Concentration Index of Inequality
ADE: Antibody Dependent Enhancement
AR1: Autoregressive specifications of first order
AR2: Autoregressive specifications of second order
CSDH: Commission on Social Determinants of Health conceptual framework
BYM: Besag-York-Mollie spatial model espacification
DAG: Directed Acyclic Graph
DIC: Deviation Information Criterion
ELISA: Enzyme Linked Immunosorbent Assays
HDI: Human Development Index
iCAR: intrinsic Conditional Autoregressive Structure
IgM: Immunoglobulin M
IgG: Immunoglobulin G
INLA: Integrated Nested Laplace Approximation
IQR: Interquartile Range
IRR: Incidence Rate Ratio
MCMC: Monte-Carlo-Markov-Chain simulation
MC-SIMEX: Simulation Extrapolation for Misclassification
MPP: Marked Point Process Model
OR: Odds Ratio
PCR: Polymerase Chain reaction
PPM: Point Process Model
RCI: Relative Concentration Index of Inequality
RT-PCRT: real-time reverse transcription polymerase chain reaction
RR: Relative Risk
SES: Socioeconomic Status
SD: Standard Deviation
SINAN: Sistema de Informação de Agravos de Notificação/National Surveillance System Brazil
SIVIGILA: Sistema de Vigilancia Epidemiológica/National Surveillance System Colombia
SRR: Standardized Rate Ratio
RW1: models with independent increments or 'random walk' of first order
RW2: models with independent increments or 'random walk' of second order
RR: Risk Ratio
95%CI: 95% Confidence Interval
95%Cr.I: 95% Credible Interval

List of Tables

List of Figures

Acknowledgements

I am very grateful for this opportunity and for the support of a (long) list of incredible people who made this possible. First, I am immensely grateful for the academic supervision, guidance and constant mentorship of Dr. Jay S. Kaufman. Thank you for opening the door and listening to my proposal, supporting me throughout the training, cheering me up when need it (meaning all the time) and answering my many (senseless) questions. Thank you for the prompt, accurate and honest feedback. Thank you being there always and for understanding me even when I do not, whether I speak Spanish, French or as a clinician, you always understand me. Thank you for the patience!

To my thesis committee members: Thank you for your time, patience and dedication to guide me and help me with this project. To Dr. Alexandra Schmidt, thank you for the valuable and constant statistical guidance, thank you for making me appreciate and (try to) understand the intricacies of spatial statistics. To Dr. Sam Harper, thank you for your support throughout the understanding of health inequalities and helping me to overcome the methodological challenges faced during this process. To Dr. Mathieu Maheu-Giroux, thank you for your honest support and advice on the spatiotemporal epidemiological methods since the very early stages of this process. To my field collaborators, infinitely thanks to Dr. Berta Nelly Restrepo, Dr. Antonio Silva Lima-Neto, Dr. Geziel dos Santos de Sousa and Dr. Andrea Caprara for their constant support and collaborations with data access/collection and the interpretation of the research findings in context.

To the department of Epidemiology, Biostatistics and Occupational Health (EBOH), thank you for giving me the opportunity to pursue my dream in this outstanding environment. Thanks to the chair Dr. Paradis and to all faculty for creating and maintaining such an exceptional program. The quality of the coursework and research training offered by the EBOH department provided me with the skills needed to achieve this goal. I want also to thank the EBOH's administrative staff, especially to Andre-Yves Gagnon, Katherine Hayden, Dolores Coleto, and Marie Boncoeur for helping me throughout the process.

To my cohort, a.k.a: *The best cohort ever!* Thank you for this experience, for the endless in-class discussions, the potlucks, drinks, and all the fun provided to make this more bearable. I can't thank enough my musketeers: Hannah and Agustin, partners in science and fun who understood and supported me all time! Thanks to Brice for the constant constructive academic criticism. Thanks also to my gym and study buddy Julie and to my writing buddy Joanne.
Special thanks to BBK for everything!!!

To Dr. Lyda Osorio, for her mentorship, guidance at early stages of my training as epidemiologist and currently as colleagues. To Dr. Kate Zinszer for her support and trust trough the different research venues embarked during this time.

To the team of the Rheumatology and Lupus clinic at the Montreal's General Hospital, especially to my rheumatologist Dr. Louis Pierre Grenier and R.N. Marli Vilsaint, for the care provided to keep me healthy enough and functional through this process!

To my family: Gracias a mis padres por la vida, por inculcarme valores de integridad, fomentar mi curiosidad y apoyar mi amor por la ciencia y por apoyarme en todos los caminos que he decidido recorrer! Por enseñarme que el trabajo duro y la constancia son recompensados con tranquilidad y satisfacción personal. A mis hermanos Danilo y Carlos, por su amor, paciencia, constante motivación e incondicional apoyo. A todos mis ti@s en especial a Amparo, Alonso, Deicy, Jaime y a mis primas-comadres-hermanas Maira y Diana por su cariño, atenciones, tolerancia y constante presencia en mi vida. A mi familia de otá: Lic. Rigoberto Rocha, Lic. Galia García, Florencio Jiménez y Juan Carlos Lara-Ayun por el constante apoyo y soporte espiritual. A mi familia cubana, Aquiles, Nereyda, Galia y Anthony por el amor incondicional y la alegría constante en medio de la adversidad. A mi hermana de la vida, Lic. Alicia Reyes, por sus oportunas palabras y consejos sobre la economía de los arbovirus y la satisfacción del alma de forma simultanea.

A todos mis amig@s, en especia al Dr. Andres Felipe Sanchez por conocerme al derecho y al revés, motivarme y caminar conmigo desde que decidimos ser médicos; a la Lic. Lorena Segura por mantener mi hemisferio izquierdo funcionando; a Gina Tatiana Vásquez, por su sabiduría, amistad honesta y transparente que ha superado situaciones extremas. A la Dra. Neila Julieth Mina, Dra. Oriana Bueno, Dra. Adriana Cárdenas y Dra. Lissette Chan por su apoyo y por caminar y volar conmigo a cualquier lugar del mundo. A Islene y Ceci por las cenas románticas en medio del frio canadiense. Al Dr. Ronald M. Asprilla y a Vanessa Orejuela, por estar siempre presentes.

To honor to the past, this thesis is *in memoriam* of all the people I have lost through life and during this training, specially, Angelino Carabali and Cecilia Diaz.

To honor the future, this thesis is dedicated to my three life-engines: Vanessa, Maria José y Mariana. I want to show you that education is power and that being black, immigrant, and woman should not be barriers to achieve your dreams. You are the reminder that life has delightful things to offer and the reason to smile and be hopeful when things are rough.

Statement of financial support

I was very fortunate to receive several sources of financial support for my doctoral training. For my first two years (2015-2016), I received a Tomlinson Doctoral Fellowship from the Faculty of Medicine. For the remaining of my training (2017-2020) I received the Canadian Institutes of Health Research's (CIHR) Frederick Banting and Charles Best Canada Graduate Scholarship Doctoral Award.

I presented my work at nine international scientific conferences with the funding received from various Graduate Research Enhancement and Travel (GREAT) Awards from the Department of Epidemiology, Biostatistics, and Occupational Health (EBOH) and travel allowances included in the Best-Banting CIHR Doctoral Award. For data collection and field visits I received the McGill's GPS's Graduate Mobility Award (2018) and I also received travel funding from the Society for Epidemiologic Research (SER) through my participation in their 2019 doctoral dissertation workshop.

During my doctoral training, I created a social determinants and health inequalities on arboviruses course, taught at the international dengue course in Havana, Cuba in 2017 and 2019. For the creation and participation of this event, I received partial funding from the Quebec-Cuba cooperation and the Ministère des Relations Internationales et de la Francophonie (MRIF) (Grant: RI000456).

Contribution of Authors

The individual manuscripts that constitute this thesis and the respective contribution of the co-authors are listed below. The data used in the analyses were routinely collected surveillance data from the study sites and outlined in greater detail in the methods overview and each manuscript. I developed the research questions, research plan and study design for all of the included manuscripts. I received substantive support from my academic supervisor Dr. Jay S. Kaufman, who provided indispensable guidance for the protocol development, assisted me to refine the research questions, and help me to build suitable analysis plans for all manuscripts. I was solely responsible for collecting and merging the data, developing the statistical functions or methodological modifications to conduct the analyses, and generating the manuscripts.

My committee members, Dr. Alexandra M. Schmidt, Dr. Sam Harper and Dr. Mathieu Maheu-Giroux contributed substantially to the development of the protocol and in each of the manuscripts as described below:

Manuscript 1: Carabali M, Schmidt AM, Restrepo BN, Kaufman JS. "A joint spatial model for dengue and severe dengue in Medellin, Colombia". Currently under consideration at the Spatial and Spatio-temporal Epidemiology journal SSTE-D-20-00079 (June 2020)

I reviewed the literature on spatial point process models, conceived and designed the study, conducted data collection, data management (cleaning and geocoding), completed the data analysis, interpreted the results and wrote the original manuscript. Dr. Schmidt provided substantive statistical and analytical guidance and support for the innovative use and application of this methodology in the field of arboviral diseases. Dr. Restrepo provided the data, facilitated the communication with the local health agencies, and participated in the epidemiological analysis and interpretation of the results. Dr. Kaufman guided the epidemiological design and contributed substantially to the analysis and interpretation of the results. All authors provided critical feedback and approved the final version of the manuscript.

Manuscript 2: Carabali M, Harper S, Lima-Neto A, de Sousa GdS, Caprara A, Restrepo BN, Kaufman JS. "Spatiotemporal Distribution and Socioeconomic Disparities in Dengue, Chikungunya and Zika in Latin America". Currently under consideration at Tropical Medicine and International Health ID: TMIH-D-20-00380 (June 2020).

I reviewed the literature on health inequalities indexes, conceived and designed the study, conducted data collection, data management, developing the statistical functions for the application of the analytical methods in this context, completed the data analysis, interpreted the results and wrote the original manuscript. Dr. Harper provided substantive analytical and methodological inputs for the assessment and interpretation of the inequalities. Dr. Lima-Neto, Dr. de Sousa, Dr. Caprara, and Dr. Restrepo facilitated data collection, the communication with the local health agencies, and participated in the epidemiological analysis and interpretation of the results. Dr. Kaufman guided the epidemiological design and contributed substantially to the analysis and interpretation of the results. All authors provided critical feedback and approved the final version of the manuscript.

Manuscript 3: Carabali M, Harper S, Lima-Neto A, Caprara A, Restrepo BN, Kaufman JS. "Decomposition of Socioeconomic Inequalities on Arboviral Diseases in Brazil and Colombia (2007-2017)". Submitted to the Pan American Journal of Public Health ID: 2020-00959.

I conceived and designed the study, developed the statistical functions for the application of the methods in this context, completed the data analysis, interpreted the results and wrote the original manuscript. Dr. Harper provided substantive analytical and methodological inputs for the interpretation of the inequalities. Dr. Lima-Neto, Dr. Caprara, and Dr. Restrepo participated in the epidemiological analysis and interpretation of the results. Dr. Kaufman guided the epidemiological analysis and contributed substantially to the interpretation of the results. All authors provided critical feedback and approved the final version of the manuscript.

Manuscript 4: Carabali M, Maheu-Giroux M, Kaufman JS. "Dengue, Severity Paradox and Socioeconomic Distribution among Afro-Colombians: Spatiotemporal Analysis of surveillance-reported dengue cases in Cali, Colombia (2012-2017)". Submitted to Epidemiology journal ID: EDE20-0461 (June 2020). Revisions sent on August 2020.

I reviewed the literature on ethnic disparities and severity on dengue, conceived and designed the study, conducted data collection, data cleaning, completed the data analysis, interpreted the results and wrote the original manuscript. Dr. Maheu-Giroux provided substantive guidance on spatiotemporal analysis using the Bayesian framework on vector-borne diseases, the epidemiological analysis and the interpretation of the data. Dr. Kaufman guided the epidemiological design and contributed substantially to the analysis and interpretation of the ethnic-inequalities results. All authors provided critical feedback and approved the final version of the manuscript.

Statement of originality

The work presented in this dissertation thesis is based on original and timely contributions to the field of health inequalities on arboviral diseases, achieved in four manuscripts conceived and executed by me. I received indispensable support and guidance from my supervisor and thesis advisory committee members, but the research questions, objectives and the analyses performed to answer my overall research goal are of my own. I addressed the identified research gaps in the literature of health inequalities on arboviral diseases, by adapting and applying existent health inequalities methods and spatiotemporal methods in this context, which have not been done before.

Specifically, in manuscript 1, I applied a spatial statistics method from the field of ecology to assess the spatial distribution of severe and non-severe cases using individual data simultaneously. This approach has not been used before in the field of arboviruses and presents several advantages due to the use of granular data, accounting for uncertainty of data to improve precision. In manuscript 2 and 3, I present one of the first quantification of the socioeconomic inequalities on arboviral diseases and the contribution of key determinants, accounting for the spatiotemporal distribution and documenting the trend over time. For this purpose, I created and modified software functions for the estimation and decomposition of health inequalities from Bayesian spatiotemporal models. In manuscript 4, I presented the analysis of ethnic disparities accounting for different sources of bias including underreporting and misclassification. Overall, I used several methods to improve the use and interpretability of surveillance data for the analysis of arboviruses and health inequalities.

In addition to generating new evidence on health inequalities on arboviruses, this dissertation offers examples of novel application of existent methods and provides the scientific community with methodological tools that could be used to facilitate the integration of spatiotemporal models in the assessment of health inequalities.

Chapter 1: **Introduction**

During the last decade an increased burden of morbidity and mortality due to dengue, chikungunya, and Zika have been observed in Latin America[1-8]. Among these three diseases, dengue has the highest incidence worldwide[6,7,9]. However, the Americas region experiences the second largest burden of symptomatic notified cases, with outbreaks occurring every three to five years[6,9]. In 2016, there were 2.38 million dengue cases reported and in 2019, the region recorded the highest outbreak in the history with 3.1 million cases and over 25,000 severe cases[6,7,9]. Chikungunya and Zika are considered re-emergent arboviruses and were introduced in the Americas in 2013 and 2015, respectively[1,3]. Chikungunya's outbreak included over a million cases in 2014 and Zika, with over 650,000 cases in 2016 was considered a Public Health Emergency of International Concern (PHEIC), due its association to congenital malformations[1-3,7,10].

Dengue, chikungunya, and Zika are three different arboviruses but transmitted to humans by the same vector: *Aedes sp*. mosquitoes[1,5]. These three diseases have a similar symptomatology, are illnesses for which specific curative treatments do not exist and for which sufficiently safe and effective vaccines are not yet introduced[1,2]. Apart from the shared clinical symptomatology and management challenges, the social and environmental characteristics associated to their presence are also similar. However, in the Americas region, where Colombia and Brazil experience up to 60% of the overall arboviral burden[1,2,4,10]; a heterogeneous distribution of arboviral diseases across socioeconomic and ethnic groups has been observed[2,11,12].

In Colombia, 80% of the dengue deaths during 2017 were among people in low socioeconomic position[12-14]. Likewise, in Brazil, the risk of dengue death is 44% higher among people with low or no education compared to people with more than four years of schooling. Interestingly, despite that African ancestry is considered a protective factor for dengue severity[2,15-17]; severity and lethality are reported to be higher among Black people or people with African ancestry in Colombia and Brazil[15,18-22]. Therefore, in addition to the presence of shorter interepidemic periods (i.e.,

outbreaks occurring more often), the identification of places and the type of populations in which these diseases concentrate is a major public health concern. However, despite the differential distribution of arboviruses within and across different places, socioeconomic strata, and ethnic groups, the analysis of the presence and trends of inequalities on dengue, chikungunya and Zika in Latin America is still limited[23,24].

Important barriers for the identification and understanding of health inequalities on arboviral diseases relate to data availability and methodological challenges, including limited adjustment for the endemic-epidemic character of these diseases, among others. Although some studies have investigated the relationship between dengue and poverty and arboviruses and low socioeconomic status, most of the literature is only descriptive or has not been analyzed within the specific lenses of health inequality. This dissertation aims at identifying, quantifying and assessing the trend of inequalities associated to dengue, Zika and chikungunya, addressing methodological challenges related to the use of surveillance data and accounting for the spatiotemporal distribution of these diseases simultaneously.

1.1 Research Objectives

The overall goal of this doctoral dissertation was to generate relevant evidence on the presence and magnitude of inequalities in dengue, chikungunya and Zika in Latin America, while accounting for their spatiotemporal distribution. In addition, this dissertation was intended to contribute to advance scientific knowledge in the analysis of inequalities on infectious diseases through the application, modification and integration of existent epidemiological and spatiotemporal methods in the context of arboviral diseases. The four specific research objectives of this thesis were:

1. To identify high-risk dengue areas while modelling simultaneously the overall distribution of dengue cases and their severity in Colombia, using individual data within a joint spatial point process model (Manuscript 1).

2. To assess the presence, magnitude and pattern of socioeconomic disparities in arboviral diseases at the neighborhood level in Brazil and Colombia, using relative and absolute indexes of inequality (Manuscript 2).

3. To identify the social determinants that are contributing the most to socioeconomic inequalities in arboviral diseases in Brazil and Colombia, by decomposing the concentration index of inequality (Manuscript 3).

4. To assess the presence of disparities in the distribution of overall and severe notified dengue cases across and within ethnic groups, accounting for the socioeconomic status using individual and aggregated data from Colombia (Manuscript 4).

1.2 Thesis Structure

This manuscript-based dissertation contains eight chapters. In chapter 1, I present the overall rationale and research objectives. In chapter 2, I provide information on the current state of the literature on arboviruses in the Americas region and other relevant contextual information for the specific research objectives discussed in this thesis. In chapter 3, I present an overview of the data and the analytical methods used to address each of my research objectives. Specifically, I present the use of Bayesian hierarchical spatiotemporal models, the approach and challenges on the use of surveillance data and the assessment of health inequalities. In chapter 4, I present the novel application of a point process analysis on dengue and severe dengue in Colombia. In chapter 5, I present the identification and quantification of socioeconomic inequalities in Brazil and Colombia using relative and absolute index of inequality. In Chapter 6, I present the decomposition of the relative index of inequality, with the identification of the social determinants contributing to the overall inequality on arboviruses in Brazil and Colombia. In Chapter 7, I present the analysis of patterns of dengue and severe dengue distribution across and within ethnic groups and socioeconomic strata in Colombia. Finally, in Chapter 8, I summarize the key findings from this dissertation and the points for discussion, offering suggestions for the application of methods and the opportunity for future research on the field.

Chapter 2: **Review of the literature**

Infectious diseases account for an important proportion of disease burden in developing countries, being the third cause of morbidity after cardiovascular diseases and cancer or violence[21,25-28]. Between 2013 to 2016, there were 11 million arboviral cases in the Americas region, with Colombia and Brazil being among the top five countries most affected by the presence of arboviruses[4,6,7,9,27,29-38]. In Colombia, in 2015, arboviral burden exceed the burden of tuberculosis and HIV/AIDS combined[33,34]. To provide background information on the current body of literature about arboviruses and the research gaps addressed in this dissertation, this chapter summarizes aspects related to the presence and distribution of dengue, chikungunya and Zika, advances in research and shortcomings faced to control these arboviruses and the identified knowledge gaps.

2.1 Arboviral Diseases

Dengue, chikungunya and Zika are arthropod-borne viral diseases transmitted to humans by the bites of *Aedes* sp. mosquitoes[1,3,9,39]. These three diseases are important public health concerns worldwide given the widespread geographical distribution and disease burden[7,39,40]. Despite the increased global burden of all three arboviruses during the last decade[1,3,4]; dengue has the highest incidence and is considered a major cause of morbidity and mortality in tropical and sub-tropical areas with 128 countries at risk[1,4,40]. Recent studies estimated that there are between 105 to 390 million dengue infections expected every year with lower-bound credible intervals ranging from 94.5 to 284 million and upper-bound credible intervals ranging from 113.6 to 528 million cases[6,8,9,41,42]. These estimates include the presence of 96 million symptomatic cases, 3.6 million hospitalizations (95% confidence Interval (95%CI) = 2.3, 4.6 million) and 20,000 deaths[8,40,42,43]. The Americas region experiences the second largest burden of dengue worldwide[40,42,43]. In 2015, from the 3.2 million cases reported, 2.35 million cases were from the Americas region alone, including 10,200 severe cases and 1,181 deaths[2,5]. In 2016, from the 2.38 million cases reported in Latin America, 1.5 million where from Brazil[5,44]. And, in 2019, the Americas recorded the highest number of dengue cases in its history, with 3.1 million notified cases and more than 25,000 severe cases[38]. Likewise, the potentially hospitalized dengue cases was estimated to be 556,800 cases (95%CI= 361,200, 702,300) in the Americas region[8].

Chikungunya virus is present in more than 60 countries and was introduced in the Americas region in 2013 from Asia and Oceania[1,3,45]. There were more than 1.5 million cases reported during its introduction in the Americas between 2013-2014[3,45,46]. Since then, there are more than 300,000 symptomatic cases reported every year[1,4,45,46]. Similarly, Zika virus was recently introduced in the Americas region (July 2015), from the French Polynesian islands[1,10,44]. Although it was considered an innocuous virus previously, Zika constituted a public health emergency of international concern in 2016 due to a massive outbreak of microcephaly and the rapid spread via sexual intercourse[47]. There are currently 84 countries reporting Zika presence and by the end of 2017 there were 221,520 confirmed cases. About 60% of cases were from Brazil and 87% of the microcephaly and congenital Zika syndrome were from Brazil and Colombia[10].

The clinical presentation for all three diseases ranges from self-limited mild febrile illness including fever, headache, lethargy, and myalgia typically lasting one to two weeks (**Figure 2.1**)[19-21]. Distinguishing features of dengue infection include nausea, vomiting, and headaches whereas for chikungunya infection include incapacitating severe joint pain, and for Zika infection, the presence of conjunctivitis[22,23]. A significant portion of infected individuals with chikungunya, dengue, and Zika remain asymptomatic, with reports of up to 30% of asymptomatic infections for chikungunya, 50% for dengue, and 80% for Zika[24,25]. These diseases provide life-long lasting immunity after infection, so that each disease can only occur once in a lifetime[1-3]. In the case of dengue, however, the disease could occur theoretically up to four times because it is caused by four serotypes (DENV1, DENV2, DENV3,and DENV4), but it is estimated that the majority of the people get only two infections in their lifetime[2,48].

Severe outcomes resulting from dengue infection include hemorrhage, shock, and death[1,2]. Secondary or subsequent infections from different dengue serotypes increase the risk of severe dengue through antibody-dependent enhancement (ADE), and severe dengue occurs in approximately 1% of all dengue cases with a 20% mortality rate[9,19,20,49]. There is evidence of post-dengue syndrome in recovering patients that is characterized by persistent fatigue, joint pain, myalgia, and malaise for up to two years following their illness. For chikungunya, adverse

outcomes include long-term arthritis and death[26], and for Zika, severe outcomes include the presence of Guillain-Barre syndrome and congenital Zika syndrome[2-4,45,46,50]. Congenital Zika syndrome is a set of congenital malformations, primarily of the Central Nervous System including microcephaly, anencephaly, ophthalmologic and musculoskeletal malformations, that has been causally related to Zika virus exposure in fetuses[44,51,52].

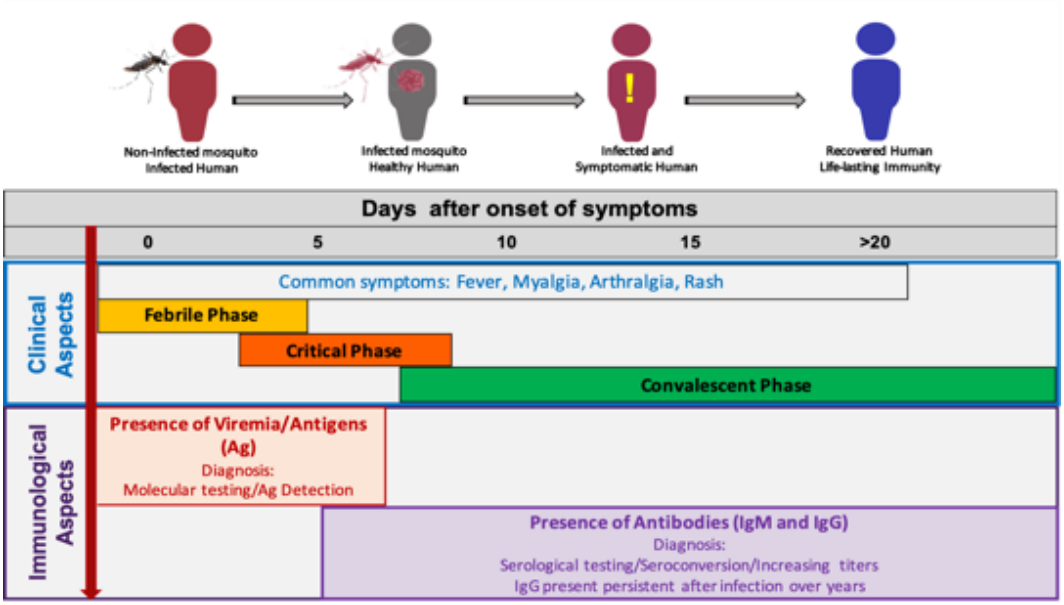


Figure 2-1 Schematic of arboviral transmission and key clinical and immunological aspects common to dengue, chikungunya and Zika.

Together, dengue, chikungunya and Zika represent an important economic problem worldwide[29-31,34,53-59]. At the global and country level, the estimated economic burden is approximately US$8.9 billion for dengue[31,53,55], US$73.6 million for the cost of a chikungunya outbreak[57], and around US$2.3 billion per year for Zika[60]. At the individual level, the cost of non-severe dengue in low and middle income countries could range from US$13 to US$385 per episode, depending largely on the aspects (direct medical costs) covered by the health system in each country[31,53,54,56]. The median direct medical cost for chikungunya in children was reported to be US$257.9 in Colombia but was reported to be US$2,105 in the Latin American Caribbean[29,34,57,58]. In addition, the median cost for Zika could be up to US$6,751 for children born with microcephaly and up to US$12,368.28 for Zika patients with Guillain-Barre syndrome[34,58,59,61].

2.2 Previous advances & shortcomings

Given the important morbidity and economic burden of dengue, chikungunya and Zika, and the social impact that these arboviruses pose, there is interest in finding alternatives to prevent, manage and control these diseases. Over the years, several approaches have been considered as alternatives for arboviral control. Such alternatives include improvement in diagnosis (using sensitive criteria for case definition and developing diagnostic tests that are accurate, affordable, and easy to use), promoting environmental and sanitation strategies to decrease the presence of vectors, promoting changes in behaviors across communities, and using integrative strategies with eco-bio-social approaches intended for a holistic disease control. However, despite the numerous efforts made to do so, there are several important challenges that are important to mention and therefore synthetized below.

To begin with, one of the main challenges for arboviral management is that there is no current specific curative treatment for any of these three conditions[2,3,48,62,63]. The clinical management relies on controlling symptoms, including antipyretics for fever and oral or intravenous fluids to avoid shock or death[1,2,43]. Anti-pain medication should be cautiously administered because the use of certain non-steroids anti-inflammatories may aggravate the condition by increasing the likelihood of hemorrhage[1,2,4,5]. Although a dengue vaccine has been licensed recently in some South-East Asian and Latin American countries[62,64], several epidemiological and clinical aspects should be taken into account before considering it as an alternative[65]. First, dengue vaccine's efficacy is low and varies widely across countries[66]. Second, its utilization has not been approved for large-scale administration because it is not considered safe in children under 9 years of age, requires a dengue seroprevalence of 70%, and the presence of a strong pharmaco-surveillance system, among others[65-68]. In the case of Zika and chikungunya, there are several vaccine candidates in the pipeline but none available for use so far[69].

Given the limitations associated with the clinical management of these arboviruses and the absence of a vaccine that could efficiently control the burden of these diseases[2], vector control has been considered the main mitigating strategy[5,70]. The principal mosquito vectors, *Aedes*

aegypti and *Aedes albopictus*, are opportunist mosquitoes adapted to urban environments and require stagnant water, usually formed by small containers, for different stages of their life cycles[7,39,71]. Although there are guidelines and established integrated vector-control and management strategies for dengue, chikungunya, and Zika, the implementation and evaluation of their impact is challenging, context dependent (e.g., geographic, cultural, and epidemiological), and in certain cases expensive (ranging from US$5.62 to US$73.5 million)[30,34,55,57,61,72-74]. Such vector control efforts include i) the implementation of mass insecticide spraying campaigns; ii) community social mobilization where population is encouraged to prevent the apparition of mosquitoes breeding sites by cleaning and covering systematically any water containers; iii) preventive measures such as the use of repellents, insecticide-impregnated curtains (screens)[73,75]; and during the last five years iv) the release of genetically-modified and sterile mosquitoes[76,77].

Although there are no randomized trials evaluating the effect of fogging on the incidence of these arboviruses, a meta-analysis showed that the measurement of the effect of the impact of fogging or indoor spraying on dengue incidence is imprecise (RR=0.67; 95%CI=0.22, 2.11) and that entomological indices (number or proportion of mosquitoes) are not consistently associated to disease rates[78]. Moreover, the evaluation of *Aedes sp*. mosquito's resistance to insecticides is not systematically addressed[71,72,79]. Community social mobilization and house screening have shown a non-precise reduction of incident dengue cases only under strict controlled conditions (RR=0.22; 95%CI=0.05, 0.93), with non-significant reduction of entomological indexes[71,78]. Despite few successful cases of reduced incidence of arboviruses due to the use of genetically-modified and sterile mosquitoes, its use has raised concerns about the environmental and ecological impact[72,79,80].

2.3 Differential burden of diseases among different populations

In addition to the described shortcomings on strategies aimed at prevention, management or better control of arboviruses, there are issues associated to their distribution across different populations that call the attention for further research. Although dengue, chikungunya and Zika

are important public health problems worldwide[1,3,4,7,40,74], there is an observed differential distribution of the morbidity and mortality, not only geographically but socially too[6,23,36,39-42,70]. Worldwide, the majority of cases are reported in low socioeconomic settings, areas where access to potable water is limited, areas of high population density, and regions where the environmental conditions favors the presence of *Aedes* mosquitoes[1,3,23,70]. In Colombia, the burden of dengue concentrates in 50 (out of 1,101) municipalities, and chikungunya and Zika have been consistently notified -since its introduction- on the same 30 municipalities[4,6,10,12,18,36,81,82]. In Brazil, although the distribution of dengue cases varies across the country, the Northwest and Central-Eastern region (Coastal areas) concentrates the majority of cases of Zika and chikungunya[11,21-23,51,83-90].

Despite the incidence of notified diseases being the same across different socioeconomic groups, mortality rates and some severe outcomes are higher among people at the bottom of the socioeconomic distribution[30,49,81,83,86,87,89,91-93]. For instance, 80% of dengue and 70% of chikungunya deaths in Colombia were people under the government subsidized health program (i.e., people at low socioeconomic position)[13,14,18,82]. In Brazil, there is an increased risk of dengue death (case fatality rates) ranging from 40-70% among people without or with less than four years of schooling[19,21,22,89]. Also, a reduction of dengue death of 10% has been observed among people with a median income over 350$USD per month (~USD$ 100 above the minimum wage)[21,22].

Paradoxically, despite that being black or of African ancestry is considered a protective factor for severe dengue[15-17,19,94-96], in Brazil, the risk for dengue death is between 34-75% higher among blacks or people from African ancestry[19,22,97]. Similarly, 70% of congenital Zika syndrome were observed in children born to young black or brown women living in low socioeconomic settings [11,83,98]. Likewise, according to the national surveillance system of Colombia (SIVIGILA) the proportion of dengue cases among Afro-Colombians ranges from 1.7% to 3.4% of all reported cases[18]. Yet, the proportion of reported severe cases among Afro-Colombians (2.7% in 2013 and 1.6% in 2017) is similar to the reported proportion of severe cases among Non-Afro-Colombians (2.5% in 2013 and 1.0% in 2017)[18,99]. And the fatality rates among Afro-Colombians (23.1% in 2016 and 14.3% in 2017) is only slightly higher than the observed severity among Non-Afro-Colombians

(13.8 in 2016 and 8.8% in 2017)[18,81]. Also, a study showed that 98% of Guillain Barre syndrome associated with Zika was among Afro-Colombians[50]. In summary, people in low socioeconomic status and non-whites are contributing differentially to the burden of severe outcomes of dengue, chikungunya and Zika[13,14,49].

2.4 Knowledge Gaps

Although it is of interest to know where and in which populations dengue, chikungunya, and Zika occur, there are still significant knowledge gaps that I am interested to fill with my PhD dissertation. First, the identification of the places where these diseases occur have been traditionally addressed using aggregated data and using fixed effects for the spatial areas, assuming homogeneity of distribution within the areas. Such identification of spatial areas includes dimensions as large as provinces and countries, with very few cases of disaggregation into small scales (e.g., neighborhoods, census blocks, or exact household level) and usually without a social characterization of the spatial area[100-106]. Although the spatial analysis of these arboviruses is guided and limited by the data availability, the traditional study of areas of high concentration using aggregated level fixed effects poses additional challenges. Namely, the potential risk of ecological fallacy, lack of proper adjustment for the effect of neighboring areas, and the lack of applicability to public health activities associated with vector control[101-107]. The latter, due to the fact that in Latin America vector control is usually decentralized and budget limited. Therefore, the identification of large areas (i.e., regions or district-level analysis) without the identification of hotspots neighborhoods, implies the administration of different measures without targeting the problematic areas[72,74]. This approach represents large costs for the vector control program and limits the effectivity of any applied measure[7,27,31,34,55,57,79,108,109], contributing to the endemicity (sustained presence of the arboviruses), and the presence of shorter interepidemic periods in concentrating areas[71,74,75,78].

Second, dengue, chikungunya, and Zika have not been addressed specifically through the lenses of health inequalities. Despite the known relationship between infectious diseases and social

disparities, the majority of the studies concentrate on the biological determinants of these conditions[23,24,26]. Social determinants are usually used on epidemiological analysis as additional covariates, that if available, will be used to adjust for possible confounding without specifically analyzing the presence of inequalities[49,86,87,89,92,100,104]. Therefore, the identification of the populations in which these diseases occur is limited to the distribution of cases according to biological factors such as age, sex and presence of comorbidities[20,49,110]; with very few studies interested on the understanding of ethnicity and socioeconomic conditions as driver of disease presence[15,21,49,50].

Although it is assumed that the paradoxical effect of ethnicity observed in Brazil and Colombia is due to confounding of socioeconomic status[16,19,21,22,111,112], no systematic analysis has been conducted to identify or refute the presence of such disparity. Moreover, to the best of my knowledge, there has not been a systematic analysis intended to understand what the main contributors to this disparity are; or whether the ethnic disparity is completely explained by the differences on the socioeconomic distribution across ethnic groups. This information is key to contribute to the design and implementation of disease control strategies that could target the most affected groups, as with interventions aimed at decreasing socioeconomic disparities and arboviral-disease control in both Colombia and Brazil.

2.5 Conclusions

Although some studies have investigated the relationship between dengue and poverty and arboviruses and low socioeconomic status, most of the literature is only descriptive or have not been analyzed within the specific lenses of health inequality[24,36,89,92,93]. Important barriers for the identification and understanding of health inequalities on arboviral diseases relate to data availability and methodological challenges, including limited adjustment for the endemic-epidemic character of these diseases, the type and interpretation of inequality measure used, and presence of structural biases such underreporting and misclassification[6,49,70,74,78,93,103,105].

Therefore, the above-mentioned research gaps about dengue, chikungunya, and Zika in Brazil and Colombia generated the following research questions: 1) Are dengue and severe dengue high-risk concentration areas similar when accounting for socioeconomic distribution? 2) What is the magnitude and trend of socioeconomic inequalities on the burden of dengue, chikungunya, and Zika? 3) Which social determinants of health are contributing the most to the observed socioeconomic inequality on arboviruses in Brazil and Colombia? And 4) Is there evidence of ethnic disparities among dengue and what is the role of socioeconomic status on the disparity?

I postulate that beyond the well-known and studied biological factors (e.g., virus serotypes/strains, immunological factors, etc.) the burden (morbidity and mortality) of dengue, Zika, and chikungunya in Latin America is affected by the large presence of social inequalities. Also, that ethnic disparities, if present, are not related to a biological difference determined by race but explained at least partially by socioeconomic inequalities. This dissertation aims at identifying, quantifying and assessing the presence and trend of inequalities associated to dengue, Zika and chikungunya, accounting for their spatiotemporal distribution and addressing some of the methodological challenges that have prevented the analysis of this phenomenon in the current body of the literature on arboviruses.

Chapter 3:**Overview of Methods**

To answer the above-mentioned research gaps and questions, this dissertation involves the development and application of a diverse range of quantitative research methods, integrating methodologies used for spatiotemporal analysis and social epidemiology, particularly from the study of health inequalities. Specifically, in order to contribute to the identification of places and populations in which these diseases concentrate, I conducted spatiotemporal and health inequality assessment of the burden of notified dengue, chikungunya and Zika in Latin America. Although the detailed methodology is presented in each manuscript chapter, here I provide an overview of the general framework, the surveillance-based analysis in the context of arboviruses in Latin America, some generalities about spatiotemporal analysis and the measures of health inequalities. In this section, I also included general information of the data used for this dissertation, including study sites and data sources.

3.1 General Framework

Considering the health of the population as a result of the interaction between individual, social and environmental factors, the analyses conducted in this dissertation included the assessment of individual and socio-environmental factors on the presence and distribution of arboviral diseases and the identification of health inequalities according to such factors. I integrated a series of analytical approaches, following the guide of the Commission on Social Determinants of Health (CSDH) conceptual framework[110,113]. The CSDH framework considers two main types of social determinants that impact in equity in health and well-being: 1) Structural determinants, including socioeconomic position, ethnicity, education, occupation and income; and 2) Intermediary determinants, including material circumstances and behavior and biological conditions (**Figure 3.1**). In this dissertation, the CSDH framework was instrumental in the identification of the structural determinants of health inequality included in the spatial analyses and contributed to the interpretation of the association of the measured determinants with the different outcomes.

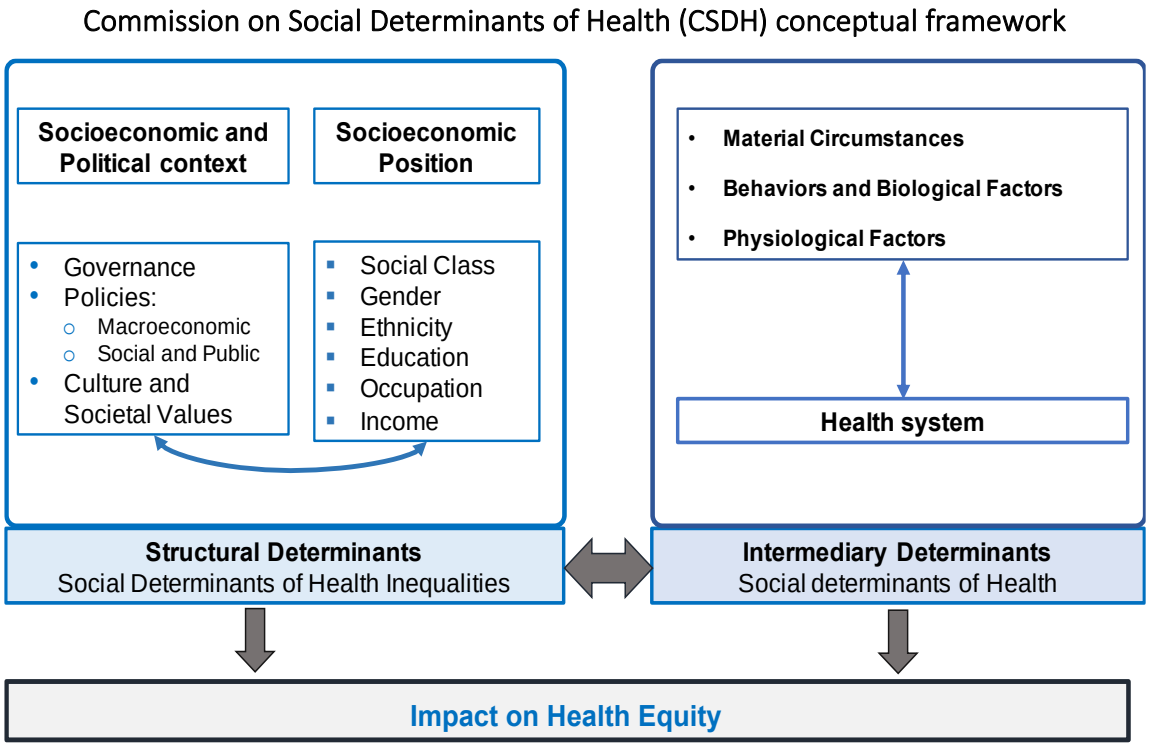


Figure 3-1 Commission on Social Determinants of Health conceptual framework.

Identification of structural and intermediary determinants of health that affect individuals and the community. Source: Modified from Commission of Social Determinants of Health-CSDH. Closing the gap in a generation. World Health Organization (2008).

3.2 Surveillance of Arboviruses

Given the epidemic character of arboviruses and the logistic limitations, including the cost of conducting cohort studies, a large proportion of the analyses on arboviruses are done with surveillance data[6,8,18,25,33,45,82,84,103,114-120]. Crucially, surveillance data is often the main source of data for decision making on arboviruses and overall for other endemic conditions or immuno-preventable diseases[2,6,7,116]. Overall, epidemiological and entomological surveillance programs in endemic countries routinely collect human case data and mosquito indicators to understand disease trends, detect outbreaks, and for the design and evaluation of interventions or programs [6,25,116,119]. A surveillance system is an essential component of disease control programs and surveillance data is also used to estimate the disease burden and to monitor long-term trends[6,43].

As for other diseases, the surveillance of arboviruses often relies on two main broader types: First, a passive surveillance, where the diagnosed (clinically or laboratory confirmed) cases are reported by health care facilities that identified the case[2,6,7]. However, the identification of these cases is done passively because the diagnosis and therefore the reporting, is done among individuals who seek health care attention[18,115,121]. Thus, indicating that the population is composed mostly of cases displaying symptoms and who were able to access (geographically, economically and physically) the health care system[25,116,119,122-124]. Second, a sentinel surveillance, implemented mainly in endemic areas, targeting populations, geographic areas and seasons at risk[7,25,43,116]. The sentinel surveillance could be passive or active depending on the resources from the health system and the presence of outbreaks[25,119]. Sentinel surveillance are often decentralized activities that allows the flexibility of diagnosis and or confirmation of cases locally[25,116,118,119]. For the passive-sentinel surveillance, individuals who seek care and present the specific profiled symptomatology under surveillance, are assessed specifically for the targeted disease. In the case of an active-sentinel surveillance, healthcare practitioners seek for individuals presenting the targeted symptomatology among individuals who seek care and among the general community[6,7,25,43,116,118,119].

Ascertainment of Arboviruses

Cases reported to the surveillance systems are laboratory and clinically confirmed cases[2,98,125-130]. Briefly, laboratory confirmation for dengue and chikungunya is based upon a positive result from antigen, antibody, or virus detection and or isolation[2,131-133]. Antigen and virus detection range in specificity (87% to 92%) and sensitivity (47% to 79%) according to the day of the sample collection, relative to the symptom's onset date[2,134-136]. In the case of Zika, sensitivity (37% to 65%) and specificity (66% to 100%) of the serological tests should be supplemented with molecular diagnostics to improve specificity and rule out other arboviruses[137-139]. The sensitivity (82% to 100%) and specificity (96% to 100%) of real-time reverse transcription polymerase chain reaction (PCR) for Zika, depends on the type of the sample (e.g., serum, urine, seminal fluids or other tissues) and the day of the sample collection[136-139].

Although all cases should be laboratory confirmed, given the potential limited resources (cost, materials and infrastructure), health systems prioritize the laboratory confirmation of severe dengue, dengue-related deaths and chikungunya related deaths[127,130]. For Zika, testing is mandatory for pregnant women and neonates, children < 1 year, adults >65 years, people with co-morbidities, people with neurological symptoms, deaths, stillbirths or children with congenital malformations in endemic areas or if they are suspected cases[126]. Confirmation of probable dengue cases is largely based on clinical diagnosis plus at least one serological test, usually an Enzyme Linked Immunosorbent Assays (ELISA) for immunoglobulin M test (IgM) with or without immunoglobulin G test (IgG) or and a confirmed epidemiological nexus[131-133]. The confirmation of chikungunya and Zika includes ruling out the presence of dengue, before conducting any test for chikungunya or Zika if not done simultaneously[2].

Reported cases to the surveillance systems include a non-negligible proportion of clinically confirmed arboviral cases[2,6,10,11,125-127]. This occurs mostly for dengue, which given the high burden of cases and the large magnitude of its outbreaks, is not expected to have the totality of cases confirmed by laboratory[6,38,125,130]. Although the laboratory confirmation for chikungunya is challenging, its distinctive clinical characteristics often help with the efficiency of the testing procedures and therefore confirmation[127,135]. The main laboratory-confirmation related challenges are for dengue and Zika because they are virologically closer, which could lead to cross-reactivity[63,133,137,139]. Usually the confirmation of Zika requires a more elaborated set of diagnostic tests, also increasing the cost and the time required for the diagnosis but improving its overall ascertainment [134,136,137]. The sensitivity of clinical diagnosis for arboviruses depends on the criteria used for the case identification, and compared to dengue, the specificity of the clinical diagnosis could be higher for Zika and chikungunya[2,115,132,134-136]. For overall clinical diagnosis of arboviruses, the estimates of sensitivity range from 11% to 89% and for specificity ranges from 80% to 95%[2,115,134]. The variability depending largely on the prevalence of arboviruses and the presence of outbreaks, with higher positive predictive values in endemic areas[2,46,115,120,123,134-136]. The proportion of laboratory confirmation of arboviruses ranges from 10% to 85%, depending on

whether there is an outbreak or not and with the remainder considered confirmed based on clinical criteria[12,126,140].

Notification, Underreporting and Misclassification of Arboviruses

Notification of dengue, chikungunya and Zika is mandatory in Brazil and Colombia through their respective national passive surveillance systems[11,98,125,128]. In Brazil, these arboviruses are individually registered weekly in the national surveillance system platform (SINAN). In addition, Zika and congenital Zika syndrome should be registered in the Public Health surveillance notification system (RESP), and every mortality outcome should be reported in the National Vital Statistics System (SINASC), as well[98,125,129,141]. In Colombia, the epidemiological surveillance system is managed by the National Health Institute and all events of public health interest, including arboviruses, are notified through the Public Health Surveillance System (SIVIGILA) and its homonymous software[126,128,130,142].

According to regional and national estimates from the surveillances systems in the Americas region, Colombia and Brazil account for up to 40% of the overall dengue burden[6,7,36,38] and had the highest burden of chikungunya and Zika in 2014 and 2016, respectively[10,45]. However, despite the large number of cases reported, there is an important number of cases that were not captured by the respective surveillance systems. Although the objective of the epidemiological surveillance systems is to provide information about the public health issues in a systematic and opportune form to inform decision making to prevent and control diseases[25,103,132,133,142], there are several reason for which the totality of arboviral cases are not captured by the surveillance systems (**Figure 3.2**)[25,122,124].

First, because of the clinical nature of arboviruses, a significant portion of infected individuals remain asymptomatic or display very mild symptoms, which contributes to the under-ascertainment of cases[1,2,4,134,143]. The proportion of asymptomatic or inapparent cases is variable and depends on several factors. In the case of dengue, symptomatic cases could be as low as 30-50% in case of DENV4 and as high as 70-85% for DENV1 or DENV3[68,144-153]. In the cases of

chikungunya, the proportion of asymptomatic cases tends to be lower and documented to be around 30% of all infected cases[32,37,88,154-156]. In the case of Zika, although the asymptomatic proportion could be larger 65-80% (compared to dengue and chikungunya), the active screening among women during reproductive age and pregnant women potentially decreases the risk of under-ascertainment[115,134,136,137]. The degree of under-ascertainment has been assessed by seroprevalence studies and these results have been used to inform estimates of the burden of diseases[114,143-145,152,157,158]. However, it is important to note that seroprevalence studies are often costly, not always feasible and therefore their application in larger populations is very limited[145,150,157]. In addition, seroprevalence studies are also affected by the sensitivity and specificity of the test and the technical challenges of cross-reactivity[8,114,144,145,157].

Second, from the cases displaying symptoms, only those seeking and receiving care are potentially captured by a passive surveillance system[2,6,36,63]. Thus, challenges in the identification of the arboviral diseases, including knowledge of the disease and availability of diagnostic tools, could affect the reporting. Finally, despite the mandatory reporting of arboviruses, reporting of cases could be differential between private or public health care providers[122,124,159,160]. Altogether leading to an expected underreporting ranging from 5% to 50%, depending on the arbovirus, the endemicity and type of health system[118,122,124,161]. Although surveillance underreporting is not uncommon, even more so among epidemic diseases including influenza and Sars-Cov-2[25,122,124,161]; underreporting represents an important challenge for the use of surveillance data because it could bias estimates and reduce their utility[25,119,122,124,161].

In addition to the potential underreporting, the notification of arboviruses in the Americas region was potentially impacted by misclassification. Given the similar clinical presentation of dengue, chikungunya and Zika, the introduction of chikungunya late in 2013 and Zika in 2014/2015 in the region, might have led to diagnosing chikungunya or Zika cases as dengue at early stages of their introduction[2,4,115,134,136]. Although laboratory diagnostic tests were made available soon after the introduction of chikungunya and Zika and the sensitivity of the clinical diagnosis of these diseases

improved with the awareness of the diseases[2,115,133-136]; it is still possible that a proportion of only clinically confirmed cases were misclassified.

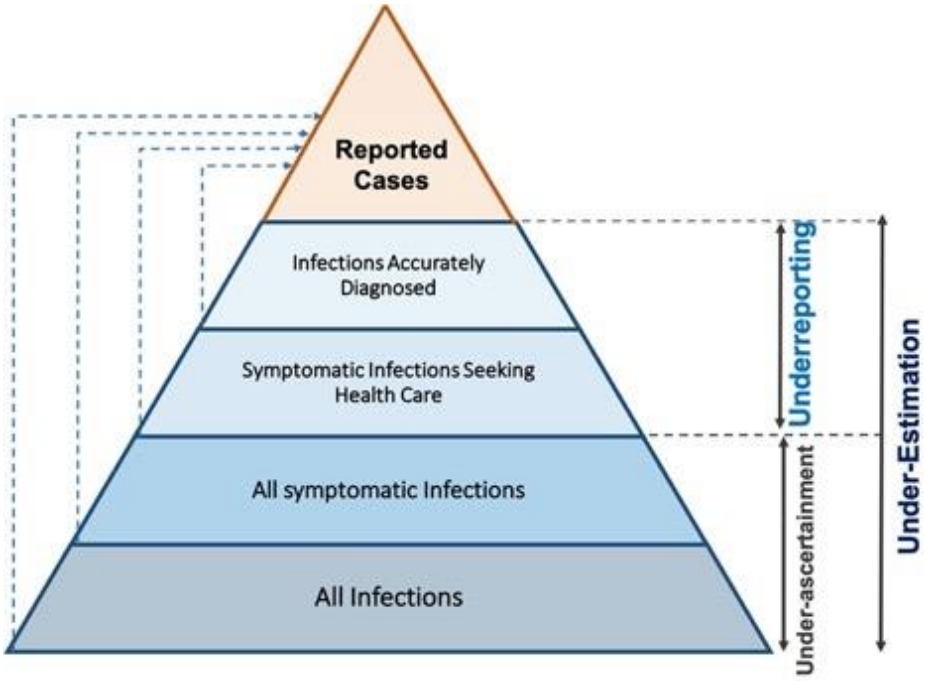


Figure 3-2 Schematic of ascertainment and notification of arboviruses.
Source: Modified from Gibbons CL, et al. (2014) Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods.

3.3 Identification of Places (spatial areas) of Arboviral concentration

Given the spatial heterogeneity of arboviral distribution worldwide and their rapid spread, particularly in Latin America, there is a special interest on identifying spatial areas of disease concentration[6,8,45,103]. Endemicity, understood as the constant presence of a given condition in a spatial area, has been widely studied for arboviruses[6,36,41,162]. Different approaches to identify places of disease concentration are broadly classified as spatial and non-spatial approaches. The latter uses mainly a fixed effect for a given geographic area (e.g., continent, region, country, province, city, district or neighborhoods, etc.) with or without adjusting for clustering effects[101,103]. Non-spatial approaches assessing differences between geographical areas are often appropriate when studying a small set of geographical units, when comparing specific characteristics such as

rurality vs urban areas, and when it is expected that other characteristics included in the analysis (if any) have the same distribution across spatial areas[101,103,107]. Given these considerations, non-spatial approaches include the comparison between (often arbitrarily) selected baseline-geographic units and other geographic units included in the analysis. This comparison could be done, as well, with or without accounting for additional characteristics of the area and or the outcome of interest. Although non-spatial approaches produce useful information, they do not necessarily account for spatial autocorrelation and are not suitable for the identification of within-area variability[101,107].

Spatial analyses are intended to provide insights about spatial distribution of the outcomes accounting for the spatial correlation of the geographic units[101,103,107,163]. Spatial approaches are also suitable for analysis of between and within geographic and spatial area variability. In addition, spatial analyses allow the comparison of large numbers of geographic units, and the comparison of the outcome's distribution across different spatial structures using different types of referent units[101,103,107]. Thus, comparing non-spatial vs spatial approaches, the spatial analyses provide robustness and flexibility to the analysis aimed at identifying areas of arboviral concentration. Although previous studies have used aggregated and area-level data, I consider critical the advantage of using individual-level location and integrating a set of individual-level and area-level covariates to improve the assessment of arboviral distribution, while accounting for spatial autocorrelation and accounting for the between and within area variability[101,107].

Hierarchical Spatiotemporal analysis

Given the need to account for the heterogeneous spatial distribution, the contribution of individual and socio-environmental spatial characteristics, and the epidemic nature of arboviruses, simultaneously; all the analysis included in this dissertation are conducted using hierarchical spatial-based approaches. Particularly, I analyzed individual and aggregated surveillance data of all dengue, chikungunya and Zika notified cases between 2007-2017 in two cities in Brazil and Colombia. Within the cities, I assessed the distribution of cases across districts, the neighborhood and the exact location of residence of the cases.

Broadly, the use of hierarchical or multilevel approaches allow the analysis of individual and area-specific effects, while simultaneously investigating the between and within-area variability[101,164,165]. This is achieved by the inclusion of random effects for the spatial/geographic areas under study, in addition of the adjustment for the individual and area-level covariates or fixed effects. Given that it is unrealistic to assume that the presence of arboviral cases are independent within the same spatial structure, I made use of random intercepts for the spatial area, which accounted for the dependence within spatial units. This approach allows each spatial area to have a specific baseline disease distribution, that is conditionally independent, given the random intercept and the covariates[164-168]. The advantage of this approach is the estimation of area-specific coefficients that could also be interpreted as marginal or population averaged effects, while accounting for clustering effect and improving precision using a more parsimonious model than a non-spatial fixed effect model[164,165,167,168].

Although standard hierarchical models could be fitted using a frequentist approach, given the nature of the surveillance data, the need to account for the uncertainty of the estimates and the flexibility required to account for the spatial correlation, I conducted all analysis under the Bayesian paradigm[163,169,170]. The Bayesian approach offers a number of advantages including flexibility, adaptability to data availability constraints, and the opportunity to account for available information to inform the estimation of the parameters of interest[107,164,167-169]. The inference under the Bayesian analyses allow the estimation of parameters under a single framework where uncertainty is naturally accounted for. The advantages presented by Bayesian models made of this approach a convincing method to study disease distribution and forecasting for several surveillance programs[101,107,163,164,166-172].

Bayesian Hierarchical Models Accounting for Neighboring Structures and Time Trends

In order to assess the diseases distribution accounting for the correlation across spatial areas and over time, I made use of structured-random effects for space and time, respectively. The use of

structured random effects under the Bayesian framework has the potential to borrow information across space and time, improving the precision of estimates[164,167].

I implemented the Bayesian analysis using Integrated Nested Laplace Approximations (INLA)[163,173,174]. The INLA program is a Bayesian-based method, using probabilistic assumptions iteratively allowing variability of the data, as it is required in this type of analysis. INLA approximates Bayesian inference in a subclass of structured additive regression models which are called latent Gaussian models[163,173,174]. As in any Bayesian approach, the procedures of INLA are iterative accounting for the main three parameters of Bayesian framework (Posterior distributions, Priors and the likelihood function). However, instead of the known Monte-Carlo-Markov-Chain (MCMC) simulation and sampling procedures, INLA uses approximation methods including components of the precision of the covariates of interest[163,173,174]. Specifically, the approximations to the posteriors are based on Laplace approximations of the precision parameters of the covariates (hyperparameters)[174]. The approximation of the conditional distributions is done through Gaussian distributions, accounting for the precision, which is also corrected by the use of a mixing term (spline) to adjust the fit[173,174]. The method explores the joint posterior of hyperparameters, proceeding with the identification of relevant points and its weights, the evaluation of conditional posteriors, and the estimation of marginal posteriors via numerical integration[173,174]. Given that the inference is made using approximations, it is possible to have somehow imprecise parameters than those obtained using MCMC methods and which could be affected with increased number of hyperparameters[163,173,174]. However, this approach, compared to standard methods is computationally efficient, solving issues of convergence and low autocorrelation observed with MCMC. The overall approach provides robust estimates with credible limits, without selecting arbitrary referent categories and accounting for clustering at the same time[163,164].

The correlation of either the spatial or temporal structure in the latent Gaussian models is specified as functions using a broad range of variance-covariance correlation structures[173]. The choice mainly depends on the known distribution of the effect and the type of correlation deemed

adequate in each case, which could be data, model, or substantive knowledge driven[163,173]. One option is to use independent identically distributed models (i.i.d), which provide a random effect for each unit, without accounting for any spatial or temporal structured or *a priori* specified correlation[101,173].

The choice for the spatial latent effects used to account for spatial correlation, depends also on whether the spatial structure is a regular lattice (e.g., a constructed regular grid with columns and rows) or an irregular lattice (e.g., polygons or maps with non-regular internal divisions)[101,173]. However, the random spatial effects are generally the results of multivariate Gaussian distributions with mean zero and a precision matrix defined by the presence of adjacency between spatial areas (i.e., identification of neighboring structures or vicinity), which in turn is determined by a specified precision parameter (i.e., the hyperparameter)[101]. For regular lattices, recommended models include those that parameterize the spatial structure as a Gaussian random field with a Matérn correlation function that changes by a given fixed parameter; or models assuming independent increments across the spatial units such as the 'random walk of order 1 (RW1)[101,173].

To account for the spatial correlation using an irregular lattice, it is necessary to use models accounting for the irregular structure of the spatial area and variability in the levels of correlation across neighboring areas[101,173]. Available models include those accounting for the neighboring structures using a conditional parameter for the precision, which depends on the presence or not of adjacency between two spatial units[101,107]. The most known model is the one described by Besag in 1974, which is described as an intrinsic conditional autoregressive structure (iCAR)[101,175,176]. This model accounts for vicinity but does not account for the presence of pure overdispersion within the spatial area outside the parameters for the structured component. Therefore, another option is the Besag-York-Mollie (BYM) model, which decompose the spatial effect into a structured component determined by the Besag structure and an unstructured component that accounts for pure overdispersion[101,166,177]. To account for the correlation over time, including the identification of seasonal patterns, some available models include the use of structured-random effects with

autoregressive specifications of first or second order (e.g., AR1, AR2) and models with independent increments or 'random walk' structures of first or second order as well (RW1, RW2)[163,178].

Spatial Bayesian Hierarchical Models Using Individual Data or Point Process Models

Another approach to identify the spatial distribution of disease outcome is the use of individual data location, modeled as single spatial points[101,102,107,179]. This type of analyses is broadly called point process models, because they assess the distribution of the individual location of an outcome (denominated points) in a given spatial structure[101].

A point process model (PPM), estimates the spatial distribution of individual cases as a function of a latent Gaussian random field[178-180]. This estimation describes the intensity of the point pattern (distribution of each point/cases), assuming conditional independence of the points presented in spatial area[102,180,181]. Specifically, conditional of the random latent field (spatial structure), the distribution of the points or point pattern of the outcome follows a Poisson distribution [178-180]. In this case, the random latent field is continuous in space and is used to express the spatial variation and autocorrelation of the outcome during a specific point in time. Therefore, the PPMs are used to identify the presence and degree of clustering within a given spatial structure, pointing out regions of higher and lower intensity of disease[102,181]. This, with the added advantage of accounting for both, observed and unobserved variation, in the assessment of the outcome's distribution at individual level.

In addition to the possibility to assess disease distributions using individual data, it is also possible to assess the spatial distribution of a given characteristic of the points. A modification of the PPM is a Marked Point Process model (MPP)[102,107,178-180]. The MPP approach uses an individual characteristic of the point (i.e., a mark) to inform the estimation of the distribution of the point given the specified characteristic, simultaneously[180]. This approach results appealing given that in addition to modeling the spatial distribution of any given arbovirus, it would be also possible to model the presence of severity simultaneously. In this case, the mark will be the presence or

absence of severity, which works as well as a response variable. In this case, the procedure is understood as a "labeling" of the Poisson process[180]. This approach allows the joint modelling of the severity 'marks' and the overall spatial pattern, accounting at the same time for their dependence[180,181]. Nonetheless, despite the compelling advantage proposed by this method, it is rarely used because the limited availability of data on exact location of cases and because it tends to be computationally intensive with large datasets.

For this dissertation, I considered a wide range of spatial and temporal structures to assess the distribution of arboviruses in Brazil and Colombia. However, in order to answer each research question optimizing the use of the data available, I conduct a spatial analysis for manuscript one and spatiotemporal analyses for manuscripts 2 to 4. Specifically, in manuscript 1, given the availability of individual location data I implemented a joint spatial marked point process model to assess overall and severity distribution of dengue in Colombia. For manuscripts 2 to 4, it was not possible to geocode each case at the exact location (latitude and longitude of the address of residence) for the entire study period. Therefore, the analyses were done at the neighborhood level as follows: For manuscripts 2 and 3, I assessed the presence of socioeconomic inequalities using aggregated data and the neighborhood of residence as the main spatial unit. For these analyses, the spatial structured random effects were specified as BYM models and for the time (month of notification), I used random walks of first and second order (RW1 or RW2) structures. Finally, in Manuscript 4 I assessed the distribution of arboviruses across ethnic groups using aggregated and individual level data using BYM and RW1 structures for space and time, respectively.

Accounting for Underreporting and Misclassification of the Outcome

In order to ensure the robustness of the results, a series of sensitivity analysis were conducted to address some of the possible limitations associated to the use of surveillance data. Although the details are presented in each manuscript and their respective supplementary material, here I provide an overview of the measures taken to account for the possible underreporting and or misclassification of the arboviral diagnosis.

As described above, the use of surveillance data is potentially affected by underreporting. In the cases of arboviruses, several sources of underreporting could affect our estimates and interpretation. Methods used to correct for underreporting in the estimation of the overall burden of disease presence include the use of expansion factors, which are the inverse of the rate of a known underreporting rate[124,143]. This method is often used to account for under-ascertainment and underestimation due to mild or asymptomatic infections not captured by a passive surveillance system, compared to an active sentinel surveillance or a cohort study [122,124,143]. The expansion factors methods are used to update estimates of the burden of disease presence, multiplying surveillance estimates by the obtained expansion factors.

In this dissertation, I focus on the burden of notified cases i.e., only cases captured by the surveillance system. Therefore, the source of underreporting includes the potential differential reporting associated to the ascertainment and reporting practices across different health systems and or populations. Given the flexibility provided by the Bayesian framework, accounting for underreporting could be done in a single step, by updating the priors in the disease models. Specifically, I used rates of underreporting in the same and similar settings[143] to update the priors for the overall distribution models and for differential reporting by socioeconomic status, insurance scheme and ethnicity. However, the use of informative priors does not completely account for the extent of the potential selection bias or measurement error. In addition, the correction for measurement error among non-normal distributions, such as the count of cases (Poisson distribution), presents important methodological challenges and methods to account for this are still under development[173,174,182]. Therefore, as an alternative to overcome the presence of underreporting as (miscounting of cases)[173,182,183], I simulated new datasets accounting for the underreporting by socioeconomic status, insurance scheme and ethnicity and fit the same set of models used for the main analysis.

To account for the potential misclassification of chikungunya and Zika with dengue at early stages of their introduction, I used two types of approaches. First, in addition to assess disease-specific

outcomes, I also conducted the analyses of all arboviruses. Specifically, I grouped dengue, chikungunya and Zika cases and created an additional outcome variable called "all arboviruses" and performed the same set of analysis that for each specific arboviral condition. Although this misclassification is expected to be non-differential and will likely pull the estimates towards the null, it is possible to consider the "all arboviruses" group as an outcome that is less sensitive to potential misclassification.

Second, since by the time of conducting the analyses for this dissertation there was no available literature on the scope or magnitude of misclassification in all the study sites[2], I used the results from a parallel study to inform a simulation-based sensitivity analysis. Specifically, to correct for the potential underreporting and misclassification, I used obtained rates from an observational capture-recapture study[122,161] in three Colombian cities from 2014-2017. The study used registries of cases from healthcare facilities and surveillance offices in each city, to identify the scope and degree of reporting of arboviruses. In order to identify predictors of reporting and estimate the probability of reporting by disease and year, the study compared cases that were diagnosed at the healthcare facilities (capture) and those that were reported to SIVIGILA (recapture). The potential misclassification of clinical diagnosis was assessed for different parameters of sensitivity and specificity of diagnosis, using simulation extrapolation for misclassification method (MC-SIMEX)[183-185]. I used the obtained results to simulate new datasets, using parameters from the distribution in the original data set and correcting specifically the outcomes (count of cases per neighborhood per month and presence of severity) accounting for the overall, ethnic, and insurance specific underreporting and misclassification.

3.4 Assessment of Arboviral distribution among Different Populations

Given the heterogeneous distribution of arboviruses across the range of socioeconomic strata and different ethnic groups, it is important to characterize the populations in which these diseases concentrate [27,81,91]. However, despite the known relationship between social determinants and the presence and distribution of arboviruses, these diseases have not been addressed specifically

through the lenses of health inequalities[24,186]. Some common challenges related to the analysis of social determinants and health inequalities on arboviral studies include: i) measurement and control of confounders; ii) misclassification of the exposure and outcome (due to limitations on disease ascertainment and the assessment of socioeconomic measures); and iii) methodological limitations on the application of available inequality methods[23,24]. In this dissertation I tried to overcome these limitations by integrating the assessment of health inequalities, the use of Bayesian spatiotemporal models, and using surveillance data, as described above.

Inequality assessment

Although the differential distribution of any outcome to assess inequalities could be done in several ways, including contrasting numbers, proportions or rates between groups, either in relative or absolute terms, there are methods developed to specifically measure inequalities[24,186-188]. Broadly, there are two main groups of summary measures of inequality, absolute measures and relative measures. Absolute measures, which are presented using the units of the health indicator, reflect the magnitude of the inequality in absolute terms[187-191]. Relative measures, which are unitless, reflect the proportional inequality. Given that both measures provide complementary information, it is recommended to use both measures when reporting the presence of inequalities[188,190,191].

Furthermore, within these two groups several other types of measures could be estimated. According to the number and type of groups used in the comparison, inequality measures could be simple or complex. While simple measures are used to assess the presence of inequalities across two subgroups (e.g., rich vs poor), complex measures are used to assess the inequality across the range of more than two subgroups (e.g., low income vs middle income vs high income)[187,188,190]. Also, complex measures could be further classified as ordered or not ordered measures, depending on the presence of a natural ordering of the subgroups analyzed. Common absolute complex measures of inequality include the population attributable risk, which is considered a weighted measure because account the population size of each group, and the slope index of inequality which is a weighted regression-based estimate [186-188,190]. Complex measures of

relative inequalities include the population attributable fraction and the relative index of inequality, both considered weighted measures[186-188,190].

Concentration Index of Inequality

The concentration index of inequality is a known complex, weighted, and rank-dependent (ordered) measure of inequality[189,192-196]. The concentration index could be measured in both, absolute and relative scales, and indicates the extent which a health outcome is concentrated among people at the bottom or the top of the socioeconomic distribution. The absolute concentration index (ACI) measures the concentration of health outcomes in the absolute scale and the relative concentration index (RCI) does it on the relative scale, and both have been widely used to explore the distribution of inequalities[189,192,197]. To calculate these measures, a weighted sample of the population is ranked from the most disadvantaged (rank zero) to the most advantaged (rank one), using a socioeconomic measure (e.g., income, education, etc)[188,193-196]. The concentration index then estimates the outcome's shares according to the socioeconomic measure's rank. The magnitude of the inequality is displayed by the concentration curve, which plots the outcome's shares and the socioeconomic rank. The concentration index is defined as twice the are under the curve that lies between the line of equality (a 45° diagonal) and the concentration curve[189,192](Figure 3.3).

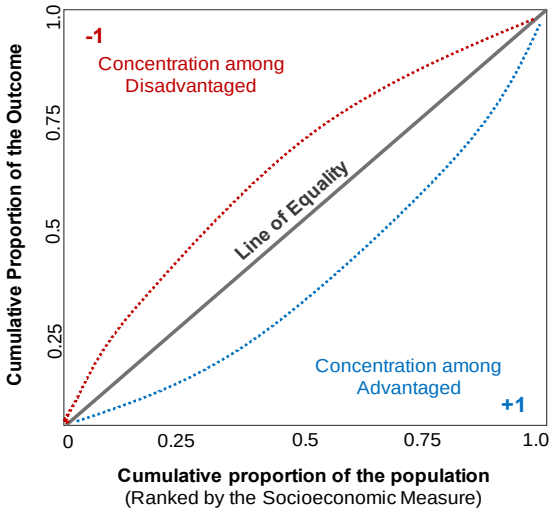


Figure 3-3 Schematic of the concentration Index of Inequality.
Display of the concentration curve, line of equality and ranges of concentration index in presence of undesirable outcomes.

The concentration index ranges from -1 to 1, where zero represents complete equality (i.e., the health outcome is distributed proportionally across the socioeconomic measures)[189,192-195]. For undesirable health outcomes, such as disease rates, negative values indicate concentration of the outcome among disadvantaged groups and positive values indicate concentration among the advantaged groups[187,188]. The RCI could be define simply as two times the covariance between a health outcome ($\boldsymbol{H}$) and the socioeconomic rank ($\boldsymbol{Y}$) over the mean health's outcome ($\boldsymbol{\mu}$), specified as: $\boldsymbol{RCI = 2cov(H, Y)/\mu}$. Then, the ACI is the multiplication of the RCI by the overall mean outcome[188,189,193,195].

However, given the need and possibility to account for confounders, it is also possible to obtain RCIs estimates via regression[187,193]. This regression-based analysis allows also the possibility to obtain the decompositions of the socioeconomic inequality indices and is given by:

$$RCI = 2\sum_{j=1}^{k} \beta_j \, Cov(x_j, d) + 2Cov(\varepsilon, d)$$

Where β_j represent the beta coefficients for the covariates x_j included in the regression, d is the socioeconomic rank measure and ε is the regression-associated error term[193,195]. This regression-based decomposition of the concentration index could be broadly understood as the identification of two main components: a deterministic component and a residual component[193,195]. The deterministic or explained part, is the first right-hand term and includes the sum of contribution of the covariates included in the regression to explain the health outcomes. The residual component or unexplained part is expected to be close to zero if the error term is uncorrelated with the socioeconomic rank, and or if all covariates included in the model could account for the inequality[193,195].

I estimated the ACI and RCIs in this dissertation because these measures allow the estimation of inequality using the entire range of the socioeconomic distribution. The ACI and RCI allowed the identification and quantifying the magnitude of socioeconomic inequalities on arboviruses. I also used a regression-based decomposition of the RCI to further identify the contribution of some

measured social determinants of health and individual and socio-environmental covariates on the overall inequality. However, these models have not been applied or modified to account for endemicity and seasonality. Therefore, I adapted or created software-functions to estimate the inequality indexes obtained from the Bayesian hierarchical spatiotemporal models as detailed in the methods section of each manuscript and their respective supplementary material.

Arboviral distribution across ethnic groups

In order to understand the ethnicity paradox for dengue severity, I assessed the distribution of arboviruses across ethnic groups. However, the definition of ethnicity result challenging in settings like Latin America, where the population presents an important degree of mixing. For this purpose, I used self-identified ethnic groups according to the notification categories in each country. This is preferred over other measures because it tend to capture the historical background, phenotype and cultural aspects inherent to the definition of ethnicity. According to the notification forms and the national census data in each county, the definitions are as follows:

Brazil: There are six ethnic groups classified as 1) Whites or Caucasian; 2) Black, including any person considered from African ancestry; 3) Yellow (Amarela), which includes people from Asian ancestry or mixed ups between whites and Asian; Pardo, which includes the majority of the population in several regions of the country and indicates any mixing between any of whites, black and or Yellow; 5) Indigenous, or aboriginal, which indicate native Americans. It is also possible that the reported ethnicity is 'ignored' and the proportion of missing data for ethnicity is not negligible.

Colombia: In Colombia the notification system includes nine possible ethnic groups, including three categories for black: 1) Black or Afro Colombians which are any inland person of African ancestry without a known African lineage, 2) Afro Colombian "Palenquero", an inland African descendant who live in or is a member of any of the protected reserves created by former slaves who escaped from slavery, 3) Afro Colombian "Raizal", people from African ancestry (with or without mixing) who live or are members of any of the communities on the Caribbean islands of Colombia; 4) Gipsy, any person from the Roma community; 5) Indigenous or aboriginal, which

indicate native Americans; 6) White or Mestizo, which in Colombia indicates whites and Caucasians or their mix only with native Americans, and accounts for the majority of the population, and 7) others, which are people from a non-mentioned category and should be otherwise specified (e.g., Asian). The identification of ethnicity is a mandatory field in Colombia, therefore missingness of this data is unlikely. Although the initial intention was to assess the ethnic disparities in both countries, due to a large proportion of missing data for ethnicity in Brazil, mostly for dengue, the focus of the ethnic disparities was conducted only in one setting in Colombia. Also, given the heterogeneity and the important mixed-up of the population, I used a binary variable indicating the two main groups as Afro-Colombians and Non-Afro-Colombians.

3.5 Study sites and Study population

Given the important burden of these three diseases in Brazil and Colombia, this dissertation uses data from both countries. In Colombia, the burden of these three diseases is concentrated in 50 out of over 1,000 cities, with Medellin and Cali included on the top five reporting cities since 1998. In 2017, Cali reported 28%, 33% and 43% of dengue, chikungunya and Zika cases of the country, respectively[12]. Although in Brazil the distribution of dengue cases varies across the country, the Northeastern region experiences 29% of Zika cases and Fortaleza, is consistently on the top 50 out of more than 5,000 cities reporting dengue and chikungunya during the last 10 years [23,98,124,198]. Municipalities were selected based on the disease burden, knowledge of the context, presence of functioning surveillance system, data availability, and expressed interest of the local health agencies on the study (**Figure 3.4**).

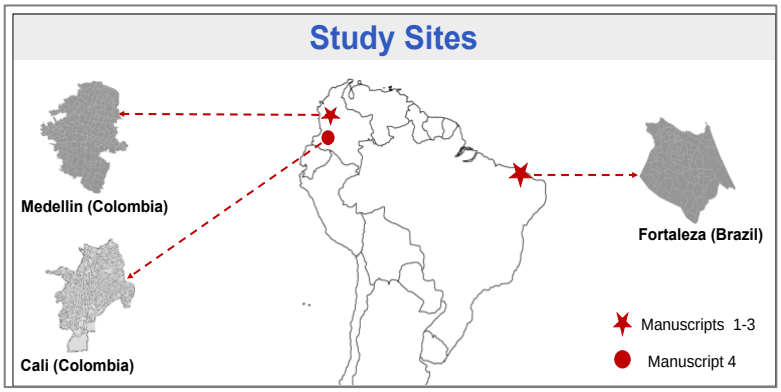


Figure 3-4 Geographical localization of the study sites.

Medellin is the second largest city in Colombia with more than 2.6 million inhabitants[199] and its annual dengue incidence ranged 161-745 cases per 100,000 inhabitants during the last 10 years [81]. The city has 21 districts (16 urban, 5 rural); 249 urban neighborhoods and 20 institutional units, distributed in 382 km2 (110 urban). Medellin's altitude ranges from 1,460 to 3,200 meters above the sea level (m.a.s.l), the average temperature 24°C, and has two rainy seasons (April – October). Although 50% of the city belongs to low socioeconomic status (SES), 98% of the city have access to potable water. The distribution of health coverage of the population is 70% contributory (employees or self-employees), 25% government subsidized, and 4% uninsured[199].

Cali is the third largest city in Colombia and dengue incidence ranged from 243 to 692 cases per 100,000 inhabitants/year over the last five years[81]. Cali has a population of 2.4 million inhabitants, from which 73.3% are considered whites or 'mestizos', 26.2% afro-descendants, and 0.5% as native Americans (68). The city has 335 neighborhoods within 22 urban districts, distributed in 566.3 km2 (120 urban). Cali's altitude ranges from 950 to 1,070 m.a.s.l, the average temperature is 25.5°C, and has two rainy seasons (April – October). Around 60% of the city belongs to low SES; 85.7% of the city have access to potable water, and 96.9% counts with a service of waste management. The distribution of health coverage of the population is 63.8% contributory, 27.1% government subsidized, and 9.1% uninsured[199].

Fortaleza is the capital of the fifth biggest state of Brazil, has a population of 2.5 million inhabitants, from which 57.2% are considered 'brown', 36.8% whites, and 4.5% blacks. The city has six districts and 119 urban neighborhoods distributed in 315 km2. Fortaleza's altitude is 21 m.a.s.l, the average temperature is 26.6°C and has one rainy season (January to May). In Brazil, social stratification is denominated according to the median per capita monthly income and is stratified into three categories as follows: Low (<R$290), Medium (R$ 291- R$ 1019), and High (> R$1,020)[198]. Brazil has a Universal Health Coverage denominated Unified Health System, which is expected to cover the health needs of the entire population[129].

Data sources: As indicated above, notification of dengue, chikungunya and Zika is mandatory in all study sites through a passive surveillance system i.e., based on symptomatic cases and/or people who sought care. The morbidity burden is defined as the incidence of notified confirmed cases and the severity/mortality burden by the severity and mortality rates among the reported cases [2,14,98,130]. Therefore, the study population was composed by all notified cases in the study sites between 2007 and 2017. Individual clinical and sociodemographic data from the National surveillance system (SINAN-Brazil and SIVIGILA-Colombia). Aggregated information about socioeconomic factors at the neighborhood was obtained using National Census data and local quality of life and basic needs surveys for socioeconomic data [198,200] (**Table 3.1**).

Table 3.1 List of covariates grouped according to the CSDH framework

DIMENSION	Variables
Demographic Factors	Urban/Rural Zone; Place of Residence & Work; Age; Sex; Ethnicity Informal Settlement; Migration; Population growth; Average time (or Distance) to closest available health facility
Environmental Factors	Temperature, humidity, rainfall; water retention sources (lakes, rivers, etc.); Entomological Indices: Breteau Index (proportion of household with larvae of *Aedes sp*.), Household Index (Proportion of household with *Aedes's sp*., larvae), Containers Index (proportion of containers with *Aedes's sp*., larvae)
Epidemiological Factors	Serotypes; Epidemiological period; Travel to endemic area; Individual clinical factors (presence of comorbidities)
Living Conditions and Lifestyle	Provision of basic services (water, waste disposal); Peridomiciliary sanitation; Water deposits (Water tanks outside the house and at ground level) Methods to control mosquitoes
Political Factors	Political commitment for the development of prevention and dengue control policies and programs (fogging, curtains, community-based programs)
Socioeconomic	Family income, median neighborhood income, education (rate of literacy and years of schooling), occupation (if applicable), insurance scheme (private or public), Human Development Index (HDI)

In summary, this dissertation includes the analysis of the burden of notified arboviral diseases including: 1) the identification of areas of high-disease risk accounting for the socioeconomic distribution of cases; 2) the identification of presence, magnitude and trend of socioeconomic inequalities; 3) the decomposition of the socioeconomic inequality; and 4) the estimation of between and within ethnic disparities. The integration of the aspects analyzed are displayed in the (over) simplified version of the directed acyclic graph (DAG) (**Figure 3.5**) were socioeconomic status (SES) was measured by median monthly household income in Brazil and by the Government-provided SES index by household in Colombia. Other information is summarized in a table including specific aspects of the method used in each manuscript (**Table 3.2)**.

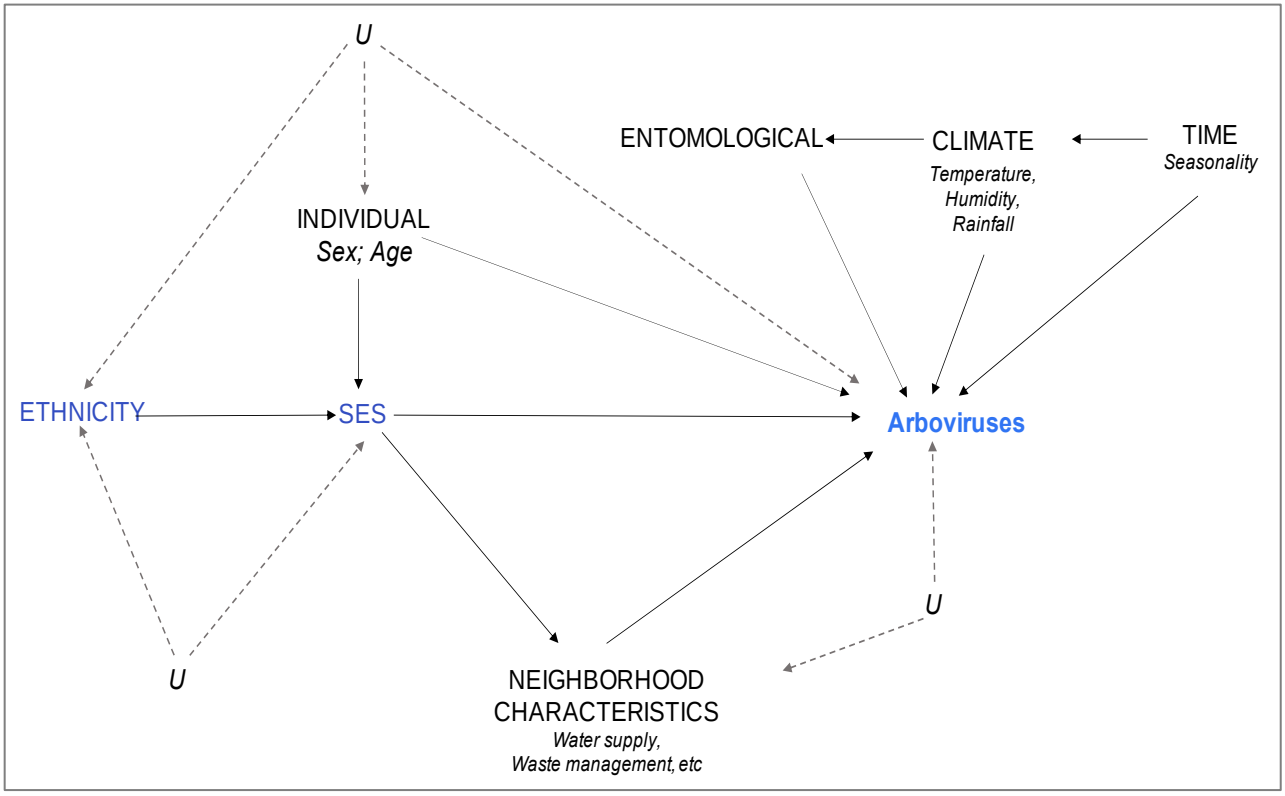


Figure 3-5 Simplified DAG for the analysis of arboviruses in Latin America

This simplified version of the DAG for the assessment of socioeconomic and ethnic inequalities, accounting for space and time, includes individual and socio-environmental characteristics and possible unmeasured confounders of the association under study.

Table 3.2 Summary of methods used in this dissertation.

	Manuscript 1	Manuscript 2	Manuscript 3	Manuscript 4
Main Objective	High-risk areas for overall and severe dengue	Presence and trend of health inequalities for arboviruses	Decomposition of socioeconomic inequality for arboviruses	Ethnic disparities for dengue and severe dengue
Outcome	Dengue rates and severity risk	Arboviral rates, RCIs, and ACIs	RCIs, ACIs and contribution of studied factors	Dengue rates and severity risk
Study Design	Cross-sectional	Longitudinal		Longitudinal
Study site (s)	Medellin, Colombia	Medellín, Colombia and Fortaleza, Brazil		Cali, Colombia
Data used	Individual level data	Aggregated data (neighborhood level)		Individual level and aggregated data
Spatiotemporal model	Joint marked Point process model (Exact location of cases)	Hierarchical with structured random effects (Spatial unit = neighborhood of residency; Time unit = month of occurrence)		
Sample size	1,793 dengue cases	281,426 arboviral cases in Fortaleza 40,887 arboviral cases in Medellin		65,402 dengue cases
Sensitivity Analysis	Informative priors	Informative priors, use of aggregated arboviral analysis; inequality analysis without spatial structure and alternative measures of socioeconomic status		Informative priors, use environmental covariates Simulation-based analysis, RCI and ACI estimation

Chapter 4:**A joint spatial marked point process model for dengue and severe dengue in Medellin, Colombia (Manuscript 1).**

4.1 Preface Manuscript 1

This manuscript is an original research about the use of spatial marked point process models to assess the distribution and concentration of dengue cases in Medellin, Colombia. Dengue is a known public health issue worldwide and specifically in Colombia and the main purpose of the manuscript is to present the methodology and to estimate quantitatively the contribution of sociodemographic factors while analyzing surveillance data for vector borne diseases in an endemic setting.

This manuscript illustrates the use of joint marked point process models to help to overcome some of the methodological issues associated with the use of surveillance data, while providing epidemiological and geographical information of high-risk areas of overall and severe dengue presence in Medellin, Colombia. The findings presented here are of the interest of the scientific community and contribute to produce useful information for different stakeholders. First, to inform the academic and research community about the possible use of the method, which has not been applied before on arboviruses, by presenting an empirical example and providing information of the current burden dengue in the study site. Second, the results obtained from this analysis could inform decision makers to address specific vector control strategies, and to help the preparedness of health services for upcoming outbreaks.

Submitted in June 2020 to the Spatial and Spatio-temporal Epidemiology Journal.
Manuscript ID: SSTE-D-20-00079

4.2 Manuscript 1

A joint spatial marked point process model for dengue and severe dengue in Medellin, Colombia.

Mabel Carabali*, Alexandra M. Schmidt, Bertha Nelly Restrepo, Jay S. Kaufman

Corresponding Author:

Mabel Carabali, MD, MSc, PhD(c): mabel.carabali@mail.mcgill.ca

Department of Epidemiology, Biostatistics and Occupational Health, McGill University

Purvis Hall, 1020 Pine Avenue West

Montreal, Quebec, Canada H3A 1A2

Tel.: 514-398-6258; Fax: 514-398-4503

Co-authors:

Alexandra M. Schmidt, PhD: alexandra.schmidt@mcgill.ca

McGill University, Montreal Quebec, Canada.

Bertha Nelly Restrepo, MD, MSc: brestrepo@ces.edu.co

Instituto Colombiano de Medicina Tropical (ICMT); Universidad CES, Sabaneta, Colombia.

Jay S. Kaufman, PhD: jay.kaufman@mcgill.ca

McGill University, Montreal, Quebec, Canada.

Word counts for abstract: 97

Word count for text: 3,673

Abstract

Surveillance-reported dengue cases and severity are usually analyzed separately, assuming independence between the overall dengue distribution and the presence of severity. To model surveillance-notified dengue cases and severity simultaneously, while identifying the spatial patterns of dengue distribution, we conducted a cross-sectional study using individual and area level covariates within a hierarchical Bayesian model. Results showed that age and socioeconomic status were associated with dengue presence, and there was evidence of clustering for overall cases but not for severity. Our findings inform decision making to address the preparedness or implementation of dengue control strategies at the local level.

Highlights

- A model to jointly assess the spatial distribution of reported dengue and severity
- This approach accounts for uncertainty in the surveillance-reported dengue cases
- We assessed spatial clustering using individual locations of dengue cases in Medellin.
- Young age was associated with higher dengue rates and older age to dengue severity.
- Non-monotonic distribution of reported dengue cases across socioeconomic status.

Key words: Dengue, surveillance, spatial analysis, point processes models, dengue transmission, Colombia, socioeconomic status.

BACKGROUND

Dengue is a vector-borne viral disease transmitted to humans by *Aedes* mosquitoes and an important public health problem worldwide[1-4]. The clinical presentation of dengue ranges from a self-limited mild febrile illness to severe outcomes[5,6]. Although lifelong immunity can be developed for each one of the four dengue serotypes[7-9]; secondary or subsequent infections from different dengue serotypes increase the risk of severe dengue, which is expected to occur in approximately 1% of all dengue cases, with reported mortality rates up to 20% worldwide[3,10,11].

Colombia is one of the Latin American countries with the highest burden of dengue[3,4] and within Colombia, dengue burden concentrates geographically in 50 of the 778 municipalities that routinely report dengue cases[1,2,12,13]. Given the spatial heterogeneity of dengue distribution, including its concentration in low socioeconomic settings, and the limitations of current dengue control strategies[14-18], it is important to investigate the spatial distribution of dengue cases. For instance, it is necessary to understand how individual level characteristics, in addition to area level covariates, are associated with the distribution of overall dengue cases. In addition, the analysis of severe cases is usually performed separately from the analysis of overall reported cases, assuming independence between overall presence of dengue and the presence of severity, often ignoring the potential underreporting associated with the use of surveillance data[4,15,19-25]. This unrealistic assumption potentially leads to further underestimation of the severity and the uncertainty associated to the individual factors related to severe cases[6,14,18,25]. Moreover, while analyzing severe cases, it is important to identify whether the distribution of severe cases follows a different spatial distribution from that of overall notified dengue cases. However, such analyses are rare and limited because they are: i) often constrained by data availability, ii) mainly conducted using aggregated area level data only, iii) often lacking proper adjustment of neighboring areas, and iv) usually computationally intensive[4,12,15,16,19-23].

Given the availability of individual dengue case locations (exact longitude and latitude), and in order to identify high-risk dengue areas while modelling *simultaneously* the cases and their severity, we conducted a single joint spatial marked-point-processes model of notified dengue

cases in Medellin, Colombia. We were motivated by the advantage of using individual level location and area level information to identify spatial patterns for clustering areas while properly accounting for spatial autocorrelation[20,21]. Hence, the main purpose of this study is to present the methodology and to estimate quantitatively the contribution of area-level and individual characteristics while analyzing surveillance data for vector borne diseases in endemic areas.

METHODS

Study site

Medellin is the second largest city in Colombia with 2.6 million inhabitants[26]. Annual dengue incidence ranged between 161 and 745 cases per 100,000 inhabitants over the last 10 years[1] and is consistently included on the top five dengue-reporting cities since 1998[2]. Medellin's urban area is composed of 269 neighborhoods, including 20 institutional units such as university campuses, jail facilities and military compounds, distributed over 110 km^2. Medellin's altitude ranges from 1,460 to 3,200 meters, the average temperature is 24^oC, and it has two rainy seasons (April and October). Although 50% of the city is classified as low socioeconomic status (SES), 98% of the city has access to potable water. The distribution of health coverage of the population is 70% contributory (employees or self-employees), 25% government subsidized, and 4% uninsured[26].

Data description

The data set comprises observations of individual location (exact longitude and latitude) of all notified dengue cases in Medellin in 2013 (n=1,793). Dengue notification in Colombia is mandatory and cases are individually registered in the national surveillance system (SIVIGILA), using the locally validated and standardized codes 210 and 220 for dengue and severe dengue, respectively[27].

Individual level covariates: Each row of the dataset included individual sociodemographic and clinical information for each notified case, including sex, age, residential and work/study addresses, date of notification, date of symptoms' onset, severity status, insurance scheme (subsidized vs contributory schemes)[28], and neighborhood of residence, all collected routinely in SIVIGILA's notification form[27].

Area level covariates: The neighborhood's population and socioeconomic status index (SES) were obtained through the office of development and planning at the local ministry of health and the Colombian Administrative Department of Demographic Survey (DANE)[26,29]. Entomological information, including the Breteau Index (IB) categorized as low, medium or high[27], was used to determine the neighborhood specific level of *Aedes* infestation and obtained from Medellin's local secretary of health.

Study design

We performed a cross-sectional study using a single joint spatial marked point process model, to simultaneously estimate the underlying process leading to the spatial patterns of overall and severe dengue cases[22,30].

Spatial point process model

A spatial point process assesses the distribution of the individual location of an outcome, over a spatial region[22,23]. Here, the individual spatial location (exact longitude and latitude) of an outcome is denominated by a point pattern[22,23,31]. We estimated the spatial distribution of individual notified dengue cases as a function of a continuous latent Gaussian random field[22,30,31]. Specifically, we proposed a Log-Gaussian Cox process discretizing the points (individual cases) using the neighborhood structure(14), which given the nature of the point process follows a Poisson distribution[22,23,31,32]. To identify whether there is an underlying mechanism leading to a different distribution of severe cases, we considered the presence of severity as an individual characteristic of each case and attributed it as a "mark" of the individual point. Since the presence of severity is conditional on being a case, we cannot assume independence between overall notified cases and severe cases. Therefore, the number of severe cases, conditioned on the total number of reported cases in each neighborhood, is assumed to follow a Binomial distribution. For this Binomial distribution, the probability of presence of severity is described by individual level fixed effects and an area latent spatial effect, which is proportional to the one used in the mean of the Poisson distribution for overall dengue cases. Here, the proposed approach has the advantage of i) *simultaneously* assessing the spatial distribution of overall dengue cases and

severe cases, by considering the spatial autocorrelation between and within spatial units and ii) accounting for the uncertainty associated with the reported number of dengue cases in the surveillance-based data [4,19,25,30,31] (Supplementary Material).

Model description

To fit a joint spatial marked-point-processes model we first constructed a model for each latent random field, one for the "pattern": overall cases and other for the "marks": severe cases[31], which *a priori* were specified as follows:

$$y_i|\eta_i^{(1)} \sim Po\left(E_i\,\eta_i^{(1)}\right), \text{ Equation (1)}$$

$$Overall\ cases\ model: log\left(\eta_i^{(1)}\right) = \beta_0^{(1)} + \beta_1^{(1)} IB(s_i) + \beta_2^{(1)} UNDER20(s_i) + \beta_3^{(1)} P.FEMALE(s_i) + \beta_4^{(1)} SES(s_i) + f_s^j(s_i) + u(s_i), \text{ Equation (2)}$$

$$m_i|\eta_i^{(2)} \sim Bin\left(y_i, p_{ij}\right), \text{ Equation (3)}$$

$$Severity\ model: logit\left(\eta_i^{(2)}\right) = \beta_0^{(2)} + \beta_1^{(2)} AGE(s_{ij}) + \beta_2^{(2)} SEX(s_{ij}) + \beta_3^{(2)} INSURANCE(s_{ij}) + \beta_4^{(2)} DISTANCEKm(s_{ij}) + \beta_s f_s^i(s_i) + v(s_i), \text{ Equation (4)}$$

Here, we assume that y_i, the total number of dengue cases observed in each neighborhood *i*, follows a Poisson distribution with mean ($E_i\,\eta_i^{(1)}$), where E_i is the expected count of cases in neighborhood *i*, obtained via indirect standardization using the city's disease rate[33] and $\eta_i^{(1)}$ is the Standardized Rate Ratio (SRR) for neighborhood *i*. Following equation (2), the SRR is decomposed as the sum of areal effects, spatially structured ($f_s^{(1)}(s_i)$) and independent random effects $u(s_i)$, modelled following the Besag-York-Mollie (BYM) specification[34]. The component ($f_s^{(1)}(s_i)$) is a Gaussian Markov Random Field (GMRF) *a priori* and works as the spatially structured effect for the pattern, which reflects the spatial autocorrelation (neighboring structure or vicinity) in the latent field that is not explained by the covariates (i.e., fixed effects)[20,31,35]. Other components of the overall cases model included $\beta_0^{(1)}$ which is the pattern's intercept and as fixed effects for the pattern of overall dengue cases we included the following neighborhood level covariates with their corresponding $\beta^{(1)}$ coefficients: Breteau Index (IB_i) categorized as low or medium; the proportion

of dengue cases under 20 years of age ($UNDER20_i$); the proportion of female dengue cases ($P.FEMALE_i$), and the socioeconomic status level (SES_i), a categorical variable with three levels (low SES level, medium SES level, and high SES level).

For the analysis of the severity "marks", in equation (3) m_i is the number of severe cases in each neighborhood *i*, which conditional on the value of a second random field $\eta_i^{(2)}$ follows a Binomial distribution, where $p_{ij} = exp\left(\eta_i^{(2)}\right)/\left(1 + exp\left(\eta_i^{(2)}\right)\right)$ is the probability of and individual *j* of being a severe case among the overall number of dengue cases y_i in neighborhood *i*. The $\text{logit}\left(\eta_i^{(2)}\right)$ is the random field for the marks (severity) at the individual level and $exp\left(\eta_i^{(2)}\right)$ is the odds ratio (OR) of severity; $\beta_0^{(2)}$is the marks' intercept and the individual level fixed effects covariates with their corresponding $\beta^{(2)}$ coefficients for the severity included a categorical variable for age: $AGE(s_{ij})$; and indicator for female sex: $SEX(s_{ij})$; the type of insurance: $INSURANCE(s_{ij})$ with 0 indicating subsidized scheme and 1 indicating a contributory scheme; and the minimum distance between severe cases per neighborhood $DISTANCEKm(s_{ij})$, which is the standardized nearest-neighbor (Euclidean) distance (km) between severe cases in each neighborhood. The component $\beta_s f_s^i(s_i)$ in equation (4) represents a single (common) random field that makes the structured spatial effect for the severity proportional to the spatial effect of the pattern[31], which is justified given that the presence of a severe case is conditional on the presence of a case, and $v(s_i)$ is the spatially unstructured random effect for the distribution of severe cases.

Data analysis

We calculated the respective descriptive statistics, and continuous estimates were presented as mean and standard deviation (SD) or as median and Interquartile Range (IQR), while categorical variables were presented as proportions. To inspect the observed distribution of cases, we plot the kernel density of the individual overall and severe dengue cases using a 5 km bandwidth[36].

The proposed joint spatial marked-point-processes model represents the two outcomes (overall reported dengue cases and severity) simultaneously in a hierarchical mixed-effects Bayesian

model. The overall disease pattern and the severity marks constitute a matrix outcome of two link functions (i.e., Poisson for overall dengue cases, and Binomial for severity); each one on a separate latent field η_i^j, which were jointly analyzed in relation to the vector of the sociodemographic covariates described above [31]. For the overall dengue pattern, we estimated the crude and adjusted Standardized Rate Ratio (SRR). For the severity marks, we estimated the odds ratio (OR), the respective probability of severity, and the overall and neighborhood-specific Relative Risk (RR) of severity. We assigned non-informative priors for the precision parameters of the random effects. The posterior distributions of the parameters and respective 95% Credible Intervals (95% Cr.I) were estimated via Integrated Nested Laplace Approximation (INLA)[31,35,37]. Model assessment to identify the variables included in the full models: equation (5) and (6), was performed through the Deviation Information Criterion (DIC)[20,22,31].

All analyses were fitted using R-INLA (R Core Team (2019); R Studio version 3.3.3)[23,37,38]. We followed the REporting of studies Conducted using Observational Routinely collected health Data (RECORD) statement guideline[39] (Supplementary Material).

Ethics statement: This study analyzed secondary data without identifying information, and therefore, did not require informed consent. The protocol was reviewed and approved by McGill's Institutional Review Board (Study No. A02-E05-18A) and by the ethics committee of the Secretary of Health of Medellin, Colombia.

RESULTS

In 2013, there were 1,793 cases reported in Medellin. In total, 1,719 (95.9%) were geocoded and were used for this analysis. There were 247 (14.4%) severe cases. Median age was 28 years (IQR=16 - 45) for overall dengue cases and 29 years (IQR=17 - 49) for severe cases. A descriptive analysis of notified cases and neighborhood characteristics is presented in **Table 4.1**.

Table 4.1 Individual and Neighborhood Characteristics of Dengue Cases Reported in Medellin, 2013.

	Overall cases	Severe cases
Individual Characteristics	**n (%)**	**n(%)**
Complete cases	1,719	247 (14.4)
Age, median (IQR)	28 (16, 45)	29 (17, 49)
Under 15 years	370 (21.5)	49 (19.8)
15-35 years	710 (41.3)	100 (40.5)
35-55 years	424 (24.7)	58 (23.5)
>55 Years	215 (12.5)	40 (16.2)
Sex (Male)	884 (51.4)	130 (52.6)
Insurance		
Subsidized	567 (33)	90 (36.4)
Contributory	1,152 (67)	157 (63.6)
DENV Classification		
Severe Dengue	247 (14.4)	-
Hospitalization	384 (22.4)	185 (74.9)
Distance between cases in Km, median (IQR)	5.15 (0.0, 12.1)	0.2 (0.1, 0.3)
Neighborhood Characteristics	**n (%)**	**n (%)**
Low SES Level	641 (37.3)	90 (36.4)
Medium SES Level	885 (51.5)	128 (51.8)
High SES Level	193 (11.2)	29 (11.7)
Breteau Index (Low)	999 (58.2)	132 (53.4)
Breteau Index (Medium)	717(41.8)	115 (46.6)
Proportion of Female cases*, median (IQR)	50.0 (40.0, 60.0)	50.0 (37.5, 56.0)
Proportion of cases* <20 years old, median (IQR)	32.0 (20.0, 44.8)	32.0 (20.0, 43.8)

**From all cases reported per neighborhood.*

The overall crude rate for reported dengue was 78 cases per 100,000 inhabitants. The median number of cases per neighborhood was four (IQR= 1- 10; range= 0- 57). The mean crude SRR was 1.3, standard deviation (SD: 2.4; range= 0- 20.6). The median number of severe cases per neighborhood was one (IQR=0- 2; range= 0, 10). There was an apparent concentration of both dengue and severe dengue cases on the northeastern neighborhoods that was observed by the crude distribution of geocoded cases (**Figure 4.1**) and the unadjusted (i.e., without accounting for population size) estimated density of cases (**Figure 4.2**).

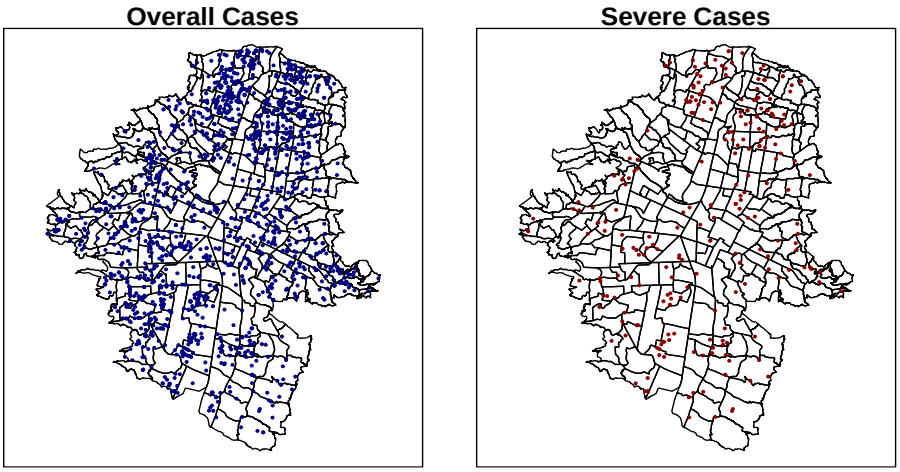


Figure 4-1 Distribution of dengue cases notified in Medellin in 2013.
Plot of observed location for overall and severe dengue cases.

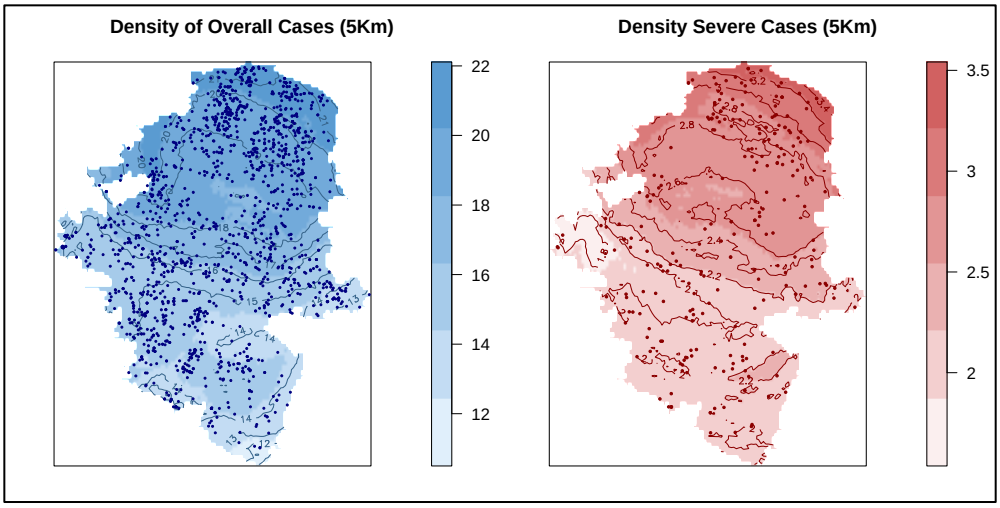


Figure 4-2 Distribution of dengue cases notified in Medellin in 2013.
Kernel density of overall and severe dengue cases using a 5 Km bandwidth. Label indicates the number of overall and severe cases per 5 Km^2.

The joint model showed that for the pattern of overall dengue cases, the mean baseline adjusted SRR per neighborhood was 0.78 (95% Cr.I=0.60, 1.01) and the average adjusted probability of severity per neighborhood was 0.92% (95% Cr.I=0.62%, 1.28%). Overall dengue rates increased with every 10% increase in the proportion of cases under 20 years old per neighborhood (SRR=1.06; 95% Cr.I=1.01, 1.10) and compared to people below 15 years old, severity tend to increase among people over 55 years old (OR=1.53; 95% Cr.I=1.00, 2.35). Female sex was associated to increased dengue rates (SRR= 1.05; 95% Cr.I=1.01, 1.09) but not to severity (OR=0.88; 95% Cr.I=0.68, 1.14). Just over half of reported cases were from neighborhoods with medium SES levels, and compared to these, dengue rates among neighborhoods in the Low SES level were on average 55% lower (SRR=0.45; 95% Cr.I=0.34, 0.59) and rates among neighborhoods with high SES level were on average 22% lower (SRR=0.78; 95% Cr.I=0.56, 1.09). Compared to neighborhoods with low Breteau Index (i.e., low *Aedes* presence), neighborhoods with a medium level of Breteau Index had slightly higher rates of dengue cases (SRR=1.12; 95% Cr.I=0.89, 1.40). Increased severity was observed with contributory insurance scheme and distance, but the estimates were imprecise (Table 4.2).

The spatial distribution of the crude SRR for overall dengue cases indicated the presence of dengue in the entire city with some concentration of dengue cases among neighborhoods in the central and the North-Eastern regions of the city. Likewise, compared to the overall odds of severity in the entire city, the distribution of severe cases indicated increased odds of severity among Southern and Eastern neighborhoods of the city (Figure 4.3). After adjusting for other covariates and comparing to the overall rate of dengue in the city, the spatially structured effect indicating the residual spatial autocorrelation not explained by the fixed effects, showed a widespread distribution of cases with some concentration in central and Northern parts of the city. For severity marks, the residual spatial effect showed a homogeneous distribution of severe cases without indication of concentration of cases in any particular neighborhood (Figure 4.4). The beta coefficient for the spatial effect of severe cases indicated that after accounting for the other covariates in the model, and given the distribution of overall dengue cases, there is no indication

that severe dengue shares the same latent spatial effect as the one for the dengue cases (OR= 0.84; 95% Cr.I=0.60, 1.27).

As a sensitivity analysis we fit the joint model using separate spatial structures for patterns and marks. The results from the fixed effects were similar to the main results presented here, however both the fixed effects and the hyperparameters for the spatial effect were less precise, given that this approach for severe cases does not borrow strength across the dengue cases (Supplemental Material).

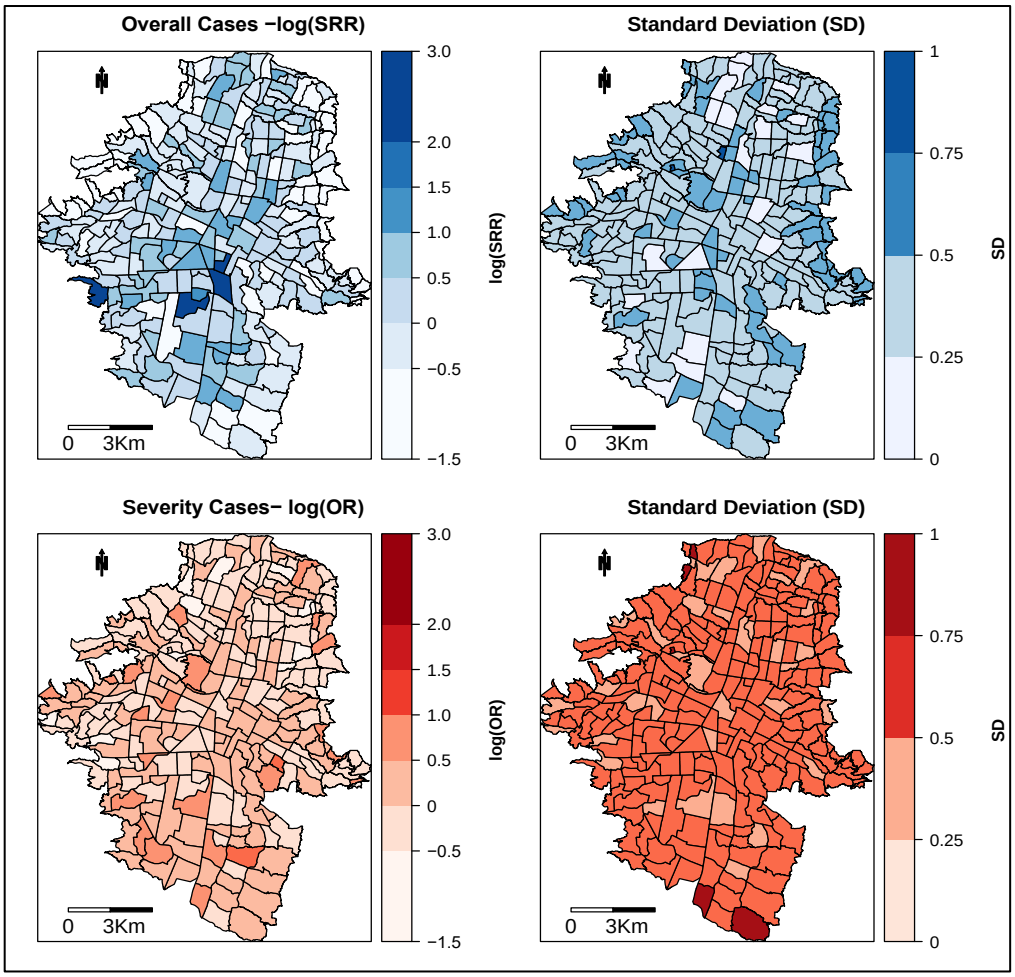


Figure 4-3 Spatial distribution of overall and severe dengue cases in Medellin, 2013.
(Top) Neighborhood specific crude Standardized Rate Ratios (SRR) and standard deviation (SD) for overall dengue cases. (Bottom) Neighborhood specific crude Odds Ratios (OR) and standard deviation (SD) for severe dengue cases.

Table 4.2. Posterior mean of the Standardized Rate Ratio (SRR), the Odds Ratio (OR) and 95% Credible Intervals (95% Cr.I) for covariates (fixed effects) in the final joint model for dengue cases in Medellin, 2013 (DIC=2762.90).

Fixed Effects Covariates	Standardized Rate Ratio or Odds Ratio and 95% Credible Intervals	
Overall Cases (Pattern)	**SRR**	**95% Cr.I**
Proportion of Female Cases	1.05	[1.01, 1.09]
Proportion of cases <20 years old	1.06	[1.01, 1.10]
Entomological Index		
Low Breteau Index	Ref.	-
Medium Breteau Index	1.12	[0.89, 1.40]
SES Level: Medium	Ref	-
SES Level: Low	0.45	[0.34, 0.59]
SES Level: High	0.78	[0.56, 1.09]
Severe Cases (Marks)	**OR**	**95% Cr.I**
Sex		
Male	Ref.	-
Female	0.88	[0.68, 1.14]
Age: <15 years	Ref.	-
Age: 15-34 Years	0.96	[0.68, 1.37]
Age: 35-54 Years	1.08	[0.73, 1.59]
Age: >55 Years	1.54	[1.01, 2.36]
Health Insurance: Subsidized	Ref.	-
Health Insurance: Contributory	1.03	[0.79, 1.35]
Distance to Severe cases (Km)	1.21	[0.66, 2.14]
Spatial Effect (β_s)	0.84	[0.60, 1.27]

*****Area level covariates: ***Proportion of Female Cases:*** *indicates 10% increase in the proportion of cases reported per neighborhood;* ***Proportion of cases <20 years old:*** *indicates 10% increase in the proportion of reported cases <20 years old per neighborhood.* ***Breteau Index:*** *Comparing the Low Breteau Index level (Reference group) to Medium Breteau Index level;* ***SES Level:*** *Comparing cases in the Medium SES (Reference group) to Cases in the Low and High SES levels.* ******Individual covariates: ***Sex:*** *Comparing Female cases to Males;* ***Age group:*** *<15 years of age (Reference group);* ***Health Insurance****: Comparing cases on the Subsidized scheme (Reference group) to the Contributory system;* ***Distance to Severe cases (Km):*** *indicates the minimum distance between the nearest-neighbor severe cases per neighborhood in Kilometer.*

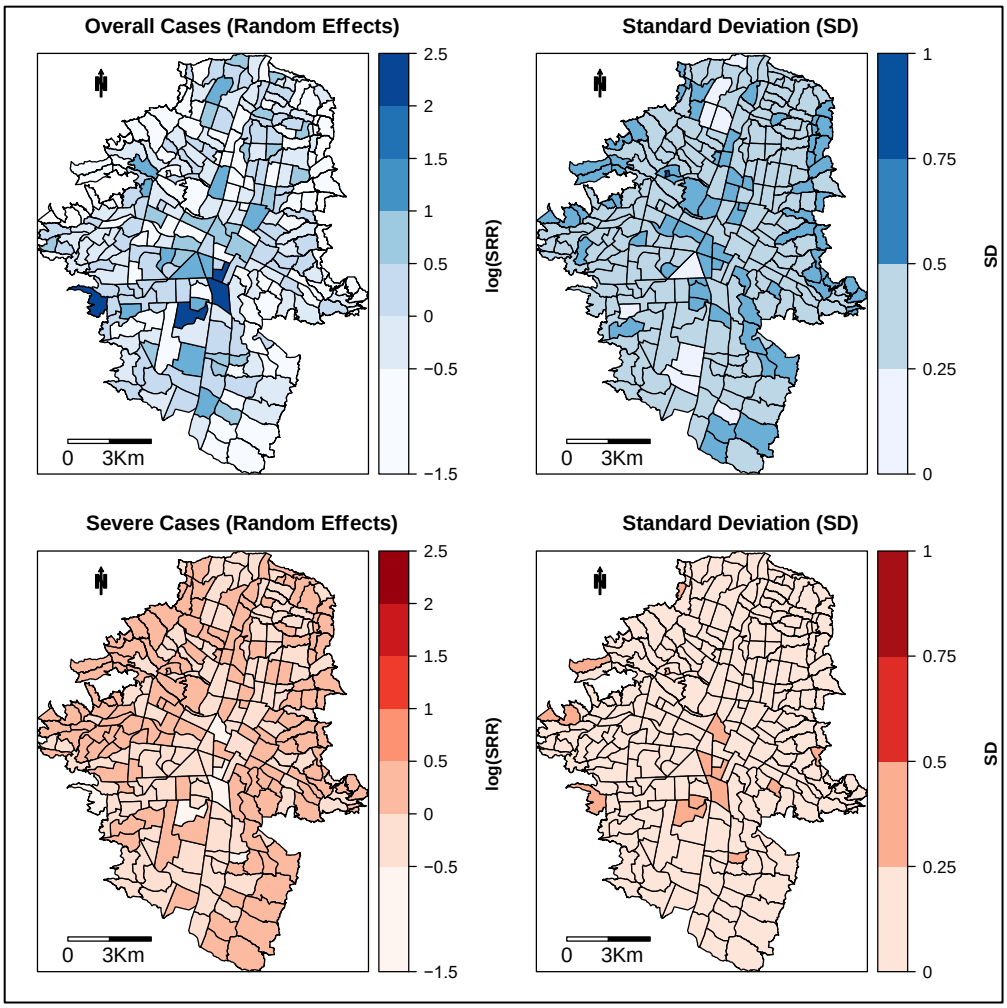


Figure 4-4 Estimated common spatial trend for overall dengue and Severe dengue cases in Medellin, 2013.
(Top) Neighborhood specific residual (Random Effects) Standardized Rate Ratio (nSRR) and Standard Deviation (SD). (Bottom) Neighborhood specific residual (Random Effects) Odds Ratio (nOR) and Standard Deviation (SD).

DISCUSSION

We presented an analysis of a joint spatial marked point processes model on routinely collected dengue data. Our study shows the possibility of simultaneously estimating the distribution of overall dengue cases and the distribution of severity, accounting for the uncertainty associated to the reporting of dengue cases in surveillance-based data, allowing for spatial autocorrelation, and using individual sociodemographic covariates to explain such outcomes.

Dengue discussion

Colombia is an endemic country and Medellin is one of the municipalities consistently reporting a high burden of cases during the last decade[1,2,27]. Our study shows that during 2013, dengue was present in the entire city, with concentration at the Northeastern neighborhoods, which are known for being densely populated areas[18,26,40]. The concentration of cases in the Eastern region has been previously explored in the context of serological surveys [41] and among children attending different schools in the city[18]. However, previous approaches did not include latent spatial structured effects that account for the neighboring structure after adjusting for available covariates. Also, previous approaches either used aggregated data, fit fixed effects for the spatial structure or modeled separately the spatial effects and the contribution of sociodemographic covariates[14,18,40,41].

In our study, there were no neighborhoods with high Breteau Index and there was no association between the SRR and Breteau Index. Although the Breteau Index is considered a useful indicator of *Aedes* infestation, there is conflicting evidence about the concordance with presence of dengue cases[1,15]. This could arise in our data because entomological information was collected at regular intervals throughout the year in different neighborhoods and households[18,27,40]. The value of entomological indexes changes over time, but the timing of exposure assessment and incident cases may not be aligned[42-44].

Although the proportion of female cases was associated with a slight increased rate in the overall distribution of cases, being female was not associated with severity in our study. Increased proportion of female dengue cases has been also reported in Medellin previously[18]. However, associations between sex, dengue and dengue severity have been inconsistent in the literature[11,40,41]. Age, specifically the proportion of people under 20 years of age, was associated with increased rates of overall dengue cases across neighborhoods and an increased OR for severity was observed among people over 55 years old. These findings could be associated with a high seroprevalence of dengue in the city and a limited presence of secondary infections[2,11,40]. In Medellin, the overall dengue seroprevalence was estimated at 61%, with a mean age of 30 years

among dengue seropositive cases. The overall seroconversion rates were estimated to increase with age, with the highest seroconversion rate (17.9 per 1,000 people) observed among subjects between 31 and 40 years of age[41]. Likewise, among school children under 19 years old, a trend of increased dengue seroprevalence and seroconversion with age has been reported[18]. However, it is also possible that the observed trend of severity by age could be related to comorbidities in older patients and the possibility of secondary infections in people over 55 years ol [6,9,11]. These characteristics have been described in other Colombian municipalities and in other Latin-American contexts[1,2,11,14,16,24,40,45]; and may contribute to an understanding of the age-related findings in this study.

Health insurance was modeled as a proxy of socioeconomic status at the individual level[29,46,47]. In Colombia, specifically, the subsidized system corresponds to individuals for whom the state pays for health coverage. The contributory system corresponds to employed individuals or people with capacity to pay for their health system coverage (affiliated to a private insurance plan or out-of-pocket). In our study, there was no association between insurance scheme and severity [1,2,46-48]. According to the SES level of the neighborhood of residency, findings from the joint model suggests a non-monotonic distribution of cases across SES levels, with fewer cases at low and high SES levels. The presence of fewer cases among neighborhoods at low SES level could be attributed to limited access (physical and financial) to health care, compared to people living in neighborhoods with medium- or higher SES levels[1,2,46-48]. Although the rate of dengue cases seems to decrease in neighborhoods at the high SES level, the lack of precision of the estimates could be attributed to the small number of cases in this stratum (n=193 cases). Nonetheless, reporting bias and spatial confounding associated to the SES level could not be completely ruled out.

Implications of routinely collected data

We used passive surveillance data, which implies a potential risk of under reporting and measurement error[17,19,25,27,47,49]. Notification depends on health seeking behavior, which in turn depends on presence and severity of symptoms and access to health care (insurance scheme, availability of health care facility, etc.) that altogether could also depend on other socioeconomic

factors[15,47,49]. Therefore, the findings from this analysis should be restricted to the subset of notified cases. For this analysis we worked closely with the municipality's secretary of health, which is considered one of the strongest surveillance systems in the country and for which dengue is a disease of mandatory notification[18,27,41]. The diagnostic system in place, including serological and clinical confirmation, decreased the risk of misclassification of the outcome but did not ruled it out completely.

Methodological discussion

This joint spatial marked point process analyses the distribution of individual-level data on dengue cases, adjusting for neighboring effects via spatial structured effects, and accounting for area level covariates simultaneously. The advantage of using a joint model to assess the spatial distribution of severe cases relies on three main aspects i) the opportunity to use individual location data for overall and severe cases to assess their distribution, ii) the opportunity to account for the uncertainty associated with the number of overall dengue cases in the surveillance-based data, and iii) the opportunity of identifying the presence of clustering of severe cases that will otherwise not be identified with separated models. This approach assumes that there is a spatial trend in the data that cannot be explained by the measured covariates and that such trend is a random field[22,23,31]. In our study, the addition of a covariate for the minimum nearest-neighbor Euclidean distance between cases allows the investigation of possible clustering within a neighborhood[31]. Although the minimum distance between severe cases indicated a local clustering for severe cases, the results were imprecise possibly due to the small number of severe dengue cases overall and within neighborhood. Nonetheless, the joint model made the spatial distribution of severe cases proportional to the distribution of overall cases and allowed the identification of the spatial distribution of severe cases and improved the precision.

Typically, point process models are fitted through the use of a regular spatial grid which approximates the latent field and the spatial pattern [20,31]. However, for ease of applicability among the public health community, data availability, and to avoid issues associated with the interpolation of population offsets, we used the actual neighborhood map and population

information as the spatial grid. This approach facilitated the fitting by providing the real neighboring boundaries and used the actual information of the population, area, and density to improve accuracy. The use of this dataset favors the use and application of research results in the context of surveillance and disease control by decision makers and other stakeholders.

Conclusion

These findings provide epidemiological and geographical information of high-risk areas of overall and severe dengue presence in Medellin, Colombia. Age, insurance scheme, and distance between cases are key sociodemographic and spatial factors associated with the presence of dengue in the city. The use of joint marked point process models improves the evidence obtained from surveillance data by accounting for the uncertainty of overall reported dengue cases and by favoring its analysis at the individual level when data is available. This application contributes to the production of public health information for decision makers to address specific disease control strategies, and to help the preparedness of health services for upcoming outbreaks at the local level.

Data accessibility: Case-specific data, which is routinely collected using the national surveillance system of Colombia (SIVIGILA; http://portalsivigila.ins.gov.co/sivigila/index.php) was obtained directly form the Local Surveillance office (Secretaria de Salud Municipal de Medellin); Socioeconomic information at neighborhood level was obtained from the website of the municipality (https://www.medellin.gov.co) and an open data source for socioeconomic information (https://www.datos.gov.co/).

Acknowledgements: We would like to acknowledge the collaboration of the Secretaria de Salud Municipal de Medellin, especially Dr. Rita Almanza and Dr. Gabriel Parra Henao for their collaboration on accessing the datasets. Mabel Carabali holds a Banting-Best Canadian Graduate Studies Doctoral Scholarship from the Canadian Institutes of Health Research (CIHR).

Conflict of Interest: The authors declare no conflict of interest

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

1. Instituto Nacional de Salud (INS). Informe Epidemiologico de Evento Dengue. In: Vigilancia y Análisis del Riesgo en Salud Pública., ed. Dengue. Vol. FOR-R02.4000-001. Bogota, Colombia: Instituto Nacional de Salud, 2016.
2. Villar LA, Rojas DP, Besada-Lombana S, Sarti E. Epidemiological trends of dengue disease in Colombia (2000–2011): a systematic review. *PLoS Negl Trop Dis* 2015;**9**.
3. Pan American Health Organization / World Health Organization. Epidemiological Update: Dengue. Washington, D.C.: PAHO/WHO, 2019;15.
4. World Health Organization (WHO). A Toolkit for National Dengue Burden Estimation. In: World Health Organization (WHO), ed. Geneva, 2018.
5. Pan American Health Organization (PAHO). *Tool for the diagnosis and care of patients with suspected arboviral diseases*. 1st Edition ed. Vol. PAHO Strategic Plan 2014-2019. Washington: PAHO, 2017.
6. Mohanty B, Sunder A, Pathak S. Clinicolaboratory profile of expanded dengue syndrome - Our experience in a teaching hospital. *Journal of family medicine and primary care* 2019;**8**(3):1022-1027.
7. Waggoner JJ, Gresh L, Mohamed-Hadley A, Balmaseda A, Soda KJ, Abeynayake J, Sahoo MK, Liu Y, Kuan G, Harris E, Pinsky BA. Characterization of Dengue Virus Infections Among Febrile Children Clinically Diagnosed With a Non-Dengue Illness, Managua, Nicaragua. *J Infect Dis* 2017;**215**(12):1816-1823.
8. Halstead SB. Antibodies Determine Virulence in Dengue. *Annals of the New York Academy of Sciences* 2009;**1171**:E48-E56.
9. Wearing HJ, Rohani P. Ecological and immunological determinants of dengue epidemics. *Proceedings of the National Academy of Sciences* 2006;**103**(31):11802-11807.
10. World Health Organization (WHO). Global strategy for dengue prevention and control 2012-2020. Geneva: WHO. World Health Organization 2012;vi, 43p.
11. Carabali M, Hernandez L, Arauz M, Villar L, Ridde V. Why are people with dengue dying? A scoping review of determinants for dengue mortality. *BMC Infectious Diseases* 2015;**15**(1):301.

12. Rico-Mendoza A, Alexandra P-R, Chang A, Encinales L, Lynch R. Co-circulation of dengue, chikungunya, and Zika viruses in Colombia from 2008 to 2018. *Revista panamericana de salud publica = Pan American journal of public health* 2019;**43**:e49-e49.

13. Martínez-Bello DA, López-Quílez A, Prieto AT. Joint Estimation of Relative Risk for Dengue and Zika Infections, Colombia, 2015-2016. *Emerging infectious diseases* 2019;**25**(6):1118-1126.

14. Martínez-Bello DA, López-Quílez A, Torres Prieto A. Relative risk estimation of dengue disease at small spatial scale. *International Journal of Health Geographics* 2017;**16**(1):31.

15. Vanlerberghe V, Gómez-Dantés H, Vazquez-Prokopec G, Alexander N, Manrique-Saide P, Coelho G, Toledo ME, Ocampo CB, Stuyft PVd. Changing paradigms in Aedes control: considering the spatial heterogeneity of dengue transmission. *Rev Panam Salud Publica* 2017;**41**(e16):1-6.

16. Vincenti-Gonzalez MF, Grillet M-E, Velasco-Salas ZI, Lizarazo EF, Amarista MA, Sierra GM, Comach G, Tami A. Spatial Analysis of Dengue Seroprevalence and Modeling of Transmission Risk Factors in a Dengue Hyperendemic City of Venezuela. *PLOS Neglected Tropical Diseases* 2017;**11**(1):e0005317.

17. Fritzell C, Rousset D, Adde A, Kazanji M, Van Kerkhove MD, Flamand C. Current challenges and implications for dengue, chikungunya and Zika seroprevalence studies worldwide: A scoping review. *PLOS Neglected Tropical Diseases* 2018;**12**(7):e0006533.

18. Piedrahita LD, Agudelo Salas IY, Marin K, Trujillo AI, Osorio JE, Arboleda-Sanchez SO, Restrepo BN. Risk Factors Associated with Dengue Transmission and Spatial Distribution of High Seroprevalence in Schoolchildren from the Urban Area of Medellin, Colombia. *Canadian Journal of Infectious Diseases and Medical Microbiology* 2018;**2018**:11.

19. Louis VR, Phalkey R, Horstick O, Ratanawong P, Wilder-Smith A, Tozan Y, Dambach P. Modeling tools for dengue risk mapping - a systematic review. *International Journal of Health Geographics* 2014;**13**(1):1-15.

20. Blangiardo M, Cameletti M. Spatial modeling. *Spatial and Spatio-temporal Bayesian Models with R-INLA* John Wiley & Sons, Ltd, 2015;173-234.

21. Congdon P. Models for spatial outcomes and geographical association. *Applied Bayesian Modelling* John Wiley & Sons, Ltd, 2014;312-363.

22. Diggle PJ, Moraga P, Rowlingson B, Taylor BM. Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm. *Statistical Science* 2013;**28**(4):542-563.

23. Illian JB, Sørbye SH, Rue H. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics* 2012:1499-1530.

24. Imai N, Dorigatti I, Cauchemez S, Ferguson NM. Estimating Dengue Transmission Intensity from Sero-Prevalence Surveys in Multiple Countries. *PLoS Negl Trop Dis* 2015;**9**(4):e0003719.

25. Vong S, Goyet S, Ly S, Ngan C, Huy R, Duong V, Wichmann O, Letson GW, Margolis HS, Buchy P. Under-recognition and reporting of dengue in Cambodia: a capture–recapture analysis of the National Dengue Surveillance System. *Epidemiol Infect.* 2011;**140**(3):491-499.

26. Departamento Nacional de Estadistica de Colombia (DANE). Estimaciones de población 1985 - 2005 y proyecciones de población 2005 - 2020 total municipal por área. *Departamento Nacional de Estadistica de Colombia* DANE, 2015.

27. Instituto Nacional de Salud (INS). Protocolo de Vigilancia en Salud Pública, Dengue (Surveillance Protocol in Public Health, Dengue). . In: Instituto Nacional de Salud (INS) C, ed. *Surveillance*. Dirección de Vigilancia y Análisis de Riesgo en Salud Pública ed. Vol. FOR-R02.0000-059 V02. Santafé de Bogota, Colombia. : Instituto Nacional de Salud (National Institute of Health). , 2014;19.

28. Ministerio de la protección Social. Ley 1438: Por medio de la cual se reforma el Sistema General de Seguridad Social en Salud y se dictan otras disposiciones. In: Social. Mdlp, ed. *IV Aseguramiento.* Bogotá, Colombia.: Congreso de la Republica de Colombia, 2011.

29. Departamento Administrativo Nacional de Estadistica (DANE). Metodologia de la Estratificacion Socioeconomica Urbana para Servicios Publicos Domicioliarios. In: Direccion de Geoestadistica, ed. Grupo de Estratificacion ed. Santa Fe de Bogota, Colombia.: DANE, 2015;96.

30. Simpson D, Illian JB, Lindgren F, Sorbye S, Rue H. Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *arXiv* 2017:1-26.

31. Illian JB, Martino S, Sorbye S, Gallego-Fernandez JB, Zunzunegui M, Esquivias MP, Travis JMJ. Fitting complex ecological point process models with Integrated Nested Laplace Approximation. *Methods in Ecology and Evolution* 2013;**4**:305-315.

32. Pinto Junior JA, Gamerman D, Paez MS, Alves Regina HF. Point pattern analysis with spatially varying covariate effects, applied to the study of cerebrovascular deaths. *Statistics in Medicine* 2015;**34**:1214-1226.

33. Waller LA, Gotway CA. Analyzing Public Health Data. . In: Shewhart WA, Wilks SS, eds. *Applied Spatial Statistics for Public Health Data*. Vol. 368 John Wiley & Sons, 2004;7-37.

34. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991;**43**(1):1-20.

35. Riebler A, Sørbye SH, Simpson D, Rue H. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research* 2016;**25**(4):1145-1165.

36. Baddeley A, Rubak E, Turner R. *Spatial point patterns: methodology and applications with R* CRC Press, 2015.

37. Rue H, Riebler A, Sørbye SH, Illian JB, Simpson DP, Lindgren FK. Bayesian computing with INLA: a review. *Annual Review of Statistics and Its Application* 2017;**4**:395-421.

38. R Core Team. R Foundation for Statistical Computing. Desktop 1.0.143. R version 3.6.0 - "Planting of a Tree" ed. Vienna, Austria: R Foundation for Statistical Computing, 2019.

39. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, Committee RW. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Medicine* 2015;**12**(10):e1001885.

40. Restrepo BN, Beatty ME, Goez Y, Ramirez RE, Letson GW, Diaz FJ, Piedrahita LD, Osorio JE. Frequency and Clinical Manifestations of Dengue in Urban Medellin, Colombia. *Journal of Tropical Medicine* 2014;**2014**:8.

41. Carabali M, Lim JK, Velez DC, Trujillo A, Egurrola J, Lee KS, Kaufman JS, DaSilva LJ, Velez ID, Osorio JE. Dengue virus serological prevalence and seroconversion rates in children and adults in Medellin, Colombia: implications for vaccine introduction. *International Journal of Infectious Diseases* 2017;**58**:27-36.

42. Boyer S, Foray C, Dehecq J-S. Spatial and temporal heterogeneities of Aedes albopictus density in La Reunion Island: rise and weakness of entomological indices. *PLoS One* 2014;**9**.

43. Chiaravalloti-Neto F, Pereira M, Fávaro EA, Dibo MR, Mondini A, Rodrigues-Junior AL, Chierotti AP, Nogueira ML. Assessment of the relationship between entomologic indicators of Aedes aegypti

and the epidemic occurrence of dengue virus 3 in a susceptible population, São José do Rio Preto, São Paulo, Brazil. *Acta Tropica* 2015;**142**:167-177.

44. Ocampo CB, Mina NJ, Carabalí M, Alexander N, Osorio L. Reduction in dengue cases observed during mass control of Aedes (Stegomyia) in street catch basins in an endemic urban area in Colombia. *Acta Tropica* 2014;**132**(0):15-22.

45. Carabali M, Lim JK, Palencia DC, Lozano-Parra A, Gelvez RM, Lee KS, Florez JP, Herrera VM, Kaufman JS, Rojas EM, Villar LA. Burden of dengue among febrile patients at the time of chikungunya introduction in Piedecuesta, Colombia. *Tropical Medicine & International Health* 2018;**23**(0):1231-1241.

46. Ardila Pinto F, Martínez S, Fuentes M, Borrero E. Análisis de las demoras en salud en personas que enfermaron de gravedad o fallecieron por dengue en cinco ciudades de Colombia. *Physis: Revista de Saúde Coletiva* 2015;**25**:571-592.

47. Carabalí JM, Hendrickx D. Dengue and health care access: the role of social determinants of health in dengue surveillance in Colombia. *Glob Health Promot* 2012;**19**.

48. Arauz MJ, Ridde V, Hernández LM, Charris Y, Carabali M, Villar LÁ. Developing a Social Autopsy Tool for Dengue Mortality: A Pilot Study. *PLoS ONE* 2015;**10**(2):e0117455.

49. Wichmann Ole, Yoon IK, Vong S, Limkittikul K, Gibbons RV, Mammen PM, Sowath Ly, Buchy P, Sirivicgayakul C, Buathong R, R H, Letson GW, Sabchareon A. Dengue in Thailand and Cambodia: An Assessment of the Degree of Underrecognized Disease Burden Based on Reported Cases. *PLoS Negl Trop Dis* 2011;**5**(3):e996.

4.3 Supplementary Material Manuscript 1

A joint spatial model for dengue and severe dengue in Medellin, Colombia.

1. Spatial point process model:

A log-Gaussian Cox point process model assesses the distribution of the individual location of the outcome points (dengue cases) in a spatial structure, and it is used in order to consider both, observed and unobserved variation in the assessment of such distribution[1]. It estimates the spatial distribution of the individual cases as a function of a continuous latent Gaussian random field, assuming conditional independence of the points presented on the field. This indicates that conditional on the latent field, the distribution of the point pattern (dengue cases), follows a Poisson process[1-3]. This analysis uses individual level information and allows covariates at the spatial level to vary according to the random field. Thus, providing information about the presence and degree of clustering within the spatial structure, while considering simultaneously the spatial autocorrelation between and within spatial units[4].

A Marked log-Gaussian Cox point process model uses an individual characteristic of the point, the 'mark', to assess the individual distribution of an event given such specific characteristic. In our case, each point represents an individual case, and the "mark" is the presence/absence of severity for each dengue case. Thus, the information about the "mark" is used to estimate two aspects simultaneously: 1) the overall distribution of the point pattern (i.e., all points/cases) and 2) the distribution of the point pattern given the mark (i.e., severe cases). The model could also be considered a "labeling" of the Poisson process, where the 'marks' work also as a response variable[2]. This approach is intended to identify whether there is an underlying mechanism leading to a differential distribution of severe cases, with the added advantage of modelling simultaneously an individual feature of the point (the severity) and the spatial distribution, while accounting for its dependence[2,3].

To conduct a point process analysis, it is necessary to consider each case-location as $x_i: i = 1, \ldots \mathrm{n}$, where, x indicates the location and i indicates the dengue case identifier that in theory could have occurred in any location inside a given spatial region[1] $\mathrm{A} \subset \Re^2$. Likewise, it is important to accept two main structural assumptions: i) that the function of the Cox-Process is a stochastic non-negative process $\Lambda = \{\Lambda(x): \mathrm{x} \in \Re^2\}$, and ii) that conditional on the realization $\Lambda(x) = \{\lambda(x): x \in \Re^2\}$, the point process is an inhomogeneous Poisson process with intensity $\lambda(\mathrm{x})$. Also, it is necessary to consider the distribution of dengue cases as a phenomenon $S(x)$ that is incompletely observed and spatially continuous, given that $S = \{S(x): x \in \Re^2\}$ is a Gaussian stochastic process and that S determines $\lambda(x)$ which is the intensity of the distribution[1,3102,181], identified by $\lambda(x) = exp\{S(x)\}$. To analyze this intensity, it is necessary to approximate the spatially continuous random field to a constructed grid[1,3,5]. Then, considering $\{y_i\}$ the observed number of points in the neighborhood, we assume that the number of points (cases) in a grid-cell/neighborhood i follows a Poisson distribution conditional on a first latent field, $\eta_i^{(1)}$:

$$y_i|\eta_i^{(1)} \sim Po\left(E_i\, \eta_i^{(1)}\right), \text{ Equation (1)}$$

The offset of the pattern E_i is specified as the expected count of cases in each neighborhood and obtained via indirect standardization[6]. To model the marked point process, we add to equation 1, the analysis of the marks (severe or not severe). For that we let m_i be the number of patients with severe dengue in each spatial unit i . Then, conditional on the value of a second latent field $\eta_i^{(2)}$ in the same neighborhood, m_i follows a binomial distribution:

$$m_i|\eta_{ij}^{(2)} \sim Bin(y_i, p_{ij}), \text{ Equation (2)}$$

where $p_{ij} = exp\left(\eta_{ij}^{(2)}\right) / \left(1 + \exp\left(\eta_{ij}^{(2)}\right)\right)$ is the probability of individual j being a severe case in a given neighborhood i while y_i is the total number of cases of dengue in neighborhood i. This, constituting a matrix outcome of two links (i.e., Poisson for overall point patterns, and Binomial for severity), each one on a separate latent field η, that are jointly analyzed in relation to a vector of sociodemographic covariates[180]. We constructed the final model for each latent field $\eta_i^{(1)}$ and

$\eta_{ij}^{(2)}$, including empirical covariates for neighborhood level and individual level as fixed effects, and spatial structures as random effects as follows:

$$log\left(\eta_i^{(1)}\right) = \beta_0^{(1)} + \beta_1^{(1)} IB(s_i) + \beta_2^{(1)} UNDER20(s_i) + \beta_3^{(1)} P.FEMALE(s_i) + \beta_4^{(1)} SES(s_i) + f_s^j(s_i) + u(s_i), \text{ Equation (3)}$$

$$logit\left(\eta_{ij}^{(2)}\right) = \beta_0^{(2)} + \beta_1^{(2)} AGE(s_{ij}) + \beta_2^{(2)} SEX(s_{ij}) + \beta_3^{(2)} INSURANCE(s_{ij}) + \beta_4^{(2)} DISTANCEKm(s_{ij}) + g_s^j(s_i) + v(s_i), \text{ Equation (4)}$$

$\beta_0^{(1)}$ and$\beta_0^{(2)}$ are the pattern and marks intercepts. $\beta_1^{(1)}$ … $\beta_4^{(1)}$ are the coefficients associated to the empirical covariates for the distribution of cases at the neighborhood level; and $\beta_1^{(2)}$ … $\beta_4^{(2)}$ are the coefficients associated to the empirical covariates for severity at the individual level, as described in the main text. $DISTANCEKm(s_{ij})$ is the standardized minimum nearest-neighbor Euclidean distance between overall and severe cases, parameterized as a continuous variable. The components $f_s^j(s_{ij})$ and $g_s^j(s_{ij})$ are the Gaussian Markov Random Field (GMRF), reflecting separately the spatial autocorrelation in the latent field, working as spatially structured effects for the pattern and the marks, respectively. $u(s_i)$is the spatially unstructured random effect for the pattern and $v(s_i)$ is the spatially unstructured random effect for the marks. To express the dependence between the pattern and the marks, we used a single (common) random field replacing equation (4) as follows:

$$logit\left(\eta_{ij}^{(2)}\right) = \beta_0^{(2)} + \beta_1^{(2)} AGE(s_{ij}) + \beta_2^{(2)} SEX(s_{ij}) + \beta_3^{(2)} INSURANCE(s_{ij}) + \beta_4^{(2)} DISTANCEKm(s_{ij}) + \beta_s f_s^i(s_i) + v(s_i), \text{ Equation (5)}$$

which makes the spatial effect for the severity proportional to the spatial effect of the pattern of case distribution[2]. The spatially structured and unstructured effects were modeled following the Besag-York-Mollie (BYM) specification[7]; where after adjusting for the fixed effects, the structured component is modeled using an intrinsic conditional autoregressive structure (iCAR) and the unstructured effect is modeled using an independent prior[6,8,9].

1.1 Fitting separated models for each random field

Table S1a. Posterior mean of the **Incidence Rate Ratio** and 95% credible intervals for covariates (fixed effects) on the single-separated model for **overall dengue** cases in Medellin, 2013

Characteristic	IRR	2.5%	97.5%
(Intercept)	0.78	0.60	0.42
Proportion of Female cases	1.07	1.01	1.12
Proportion of cases <20 years old	1.06	1.01	1.12
Medium Breteau Index	1.10	0.84	1.44
Low SES	0.43	0.33	0.56
High SES	0.77	0.55	1.07

Proportion of female cases and proportion of cases under 20 years old indicate 10% increase in the proportion of cases per neighborhood. SES reference group = Medium SES; Breteau Index reference group = Low.

Table S1b. Posterior mean and 95% credible intervals of Hyperparameters:

Parameter	Mean	SD	2.5%	97.5%	mode
Precision for $f_s^j(s_{ij}) + u_{(s_{ij})}$	1.69	0.26	1.23	2.28	1.63
Phi for $u_{(s_{ij})}$	0.38	0.14	0.13	0.68	0.34

Table S2a. Posterior mean of the Odds Ratio and 95% credible intervals for covariates (fixed effects) on the single-spearated model for **severe dengue** cases in Medellin, 2013

Characteristic	OR	2.5%	97.5%
Age			
15-34 Years	1.09	0.78	1.55
35-54 Years	1.08	0.74	1.59
>55 Years	1.49	0.98	2.27
Sex (Female)	0.92	0.72	1.19
Contributory Insurance	0.86	0.66	1.12
Distance to Severe cases (Km)	0.50	0.27	0.89

Age group reference= <15 years of age; Sex reference= Male; Health Insurance reference is Subsidized scheme.

Table S2b. Posterior mean and 95% credible intervals of Hyperparameter:

Parameter	Mean	SD	2.5%	97.5%	mode
Precision $g_s^j(s_{ij}) + v_{(s_{ij})}$	514.20	3018.98	12.39	3360.07	26.67
Phi for $v_{(s_{ij})}$	0.31	0.26	0.01	0.90	0.02

1.2 Sensitivity Analyses: Joint models

In addition to the full model, we fitted other joint models including a model without the SES covariate and the full joint model using a different spatial effect for severity. Models were examined using Deviation Information Criteria (DIC) and are summarized in Table S3 and results are presented in Table S4 and S5 respectively.

Table S3. Summary of Deviance Information Criterion (DIC) values and specification for joint models of overall (pattern) and severe (marks) dengue cases.

Model	Model Components	DIC
Full without the SES covariate.	$log\left(\eta_i^{(1)}\right) = \beta_0^{(1)} + \beta_1^{(1)} IB(s_i) + \beta_2^{(1)} UNDER20(s_i) + \beta_3^{(1)} P.FEMALE(s_i) + f_s^j(s_i) + u(s_i);$ $logit\left(\eta_i^{(2)}\right) = \beta_0^{(2)} + \beta_1^{(2)} AGE(s_i) + \beta_2^{(2)} SEX(s_i) + \beta_3^{(2)} INSURANCE(s_i) + \beta_4^{(2)} DISTANCEKm(s_i) + \beta_s f_s^i(s_i) + v(s_i),$	2849.2
Full model with two separate spatial structures.	$log\left(\eta_i^{(1)}\right) = \beta_0^{(1)} + \beta_1^{(1)} IB(s_i) + \beta_2^{(1)} UNDER20(s_i) + \beta_3^{(1)} P.FEMALE(s_i) + \beta_4^{(1)} SES(s_i) + f_s^j(s_i) + u(s_i);$ $logit\left(\eta_i^{(2)}\right) = \beta_0^{(2)} + \beta_1^{(2)} AGE(s_i) + \beta_2^{(2)} SEX(s_i) + \beta_3^{(2)} INSURANCE(s_i) + \beta_4^{(2)} DISTANCEKm(s_i) + g_s^{(2)}(s_i) + v(s_i)$	2762.9

Table S4a. Incidence Rate Ratio (IRR), Odds Ratios (OR) and 95% credible intervals of covariates (fixed effects) for the model without the SES covariates.

Fixed Effects	Estimates and 95% Credible intervals		
Overall Cases (Pattern)			
	IRR	2.5%	97.5%
Intercept (Pattern)	0.60	0.46	0.78
Proportion of Female Cases	1.04	1..00	1.09
Proportion of <20 years old	1.04	0.99	1.09
Breteau Index (Medium)	1.05	0.83	1.33
Severity (Marks)			
	OR	2.5%	97.5%
Sex (Female)	0.88	0.68	1.14
Age			
15-34 Years	0.97	0.69	1.38
35-54 Years	1.07	0.73	1.58
>55 Years	1.52	0.99	2.33
Contributory Insurance	1.00	0.77	1.31

Proportion of woman and cases under 20 years old indicate 10% increase. Breteau Index reference group = Low; Age group reference= <15 years of age; Sex reference= Male; Health Insurance reference is Subsidized scheme.

Table S4b. Posterior mean and 95% credible intervals of hyperparameters for the model without the SES covariates.

Parameter	mean	SD	2.5%	97.5%	mode
Precision for $f_s^j(s_i) + u(s_i)$	1.39	0.21	1.03	1.84	1.35
Phi for $u(s_i)$	0.41	0.14	0.17	0.69	0.39
β Coefficient for Severity for $g_s^j(s_{ij})$	-0.31	0.11	-0.53	-0.09	-0.32

Table S5a. Incidence Rate Ratio (IRR), Odds Ratios (OR) and 95% credible intervals of covariates (fixed effects) for the model with two separate spatial structures.

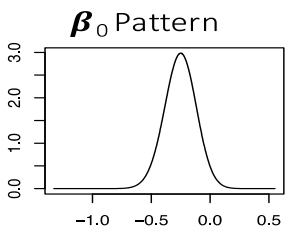
Fixed Effects	Posterior mean and 95% Credible intervals		
Overall Cases (Pattern)			
	IRR	2.5%	97.5%
Intercept (Pattern)	0.78	0.60	1.01
Proportion of Female cases	1.05	1.01	1.09
Proportion of cases <20 years old	1.06	1.01	1.11
Socio Economic Status (SES)			
Low SES	0.43	0.33	0.56
High SES	0.77	0.55	1.07
Breteau Index (Medium)	1.12	0.89	1.40
Severity (Marks)			
	OR	2.5%	97.5%
Sex (Female)	0.92	0.71	1.19
Age			
15-34 Years	1.05	0.74	1.5
35-54 Years	1.13	0.76	1.69
>55 Years	1.53	0.98	2.37
Contributory Insurance	0.89	0.67	1.17

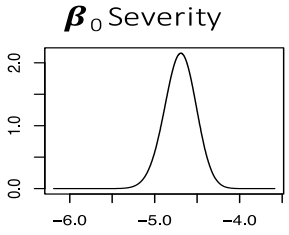
Proportion of woman and cases under 20 years old indicate 10% increase. SES reference group = Medium SES Level; Breteau Index reference group = Low; Age group reference= <15 years of age; Sex reference= Male; Health Insurance reference is Subsidized scheme

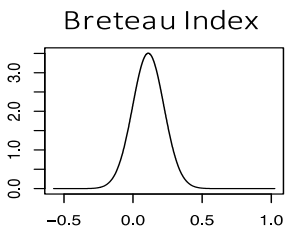
Table S5b. Posterior mean and 95% credible intervals of hyperparameters for the model with two separate spatial structures

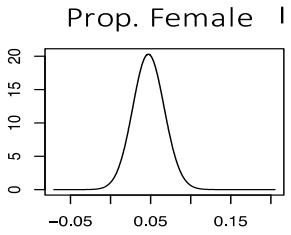
Parameter	mean	SD	2.5%	97.5%	mode
Precision for $f_s^j(s_i) + u(s_i)$	1.69	0.26	1.24	2.26	1.63
Phi for $u(s_i)$	0.38	0.14	0.13	0.68	0.35
Precision for $g_s^{(2)}(s_i) + v(s_i)$	2.02	0.59	1.13	3.43	1.75
Phi for $v(s_i)$	0.09	0.09	0.00	0.35	0.01

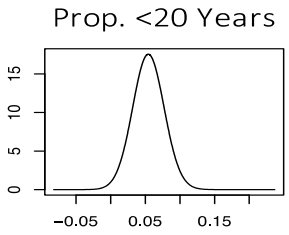
1.3 Posterior density of fixed and random effects of the final joint model using a single spatial component.

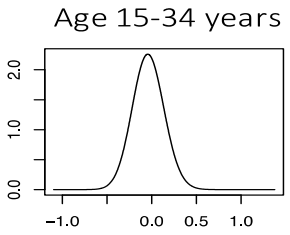


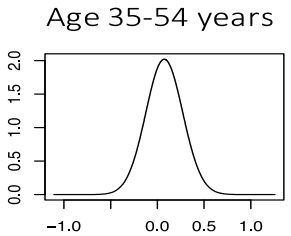


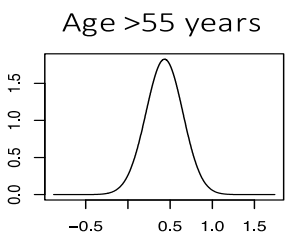


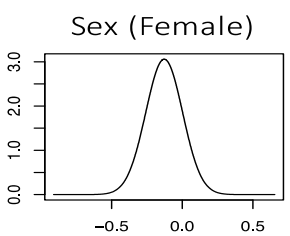


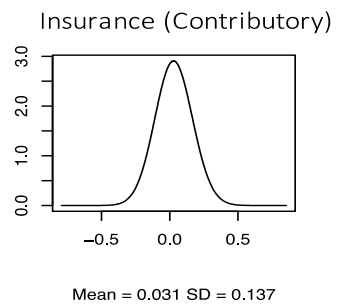


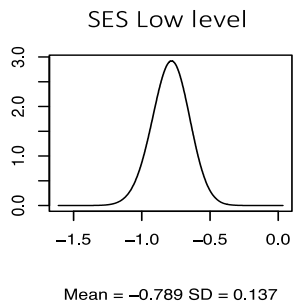


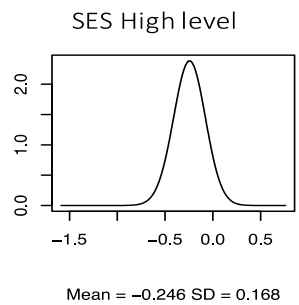


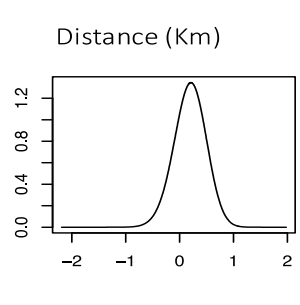


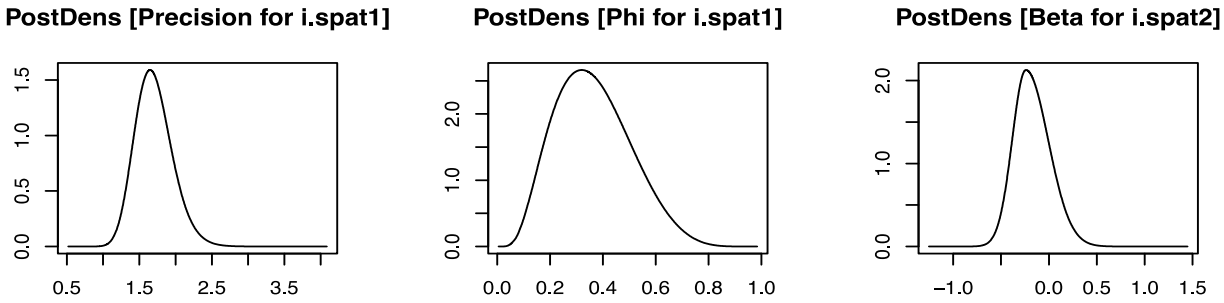












i.spat1= Spatial Effect for overall cases; i.spat2= Spatial effect for severe cases

References

1. Diggle PJ, Moraga P, Rowlingson B, Taylor BM. Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm. *Statistical Science* 2013;**28**(4):542-563.

2. Illian JB, Martino S, Sorbye S, Gallego-Fernandez JB, Zunzunegui M, Esquivias MP, Travis JMJ. Fitting complex ecological point process models with Integrated Nested Laplace Approximation. *Methods in Ecology and Evolution* 2013;**4**:305-315.

3. Simpson D, Illian JB, Lindgren F, Sorbye S, Rue H. Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *arXiv* 2017:1-26.

4. Illian JB, Sørbye SH, Rue H. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics* 2012:1499-1530.

5. Pinto Junior JA, Gamerman D, Paez MS, Alves Regina HF. Point pattern analysis with spatially varying covariate effects, applied to the study of cerebrovascular deaths. *Statistics in Medicine* 2015;**34**:1214-1226.

6. Waller LA, Gotway CA. Analyzing Public Health Data. . In: Shewhart WA, Wilks SS, eds. *Applied Spatial Statistics for Public Health Data*. Vol. 368 John Wiley & Sons, 2004;7-37.

7. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991;**43**(1):1-20.

8. Blangiardo M, Cameletti M. Spatial modeling. *Spatial and Spatio-temporal Bayesian Models with R-INLA* John Wiley & Sons, Ltd, 2015;173-234.

9. Congdon P. Models for spatial outcomes and geographical association. *Applied Bayesian Modelling* John Wiley & Sons, Ltd, 2014;312-363.

Chapter 5: **Spatiotemporal Distribution and Socioeconomic Disparities in Dengue, Chikungunya and Zika in Latin America (Manuscript 2).**

5.1 Preface Manuscript 2

This manuscript addresses the burden of arboviral diseases in Brazil and Colombia, using Bayesian Spatiotemporal analysis with hierarchical mixed models to assess the nature, pattern and magnitude of disparities on disease burden in two cities between 2007-2017. Although literature on health inequalities on dengue do exists, this manuscript includes the analysis of chikungunya and Zika. Likewise, the current literature on the subject do not offer measures of inequality or do not integrate time into the spatial analysis or used simultaneously a spatiotemporal model to estimate inequality indexes, as we do in this manuscript.

Using more than 300,000 cases from surveillance data, we documented quantitatively the presence and consistent concentration of arboviral diseases among low socioeconomic settings over time and changes in the disparity associated to the presence of outbreak or during the introduction of new diseases. We also identify the presence of heterogeneity and discussed and addressed challenges presented by traditional methods for estimating inequalities, such as the case of the concentration index.

We developed functions that facilitate the estimation of inequalities from Bayesian Hierarchical models that used structured random effects to account for the spatiotemporal distribution of endemo-epidemic conditions such arboviruses. We obtained one of the firsts socioeconomic-specific attributable estimates of disease distribution in both study sites. Furthermore, the methods that we propose and the use of the developed functions to extract inequality indexes could be used in both communicable and non-communicable diseases, contributing substantively to the current body of literature on health inequalities.

This manuscript is currently under review at the Tropical Medicine and International Health journal manuscript ID TMIH-D-20-00380.

5.2 Manuscript 2

Tittle: Spatiotemporal Distribution and Socioeconomic Disparities on Dengue, Chikungunya and Zika in Latin America from 2007 to 2017

Authors: Mabel Carabali[1], Sam Harper[1], Antonio S. Lima-Neto[2,3], Geziel dos Santos de Sousa[3], Andrea Caprara[4], Berta Nelly Restrepo[5], Jay S. Kaufman[1].

Affiliations:

[1]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada;

[2]Fortaleza's Secretary of Health, Fortaleza, Brazil;

[3]University of Fortaleza, Fortaleza, Brazil;

[4]Universidad Estadual de Ceara, Fortaleza, Brazil

[5]Instituto Colombiano de Medicina Tropical, Universidad CES, Sabaneta, Antioquia, Colombia

Corresponding author: Mabel Carabali (mabel.carabali@mail.mcgill.ca)

McGill University, Purvis Hall, 1020 Pine Avenue West, Montreal, Quebec, Canada (H3A 1A2).

Tel.: 514-398-6258; Fax: 514-398-4503

Highlights:

- Presence, pattern and magnitude of SES disparities on arboviruses in Latin America
- Relative and Absolute Concentration inequality indexes using spatiotemporal models
- Non-monotonic association between disease rates and socioeconomic distribution
- Changes in the inequalities during outbreaks or introduction of new arboviruses

ABSTRACT:

Objective: To assess the presence, pattern and magnitude of socioeconomic (SES) inequalities on arboviruses in Latin America, accounting for their spatiotemporal distribution.

Methods: Using longitudinal surveillance data from Fortaleza, Brazil and Medellin, Colombia between 2007-2017; we fit Bayesian hierarchical models to estimate: Relative Concentration Index (RCI) and Absolute Concentration Index (ACI) of inequality; Temporal trends in RCIs; and SES-specific estimates of disease distribution.

Results: We observed greater concentration of dengue among the lower-SES residents of both cities: RCI= -0.12 (95% CI= -0.13, -0.10) in Fortaleza and RCI= -0.04 (95% CI= -0.05, -0.03) in Medellin. The magnitude of inequality varied over time across sites but was larger during outbreaks. We identified a non-monotonic association between chikungunya rates and income distribution that changed over time. The SES-specific model showed increased disease rates at household incomes below US$400 in Brazil and Low-SES levels in Colombia.

Conclusions: We provide robust quantitative estimates of the socioeconomic inequalities on arboviruses in Latin America. Our findings could inform policy making by identifying spatial hotspots for arboviruses and targeting strategies to decrease disparities at the local level.

Key words: Health Inequalities, Concentration Index, Social Determinants, Latin America, Arboviruses, Bayesian Hierarchical models

INTRODUCTION

The dengue, Zika, and chikungunya arboviruses are important public health problems worldwide[1,2]. However, these viruses also show geographic and social differences in their distribution of morbidity and mortality burden[2-4]. Worldwide, the majority of cases are reported in poorer areas, areas where access to water is limited, areas of high concentration of people, and regions where the environmental conditions favor the presence of *Aedes* mosquitoes[2-4]. In addition, mortality rates and severe outcomes are higher among people at the bottom of the socioeconomic distribution[5]. For instance, 80% of dengue and 70% of chikungunya deaths in Colombia were cases from the government subsidized health program (i.e., low socioeconomic position)[6,7]. In Brazil, there is an increased risk of dengue death ranging from 40-70% among people with less than four years of schooling[8,9]. A reduction of dengue death of 10% has also been observed among people with a median monthly income over US$350 (roughly US$100 above the minimum wage)[8]. In general, people in low socioeconomic position are contributing differentially to the morbidity and mortality burden of vector borne diseases[5,7-10].

Income, health care, and education are structural social determinants of health inequality associated with infectious diseases[10,11]. However, in epidemiological studies on vector borne diseases, social determinants are usually used as covariates that are used to adjust for possible confounding without further analysis[10]. The differential distribution of dengue, chikungunya and Zika within and across neighborhoods, ethnic groups, and different levels of education, suggests that such social determinants are playing a role in the presence and expansion of these diseases in the Americas[4]. Yet there is no comprehensive study of the effects of these social determinants on the observed disparities among these arboviruses[4,12-14].

Therefore, in order to assess the presence of socioeconomic inequalities of dengue, chikungunya and Zika, we estimated the relative and absolute concentration index of inequality, accounting for the spatiotemporal distribution at the neighborhood level in Brazil and Colombia from 2007 to 2017.

METHODS

We used longitudinal surveillance data to estimate the overall arboviral and disease-specific socioeconomic inequalities and its distribution overtime.

Study Sites: Brazil and Colombia shoulder 60% of the burden of arboviral diseases in Latin America[1,2,15]. This study was therefore conducted with data from Fortaleza, Brazil and Medellin, Colombia. Municipalities were selected based on the disease burden, knowledge of the context, presence of functioning surveillance system, data availability, and expressed interest of the local health agencies on the study.

Fortaleza is the capital of the fifth biggest state of Brazil and has a population of 2.5 million inhabitants[16]. The Northeastern region of Brazil experiences 29% of Zika cases and Fortaleza has consistently been one of the top 50 municipalities (out of more than 5,000) reporting arboviruses over the last 10 years[4,17-19]. The city has six districts and 119 urban neighborhoods distributed in 315 km^2. Fortaleza's altitude is 21 meters above sea level, the average temperature is 27^oC and it has one rainy season from January to May. Brazil has Universal Health Coverage denominated Unified Health System, which is expected to cover the health needs of the entire population.

Medellin is the second largest city in Colombia with 2.6 million inhabitants[20]. Dengue incidence has been around 160-750 cases per 100,000 inhabitants during the last 10 years and has been consistently included on the top five dengue-reporting cities since 1999[5,6]. Medellin has 21 districts (16 urban, 5 rural); 249 urban neighborhoods, and 20 institutional units, distributed in 382 km^2 (110 urban). Medellin's altitude ranges from 1,460 to 3,200 meters, the average temperature 24^oC, and it has two rainy seasons, in April and October. Fifty percent of the city has low socioeconomic status (SES) and 98% of the city has access to potable water. The distribution of health coverage of the population is 70% contributory (employees or self-employees who pay for health insurance), 25% government subsidized, and 4% uninsured[20].

Data Sources: Notification of dengue, chikungunya and Zika is mandatory in all study sites through a passive surveillance system[21-23]. These arboviruses are individually registered weekly in the national surveillance system (SINAN-Brazil and SIVIGILA-Colombia) and our sample included all notified cases in the study sites between 2007 and 2017. Supplementary aggregated information about socioeconomic factors at the neighborhood level was obtained using National Census data and local quality of life and basic needs surveys for socioeconomic data[19,24].

Outcomes: We studied dengue, chikungunya and Zika cases registered in the surveillance system that were clinically and laboratory confirmed[22,25]. Given the possibility of misclassification of chikungunya and Zika with dengue at early stages of their introduction, we grouped these conditions into "all arboviruses", and performed the analysis using the combined group as well as each specific condition as dependent variables in our analysis.

Socioeconomic measures: In Brazil, social stratification is measured according to the median per capita monthly income and was stratified into three categories: Low (<R$290 ≈ US$70), Medium (R$ 291- R$ 1019 ≈ US$71 - US$244), and High (> R$1,020 ≈ >US$245)[16,19]. In Colombia, SES is estimated using an administrative summary measure (range 1–6) with 1 indicating the lowest SES and 6 indicating highest SES levels[26]. This measure is constructed according to assets and characteristics of the household, such as construction material, presence of electricity, potable water, etc. Each household has a designated SES level provided by the municipality and each neighborhood in turn possess an SES designation as well, according to a weighted mode of the household's SES within each block and neighborhood; classified as: 1(Very Low); 2(Low); 3(Low-Medium); 4(Medium); 5(Medium-High); and 6(High)[26]. For the Brazilian analysis we used the continuous measure of median monthly neighborhood income in USD and for the Colombian analyses we used the household or neighborhood's government-assigned SES index level described above.

Covariates: We included the proportion of female cases and proportion of cases by age group (<5 years, 5-9 years, 10-20 years, 20-50 years, and >50 years) reported by each disease, and at the

neighborhood level the proportion of people with secondary education, the proportion of households with adequate supply of potable water, sanitation, number of people per household, and number of health care centers per neighborhood according to the availability in each data set (Supplementary Material).

Statistical analysis

The analysis was conducted in three main steps, estimating: 1) the spatiotemporal adjusted incidence rates using notified arboviral diseases per neighborhood[27]; 2) the Relative Concentration Index of inequality (RCI) and the Absolute Concentration Index (ACI) using the spatiotemporal adjusted incidence rates[28-30]; and 3) estimating the overall and arbovirus-specific RCI trend over the time.

Spatiotemporal adjusted incidence rate: We used Bayesian hierarchical Poisson models with random effects for time and neighborhood, modeling the count of cases separately for each city as a function of the covariates, specified as:

$$log(Y_{ij}) = \log(E_{ij}) + \beta_0 + x_{ij}^T\beta_j + s_i + t_j, \text{ equation 1}$$

Where $log(Y_{ij})$ is the log incidence rate of the arbovirus in neighborhood i at time j, $\log(E_{ij})$ is the offset determined by the log of the mid-year neighborhood-specific population, x_{ij} is a vector of the neighborhood-level covariates listed above with its corresponding parameter β_j. The parameters s_i, t_j denote the spatially and temporally structured random effects, respectively. The spatial effect was modelled assuming a Besag-York-Mollie specification[31], which includes a structured component using an intrinsic conditional autoregressive structure (iCAR) and an unstructured component modelled as independent and identically distributed (i.i.d). The temporal component was modelled dynamically using a scaled random walk of first order (RW1) structure for the month[32,33]. The posterior distribution of the fixed-effects, random-effects and their 95% Credible Intervals (95% Cr.I) were obtained from a model fitted using the default non-informative priors in R-INLA[32,33].

Concentration index of inequality: The RCI is a measure of relative inequality that allows the identification of differential burden of diseases among the population across the socio-economic distribution[28-30]. The RCI estimation was given by:

$$\text{RCI} = \frac{2cov(y_{ij}, R_{ij})}{\mu} = \frac{2}{n^2 \mu_h} \sum_{i=1}^{n} y_{ij} R_{ij}, \text{ equation 2}$$

where y_{ij} is the health outcome variable (disease rates in neighborhood i at time j), R_{ij} the rank of socioeconomic measure for neighborhood i at time j, and μ is the mean of the spatiotemporally adjusted disease rate. We also estimated the absolute concentration index (ACI), as the product of the RCI and the mean disease rate for each outcome $ACI = RCI * \mu$. The RCI could be understood as a weighted mean of morbidity shares, with the weights depending on the fractional socioeconomic measure rank[28,29]. The RCI equals zero in absence of inequality, is negative when the disease rate is more concentrated among the poor, and positive when it is the opposite[30]. To obtain aggregated, diseases-specific and year-specific RCIs, and their point-wise estimates, we applied equation 2 to the rates obtained from the model in equation 1. For the adjusted estimates, the model fitted to obtain the rates (equation 1), did not include the socioeconomic covariate as an explanatory variable[30]. Estimates were obtained using modified functions from the 'decomp' R-packages for health inequalities[34] in RStudio (R version 3.6.1, R Core Team; Vienna, Austria; 2019) (Supplementary Material).

Sensitivity Analysis: Given that the RCI is a relative measure that depends on a rank, and because for each disease and each city the starting point of the socioeconomic rank changed by time, we conducted stratified analysis by year. Given that each neighborhood had different baseline disease rates and different baseline socioeconomic distributions, to generate socioeconomic-specific estimates while simultaneously adjusting for space and time, we fit spatiotemporal adjusted disease models including the socioeconomic variables as unstructured random effects. Specifically, we added to "equation 1", the income or SES measure, and model it using an independent and identically distributed random effect (i.i.d)[32]. To test the strength of our analysis with other available socioeconomic measures, we estimated the RCI using the Human

Development Index in Fortaleza and an extrapolated proxy measure of the mean income per neighborhood obtained from the quality of life survey from Medellin[35]. Models including entomological information and environmental covariates were fitted when such data was available (Supplementary Material).

Ethics: This study analyzed secondary data without human samples analyses, therefore, it did not require consent to participation. The protocol was reviewed and approved by McGill's Institutional Review Board (Study No. A02-E05-18A) and by the ethics committee of the Brazilian Ministry of Health (Code: 2.624.599) and Secretary of Health of Medellin, Colombia.

RESULTS

There were 281,426 arboviral cases notified in Fortaleza, and 40,887 in Medellin. In Fortaleza, 80.5% of notified cases were from the public health care system and in Medellin 76.6% of the cases belonged to the contributory insurance scheme. The socioeconomic distribution of cases was as follows: In Fortaleza, overall median household income was US$232; Inter Quartile Range (IQR) = $168 - $268. The median household income for dengue cases was US$197 (IQR= $146 - $253); for chikungunya US$253 (IQR=$237 - $326) without cases in households with income below US$212; and for Zika US$237 (IQR= $237 - $281). In Medellin, the overall median SES index was 3 (IQR=2 – 3), with arboviruses at all SES levels during the study period (Table 5.1).

Arboviral distribution

The overall median crude rate of notified arboviral diseases from 2007 to 2017 in Fortaleza was three cases per 10,000 inhabitants per neighborhood per month (IQR=1.2- 10.2). Compared to the overall disease-specific and spatiotemporally adjusted rates in the city during the study period, the highest dengue rate over time was observed during 2012 (IRR=29.1; 95% Cr.I= 28.6, 29.6). Chikungunya was first reported at the end of 2014 with the highest rate occurring in 2017 (IRR= 65; 95% Cr.I = 55.8, 76.9). Zika was first reported in 2015 and the highest rate was observed in 2016 (IRR 3.7; 95% Cr.I= 3.1, 4.6) (Figure 1). Dengue and chikungunya cases were concentrated slightly in the Northwestern neighborhoods, while Zika showed a similar spatial distribution across the entire city (Figure 5.2).

Table 5.1 Descriptive characteristics of notified arboviral diseases in Fortaleza, Brazil and Medellin, Colombia Between 2007- 2017.

	Fortaleza (Brazil)			Medellin (Colombia)		
Variable	Dengue	Chikungunya	Zika	Dengue	Chikungunya	Zika
Total cases	199,911	79,856	1,659	39,509	722	656
Cases [a] (IQR)	5 (2, 15)	6 (2, 27)	2 (1, 3)	8 (3, 17)	1 (1, 3)	1 (1, 2)
Hospitalization	4,671 (7.6%)	662 (1.8%)	-	8,535 (21.6%)	86 (11.9%)	174 (26.5%)
Deaths	219 (0.1%)	168 (0.2%)	0 (0)	6 (0.02%)	1 (0.1%)	1 (0.2%)
Age, Median Years (IQR)	25 (15, 39)	38 (24, 53)	29 (20, 43)	28 (16, 45)	34 (24, 49)	29 (21, 42)
Sex: Female	110,037 (55.0%)	49,188 (61.6%)	1,110 (66.9%)	18,973 (48.0%)	432 (59.8%)	402 (61.3%)
Sex: Male	89,836 (44.9%)	30,629 (38.4%)	548 (33.0%)	20,536 (52.0%)	290 (40.2%)	254 (38.7%)
Not disclosed	37 (0.02%)	37 (0.05%)	1 (0.06%)			
SES measure [b], median (IQR)	197 (146, 253)	253 (237, 326)	237 (237, 281)	3 (2, 3)	4 (3, 5)	3 (2, 5)
Year of Report						
2007	10,398 (5.2%)	-	-	-	-	-
2008	31,521 (15.8%)	-	-	576 (1.5%)	-	-
2009	3,450 (1.7%)	-	-	622 (1.6%)	-	-
2010	3,703 (1.9%)	-	-	13,064 (33.1%)	-	-
2011	34,659 (17.3%)	-	-	533 (1.3%)	-	-
2012	39,100 (19.6%)	-	-	473 (1.2%)	-	-
2013	8,830 (4.4%)	-	-	1,630 (4.1%)	-	-
2014	5,150 (2.6%)	3 (<0.01%)	-	2,859 (7.2%)	51 (7.0%)	-
2015	26,850 (13.4%)	4 (0.01%)	15 (0.9%)	3,555 (9.0%)	474 (65.5%)	51 (7.8%)
2016	21,840 (10.9%)	17,620 (22.1%)	1,343 (81.0%)	16,089 (40.7%)	164 (22.7%)	578 (88.1%)
2017	14,410 (7.2%)	62,229 (77.9%)	301 (18.1%)	108 (0.3%)	33 (4.6%)	27 (4.1%)

[a] Cases per month and neighborhood. [b] Socioeconomic measure: $USD for Fortaleza, Brazil and SES Index (range 1-6; 1 Lowest/Poorest and 6 Highest/Richer) for Medellin, Colombia.

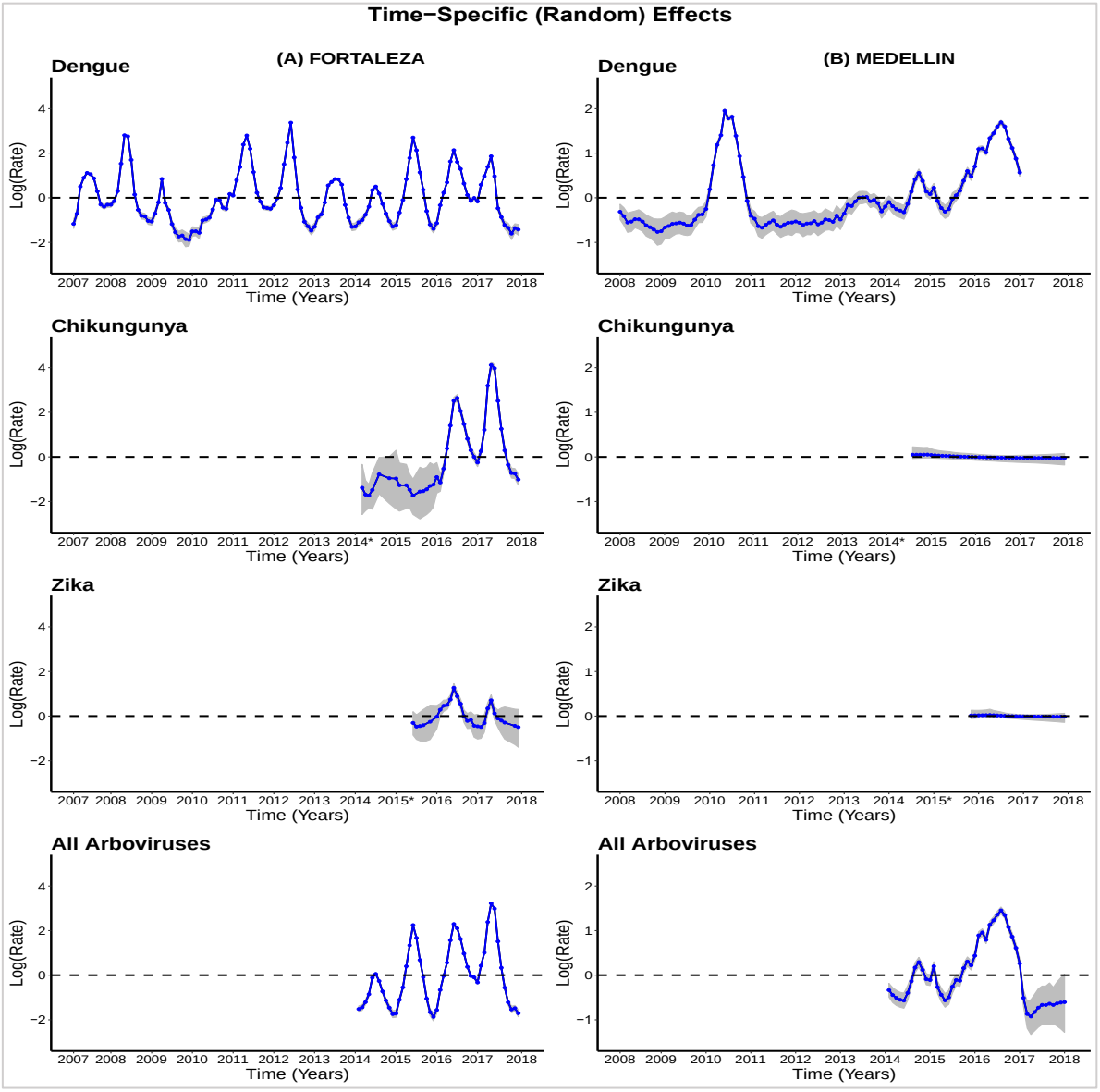


Figure 5-1 Temporal (Month-specific) random effects.

Disease-specific and aggregated arbovirus distribution in Fortaleza, Brazil (2007-2017) and Medellin, Colombia (2008-2017). Posterior mean of the time-specific random effects (log scale) and their 95% Credible Intervals.

In Medellin, the median crude rate of arboviral diseases between 2008-2017 was one case per 10,000 inhabitants per neighborhood per month (IQR=1.2 - 5.7). Compared to the adjusted overall disease rates during the entire study period, the highest dengue rates over time were observed during 2010 (IRR =7.1; 95% Cr.I= 6.8, 7.4) and 2016 (IRR=5.5; 95% Cr.I =5.2, 5.7). Chikungunya was first reported in 2014 and Zika in 2015, both diseases had relatively small rates throughout the

study period, with small peaks at the end of 2014 for chikungunya and the beginning of 2016 for Zika (Figure 1). Dengue presence was observed in the entire city with the majority of cases in the Northeastern neighborhoods, while chikungunya and Zika presented a smaller number of cases evenly distributed across the city (**Figure 5.2**).

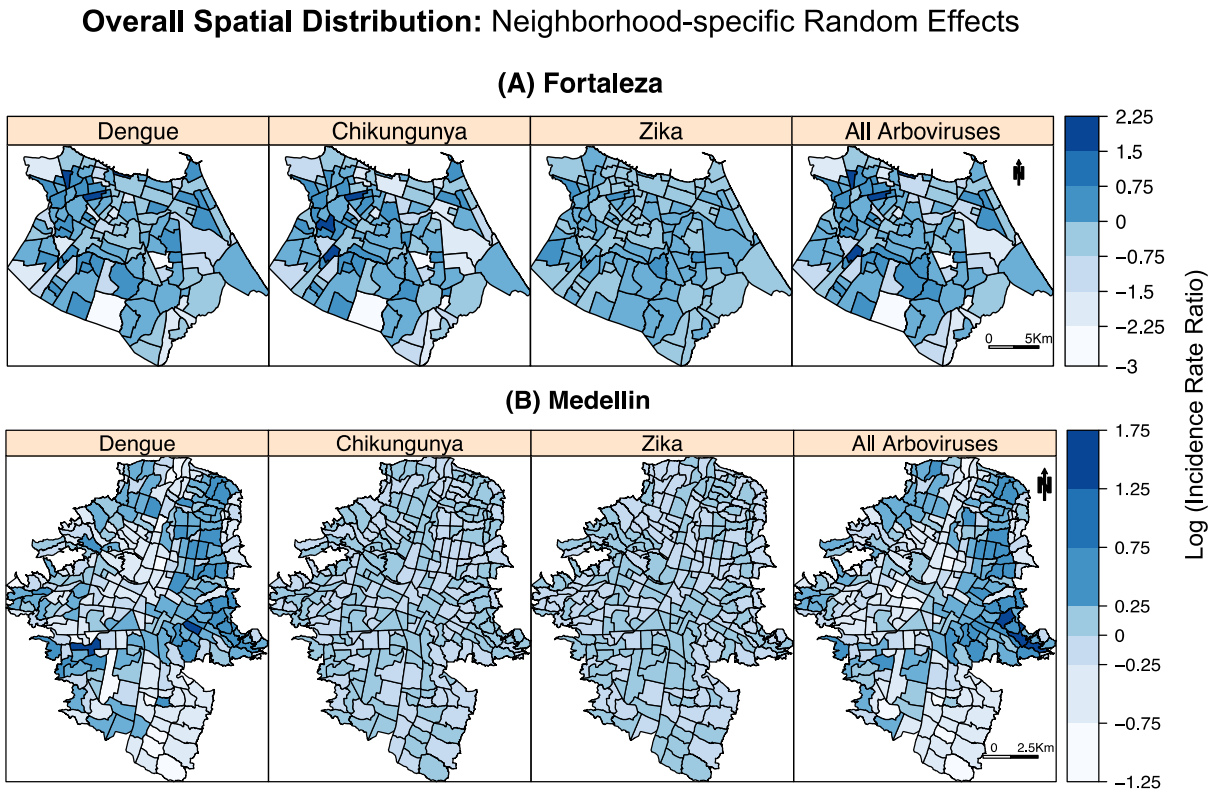


Figure 5-2 Spatial (Neighborhood-specific) random effects.
Disease-specific and aggregated arbovirus distribution. (A)Cumulative neighborhood-specific log(Incidence Rate Ratios), compared to the spatiotemporally adjusted overall rate in Fortaleza, Brazil (2007-2017). (B) Cumulative neighborhood-specific log(Incidence Rate Ratio), compared to the spatiotemporally adjusted overall rate in Medellin, Colombia (2008-2017).

Relative and Absolute Concentration Index and Trend Over Time

In Fortaleza, the adjusted overall RCI for dengue was -0.12 (95% CI= -0.14, -0.10); for chikungunya 0.03 (95% CI= -0.01, 0.07), and for Zika -0.07 (95% CI= -0.10, -0.04) (**Table 5.2**; **Figure 5.3**). Dengue adjusted RCIs and credible intervals were consistently less than zero, indicating a concentration of dengue cases among people living in low income neighborhoods. Chikungunya's RCIs ranged from -0.09 to 0.002 and Zika's RCIs ranged from -0.04 to -0.07, with credible intervals including equality in 2015 and 2017 (Figure 4). The ACI for dengue was -0.22, ranging from -0.04 to -0.36, indicating

greater absolute inequalities on dengue rates, mostly during epidemic periods. The overall ACI for chikungunya was 0.10, however, the stratified analysis by year showed ACIs ranging from -0.37 to less than 0.01 with the largest absolute inequality observed for chikungunya in 2017. The ACI for Zika was -0.02 indicating a small absolute inequality over time (Table 5.3).

Table 5.2 Overall Crude and Adjusted Relative Concentration Index (RCI) and the Absolute Concentration Index (ACI) of Inequality for Arboviral Diseases in Fortaleza, Brazil (2007-2017) and Medellin, Colombia (2008-2017).

City/Disease	Crude RCI		Adjusted RCI		Absolute CI
Fortaleza	RCI	95% CI	RCI	95% CI	ACI
All arboviruses	0.091	(0.064, 0.119)	0.105	(0.083, 0.127)	0.334
Dengue	-0.116	(-0.138, -0.095)	-0.117	(-0.135, -0.099)	-0.221
Chikungunya	0.022	(-0.020, 0.063)	0.028	(-0.010, 0.066)	0.101
Zika	-0.097	(-0.152, -0.043)	-0.069	(-0.103, -0.036)	-0.018
Medellin	RCI	95% CI	RCI	95% CI	ACI
All arboviruses	-0.083	(-0.10, -0.06)	-0.072	(-0.085, -0.059)	-0.033
Dengue	-0.049	(-0.064, -0.034)	-0.040	(-0.051, -0.029)	-0.017
Chikungunya	-0.023	(-0.045, -0.001)	-0.005	(-0.016, 0.005)	-0.001
Zika	0.008	(-0.014, 0.031)	0.015	(0.003, 0.027)	0.003

The RCI and ACI for the aggregated arboviruses were estimated using only data from 2014-2017, which is the period where the four arboviral conditions were present simultaneously. The overall adjusted RCI in Fortaleza was 0.10 (95% CI=0.08, 0.13) and the overall ACI was 0.33, indicating a positive absolute inequality concentrating arboviral diseases among people in neighborhoods at the top of the socioeconomic distribution. However, the stratified analysis by year showed negative RCIs that ranged from -0.07 (95%CI = -0.10, -0.04); ACI= -0.03 in 2014 to -0.09 (95%CI= -0.13, -0.04); ACI = -0.48 in 2017, indicating a non-monotonic relationship with heterogeneity by year and confounding by disease, given the differential distribution of income among each diseases and year (Supplementary Material).

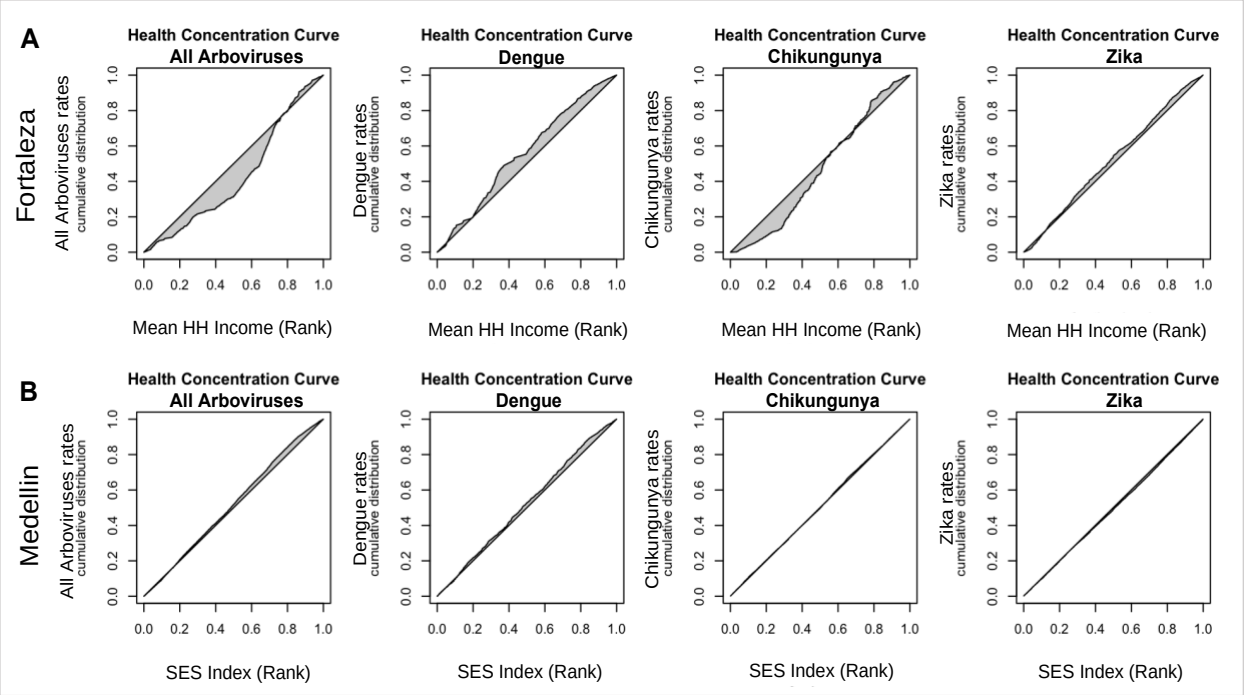


Figure 5-3 Overall Relative Concentration Index (RCI) for aggregated and disease-specific arbovirus distribution.
(A) Health concentration curves for Fortaleza, Brazil (2007-2017), using cumulative distribution of the disease-specific rates and the mean household (HH) income rank. (B) Health concentration curves for Medellin, Colombia (2008-2017), using cumulative distribution of the disease-specific rates and the Socioeconomic (SES) index rank.

In Medellin, the adjusted overall RCI for dengue was -0.04 (95% CI= -0.05, -0.03), for Chikungunya -0.005 (95% CI= -0.016, 0.005), and for Zika 0.015 (95% CI= 0.003, 0.027) (Table 5.2). Dengue RCIs ranged from -0.06 to -0.02, while chikungunya and Zika RCIs covered the line of equality throughout the study period (Figure 5.4). The overall ACI for dengue was -0.017, with the highest ACIs values being -0.027 and -0.037 during 2010 and 2016, respectively. Chikungunya and Zika ACIs indicated small absolute inequalities with diseases concentrating similarly across the spectrum of SES. The adjusted RCI for aggregated arboviruses in Medellin ranged from -0.07 (2015) to -0.10 (2017), and their highest ACI was -0.05 observed in 2016 a dengue and Zika epidemic year. The spatiotemporal adjusted model including random effects for socioeconomic measures showed increased arboviral rates at household's income below US$400 in Fortaleza and among people below SES level 4 in Medellin (Figure 5.5).

Table 5.3 Adjusted Relative Concentration Index (RCIs) and Absolute Concentration Index (ACIs) of Inequality for Arboviral Diseases per year in Fortaleza, Brazil (2007-2017) and Medellin, Colombia (2008-2017).

	Fortaleza			Medellin		
Disease/Year	Relative Concentration Index		Absolute Concentration Index	Relative Concentration Index		Absolute Concentration Index
Dengue	**RCI**	**95% CI**	**ACI**	**RCI**	**95% CI**	**ACI**
2007	-0.09	(-0.12, -0.06)	-0.11	-	-	-
2008	-0.08	(-0.13, -0.02)	-0.28	-0.03	(-0.06, -0.004)	-0.005
2009	-0.09	(-0.14, -0.05)	-0.05	-0.04	(-0.07, -0.01)	-0.006
2010	-0.12	(-0.15, -0.09)	-0.06	-0.04	(-0.06, -0.02)	-0.027
2011	-0.13	(-0.17, -0.08)	-0.36	-0.02	(-0.06, 0.01)	-0.003
2012	-0.10	(-0.15, -0.04)	-0.37	-0.06	(-0.08, -0.03)	-0.007
2013	-0.11	(-0.14, -0.08)	-0.09	-0.02	(-0.04, 0.002)	-0.004
2014	-0.10	(-0.13, -0.07)	-0.05	-0.06	(-0.08, -0.04)	-0.014
2015	-0.10	(-0.15, -0.05)	-0.25	-0.03	(-0.06, -0.01)	-0.009
2016	-0.11	(-0.14, -0.07)	-0.19	-0.05	(-0.07, -0.04)	-0.037
2017	-0.08	(-0.12, -0.03)	-0.11	-	-	-
Chikungunya	**RCI**	**95% CI**	**ACI**	**RCI**	**95% CI**	**ACI**
2014	0.002	(-0.14, 0.15)	<0.001	0.04	(-0.01, 0.09)	0.005
2015	-0.09	(-0.23, 0.06)	-0.01	0.003	(-0.02, 0.01)	<0.001
2016	-0.07	(-0.11, -0.02)	-0.12	-0.01	(-0.03, 0.01)	-0.001
2017	-0.07	(-0.13, -0.01)	-0.37	-0.01	(-0.03, 0.01)	-0.001
Zika	**RCI**	**95% CI**	**ACI**	**RCI**	**95% CI**	**ACI**
2015	-0.04	(-0.12, 0.05)	0.004	<-0.01	(-0.03, 0.03)	0.001
2016	-0.07	(-0.11, -0.03)	-0.02	0.02	(0.003, 0.03)	0.002
2017	-0.04	(-0.08, 0.01)	-0.01	0.02	(-0.002, 0.04)	0.002
All arboviruses	**RCI**	**95% CI**	**ACI**	**RCI**	**95% CI**	**ACI**
2014	-0.07	(-0.10, -0.05)	-0.03	-0.08	(-0.09, -0.06)	-0.018
2015	-0.08	(-0.13, -0.04)	-0.19	-0.07	(-0.09, -0.05)	-0.019
2016	-0.09	(-0.13, -0.06)	-0.27	-0.07	(-0.09, -0.06)	-0.052
2017	-0.09	(-0.13, -0.04)	-0.48	-0.10	(-0.15, -0.05)	-0.015

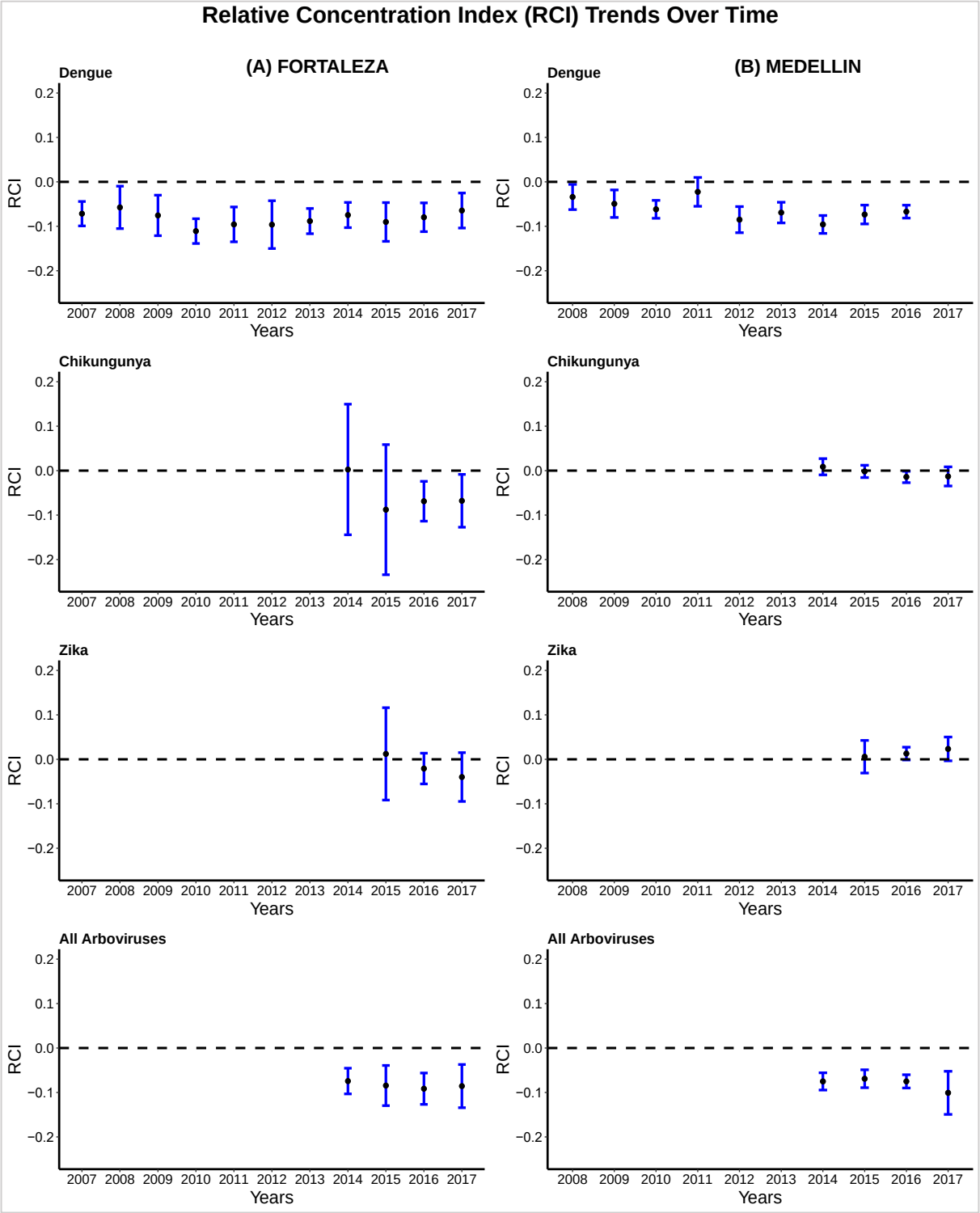


Figure 5-4 Relative Concentration Index (RCI) trend, overall and disease-specific arbovirus distribution in Fortaleza, Brazil (2007-2017) and Medellin, Colombia (2008-2017).
The dashed black line indicates the line of equality (i.e., zero) and the bars the year specific RCI estimates and their pointwise credible Intervals.

Given that income and the SES index could be correlated with the spatial structure, we fit models without the spatial random effect and results were also consistent, showing a concentration of arboviruses among people in the lower range of the socioeconomic distribution. Sensitivity analysis using Human development index (HDI) in Fortaleza and the proxy for monthly income as the socioeconomic measure in Medellin were consistent with the estimates obtained in the main analysis. Entomological information at the neighborhood level was available only for 2008-2013 in Medellin and environmental covariates for 2014-2017. However, the resulting RCIs and ACIs from these models were not substantially different from the results presented here (Supplementary Material).

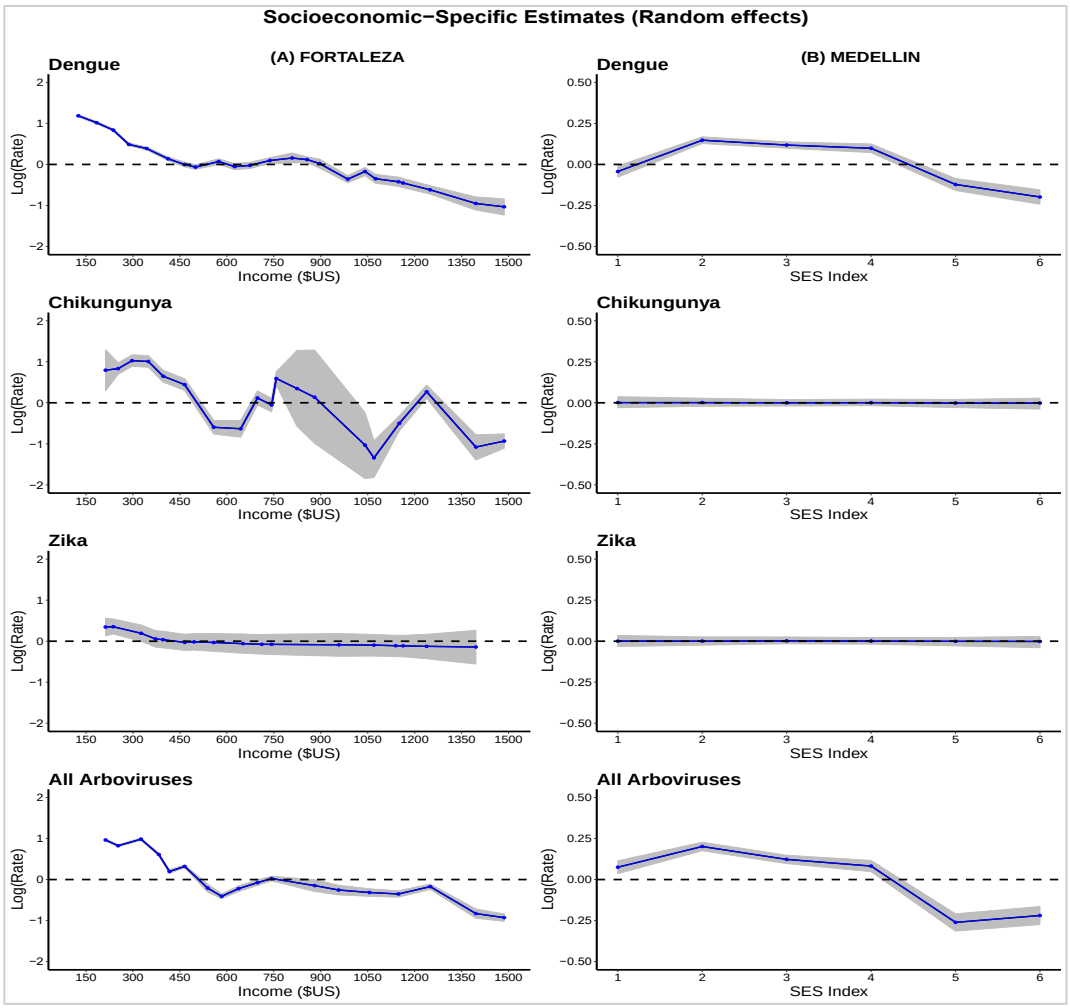


Figure 5-5 Socioeconomic-specific Random Effects for aggregated and each specific arbovirus distribution in Fortaleza, Brazil (2007-2017) and Medellin, Colombia (2008-2017).
Posterior mean of the Income/SES-specific random effect (log scale), and their respective 95% Credible Intervals for disease distribution, obtained from models adjusting by age, sex, and sanitation as fixed effects and structured random effects for time.

DISCUSSION

This surveillance-based study of arboviral diseases in Brazil and Colombia from 2007 to 2017, provides novel evidence about the presence of socioeconomic inequalities in the distribution of arboviruses in two Latin American cities. This study is consistent with previous observations of dengue as an endemic disease, with outbreaks occurring every 2-3 years in the two study settings[1,2,15,36]. Likewise, it was also possible to identify the introduction of two new arboviruses, chikungunya and Zika, and their initial notification in 2014 and 2015, respectively[15,22]. The highest rate of chikungunya occurred in 2014 in Colombia and during 2017 in Fortaleza, while the highest rates of Zika occurred in 2016 in both study sites. The spatial assessment showed the presence of all arboviral diseases in almost all neighborhoods in both study settings. However, it was possible to identify spatial clusters for dengue presence and a spatial dispersion of chikungunya and Zika. The occurrence of outbreaks of chikungunya and Zika in both study sites could explain the widespread presence of cases in each city[15,22,25,37,38]. However, the magnitude of chikungunya's outbreak in Fortaleza was larger compared to the outbreak of Zika in the same city[38] and to chikungunya and Zika in Medellin.

The adjusted and yearly-stratified RCI estimates for dengue consistently showed a greater concentration of the diseases among people at the lower end of the socioeconomic range in both cities. The cumulative RCI for the aggregated arboviruses and for chikungunya in Brazil were either positive or covered the line of equality. However, stratified analyses by year showed consistently negative RCIs, indicating concentration among people in the lower end of the socioeconomic range. This could be because the distribution of disease rates across the socioeconomic spectrum varied by year. Specifically, there were few cases of chikungunya in 2014 and 2015. During 2016 there were no cases of chikungunya or Zika at levels of mean household income below US$212 and during 2017 below US$237. Therefore, each disease had a different starting point for the socioeconomic rank per year, which is not accounted for while estimating a cumulative RCI measure. In addition, for chikungunya, we observed a non-monotonic association between disease rates and income distribution, with more cases around the second and fourth quartile of the income distribution. This highlights a limitation of the rank-based nature of the RCI

estimation[28-30], and the importance of evaluating heterogeneity using stratified analysis in situations with the possibility of non-monotonic associations[39], as in this instance. Also, reinforcing the importance of accounting for time when conducting inequality analysis on non-stationary conditions, such is the cases of endemo-epidemic communicable diseases.

We observed changes in the RCI trend by year that were consistent with changes in the temporal distribution of arboviruses in both study sites. Specifically, the RCIs for dengue moved towards the line of equality during epidemic years in Fortaleza while during the interepidemic years with low rates, the RCIs indicate concentration of diseases among people at the bottom of the socioeconomic ranking. In the case of chikungunya and Zika, which are considered newly introduced arboviruses, the RCIs and their point-wise credible intervals either covered the line of equality or indicated concentration of diseases among people at the upper range of the SES rank during their introduction. This could be explained by the fact that during uncontrolled outbreaks, the entire population is similarly at risk of being affected by these arboviruses, regardless their socioeconomic position[2,12,14].

The ACI allowed the identification of very small inequalities as in the case of Zika and large absolute inequalities as in the case of dengue or chikungunya, especially during outbreaks of endemic diseases. The magnitude of the absolute inequalities between Colombia and Brazil for dengue are different, possibly indicating larger inequalities in Brazil than in Colombia. However, the direction of the inequality in both study sites is similar, indicating lower rates of dengue among people in neighborhoods at the top of the socioeconomic distribution and larger absolute differences during epidemic years. Dengue is endemic in both sites and this study shows substantial disease concentration among limited resources and low-SES neighborhoods, emphasizing the association of socio-environmental factors in the presence of arboviruses[4,12-14]. Nonetheless, the different absolute changes for chikungunya and Zika in the two study sites could be explained by the differential magnitude of each outbreak in each site. While chikungunya represented a major public health issue in Fortaleza, with higher rates and presence of spatial clustering, the rates in Medellin were low and evenly distributed. In fact, Fortaleza presented one of the highest rates of

chikungunya in Brazil during 2017, with the highest reported mortality rate in the country[22,38,40]. Different from Medellin, where chikungunya rates were below the national average in Colombia[25]. Likewise, Zika's outbreak was of greater magnitude in Fortaleza than in Medellin[22,37]. However, despite the differential magnitude in terms of number of cases and rates, the estimated ACIs showed very small inequalities across the SES spectrum. Again, this stresses the epidemic character of the diseases and their time-dependent impact on the entire population.

Finally, the difference between the crude and adjusted estimates could indicate the presence of additional environmental and/or individual factors contributing to the inequality. Age and sanitation variables were associated with increased rates in Fortaleza, and number of people per household and sanitation variables in Medellin. These characteristics, which are considered social determinants of health, have been widely described as drivers of the presence of arboviral diseases in the two settings[2,4,10,12,13]. However, further analysis to estimate the specific contribution of each determinant on the socioeconomic inequality on vector-borne diseases would be informative.

Limitations:

Although this study presents novel information on the distribution of socioeconomic inequalities in arboviruses in Latin America, it is important to acknowledge the presence of some methodological limitations. Since this is a surveillance-based study, the interpretation should be conditional on subjects who were symptomatic and/or sought care and were reported to each national surveillance system. Underreporting is an important issue among surveillance systems in Latin America, and as an attempt to correct for the possibility of this bias, we conducted sensitivity analysis using informative priors with different values from the literature. The estimated parameters did not change substantially the magnitude of the RCI estimates, which was the focus of this paper. Since the notification is likely affected by the presence or magnitude of symptoms, the individual's capacity (physical or economic) to access health care, and the reporting by the healthcare providers (mostly from Fortaleza where underreporting is higher among private

institutions); conditioning on surveillance data would result in an underestimation of the rates, likely moving our estimates further towards the null.

Likewise, given the introduction of chikungunya and Zika in 2014 and 2015 respectively, there is a potential risk of misclassification of the outcome. Specifically, some early cases of chikungunya and Zika could have been misclassified as dengue cases. However, there is no available literature on the scope or magnitude of this potential bias in our study settings[41]. Therefore, as an attempt avoid over or underestimation of the inequality due to the misclassification, we grouped dengue, chikungunya and Zika cases and estimated the aggregated arboviral distributions, RCIs, ACIs and its trend overtime, and we consider this outcome as less sensitive to potential misclassification.

Entomological information used in the sensitivity analyses did not change the results, possibly because entomological surveys are conducted sporadically (e.g., four times per year in Colombia), in different seasons in the year and at different locations in each neighborhood. In addition, Fortaleza had different types of entomological measures and in Medellin the entomological indexes were consistently low throughout the study period. Nonetheless, since entomological aspects are spatial level characteristics, we expect to have captured some of these effects by adjusting for neighborhoods as structured spatial-random effects. Likewise, the effect of environmental covariates such as precipitation and temperature when available did not impact the inequalities estimates and any effect of such covariates is expected to be captured by the structured temporal-random effects.

Conclusion

Our study documents the presence, pattern and magnitude of socioeconomic disparities on arboviral diseases in two Latin American cities from 2007-2017. We documented the consistent concentration of arboviral diseases among poorer neighborhoods over time and changes in socioeconomic inequality associated to the presence of outbreaks or during the introduction of new diseases. Our study also highlights the need of careful estimation of inequalities, accounting for time, space and examining the presence of heterogeneity when using traditional methods. We

contribute to the literature by providing robust quantitative estimates of the socioeconomic disparities on arboviral diseases that to the best of our knowledge, have not been presented before. Given that disease control strategies in endemic countries are mainly informed by analysis of surveillance data, our results could be used for policy making to identify areas of constant presence of arboviruses and targeting strategies to decrease disparities at the local level.

Conflict of interest: None declared.

Funding: This study did not receive any specific funding.

Computing code: The R-script used to conduct the analysis in this manuscript is provided as supplementary material.

Data availability: Data used in this manuscript could be obtained by official requests to the public health offices/ local Ministry of Health from Fortaleza, Brazil and in Medellin, Colombia.

Acknowledgments: We acknowledge the collaboration of the Public Health and the Surveillance office in Fortaleza, Brazil and in Medellin, Colombia. Mabel Carabali holds a CIHR-Banting-Best Doctoral Award fellowship.

REFERENCES

1. World Health Organization (WHO). A Toolkit for National Dengue Burden Estimation. In: World Health Organization (WHO), ed. Geneva, 2018.
2. Mayer SV, Tesh RB, Vasilakis N. The emergence of arthropod-borne viral diseases: A global prospective on dengue, chikungunya and Zika fevers. *Acta Trop* 2017;**166**(Supplement C):155-163.
3. Liang G, Gao X, Gould EA. Factors responsible for the emergence of arboviruses; strategies, challenges and limitations for their control. *Emerging Microbes & Infections* 2015;**4**(1):e18 (5).
4. Caprara A, Lima JW, Marinho ACP, Calvasina PG, Landim LP, Sommerfeld J. Irregular water supply, household usage and dengue: a bio-social study in the Brazilian Northeast. *Cad Saude Publica* 2009;**25**:S125 - S136.
5. Instituto Nacional de Salud (INS). Informe Epidemiologico de Evento Dengue. In: Vigilancia y Análisis del Riesgo en Salud Pública., ed. Dengue. Vol. FOR-R02.4000-001. Bogota, Colombia: Instituto Nacional de Salud, 2016.
6. Villar LA, Rojas DP, Besada-Lombana S, Sarti E. Epidemiological trends of dengue disease in Colombia (2000–2011): a systematic review. *PLoS Negl Trop Dis* 2015;**9**.
7. Ardila Pinto F, Martínez S, Fuentes M, Borrero E. Análisis de las demoras en salud en personas que enfermaron de gravedad o fallecieron por dengue en cinco ciudades de Colombia. *Physis: Revista de Saúde Coletiva* 2015;**25**:571-592.
8. Paixão ES, Costa MdCN, Rodrigues LC, Rasella D, Cardim LL, Brasileiro AC, Teixeira MGLC. Trends and factors associated with dengue mortality and fatality in Brazil. *Revista da Sociedade Brasileira de Medicina Tropical* 2015;**48**:399-405.
9. Moraes GH, de Fátima Duarte E, Duarte EC. Determinants of Mortality from Severe Dengue in Brazil: A Population-Based Case-Control Study. *The American Journal of Tropical Medicine and Hygiene* 2013;**88**(4):670-676.
10. Carabali M, Hernandez L, Arauz M, Villar L, Ridde V. Why are people with dengue dying? A scoping review of determinants for dengue mortality. *BMC Infectious Diseases* 2015;**15**(1):301.
11. Comision of Social Determinants of Health-CSDH. Closing the gap in a generation : health equity through action on the social determinants of health. Final report of the commission on social determinants of health. Geneva, Switzerland: World Health Organization., 2008;256.

12. Mulligan K, Dixon J, Joanna Sinn C-L, Elliott SJ. Is dengue a disease of poverty? A systematic review. *Pathogens and Global Health* 2015;**109**(1):10-18.

13. Teixeira MdG, Barreto ML, Costa MdCN, Ferreira LDA, Vasconcelos PFC, Cairncross S. Dynamics of dengue virus circulation: a silent epidemic in a complex urban area. *Tropical Medicine & International Health* 2002;**7**(9):757-762.

14. Moloughney B. What can public health do to address inequities in infectious disease? *Canada Communicable Disease Report* 2016;**42**(S1):S1_14.

15. Rico-Mendoza A, Alexandra P-R, Chang A, Encinales L, Lynch R. Co-circulation of dengue, chikungunya, and Zika viruses in Colombia from 2008 to 2018. *Revista panamericana de salud publica = Pan American journal of public health* 2019;**43**:e49-e49.

16. Perfeitura de Fortaleza. A cidade. https://www.fortaleza.ce.gov.br/, 2017.

17. Ministério da Saúde, Secretaria de Atenção à Saúde. Protocolo de atenção à saúde e resposta à ocorrência de microcefalia relacionada à infecção pelo vírus Zika *Atenção à Saúde* Vol. CDU 616-022. Brasilia, Brazil: Ministério da Saúde, Secretaria de Atenção à Saúde, 2016;42.

18. Toan NT, Rossi S, Prisco G, Nante N, Viviani S. Dengue epidemiology in selected endemic countries: factors influencing expansion factors as estimates of underreporting. *Trop Med Int Health.* 2015;**20**(7):840-863.

19. IBGE, Instituto Brasileiro de Geografia e Estatística. Atlas do Censo Demografico Brasileiro 2010. . In: SIDRA, ed. Vol. Banco de Tabelas Estadisticas. Rio de Janeiro, Brazil: Ministerio do Planejamento, Orçamento e Gestao., 2010.

20. Departamento Nacional de Estadistica de Colombia (DANE). Estimaciones de población 1985 - 2005 y proyecciones de población 2005 - 2020 total municipal por área. *Departamento Nacional de Estadistica de Colombia* DANE, 2015.

21. Ministerio de la Protección Social. Decreto Numero 3518 de 2006. Creacion y reglamentacion del Sistema de Vigilancia en Salud Publica- SIVIGILA. In: Social. MdlP, ed. Vol. 3518/2006. Bogota, Colombia., 2006;1-17.

22. Governo do Estado do Ceara. Boletim Epidemiologico: Dengue, Chikungunya e Zika. In: Fortaleza RdS, ed. *Coordenadoria de Promoção e Proteção à Saúde* Fortaleza: Secretaria da Saúde do Estado do Ceará – Região de Saúde de Fortaleza, 2017;12.

23. Brasil, Ministerio da Saude. Portaria No 47, 3 de Maio de 2016. Define os parâmetros para monitoramento da regularidade na alimentação do Sistema de Informação de Agravos de Notificação (SINAN). In: Ministerio da Saude SdVeS, ed. Vol. Portaria No. 47, 2016. Brasilia, 2016.
24. Departamento Administrativo Nacional de Estadística (DANE). Estimaciones de población 1985–2005 y proyecciones de población 2005–2020 total municipal por área. . Vol. Vital Statistics. Bogota, Colombia: DANE, 2016.
25. Instituto Nacional de Salud (INS). Protocolo de Vigilancia en Salud Publica Enfermedad por Virus Chikungunya. In: Instituto Nacional de Salud (INS), ed. Vol. . Bogota, Colombia: VIGILANCIA Y ANÁLISIS DEL RIESGO EN SALUD PÚBLICA, 2017.
26. Departamento Administrativo Nacional de Estadistica (DANE). Metodologia de la Estratificacion Socioeconomica Urbana para Servicios Publicos Domicioliarios. In: Direccion de Geoestadistica, ed. Grupo de Estratificacion ed. Santa Fe de Bogota, Colombia.: DANE, 2015;96.
27. Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology* 2013;**7**(Supplement C):39-55.
28. Konings P, Harper S, Lynch J, Hosseinpoor AR, Berkvens D, Lorant V, Geckova A, Speybroeck N. Analysis of socioeconomic health inequalities using the concentration index. *International Journal of Public Health* 2010;**55**(1):71-74.
29. Wagstaff A. The concentration index of a binary outcome revisited. *Health Economics* 2011;**20**(10):1155-1160.
30. Erreygers G, Kessels R. Regression-Based Decompositions of Rank-Dependent Indicators of Socioeconomic Inequality of Health. In: Kessels R, ed. *Health and Inequality*. Research on Economic Inequality. Vol. 21 Emerald Group Publishing Limited, 2013;227-259.
31. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991;**43**(1):1-20.
32. Blangiardo M, Cameletti M. Bayesian regression and hierarchical models. *Spatial and Spatio-temporal Bayesian Models with R-INLA* John Wiley & Sons, Ltd, 2015;127-172.
33. Rue H, Riebler A, Sørbye SH, Illian JB, Simpson DP, Lindgren FK. Bayesian Computing with INLA: A Review. *Annual Review of Statistics and Its Application* 2017;**4**(1):395-421.

34. Konings P, Speybroeck N. decomp: Various functions to quantify and decompose health inequalities. 0.3 ed CRAN, 2012;This package contains functions to calculate the relative concentration index, plot Lorenz curves and decompose inequality using (generalized) linear models and survival models while taking survey design into account.

35. Alcaldia de Medellin. Encuesta Calidad de Vida 2004 - 2018. *Transparencia y acceso a la información pública*. Medellin, Colombia, 2019.;www.medellin.gov.co.

36. MacCormack-Gelles B, Lima Neto AS, Sousa GS, Nascimento OJ, Machado MMT, Wilson ME, Castro MC. Epidemiological characteristics and determinants of dengue transmission during epidemic and non-epidemic years in Fortaleza, Brazil: 2011-2015. *PLOS Neglected Tropical Diseases* 2018;**12**(12):e0006990.

37. Instituto Nacional de Salud (INS). Informe epidemiologico de evento enfermedad por virus Zika. In: Instituto Nacional de Salud (INS), ed. Vol. FOR-R02.4000-001. Bogota, Colombia: VIGILANCIA Y ANALISIS DEL RIESGO EN SALUD PÚBLICA, 2017;17.

38. Simião AR, Barreto FKdA, Oliveira RdMAB, Cavalcante JW, Lima Neto AS, Barbosa RB, Lins CdS, Meira AG, Araújo FMdC, Lemos DRQ, Alencar CH, Cavalcanti LPdG. A major chikungunya epidemic with high mortality in northeastern Brazil. *Revista da Sociedade Brasileira de Medicina Tropical* 2019;**52**.

39. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. *International Journal of Epidemiology* 2012;**41**(2):514-520.

40. Lima Neto AS, Sousa GS, Nascimento OJ, Castro MC. Chikungunya-attributable deaths: A neglected outcome of a neglected disease. *PLoS neglected tropical diseases* 2019;**13**(9):e0007575-e0007575.

41. Pan American Health Organization (PAHO). *Tool for the diagnosis and care of patients with suspected arboviral diseases*. 1st Edition ed. Vol. PAHO Strategic Plan 2014-2019. Washington: PAHO, 2017.

5.3 Supplementary Material Manuscript 2

Spatiotemporal Distribution and Socioeconomic Disparities on Dengue, Chikungunya and Zika in Latin America from 2007 to 2017

Appendix 1. Descriptive characteristics

Table S.1. Ethnic distribution of arboviruses in Fortaleza, Brazil (2007-2017) and Medellin, Colombia (2008-2017).

	Fortaleza (Brazil)			Medellin (Colombia)		
Variable	Dengue	Chikungunya	Zika	Dengue	Chikungunya	Zika
Total cases	199,911	79,856	1,659	39,509	722	656
Ethnicity	n(%)	n(%)	n(%)	n(%)	n(%)	n(%)
White	17,399 (10.0%)	5,452 (6.8%)	302 (19.2%)	37992 (96.2%)	706 (97.5%)	636 (97.0%)
Black	3,689 (2.1%)	1,320 (1.7%)	47 (3.0%)	1327 (3.4%)	12 (1.7%)	13 (2.0%)
Indigenous	365 (0.2%)	143 (0.2%)	0 (0)	189 (0.5%)	6 (0.8%)	7 (1.1%)
Yellow	2,121 (1.2%)	1,849 (2.3%)	14 (0.9%)	-	-	-
Pardos	108,618 (62.4%)	57,036 (71.5%)	1,067 (67.9%)	-	-	-
Not disclosed	41,744 (24.0%)	14,005 (17.5%)	142 (9.0%)	-	-	-

Note: This information is merely descriptive to illustrate the ethnicity distribution in both study sites, but ethnicity was not included as covariate in any of the inequality analysis. Brazil has five ethnic groups classified as 1) Whites or Caucasian; 2) Black, including any person considered from African ancestry; 3) Yellow (*Amarela*), which includes people from Asian ancestry or mixed ups between whites and Asian; 4) Pardo, which includes the majority of the population in several regions of the country and indicates any mixing between any of whites, black and or Yellow; and 5) Indigenous, or aboriginal, which indicate native Americans. It is also possible that the reported ethnicity is 'ignored' or non-disclosed.

Table S.2a. Descriptive neighborhood characteristics for notified arboviral cases in Medellin, Colombia Between 2007- 2017.

Medellin	Dengue	Chikungunya	Zika
Total cases	N=39509	N=722	N=656
%Electricity, median (IQR)	98 (96, 100)	97 (95, 100)	100 (97, 100)
%Water supply, median (IQR)	98 (95, 100)	96 (92, 100)	100 (96, 100)
Sewage, median (IQR)	97 (94, 100)	95 (92, 98)	98 (94, 100)
Waste management, median (IQR)	96 (91, 99)	94 (87, 99)	97 (93, 100)
People per Household, median (IQR)	4.4 (4.0, 4.7)	3.8 (3.5, 4.3)	3.9 (3.6, 4.4)
US$ HH Income (proxy), median (IQR)	283 (264, 361)	279 (268, 396)	279 (266, 357)

Table S.2b. Descriptive neighborhood characteristics for notified arboviral cases in Fortaleza, Brazil between 2007- 2017.

Fortaleza	Dengue	Chikungunya	Zika
Total cases	N=199,911	N=79,856	N=1,659
%Electricity, median (IQR)	99 (98, 100)	99 (98, 100)	100 (99, 100)
%Water supply, median (IQR)	95 (91, 97)	95 (93, 98)	95 (92, 97)
%Literacy, median (IQR)	94 (91, 96)	94 (91, 96)	94 (91, 95)
%Waste management, median (IQR)	99 (97, 99)	99 (98, 99)	99 (97, 99)
Number of health care facilities, median (IQR)	2.0 (1.0, 3.0)	2.0 (1.0, 3.0)	2.0 (1.0, 3.0)
Number of Educative Institutions, median (IQR)	5.0 (2.0, 9.0)	4.0 (2.0, 8.0)	5.0 (2.0, 9.0)
Human development Index, median (IQR)	3.4 (2.5, 4.5)	3.4 (2.5, 4.7)	3.4 (2.3, 4.1)

Appendix 2. Concentration Index of Inequality

The Relative Concentration Index (RCI) is a measure of relative inequality that allows the identification of differential burden of diseases among the population in different socio-economic strata[1-4]. The RCI estimation was given by:

$$\text{RCI} = \frac{2cov(y_{ij}, R_{ij})}{\mu} = \frac{2}{n^2 \mu_h} \sum_{i=1}^{n} y_{ij} R_{ij}$$

where y_{ij} is the health outcome variable (disease rates in neighborhood i at time j), R_{ij} the rank of socioeconomic measure for neighborhood i at time j, and μ is the spatiotemporally adjusted mean disease rate. The RCI could be interpreted as a weighted mean of morbidity shares, with the weights depending on the fractional socioeconomic measure rank[2,3]. The RCI equals zero in absence of inequality, is negative when the disease rate concentrates among the poor, and positive when it is the opposite[1]. The concentration index corresponds to the area under the curve that lies above the 45° line of 'equality'. The concentration index equals zero in absence of inequality, is negative when the rate of morbidity/mortality concentrates among the poor and positive when it is the opposite[1]. We also estimated the absolute concentration index (ACI), as the product of the RCI and the mean disease rate for each outcome $ACI = RCI * \mu$.

Given that the RCI is a relative measure that depends on a rank, and because for each disease and each city the starting point of the socioeconomic rank changed by year, we conducted stratified analysis by yea (Figure S.1-S.2). Crude RCI were estimated with unadjusted rates and the adjusted RCIs were estimated by including age, sex, sanitation, and number of people per household in each disease-specific model, without including the socioeconomic covariate as an explanatory variable[4] (Table S.3; Figure S.3). Age and sanitation variables were associated with increased rates in Fortaleza, and number of people per household and sanitation variables in Medellin. RCI and ACI trends were estimated using adjusted models (Table 3 in main text).

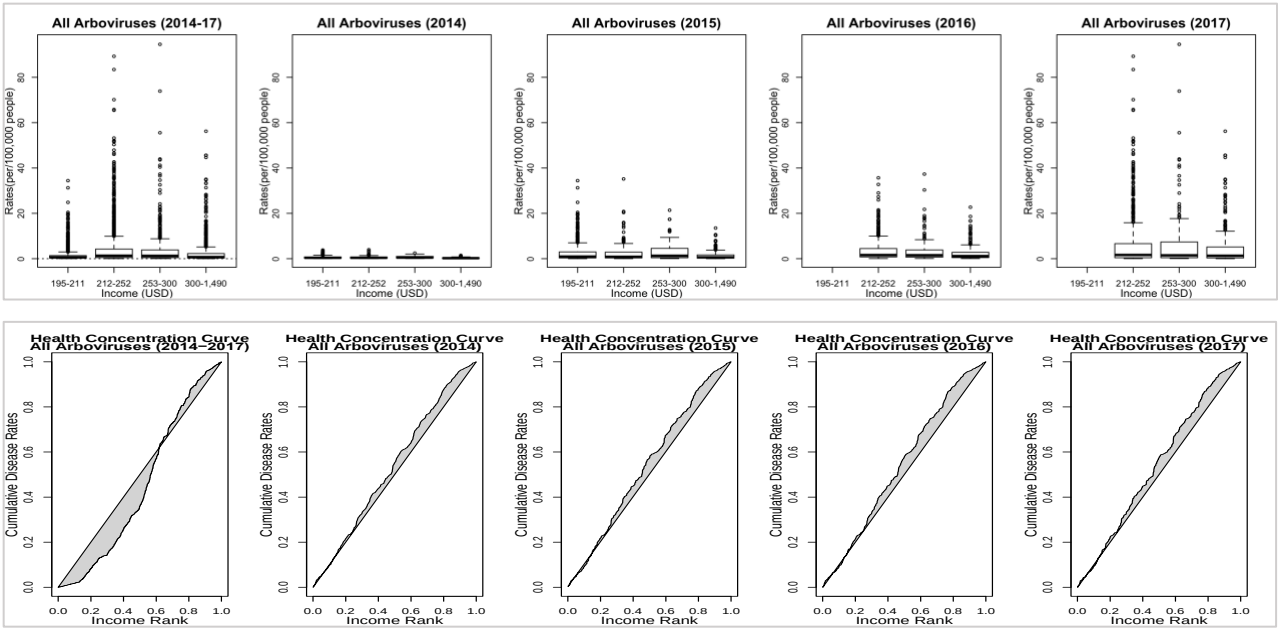


Figure S.1. Top: Distribution of aggregated arboviruses by income (Quartile) 2014-2017.Bottom: Health concentration curve for aggregated arboviruses cumulative and by year.

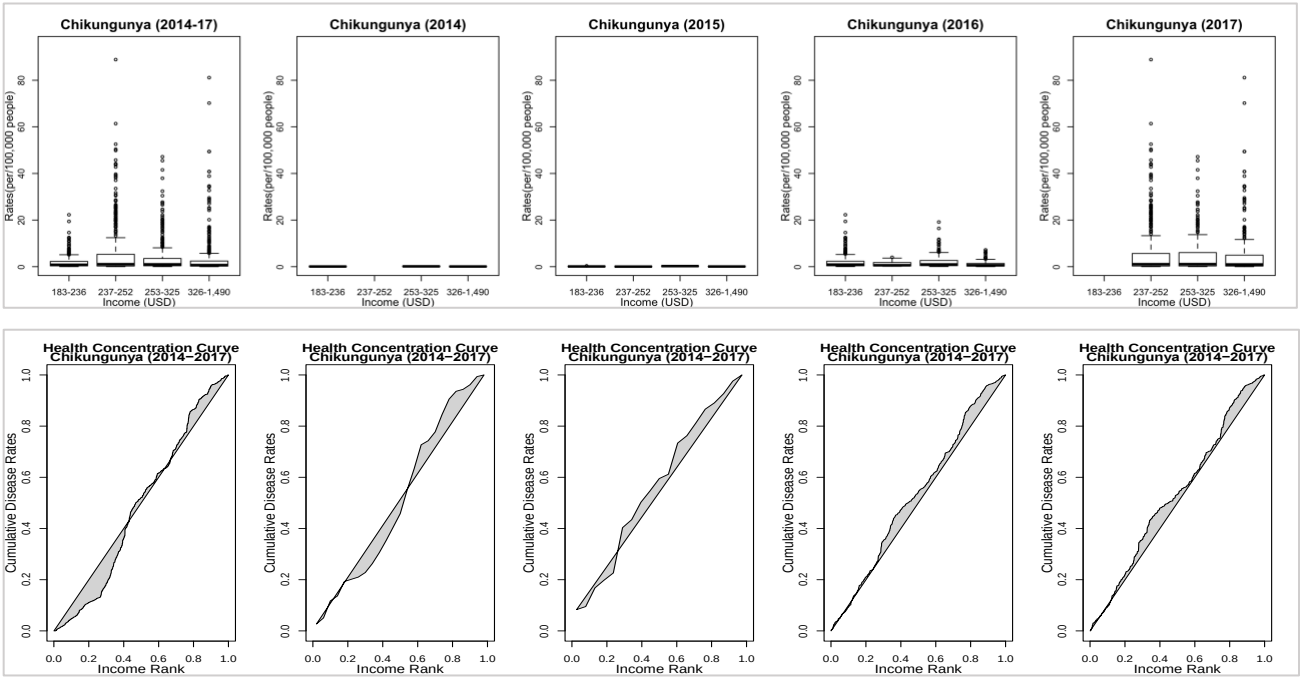


Figure S.2. Top: Distribution of chikungunya by income (Quartile) 2014-2017.Bottom: Health concentration curve for chikungunya cumulative and by year.

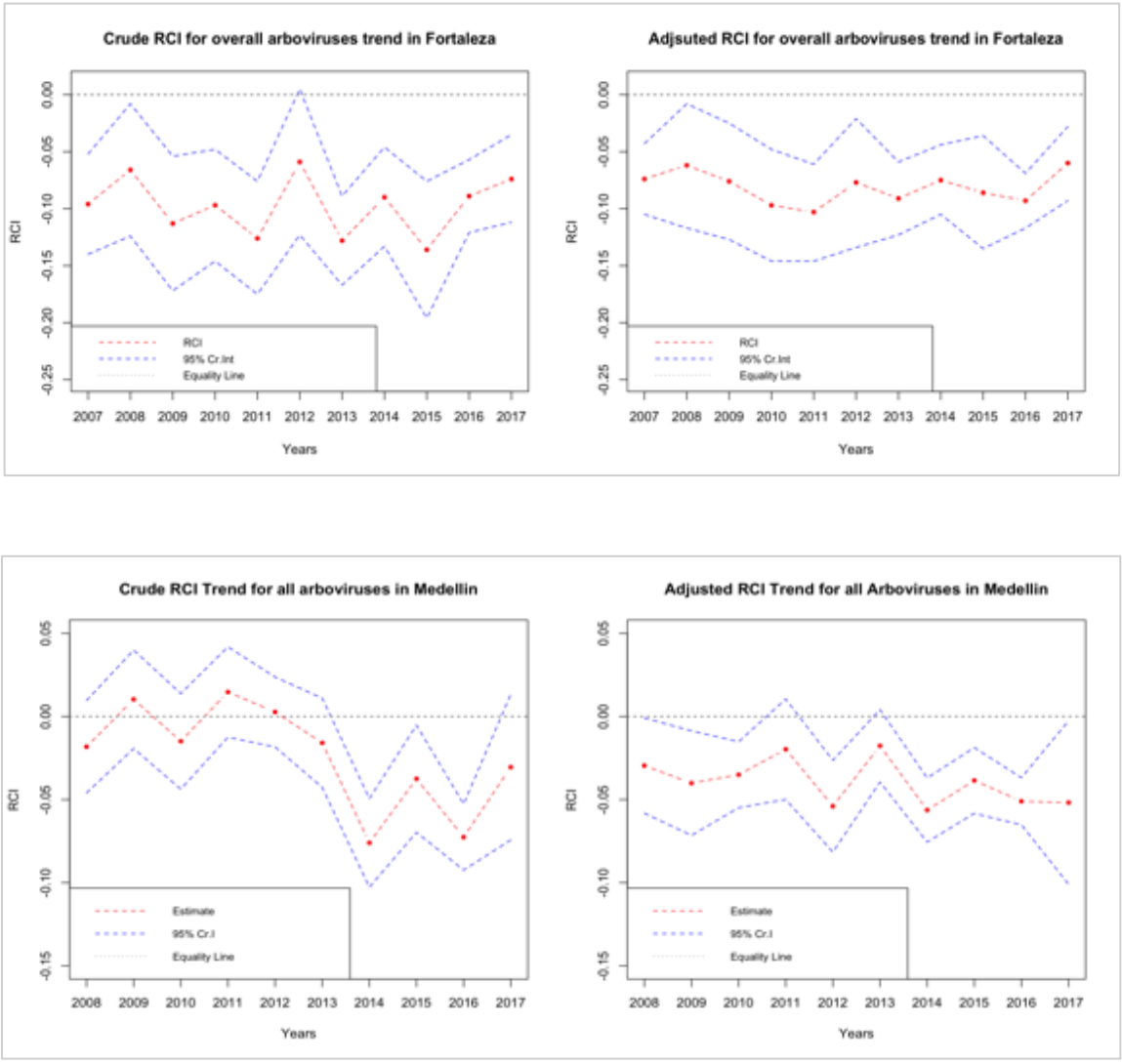


Figure S.3. Crude and Adjusted Relative Concentration Index (RCI) by year for aggregated arboviruses in Fortaleza, Brazil (Top) and Medellin, Colombia (Bottom).

Table S.3. Adjusted posterior mean of the fixed effect covariates and their 95% Credible Intervals for aggregated arboviruses distribution in Fortaleza, Brazil (2007-2017) and Medellin, Colombia (2008-2017).

Fortaleza	IRR	95%Cr.I
% Female	1.01	(0.99, 1.01)
%Under 5 years	0.95	(0.93, 0.96)
%6-9 years	0.98	(0.96, 0.99)
%10-19 years	0.99	(0.97, 1.01)
%20-49 years	0.97	(0.96, 0.99)
%Over 50 years	0.99	(0.97, 1.01)
%Waste management	0.98	(0.97, 0.99)
%Water supply	0.99	(0.98, 1.01)
Health Institutions (<3)	0.98	(0.97, 1.00)
Health Institutions (3-5)	1.00	(0.98, 1.02)
Health Institutions (>6)	0.94	(0.93, 0.96)
Medellin	IRR	95%Cr.I
%Electricity	0.95	(0.72, 1.26)
% Water Supply	0.93	(0.71, 1.23)
%Waste management	1.86	(1.53, 2.26)
Number of people per household	1.03	(1.00, 1.06)
%Female	0.99	(0.98, 1.00)
% Under 20 years	0.99	(0.99, 1.00)
% 20 – 49 years	0.99	(0.99, 1.00)
% Older than 50 years	0.99	(0.99, 1.00)

Incidence Rate Ratio (IRR) = Exponentiated adjusted posterior mean of the fixed effect covariates from the model adjusting for covariates included in the table.

Appendix 3. Socioeconomic Measures as Random Effects

Given that the Relative Concentration Index (RCI) is a relative measure that depends on a rank, and because for each disease and each city the starting point of the socioeconomic rank changed by year, we also fit models including the socioeconomic variables as structured random effects (rw1) to estimate the adjusted attributable effect of SES in the distribution of diseases. The model was specified as follows:

$$Y_{ij} \sim \text{Poisson}(E_{ij}\lambda_{ij})$$
$$log(Y_{ij}) = \beta_0 + x_{ij}^T\beta_j + s_i + t_j + ses_{ij}$$
$$[s_1 \dots s_n]^T \sim \text{BYM}(\sigma, \phi)$$
$$[t_1 \dots t_n]^T \sim \text{RW1}(\gamma)$$

Where λ_{it} is the disease rate in time t for neighborhood i, E_{ij} the log of the neighborhood-specific population or offset, $log(Y_{ij})$ is the log incidence rate of the arbovirus in neighborhood i at time j given the offset (E_{ij}), x_{ij} is a vector of neighborhood-level covariates with its corresponding parameter β_j; with the exponentiated coefficients indicating an incidence rate ratio (IRR). The parameters s_i, t_j and ses_{ij} denote the spatial, temporal, and socioeconomic structured random effects, respectively. The spatial effect ($s_i = s_{1\dots}s_{119}$ in Fortaleza and $s_i = s_{1\dots}s_{249}$ in Medellin) was modelled assuming a Besag-York-Mollie (BYM) specification[5], which includes a structured component using an intrinsic conditional autoregressive structure (iCAR) and an unstructured component modelled as independent and identically distributed (i.i.d). The σ parameter is a standard deviation parameter with σ^2 a weighted sum of the spatial and independent variance terms. The ϕ parameter is similar to a correlation, being the proportion of the variance attributable to the spatially structured term by neighborhood[6].

The temporal component was modelled dynamically using a scaled random walk of first order structure (RW1) for the month. As sensitivity analysis, the socioeconomic level per neighborhood (i.e., each SES level and income group in each city) was modeled as unstructured i.i.d random effect. Given that income and the SES index could be corelated to the spatial structure, we fit models without the spatial random effect (presented in the main text), as follows:

$$log(Y_{ij}) = \beta_0 + x_{ij}^T \beta_j + t_j + ses_{ij}$$

Both results were consistent, showing a concentration of arboviruses among people in the lower spectrum of the socioeconomic distribution. However, as expected, the posterior mean of the socioeconomic-specific random effects from the model adjusting for the spatial unit (neighborhood) as random effects were slightly less precise than those obtained from the models without adjusting for spatial unit (Figure S.4). Exponentiated fixed effects from the model log(rates) are Incidence Rate Ratios (IRR) for each specific covariate and exponentiated log(rates) from the random component are random-effect specific (SES-specific) IRRs.

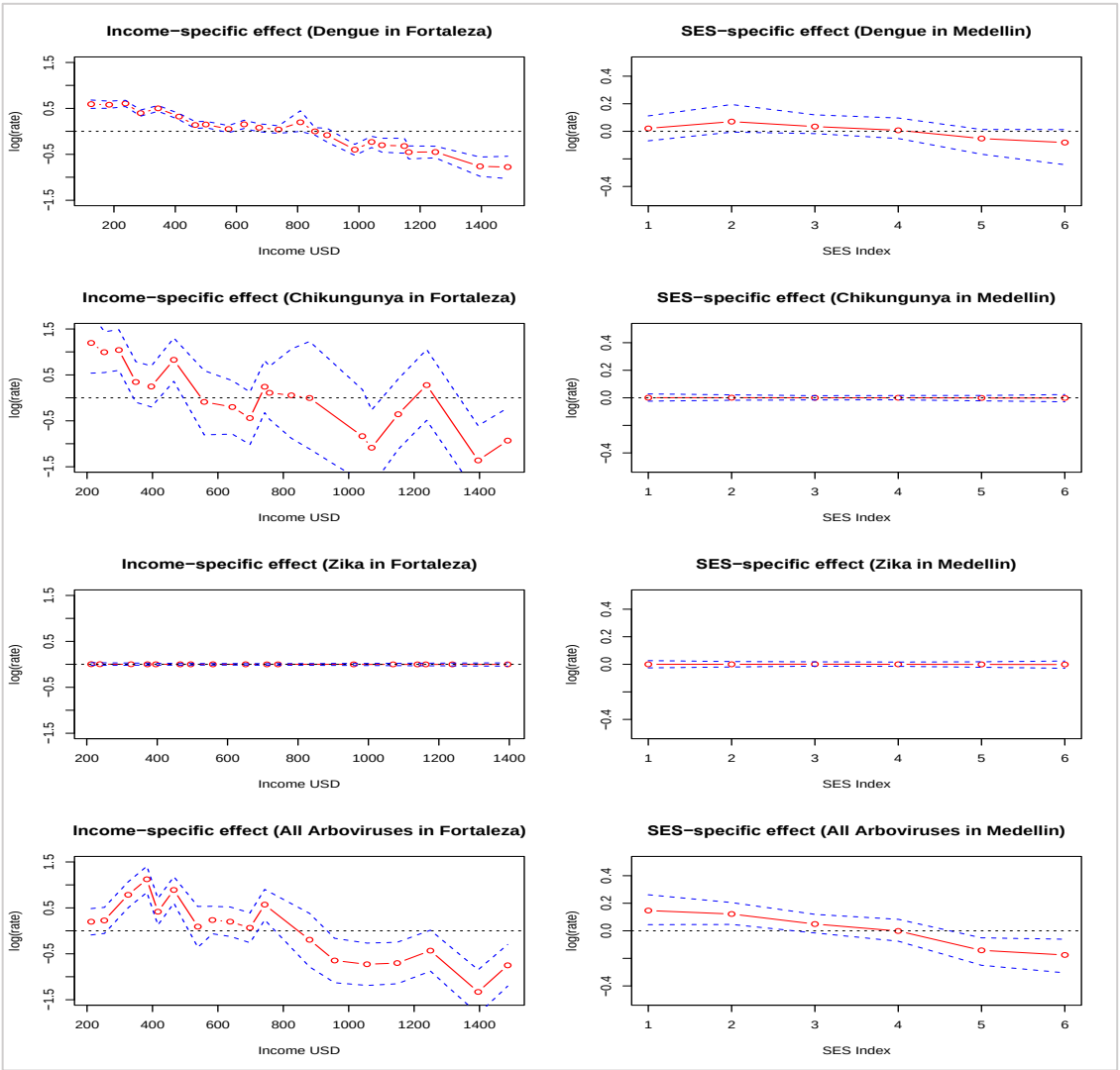


Figure S.4. Posterior mean and 95% Credible Intervals (log scale) of the Socioeconomic-specific Random Effects for overall and each specific arbovirus distribution in Fortaleza, Brazil (2007-2018) and Medellin, Colombia (2008-2017).

Appendix 4. Relative concentration index using Human Development index (HDI) in Fortaleza and Income-proxy in Medellin.

To test the strength of our analysis with other available socioeconomic measures, we estimated the RCI using the Human Development Index in Fortaleza (removing the education-related variables form the disease-model), and used an extrapolated proxy measure of the mean income per neighborhood obtained from the quality of life survey from Medellin[7].

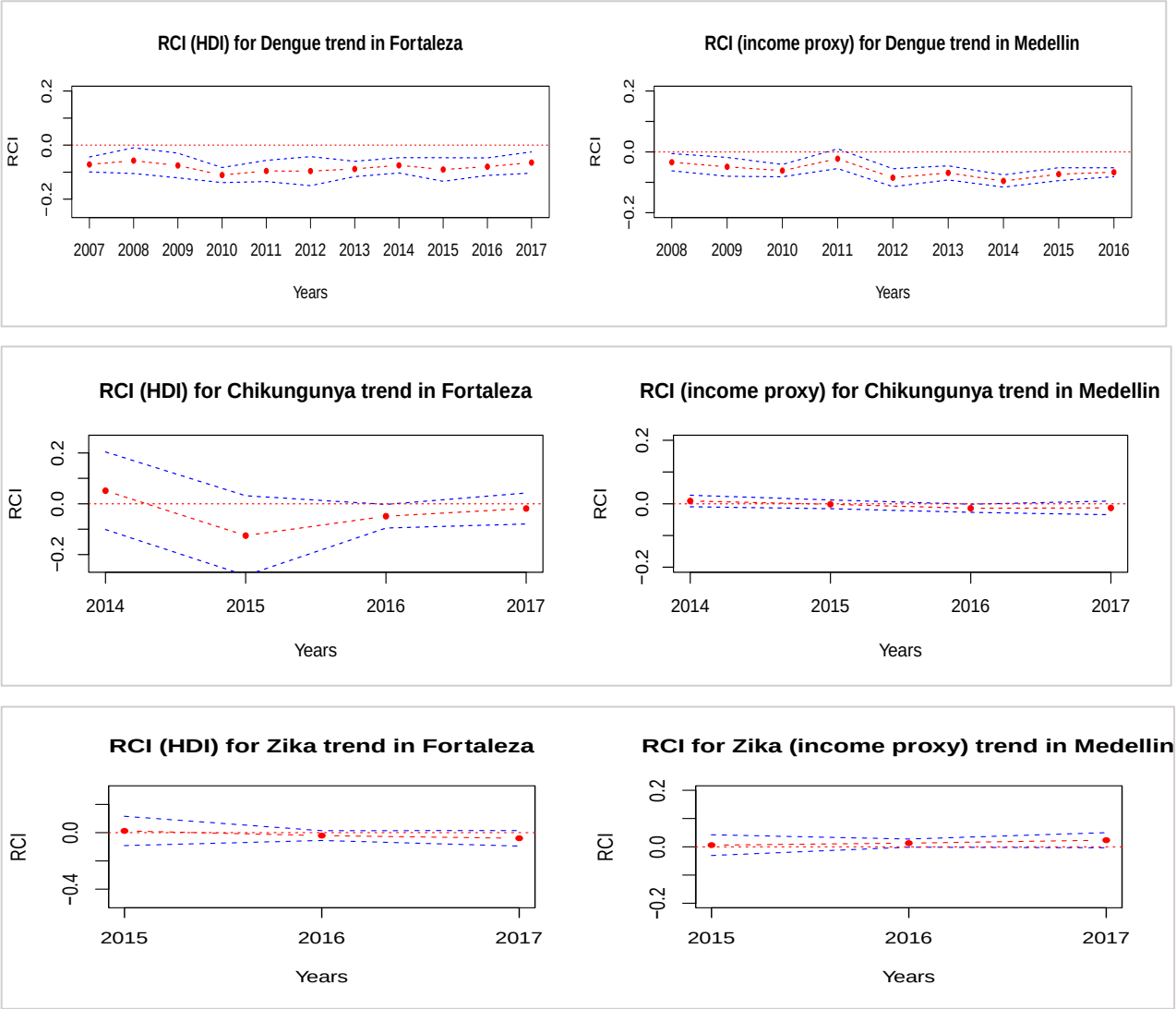


Figure S.5. Relative Concentration Index (RCI) trend overtime for dengue, chikungunya and Zika distribution in Fortaleza, Brazil (2007-2018) and Medellin, Colombia (2008-2017), using Human Development index (HDI)in Fortaleza and Income-proxy in Medellin as the socioeconomic measure. RCI estimates (Red dots line) and 95% Credible Intervals (Blue dashed line) and red dotted line indicating the equality line.

Appendix 5. Codes

Estimates were obtained using modified functions from the 'decomp' R-packages for health inequalities[8], to specifically extract parameters from INLA models, using RStudio (R version 3.6.1, R Core Team; Vienna, Austria; 2019).

Functions are in the attached in a R-Script document.

Additional References:

1. Erreygers G. Correcting the Concentration Index. Journal of Health Economics 2009;28(2):504-15.
2. Konings P, Harper S, Lynch J, et al. Analysis of socioeconomic health inequalities using the concentration index. International Journal of Public Health 2010;55(1):71-4.
3. Wagstaff A. The concentration index of a binary outcome revisited. Health Economics 2011;20(10):1155-60.
4. Erreygers G, Kessels R. Regression-Based Decompositions of Rank-Dependent Indicators of Socioeconomic Inequality of Health. In: Kessels R, ed. Health and Inequality: Emerald Group Publishing Limited, 2013:227-59.
5. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics 1991;43(1):1-20.
6. Blangiardo M, Cameletti M, Baio G, et al. Spatial and spatio-temporal models with R-INLA. Spatial and Spatio-temporal Epidemiology 2013;7(Supplement C):39-55.
7. Alcaldia de Medellin. Encuesta Calidad de Vida 2004 - 2018. Transparencia y acceso a la información pública. Medellin, Colombia, 2019.:www.medellin.gov.co.
8. Konings P, Speybroeck N. decomp: Various functions to quantify and decompose health inequalities. CRAN, 2012:This package contains functions to calculate the relative concentration index, plot Lorenz curves and decompose inequality using (generalized) linear models and survival models while taking survey design into account.

Chapter 6: **Decomposition of Socioeconomic Inequalities in Arboviral Diseases in Brazil and Colombia (2007- 2017) (Manuscript 3).**

6.1 Preface Manuscript 3

Once the presence of inequalities on arboviral diseases was documented, it was important to identify the factors contributing to the inequality. This manuscript assesses the contribution of some socio-environmental and individual covariates to the overall inequality. This analysis allowed the identification of the time of notification and sanitation variables as the main contributors to the overall inequality on the studied arboviruses in both Brazil and Colombia. While time of notification, availability of healthcare facilities and age of the case were the main contributors in Fortaleza, age and waste management were the main contributors to the socioeconomic inequality in Medellin. Overall, these findings stress that the presence and magnitude of the inequalities and that the role and contribution of each determinant is context specific.

This manuscript is in preparation for submission to the *Pan American journal of Public Health (PAHO Journal)*. I am targeting this journal given their scope, audience (public health decision makers, program coordinators and other Latin American stakeholders), and regional relevance to the topic discussed here. This manuscript is prepared mainly for a public health audience, providing descriptive quantitative information on the health inequalities with some qualitative interpretation of the results. The main objective to submit this manuscript to the *PAHO* journal is to call the attention on the presence of inequalities and provide quantitative estimates of their contributors, highlighting the role that outbreaks and introduction of new diseases have on the presence of inequalities.

6.2 Manuscript 3

Tittle: Decomposition of Socioeconomic Inequalities in Arboviral Diseases in Brazil and Colombia (2007- 2017)

Authors: Mabel Carabali[1], Sam Harper[1], Antonio S. Lima-Neto[2,3], Andrea Caprara[4], Berta Nelly Restrepo[5], Jay S. Kaufman[1].

Affiliations:

[1] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada;

[2]Fortaleza's Secretary of Health, Fortaleza, Brazil;

[3]University of Fortaleza, Fortaleza, Brazil;

[4]Universidad Estadual de Ceara, Fortaleza, Brazil

[5]Instituto Colombiano de Medicina Tropical, Universidad CES, Sabaneta, Antioquia

Corresponding author: Mabel Carabali (mabel.carabali@mail.mcgill.ca)

McGill University, Purvis Hall, 1020 Pine Avenue West, Montreal, Quebec, Canada (H3A 1A2).

Tel.: 514-398-6258; Fax: 514-398-4503

ABSTRACT

Objectives: To identify social determinants of health contributing to socioeconomic inequalities among arboviral diseases in Latin America.

Methods: We used surveillance data from Brazil and Colombia between 2007-2017, to assess the presence of socioeconomic inequalities on dengue, chikungunya and Zika at the neighborhood level. We estimated and decomposed the relative concentration index of inequality (RCI) accounting for the spatiotemporal distribution of the diseases.

Results: There were 281 426 arboviral cases notified in Fortaleza, Brazil and 40 889 in Medellin, Colombia. The RCI indicated greater concentration of dengue cases among people living in low socioeconomic settings in Fortaleza and Medellin. The RCIs for chikungunya in Fortaleza covered the line of equality during their introduction in 2014 while the RCIs for Zika and chikungunya in Medellin indicated the presence of a small inequality. The RCI decomposition showed that the year of notification and age of the cases were the main contributors to this small inequality. In Medellin, the RCI decomposition showed that main contributors were age and access to waste management, accounting for 75.5%, 72.2% and 54.5% to the overall inequality towards the poor for dengue, chikungunya and Zika, respectively.

Conclusions: Our study presents estimates of the socioeconomic inequality of arboviruses and its decomposition in two Latin American cities. We corroborate the concentration of arboviral diseases in low socioeconomic neighborhoods and identify that year of occurrence, age, presence of healthcare facilities, and waste management are key determinants of the heterogenous distribution of endemic arboviruses across the spectrum of socioeconomic status.

Key words: arboviruses; dengue; chikungunya; Zika; Social Inequity; Collective Effects of Health Disparities; Social Determinants of Health; Latin America; Brazil; Colombia

INTRODUCTION

Dengue, chikungunya and Zika are main public health concerns in the Americas region[1-4]. These three arboviruses have a similar symptomatology, are illnesses for which specific curative treatments do not exist, and diseases for which sufficiently safe and effective vaccines are not yet introduced[1,2]. Among these three diseases, dengue has the highest incidence worldwide with the Americas region experiencing the second largest burden of notified cases. The largest dengue outbreaks in the region occurred in 2016 and 2019, with 2.38 and 3.1 million cases, respectively[1,5]. Chikungunya and Zika are considered re-emergent diseases and were introduced in the Americas in 2013 and 2015, respectively[1,5]. Chikungunya's outbreak included over a million cases in 2014 and Zika was considered a Public Health Emergency of International Concern (PHEIC) in 2016, due its association to congenital malformations. In the Americas, Brazil and Colombia account for 40-60% of the overall burden of arboviruses in the region[1,2,5,6]. In Colombia, the burden of dengue concentrates in 50 out of 1 101 municipalities, and chikungunya and Zika have been consistently notified on the same 30 municipalities[7-9]. In Brazil, although the distribution of dengue cases varies across the country, the Northwest and Central-Eastern region concentrates the majority of cases of Zika and chikungunya as well[10-13].

Although infectious diseases are well-known to demonstrate important social inequalities, the literature on arboviral diseases with the specific focus of health inequalities is limited[14,15]. Given the heterogeneous distribution of arboviral cases across different strata of socioeconomic status (SES)[1,11,12,14-16] and having acknowledged the presence of socioeconomic inequalities on arboviruses in Latin America[14,16], it is imperative to understand what are the main contributors to the inequality. This information is key to contribute to the design and implementation of disease control strategies that could target the most affected groups, and to interventions aimed at decreasing socioeconomic inequalities at local level.

Therefore, with the objective to identify which social determinants contribute most to socioeconomic inequalities in arboviral diseases in Latin America; we used longitudinal surveillance data of dengue, chikungunya and Zika and assessed the presence of socioeconomic

inequality at the neighborhood level in Brazil and Colombia, by estimating and decomposing the relative concentration index of inequality.

METHODS

Study Sites: This study was conducted with data from Fortaleza, Brazil and Medellin, Colombia. Municipalities were selected based on the disease burden, knowledge of the context, presence of functioning surveillance system, data availability, and expressed interest of the local health agencies on the study.

Fortaleza is the capital of Ceara, the fifth biggest state of Brazil, and has a population of 2.5 million inhabitants[17]. The Northeastern region of Brazil experiences 29% of Zika cases and Fortaleza is consistently one of the top 50 municipalities (out of more than 5 000) reporting dengue and chikungunya during the last 10 years[17-19]. The city has six districts and 119 urban neighborhoods distributed in 315 km^2. Fortaleza's altitude is 21 meters above the sea level, the average temperature is 26.6^oC and has one rainy season (January to May). Brazil has Universal Health Coverage denominated Unified Health System, which is expected to cover the health needs of the entire population[18].

Medellin is the second largest city in Colombia with 2.6 million inhabitants[20]. Dengue incidence ranged 161-745 cases per 100 000 inhabitants during the last 10 years and has been consistently included on the top five reporting cities since 1998[8,9]. The city has 16 urban districts and 249 urban neighborhoods distributed in 110 km^2. Medellin's altitude ranges from 1 460 to 3 200 meters, the average temperature 24^oC, and has two rainy seasons (April and October). Although 50% of the city belongs to low SES, 98% of the city has access to potable water. Health coverage of the population is as follows: 70% contributory (employees or self-employees with capacity to pay for health coverage), 25% government subsidized, and 4% uninsured[20].

Data Sources: Notification of dengue, chikungunya and Zika is mandatory in all study sites through a passive surveillance system[21,22]. All cases of *Aedes*-transmitted diseases are individually

registered in the national surveillance system (SINAN-Brazil and SIVIGILA-Colombia) and our sample included all notified cases in the study areas between 2007 and 2017. Supplementary aggregated information about socioeconomic factors at the neighborhood level was obtained using National Census data and local quality of life and basic needs surveys for socioeconomic data[17,23].

Outcomes: Clinically and laboratory confirmed dengue, chikungunya and Zika cases registered in the study site's surveillance system were used for this analysis. Given the possibility of misclassification of chikungunya and Zika with dengue at early stages of their introduction (e.g., chikungunya or Zika could have been misdiagnosed as dengue in 2014 or 2015, respectively)[2]; we also grouped these diseases into a single variable denominated "*all arboviruses*". We estimated the aggregated and disease-specific spatiotemporally adjusted rate of disease by month and neighborhood and used it as the health outcome for the estimation of the inequality.

Socioeconomic measures: In Brazil, the socioeconomic measure used was the median monthly household income in USD as a continuous variable[17]. In Colombia, we used the national Socioeconomic Status (SES) index, which is an administrative summary ordinal measure (range 1–6) with one indicating the lowest SES and six indicating highest SES[32]. The Colombian SES is a standardized and validated measure constructed by the National Department of Planning, according to the characteristics of the household including construction material, presence of assets and household conditions. Each house has a designated SES level provided by the municipality and each neighborhood in turn possess an SES designation according to a weighted mode of the household's SES within each block and neighborhood[24].

Other covariates: The *a priori* covariates included into the decomposition were: the proportion of disease-specific female cases and the proportion of cases by age group (<5 years, 5-9 years, 10-20 years, 20-50 years, and >50 years) per neighborhood and per month. As SDH covariates we included the proportion of people with secondary education, proportion of households with adequate supply of potable water, waste management, number of people per household, type of

health insurance and number of healthcare centers per neighborhood, according to the availability in each dataset.

Statistical analysis

Descriptive statistics for the data are presented as mean and standard deviation (SD) or median and interquartile ranges (IQR) for continuous variables and as proportions for binary or categorical variables.

Relative Concentration Index Decomposition: The assessment of socioeconomic inequality was conducted by the estimation of crude and adjusted relative index of inequality with a regression-based decomposition of the latter[25,26]. The Relative Concentration Index (RCI) is a measure of relative inequality that allows the identification of differential burden of diseases among the population in different socio-economic strata[26]. To estimate the RCI, we used disease-specific rates and the socioeconomic rank of the population in each area to estimate a relative concentration curve, which plots the cumulative fraction of cases on the y-axis against the cumulative fraction of the population ranked by SES on the x-axis. The RCI is twice the area that lies between the 45° line of equality and the concentration curve[26]. The concentration index equals zero in absence of inequality, is negative when disease rates are more concentrated among the poor, and positive when rates are more concentrated among the rich[25,26]. The regression-based decomposition was conducted using a Bayesian hierarchical Poisson or Negative binomial model of notified arboviruses. The regression model, the RCI estimation and the SDH-specific contribution were specified as follows:

Disease model: $log\left(\, y_{ij}\right) = \log\left(E_{ij}\right) + \beta_0 + \, x_{ij}^{n}\beta_n + s_i \, + t_j$, equation 1

Relative Concentration Index (RCI): $\mathbf{RCI} = \frac{2cov\left(\hat{y}_{ij}, R_i\right)}{\mu} = \frac{2}{n^2\mu_h} \sum_{i=1}^{n} \hat{y}_{ij} \, R_i$, equation 2

SDH contribution: $SDH_n = [(\beta_n \bar{x}_n/\mu) * RCI_n]/\mathbf{RCI}$, equation 3

Equation 1 models the count of cases y_{ij} in neighborhood i at month j[27]; E_{ij} is the mid-year neighborhood specific population as the offset; β_0 is the intercept indicating the average log rate of disease in all neighborhoods; $x_{ij}^n \beta_n$ is a vector of fixed-effects covariates including sex, age, sanitation and the other neighborhood level covariates listed above. To account for the spatiotemporal distribution of the diseases, we included structured random effects for neighborhoods (s_i) and month of notification (t_j), and indicator variables for the year of notification in the disease model. The RCI was estimated using equation 2, where $\hat{y}_{ij}$ is the covariates and spatiotemporally adjusted disease rate obtained from equation 1, R_i is the rank of socioeconomic measure at neighborhood i, and μ is the mean rate of reported cases. The SDH-specific contribution was estimated using equation 3, where the component $\beta_n \bar{x}_n / \mu$ is the elasticity, a parameter that determines the strength of the association between covariates and the inequality; β_n is the beta coefficient for each fixed effects covariate in the disease model regression, $\bar{x}_n$ is the covariate-specific mean, and RCI_n is the covariate-specific concentration index. The specific relative contribution to the overall inequality is obtained by dividing the covariate-specific contribution by the overall RCI[25,28].

Given that the RCI could change over time and that socioeconomic measures could be correlated with the neighborhood, we fit models stratified by year and a set of models without the spatial random effect as sensitivity analysis. To illustrate the magnitude of the inequality in presence of a monotonic relationship (with concentration of disease rates at the lower end of the SES distribution), we conducted a sensitivity analysis using the same rates from our study but scaling the SES measure. We calculated the aggregated and disease-specific RCIs, the point-wise intervals, and contributions using modified functions from the 'decomp' R-packages for health inequalities[29] (Supplementary Material). All analyses were performed using RStudio (R version 3.6.1, R Core Team; Vienna, Austria; 2019).

Ethics statement: This study analyzed secondary data without human samples analyses, therefore, it did not require consent to participation. The protocol was reviewed and approved by the

respective Institutional Review Boards (Study No. A02-E05-18A) and by the ethics committee of the Brazilian Ministry of Health (Code: 2.624.599) and Secretary of Health of Medellin, Colombia.

RESULTS

There were 281 426 arboviral cases notified in Fortaleza, Brazil and 40 889 in Medellin, Colombia. Overall, crude and adjusted disease rates showed a non-monotonic relationship with socioeconomic measures. In Fortaleza, we observed higher dengue rates among low-income settings, Zika rates were low across the median-household income spectrum and chikungunya rates were highly variable. In Medellin, dengue rates were higher among middle SES-level neighborhoods while chikungunya and Zika rates were similar across the SES-strata (Figure 6.1A and 6.1B). Descriptive characteristics are presented in Table 6.1 and Table 6.2.

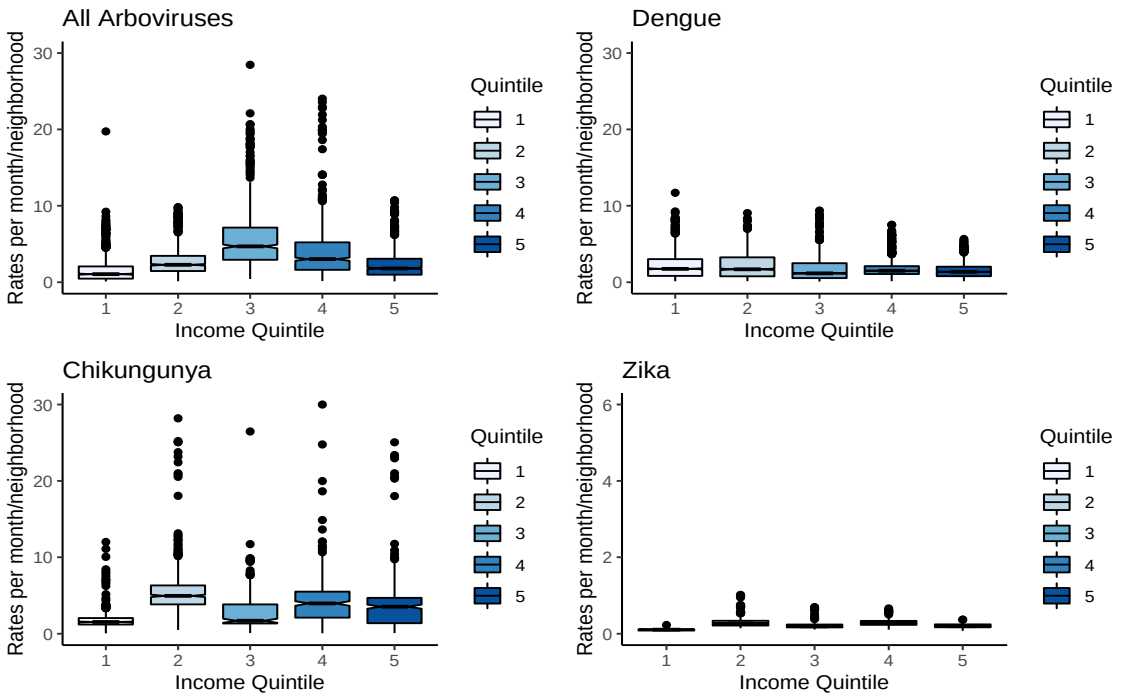
Table 6.1 Descriptive Characteristics of notified arboviral cases in Fortaleza, Brazil between 2007-2017.

Fortaleza	Dengue	Chikungunya	Zika
Total cases	N=200,832	N=80,408	N=1,681
Age, Median Years (IQR)	25 (15, 39)	38 (24, 53)	30 (20, 43)
Crude neighborhood rates per 100,000, Median (IQR)	27.6 (10.9, 79.7)	34.9 (11.4, 151.9)	8.4 (4.6, 15.7)
Sex			
Female	110546 (55.0%)	49544 (61.6%)	1125 (66.9%)
Male	90286 (44.9%)	30864 (38.4%)	556 (33.1%)
Not disclosed	38 (<1%)	37 (<1%)	1 (0.1%)
%Water supply, median (IQR)	95.1(91.5, 97.1)	95.0 (91.1, 97.0)	95.0 (91.6, 97.1)
%Literacy, median (IQR)	93.8 (90.8, 95.6)	94.0 (91.0, 95.7)	93.9 (90.8, 95.4)
%Waste management, median (IQR)	99.6 (98.6, 99.9)	99.7 (98.7, 99.9)	99.6 (98.6, 99.9)

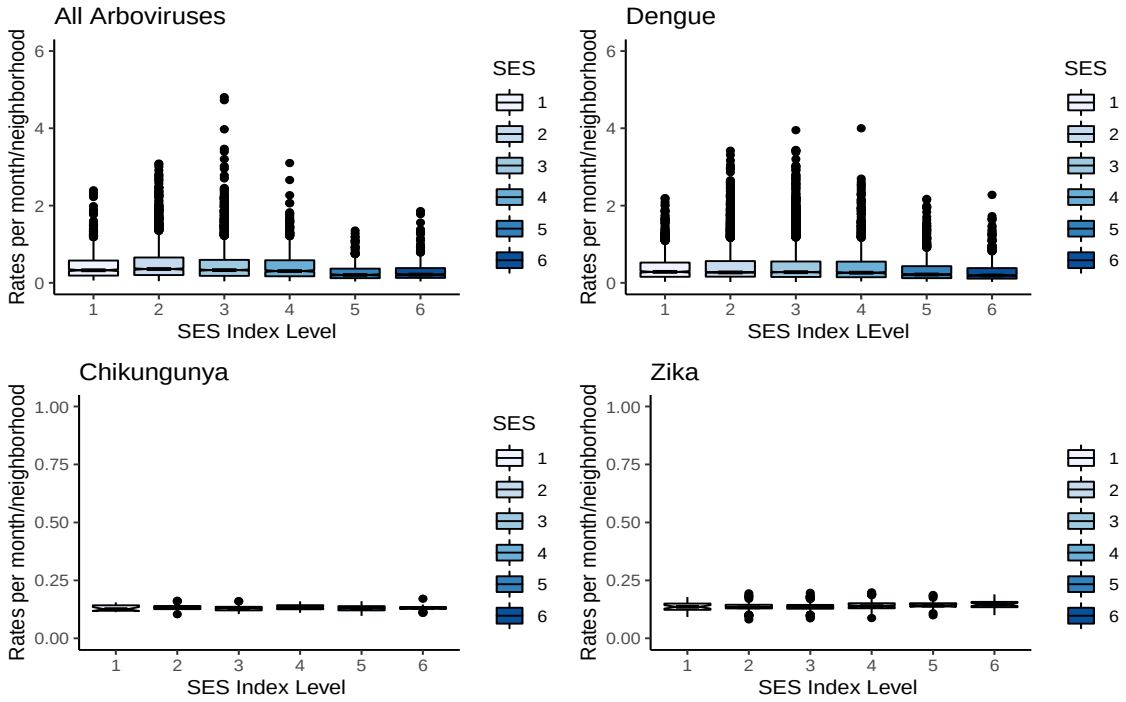
Number of health care facilities, median (IQR)	2.0 (1.0, 3.0)	2.0 (1.0, 3.0)	2.0 (1.0, 3.0)
Number of Educative Institutions, median (IQR)	5.0 (2.0, 9.0)	4.0 (2.0, 8.0)	5.0 (2.0, 9.0)
Human development Index, median (IQR)	3.4 (2.5, 4.5)	3.4 (2.5, 4.7)	3.4 (2.3, 4.1)

Table 6.2 Descriptive Characteristics of notified arboviral cases in Medellin, Colombia Between 2008- 2017.

Medellin	Dengue	Chikungunya	Zika
Total cases	N=39509	N=724	N=656
Age, Median Years (IQR)	28 (16, 45)	34 (24, 49)	29 (21, 42)
Crude neighborhood rates per 100,000, Median (IQR)	24.5 (12.3, 127.8)	10.1 (8.9, 17.4)	13.0 (8.9, 20.4)
Sex			
Female	18,973 (48.0%)	433 (59.8%)	402 (61.3%)
Male	20,536 (52.0%)	291 (40.2%)	254 (38.7%)
Insurance			
Subsidized Scheme	9,275 (23.5%)	144 (19.9%)	137 (20.9%)
Contributory Scheme	30,230 (76.5%)	580 (80.1%)	519 (79.1%)
%Electricity, median (IQR)	98.9 (96.4, 100)	97.5 (93.8, 100)	99.2 (95.7, 100)
%Water supply, median (IQR)	98.1 (94.8, 100)	95.8 (91.9, 100)	98.5 (95.1, 100)
Sewage, median (IQR)	97.5 (93.6, 100)	95.4 (92.3, 98.4)	97.7 (93.9, 100)
Waste management, median (IQR)	95.9 (91.2, 99.2)	93.6 (86.9, 98.6)	97.2 (93.3, 100)
People per Household, median (IQR)	4.4 (4.0, 4.7)	3.8 (3.5, 4.3)	3.9 (3.6, 4.4)
Mean monthly US$ Household Income (proxy), median (IQR)	283 (264, 361)	279 (268, 396)	279 (266, 357)



(A) FORTALEZA



(B) MEDELLIN

Figure 6-1 Disease rates by SES status.
(A) Adjusted disease rates across median household income quintiles in Fortaleza, Brazil (2007-2017).
(B) Adjusted disease rates across SES Index level in Medellin, Colombia (2008- 2017).

RCI Estimation and Decomposition

In Fortaleza, the overall adjusted RCIs were -0.02 (95% CI= -0.02, -0.01), 0.02 (95% CI= 0.02, 0.03), and -0.03 (95% CI= -0.04, -0.02), for dengue, chikungunya and Zika, respectively (**Figure 6.2**). The RCI decomposition showed that the year of notification contributed importantly to the overall inequality, either towards the poor or the rich. For dengue, aside from the year of notification, age was the main contributor to the overall inequality towards the poor with 102.7% from cases between 20-49 years old, followed by 11.8% from waste management. For chikungunya, we observed that presence of three to five healthcare units in the neighborhood contributed 23.3% and for Zika the presence of more than six healthcare units contributed 13% to the overall inequality towards the poor. For the aggregated arboviruses, only using data from 2014 to 2017, the adjusted RCI was 0.06 (95% CI = 0.05, 0.08). The age of cases and presence of health care institutions in the neighborhoods contributed 49.3% to the overall socioeconomic inequality towards the poor, while the years 2014 and 2017 contributed 75.1% and 52.3%, respectively, to the overall inequality towards the rich (**Figure 6.3A** and Supplementary Material).

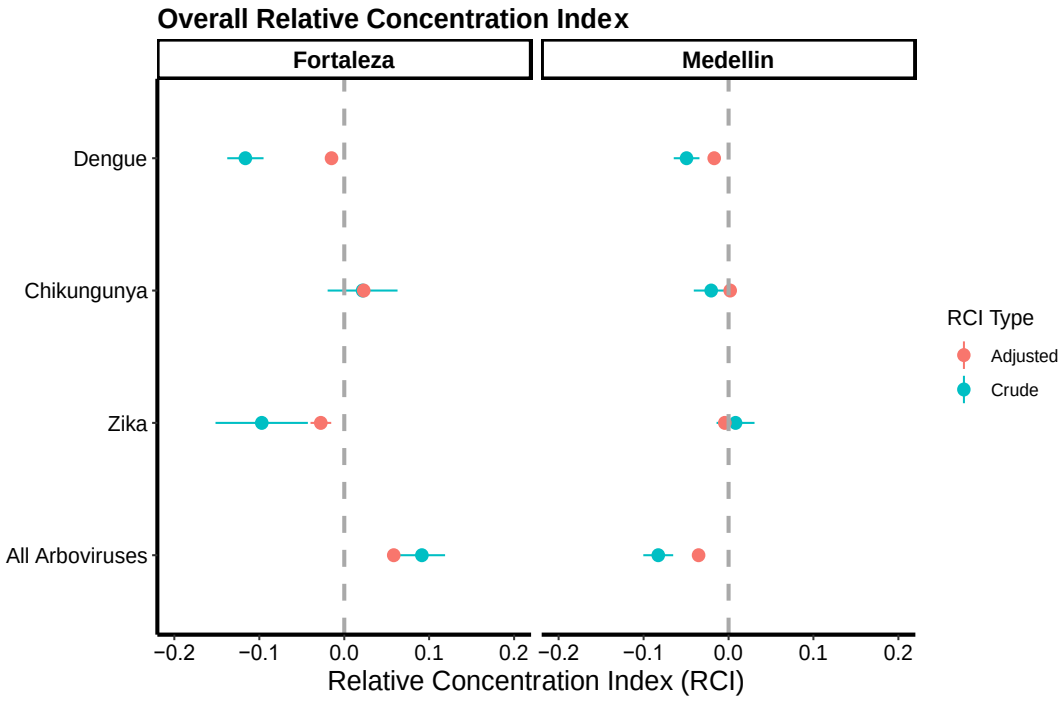


Figure 6-2 Crude and adjusted RCI for each arboviral disease in Fortaleza and Medellin, Colombia (2007-2017).

The stratified analysis by year for chikungunya did not show evidence of socioeconomic inequalities in 2014 and 2015. During 2016 and 2017, the RCIs indicated modest inequality with concentration of cases among people in low income neighborhoods RCIs= -0.03 (95% CI= -0.04, -0.01) and -0.01 (95% CI= -0.02, 0.00), respectively. Age and waste management were the main contributors to the inequalities towards the poor and presence of healthcare institutions in the neighborhood were the main contributor to the inequality towards the rich. The stratified analysis for aggregated arboviruses showed small RCIs consistently below the line of equality, ranging from -0.01 in 2014 to -0.04 in 2016. As contributors to the inequality towards the poor, waste management accounted for about 10% every year and presence of healthcare institutions contributed to 43% in 2017 (Supplementary Material).

In Medellin, the overall adjusted RCI showed the presence of small inequalities for dengue -0.02 (95% CI= -0.02, -0.01) and very small or no inequalities for chikungunya (Figure 6.2). The decomposition showed that main contributors to the overall inequality towards the poor on dengue were age and waste management. Contributory insurance accounted for 5.8% and 26.4% to the inequality towards the rich on dengue and chikungunya, respectively. The adjusted RCI for aggregated arboviruses was -0.03 (95% CI = -0.04, -0.03) with age contributing 24.5% to the overall socioeconomic inequality towards the poor (Figure 6.3B and Supplementary Material).

Overall, models fitted without spatial random effects did not change the results for the RCIs but showed larger contributions for age and sanitation covariates. RCIs from the sensitivity analysis using a monotonic relationship between socioeconomic measures and disease rates were on average larger in magnitude than those presented in the main analysis (Supplementary Material).

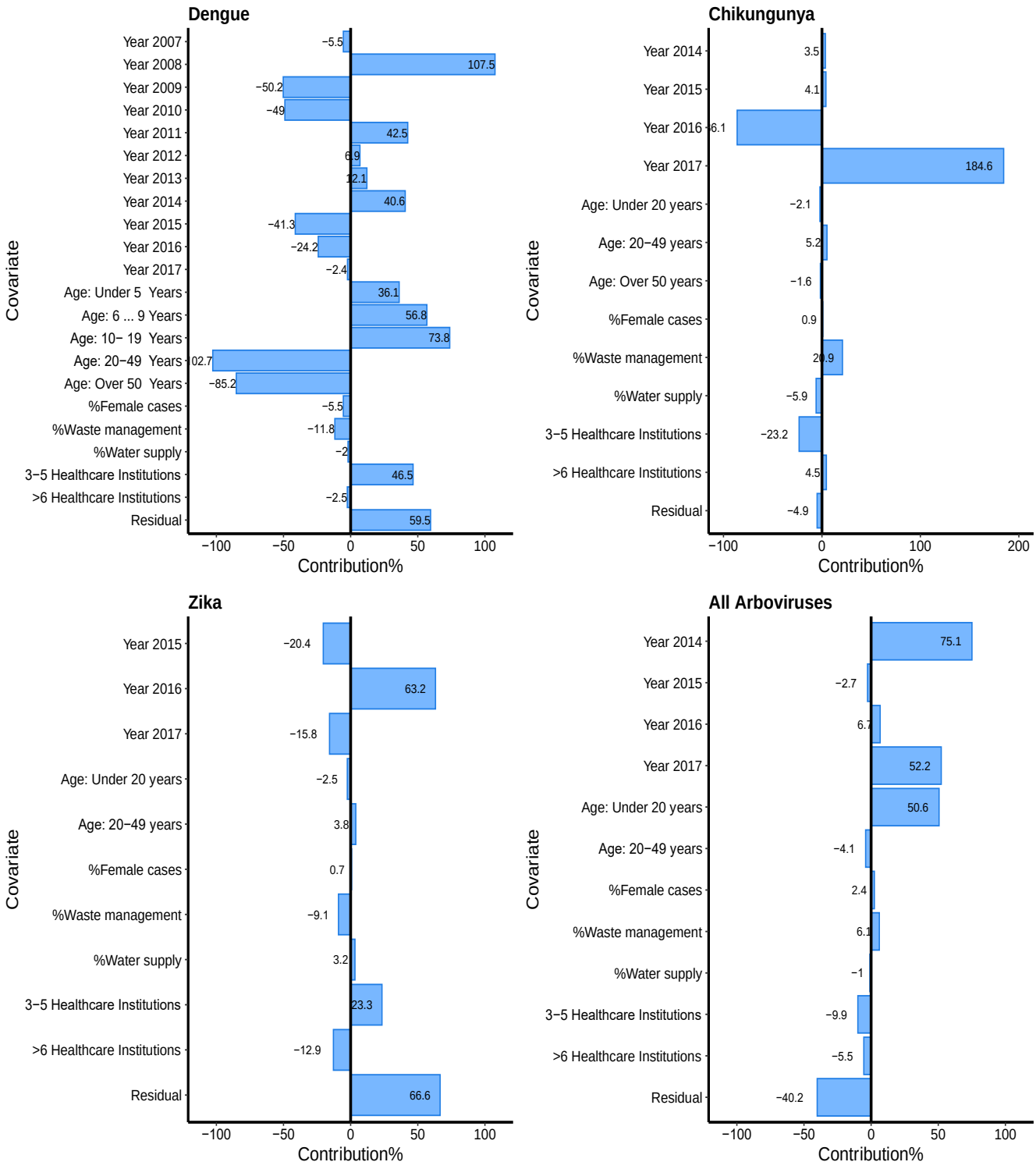


Figure 6.3. A. Contribution of covariates to the overall relative inequality on Dengue, Chikungunya and Zika in Fortaleza

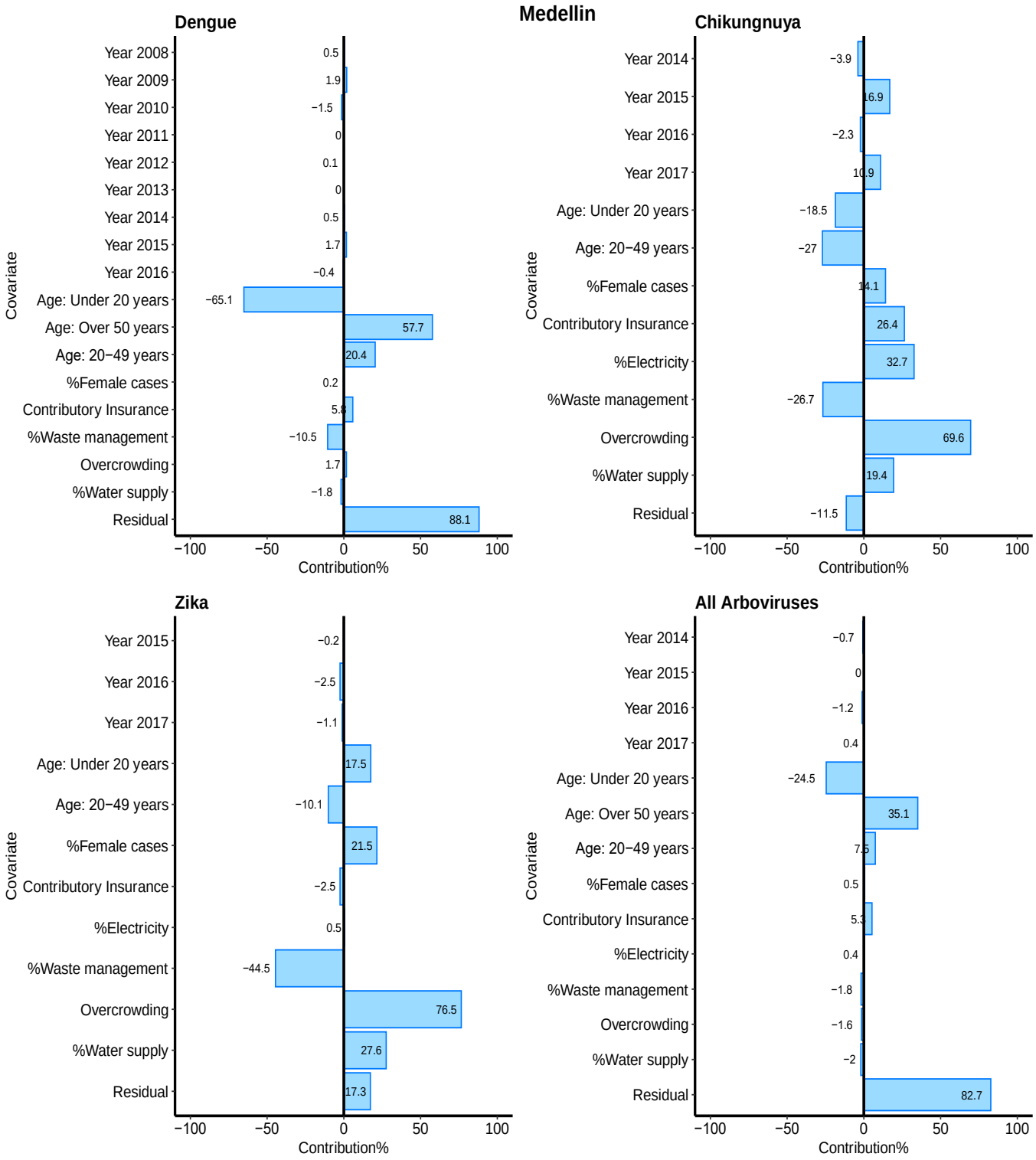


Figure 6-3. B. Contribution of covariates to the overall relative inequality on Dengue, Chikungunya and Zika in Medellin

DISCUSSION

We assessed the presence of socioeconomic inequalities on arboviruses in Brazil and Colombia using surveillance data from 2007 to 2017, providing estimates of the contributions of some measured socio-environmental factors to the presence of the socioeconomic inequality.

We observed a constant presence of all arboviruses in both municipalities, but disease rates were higher in Brazil compared to Colombia. We corroborate the endemic character of dengue in both study sites[6,9,11]. However, highlighting the increased burden of notified arboviruses in Brazil, possibly due to higher temperatures, population density and limited waste management, as previously described[10,11,13]. Despite the observation that chikungunya and Zika affected Colombia significantly, Medellin had fewer cases than the national average and showed a wide disease distribution across the city[9]. Fortaleza, in contrast, presented one of the largest chikungunya outbreaks in Brazil while the number of Zika cases were comparable to those reported in the country[10-13]. Despite the small magnitude of the RCIs, we consistently observed a greater concentration of dengue and Zika cases among the poor in both municipalities. However, the relative inequality in Fortaleza was larger than for Medellin. This difference could be attributed to a more heterogenous socioeconomic distribution in Fortaleza compared to Medellin but also to the fact that the burden of notified cases in Fortaleza is on average higher than in Medellin[8-10].

The year of notification impacted the measures of inequality for all outcomes and in both study sites, but of larger magnitude in Fortaleza. The yearly stratified analysis for chikungunya and aggregated arboviruses showed RCIs covering the line of equality during 2014 and 2015, indicating a non-differential distribution of disease rates across the socioeconomic distribution. In Medellin, chikungunya and Zika were widely spread across the city and SES strata, with very few cases and small relative inequality. These findings could be attributed to the fact that chikungunya and Zika were newly introduced arboviruses that started with outbreaks, where everyone was exposed in the same way, regardless their SES or other individual and neighborhood-specific characteristics[4,11,13]. The changes across time and magnitude of the relative inequality are also indicators of the epidemic nature of these arboviruses[6,11]. Thus, indicating that the presence of

inequalities (when they exist), are more evident during outbreaks of already stablished diseases such as dengue, and showing a broader scope across the socioeconomic strata during the introduction of new diseases as in the case of chikungunya and Zika. These findings stress the need to account for the temporal distribution of the outcome when analyzing the presence of health inequalities of epidemic diseases.

Age and the presence of more than three healthcare facilities in the neighborhood accounted for the majority of the measured inequality in Fortaleza. In Medellin, the main contributors to the overall inequality towards the poor were age and waste management. The contribution of age could be attributed to the known differential pattern of arboviruses across age groups: dengue is most likely present among children and young adults in endemic populations, which is the case of our study sites[1,3,6], chikungunya among mid-age and older adults, and Zika being a concern among women during reproductive age [2,3,5,12].

The contribution of the presence of healthcare facilities in Fortaleza and the contributory insurance scheme in Medellin could be explained by their role as healthcare access indicators. It has been reported that physical access (geographical distance and healthcare facilities availability) and the type of health system (public vs. private) are related to differential outcomes for dengue[2,4,8,11,15,16,30,31]. However, we used surveillance data and our study population included individuals who were able to seek and receive care, which could indicate a differential ascertainment and reporting of cases to national surveillance system[32]. Although the presence of healthcare facilities as an indicator of healthcare access could contribute in either direction to the inequality, the use of surveillance data likely underestimates the true burden of cases among the poor[19]. Similarly, although the contributory insurance (which is similar to a private health care system) could contribute to the inequality towards the better-off, the potential underreporting due to the use of surveillance data possibly moves our estimates towards the line of equality.

The contribution of sanitation factors such as waste management and water supply is justified given their main roles in the presence and distribution of *Aedes* mosquitoes[3,33]. The presence of

arboviruses in areas of limited waste management and water supply is widely known given their potential for increasing breeding sites of *Aedes* mosquitoes[3,33]. Likewise, the contribution of overcrowding, which is more likely seen in low-SES settings, has been associated to disease transmission even in areas with low entomological indexes, which is the case of Medellin[3-6,8,14,15,30].

The small contribution of female cases to the overall inequality could be explained by a differential pattern of health seeking behavior. Specifically, pregnant women or women in reproductive ages may have consulted more, increasing the likelihood of diagnosis and therefore its notification, particularly after the launch of the PHEIC in 2016 for Zika[4,5,18,34]. However, although overreporting among women is a plausible explanation, an actual increased risk of negative outcomes among women could not completely be ruled-out[4,11,32].

Strengths and Limitations:

As indicated above, the use of surveillance data in our analysis limits its generalizability and the interpretation should be conditional on individuals who seek care and who were notified to either surveillance system. There is also a concern related to underreporting, mostly due to the lack of notification from private institutions. Although notification is mandatory, private institutions were reported as incompletely compliant in previous studies [12,32,34]. Since receiving healthcare at private institutions is positively related to the SES measure evaluated here, conditioning on surveillance in the presence of differential underreporting from private healthcare providers would have resulted in biased estimates towards equality, altogether potentially underestimating the inequality.

The introduction of chikungunya in 2014 and Zika in 2015 presents a risk of misclassification of the outcome. Given the similarity of symptoms between the studied arboviruses and the endemicity of dengue, newly introduced chikungunya and Zika cases could have been misdiagnosed as dengue early during their introduction[2,32,34]. In order to avoid an over or underestimation of the inequality due to the misclassification, we grouped dengue, chikungunya and Zika cases and estimated the

aggregated arboviral distributions, RCIs and contributions, considering this outcome as less sensitive to potential misclassification.

To avoid some methodological issues presented in other analysis of socioeconomic distribution and arboviruses [11,14-16]; we used a large sample size, accounted for the spatiotemporal nature of the diseases and tried to account for the use of surveillance data and misclassification. However, it is possible that our models did not capture completely all possible determinants of the inequality, leaving some of the inequality unexplained, which is a well-known limitation for the regression-based decompositions[25,26,28]. Likewise, we observed a non-monotonic disease distribution with instances were more cases were around the middle SES measures, moving the RCIs towards zero, indicating small or no disparity[14,19,35]. This was corroborated with our sensitivity analysis, where we observed that indeed, the RCI were larger in magnitude in presence of monotonic associations between SES and disease rates.

Nonetheless, our approach presents relevant information that would have been otherwise missed by standard analysis [14,15,31] and we consider that our study provides insightful information about the presence of health inequalities and offers quantitative estimates of the contribution of some known determinants on arboviral distribution in Latin America that have not been presented before.

Conclusion

Our study presents quantitative estimates of the socioeconomic inequality among arboviruses and its decomposition, accounting for the spatiotemporal distribution in two Latin American cities. We corroborate the concentration of arboviral diseases on low socioeconomic neighborhoods and identify that year of occurrence, age, presence of healthcare facilities, and waste management are key determinants of the heterogenous distribution of endemic arboviruses across the spectrum of socioeconomic status. Our results contribute to the body of evidence on health inequalities and could be used to design and implement targeted strategies to decrease health inequalities and or for disease control at local level.

Conflict of interest: None declared.

Funding: This study did not receive any specific funding.

Computing code: The R-script used to conduct the analysis in this manuscript is provided as supplementary material.

Data availability: Data used in this manuscript could be obtained by official requests to the public health offices/ local Ministry of Health from Fortaleza, Brazil and in Medellin, Colombia.

Acknowledgments: We acknowledge the collaboration of the Public Health and the Surveillance office in Fortaleza, Brazil and in Medellin, Colombia. Mabel Carabali holds a CIHR-Banting-Best Doctoral Award fellowship.

REFERENCES

1. Mayer SV, Tesh RB, Vasilakis N. The emergence of arthropod-borne viral diseases: A global prospective on dengue, chikungunya and Zika fevers. *Acta Trop* 2017;**166**(Supplement C):155-163.

2. Pan American Health Organization (PAHO). *Tool for the diagnosis and care of patients with suspected arboviral diseases*. 1st Edition ed. Vol. PAHO Strategic Plan 2014-2019. Washington: PAHO, 2017.

3. MacCormack-Gelles B, Lima Neto AS, Sousa GS, Nascimento OJ, Machado MMT, Wilson ME, Castro MC. Epidemiological characteristics and determinants of dengue transmission during epidemic and non-epidemic years in Fortaleza, Brazil: 2011-2015. *PLOS Neglected Tropical Diseases* 2018;**12**(12):e0006990.

4. Borchering RK, Huang AT, Mier-y-Teran-Romero L, Rojas DP, Rodriguez-Barraquer I, Katzelnick LC, Martinez SD, King GD, Cinkovich SC, Lessler J, Cummings DAT. Impacts of Zika emergence in Latin America on endemic dengue transmission. *Nature Communications* 2019;**10**(1):5730.

5. Patterson J, Sammon M, Garg M. Dengue, Zika and Chikungunya: Emerging Arboviruses in the New World. *Western Journal of Emergency Medicine* 2016;**17**(6):671-679.

6. Pan American Health Organization / World Health Organization. Epidemiological Update: Dengue. Washington, D.C.: PAHO/WHO, 2019;15.

7. Villar LA, Rojas DP, Besada-Lombana S, Sarti E. Epidemiological trends of dengue disease in Colombia (2000–2011): a systematic review. *PloS Negl Trop Dis* 2015;**9**.

8. Instituto Nacional de Salud (INS). Informe epidemiologico de Evento Dengue. In: Vigilancia y Análisis del Riesgo en Salud Pública., ed. Vol. FOR-R02.4000-001. Bogota, Colombia: Instituto Nacional de Salus, 2018;22.

9. Rico-Mendoza A, Alexandra P-R, Chang A, Encinales L, Lynch R. Co-circulation of dengue, chikungunya, and Zika viruses in Colombia from 2008 to 2018. *Pan American journal of public health* 2019;**43**:e49-e49.

10. Governo do Estado do Ceara. Boletim Epidemiologico: Dengue, Chikungunya e Zika. In: Fortaleza RdS, ed. *Coordenadoria de Promoção e Proteção à Saúde* Fortaleza: Secretaria da Saúde do Estado do Ceará – Região de Saúde de Fortaleza, 2017;12.

11. Paixão ES, Teixeira MG, Rodrigues LC. Zika, chikungunya and dengue: the causes and threats of new and re-emerging arboviral diseases. *BMJ Global Health* 2018;**3**(Suppl 1):e000530.

12. Lima Neto AS, Sousa GS, Nascimento OJ, Castro MC. Chikungunya-attributable deaths: A neglected outcome of a neglected disease. *PloS neglected tropical diseases* 2019;**13**(9):e0007575-e0007575.

13. Simião AR, Barreto FkdA, Oliveira RdMAB, Cavalcante JW, Lima Neto AS, Barbosa RB, Lins CdS, Meira AG, Araújo FMdC, Lemos DRQ, Alencar CH, Cavalcanti LPdG. A major chikungunya epidemic

with high mortality in northeastern Brazil. *Revista da Sociedade Brasileira de Medicina Tropical* 2019;52.

14. Farinelli EC, Baquero OS, Stephan C, Chiaravalloti-Neto F. Low socioeconomic condition and the risk of dengue fever: A direct relationship. *Acta Tropica* 2018;180:47-57.

15. Rodrigues NCP, Daumas RP, de Almeida AS, dos Santos RS, Koster I, Rodrigues PP, Gomes MdF, Macedo AdF, Gerardi A, Leite IdC. Risk factors for arbovirus infections in a low-income community of Rio de Janeiro, Brazil, 2015-2016. *PLOS ONE* 2018;13(6):e0198357.

16. Mulligan K, Dixon J, Joanna Sinn C-L, Elliott SJ. Is dengue a disease of poverty? A systematic review. *Pathogens and Global Health* 2015;109(1):10-18.

17. IBGE, Instituto Brasileiro de Geografia e Estatística. Atlas do Censo Demografico Brasileiro 2010. . In: SIDRA, ed. Vol. Banco de Tabelas Estadisticas. Rio de Janeiro, Brazil: Ministerio do Planejamento, Orçamento e Gestao., 2010.

18. Ministério da Saúde, Secretaria de Atenção à Saúde. Protocolo de atenção à saúde e resposta à ocorrência de microcefalia relacionada à infecção pelo vírus Zika *Atenção à Saúde* Vol. CDU 616-022. Brasilia, Brazil: Ministério da Saúde, Secretaria de Atenção à Saúde, 2016;42.

19. Toan NT, Rossi S, Prisco G, Nante N, Viviani S. Dengue epidemiology in selected endemic countries: factors influencing expansion factors as estimates of underreporting. *Trop Med Int Health.* 2015;20(7):840-863.

20. Departamento Nacional de Estadistica de Colombia (DANE). Estimaciones de población 1985 – 2005 y proyecciones de población 2005 – 2020 total municipal por área. *Departamento Nacional de Estadistica de Colombia* DANE, 2015.

21. Ministerio de la Protección Social. Decreto Numero 3518 de 2006. Creacion y reglamentacion del Sistema de Vigilancia en Salud Publica- SIVIGILA. In: Social. MdlP, ed. Vol. 3518/2006. Bogota, Colombia., 2006;1-17.

22. Brasil, Ministerio da Saude. Portaria No 47, 3 de Maio de 2016. Define os parâmetros para monitoramento da regularidade na alimentação do Sistema de Informação de Agravos de Notificação (SINAN). In: Ministerio da Saude SdVeS, ed. Vol. Portaria No. 47, 2016. Brasilia, 2016.

23. Departamento Administrativo Nacional de Estadística (DANE). Estimaciones de población 1985–2005 y proyecciones de población 2005–2020 total municipal por área. . Vol. Vital Statistics. Bogota, Colombia: DANE, 2016.

24. Departamento Administrativo Nacional de Estadistica (DANE). Metodologia de la Estratificacion Socioeconomica Urbana para Servicios Publicos Domicioliarios. In: Direccion de Geoestadistica, ed. Grupo de Estratificacion ed. Santa Fe de Bogota, Colombia.: DANE, 2015;96.

25. Speybroeck N, Konings P, Lynch J, Harper S, Berkvens D, Lorant V, Geckova A, Hosseinpoor AR. Decomposing socioeconomic health inequalities. *International Journal of Public Health* 2010;55(4):347-351.

26. Erreygers G, Kessels R. Regression-Based Decompositions of Rank-Dependent Indicators of Socioeconomic Inequality of Health. In: Kessels R, ed. *Health and Inequality*. Research on Economic Inequality. Vol. 21 Emerald Group Publishing Limited, 2013;227-259.

27. Blangiardo M, Cameletti M. Bayesian regression and hierarchical models. *Spatial and Spatio-temporal Bayesian Models with R-INLA* John Wiley & Sons, Ltd, 2015;127-172.

28. Wagstaff AD, van Eddy Watanabe, Naoko. On Decomposing the Causes of Health Sector Inequalities with an Application to Malnutrition Inequalities in Vietnam: The World Bank; 2001. 26 p.

29. Konings P, Speybroeck N. decomp: Various functions to quantify and decompose health inequalities. 0.3 ed CRAN, 2012;This package contains functions to calculate the relative concentration index, plot Lorenz curves and decompose inequality using (generalized) linear models and survival models while taking survey design into account.

30. Piedrahita LD, Agudelo Salas IY, Marin K, Trujillo AI, Osorio JE, Arboleda-Sanchez SO, Restrepo BN. Risk Factors Associated with Dengue Transmission and Spatial Distribution of High Seroprevalence in Schoolchildren from the Urban Area of Medellin, Colombia. *Canadian Journal of Infectious Diseases and Medical Microbiology* 2018;**2018**:11.

31. Bavia L, Melanda FN, de Arruda TB, Mosimann ALP, Silveira GF, Aoki MN, Kuczera D, Sarzi ML, Junior WLC, Conchon-Costa I, Pavanelli WR, Duarte dos Santos CN, Barreto RC, Bordignon J. Epidemiological study on dengue in southern Brazil under the perspective of climate and poverty. *Scientific Reports* 2020;**10**(1):2127.

32. Chow A, Ho H, Win M-K, Leo Y-S. Assessing Sensitivity and Specificity of Surveillance Case Definitions for Zika Virus Disease. *Emerging infectious diseases* 2017;**23**(4):677-679.

33. Longbottom J, Krause A, Torr SJ, Stanton MC. Quantifying geographic accessibility to improve efficiency of entomological monitoring. *PLOS Neglected Tropical Diseases* 2020;**14**(3):e0008096.

34. Braga JU, Bressan C, Dalvi APR, Calvet GA, Daumas RP, Rodrigues N, Wakimoto M, Nogueira RMR, Nielsen-Saines K, Brito C, Bispo de Filippis AM, Brasil P. Accuracy of Zika virus disease case definition during simultaneous Dengue and Chikungunya epidemics. *PLOS ONE* 2017;**12**(6):e0179725.

35. Moloughney B. What can public health do to address inequities in infectious disease? *Canada Communicable Disease Report* 2016;**42**(S1):S1_14.

6.3 Supplementary Material Manuscript 3

Decomposition of Socioeconomic Inequalities on Arboviral Diseases in Brazil and Colombia (2007- 2017)

Table S1.A. Overall, Partial Relative concentration Index (RCIs) and their respective percentage contributions to the overall inequality on Dengue in Fortaleza, Brazil (2007-2017)

Dengue – Fortaleza, Brazil		
Overall Concentration Index	RCI = -0.015 (-0.018, -0.012)	
Covariate	Partial Concentration Index (95%Cis)	Contribution (%)
Residual	-	59.5
Year of notification		
2007	-0.694 (-0.725, -0.663)	-5.5
2008	-0.584 (-0.616, -0.553)	107.5
2009	-0.444 (-0.483, -0.405)	-50.2
2010	-0.326 (-0.363, -0.290)	-49.0
2011	-0.205 (-0.233, -0.176)	42.5
2012	-0.030 (-0.058, -0.002)	6.9
2013	0.108 (0.082, 0.134)	12.1
2014	0.186 (0.162, 0.211)	40.6
2015	0.342 (0.321, 0.362)	-41.3
2016	0.511 (0.495, 0.527)	-24.2
2017	0.613 (0.598, 0.627)	-2.4
Age: Under 5 years	-0.153 (-0.175, -0.131)	36.1
Age: 6 – 9years	-0.193 (-0.219, -0.168)	56.8
Age: 10- 19 years	-0.071 (-0.083, -0.058)	73.8
Age: 20-49 years	0.044 (0.037, 0.051)	-102.7
Age: Over 50 years	0.142 (0.125, 0.160)	-85.2
%Female cases	0.008 (0.002, 0.015)	-5.5
%Waste management	0.002 (0.002, 0.002)	-11.8
%Water supply	0.002 (0.002, 0.003)	-2.0
3-5 Healthcare Institutions	-0.170 (-0.189, -0.151)	46.5
>6 Healthcare Institutions	0.035 (-0.008, 0.078)	-2.5

Table S1.B. Overall, Partial Relative concentration Index (RCIs) and their respective percentage contributions to the overall inequality on Chikungunya in Fortaleza, Brazil (2007-2017).

Chikungunya – Fortaleza, Brazil		
Overall Concentration Index	RCI=0.023 (95% CI= 0.017, 0.029)	
Covariate	Partial Concentration Index (95%Cis)	Contribution (%)
Residual	-	-4.9
Year 2014	-0.175 (-0.470, 0.119)	3.5
Year 2015	-0.260 (-0.628, 0.107)	4.1
Year 2016	-0.157 (-0.185, -0.130)	-86.1
Year 2017	0.153 (0.128, 0.179)	184.6
Age: Under 20 years	-0.009 (-0.045, 0.028)	-2.0
Age: 20-49 years	-0.031 (-0.045, -0.018)	5.2
Age: Over 50 years	0.057 (0.035, 0.079)	-1.6
%Female cases	-0.011 (-0.023, 0.000)	0.9
%Waste management	0.003 (0.002, 0.003)	20.9
%Water supply	0.006 (0.005, 0.008)	-5.9
3-5 Healthcare Institutions	-0.247 (-0.286, -0.207)	-23.2
>6 Healthcare Institutions	0.124 (0.020, 0.227)	4.5

Table S1.C. Overall, Partial Relative concentration Index (RCIs) and their respective percentage contributions to the overall inequality on Zika in Fortaleza, Brazil (2007-2017).

Zika – Fortaleza, Brazil		
Overall Concentration Index	RCI=-0.028 (95% CI= -0.040, -0.015)	
Covariate	Partial Concentration Index (95%Cis)	Contribution (%)
Residual		66.6
Year 2015	-0.500 (-0.822, -0.178)	-20.4
Year 2016	-0.086 (-0.115, -0.058)	63.2
Year 2017	0.386 (0.322, 0.451)	-15.8
Age: Under 20 years	-0.055(-0.119, 0.008)	-2.5
Age: 20-49 years	0.022 (-0.008, 0.053)	3.8
Age: Over 50 years	0.005 (-0.087, 0.097)	0.0
%Female cases	0.004 (-0.022, 0.030)	0.7
%Waste management	0.002 (0.002, 0.003)	-9.1
%Water supply	0.004 (0.001, 0.007)	3.2
3-5 Healthcare Institutions	-0.172 (-0.245, -0.100)	23.3
>6 Healthcare Institutions	0.380 (0.217, 0.543)	-12.9

Table S1.D. Overall, Partial Relative concentration Index (RCIs) and their respective percentage contributions to the overall inequality on All Arboviruses in Fortaleza, Brazil (2007-2017).

All Arboviruses – Fortaleza, Brazil		
Overall Concentration Index	RCI = 0.058 (95% CI = 0.052, 0.065)	
Covariate	Partial Concentration Index (95%Cis)	Contribution (%)
Residual	-	-40.2
Year 2014	-0.425 (-0.460, -0.391)	75.1
Year 2015	-0.179 (-0.210, -0.149)	-2.7
Year 2016	0.145 (0.121, 0.169)	6.7
Year 2017	0.363 (0.342, 0.385)	52.2
Age: Under 20 years	-0.139 (-0.157, -0.121)	50.6
Age: 20-49 years	0.005 (-0.004, 0.015)	-4.1
Age: Over 50 years	0.166 (0.146, 0.186)	-29.8
%Female cases	0.013 (0.004, 0.023)	2.4
%Waste management	0.003 (0.003, 0.004)	6.1
%Water supply	0.004 (0.003, 0.005)	-1.0
3-5 Healthcare Institutions	0.054 (0.045, 0.064)	-9.9
>6 Healthcare Institutions	-0.211 (-0.239, -0.184)	-5.5

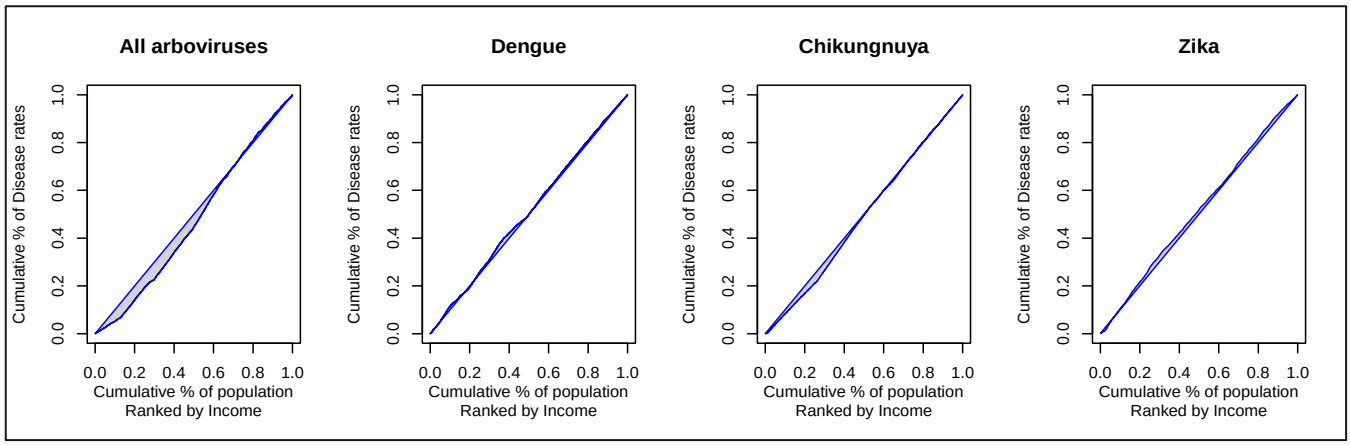


Figure S1. Relative Concentration curve (RCI) for each arboviral disease in Fortaleza

Table S2.A. Overall, Partial Relative concentration Index (RCIs) and their respective percentage contributions to the overall inequality on Dengue in Medellin, Colombia (2008-20016).

Dengue – Medellin, Colombia		
Overall Concentration Index	RCI = -0.017 (95% CI= -0.021, -0.013)	
Covariate	Partial Concentration Index (95%Cis)	Contribution (%)
Residual	-	88.1
Year 2008	0.014 (-0.036, 0.064)	0.5
Year 2009	0.057 (0.006, 0.108)	1.9
Year 2010	0.034 (0.011, 0.057)	-1.5
Year 2011	-0.045 (-0.097, 0.007)	0.0
Year 2012	0.040 (-0.015, 0.095)	0.1
Year 2013	0.001 (-0.034, 0.037)	0.0
Year 2014	-0.024 (-0.053, 0.004)	0.5
Year 2015	-0.047 (-0.075, -0.019)	1.7
Year 2016	0.002 (-0.017, 0.021)	-0.4
Age: Under 20 years	-0.095 (-0.113, -0.077)	-65.1
Age: 20-49 years	0.021 (0.009, 0.033)	20.4
Age: Over 50 years	0.147(0.119, 0.176)	57.7
%Female cases	0.012 (0.003, 0.021)	0.2
%Waste management	0.011 (0.010, 0.012)	-10.5
%Water supply	0.007 (0.006, 0.008)	-1.8
Contributory Insurance	0.052 (0.047, 0.057)	5.8
Overcrowding	-0.159 (-0.170, -0.148)	1.7

Table S2.B. Overall, Partial Relative concentration Index (RCIs) and their respective percentage contributions to the overall inequality on Chikungunya in Medellin, Colombia (2008-20016).

Chikungunya – Medellin, Colombia		
Overall Concentration Index	RCI=0.002 (95% CI= -0.001, 0.004)	
Covariate	Partial Concentration Index (95%Cis)	Contribution (%)
Residual	-	-11.5
Year 2014	-0.175 (-0.470, 0.119)	-3.9
Year 2015	-0.260 (-0.628, 0.107)	16.9
Year 2016	-0.157 (-0.185, -0.130)	-2.3
Year 2017	0.153 (0.128, 0.179)	10.9
Age: Under 20 years	-0.042 (-0.148, 0.064)	-18.5
Age: 20-49 years	-0.015(-0.052, 0.023)	-27.0
%Electricity	0.005 (0.003, 0.007)	32.7
%Female cases	-0.034 (-0.072, 0.004)	14.1
%Waste management	0.005 (-0.001, 0.011)	-26.7
%Water supply	0.002 (-0.002, 0.006)	19.4
Contributory Insurance	0.046 (0.023, 0.068)	26.4
Overcrowding	-0.259 (-0.304, -0.215)	69.6

Table S2.C. Overall, Partial Relative concentration Index (RCIs) and their respective percentage contributions to the overall inequality on Zika in Medellin, Colombia (2008-20016).

Zika – Medellin, Colombia		
Overall Concentration Index	RCI=-0.004 (95% CI= -0.008, -0.001)	
Covariate	Partial Concentration Index (95%Cis)	Contribution (%)
Residual	-	17.3
Year 2015	0.008(-0.156, 0.172)	-0.2
Year 2016	-0.002 (-0.023, 0.019)	-2.5
Year 2017	0.018 (-0.234, 0.270)	-1.10
Age: Under 20 years	-0.023 (-0.109, 0.063)	17.5
Age: 20-49 years	0.007 (-0.033, 0.047)	-10.1
%Electricity	0.003 (0.001, 0.006)	0.5
%Female cases	-0.039 (-0.076, -0.001)	21.5
%Waste management	0.009 (0.006, 0.013)	-44.5
%Water supply	0.009 (0.005, 0.010)	27.6
Contributory Insurance	0.042, (0.017, 0.067)	-2.5
Overcrowding	-0.199 (-0.240, -0.158)	76.5

Table S2.D. Overall, Partial Relative concentration Index (RCIs) and their respective percentage contributions to the overall inequality on Zika in Medellin, Colombia (2008-20016).

Medellin, Colombia		
Overall Concentration Index	RCI=0.002 (95% CI= -0.001, 0.004)	
Covariate	Partial Concentration Index (95%Cis)	Contribution (%)
Residual	-	-11.5
Year 2014	-0.175 (-0.470, 0.119)	-3.9
Year 2015	-0.260 (-0.628, 0.107)	16.9
Year 2016	-0.157 (-0.185, -0.130)	-2.3
Year 2017	0.153 (0.128, 0.179)	10.9
Age: Under 20 years	-0.042 (-0.148, 0.064)	-18.5
Age: 20-49 years	-0.015(-0.052, 0.023)	-27.0
%Electricity	0.005 (0.003, 0.007)	32.7
%Female cases	-0.034 (-0.072, 0.004)	14.1
%Waste management	0.005 (-0.001, 0.011)	-26.7
%Water supply	0.002 (-0.002, 0.006)	19.4
Contributory Insurance	0.046 (0.023, 0.068)	26.4
Overcrowding	-0.259 (-0.304, -0.215)	69.6

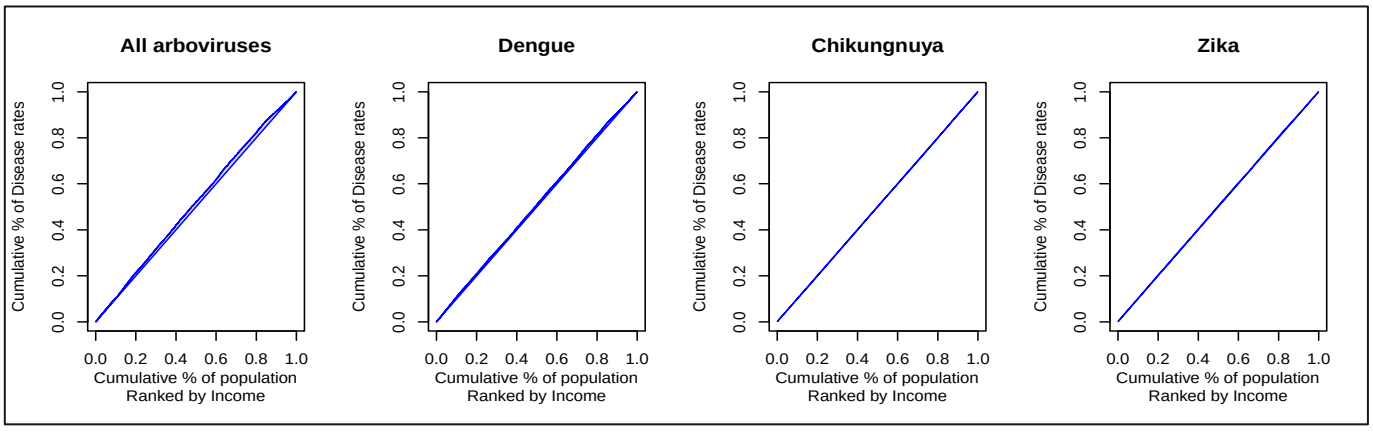


Figure S2. Relative Concentration curve (RCI) for each arboviral disease in Medellin.

Table S.3. Crude and Adjusted Overall RCI and 95% point-wise intervals from Fortaleza and Medellin.

Fortaleza	Crude RCI			Adjusted RCI		
Disease	**RCI**	**2.5%**	**97.5%**	**RCI**	**2.5%**	**97.5%**
All Arboviruses	0.091	0.064	0.119	0.058	0.052	0.065
Dengue	-0.116	-0.138	-0.095	-0.015	-0.018	-0.012
Chikungunya	0.022	-0.020	0.063	0.023	0.017	0.029
Zika	-0.097	-0.152	-0.043	-0.028	-0.040	-0.015
Medellin	**Crude RCI**			**Adjusted RCI**		
Disease	**RCI**	**2.5%**	**97.5%**	**RCI**	**2.5%**	**97.5%**
All Arboviruses	-0.083	-0.100	-0.065	-0.035	-0.041	-0.030
Dengue	-0.049	-0.065	-0.034	-0.017	-0.021	-0.013
Chikungunya	-0.020	-0.041	0.000	0.002	0.000	0.004
Zika	0.008	-0.014	0.031	-0.004	-0.008	-0.001

Table S.4. Overall RCI and 95% point-wise intervals from the stratified analysis for Chikungunya and Aggregated Arboviruses in Fortaleza.

	Overall RCI for All Arboviruses			Overall RCI for Chikungunya		
Year	**RCI**	**2.5%**	**97.5%**	**RCI**	**2.5%**	**97.5%**
2014	-0.01	-0.03	0.00	-0.01	-0.03	0.01
2015	-0.02	-0.03	0.00	-0.02	-0.03	0.00
2016	-0.04	-0.05	-0.03	-0.03	-0.04	-0.01
2017	-0.02	-0.03	-0.01	-0.01	-0.02	0.00

Table S.3. Stratified analysis for chikungunya in Fortaleza

2014	Contribution (%)	Concentration Index	2.5%	97.5%
Residual	10.3	-	-	-
%Female cases	-66.49	-0.13	-0.31	0.04
Age: Under 20 years	-63.00	0.45	0.14	0.77
Age: 20-49 years	6.76	-0.09	-0.21	0.02
Age: Over 50 years	17.27	0.09	-0.25	0.43
%Waste management	82.06	0.00	0.00	0.01
%Water supply	21.51	0.01	-0.02	0.04
3-5 Healthcare Institutions	60.06	-0.72	-0.96	-0.48
>6 Healthcare Institutions	31.55	-0.08	-0.42	0.26
2015				
Residual	-17.0	-	-	-
%Female cases	-37.32	0.06	-0.06	0.17
Age: Under 20 years	42.82	-0.26	-0.68	0.16
Age: 20-49 years	12.93	0.11	-0.14	0.35
Age: Over 50 years	1.53	0.01	-0.39	0.41
%Waste management	65.61	0.00	0.00	0.01
%Water supply	-7.94	0.01	0.00	0.03
3-5 Healthcare Institutions	2.37	0.14	-0.43	0.71
>6 Healthcare Institutions	36.98	-0.95	-1.05	-0.85
2016				
Residual	103.6	-	-	-
%Female cases	-0.32	-0.01	-0.02	0.01
Age: Under 20 years	0.62	-0.02	-0.08	0.04
Age: 20-49 years	-1.41	-0.03	-0.05	-0.01
Age: Over 50 years	-1.53	0.05	0.02	0.08
%Waste management	-9.69	0.00	0.00	0.00
%Water supply	-10.90	0.01	0.01	0.01
3-5 Healthcare Institutions	17.68	-0.26	-0.31	-0.20
>6 Healthcare Institutions	1.93	0.12	-0.04	0.27
2017				
Residual	59.2	-	-	-
%Female cases	-1.23	0.00	-0.02	0.01
Age: Under 20 years	7.31	-0.04	-0.08	0.01
Age: 20-49 years	8.49	-0.03	-0.05	-0.02
Age: Over 50 years	-9.28	0.08	0.05	0.10
%Waste management	-9.06	0.00	0.00	0.00
%Water supply	11.33	0.01	0.01	0.01
3-5 Healthcare Institutions	31.80	-0.24	-0.30	-0.19
>6 Healthcare Institutions	1.45	0.06	-0.12	0.24

Table S.4. Stratified analysis for Aggregated Arboviruses in Fortaleza

2014	Contribution (%)	Concentration Index	2.5%	97.5%
Residual	-25.48950	-	-	-
%Female cases	-0.01	0.00	-0.02	0.03
Age: Under 20 years	2.22	-0.10	-0.13	-0.06
Age: 20-49 years	0.42	0.02	-0.01	0.04
Age: Over 50 years	13.37	0.18	0.11	0.25
%Waste management	-7.00	0.00	0.00	0.00
%Water supply	3.58	0.01	0.01	0.01
3-5 Healthcare Institutions	75.93	0.06	0.04	0.09
>6 Healthcare Institutions	36.97	-0.22	-0.27	-0.16
2015				
Residual	87.11	-	-	-
%Female cases	-0.07	0.00	-0.02	0.02
Age: Under 20 years	4.24	-0.08	-0.11	-0.05
Age: 20-49 years	-4.96	0.03	0.01	0.05
Age: Over 50 years	2.19	0.06	0.01	0.11
%Waste management	-16.92	0.00	0.00	0.00
%Water supply	-5.83	0.01	0.00	0.01
3-5 Healthcare Institutions	20.17	0.06	0.04	0.08
>6 Healthcare Institutions	14.07	-0.22	-0.27	-0.16
2016				
Residual	92.23	-	-	-
%Female cases	-0.06	0.00	-0.01	0.01
Age: Under 20 years	1.35	-0.03	-0.06	0.01
Age: 20-49 years	0.87	-0.02	-0.03	0.00
Age: Over 50 years	-3.88	0.07	0.04	0.10
%Waste management	-8.30	0.00	0.00	0.00
%Water supply	-1.81	0.01	0.01	0.01
3-5 Healthcare Institutions	-19.22	0.06	0.04	0.08
>6 Healthcare Institutions	38.81	-0.23	-0.28	-0.17
2017				
Residual	47.42	-	-	-
%Female cases	-0.05	0.00	-0.02	0.01
Age: Under 20 years	2.47	-0.05	-0.09	-0.01
Age: 20-49 years	2.72	-0.03	-0.04	-0.01
Age: Over 50 years	0.54	0.10	0.06	0.13
%Waste management	-6.16	0.00	0.00	0.00
%Water supply	14.81	0.01	0.01	0.01
3-5 Healthcare Institutions	-42.60	0.07	0.05	0.09
>6 Healthcare Institutions	80.85	-0.23	-0.28	-0.18

Monotonic vs. Non-monotonic relationship between SES and disease rates

To illustrate the magnitude of the inequality in presence of a monotonic relationship (with concertation of diseases at the lower end of the SES distribution), we conducted a sensitivity analysis using the same rates from our study but using the quantiles of SES and or scaling the SES measure.

Table S6. RCI and 95% point-wise intervals in the observed scenario and what could be expected with similar rates in two monotonic scenarios of associations between SES and disease rates.

	Non-Monotonic (As observed)			Monotonic (Using quintiles of SES)			Monotonic (Scaling the SES quantiles)		
Disease	RCI	2.5%	97.5%	RCI	2.5%	97.5%	RCI	2.5%	97.5%
Dengue	-0.02	-0.02	-0.01	-0.07	-0.07	-0.06	-0.06	-0.07	-0.06
Chikungunya	0.02	0.02	0.03	0.07	0.05	0.09	-0.04	-0.06	-0.01
Zika	-0.03	-0.04	-0.02	-0.07	-0.09	-0.05	-0.10	-0.12	-0.08
All Arboviruses	0.06	0.05	0.06	0.11	0.09	0.12	-0.01	-0.02	0.01

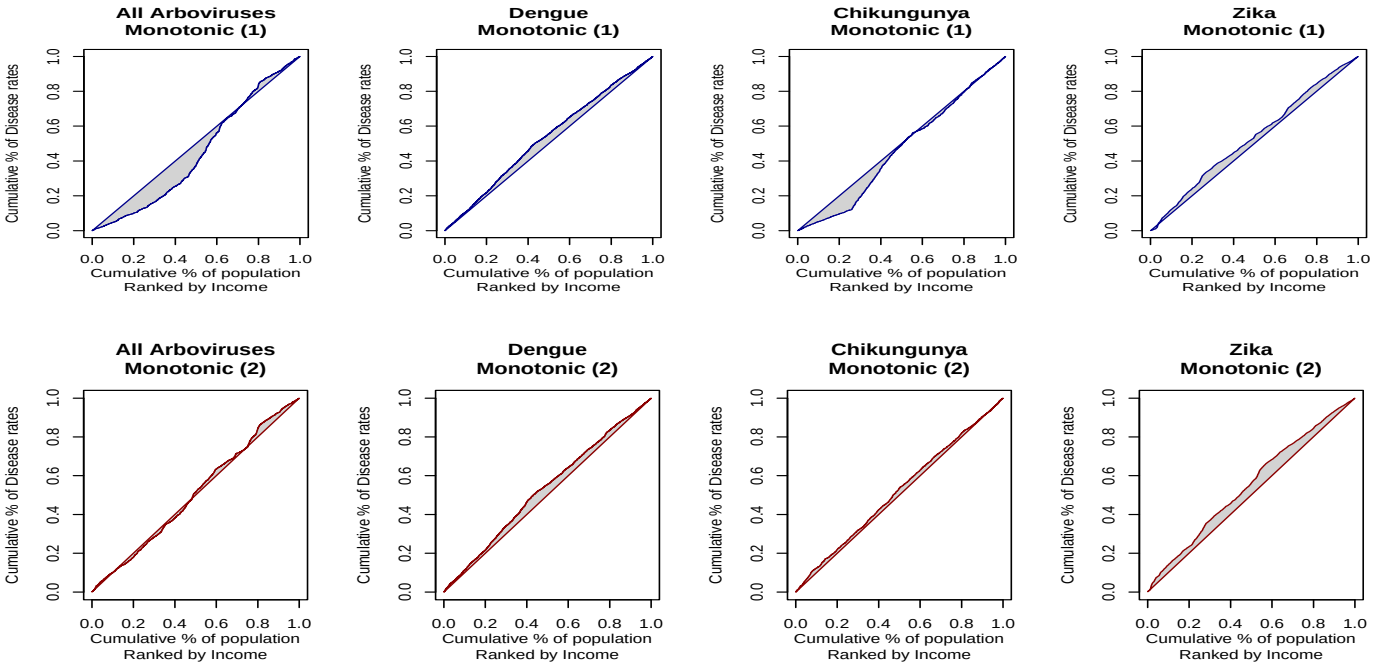
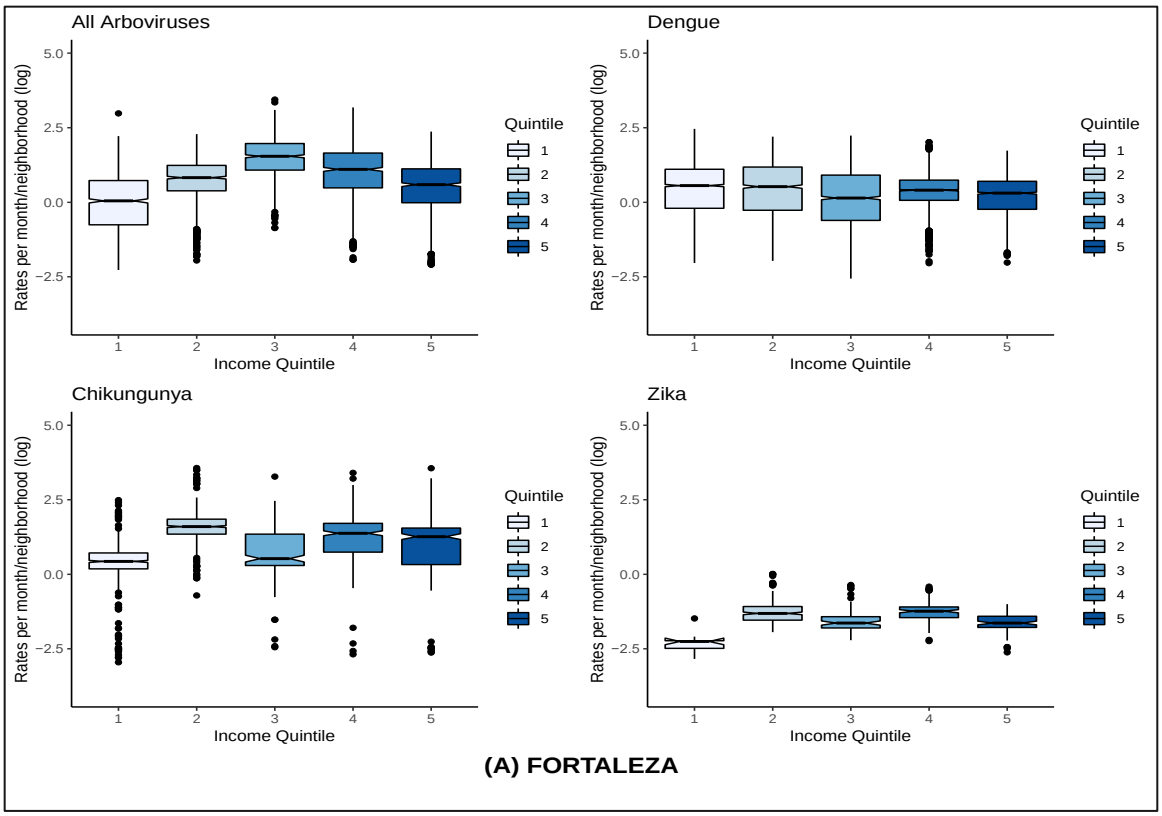


Figure S3. RCI curves for each arboviral disease in Fortaleza under different SES-monotonic scenarios. Top: Scenario (1) of Monotonicity using the quintiles of the median household income. Bottom: Scenario (2) of monotonicity scaling the distribution of the median household income to concentrate disease rates among the first quintiles.



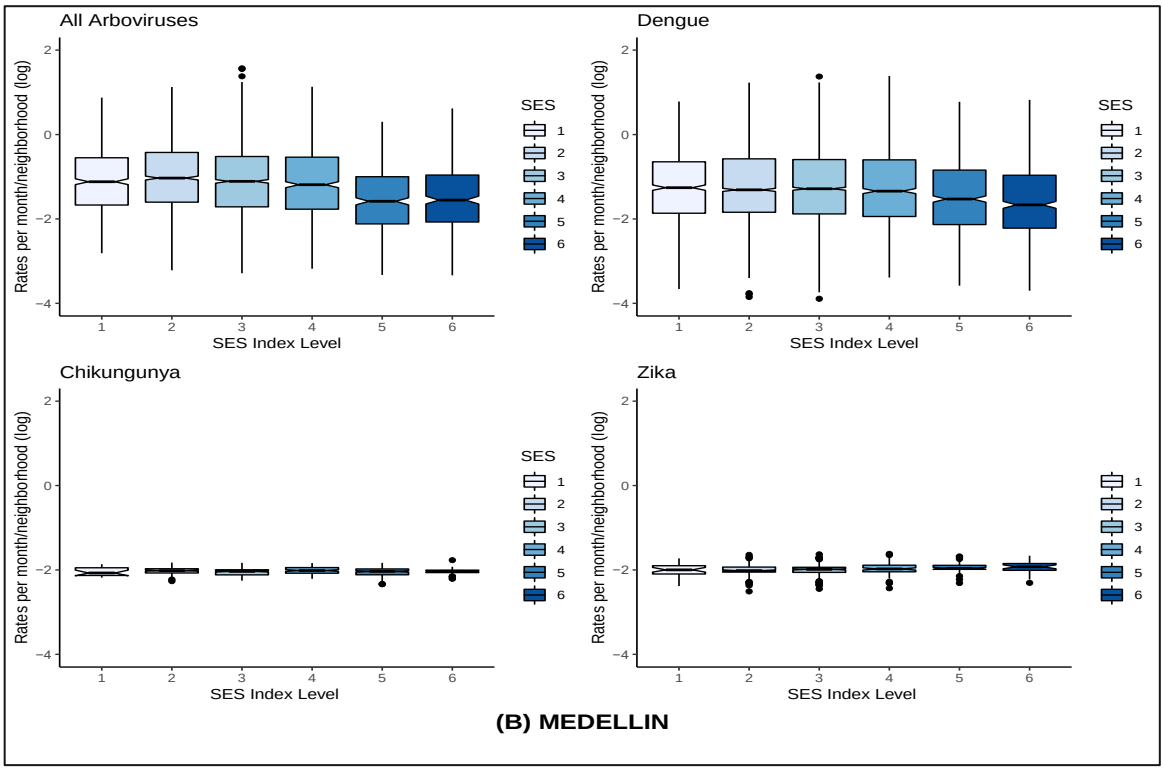


Figure S4. Distribution of adjusted diseases rates (log) by Socioeconomic Status in Fortaleza and Medellin.

Chapter 7: **Dengue, Severity Paradox and Socioeconomic Distribution among Afro-Colombians (Manuscript 4).**

7.1 Preface Manuscript 4

This manuscript examined the presence of ethnic disparities on arboviruses in the Colombian context. This manuscript was motivated by previous reports indicating a "paradoxical" increase on dengue severity among Blacks or people with African ancestry in the Americas. An increased presence of severity and mortality is considered because the literature on arboviruses, including genetic and biomarker-based studies has postulated a low presence of severe dengue among Blacks or people with African ancestry.

Being myself black, clinician and epidemiologist, I am cautious about considering ethnicity or race as a risk or protective factor. I believe that the use of ethnicity/race in epidemiological research should be well informed. Likewise, the use of ethnicity/race in research should be intended to explain the possible effects that this characteristic has, as a construct and not as a biological factor, in the health of individuals and populations beyond any arbitrary standards.

Therefore, in order to contribute to explain the ethnicity paradox in dengue I evaluated the overall and severe dengue distribution between Afro-Colombians and Non-Afro-Colombians from the perspective of intersectionality. I used spatiotemporal methods, assessed the heterogeneity across SES and by ethnic groups, and used several methodological approaches to correct for potential bias related to the use of surveillance data. The results presented here expose some methodological issues associated to the use of ethnicity/race in this type of analysis. These results indicate that the "ethnicity paradox" is likely the result of some socioeconomic factors interacting together and resulting in differential reporting among Afro-Colombians when dengue is severe. This manuscript was submitted to *Epidemiology* journal Manuscript ID: EDE20-0461.

Note: the underreporting assessment used here to inform the corrections made in this manuscript were part of a parallel work currently under peer review at Plos NTD journal (PNTD-D-20-00239).

7.2 Manuscript 4

Type of article: Original Research

Dengue, Severity Paradox and Socioeconomic Distribution among Afro-Colombians

Authors: Mabel Carabali[1*], Mathieu Maheu-Giroux[1], Jay S. Kaufman[1].

Affiliations: Department of Epidemiology, Biostatistics and Occupational Health, McGill University

Corresponding Author:

Mabel Carabali, MD, MSc, PhDI: mabel.carabali@mail.mcgill.ca

Department of Epidemiology, Biostatistics and Occupational Health, McGill University

Purvis Hall, 1020 Pine Avenue West

Montreal, Quebec, Canada H3A 1A2

Tel.: 514-398-6258; Fax: 514-398-4503

Running head: "*Dengue and severity paradox among Afro-Colombians*"

Conflict of interest: None declared.

Funding: This study did not receive any specific funding.

Computing code: The R-script used to conduct the analysis in this manuscript is provided as supplementary material.

Data availability: Data used in this manuscript could be obtained by official requests to the public health offices/ local Ministry of Health from Cali, Colombia.

ABSTRACT

Background: The clinical presentation of dengue ranges from self-limited mild illness to severe forms, including death. African ancestry is often described as protective against dengue severity. However, in the Latin American context, African ancestry has been associated with increased mortality. This "severity paradox" has been hypothesized as resulting from confounding or heterogeneity by socioeconomic status (SES). However, few systematic analyses have been conducted to investigate the presence and nature of the disparity paradox.

Methods: We fit Bayesian hierarchical spatiotemporal models using individual-level surveillance data from Cali, Colombia (2012-2017), to assess the overall morbidity and severity burden of notified dengue. We fitted overall and ethnic-specific models to assess the presence of heterogeneity by SES across and within ethnic groups (Afro-Colombian vs. Non-Afro-Colombians), conducting sensitivity analysis to account for potential underreporting.

Results: Our study included 65,402 dengue cases and 13,732 (21%) hospitalizations. Overall notified dengue incidence rates did not change across ethnic groups. Severity risk was higher among Afro-Colombians (RR=1.16; 95%Credible Interval [95%CrI]: 1.08-1.24) but after accounting for underreporting by ethnicity this association disappeared (RR=1.02; 95%CrI: 0.97-1.07). Subsidized health insurance and low-SES were associated with increased overall dengue rates and severity.

Conclusion: The paradoxical increased severity among Afro-Colombians can be attributed to differential health-seeking behaviors and reporting among Afro-Colombians. Such differential reporting can be understood as a kind of intersectionality between SES, insurance scheme, and ethnicity that requires a quantitative assessment in future studies.

Keywords: dengue, health inequalities, ethnic disparities, Flavivirus, spatiotemporal analysis, Colombia, intersectionality.

Background

Dengue is the fastest spreading arthropod-borne viral disease and a major cause of morbidity and mortality worldwide[1-4]. The Americas region experiences the second largest burden of notified dengue, with over two million cases reported in 2015 and 3.1 million cases in 2019[4-7]. Dengue's clinical presentation ranges from asymptomatic, subclinical or self-limited mild febrile illness to severe forms including hemorrhage, shock, and death[1,3,6]. The disease is caused by any of four closely related serotypes (DENV1, DENV2, DENV3, and DENV4) which provide serotype-specific life-long immunity after infection[3,6,8]. Biological factors such as age, comorbidities, biomarkers, secondary infections (by different serotypes) and circulation of new serotypes, have been associated with increased severity of dengue[6,8-10]. Likewise, social determinants of health, including income and healthcare access, have been associated with dengue incidence and severity[10-15].

Self-declared Black ethnicity and African ancestry are considered "protective" for severe forms of dengue[8,16-26]. The rare presence of severe dengue in Haiti and West Africa and ancestry studies on dengue from Cuba, Brazil and Colombia have been used to support the presence of a decreased likelihood of severe dengue among these populations[16-20]. Cuban studies reported reductions of 60% in the odds of dengue severity for a 50% increase in African ancestry[18-20]. Brazilian studies reported lower odds of dengue severity with increasing percentages of African ancestry (0.13; 95%CI=0.02, 0.69) and for self-identified Black ethnicity (OR=0.28; 95%CI=0.10, 0.81), even after adjusting for socioeconomic factors[21,22]. However, the literature on ethnicity and socioeconomic status (SES) as drivers of dengue infection remains limited, contradictory and or imprecise[10,11,26,27]. For instance, while studies indicating the "protective" effect of African ancestry showed increasing odds of severity with increasing income index (0R=9.6; 95%CI=1.2, 79.7)[21,22], other studies reported higher severity and mortality among Black people (OR=1.52; 95%CI=1.25, 1.84) and low SES individuals[11,28].

In Colombia, dengue incidence ranged 99-493 cases per 100,000 inhabitants during the last decade[9,29]. Dengue and ancestry studies in Colombia reported a 3% decrease in the odds of

dengue severity per every 1% increase in African ancestry (OR=0.97; 95%C=0.95–0.99) and an average 80% lower odds of severity (OR=0.20; 95%CI=0.06–0.64) for a 50% increase in African ancestry, even in admixture populations[24,25]. According to the national surveillance system (SIVIGILA), the proportion of dengue cases among Afro-Colombians varied between 1.7%-3.4% of all reported cases[29]. Yet, the proportion of reported severe cases among Afro-Colombians was similar to that among Non-Afro-Colombians and fatality among Afro-Colombians was between 5%-10% higher than among Non-Afro-Colombians[29]. One ecological study accounting for the proportion of Afro-Colombian population per district of residence, reported that dengue risk was nine times higher among Non-Afro-Colombians living in districts with an Afro-Colombian population >30% and four times higher for the same population living in districts with an Afro-Colombian population <30%[26]. Nevertheless, in the same study, severity rates were 3.3% and 4.2% for Non-Afro-Colombians and Afro-Colombians, and fatality rates were 2.7% and 3.4% for Non-Afro-Colombians and Afro-Colombians, respectively[26]. Altogether, showing a paradoxical effect of ethnicity on dengue severity in the Colombian contexts as well.

Although it is generally assumed that this paradoxical effect of ethnicity on dengue severity is due to confounding or effect modification by SES[21], given that other biological factors are not documented to change by ethnicity or ancestry[6,12,13,30-32], few analyses have been conducted to confirm or refute the presence of such disparity[11,21]. Previous reports have often suffered from small sample sizes, did not account properly for the spatiotemporal distribution of dengue, used only aggregated data, and were limited by incomplete surveillance data and/or misclassification of ethnicity[11-14,19,21,24,26]. To the best of our knowledge, there has not been a robust analysis with large sample sizes, integrating the spatiotemporal nature of dengue, and adjusting for underreporting and misclassification, intended to understand whether the ethnic disparity is at least partially explained by differences in SES distribution across ethnic groups. This disparity may also be the result of intersectionality[33], a theory suggesting that factors such as ethnicity and SES do not act individually but rather interact reciprocally, producing different outcomes across groups with different combinations of social characteristics[33-35]. Our overarching aim is to understand if and how ethnicity disparities affect the overall and severe burdens of dengue. Using detailed

dengue surveillance data on notified dengue cases in Cali (Colombia), we examined dengue burden across and within ethnic groups, adjusting for and assessing heterogeneity by socioeconomic status, while accounting for spatiotemporal clustering using hierarchical spatiotemporal Bayesian models.

METHODS

Study setting

Cali is the third largest city in Colombia and has a population of 2.4 million inhabitants, from which 73% are considered Non-Afro-Colombians, 26% Afro-Colombians, and <1% Indigenous[36]. Dengue incidence ranged 243-692 cases per 100,000 inhabitants/year over the last five years, contributing more than 15% all dengue cases reported in the country each year [9,30,32,37-39]. The city is administratively organized in 335 neighborhoods which are grouped within 22 urban districts, covering a 120 km^2 area. Cali has two rainy seasons (April and October), the average temperature is 26°C and altitude of approximately 1,000 meters above the sea level. In Cali, 86% of the population has access to potable water, 97% have waste management and 60% live in neighborhoods with low or medium-low SES[36,40,41]. Health coverage by insurance scheme is as follows: 64% contributory (employees or people with self-paid insurance), 27% government subsidized, and 9% uninsured[41].

Data source

Our sample comprised all individual dengue cases notified in Cali from 2012-2017, to SIVIGILA, the national surveillance system. Although dengue notification is mandatory and cases are reported using validated and standardized codes[42], cases are captured through a passive surveillance system (i.e., relying on individuals who display symptoms, seek health care attention and were diagnosed as dengue cases), which may result in underreporting. Data included sociodemographic (age, sex, ethnicity, socioeconomic status, health insurance scheme, occupation, and place of residence) and biomedical variables (laboratory test results and confirmation status), which are collected according to the national guidelines for all cases included in the surveillance system which in principle should not be differential by health insurance scheme[26,38,39,42-44]. Supplementary

aggregated information at the neighborhood level was obtained using *National Census and the Administrative Department of Planning and Infrastructure* data[41]. Entomological information was provided through the local *Secretary of health* and environmental variables through the *National Institute of Meteorology*[45].

Measures

Ethnicity: Defining ethnicity remains challenging in Latin America, where inter-ethnic mixing has historically been common[46,47]. In the Colombian context, self-reported ethnicity has shown good correlation (over 87%) with genetic ancestry, especially for African ancestry[24,25]. Therefore, in this study, we used self-identified ethnic groups, which are preferred over other measures because it also captures the historical background, phenotype and cultural aspects inherent to the definition of ethnicity [46,47]. For this analysis, we used a binary variable using the self-declared ethnicity in the surveillance notification form. Non-Afro-Colombians, indicating the population of Hispanic Whites and Caucasians, or their mix, were used as the referent group. Afro-Colombians include any of the following three possible categories: 1) Afro-Colombians, which are people of African ancestry without a known African lineage; 2) Afro-Colombian *Palenquero*, an African-descendant who lives in or is a member of any of the protected Afro-Colombian reserves or *Palenques*; 3) Afro-Colombian *Raizal*, people with African ancestry (with or without mixing) who are members or descendants from the communities in the Caribbean Islands of Colombia[42]. Although the notification system includes the identification of other ethnic groups including Roma and Indigenous, these cases were not considered in our analysis due to their small numbers.

Socioeconomic measures: In Colombia, SES is estimated using an administrative summary measure (range 1–6) with 1 indicating the lowest SES and 6 indicating the highest SES levels[40]. This proxy of SES is based on households' characteristics, such as construction materials, and assets. Each household has a SES level designated by the municipality. In turn, each neighborhood has a designated SES level that corresponds to the weighted mode of the household's SES within each block and neighborhood[40]. For this analysis, we used the six-level measure for household or

neighborhood's SES as described above, and a three-level categorical variable indicating low-SES (levels 1 and 2), medium-SES (levels 3 and 4) and high-SES (levels 5 and 6).

Outcomes: For the assessment of morbidity we included all notified, clinically and laboratory confirmed dengue cases in the city[42]. To assess the overall rate of dengue distribution by neighborhood's SES we included overall, and ethnic-specific monthly case counts per neighborhood. For the presence of severity, we used individual-level data including a binary variable for dengue cases who required hospitalization[1,6,42]. At the neighborhood level, we estimated the overall and ethnic-specific monthly counts of severe cases per neighborhood.

Statistical Analysis

Descriptive statistics were presented as medians and interquartile range (IQR), frequencies, or proportions. To estimate the overall dengue distribution, we fitted *Negative Binomial* models within a spatiotemporal hierarchical structure, as follows:

$$I_{it} \sim NB(\mu_{it}, \varphi)$$
$$log(\mu_{it}) = log(Pop_{it}) + \beta_0 + \boldsymbol{X'_{it}}\beta_x + u_i + w_t, \quad \text{(equation 1)}$$

where I_{it} is the reported monthly number of notified incident dengue cases for neighborhood i at time t; φ is the overdispersion parameter; $\mu_{it}$is the model predicted incidence; Pop_{it} is the model's offset that contains the mid-year population for neighborhood i and time t; β_0 is the intercept; X'_{it} is a vector of fixed effects covariates including a variable for each: the proportion of male cases, cases under 20 years old, cases from subsidized insurance, and Non-Afro-Colombian cases per neighborhood per month, and a categorical variable for the neighborhood's SES-level (low, medium or high), with their corresponding β_x coefficients. The spatially structured random effects u_i are specified using an intrinsic conditional autoregressive structure (iCAR)[48] using a neighborhood structure defined by: $u_i|u_j,\ i \neq j, \tau \sim \mathcal{N}\left(1/n_i \Sigma x_j, 1/n_i\tau\right),$ where n_i is the number of neighborhoods of node $i = 1 \ldots m,\ u_i \sim u_j$ indicates that the two nodes are neighbors, and τ is the conditional precision for the spatial random effect[48]. Then, w_t indicates the random effect for

time (month), specified as a first order random walk (RW1). The exponentiated coefficients of the fixed effects indicate the overall Incidence Rate Ratio (IRR) for dengue distribution after accounting for other covariates. The exponentiated coefficients for the spatially structured random effect indicate the residual neighborhood-specific IRR, or the effect not explained by the fixed-effects covariates, compared to the overall rate of disease in the city[48].

To assess the severity using individual-level data, we fitted spatiotemporal hierarchical *Poisson* models. We choose Poisson models over logistic regression to avoid overestimation of the risk when the outcome is not rare and non-collapsibility associated to the estimation of odds ratios[49,50]. Although the Poisson models provide an unbiased estimate of the risk ratio, we acknowledge the possibility of some imprecisions in the estimation of the variance that do not affect the interpretation of the uncertainty of our risk estimates[49-51]. Given the lack of independency between the overall dengue distribution and the presence of severity, and to improve the precision of the estimates of severity while accounting for the spatial autocorrelation and the uncertainty associated to the use of surveillance data, the severity risk was modeled as a function of the notified cases and other covariates specified as follows:

$$y_{ijt} \sim Poisson(p_{ijt})$$
$$log(p_{ijt}) = \log(I_{it}) + \beta_0 + \boldsymbol{S'_{ijt}}\,\beta_s + u_i + w_t, \quad \text{(equation 2)}$$

where p_{ijt} is the severity risk for an individual j in neighborhood i at time t, given the overall dengue cases reported I_{it}; the intercept is β_0 and S'_{ijt} is a vector of fixed-effects covariates at the individual level, including a categorical variable for age, sex, insurance and ethnicity, with their respective β_s coefficients. The spatially structured effect u_i and time w_t are specified as described above. The exponentiated $log(p_{ijt})$ indicate the overall severity's Risk Ratio (RR). The exponentiated spatially structured random effects coefficients indicate the residual neighborhood-specific risk not explained by the fixed-effects covariates.

Single models assume covariate's homogeneity across ethnic groups and that any ethnic-group differences will be captured by the estimated coefficients of ethnicity. We fitted ethnic-specific models to allow the identification of heterogeneity across and within ethnic groups on measured covariates, especially for SES. Ethnic-specific models included the mid-year ethnic-specific populations per neighborhood as the offsets for "*equation 1*" and the ethnic-specific number of reported cases per neighborhood/month for "*equation 2*".

The posterior distributions of the parameters, Incidence Rate Ratios (IRR) and Risk Ratios (RR) and their respective 95% Credible Intervals (95%CrI) were estimated via *Integrated Nested Laplace Approximation* (INLA)[52]. We assigned non-informative priors for the precision parameters of the random effects. Model assessment was performed through the Deviation Information Criterion (DIC)[48]. All analyses were fitted using R-INLA (R Core Team (2019); R Studio version 3.3.3)[52,53]. We followed the *REporting of studies Conducted using Observational Routinely collected health Data* (RECORD) statement guideline [54] (Appendix).

Sensitivity analyses: When available, entomological and environmental covariates were added to the model for the overall dengue distribution to assess their contribution to disease incidence. To assess the possibility of residual confounding due to the categorization of SES and or identify any additional pattern of distribution across the SES strata, we fitted models using the six-level SES index. To further assess and quantify the presence of socioeconomic inequalities in the overall and severe dengue distribution, we estimated overall and ethnic-specific *Relative* (RCI) and *Absolute Concentration Index of Inequality* (ACI)[55] (Appendix).

Since surveillance data could be affected by underreporting and given limitations in methods to correct for underreporting on non-normal distributions and misclassification on categorical variables[56], we conducted two sets of sensitivity analysis. First, we updated our priors for the precision of the spatial random effect (outcome precision) and the prior for the precision of the regression parameters (coefficients). Second, to assess the potential effect of underreporting and misclassification with other arboviruses, we fit the models described above on simulated datasets

accounting for the measurement error. Specifically, we used the distribution parameters from the original data to simulate datasets corrected for different ranges of underreporting using reporting rates from the same and similar settings. For overall underreporting, we used an average underreporting of 31% obtained from a study in an endemic Colombian city[31]. However, we also used a wide range of possible underreporting values (10-90%) to account for the large variability of underreporting estimates in the literature[31,39,57]. To account for the potential differential underreporting, we used data from a capture-recapture validation study (comparing the cases diagnosed clinically at different health care facilities to those that were included in the surveillance system) in the same setting, to estimate the underreporting by age, sex, ethnicity, health insurance, and year of notification to inform our simulations (eAppendix 4, eTable 7). Given the results from the validation study and a previous assessment of underreporting accounting for misclassification for other arboviruses (eFigure 3), the parameters of underreporting used in our simulation ranged from 5-50% for Afro-Colombians and from 1-30% for subsidized insurance scheme (eAppendix 4).

Ethics statement: This study analyzed secondary data without human samples analyses, therefore, it did not require consent to participation. The protocol was reviewed and approved by the *Institutional Review Board* of McGill University (Study No. A02-E05-18A).

RESULTS

There were 65,774 dengue cases reported to SIVIGILA between 2012-2017. Excluding 372 (0.6%) Indigenous cases, our analysis included 65,402 dengue cases of which 3,102 (5%) were Afro-Colombian. There were 13,732 (21%) cases hospitalized, of which 7% (945/13,732) were Afro-Colombians. Twenty percent (n=14,076) of all notified cases, 39% (n=4,939) of Non-Afro-Colombian hospitalized cases and 66% (n=628) of Afro-Colombians hospitalized cases, had subsidized insurance. Of all notified cases, 52% (n=33,933) of overall and 53% (n=7,309) of hospitalized cases were from low-SES neighborhoods. Among Afro-Colombians 68% (n=639) hospitalized cases were from low-SES neighborhoods (**Table 7.1**).

Table 7.1 Characteristics of the Dengue Cases reported in Cali Colombia (2012-2017).

	All Dengue Cases (n=65,402)		**Non-Afro-Colombians** (n=62,300)		**Afro-Colombians** (n=3,102)	
Characteristic	**Dengue Cases n (%)**	**Severe Cases n (%)**	**Dengue Cases n (%)**	**Severe Cases n (%)**	**Dengue Cases n (%)**	**Severe Cases n (%)**
Total Cases	65,402 (100%)	13,732 (21%)	62,300 (100%)	12,787 (20.5%)	3,102 (100%)	945 (30.5%)
Age, median (IQR)	25 (13, 41)	18 (10, 38)	25 (13, 41)	18 (10, 38)	22 (12, 39)	17 (10, 32)
Age Group: <9 years	10906 (16.7%)	3261 (23.7%)	10344 (16.6%)	3035 (23.7%)	562 (18.1%)	226 (23.9%)
10-19 years	14614 (22.3%)	4135 (30.1%)	13815 (22.2%)	3839 (30.0%)	799 (25.8%)	296 (31.3%)
20-39 years	22723 (34.7%)	3076 (22.4%)	21737 (34.9%)	2842 (22.2%)	986 (31.8%)	234 (24.8%)
40-59 years	11799 (18.0%)	1723 (12.5%)	11316 (18.2%)	1631 (12.8%)	483 (15.6%)	92 (9.7%)
>60 years	5358 (8.2%)	1537 (11.2%)	5088 (8.2%)	1440 (11.3%)	270 (8.7%)	97 (10.3%)
Sex: Male	33747 (51.6%)	6939 (50.5%)	32236 (51.7%)	6495 (50.8%)	1511 (48.7%)	444 (47.0%)
Sex: Female	31653 (48.4%)	6793 (49.5%)	30064 (48.3%)	6292 (49.2%)	1589 (51.3%)	501 (53.0%)
Insurance Scheme: Subsidized	13076 (20.0%)	5567 (40.5%)	11576 (18.6%)	4939 (38.6%)	1500 (48.4%)	628 (66.5%)
Insurance Scheme: Contributory	52324 (80.0%)	8165 (59.5%)	50724 (81.4%)	7848 (61.4%)	1600 (51.6%)	317 (33.5%)
Socio Economic Status (SES) Index						
SES 1 (Very Low)	10629 (16.3%)	2662 (19.4%)	9791 (15.7%)	2359 (18.4%)	838 (27.0%)	303 (32.1%)
SES 2 (Low)	23304 (35.6%)	4647 (33.8%)	22177 (35.6%)	4311 (33.7%)	1127 (36.3%)	336 (35.6%)
SES 3 (Low-Medium)	23867 (36.5%)	4799 (34.9%)	23028 (37.0%)	4570 (35.7%)	839 (27.1%)	229 (24.2%)
SES 4 (Medium)	3981 (6.1%)	825 (6.0%)	3812 (6.1%)	776 (6.1%)	169 (5.4%)	49 (5.2%)
SES 5 (Medium-High)	2949 (4.5%)	649 (4.7%)	2852 (4.6%)	625 (4.9%)	97 (3.1%)	24 (2.5%)
SES 6 (High)	671 (1.0%)	150 (1.1%)	640 (1.0%)	146 (1.1%)	31 (1.0%)	4 (0.4%)
Year 2012	2580 (3.9%)	831 (6.1%)	2381 (3.8%)	770 (6.0%)	199 (6.4%)	61 (6.5%)
Year 2013	19130 (29.3%)	4922 (35.8%)	17887 (28.7%)	4642 (36.3%)	1243 (40.1%)	280 (29.6%)
Year 2014	4964 (7.6%)	998 (7.3%)	4664 (7.5%)	902 (7.1%)	300 (9.7%)	96 (10.2%)
Year 2015	13433 (20.5%)	2257 (16.4%)	12911 (20.7%)	2036 (15.9%)	522 (16.8%)	221 (23.4%)
Year 2016	21281 (32.5%)	4035 (29.4%)	20611 (33.1%)	3819 (29.9%)	670 (21.6%)	216 (22.9%)
Year 2017	4012 (6.1%)	689 (5.0%)	3846 (6.2%)	618 (4.8%)	166 (5.4%)	71 (7.5%)

Overall Dengue Distribution

On average, the spatiotemporally adjusted incidence rate during the study period was 40.4 (95%CrI: 37.6-43.4) per 100,000 people/month. After adjusting for other covariates, dengue rates increased slightly with every 10% increase in the proportion of cases with subsidized insurance (IRR=1.02; 95%CrI: 1.01-1.03). Comparing to the low-SES level, dengue rates were lower among neighborhoods with high-SES level (IRR=0.68; 95%CrI: 0.55-0.84). Increases in the neighborhood proportion of Afro-Colombian cases did not change the overall distribution of dengue rates (IRR=0.99; 95%CrI: 0.98-1.00), even after adjusting for the proportion of Afro-Colombian population per neighborhood. Among Non-Afro-Colombians, dengue rates decrease slightly with the proportion of cases under 20 years old and increased with the proportion of cases with subsidized insurance. Among Afro-Colombians, dengue rates did not change by any covariate in the model (Table 7.2).

The spatially structured random effects, indicating the spatial effect not explained by other covariates, showed dengue presence in the entire city with concertation in eastern neighborhoods. We observed absence of Afro-Colombian cases in 50 neighborhoods and from those notifying cases, there was no evidence of spatial clustering within the city. Non-Afro-Colombian cases were reported in all but three neighborhoods, and a concentration of cases was observed in some outlying neighborhoods of the city (Figure 7.1). During the study period, higher dengue rates were observed during 2013 and 2016 and the same trend but of different magnitude was observed across ethnic groups (Figure 7.2).

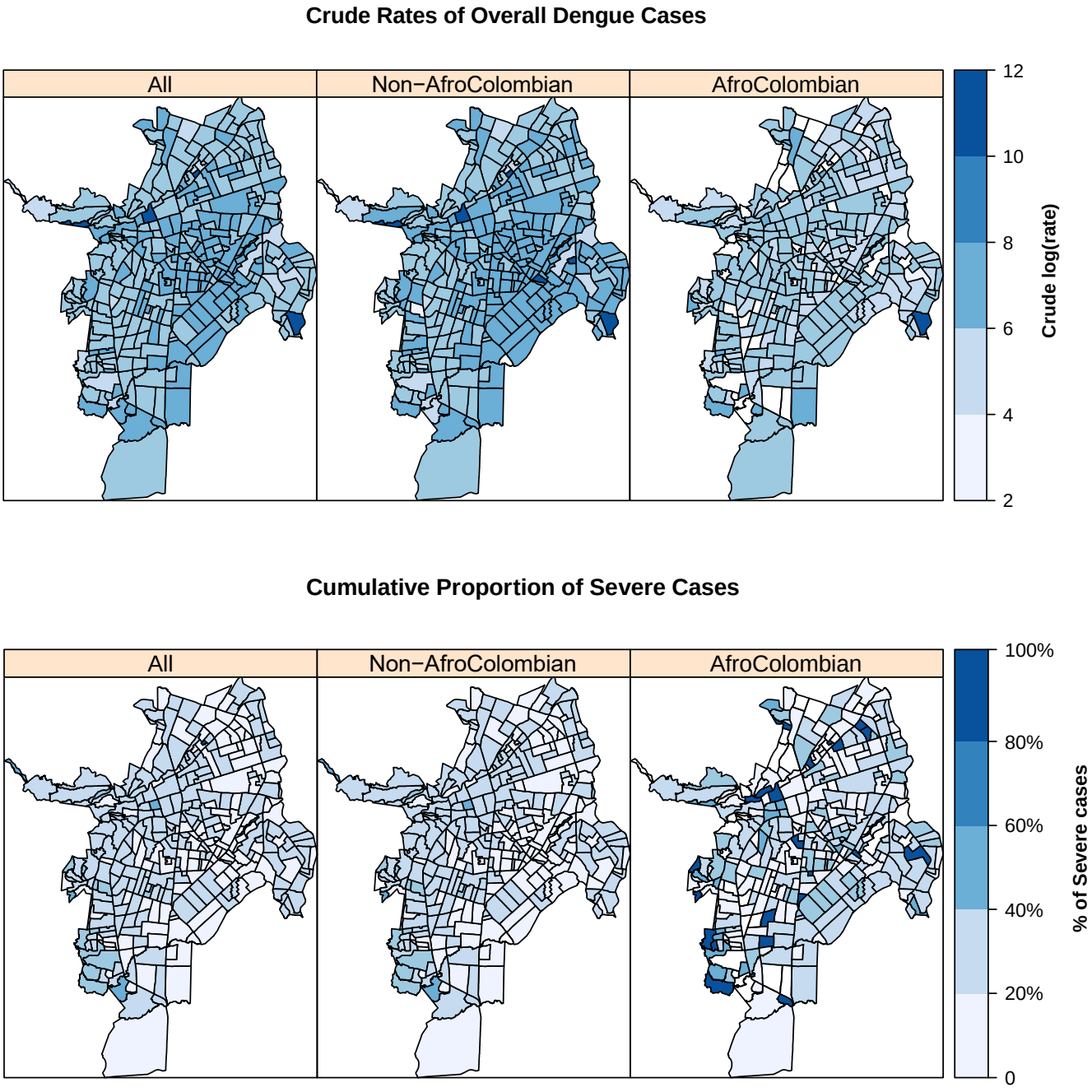


Figure 7-1. Distribution of dengue cases in Cali, Colombia (2012- 2017).
Figure 1A. Crude rates of overall dengue cases reported per neighborhood during the study period.
Figure 1B. Proportion of severe cases among the total reported dengue cases per neighborhood during the study period.

Severe Dengue Distribution

The adjusted model identified an average severity risk of 8.4% (95%CrI: 7.5%-9.2%) with increased severity risk among Afro-Colombians (RR=1.16; 95%CrI: 1.08-1.24) and subsidized insurance (RR=2.37; 95%CrI: 2.28-2.47). Compared to cases under 10 years old, severity risk was lower among cases 10 to 59 years of age. Ethnic-specific models adjusted by age, sex and insurance scheme estimated an average probability of severe dengue of 8.5% (95%CrI: 7.7%-9.4%) for Non-Afro-Colombians and 23% (95%CrI: 18%-28%) for Afro-Colombians. Subsidized insurance was associated with higher severity among both Non-Afro-Colombians (RR=2.36; 95%CrI: 2.27-2.46) and Afro-Colombians (RR=1.79; 95%CrI: 1.53-2.09) (**Table 7.3**). The residual spatial effect for severity risk showed less variability in neighborhoods with higher overall dengue rates and did not suggest clustering effects (**Figure 7.3**).

Table 7.2 Model for the spatial distribution of notified dengue cases for the overall population and ethnic-specific models.

	All Dengue Cases (n=65,402)		Non-Afro-Colombians (n=62,300)		Afro-Colombians (n=3,102)	
Covariate	**IRR**	**95% CrI**	**IRR**	**95% CrI**	**IRR**	**95% CrI**
Proportion of Non-Afro-Colombian cases [a]	0.99	(0.98, 1.00)	-	-	-	-
Proportion of Male cases [a]	1.03	(1.02, 1.04)	1.02	(1.01, 1.03)	1.00	(0.98, 1.01)
Proportion of Cases Under 20 years old [a]	1.00	(0.99, 1.01)	0.98	(0.97, 0.99)	1.00	(0.98, 1.02)
Proportion of Cases in the Subsidized Scheme [a]	1.02	(1.01, 1.03)	1.02	(1.01, 1.03)	1.00	(0.98, 1.01)
Socio Economic Status (SES) Index [b]						
Low SES (Levels 1 and 2)	Ref	-	Ref	-	Ref	-
Medium SES (Levels 3 and 4)	0.97	(0.84, 1.11)	0.94	(0.82, 1.09)	0.96	(0.88, 1.06)
High SES (Levels 5 and 6)	0.68	(0.55, 0.84)	0.66	(0.53, 0.81)	1.05	(0.86, 1.28)

[a] Indicates a 10% increase in the proportion of cases. [b] Compares to the Low SES level

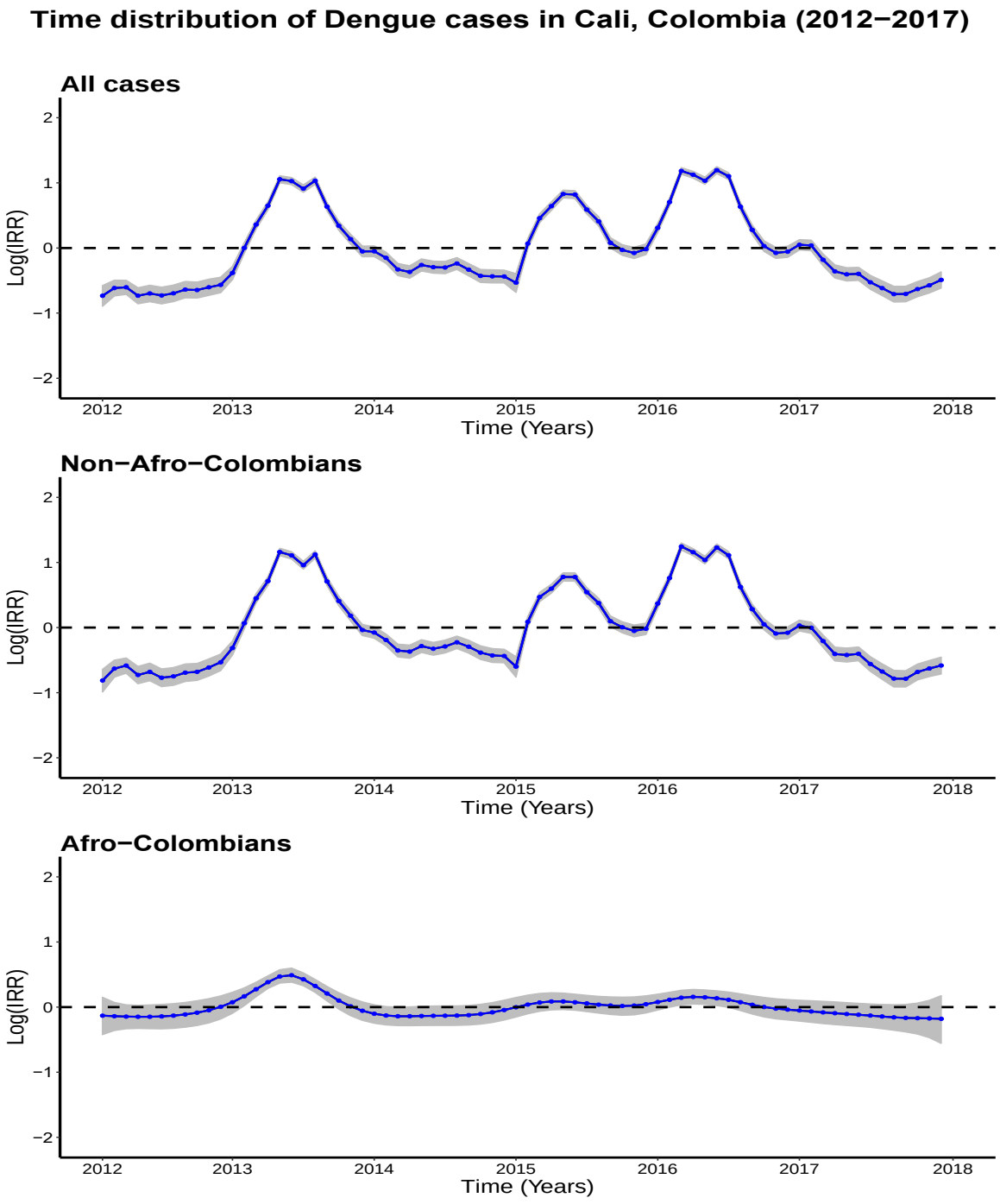


Figure 7-2 Temporal distribution of dengue Cases in Cali, Colombia (2012- 2017).
Log Incidence rate Ratios (log(IRR)), compared to the average rate in the city during the study period.

Table 7.3 Model for the spatial distribution of severe dengue cases for the overall population and ethnic-specific models.

	All Dengue Cases (n=65,402)		Non-Afro-Colombians (n=62,300)		Afro-Colombians (n=3,102)	
Covariate	RR	95% CrI	RR	95% CrI	RR	95% CrI
Afro-Colombians	1.16	1.08, 1.24	-	-	-	-
Age Group						
Under 9 years	Ref	-	Ref	-	Ref	-
10-19 years	0.86	0.82, 0.90	0.85	0.81, 0.90	0.92	0.77, 1.11
20-39 years	0.52	0.50, 0.55	0.52	0.49, 0.55	0.63	0.52, 0.76
40-59 years	0.56	0.53, 0.60	0.56	0.53, 0.60	0.56	0.43, 0.71
>60 years	1.01	0.95, 1.07	1.02	0.96, 1.09	0.95	0.74, 1.21
Sex						
Female	Ref	-	Ref	-	Ref	-
Male	0.97	0.94, 1.00	0.97	0.94, 1.01	0.96	0.84, 1.09
Insurance						
Contributory	Ref	-	Ref	-	Ref	-
Subsidized Scheme	2.37	2.28, 2.47	2.36	2.27, 2.46	1.79	1.53, 2.09

Sensitivity analyses

To assess the contribution of the entomological data, we used the log-transformed proportion of *Aedes*-positive catch basins (only available at the district level (n=22)). This model showed that dengue rates increased with rising proportion of *Aedes*-positive catch basins per district (IRR=1.12; 95%CrI: 1.06-1.19). To assess the contribution of precipitation and temperature (only available from 2014-2017), we modeled both as random effects and continuous variables, but did not find any association with our outcomes.

The analyses using the six-level SES index showed that compared to neighborhoods at the lowest SES (SES level 1), dengue rates were around 43% higher at low and medium-SES (SES levels 2 and 3). While rates in neighborhoods at high SES (SES levels 4 to 6) were on average 20% lower than rates at the lowest SES. Neither the six-level-SES variable nor the entomological index or environmental covariates changed the main results or improved the precision or the DIC values (Supplementary Material).

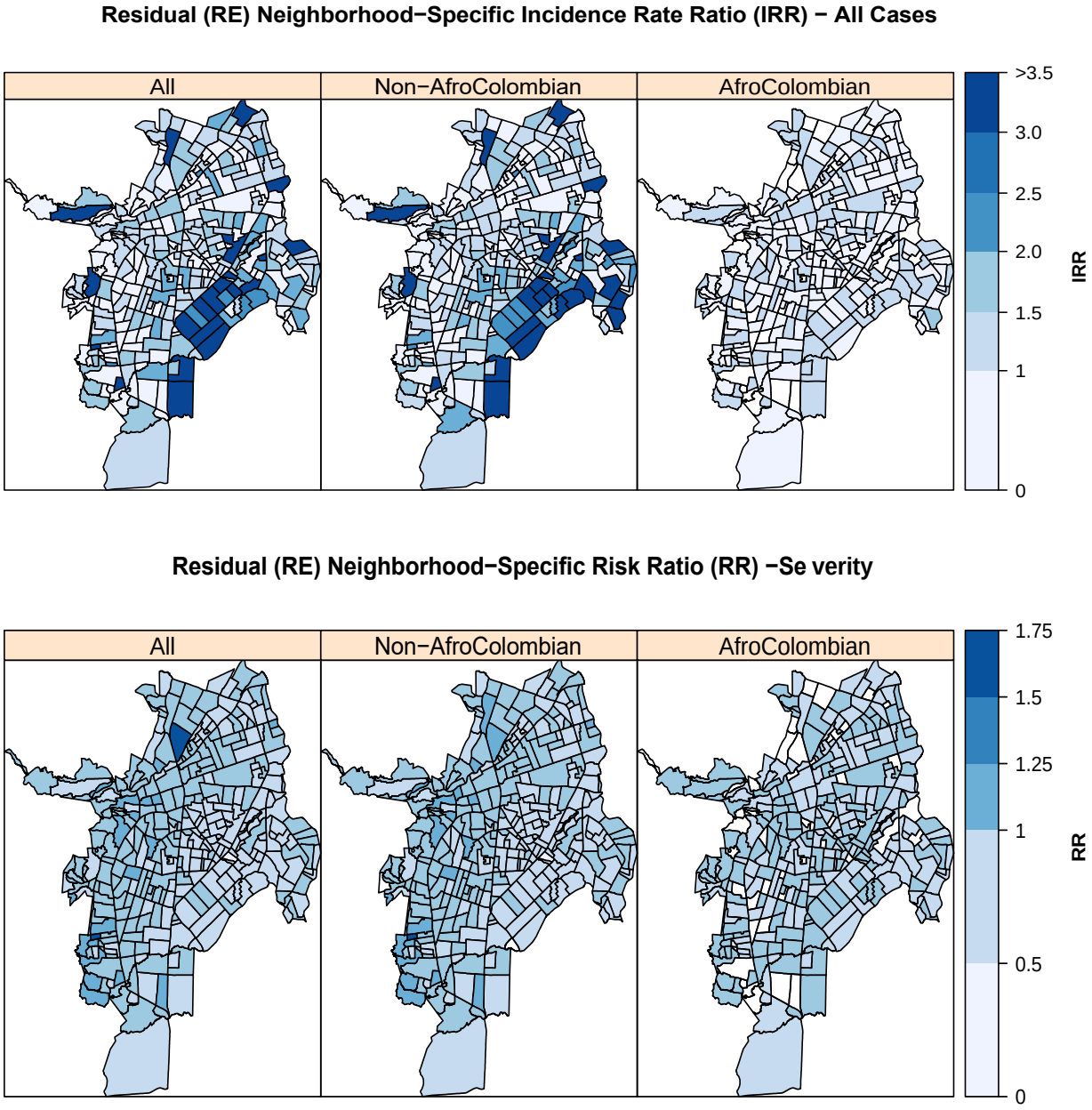


Figure 7-3 Residual Spatial effects (spatially structured random effects) for dengue distribution in Cali, Colombia (2012- 2017).

Results obtained from models adjusting for age, sex, insurance scheme and SES. **(A) Top.** Spatial effects for all reported dengue cases, indicating the neighborhood-specific random effect incidence rate ratio (IRR), comparing the neighborhood's specific incidence rate to the to the overall incidence rate of dengue cases in the city, after adjusting for other covariates, during the study period. **(B) Bottom.** Spatial effects for severe dengue cases, indicating the neighborhood-specific random effect risk ratio (RR), comparing the neighborhood's specific risk of severity to the to the overall risk of severe dengue in the city, after adjusting for other covariates, during the study period.

The relative concentration index (RCI) for all dengue cases, calculated using the six-level SES index was -0.05 (95%CrI: -0.06 to -0.04); indicating the presence of a modest socioeconomic inequality with concentration of dengue cases in low-SES neighborhoods. Ethnic-specific RCIs for overall distribution of cases showed the presence of larger inequalities among Non-Afro-Colombians. For severity, the RCI among Afro-Colombians was -0.04 (95%CrI: -0.07 to -0.01) indicating the concentration of severe Afro-Colombians cases among low-SES neighborhoods (Supplementary Material, eFigure 3).

The analysis of the overall distribution with informative priors showed that across ethnic groups, a marked effect of SES was observed among Non-Afro-Colombians while similar rates were observed across SES-level among Afro-Colombians. Using the ranges of underreporting derived from the validation study (eTable 7 and eFigure 4), the simulations adjusting for underreporting (overall, and by insurance and ethnicity) and for misclassification with other arboviruses, did not change the magnitude of the estimated for overall dengue presence but affected their precision. However, the adjusted and corrected individual model for severe dengue showed an average severity risk of 7.8% (95%CrI: 6.6%-9.2%) and no difference across ethnic groups (RR=1.02; 95%CrI: 0.97-1.07) (eTable 8).

DISCUSSION

We investigated the presence of ethnic disparities in notified dengue cases and the paradoxical increase of dengue severity among people of African ancestry, despite the consideration that African ancestry is "protective" for severe forms of the disease, in the Colombian context. Using data from over 65,000 cases in Cali-Colombia, our findings suggest that overall dengue distribution did not change across ethnic groups, but severity was on average 16% higher among Afro-Colombians. However, this association disappeared when we corrected for underreporting by ethnicity and potential misclassification of the outcome. Likewise, ethnic-specific models allowed the identification of differential dengue patterns across SES and within ethnic groups[26,46,47,58]. The results for sex and age distribution across and within ethnic groups were consistent with previous reports about overall distribution and severity[9,13,28,57].

In this context, the increased severity among Afro-Colombians could be understood as the result of intersectionality between SES, health seeking behavior by insurance scheme and ethnicity[33,34]. In Cali, Afro-Colombian cases were only a small proportion of all dengue cases but were also generally from low-SES settings and had subsidized insurance. This combination of factors may have led to differential health-seeking behaviors among Afro-Colombians and, therefore, to a differential reporting. First, cases from low-SES settings might face several financial barriers accessing health care (e.g., pay for transportation, copayments or indirect-medical costs, etc.) as described in other Latin American settings[14,15,30,32,59]. Second, cases with subsidized insurance do not benefit from paid medical leaves, which is a known barrier for seeking and accessing health care[10,11,13,15,28,43,44,59,60]. Third, in addition to health insurance, there are other (unmeasured) socio-cultural characteristics associated with the disadvantage of being Afro-Colombian living in low-SES settings in Colombia[26,30,46,47,59]. Altogether, these factors likely lead Afro-Colombians with dengue to consult -and being reported- only when their symptoms became severe [10,11,44,46,47,60].

Our postulate is also supported by other findings, including that among Afro-Colombians, some neighborhoods only reported severe cases. Moreover, the sensitivity analyses correcting for overall and ethnic-specific underreporting did not identify any association between severity and ethnicity. Likewise, the estimation of RCIs for severity indicated the presence of socioeconomic inequalities towards the poor mostly among Afro-Colombians. Taken together, these observations are consistent with an interpretation of these disparities using an intersectionality lens[10,15,33-35,46,47]. Although intersectionality has not been systematically assessed for arboviruses in the Colombian context, access to healthcare and self-reported health are generally considered to be worse for Afro-Colombians[46,47].

Importantly, our study only allows the consideration of a non-differential distribution of dengue severity across ethnic groups in this study population. Although our analyses allow the consideration of the increased severity among Afro-Colombians as the result of the intersectionality described above, the scope of this study does not allow the assessment or

confirmation of the said protective effect of African ancestry for dengue severity. On the one hand, despite a reported good correlation between self-reported ethnicity and ancestry in certain Colombian settings[24,25], the degree of correlation changes according to the setting, the degree of admixture and the socioenvironmental conditions[19-22,25]. Therefore, any assessment on the reported ethnicity presented in our study should be interpreted in this context as a sociodemographic construct and not as a mere biological characteristic[33,47]. On the other hand, the ascertainment of any biological protective effect of ancestry would likely require the assessment of biomarkers or immunological analyses that were out of the scope of this study. In addition, although the presence of primary vs secondary infections and the serotypes-interaction are factors associated to severity and could be postulated drivers of the paradox, we did not account for the type of infection nor the serotype-specific infection in our study due to data availability. However, such factors are only potential candidates to explain the paradox if considered to behave differentially by ethnicity or ancestry, which has not been yet reported literature[6,19-21,23-25].

The analysis of neighborhood SES and dengue rates showed a non-monotonic association, with an overall larger presence of dengue cases among low and medium-SES neighborhoods. These findings are consistent with other analyses in Latin America, where increased dengue among low-SES settings is attributed to socio-environmental factors such as limited water supply and overcrowding that facilitate *Aedes* presence[12-15,26 10,12,13]. Interestingly, these results were not parallel across ethnic groups. Among Afro-Colombians, the absence of dengue trends across SES and lack of precision could be attributed to their relatively small sample size. However, these findings could be attributed also to a violation of the positivity assumption that results from structural confounding, given that only 4% of the Afro-Colombian cases lived in high-SES settings [58,61]. To mitigate this potential issue, ethnic-specific models only included neighborhoods reporting cases and the main analysis used the three-categories-level SES variable. However, the spatial nature of the analysis could have led to data sparsity and the results for dengue rates among high-SES neighborhoods are considered "off-support"[61].

Despite the majority of reported cases having contributory insurance, the overall distribution of dengue and the risk of severity were consistently associated with subsidized insurance across and within ethnic groups, even after adjusting for the neighborhood SES. Subsidized insurance in Colombia is provided by the government for individuals with income below the minimum wage or with no income, while the contributory scheme corresponds to the coverage provided to individuals with capacity to contribute for themselves [37,43,44,60]. Considering the insurance scheme as a proxy of SES at the individual level, our findings corroborate a larger dengue presence among disadvantaged populations[37,43,44,60]. This is consistent with an increased presence of other communicable diseases among individuals with subsidized insurance in Colombia[32,44,47,60,62].

Our study also confirmed the endemic and epidemic character of dengue in Cali, Colombia[9,26,30,32,38,39]. We also identified the presence of dengue clustering in some eastern neighborhoods and a similar spatial risk for severity across the city. The spatial clustering was observed among previously described "dengue vulnerable zones"[30,32,59], reinforcing again the presence of dengue among low-SES settings. Although entomological indexes could have contributed to the spatial heterogeneity, this data was only available at the district-level and amongst external catch basins, for which results and level of agreement with dengue rates have been debated[29,42]. Nonetheless, the effect of any unmeasured spatial-level covariate or any residual seasonal effect, are expected to be captured by the residual spatial and temporal random effects, respectively[48,63,64].

Strengths and Limitations

Surveillance data could be affected by measurement error and selection bias. To account for this, we conducted a series of sensitivity analyses accounting for overall and ethnic-specific underreporting and misclassification. The simulated results for the overall distribution of cases reinforced the trend of increased rates among low-SES while the simulated results for severity, showed an attenuation of the severity risk for Afro-Colombians. Our sensitivity analyses were intended to assess the scope and direction of the bias. However, further research is required to quantify and determine the impact of underreporting, probably using other source of validation

data and parameters of the simulations to assess consistency and or using other approaches to assess quantitatively the magnitude of the intersectionality[34,35]. Nonetheless, it is important to note that surveillance data is the main source of information used for policy-making and disease control in many endemic settings[2,27,42]. Therefore, the use of surveillance-based data in our analysis is justified, as long as efforts are made to account for potential biases and ensure the interpretability of the results conditional on the reported cases.

Strengths of our study include its large sample size, the assessment of the severity risk proportional to the spatiotemporal distribution of overall cases, the assessment of heterogeneity across SES and by ethnic groups, and the efforts made to correct for potential bias related to the use of surveillance data.

Conclusions

Overall dengue rates do not change across ethnic groups. Although we also observed a small increase in severity among Afro-Colombians, this paradoxical finding could be attributed to differential health-seeking behaviors and reporting among Afro-Colombians. More generally, the presence of intersectionality between diverse socioeconomic factors determining access to health care across ethnic groups, could help explain the increased severity observed in our study and elsewhere in the Americas regions. However, further research is necessary to assess quantitatively the effect of intersectionality between factors such as income, insurance schemes and ethnicity. Finally, our study also confirms the overall concentration of dengue cases among vulnerable populations in low socioeconomic settings. These results contribute to the body of evidence about health inequalities in arboviruses by providing robust estimates of the ethnic, socioeconomic, and spatial distribution of dengue cases.

Conflict of interest: None declared.

Funding: This study did not receive any specific funding.

Computing code: The R-script used to conduct the analysis in this manuscript is provided as supplementary material.

Data availability: Data used in this manuscript could be obtained by official requests to the public health offices/ local Ministry of Health from Cali, Colombia.

Acknowledgments: We acknowledge the collaboration of the Public Health and the Surveillance office in Cali, Colombia. Mabel Carabali holds a CIHR-Banting-Best Doctoral Award fellowship.

REFERENCES

1. World Health Organization (WHO). Dengue and severe dengue. In: WHO, ed. *Health Topics. Fact sheets [Dengue]*. Geneva, Switzerland: WHO, 2020;https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue.
2. World Health Organization (WHO). A Toolkit for National Dengue Burden Estimation. In: World Health Organization (WHO), ed. Geneva, 2018.
3. Mayer SV, Tesh RB, Vasilakis N. The emergence of arthropod-borne viral diseases: A global prospective on dengue, chikungunya and zika fevers. *Acta Trop* 2017;166(Supplement C):155-163.
4. Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda KA, Bhatt S, Katzelnick L, Howes RE, Battle KE, Simmons CP, Hay SI. Global spread of dengue virus types: mapping the 70 year history. *Trends Microbiol.* 2014;22(3):138-146.
5. World Health Organization (WHO). Global strategy for dengue prevention and control 2012-2020. Geneva: WHO. World Health Organization 2012;vi, 43p.
6. Pan American Health Organization (PAHO). *Tool for the diagnosis and care of patients with suspected arboviral diseases*. 1st Edition ed. Vol. PAHO Strategic Plan 2014-2019. Washington: PAHO, 2017.
7. World Health Organization (WHO). Dengue: prevention and control. *Sixty-eight World Health Assembly*. Geneva, Switzerland: World Health Organization, 2015.
8. World Health Organization, Special Programme for Research and Training in Tropical Diseases. *Dengue: guidelines for diagnosis, treatment, prevention and control*. WHO Library Cataloguing-in-Publication Data ed. Vol. New Edition. Geneva: World Health Organization, 2009.
9. Villar LA, Rojas DP, Besada-Lombana S, Sarti E. Epidemiological Trends of Dengue Disease in Colombia (2000-2011): A Systematic Review. *PLoS Negl Trop Dis* 2015;9(3):e0003499.
10. Carabali M, Hernandez L, Arauz M, Villar L, Ridde V. Why are people with dengue dying? A scoping review of determinants for dengue mortality. *BMC Infectious Diseases* 2015;15(1):301.

11. Moraes GH, de Fátima Duarte E, Duarte EC. Determinants of Mortality from Severe Dengue in Brazil: A Population-Based Case-Control Study. *The American Journal of Tropical Medicine and Hygiene* 2013;88(4):670-676.
12. Farinelli EC, Baquero OS, Stephan C, Chiaravalloti-Neto F. Low socioeconomic condition and the risk of dengue fever: A direct relationship. *Acta Tropica* 2018;180:47-57.
13. Rodrigues NCP, Daumas RP, de Almeida AS, dos Santos RS, Koster I, Rodrigues PP, Gomes MdF, Macedo AdF, Gerardi A, Leite IdC. Risk factors for arbovirus infections in a low-income community of Rio de Janeiro, Brazil, 2015-2016. *PLOS ONE* 2018;13(6):e0198357.
14. Bavia L, Melanda FN, de Arruda TB, Mosimann ALP, Silveira GF, Aoki MN, Kuczera D, Sarzi ML, Junior WLC, Conchon-Costa I, Pavanelli WR, Duarte dos Santos CN, Barreto RC, Bordignon J. Epidemiological study on dengue in southern Brazil under the perspective of climate and poverty. *Scientific Reports* 2020;10(1):2127.
15. Mulligan K, Dixon J, Joanna Sinn C-L, Elliott SJ. Is dengue a disease of poverty? A systematic review. *Pathogens and Global Health* 2015;109(1):10-18.
16. Jaenisch T, Junghanss T, Wills B, Brady OJ, Eckerle I, Farlow A, Hay SI, McCall PJ, Messina JP, Ofula V, Sall AA, Sakuntabhai A, Velayudhan R, Wint GRW, Zeller H, Margolis HS, Sankoh O, Dengue in Africa Study G. Dengue expansion in Africa-not recognized or not happening? *Emerging infectious diseases* 2014;20(10):e140487.
17. Halstead SB, Streit TG, Lafontant JG, Putvatana R, Russell K, Sun W, Kanesa-Thasan N, Hayes CG, Watts DM. Haiti: absence of dengue hemorrhagic fever despite hyperendemic dengue virus transmission. *The American Journal of Tropical Medicine and Hygiene* 2001;65(3):180-3.
18. Sierra BDLC, García G, Pérez AB, Morier L, Alvarez M, Kourí G, Guzmán MG. Ethnicity and Difference in Dengue Virus-Specific Memory T Cell Responses in Cuban Individuals. *Viral Immunology* 2006;19(4):662-668.
19. Sierra BdlC, Kourí G, Guzmán MG. Race: a risk factor for dengue hemorrhagic fever. *Archives of Virology* 2007;152(3):533-542.
20. Sierra B, Triska P, Soares P, Garcia G, Perez AB, Aguirre E, Oliveira M, Cavadas B, Regnault B, Alvarez M, Ruiz D, Samuels DC, Sakuntabhai A, Pereira L, Guzman MG. OSBPL10, RXRA and

lipid metabolism confer African-ancestry protection against dengue haemorrhagic fever in admixed Cubans. *PLOS Pathogens* 2017;**13**(2):e1006220.

21. Blanton RE, Silva LK, Morato VG, Parrado AR, Dias JP, Melo PRS, Reis EA, Goddard KAB, Nunes MRT, Rodrigues SG, Vasconsuelos PFC, Castro JM, Reis MG, Barreto ML, Teixeira MG. Genetic ancestry and income are associated with dengue hemorrhagic fever in a highly admixed population. *European Journal of Human Genetics* 2008;**16**(6):762-765.
22. Silva LK, Blanton RE, Parrado AR, Melo PS, Morato VG, Reis EAG, Dias JP, Castro JM, Vasconcelos PFC, Goddard KAB, Barreto ML, Reis MG, Teixeira MG. Dengue hemorrhagic fever is associated with polymorphisms in JAK1. *European Journal of Human Genetics* 2010;**18**(11):1221-1227.
23. Restrepo BN, Ramirez RE, Arboleda M, Alvarez G, Ospina M, Diaz FJ. Serum Levels of Cytokines in Two Ethnic Groups with Dengue Virus Infection. *The American Journal of Tropical Medicine and Hygiene* 2008;**79**(5):673-677.
24. Chacón-Duque JC, Adhikari K, Avendaño E, Campo O, Ramirez R, Rojas W, Ruiz-Linares A, Restrepo BN, Bedoya G. African genetic ancestry is associated with a protective effect on Dengue severity in colombian populations. *Infection, Genetics and Evolution* 2014;**27**:89-95.
25. Avendaño-Tamayo E, Campo O, Chacón-Duque JC, Ramírez R, Rojas W, Agudelo-Flórez P, Bedoya G, Restrepo BN. Variantes en los genes TNFA, IL6 e IFNG asociadas con la gravedad del dengue en una muestra de población colombiana. *Biomédica* 2017;**37**(4):486-497.
26. Rojas Palacios JH, Alzate A, Martínez Romero HJ, Concha-Eastman AI. AfroColombian ethnicity, a paradoxical protective factor against Dengue. *Colombia Médica* 2016;**47**:133-141.
27. Martins-Melo FR, Ramos AN, Alencar CH, Heukelbach J. Mortality from neglected tropical diseases in Brazil, 2000–2011. *Bulletin of the World Health Organization* 2016;**94**(2):103-110.
28. Paixão ES, Costa MdCN, Rodrigues LC, Rasella D, Cardim LL, Brasileiro AC, Teixeira MGLC. Trends and factors associated with dengue mortality and fatality in Brazil. *Revista da Sociedade Brasileira de Medicina Tropical* 2015;**48**:399-405.
29. Instituto Nacional de Salud (INS). Informes de Evento [Surveillance Reports]. *Evento: Dengue*. Santa Fe de Bogota, Colombia: Publicaciones: Instituto Nacional de Salud de Colombia, 2019;https://www.ins.gov.co.

30. Delmelle E, Hagenlocher M, Kienberger S, Casas I. A spatial model of socioeconomic and environmental determinants of dengue fever in Cali, Colombia. *Acta Tropica* 2016;**164**(Supplement C):169-176.
31. Carabali M, Lim JK, Palencia DC, Lozano-Parra A, Gelvez RM, Lee KS, Florez JP, Herrera VM, Kaufman JS, Rojas EM, Villar LA. Burden of dengue among febrile patients at the time of chikungunya introduction in Piedecuesta, Colombia. *Tropical Medicine & International Health* 2018;**23**(0):1231-1241.
32. Desjardins MR, Casas I, Victoria AM, Carbonell D, Dávalos DM, Delmelle EM. Knowledge, attitudes, and practices regarding dengue, chikungunya, and Zika in Cali, Colombia. *Health & Place* 2020;**63**:102339.
33. Collins PH. Intersectionality's Definitional Dilemmas. *Annual Review of Sociology* 2015;**41**(1):1-20.
34. Jackson JW, VanderWeele TJ. Intersectional decomposition analysis with differential exposure, effects, and construct. *Social Science & Medicine* 2019;**226**:254-259.
35. Jackson JW. Explaining intersectionality through description, counterfactual thinking, and mediation analysis. *Social Psychiatry and Psychiatric Epidemiology* 2017;**52**(7):785-793.
36. Alcaldia de Cali. Cali en Cifras. In: Departamento Administrativo de Planeacion Municipal, ed. *Estratificacion Socioeconomica*. Santiago de Cali, Colombia: Alcaldia de Cali, 2018;www.cali.gov.co.
37. Instituto Nacional de Salud (INS). Informe Epidemiologico de Evento Dengue. In: Vigilancia y Análisis del Riesgo en Salud Pública., ed. Dengue. Vol. FOR-R02.4000-001. Bogota, Colombia: Instituto Nacional de Salud, 2016.
38. Rico-Mendoza A, Alexandra P-R, Chang A, Encinales L, Lynch R. Co-circulation of dengue, chikungunya, and Zika viruses in Colombia from 2008 to 2018. *Revista panamericana de salud publica = Pan American journal of public health* 2019;**43**:e49-e49.
39. Mora-Salamanca AF, Porras-Ramírez A, De la Hoz Restrepo FP. Estimating the burden of arboviral diseases in Colombia between 2013-2016. *International Journal of Infectious Diseases* 2020.

40. Departamento Administrativo Nacional de Estadistica (DANE). Metodologia de la Estratificacion Socioeconomica Urbana para Servicios Publicos Domicioliarios. In: Direccion de Geoestadistica, ed. Grupo de Estratificacion ed. Santa Fe de Bogota, Colombia.: DANE, 2015;96.
41. Departamento Nacional de Estadistica de Colombia (DANE). Estimaciones de población 1985 - 2005 y proyecciones de población 2005 - 2020 total municipal por área. *Departamento Nacional de Estadistica de Colombia* DANE, 2015.
42. Instituto Nacional de Salud (INS). Protocolo de Vigilancia en Salud Pública, Dengue (Surveillance Protocol in Public Health, Dengue). . In: Instituto Nacional de Salud (INS) C, ed. *Surveillance*. Dirección de Vigilancia y Análisis de Riesgo en Salud Pública ed. Vol. FOR-R02.0000-059 V02. Santafé de Bogota, Colombia. : Instituto Nacional de Salud (National Institute of Health). , 2014;19.
43. Carabalí JM, Hendrickx D. Dengue and health care access: the role of social determinants of health in dengue surveillance in Colombia. *Glob Health Promot* 2012;19.
44. Ardila Pinto F, Martínez S, Fuentes M, Borrero E. Análisis de las demoras en salud en personas que enfermaron de gravedad o fallecieron por dengue en cinco ciudades de Colombia. *Physis: Revista de Saúde Coletiva* 2015;25:571-592.
45. Instituto de Hidrologia Meteorologia y Estudios Ambientales (IDEAM). [www.ideam.gov.co]. http://institucional.ideam.gov.co/jsp/index.jsf Accessed January 2020, 2020.
46. Bernal R, Cárdenas M. Race and ethnic inequality in health and health care in Colombia. *Documentos de Trabajo (Working Papers). No. 29. Enero 2005*. http://hdl.handle.net/11445/811. Vol. Desempeño de la Salud. Bogota, Colombia: Fundación para la Educación Superior y el Desarrollo (Fedesarrollo). 2005;41.
47. Agudelo-Suárez AA, Martínez-Herrera E, Posada-López A, Rocha-Buelvas A. Ethnicity and Health in Colombia: What Do Self-Perceived Health Indicators Tell Us? *Ethnicity & Disease* 2016;26(2):147-156.
48. Blangiardo M, Cameletti M. Spatial modeling. *Spatial and Spatio-temporal Bayesian Models with R-INLA* John Wiley & Sons, Ltd, 2015;173-234.

49. Chu H, Cole SR. Estimation of Risk Ratios in Cohort Studies With Common Outcomes: A Bayesian Approach. *Epidemiology* 2010;**21**(6):855-862.
50. Shrier I, Pang M. Confounding, effect modification, and the odds ratio: common misinterpretations. *Journal of clinical epidemiology* 2015;**68**(4):470-474.
51. Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *American Journal of Epidemiology* 2004;**159**(7):702-706.
52. Rue H, Riebler A, Sørbye SH, Illian JB, Simpson DP, Lindgren FK. Bayesian computing with INLA: a review. *Annual Review of Statistics and Its Application* 2017;**4**:395-421.
53. R Core Team. R Foundation for Statistical Computing. Desktop 1.0.143. R version 3.6.0 - "Planting of a Tree" ed. Vienna, Austria: R Foundation for Statistical Computing, 2019.
54. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, Committee RW. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Medicine* 2015;**12**(10):e1001885.
55. Erreygers G, Van Ourti T. Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: A recipe for good practice. *Journal of Health Economics* 2011;**30**(4):685-694.
56. Muff S, Riebler A, Held L, Rue H, Saner P. Bayesian analysis of measurement error models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2015;**64**(2):231-252.
57. Pan American Health Organization / World Health Organization. Epidemiological Update: Dengue. Washington, D.C.: PAHO/WHO, 2019;15.
58. Nuru-Jeter AM, Michaels EK, Thomas MD, Reeves AN, Thorpe RJ, LaVeist TA. Relative Roles of Race Versus Socioeconomic Position in Studies of Health Inequalities: A Matter of Interpretation. *Annual Review of Public Health* 2018;**39**(1):169-188.
59. Hagenlocher M, Delmelle E, Casas I, Kienberger S. Assessing socioeconomic vulnerability to dengue fever in Cali, Colombia: statistical vs expert-based modeling. *Int J Health Geogr* 2013;**12**.
60. Arauz MJ, Ridde V, Hernández LM, Charris Y, Carabali M, Villar LÁ. Developing a Social Autopsy Tool for Dengue Mortality: A Pilot Study. *PLoS ONE* 2015;**10**(2):e0117455.

61. Messer LC, Oakes JM, Mason S. Effects of Socioeconomic and Racial Residential Segregation on Preterm Birth: A Cautionary Tale of Structural Confounding. *American Journal of Epidemiology* 2010;**171**(6):664-673.
62. Guarnizo-Herreño CC, Watt RG, Garzón-Orjuela N, Suárez-Zúñiga E, Tsakos G. Health insurance and education: major contributors to oral health inequalities in Colombia. *Journal of Epidemiology and Community Health* 2019;**73**(8):737-744.
63. Honorato T, Lapa PPDA, Sales CMM, Reis-Santos B, Tristão-Sá R, Bertolde AI, Maciel ELN. Spatial analysis of distribution of dengue cases in Espírito Santo, Brazil, in 2010: use of Bayesian model. *Revista Bras Epidemiol* 2014;**17**.
64. Riebler A, Sørbye SH, Simpson D, Rue H. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research* 2016;**25**(4):1145-1165.

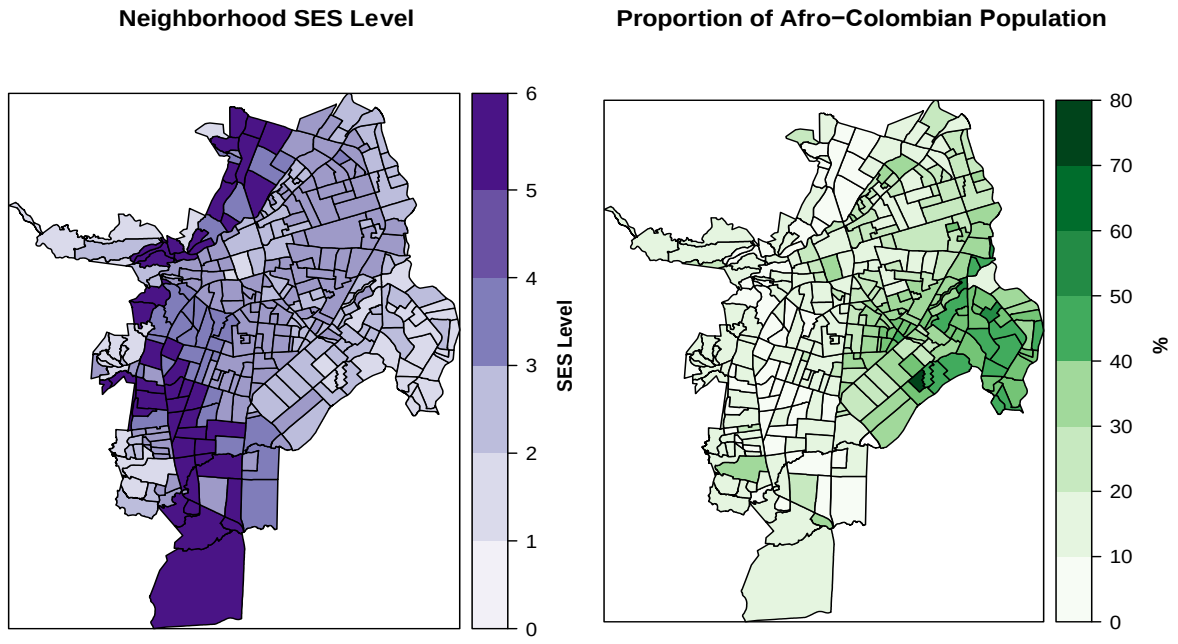
7.3 Supplementary Material Manuscript 4

Supplementary Material: Dengue, Severity Paradox and Socioeconomic Distribution among Afro-Colombians

1. Descriptive information

eTable 1a. Descriptive characteristics of additional covariates at the neighborhood level

Covariate	Median (IQR)	Min- Max
Overall cases per neighborhood per month	2 (1- 5)	1- 101
Non-Afro-Colombian cases per neighborhood per month	1 (1- 5)	1- 98
Afro-Colombian Cases per neighborhood per month	1 (1- 1)	1- 11
Population per neighborhood	5,684 (3,439 - 11-220)	426- 52,853
% of *Aedes*-Positivity catch basins per district	2.4 (1.3- 3.6)	0.4- 6.0
Temperature (highest °C) per month	34.9 (33.7- 35.8)	32.5- 37.3
Relative Humidity	99 (98- 99)	96- 100
Total Precipitation per month (ml)	91.5 (34.8- 153.5)	3.9- 212.1

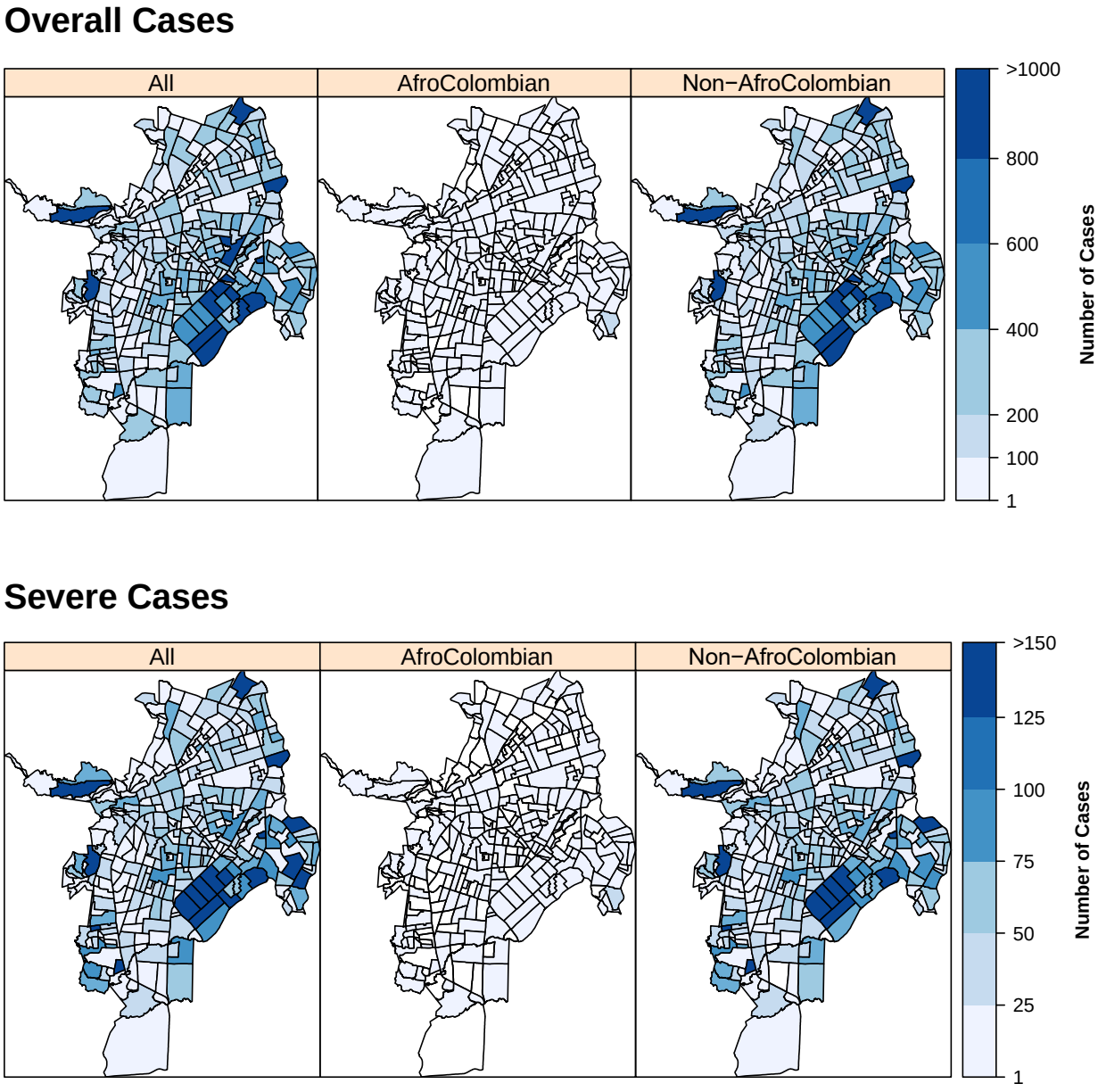


eFigure 1. Neighborhood's socioeconomic status (SES) and proportion of Afro-Colombians per neighborhood in Cali, Colombia (2012- 2017).

eTable 1b. Descriptive characteristics (proportions) of covariates distribution in the entire population of Cali, Colombia.

Covariates	All Population (n=2,496,442)	Non-Afro-Colombians (n=1,842,374; 73.8%)	Afro-Colombians (n=654,068; 26.2%)
Sex			
Males	47.8	47.7	47.9
Females	52.2	52.3	52.1
Insurance			
Contributory Insurance	67.0	64.2	50.7
Subsidized Insurance	33.0	35.8	49.3
Age Group			
<9 years	14.3	17.0	19.0
10-19 years	15.4	18.3	20.5
20-39 years	31.9	32.3	33.9
40-59 years	24.4	22.4	20.2
>60 years	14.0	10.0	6.4
SES Level*			
SES 1 (Lowest)	15.4	-	-
SES 2 (Low)	20.8	-	-
SES 3 (Medium - Low)	25.7	-	-
SES 4 (Medium)	15.8	-	-
SES 5 (Medium High)	15.4	-	-
SES 6 (High)	6.9	-	-

*SES Level is measured at the household and neighborhood levels not at the individual level or discriminated by ethnicity.



eFigure 2. Distribution of dengue cases in Cali, Colombia (2012- 2017). Top: Total number of dengue cases reported per neighborhood during the study period. Bottom: Total number of severe cases reported per neighborhood during the study period.

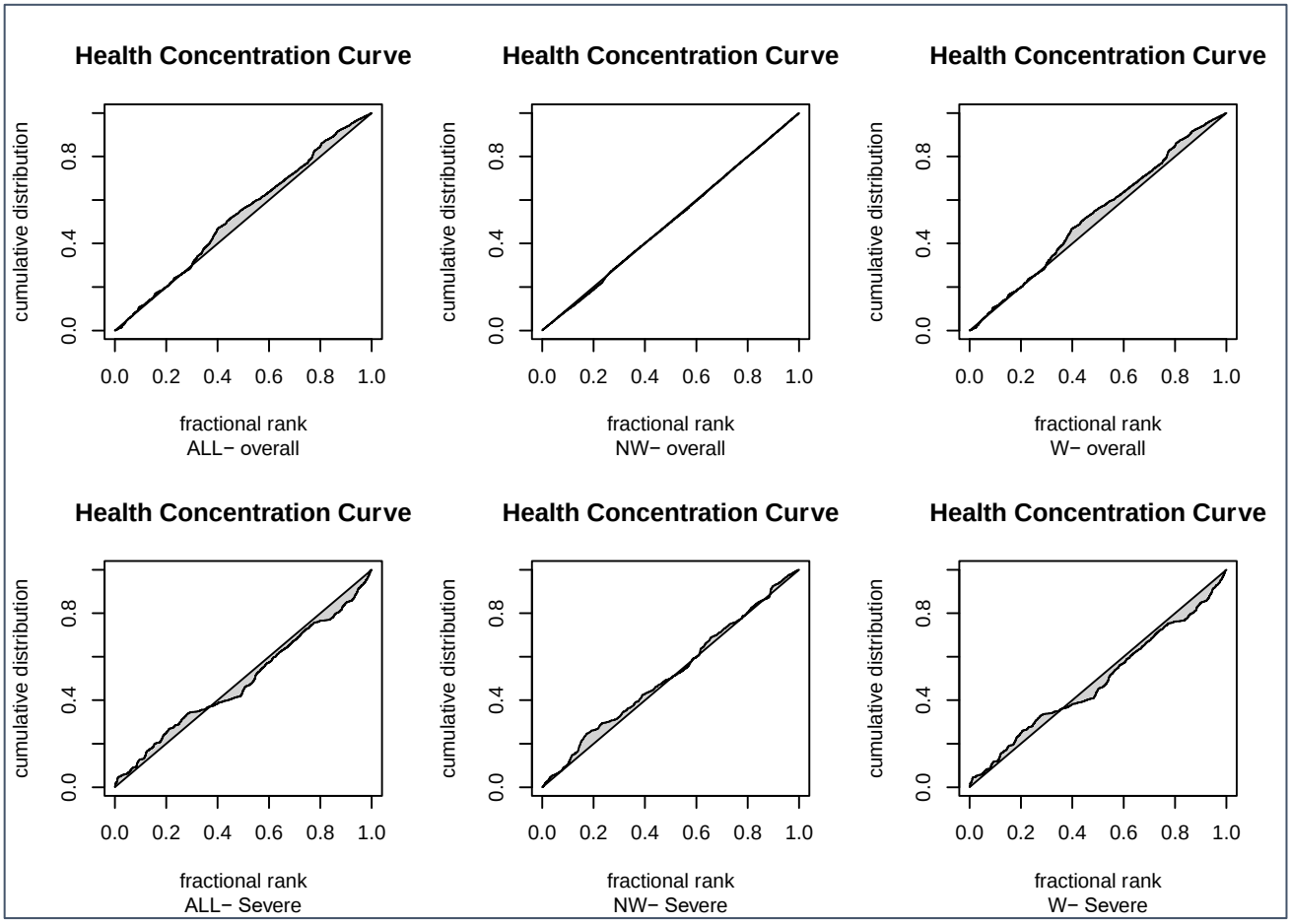
2. Results from the Sensitivity analysis and additional models

eTable 2. Model for the overall spatial distribution of notified dengue cases for all dengue cases and ethnic-specific models using the six-level SES covariate.

	All Dengue Cases (n=65,402)			Non-Afro-Colombians (n=62,300)			Afro-Colombians (n=3,102)		
Covariate	IRR	2.5 %	97.5%	IRR	2.5 %	97.5%	IRR	2.5 %	97.5%
Proportion of Non-Afro-Colombian cases	0.99	0.99	1.00						
Proportion of Male cases	1.03	1.02	1.03	1.02	1.02	1.03	1.00	0.98	1.01
Proportion of Cases Under 20 years old	1.00	1.0	1.01	0.98	0.97	0.98	1.00	0.98	1.02
Proportion of Cases in the Subsidized Scheme	1.02	1.02	1.02	1.01	1.01	1.02	1.00	0.98	1.02
Socio Economic Status (SES) Index									
SES 1 (Very Low)	Ref.	-							
SES 2 (Low)	1.46	1.19	1.79	1.50	1.22	1.83	1.00	0.89	1.12
SES 3 (Low-Medium)	1.27	1.05	1.54	1.27	1.05	1.54	0.95	0.84	1.07
SES 4 (Medium)	1.00	0.76	1.3	0.95	0.73	1.25	1.03	0.85	1.25
SES 5 (Medium-High)	0.88	0.67	1.14	0.87	0.66	1.13	1.01	0.8	1.28
SES 6 (High)	0.81	0.55	1.19	0.77	0.53	1.14	1.19	0.8	1.72

eTable 3. Relative Concentration Index (RCI)and Absolute Concentration Index of overall and severe cases distribution across SES-level

	Overall Dengue			Severe Dengue		
Population	RCI	95% CIs	ACI	RCI	95% CIs	ACI
All Dengue Cases	-0.05	(-0.06, -0.04)	-0.03	0.02	(0.01, 0.03)	0.001
Non-Afro-Colombians	-0.06	(-0.07, -0.05)	-0.03	0.03	(0.02, 0.04)	0.001
Afro-Colombians	0.004	(-0.001, 0.01)	0.001	-0.04	(-0.07, -0.01)	-0.01



eFigure 3. RCI for overall distribution (top) and severe cases (bottom). The curve above the diagonal indicates concentration among the poor. Curves crossing the diagonal are indicative of non-monotonic relationship between rates of diseases and socioeconomic status. ALL= All the population; NW= Non-Whites/Afro-Colombians; W=Whites/Non-Afro-Colombians.

eTable 4. Model for the overall spatial distribution of notified dengue cases including entomological covariates.

Covariate	IRR	2.5%	97.5%
Proportion of Non-Afro-Colombian cases	0.99	0.98	1.01
Proportion of Male cases	1.03	1.02	1.03
Proportion of Cases Under 20 years old	1.00	0.99	1.01
Proportion of Cases in the Subsidized Scheme	1.02	1.02	1.03
% Aedes-Positivity of catch basins	1.12	1.06	1.19
Low SES (Levels 1 and 2)	Ref.	-	-
Medium SES (Levels 3 and 4)	0.98	0.86	1.13
High SES (Levels 5 and 6)	0.72	0.59	0.89

eTable 5. Model for the overall spatial distribution of notified dengue cases including precipitation and temperature covariates as random effects.

Covariate	IRR	2.5%	97.5%
Proportion of Non-Afro-Colombian cases	0.99	0.98	1.00
Proportion of Male cases	1.02	1.02	1.03
Proportion of Cases Under 20 years old	1.00	1.00	1.01
Proportion of Cases in the Subsidized Scheme	1.02	1.01	1.02
Low SES (Levels 1 and 2)	Ref.	-	-
Medium SES (Levels 3 and 4)	0.91	0.78	1.05
High SES (Levels 5 and 6)	0.68	0.54	0.84
Random-Effect Covariate	Precision	2.5%	97.5%
Precipitation	21847.9	1483.5	73998.0
Temperature	2261.479	25.4	14302.6

3. Results from models accounting for underreporting with updated Priors

eTable 6. Model for the overall spatial distribution of notified dengue cases using informative priors.

Covariate	IRR	2.5%	97.5%
Updated Proportion of Non-Afro-Colombian cases	1.00	0.99	1.00
Proportion of Afro-Colombians per neighborhood	1.07	1.04	1.11
Proportion of Male cases	1.01	1.01	1.02
Proportion of Cases Under 20 years old	1.00	1.00	1.00
Updated Proportion of Cases in the Subsidized Scheme	1.00	1.00	1.00
Low SES (Levels 1 and 2)	Ref.	-	-
Medium SES (Levels 3 and 4)	1.02	0.93	1.12
High SES (Levels 5 and 6)	0.90	0.78	1.04

4. Underreporting and Misclassification correction

Surveillance systems are vital components of disease control programs to understand disease burden, trends, and to detect outbreaks. However, underreporting can bias estimates and greatly reduce surveillance data utility. Dengue is a major public health issue in Colombia and despite its mandatory reporting to the National Surveillance System in Colombia (SIVIGILA), it has been reported that the system captures a small proportion of the actual burden. To correct for the potential underreporting and misclassification of dengue, we used rates from an observational capture-recapture study in three Colombian cities from 2014-2017 (Carabali M, et al., 2020. Unpublished Manuscript).

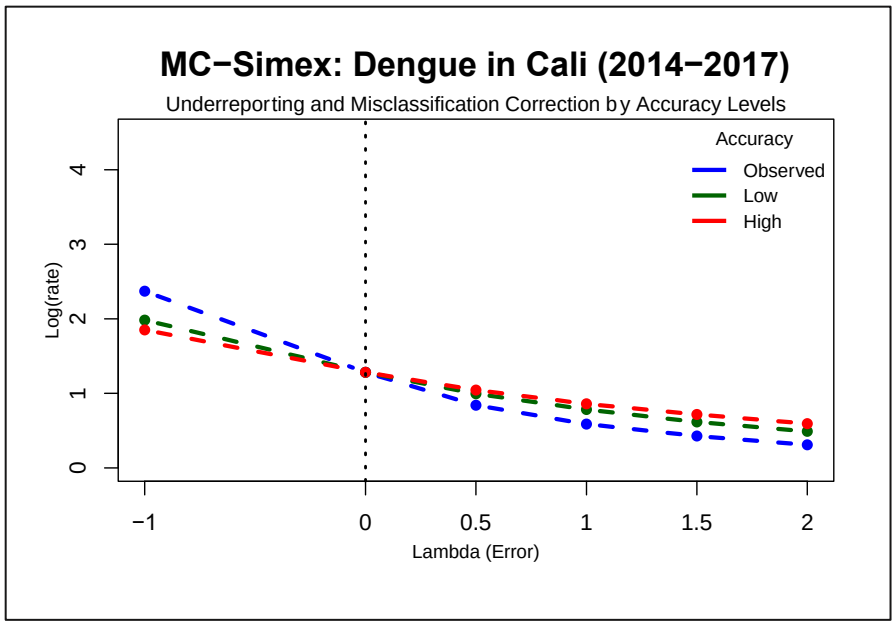
The study used registries of cases from healthcare facilities and surveillance offices in each city, to identify cases that were diagnosed at the healthcare facilities (capture) and those that were reported to SIVIGILA (recapture). To identify predictors of reporting and estimate the probability of reporting by disease and year, robust Poisson regressions were fit adjusting for age, sex, insurance scheme, and year of notification. To account for the potential misclassification of clinical diagnosis of other arboviruses, a simulation extrapolation for misclassification (MC-SIMEX)[1-3] method was used. The results for Cali indicated that dengue reporting ranged from (21-70%) depending on year of diagnosis and type of insurance. The MC-SIMEX analysis indicated that naïve estimates were consistently lower (log-rate=1.28) than the corrected results (log-rates range=1.84 to 2.36), indicating underestimation of the reporting due to the potential misclassification bias.

To correct for the underreporting in the current study and in order to capture the variability described in the literature[4-8], we simulated data using variable ranges of possible overall underreporting (min=10, max=90%). Based on the estimates of underreporting from a similar Colombian setting (31%), we also explored a narrower range of values for overall underreporting (15-40%). We used the validation data from Cali to assess the likelihood of dengue reporting by age, sex, insurance scheme, year of notification and ethnicity (eTable 7). Given that the capture-

recapture study showed differential underreporting by ethnicity and insurance scheme accounting for the potential misclassification (eFigure 3), we used the estimated parameters to inform the range of values used to account for underreporting by Afro-Colombians (5-50%) and by Subsidized insurance scheme (1-30%) in the current study. Overall the simulated data was constructed using parameters from the distribution in the original data set, correcting specifically the outcomes (count of cases per neighborhood per month and presence of severity) accounting for the overall, ethnic, and insurance specific underreporting and misclassification. Examples with the parameters of the simulation are specified in the coding appendix.

eTable 7. Results from the underreporting analysis in Cali from 2014-2017, indicating the rate of dengue reporting by each covariate in the model.

Covariate	IRR	2.5%	97.5%
Under 18 years old	Ref.	-	-
Over 18 Years	4.1	3.8	4.43
Female Sex	0.67	0.64	0.71
2014	Ref.	-	-
2015	1.22	1.12	1.32
2016	0.95	0.88	1.03
2017	0.92	0.83	1.03
Afro-Colombians	0.77	0.67	0.89
Subsidized Insurance	0.91	0.85	0.97



eFigure 4. Simulation Extrapolation results: Measurement Error estimation of dengue reporting, adjusted by age and year of notification in Cali, Colombia between 2014-2017.

The estimated rates on the log scale are presented on the Y axis and the degree/change of misclassification error (Lambda) on the X axis. The dotted line at error level zero in the X axis indicates the observed rates (i.e: uncorrected rates) and the estimates at the error level -1, indicates the extrapolated (corrected) rates accounting for the misclassification error. Estimates from error level 0.5 to error level 2, indicate the change in the estimated rates of reporting by adding different levels of misclassification error.

Results of the correction for misclassification are presented according to different values of accuracy (sensitivity) for the diagnosis of dengue, compared to other arboviruses present during that time (2014-2017). The MC-SIMEX procedure for the validation study included the use of three sets of correlation matrixes for each city: 1) a low sensitivity/specificity matrix were the correlation with the main diagnosis was set to 70%; 2) a high sensitivity/specificity matrix were the correlation to the main diagnosis was set to 90%; and 3) a correlation matrix using the observed data, this is, the correction was made using the actual observed correlation between the institutional diagnosis (capture) and the notified diagnosis (recapture), having the clinical institutional diagnosis as the reference.

eTable 8. Model for the spatial distribution of severe dengue cases for the overall population and ethnic-specific models using the simulated data accounting for underreporting and misclassification.

Covariate	RR	2.5%	97.5%
Afro-Colombian (Corrected)	1.02	0.97	1.07
Under 9 years	Ref.		
10-19 years	1.00	0.94	1.06
20-39 years	0.98	0.93	1.04
40-59 years	0.97	0.92	1.03
>60 years	0.98	0.90	1.06
Male Sex	1.02	0.98	1.05
Subsidized Insurance (Corrected)	1.01	0.97	1.05
Low SES (Levels 1 and 2)	Ref.	-	-
Medium SES (Levels 3 and 4)	1.07	0.85	1.35
High SES (Levels 5 and 6)	0.72	0.49	1.07

5. RECORD checklist (Attached)
6. Coding appendix (R-Markdown attached)

References

1.Küchenhoff H, Mwalili SM, Lesaffre E. A General Method for Dealing with Misclassification in Regression: The Misclassification SIMEX. *Biometrics.* 2006;62(1):85-96.

2.Lederer W, Küchenhoff H, Lederer MW, by Küchenhoff M. Package 'simex'. 2019.

3.Lederer W, Küchenhoff H. A short Introduction to the SIMEX and MCSIMEX. *The Newsletter of the R Project Volume 6/4, October 2006.* 2006;6:26.

4.Vong S, Goyet S, Ly S, et al. Under-recognition and reporting of dengue in Cambodia: a capture–recapture analysis of the National Dengue Surveillance System. *Epidemiol Infect.* 2011;140(3):491-499.

5.Gibbons CL, Mangen M-JJ, Plass D, et al. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. *BMC Public Health.* 2014;14(1):147.

6.Toan NT, Rossi S, Prisco G, Nante N, Viviani S. Dengue epidemiology in selected endemic countries: factors influencing expansion factors as estimates of underreporting. *Trop Med Int Health.* 2015;20(7):840-863.

7.Carabali M, Lim JK, Palencia DC, et al. Burden of dengue among febrile patients at the time of chikungunya introduction in Piedecuesta, Colombia. *Tropical Medicine & International Health.* 2018;23(0):1231-1241.

8.Romero-Vega L, Pacheco O, la Hoz-Restrepo Fd, Díaz-Quijano FA. Evaluation of dengue fever reports during an epidemic, Colombia. *Revista de saude publica.* 2014;48:899-905.

Chapter 8: **Overall Discussion**

The overall goal of this dissertation was the identification of places of arboviral concentration and arboviral distribution across populations with respect to health inequalities. The collective results from this dissertation corroborate the endemic and epidemic character of the arboviral distribution in Latin America. Likewise, I provided robust evidence of the presence and quantified the socioeconomic disparities in Brazil and Colombia. I found that factors such as time (year of notification) and sanitation as socio-environmental variables, and age as individual characteristic, are the determinants contributing the most to the overall socioeconomic inequality towards the poor. Finally, the analysis of dengue distribution across ethnic groups, indicate that the paradoxical increase of severity among Afro-Colombians is likely the result of a differential health seeking behavior among Afro-Colombians and intersectionality between SES, insurance scheme and ethnicity.

In addition to the findings presented here, I contributed to the field of health inequalities on infectious diseases by adapting and applying existent epidemiological and spatiotemporal methods in the context of arboviral diseases. Below I provide an overall discussion on the findings related to the overall goal of this dissertation and a methodological discussion related to the advantages and limitations of the approaches used in this dissertation.

8.1 Summary of Findings

Identification of Places (Spatial Areas) of Arboviral Concentration

All the findings presented in this dissertation confirm dengue endemicity and presence of outbreaks every two to three years in the study sites [6,8,9,35,40,60,114,120,201]. Chikungunya and Zika were introduced as outbreaks but later established as endemic conditions as well [32,33,37,88,120]. In addition, the spatial analyses included in all manuscripts allowed the identification of disease clustering within the study sites, the identification of spatial patterns of distribution by disease and across ethnic groups. Overall, indicating that the spatial distribution of arboviruses does not only depend on spatial/area-level specific characteristics[8,33,89,92,100,202-204]. Dengue, as a known condition, was clustered among the known disadvantaged areas in Fortaleza, Medellin and Cali.

The widespread occurrence of Zika in Fortaleza, and chikungunya and Zika in Medellin indicated that the newly introduced arboviruses affected all spatial areas similarly. These findings demonstrate that the spatial distribution of these arboviruses is likely the result from the interaction of several aspects including the endemicity of the arbovirus, the introduction of new diseases and or presence of outbreaks, the presence of some socio-environmental characteristics of the spatial areas such as sanitation, and characteristics of the populations residing in those spatial areas [6,88,89,92,100].

Assessment of Arboviral Distribution Among Different Populations

The spatial areas of disease clustering identified above were also considered disadvantaged or vulnerable areas. This was expected, given that the body of literature on arboviruses highlights the presence of arboviruses in low socioeconomic settings [14,24,28,49,83,87,92,100,205-207]. This dissertation went beyond the localization of clustering areas and provide the quantification of the socioeconomic inequality on dengue, chikungunya and Zika, estimate the contribution of some measured socio-environmental covariates to the inequality, and analyses the distribution across different populations.

Presence, Quantification and Trend of Socioeconomic Inequalities

The Relative and Absolute Concentration index of Inequality (RCI and ACI) confirmed the concentration of arboviral diseases among people at the bottom of the socioeconomic distribution in Brazil and Colombia. However, the magnitude of the socioeconomic inequality changed across study sites, by disease and over time.

The RCIs for dengue were consistently less than zero with ACIs ranging from -0.05 to -0.37 in Fortaleza and from -0.01 to -0.04 in Medellin. These results showed a constant presence of socioeconomic inequality towards the most disadvantaged groups for dengue. The inequality ranged in magnitude during outbreaks and during interepidemic periods. Likewise, although conducted as a sensitivity analysis, I observed the presence of inequalities towards the poor on

dengue for the overall population (RCI= -0.05; 95%Cr.I = -0.06, -0.04) and for severity (RCI= -0.04; 95%Cr.I= -0.07, -0.01) among Afro-Colombians.

The largest variation in the magnitude of inequality was observed for chikungunya in Fortaleza. The ACIs ranged from less than 0.001 in 2014 to -0.37 in 2017. These results indicated changes from absence of inequality or even slight concentration among the better-off populations during 2014, to the presence of inequality towards the poor during the outbreak period in 2016 and 2017. Interestingly, the arboviruses that were widely spread or with outbreaks of lesser magnitude did not display larger socioeconomic inequalities, as in the case of Zika in Fortaleza and chikungunya and Zika in Medellin. Thus, corroborating that at introduction or during outbreaks, an infectious disease could affect the entire population. However, once the disease is established, the most affected are often the most disadvantaged populations [32,33,88,89,120].

The importance of the endemic and epidemic contribution to the socioeconomic inequality was also evidenced with the decomposition analysis conducted in the third manuscript. This analysis allowed the identification of the year of notification and sanitation variables as the main contributors to the overall inequality on the studied arboviruses in both study sites. While year of notification, availability of healthcare facilities and age of the case were the main contributors in Fortaleza, age and waste management were the main contributors to the socioeconomic inequality in Medellin. Overall, this indicates the importance of general socio-environmental conditions in the presence of arboviruses, but stresses that the role and contribution of each determinant is context-specific[24,28,100,204,207].

Arboviral Distribution Across Different Socioeconomic Strata

In all manuscripts, it was evidenced the presence of a non-monotonic relationship between disease distribution and socioeconomic distribution. The analyses from Brazil showed concentration of arboviruses at lower levels of income but the pattern of disease distribution changed by disease. Specifically, dengue rates in Fortaleza increased at median monthly household income below US$450 and decreased at median monthly household income above

US$950. Zika rates were similar across the spectrum of income and chikungunya rates were variable with several fluctuations across the spectrum of income.

Analyses from Colombia showed that although the majority of reported cases were from low-SES and medium-SES level neighborhood, when comparing those two groups (low vs. medium), disease rates were lower among the lowest-SES strata or not different across the two levels. In the first manuscript, compared to medium-SES level, adjusted standardized rate ratios (SRR) of dengue in Medellin were lower among low-SES neighborhoods (SRR=0.45; 95% Cr.I= 0.34, 0.59) and not different among high-SES level neighborhoods (SRR=0.78; 95% Cr.I= 0.56, 1.09).

In the second manuscript, the non-monotonic relationship between SES and disease distribution in Medellin was more evident for dengue. Specifically, dengue rates were higher among neighborhoods with SES levels 2 to 4 (i.e., low to medium SES) and lower at the SES levels 1 and 6, the lowest and highest-SES levels, respectively. While chikungunya and Zika rates were similar across the spectrum of the SES level. In the fourth manuscript, compared to the low-SES level, incidence rate ratios (IRR) for dengue were lower among neighborhoods with high-SES levels (IRR=0.68; 95%Cr.I=0.55, 0.84) and not different among neighborhoods with medium-SES levels (IRR=0.97; 95%Cr.I=0.84, 1.11). Interestingly, analyses using the six-level SES index showed similar results than those observed in Medellin (i.e., lower dengue rates at the lowest and highest-SES levels).

As indicated in each manuscript, lower disease rates among the people at the bottom (lowest-SES) of the socioeconomic distribution could be attributed to several factors, including barriers to accessing health care. As barriers, I considered not only the presence of health coverage, given that a sort of universal health care coverage exist in Colombia and the analysis were adjusted for the type of insurance. Additional barriers include, geographical distance and availability of health care facilities, which was considered one of the main contributors to the inequality in the third manuscript. Likewise, despite the presence of health coverage, individuals with subsidized insurance do not enjoy the benefits of a contributory insurance (e.g., paid medical leaves,

coverage of prescriptions) or simply, most disadvantaged individuals may not count with the financial means to pay for transportation and other costs associated to the seek medical care[14,37,49,88,204,206]. Contrary to what is expected from people living in areas at medium-SES levels. Because, regardless of the magnitude of arboviral presence among these populations, benefits related to their socioeconomic position (e.g., contributory insurance, financial ability to afford co-payments, or to pay for medications, etc.) may increase the likelihood of their ascertainment and therefore their notification[14,33,37,49,120]. Hence, the relative higher presence of arboviruses among people living in areas at medium-SES levels.

The Paradox of Ethnic Inequalities on Dengue Severity

The fourth manuscript examined the overall and severe dengue distribution between Afro-Colombians and Non-Afro-Colombians. Previous reports indicated increased risk of severity and mortality among Black people or people with African Ancestry, despite it is considered protective for severe forms of dengue[15,19,21,96,208]. In my analysis, I observed a similar rate of non-severe dengue distribution across ethnic groups and a small increase of dengue severity among Afro-Colombians. However, after accounting for overall underreporting and differential underreporting by ethnicity and insurance scheme, the relative increased severity among Afro-Colombians disappeared. These findings, therefore, indicate that the "ethnicity paradox" for dengue severity could be attributed to a differential reporting among Afro-Colombians. More generally, these results indicate that the reported increased dengue severity observed in this study and elsewhere in the Americas regions, is likely the result of "intersectionality"[209-211]. Specifically, these findings suggest that neighborhood's SES-level, insurance scheme and ethnicity, interact resulting in a differential health seeking behavior among Afro-Colombians[211-214].

8.2 Methodological Discussion: Strengths and Limitations

This dissertation comprises a large set of methodological approaches aimed to avoid or mitigate the potential threats to validity posed by the study designs and or the nature of the data. Here, I provide an overview of the methodological issues addressed throughout the development of this dissertation and highlight my contribution to the scientific knowledge in each case.

Despite that my analyses rely heavily on modeling techniques and robust statistical approaches, the body of the evidence presented here is still mainly descriptive. Although extremely useful and informative, these results are not intended to be interpreted as causal and are not necessarily conclusive. Not only because of the type of analyses or methods used, but because the presence and distribution of arboviral diseases is multifactorial, with socio-environmental factors being only a part of these multifactorial determinants [9,28,49,204].

Use of Surveillance Data

As mentioned throughout the manuscripts, the use of surveillance data possesses several challenges. First, the potential for selection bias by conditioning only on individuals reported in the surveillance system[8,25,159,161,215]. Second, the potential of measurement error due to underreporting and misclassification of the arboviral disease[215,216]. To account for this, in all manuscripts I conducted sensitivity analysis using either informative priors or simulated-based corrections or both. In addition, for the misclassification, the analysis of the second and third manuscripts included the assessment of the inequality for all arboviruses combined. However, I could not completely rule-out any additional underreporting associated to specific practices by health care providers, mostly for Brazil [85,136]. In Colombia, however, the use of priors and the simulated-based analysis were informed by a parallel study on underreporting in the same or similar settings.

Nonetheless, as reflected in each analysis, the potential effect of the use of surveillance data is therefore an underestimation of the effects[161,215]. Specifically, it is probable that the measures of RCI and ACI obtained here are biased towards the line of equality and that the rates of disease across SES-levels are biased towards the null. Thus, indicating that the magnitude of the socioeconomic inequality towards the poor is even larger than the one identified here and that the disease rates are even higher among low-SES settings. Nonetheless, these postulates could be only assessed by the means of cohort studies covering epidemic and interepidemic periods, large populations and using reliable measures for socioeconomic identification and disease

ascertainment. However, such studies are limited or non-existent due high economic and logistic cost that such studies embody [143,157,202,216].

Spatiotemporal analysis

The use of Bayesian hierarchical spatial and spatiotemporal analysis improved the assessment of spatial areas of concentration by handling uncertainty, borrowing information from neighboring areas, and accounting for the vicinity of neighborhoods, a special aspect to account for in communicable diseases[164,174,178]. The use of spatiotemporal structured random effects provided extra flexibility to the analysis by allowing each spatial area to have their own baseline disease rates and accounting for their changes over time.

The first manuscript included the application of a joint spatial marked point process model, a method borrowed from the field of ecology[180]. To the best of my knowledge, this approach has not been used before in the field of arboviruses. Here the use of this approach, allowed the use of individual data for spatial inference of disease distributions. The benefits of this method included increase of precision, decrease the potential for ecological fallacy, and the assessment of two conditional outcomes simultaneously, accounting for the uncertainty associated to the use of surveillance data. However, despite the observed methodological benefits, the application of this method in the rest of the dissertation was not possible due to the lack of individual location (geocoded) data for all arboviral cases during the entire study period in all study sites.

The second and third manuscripts used aggregated data at the neighborhood level. However, the use of structured spatial and temporal random effects and the inclusion of several area-level and disease-related covariates contributed to mitigate the potential residual confounding and potential ecological fallacy. The fourth manuscript used both, individual and aggregated data, although not as a joint model. For these analyses, I integrated the overall reported cases in the analysis of severity by modelling severe cases as a function of the reported cases, their spatial distribution and individual covariates in a separated model. Thus, with the benefit of modeling

dengue severity using individual data, assessing the spatial distribution and accounting for the uncertainty associated to the surveillance data without using the exact location of cases.

Integration of Socioeconomic Inequality methods and Spatiotemporal analyses

To analyze the presence of socioeconomic inequalities on arboviruses it is important to account for their epidemic nature [8,201-203,205]. For this, it was necessary to integrate the inequality assessment within the spatiotemporal framework as presented in the second and third manuscripts. Specifically, in order to account for the spatiotemporal distribution of arboviruses, I estimated the presence of inequalities using the outputs from the spatiotemporal models. This approach allowed the use of (properly) adjusted disease rates to estimate the measures of inequality. By not adjusting for the spatiotemporal distribution of diseases, assuming static or invariant disease rates across the spatial units and over time, the estimates of inequality would likely be biased.

The measures of inequality do not account for the time-varying nature of the outcome, unless a stratified analysis is used or if a fixed-effects variable is included in the disease model [193,195]. However, such approaches would not account for seasonality and would rely on the (arbitrary) selection of a baseline point in time, making the interpretation of the inequality also conditional on the selected departure point in time. To mitigate these methodological limitations, all my analyses included a disease model using structured random effects for the month of occurrence and notification, which accounted for seasonality. In addition, for the overall assessment of inequality, I also conducted stratified analysis by year of notification and for the decomposition analysis I also included indicator variables for year of notification. Thus, the estimation of the inequality is based on a robustly adjusted disease distribution and the interpretation of the inequality could be done for specific points in time or across and over time. Consequently, allowing the estimation of the contribution of the temporal effect to the overall inequality.

Nonetheless, it is important to note that although the use of the spatiotemporal random effects allows the identification of trends overtime and spatial clustering, the residual spatial and

temporal random effects could also capture some additional effects that are not explained by the covariates included in the model [193]. It is expected that some environmental effects such as temperature and precipitation are captured by the residual temporal random effects, and that unmeasured area level specific characteristics are included in the residual spatial effects. However, these residuals could also include some of the variability due to model misspecification [178,193,195].

This consideration is important, mostly for the interpretation of the decomposition analysis. The decomposition is estimated based on the contribution of the fixed-effects covariates [193,195], which in this case also includes indicator variables for the year of notification. Therefore, in addition to account for the seasonality, the specific contribution of the year of notification is estimated and the residual effect due to time is likely reduced. However, the proportion of the inequality that is unexplained and usually captured by the residual effect the fixed-effects component, here will also include some of the residual spatial effect. The interpretation of the decomposition should therefore be made considering that i) the contribution of the covariates to the proportion explained is conditional on the measured covariates, and ii) the unexplained proportion includes both, the proportion unexplained by the fixed-effects and by the random-effects component.

Overall Strengths

Despite of the mentioned limitations and mitigation strategies mentioned above, this dissertation presents several strengths that are worth highlighting. First, this dissertation includes one of the first quantification of the socioeconomic inequalities on arboviral diseases, accounting for the spatiotemporal distribution and documenting the trend over time. Second, even if the use of surveillance data is seen as a disadvantage, surveillance data is the main source of data for policy making on arboviral diseases. Here, I used several methods to improve the use and interpretability of surveillance data for the analysis of arboviruses and health inequalities. Likewise, the use of surveillance data allowed the use of a non-negligible amount of information. I analyzed over 300,000 cases reported over a decade in three cities in Colombia and Brazil. This amount of data could not be (as easily) obtained from cohort or case-control studies.

Finally, as evidenced in the manuscripts and in the methodological discussion, one of the main strengths of this dissertation is the large, robust and comprehensive set of methodological approaches intended to overcome issues such as confounding, selection bias and measurement error. This dissertation provides examples of novel application of existent methods, facilitated the creation and modification of software functions for the estimation and decomposition of health inequalities from Bayesian spatiotemporal models, and showed the need and possibility of integrating spatiotemporal models and methods for the assessment of inequalities.

8.3 Relevance, Implications and Opportunities for future research

There could not be a more relevant time to share the results of my dissertation with the scientific community. The year 2019 included one of the largest dengue outbreaks in the Americas. During 2019, Brazil reported over two million dengue cases and Colombia more than 100 thousand cases. This epidemic was extended to April 2020, where Brazil had reported over a million cases and Colombia more than 55 thousand dengue cases. Unfortunately, this does not indicate that the dengue outbreak ended there. The dengue outbreak was only overshadowed by the pandemic originated by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which cause the coronavirus disease (COVID-19)[217].

The collective results of my dissertation provide the motivation and plausible support to conduct further analysis on the field of inequalities on arboviruses and for other infectious diseases. I could see several venues in which this dissertation opens the door to future research on the field. For instance, I consider that to corroborate the findings presented in this dissertation, all the analyses conducted here could and should be used for the estimation of inequalities with data from cohort studies or using data that is not subjected to underreporting. Specifically, this would contribute to provide robust estimates that are less sensitive to the underestimation, either by under-ascertainment or underreporting of cases, and also using more accurate measures of socioeconomic position or class. These studies are required to identify and provide a more accurate figure of the magnitude of the overserved inequality and therefore the contribution of their determinants.

The use of data from cohort-studies or in general studies with less underestimation of diseases, would also provide insightful and necessary information on the ethnic disparities. Such studies would allow better characterization of overall and severe cases and would hopefully facilitate the qualitative and quantitative identification of the intersectionality effect of socioeconomic aspects and ethnicity. At early stages of the conception of this dissertation, I was motivated by the use of Oaxaca-Blinder decomposition and other statistical approaches to disentangle and understand better the presence of ethnic inequalities on arboviruses [218,219]. However, this approach required further statistical verification and modification of the methods for their use in the context of surveillance data and non-linear (Poisson or Negative Binomial) model [219-221]. Thus, I consider this analysis as a definitive next step for future research on the field.

Likewise, the estimation of inequality indexes and the analysis of socioeconomic decomposition within the mixed-effect models deserves further study. Although the estimation uses a robustly adjusted outcome, the estimation of the decomposition includes some of the residual random effects in the unexplained portion [193,195]. The characterization of this unexplained contribution or including the spatial or temporal residual contribution into the explained proportion could therefore be informative in this context. This was out of the scope for my dissertation but further development in this area of research will greatly contribute to the characterization of the socioeconomic inequalities.

Lastly, and somehow more importantly, the analyses conducted in this dissertation have the opportunity to be applied in any other infectious diseases. All caveats acknowledged, these approaches could be used either with robust data (e.g., from cohort studies) or with imperfect data (e.g., from surveillance data). Currently, lots of concerns have been raised due to the apparent inequalities around COVID-19. From socioeconomic inequalities to ethnic inequalities, the discussion includes questions about why and by how much the disease seems to affect most disadvantaged populations [217,222-225]. With a pandemic of such magnitude we could believe that the SARS-CoV-2 would be blinded for socioeconomic aspects or skin color, as it is also expected

for the viruses transmitting dengue, chikungunya and Zika. However, as evidenced in both scenarios, this is not the case. Not necessarily because the virus has a tropism for Black people and people at the bottom of the socioeconomic distributions, but because the conditions surrounding them make the presence of these viruses more evident and or more difficult to handle [37,49,92,203,223,225].

For illustration of the scope and opportunities provided by this thesis, I am allowing myself to make a parallel between the inequalities in arboviruses and the inequalities in novel coronavirus disease. At the beginning of the COVID-19 pandemic, in the Americas and other places outside China, most of the cases were travelers (business or tourism) and or people at the higher end of the socioeconomic distribution[217,222]. This was also the case of chikungunya and Zika cases during their introduction in the Americas in 2014 and 2015[35]. Once the COVID-19 pandemic was declared, we start observing cases across the entire socioeconomic distribution, the same occurred for chikungunya and Zika after the declaration of a public health emergency of international concern [32,35,47,222]. However, once the COVID-19 pandemic was stablished in settings such as New York in the United States of America, Montreal in Canada, and Cali in Colombia; a relative and absolute concentration of cases was observed among the most disadvantaged populations[223-225]. This was also the case for all arboviruses studied in this dissertation in all my study sites [14,92,100,157,204].

This rough parallel shows the opportunity that my analyses and approaches have to build upon the body of evidence on other infectious diseases. Although the use of surveillance data may be seen as a limitation, the main available source of data for arboviruses and for COVID-19, may possess some methodological challenges. The use of techniques to account for underreporting is potentially more easily applied in the COVID-19 context, given that under-ascertainment could be lower [217]. Likewise, the analysis of intersectionality for ethnic disparities could be further improved in the context of COVID-19 by including aspects such as test availability, occupation of the cases, and other factors that could interact with ethnicity and result in the higher burden of disease for ethnic minorities in some settings.

Overall, the relevance, implications and opportunities for future research derived from this dissertation are vast. It is relevant for arboviruses in Latin America and applicable to other infectious and non-infectious diseases that could change over time worldwide. The methods used here could be improved and adapted to different data and study designs and definitely useful to understand inequalities on diseases of epidemic potential.

8.4 Conclusion

Inequalities in arboviral diseases in Latin America do exist. Arboviral diseases concentrate in vulnerable and disadvantaged areas and the magnitude of the socioeconomic inequality changes by setting, disease and during epidemic periods. The presence of ethnic disparities, however, is not necessarily attributed to a biological or genetic effect of race, but likely the result of the interaction between social factors and ethnicity. In this dissertation, I provided robust quantitative estimates of the socioeconomic disparities on arboviral diseases that to the best of our knowledge, has not been presented before. The approaches presented here, which are particularly quantitative, have the capacity to expand and improve upon the current body of evidence of both communicable and non-communicable diseases, contributing substantively to the scientific knowledge on health inequalities. Nonetheless, the body of evidence presented here could and should be improved with future research, either using better quality data and or implementing statistical approaches that could better quantify and characterize the presence of inequalities. Finally, although the use of surveillance data could be seen as a disadvantage, disease control strategies in endemic countries are mainly informed by surveillance-based analysis. Therefore, the analyses presented here are justified and considered relevant. Hence, these findings could be used in other endemic and epidemic settings with similar sociodemographic characteristics for policy making, either for disease control and or for targeting strategies to decrease disparities at the local level.

Appendices

A. Notification Forms

Brazil: Notification forms are available in the following link (Portuguese):

http://portalsinan.saude.gov.br/sinan-dengue-chikungunya

Colombia: Notification forms are available in the following link (Spanish)

https://www.ins.gov.co/buscador-eventos/Lineamientos/Dengue%20PROTOCOLO.pdf

B. R-markdowns and or Codes for the analyses

Given the length of the analysis, the R-markdowns and documents for the codes used in this dissertation are also available in PDF and html format and could be accessed here:

Sample Codes for Thesis Dissertation (Password: Carabali2020).

Otherwise, copy-paste the following link: https://bit.ly/2Y4wKxh (Password: Carabali2020)

C. Ethical Approvals

McGill Institutional Review Boards

This thesis received ethical approval by McGill Institutional Review Board on March 12, 2018 (IRB Study Number A02-E05-18A).

Brazil's Ethics Committee

This thesis received ethical approval by Brazil's Ethics Committee, the University of Ceara and the surveillance office in Fortaleza on April 26, 2018 (2.624.599; CAAE: 84463318.3.0000.5534).

Colombia: We received approval to the use of secondary data from the Surveillance office in Cali and Medellin.

REFERENCES

1. Mayer SV, Tesh RB, Vasilakis N. The emergence of arthropod-borne viral diseases: A global prospective on dengue, chikungunya and zika fevers. *Acta Trop* 2017;**166**(Supplement C):155-163.

2. Pan American Health Organization (PAHO). *Tool for the diagnosis and care of patients with suspected arboviral diseases*. 1st Edition ed. Vol. PAHO Strategic Plan 2014-2019. Washington: PAHO, 2017.

3. Patterson J, Sammon M, Garg M. Dengue, Zika and Chikungunya: Emerging Arboviruses in the New World. *Western Journal of Emergency Medicine* 2016;**17**(6):671-679.

4. Rodriguez-Morales AJ, Villamil-Gómez WE, Franco-Paredes C. The arboviral burden of disease caused by co-circulation and co-infection of dengue, chikungunya and Zika in the Americas. *Travel Med Infect Dis.* 2016;**14**(3):177-179.

5. World Health Organization (WHO). Dengue: prevention and control. *Sixty-eight World Health Assembly*. Geneva, Switzerland: World Health Organization, 2015.

6. World Health Organization (WHO). A Toolkit for National Dengue Burden Estimation. In: World Health Organization (WHO), ed. Geneva, 2018.

7. World Health Organization, UNICEF/UNDP/World Bank/WHO Special Programme for Research Training in Tropical Diseases. *Global vector control response 2017-2030*. Geneva: World Health Organization, 2017.

8. Cattarino L, Rodriguez-Barraquer I, Imai N, Cummings DAT, Ferguson NM. Mapping global variation in dengue transmission intensity. *Science Translational Medicine* 2020;**12**(528):eaax4144.

9. World Health Organization (WHO). Dengue and severe dengue. In: WHO, ed. *Health Topics. Fact sheets [Dengue]*. Geneva, Switzerland: WHO, 2020;https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue.

10. Pan American Health Organization., World Health Organization. Zika suspected and confirmed cases reported by countries and territories in the Americas Cumulative cases, 2015-2016. In: PAHO/WHO, ed. Washington, D.C: PAHO/WHO, 2016.

11. Governo do Estado do Ceara. Boletim Epidemiologico: Dengue, Chikungunya e Zika. In: Fortaleza RdS, ed. *Coordenadoria de Promoção e Proteção à Saúde* Fortaleza: Secretaria da Saúde do Estado do Ceará – Região de Saúde de Fortaleza, 2017;12.

12. Instituto Nacional de Salud (INS). Informe epidemiologico de Evento Dengue. In: Vigilancia y Análisis del Riesgo en Salud Pública., ed. Vol. FOR-R02.4000-001. Bogota, Colombia: Instituto Nacional de Salus, 2018;22.

13. Ardila Pinto F, Martínez S, Fuentes M, Borrero E. Análisis de las demoras en salud en personas que enfermaron de gravedad o fallecieron por dengue en cinco ciudades de Colombia. *Physis: Revista de Saúde Coletiva* 2015;**25**:571-592.

14. Arauz MJ, Ridde V, Hernández LM, Charris Y, Carabali M, Villar LÁ. Developing a Social Autopsy Tool for Dengue Mortality: A Pilot Study. *PLoS ONE* 2015;**10**(2):e0117455.

15. Rojas Palacios JH, Alzate A, Martínez Romero HJ, Concha-Eastman AI. AfroColombian ethnicity, a paradoxical protective factor against Dengue. *Colombia Médica* 2016;**47**:133-141.

16. Blanton RE, Silva LK, Morato VG, Parrado AR, Dias JP, Melo PRS, Reis EA, Goddard KAB, Nunes MRT, Rodrigues SG, Vasconsuelos PFC, Castro JM, Reis MG, Barreto ML, Teixeira MG. Genetic ancestry and income are associated with dengue hemorrhagic fever in a highly admixed population. *European Journal of Human Genetics* 2008;**16**(6):762-765.

17. Sierra BdlC, Kourí G, Guzmán MG. Race: a risk factor for dengue hemorrhagic fever. *Archives of Virology* 2007;**152**(3):533-542.

18. Instituto Nacional de Salud (INS). Informes de Evento [Surveillance Reports]. *Evento: Dengue*. Santa Fe de Bogota, Colombia: Publicaciones: Instituto Nacional de Salud de Colombia, 2019;https://www.ins.gov.co.

19. Moraes GH, de Fátima Duarte E, Duarte EC. Determinants of Mortality from Severe Dengue in Brazil: A Population-Based Case-Control Study. *The American Journal of Tropical Medicine and Hygiene* 2013;**88**(4):670-676.

20. Gibson G, Souza-Santos R, Pedro AS, Honório NA, Sá Carvalho M. Occurrence of severe dengue in Rio de Janeiro: an ecological study. *Revista da Sociedade Brasileira de Medicina Tropical* 2014;**47**:684-691.

21. Martins-Melo FR, Ramos AN, Alencar CH, Heukelbach J. Mortality from neglected tropical diseases in Brazil, 2000–2011. *Bulletin of the World Health Organization* 2016;**94**(2):103-110.

22. Paixão ES, Costa MdCN, Rodrigues LC, Rasella D, Cardim LL, Brasileiro AC, Teixeira MGLC. Trends and factors associated with dengue mortality and fatality in Brazil. *Revista da Sociedade Brasileira de Medicina Tropical* 2015;**48**:399-405.

23. Caprara A, Lima JW, Marinho ACP, Calvasina PG, Landim LP, Sommerfeld J. Irregular water supply, household usage and dengue: a bio-social study in the Brazilian Northeast. *Cad Saude Publica* 2009;**25**:S125 - S136.

24. Azoh Barry J. Social Sciences Research on Infectious Diseases of Poverty: Too Little and Too Late? *PLoS Negl Trop Dis* 2014;**8**(6):e2803.

25. Morse SS. Public health surveillance and infectious disease detection. *Biosecur Bioterror* 2012;**10**(1):6-16.

26. Moloughney B. What can public health do to address inequities in infectious disease? *Canada Communicable Disease Report* 2016;**42**(S1):S1_14.

27. Campeau L, Degroote S, Ridde V, Carabali M, Zinszer K. Containment measures for emerging and re-emerging vector-borne and other infectious diseases of poverty in urban settings: a scoping review. *Infectious Diseases of Poverty* 2018;**7**(1):95.

28. Degroote S, Bermudez-Tamayo C, Ridde V. Approach to identifying research gaps on vector-borne and other infectious diseases of poverty in urban settings: scoping review protocol from the VERDAS consortium and reflections on the project's implementation. *Infectious Diseases of Poverty* 2018;**7**(1):98.

29. Bloch D. The Cost And Burden Of Chikungunya In The Americas [Open Access Thesis]. Yale University, 2016.

30. Castro Rodríguez R, Carrasquilla G, Porras A, Galera-Gelvez K, Lopez Yescas JG, Rueda-Gallardo JA. The Burden of Dengue and the Financial Cost to Colombia, 2010–2012. *The American Journal of Tropical Medicine and Hygiene* 2016;**94**(5):1065-1072.

31. Shepard DS, Undurraga EA, Halasa YA, Stanaway JD. The global economic burden of dengue: a systematic analysis. *The Lancet Infectious Diseases* 2016;**16**(8):935-941.

32. Paixão ES, Teixeira MG, Rodrigues LC. Zika, chikungunya and dengue: the causes and threats of new and re-emerging arboviral diseases. *BMJ Global Health* 2018;**3**(Suppl 1):e000530.

33. Mora-Salamanca AF, Porras-Ramírez A, De la Hoz Restrepo FP. Estimating the burden of arboviral diseases in Colombia between 2013-2016. *International Journal of Infectious Diseases* 2020.

34. Thompson R, Martin Del Campo J, Constenla D. A review of the economic evidence of Aedes-borne arboviruses and Aedes-borne arboviral disease prevention and control strategies. *Expert Review of Vaccines* 2020;**19**(2):143-162.

35. Carrillo-Hernández MY, Ruiz-Saenz J, Villamizar LJ, Gómez-Rangel SY, Martínez-Gutierrez M. Co-circulation and simultaneous co-infection of dengue, chikungunya, and zika viruses in patients with febrile syndrome at the Colombian-Venezuelan border. *BMC Infectious Diseases* 2018;**18**(1):61.

36. Espinal MA, Andrus JK, Jauregui B, Hull Waterman S, Morens DM, Santos JI, Horstick O, Francis LA, Olson D. Arbovirosis emergentes y reemergentes transmitidas por Aedes en la Región de las Américas: implicaciones en materia de políticas de salud. *Rev Panam Salud Publica* 2019;**43**:6.

37. Lima Neto AS, Sousa GS, Nascimento OJ, Castro MC. Chikungunya-attributable deaths: A neglected outcome of a neglected disease. *PLoS neglected tropical diseases* 2019;**13**(9):e0007575-e0007575.

38. Pan American Health Organization / World Health Organization. Epidemiological Update: Dengue. Washington, D.C.: PAHO/WHO, 2019;15.

39. Kraemer MUG, Sinka ME, Duda KA, Mylne AQN, Shearer FM, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Elyazar IRF, Teng H-J, Brady OJ, Messina JP, Pigott DM, Scott TW, Smith DL, Wint GRW, Golding N, Hay SI. The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *eLife* 2015;**4**:e08347.

40. Messina JP, Brady OJ, Scott TW, Zou C, Pigott DM, Duda KA, Bhatt S, Katzelnick L, Howes RE, Battle KE, Simmons CP, Hay SI. Global spread of dengue virus types: mapping the 70 year history. *Trends Microbiol.* 2014;**22**(3):138-146.

41. Messina JP, Brady OJ, Golding N, Kraemer MUG, Wint GRW, Ray SE, Pigott DM, Shearer FM, Johnson K, Earl L, Marczak LB, Shirude S, Davis Weaver N, Gilbert M, Velayudhan R, Jones P, Jaenisch T, Scott TW, Reiner RC, Hay SI. The current and future global distribution and population at risk of dengue. *Nature Microbiology* 2019;**4**(9):1508-1515.

42. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O. The global distribution and burden of dengue. *Nature* 2013;**496**:504–507.

43. World Health Organization (WHO). Global strategy for dengue prevention and control 2012-2020. Geneva: WHO. World Health Organization 2012;vi, 43p.

44. Posen HJ, Keystone JS, Gubbay JB, Morris SK. Epidemiology of Zika virus, 1947–2007. *BMJ Global Health* 2016;**1**(2).

45. Yactayo S, Staples JE, Millot V, Cibrelus L, Ramon-Pardo P. Epidemiology of Chikungunya in the Americas. *The Journal of Infectious Diseases* 2016;**214**(suppl_5):S441-S445.

46. Sharp TM, Ryff KR, Alvarado L, Shieh W-J, Zaki SR, Margolis HS, Rivera-Garcia B. Surveillance for Chikungunya and Dengue During the First Year of Chikungunya Virus Circulation in Puerto Rico. *J Infect Dis* 2016;**214**(suppl_5):S475-S481.

47. WHO. Statement on the first meeting of the International Health Regulations (2005) (IHR 2005) Emergency Committee on Zika virus and observed increase in neurological disorders and neonatal malformations. In: statement W, ed. Geneva, Switzerland: WHO, 2016.

48. World Health Organization, Special Programme for Research and Training in Tropical Diseases. *Dengue: guidelines for diagnosis, treatment, prevention and control*. WHO Library Cataloguing-in-Publication Data ed. Vol. New Edition. Geneva: World Health Organization, 2009.

49. Carabali M, Hernandez L, Arauz M, Villar L, Ridde V. Why are people with dengue dying? A scoping review of determinants for dengue mortality. *BMC Infectious Diseases* 2015;**15**(1):301.

50. Salinas JL, Walteros DM, Styczynski A, Garzón F, Quijada H, Bravo E, Chaparro P, Madero J, Acosta-Reyes J, Ledermann J, Arteta Z, Borland E, Burns P, Gonzalez M, Powers AM, Mercado M, Solano A, Sejvar JJ, Ospina ML. Zika virus disease-associated Guillain-Barré syndrome—Barranquilla, Colombia 2015–2016. *Journal of the Neurological Sciences* 2017;**381**:272-277.

51. Nishiura H, Mizumoto K, Rock KS, Yasuda Y, Kinoshita R, Miyamatsu Y. A theoretical estimate of the risk of microcephaly during pregnancy with Zika virus infection. *Epidemics* 2016;**15**:66-70.

52. Alvarado MG, Schwartz DA. Zika Virus Infection in Pregnancy, Microcephaly, and Maternal and Fetal Health: What We Think, What We Know, and What We Think We Know. *Arch Pathol Lab Med* 2016;**141**(1):26-32.

53. Lee J-S, Mogasale V, Lim JK, Carabali M, Lee K-S, Sirivichayakul C, Dang DA, Palencia-Florez DC, Nguyen THA, Riewpaiboon A, Chanthavanich P, Villar L, Maskery BA, Farlow A. A multi-country study of the economic burden of dengue fever: Vietnam, Thailand, and Colombia. *PLOS Neglected Tropical Diseases* 2017;**11**(10):e0006037.

54. Lee J-S, Mogasale V, Lim JK, Ly S, Lee KS, Sorn S, Andia E, Carabali M, Namkung S, Lim S-K, Ridde V, Njenga SM, Yaro S, Yoon I-K. A multi-country study of the economic burden of dengue fever based on patient-specific field surveys in Burkina Faso, Kenya, and Cambodia. *PLoS neglected tropical diseases* 2019;**13**(2):e0007164-e0007164.

55. Constenla D, Armien B, Arredondo J, Carabali M, Carrasquilla G, Castro R, Durand L, Durán-Arenas L, García ME, Gallegos RV, Gontes ML, López JG, McFarlane C, Montoya R, Sartori AM, Siqueira JB, Martelli CT. Costing Dengue Fever Cases and Outbreaks: Recommendations from a Costing Dengue Working Group in the Americas. *Value in Health Regional Issues* 2015;**8**:80-91.

56. Constenla D, Clark S. Financing dengue vaccine introduction in the Americas: challenges and opportunities. *Expert Rev Vaccines* 2016;**15**(4):547-59.

57. Cardona-Ospina JA, Villamil-Gómez WE, Jimenez-Canizales CE, Castañeda-Hernández DM, Rodríguez-Morales AJ. Estimating the burden of disease and the economic cost attributable to chikungunya, Colombia, 2014. *Transactions of The Royal Society of Tropical Medicine and Hygiene* 2015;**109**(12):793-802.

58. Alvis-Zakzuk NJ, Díaz-Jiménez D, Castillo-Rodríguez L, Castañeda-Orjuela C, Paternina-Caicedo Á, Pinzón-Redondo H, Carrasquilla-Sotomayor M, Alvis-Guzmán N, De La Hoz-Restrepo

F. Economic Costs of Chikungunya Virus in Colombia. *Value in Health Regional Issues* 2018;17:32-37.

59. Shewale JB, Ganduglia Cazaban CM, Waller DK, Mitchell LE, Langlois PH, Agopian AJ. Microcephaly inpatient hospitalization and potential Zika outbreak in Texas: A cost and predicted economic burden analysis. *Travel Medicine and Infectious Disease* 2019;30:67-72.

60. Colón-González FJ, Peres CA, Steiner São Bernardo C, Hunter PR, Lake IR. After the epidemic: Zika virus projections for Latin America and the Caribbean. *PLOS Neglected Tropical Diseases* 2017;11(11):e0006007.

61. Lee BY, Alfaro-Murillo JA, Parpia AS, Asti L, Wedlock PT, Hotez PJ, Galvani AP. The potential economic burden of Zika in the continental United States. *PLOS Neglected Tropical Diseases* 2017;11(4):e0005531.

62. Schwartz LM, Halloran ME, Durbin AP, Longini Jr IM. The dengue vaccine pipeline: Implications for the future of dengue control. *Vaccine* 2015;33(29):3293-3298.

63. Sun J, Du S, Zheng Z, Cheng G, Jin X. Defeat Dengue and Zika Viruses With a One-Two Punch of Vaccine and Vector Blockade. *Frontiers in Microbiology* 2020;11(362).

64. Vannice KS, Roehrig JT, Hombach J. Next generation dengue vaccines: A review of the preclinical development pipeline. *Vaccine* 2015;33(50):7091-7099.

65. Lim JK, Lee Y-S, Wilder-Smith A, Thiry G, Mahoney R, Yoon I-K. Points for Consideration for dengue vaccine introduction – recommendations by the Dengue Vaccine Initiative. *Expert Review of Vaccines* 2016;15(4):529-538.

66. World Health Organization (WHO). Dengue vaccine: WHO position paper – July 2016. *Weekly epidemiological record*. WHO ed. Vol. 91. Geneva: World Health Organization, 2016;349-364.

67. Villabona-Arenas CJ, Ocazionez Jimenez RE, Jimenez Silva CL. Dengue Vaccine: Considerations before Rollout in Colombia. *PLoS Negl Trop Dis* 2016;10(6):e0004653.

68. Lourenço J, Recker M. Dengue serotype immune-interactions and their consequences for vaccine impact predictions. *Epidemics* 2016;16.

69. Smalley C, Erasmus JH, Chesson CB, Beasley DWC. Status of research and development of vaccines for chikungunya. *Vaccine* 2016;**34**(26):2976-2981.

70. Liang G, Gao X, Gould EA. Factors responsible for the emergence of arboviruses; strategies, challenges and limitations for their control. *Emerging Microbes & Infections* 2015;**4**(1):e18 (5).

71. Achee NL, Gould F, Perkins TA, Reiner RC, Jr., Morrison AC, Ritchie SA, Gubler DJ, Teyssou R, Scott TW. A Critical Assessment of Vector Control for Dengue Prevention. *PLoS Negl Trop Dis* 2015;**9**(5):e0003655.

72. Horstick O, Runge-Ranzinger S, Nathan MB, Kroeger A. Dengue vector-control services: how do they work? A systematic literature review and country case studies. *Transactions of The Royal Society of Tropical Medicine and Hygiene* 2010;**104**(6):379-386.

73. Baly A, Toledo ME, Lambert I, Benítez E, Rodriguez K, Rodriguez E, Vanlerberghe V, Stuyft PVd. Cost of intensive routine control and incremental cost of insecticide-treated curtain deployment in a setting with low Aedes aegypti infestation. *Revista da Sociedade Brasileira de Medicina Tropical* 2016;**49**:418-424.

74. Vanlerberghe V, Gómez-Dantés H, Vazquez-Prokopec G, Alexander N, Manrique-Saide P, Coelho G, Toledo ME, Ocampo CB, Stuyft PVd. Changing paradigms in Aedes control: considering the spatial heterogeneity of dengue transmission. *Rev Panam Salud Publica* 2017;**41**(e16):1-6.

75. Toledo ME, Vanlerberghe V, Rosales JP, Mirabal M, Cabrera P, Fonseca V, Gómez Padrón T, Pérez Menzies M, Montada D, Van der Stuyft P. The additional benefit of residual spraying and insecticide-treated curtains for dengue control over current best practice in Cuba: Evaluation of disease incidence in a cluster randomized trial in a low burden setting with intensive routine control. *PLOS Neglected Tropical Diseases* 2017;**11**(11):e0006031.

76. Aliota MT, Peinado SA, Velez ID, Osorio JE. The wMel strain of Wolbachia Reduces Transmission of Zika virus by Aedes aegypti. 2016;**6**:28792.

77. Suh E, Mercer DR, Dobson SL. Life-shortening Wolbachia infection reduces population growth of Aedes aegypti. *Acta Tropica* 2017;**172**:232-239.

78. Bowman LR, Donegan S, McCall PJ. Is Dengue Vector Control Deficient in Effectiveness or Evidence?: Systematic Review and Meta-analysis. *PLOS Neglected Tropical Diseases* 2016;**10**(3):e0004551.

79. Esu E, Lenhart A, Smith L, Horstick O. Effectiveness of peridomestic space spraying with insecticide on dengue transmission; systematic review. *Tropical Medicine & International Health* 2010;**15**(5):619-631.

80. Murray JV, Jansen CC, De Barro P. Risk Associated with the Release of Wolbachia-Infected Aedes aegypti Mosquitoes into the Environment in an Effort to Control Dengue. *Frontiers in Public Health* 2016;**4**:43.

81. Instituto Nacional de Salud (INS). Informe Epidemiologico de Evento Dengue. In: Vigilancia y Análisis del Riesgo en Salud Pública., ed. Dengue. Vol. FOR-R02.4000-001. Bogota, Colombia: Instituto Nacional de Salud, 2016.

82. Rico-Mendoza A, Alexandra P-R, Chang A, Encinales L, Lynch R. Co-circulation of dengue, chikungunya, and Zika viruses in Colombia from 2008 to 2018. *Revista panamericana de salud publica = Pan American journal of public health* 2019;**43**:e49-e49.

83. Kikuti M, Cunha GM, Paploski IAD, Kasper AM, Silva MMO, Tavares AS, Cruz JS, Queiroz TL, Rodrigues MS, Santana PM, Lima HCAV, Calcagno J, Takahashi D, Gonçalves AHO, Araújo JMG, Gauthier K, Diuk-Wasser MA, Kitron U, Ko AI, Reis MG, Ribeiro GS. Spatial Distribution of Dengue in a Brazilian Urban Slum Setting: Role of Socioeconomic Gradient in Disease Risk. *PLOS Neglected Tropical Diseases* 2015;**9**(7):e0003937.

84. Cavalcanti LPdG, Freitas ARR, Brasil P, Cunha RVd. Surveillance of deaths caused by arboviruses in Brazil: from dengue to chikungunya. *Memórias do Instituto Oswaldo Cruz* 2017;**112**:583-585.

85. Freitas ARR, Cavalcanti L, Von Zuben AP, Donalisio MR. Excess Mortality Related to Chikungunya Epidemics in the Context of Co-circulation of Other Arboviruses in Brazil. *PLoS currents* 2017;**9**:ecurrents.outbreaks.14608e586cd321d8d5088652d7a0d884.

86. MacCormack-Gelles B, Lima Neto AS, Sousa GS, Nascimento OJ, Machado MMT, Wilson ME, Castro MC. Epidemiological characteristics and determinants of dengue transmission during epidemic and non-epidemic years in Fortaleza, Brazil: 2011-2015. *PLOS Neglected Tropical Diseases* 2018;**12**(12):e0006990.

87. Rodrigues NCP, Daumas RP, de Almeida AS, dos Santos RS, Koster I, Rodrigues PP, Gomes MdF, Macedo AdF, Gerardi A, Leite IdC. Risk factors for arbovirus infections in a low-income community of Rio de Janeiro, Brazil, 2015-2016. *PLOS ONE* 2018;**13**(6):e0198357.

88. Simião AR, Barreto FKdA, Oliveira RdMAB, Cavalcante JW, Lima Neto AS, Barbosa RB, Lins CdS, Meira AG, Araújo FMdC, Lemos DRQ, Alencar CH, Cavalcanti LPdG. A major chikungunya epidemic with high mortality in northeastern Brazil. *Revista da Sociedade Brasileira de Medicina Tropical* 2019;**52**.

89. Bavia L, Melanda FN, de Arruda TB, Mosimann ALP, Silveira GF, Aoki MN, Kuczera D, Sarzi ML, Junior WLC, Conchon-Costa I, Pavanelli WR, Duarte dos Santos CN, Barreto RC, Bordignon J. Epidemiological study on dengue in southern Brazil under the perspective of climate and poverty. *Scientific Reports* 2020;**10**(1):2127.

90. Carvalho MS, Freitas LP, Cruz OG, Brasil P, Bastos LS. Association of past dengue fever epidemics with the risk of Zika microcephaly at the population level in Brazil. *Scientific Reports* 2020;**10**(1):1752.

91. Roth A, Mercier A, Lepers C, Hoy D, Duituturaga S, Benyon E, Guillaumot L, Souares Y. Concurrent outbreaks of dengue, chikungunya and Zika virus infections-an unprecedented epidemic wave of mosquito-borne viruses in the Pacific 2012-2014. *Euro Surveill* 2014;**19**(41):20929.

92. Farinelli EC, Baquero OS, Stephan C, Chiaravalloti-Neto F. Low socioeconomic condition and the risk of dengue fever: A direct relationship. *Acta Tropica* 2018;**180**:47-57.

93. Mulligan K, Dixon J, Joanna Sinn C-L, Elliott SJ. Is dengue a disease of poverty? A systematic review. *Pathogens and Global Health* 2015;**109**(1):10-18.

94. Sierra BDLC, García G, Pérez AB, Morier L, Alvarez M, Kourí G, Guzmán MG. Ethnicity and Difference in Dengue Virus-Specific Memory T Cell Responses in Cuban Individuals. *Viral Immunology* 2006;**19**(4):662-668.

95. Restrepo BN, Ramirez RE, Arboleda M, Alvarez G, Ospina M, Diaz FJ. Serum Levels of Cytokines in Two Ethnic Groups with Dengue Virus Infection. *The American Journal of Tropical Medicine and Hygiene* 2008;**79**(5):673-677.

96. Magalhães da Silva T, Sandhya Rani MR, de Oliveira Costa GN, Figueiredo MA, Melo PS, Nascimento JF, Molyneaux ND, Barreto ML, Reis MG, Teixeira MG, Blanton RE. The correlation between ancestry and color in two cities of Northeast Brazil with contrasting ethnic compositions. *European Journal of Human Genetics* 2015;**23**(7):984-989.

97. Honorato T, Lapa PPDA, Sales CMM, Reis-Santos B, Tristão-Sá R, Bertolde AI, Maciel ELN. Spatial analysis of distribution of dengue cases in Espírito Santo, Brazil, in 2010: use of Bayesian model. *Revista Bras Epidemiol* 2014;**17**.

98. Ministério da Saúde, Secretaria de Atenção à Saúde. Protocolo de atenção à saúde e resposta à ocorrência de microcefalia relacionada à infecção pelo vírus zika *Atenção à Saúde* Vol. CDU 616-022. Brasilia, Brazil: Ministério da Saúde, Secretaria de Atenção à Saúde, 2016;42.

99. Instituto Nacional de Salud (INS). Informe Epidemiologico de Evento Dengue. In: Vigilancia y Análisis del Riesgo en Salud Pública., ed. Dengue. Vol. FOR-R02.4000-001. Bogota, Colombia: Instituto Nacional de Salud, 2013;30.

100. Hagenlocher M, Delmelle E, Casas I, Kienberger S. Assessing socioeconomic vulnerability to dengue fever in Cali, Colombia: statistical vs expert-based modeling. *Int J Health Geogr* 2013;**12**.

101. Blangiardo M, Cameletti M. Spatial modeling. *Spatial and Spatio-temporal Bayesian Models with R-INLA* John Wiley & Sons, Ltd, 2015;173-234.

102. Diggle PJ, Moraga P, Rowlingson B, Taylor BM. Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm. *Statistical Science* 2013;**28**(4):542-563.

103. Louis VR, Phalkey R, Horstick O, Ratanawong P, Wilder-Smith A, Tozan Y, Dambach P. Modeling tools for dengue risk mapping - a systematic review. *International Journal of Health Geographics* 2014;**13**(1):1-15.

104. Delmelle E, Hagenlocher M, Kienberger S, Casas I. A spatial model of socioeconomic and environmental determinants of dengue fever in Cali, Colombia. *Acta Tropica* 2016;**164**(Supplement C):169-176.

105. Khormi HM, Kumar L. The importance of appropriate temporal and spatial scales for dengue fever control and management. *Sci Total Environ* 2012;**430**.

106. Khormi HM, Kumar L, Elzahrany RA. Modeling spatio-temporal risk changes in the incidence of dengue fever in Saudi Arabia: a geographical information system case study. *Geospat Health* 2011;**6**.

107. Congdon P. Models for spatial outcomes and geographical association. *Applied Bayesian Modelling* John Wiley & Sons, Ltd, 2014;312-363.

108. Guagliardo SAJ, Ardila Roldan SC, Santacoloma L, Luna C, Cordovez Alvarez JM, Rojas Gacha JD, Mansur M, Levine RS, Lenhart A, Oviedo PF. Enhanced vector surveillance to control arbovirus epidemics in Colombia. *Revista panamericana de salud publica = Pan American journal of public health* 2019;**43**:e50-e50.

109. Harris ML, Carter ED. Muddying the waters: A political ecology of mosquito-borne disease in coastal Ecuador. *Health & Place* 2019;**57**:330-338.

110. Marmot SM. Closing the health gap in a generation: the work of the Commission on Social Determinants of Health and its recommendations. *Global Health Promotion* 2009;**16**(1 suppl):23-27.

111. Halstead SB, Streit TG, Lafontant JG, Putvatana R, Russell K, Sun W, Kanesa-Thasan N, Hayes CG, Watts DM. Haiti: absence of dengue hemorrhagic fever despite hyperendemic dengue virus transmission. *The American Journal of Tropical Medicine and Hygiene* 2001;**65**(3):180-3.

112. Cobelens FGJ, Groen J, Osterhaus ADME, Leentvaar-Kuipers A, Wertheim-van Dillen PME, Kager PA. Incidence and risk factors of probable dengue virus infection among Dutch travellers to Asia. *Tropical Medicine & International Health* 2002;**7**(4):331-338.

113. Comision of Social Determinants of Health-CSDH. Closing the gap in a generation : health equity through action on the social determinants of health. Final report of the commission on social determinants of health. Geneva, Switzerland: World Health Organization., 2008;256.

114. Imai N, Dorigatti I, Cauchemez S, Ferguson NM. Estimating Dengue Transmission Intensity from Sero-Prevalence Surveys in Multiple Countries. *PLoS Negl Trop Dis* 2015;**9**(4):e0003719.

115. Chow A, Ho H, Win M-K, Leo Y-S. Assessing Sensitivity and Specificity of Surveillance Case Definitions for Zika Virus Disease. *Emerging infectious diseases* 2017;**23**(4):677-679.

116. Scarpino SV, Meyers LA, Johansson MA. Design Strategies for Efficient Arbovirus Surveillance. *Emerging infectious diseases* 2017;**23**(4):642-644.

117. Catenacci LS, Ferreira M, Martins LC, De Vleeschouwer KM, Cassano CR, Oliveira LC, Canale G, Deem SL, Tello JS, Parker P, Vasconcelos PFC, Travassos da Rosa ES. Surveillance of Arboviruses in Primates and Sloths in the Atlantic Forest, Bahia, Brazil. *EcoHealth* 2018;**15**(4):777-791.

118. Grubaugh ND, Saraf S, Gangavarapu K, Watts A, Tan AL, Oidtman RJ, Ladner JT, Oliveira G, Matteson NL, Kraemer MUG, Vogels CBF, Hentoff A, Bhatia D, Stanek D, Scott B, Landis V, Stryker I, Cone MR, Kopp EW, Cannons AC, Heberlein-Larson L, White S, Gillis LD, Ricciardi MJ, Kwal J, Lichtenberger PK, Magnani DM, Watkins DI, Palacios G, Hamer DH, Gardner LM, Perkins TA, Baele G, Khan K, Morrison A, Isern S, Michael SF, Andersen KG. Travel Surveillance and Genomics Uncover a Hidden Zika Outbreak during the Waning Epidemic. *Cell* 2019;**178**(5):1057-1071.e11.

119. Runge-Ranzinger S, Horstick O, Marx M, Kroeger A. What does dengue disease surveillance contribute to predicting and detecting outbreaks and describing trends? *Tropical Medicine & International Health* 2008;**13**(8):1022-1041.

120. Mercado-Reyes M, Acosta-Reyes J, Navarro-Lechuga E, Corchuelo S, Rico A, Parra E, Tolosa N, Pardo L, González M, Martìn-Rodriguez-Hernández J, Karime-Osorio L, Ospina-Martinez M, Rodriguez-Perea H, Del Rio-Pertuz G, Viasus D. Dengue, chikungunya and zika virus coinfection: results of the national surveillance during the zika epidemic in Colombia. *Epidemiology and Infection* 2019;**147**:e77.

121. Camacho T, de la Hoz F, Cárdenas V, Sánchez C, de Calderón L, Pérez L, Bermúdez A. Incomplete surveillance of a Dengue-2 epidemic in Ibagué, Colombia, 1995-1997. *Biomédica* 2004;**24**:174-182.

122. Vong S, Goyet S, Ly S, Ngan C, Huy R, Duong V, Wichmann O, Letson GW, Margolis HS, Buchy P. Under-recognition and reporting of dengue in Cambodia: a capture–recapture analysis of the National Dengue Surveillance System. *Epidemiol Infect.* 2011;**140**(3):491-499.

123. Romero-Vega L. Evaluation of dengue fever reports during an epidemic, Colombia. . *Revista de Saúde Pública [online].* 2014;**48**(6):6.

124. Toan NT, Rossi S, Prisco G, Nante N, Viviani S. Dengue epidemiology in selected endemic countries: factors influencing expansion factors as estimates of underreporting. *Trop Med Int Health.* 2015;**20**(7):840-863.

125. Brasil, Ministerio da Saude. Portaria No 47, 3 de Maio de 2016. Define os parâmetros para monitoramento da regularidade na alimentação do Sistema de Informação de Agravos de Notificação (SINAN). In: Ministerio da Saude SdVeS, ed. Vol. Portaria No. 47, 2016. Brasilia, 2016.

126. Instituto Nacional de Salud (INS). Protocolo de Vigilancia en Salud Publica Enfermedad por Virus Zika. In: Instituto Nacional de Salud (INS), ed. Vol. FOR-R02.0000-059. Bogota, Colombia: VIGILANCIA Y ANÁLISIS DEL RIESGO EN SALUD PÚBLICA, 2017;22.

127. Instituto Nacional de Salud (INS). Protocolo de Vigilancia en Salud Publica Enfermedad por Virus Chikungunya. In: Instituto Nacional de Salud (INS), ed. Vol. . Bogota, Colombia: VIGILANCIA Y ANÁLISIS DEL RIESGO EN SALUD PÚBLICA, 2017.

128. Ministerio de la Protección Social. Decreto Numero 3518 de 2006. Creacion y reglamentacion del Sistema de Vigilancia en Salud Publica- SIVIGILA. In: Social. MdlP, ed. Vol. 3518/2006. Bogota, Colombia., 2006;1-17.

129. Brasil Ministerio da Saude. Sistema de Informação de Agravos de Notificação – Sinan: normas e rotinas. In: Ministério da Saúde SdVeS, Departamento de Vigilância Epidemiológica., ed. Normas e Manuais Tecnicos. Brasilia: Editora do Ministério da Saúde,, 2007;68.

130. Instituto Nacional de Salud (INS). Protocolo de Vigilancia en Salud Pública, Dengue (Surveillance Protocol in Public Health, Dengue). . In: Instituto Nacional de Salud (INS) C, ed. *Surveillance*. Dirección de Vigilancia y Análisis de Riesgo en Salud Pública ed. Vol. FOR-R02.0000-059 V02. Santafé de Bogota, Colombia. : Instituto Nacional de Salud (National Institute of Health). , 2014;19.

131. Hunsperger EA, Yoksan S, Buchy P, Nguyen VC, Sekaran SD, Enria DA, Vazquez S, Cartozian E, Pelegrino JL, Artsob H, Guzman MG, Olliaro P, Zwang J, Guillerm M, Kliks S, Halstead S, Peeling RW, Margolis HS. Evaluation of Commercially Available Diagnostic Tests for the Detection of Dengue Virus NS1 Antigen and Anti-Dengue Virus IgM Antibody. *PLoS Negl Trop Dis* 2014;**8**(10):e3171.

132. Muller DA, Depelsenaire ACI, Young PR. Clinical and Laboratory Diagnosis of Dengue Virus Infection. *J Infect Dis* 2017;**215**(suppl_2):S89-S95.

133. Ohst C, Saschenbrecker S, Stiba K, Steinhagen K, Probst C, Radzimski C, Lattwein E, Komorowski L, Stöcker W, Schlumberger W. Reliable Serological Testing for the Diagnosis of Emerging Infectious Diseases. In: Hilgenfeld R, Vasudevan SG, eds. *Dengue and Zika: Control and Antiviral Treatment Strategies*. Singapore: Springer Singapore, 2018;19-43.

134. Beltrán-Silva SL, Chacón-Hernández SS, Moreno-Palacios E, Pereyra-Molina JÁ. Clinical and differential diagnosis: Dengue, chikungunya and Zika. *Revista Médica del Hospital General de México* 2016.

135. Calvo EP, Coronel-Ruiz C, Velazco S, Velandia-Romero M, Castellanos JE. Diagnóstico diferencial de dengue y chikungunya en pacientes pediátricos. *Biomédica* 2016;**36**:35-43.

136. Braga JU, Bressan C, Dalvi APR, Calvet GA, Daumas RP, Rodrigues N, Wakimoto M, Nogueira RMR, Nielsen-Saines K, Brito C, Bispo de Filippis AM, Brasil P. Accuracy of Zika virus disease case definition during simultaneous Dengue and Chikungunya epidemics. *PLOS ONE* 2017;**12**(6):e0179725.

137. Waggoner JJ, Pinsky BA. Zika Virus: Diagnostics for an Emerging Pandemic Threat. *Journal of Clinical Microbiology* 2016;**54**(4):860-867.

138. Fischer C, Pedroso C, Mendrone A, Jr., Bispo de Filippis AM, Vallinoto ACR, Ribeiro BM, Durigon EL, Marques ETA, Jr., Campos GS, Viana IFT, Levi JE, Scarpelli LC, Nogueira ML, Bastos MdS, Souza NCS, Khouri R, Lira S, Komninakis SV, Baronti C, Charrel RN, Kümmerer BM, Drosten C, Brites C, de Lamballerie X, Niedrig M, Netto EM, Drexler JF. External Quality Assessment for Zika Virus Molecular Diagnostic Testing, Brazil. *Emerging infectious diseases* 2018;**24**(5):888-892.

139. Voermans JJ, Pas SD, van der Linden A, GeurtsvanKessel C, Koopmans M, van der Eijk A, Reusken CB. Whole-Blood Testing for Diagnosis of Acute Zika Virus Infections in Routine Diagnostic Setting. *Emerging infectious diseases* 2019;**25**(7).

140. Instuto Nacional de Salud (INS). Informe del Evento Chikunguña Periodo Epidemiologico XIII, Colombia 2016. Bogota, Colombia: INS, 2016;1-18.

141. Ministerio de Saude Publica do Brasil. Registro de Eventos em Saude Publica. Monitoramento integrado de vigilância e atenção à saúde de condições relacionadas às infecções durante a gestação, identificadas no pré-natal, parto e puericultura. http://www.resp.saude.gov.br/microcefalia#/painel.

142. Instituto Nacional de Salud (INS). Protocolo de vigilancia y control de Dengue INT- PRO-R02.003.0000-004. . Bogotá, Colombia: Instituto Nacional de Salud. Subdirección de Vigilancia y Control en Salud Pública,, 2010.

143. Carabali M, Lim JK, Palencia DC, Lozano-Parra A, Gelvez RM, Lee KS, Florez JP, Herrera VM, Kaufman JS, Rojas EM, Villar LA. Burden of dengue among febrile patients at the time of chikungunya introduction in Piedecuesta, Colombia. *Tropical Medicine & International Health* 2018;**23**(0):1231-1241.

144. Teixeira MdG, Barreto ML, Costa MdCN, Ferreira LDA, Vasconcelos PFC, Cairncross S. Dynamics of dengue virus circulation: a silent epidemic in a complex urban area. *Tropical Medicine & International Health* 2002;**7**(9):757-762.

145. Thai KTD, Binh TQ, Giao PT, Phuong HL, Hung LQ, Nam NV, Nga TT, Groen J, Nagelkerke N, de Vries PJ. Seroprevalence of dengue antibodies, annual incidence and risk factors among children in southern Vietnam. *Tropical Medicine & International Health* 2005;**10**(4):379-386.

146. Wearing HJ, Rohani P. Ecological and immunological determinants of dengue epidemics. *Proceedings of the National Academy of Sciences* 2006;**103**(31):11802-11807.

147. Usme-Ciro JA, Mendez JA, Tenorio A, Rey GJ, Domingo C, Gallego-Gomez JCJVJ. Simultaneous circulation of genotypes I and III of dengue virus 3 in Colombia. 2008;**5**(1):101.

148. Halsey ES, Marks MA, Gotuzzo E, Fiestas V, Suarez L, Vargas J, Aguayo N, Madrid C, Vimos C, Kochel TJ, Laguna-Torres VA. Correlation of Serotype-Specific Dengue Virus Infection with Clinical Manifestations. *PLoS Negl Trop Dis* 2012;**6**(5):e1638.

149. Huy NT, Van Giang T, Thuy DHD, Kikuchi M, Hien TT, Zamora J, Hirayama K. Factors Associated with Dengue Shock Syndrome: A Systematic Review and Meta-Analysis. *PLoS Negl Trop Dis* 2013;**7**(9):e2412.

150. Reiner RC, Stoddard ST, Forshey BM, King AA, Ellis AM, Lloyd AL, Long KC, Rocha C, Vilcarromero S, Astete H, Bazan I, Lenhart A, Vazquez-Prokopec GM, Paz-Soldan VA, McCall PJ, Kitron U, Elder JP, Halsey ES, Morrison AC, Kochel TJ, Scott TW. Time-varying, serotype-specific force of infection of dengue virus. *Proceedings of the National Academy of Sciences* 2014.

151. Soo K-M, Khalid B, Ching S-M, Chee H-Y. Meta-Analysis of Dengue Severity during Infection by Different Dengue Virus Serotypes in Primary and Secondary Infections. *PLOS ONE* 2016;**11**(5):e0154760.

152. Garg S, Chakravarti A, Singh R, Masthi NRR, Goyal RC, Jammy GR, Ganguly E, Sharma N, Singh MM, Ferreira G, Moureau A, Ojha S, Nealon J. Dengue serotype-specific seroprevalence among 5- to 10-year-old children in India: a community-based cross-sectional study. *International Journal of Infectious Diseases* 2017;**54**:25-30.

153. Carreño MF, Jiménez-Silva CL, Rey-Caro LA, Conde-Ocazionez SA, Flechas-Alarcón MC, Velandia SA, Ocazionez RE. Dengue in Santander State, Colombia: fluctuations in the prevalence of virus serotypes are linked to dengue incidence and genetic diversity of the circulating viruses. *Tropical Medicine & International Health* 2019;**24**(12):1400-1410.

154. Economopoulou A, Dominguez M, Helynck B, Sissoko D, Wichmann O, Quenel P, Germonneau P, Quatresous I. Atypical Chikungunya virus infections: clinical manifestations, mortality and risk factors for severe disease during the 2005–2006 outbreak on Réunion. *Epidemiol Infect.* 2008;**137**(4):534-541.

155. Tomashek KM, Lorenzi OD, Andújar-Pérez DA, Torres-Velásquez BC, Hunsperger EA, Munoz-Jordan JL, Perez-Padilla J, Rivera A, Gonzalez-Zeno GE, Sharp TM, Galloway RL, Glass Elrod M, Mathis DL, Oberste MS, Nix WA, Henderson E, McQuiston J, Singleton J, Kato C, García Gubern C, Santiago-Rivera W, Cruz-Correa J, Muns-Sosa R, Ortiz-Rivera JD, Jiménez G, Galarza IE, Horiuchi K, Margolis HS, Alvarado LI. Clinical and epidemiologic characteristics of dengue and other etiologic agents among patients with acute febrile illness, Puerto Rico, 2012–2015. *PLoS Negl Trop Dis* 2017;**11**(9):e0005859.

156. Bastos MLA, Abreu FSd, Silva Junior GBd. Inability to work due to Chikungunya virus infection: impact on public service during the first epidemic in the State of Ceará, northeastern Brazil. *Brazilian Journal of Infectious Diseases* 2018;**22**:248-249.

157. Carabali M, Lim JK, Velez DC, Trujillo A, Egurrola J, Lee KS, Kaufman JS, DaSilva LJ, Velez ID, Osorio JE. Dengue virus serological prevalence and seroconversion rates in children and

adults in Medellin, Colombia: implications for vaccine introduction. *International Journal of Infectious Diseases* 2017;58:27-36.

158. Lim JK, Carabali M, Lee J-S, Lee K-S, Namkung S, Lim S-K, Ridde V, Fernandes J, Lell B, Matendechero SH, Esen M, Andia E, Oyembo N, Barro A, Bonnet E, Njenga SM, Agnandji ST, Yaro S, Alexander N, Yoon I-K. Evaluating dengue burden in Africa in passive fever surveillance and seroprevalence studies: protocol of field studies of the Dengue Vaccine Initiative. *BMJ Open* 2018;8(1).

159. Zea D, Osorio L. The status of the dengue surveillance system in a Colombian municipality. *Revista de Salud Pública* 2011;13(5):785-795.

160. Carabalí JM, Hendrickx D. Dengue and health care access: the role of social determinants of health in dengue surveillance in Colombia. *Glob Health Promot* 2012;19.

161. Gibbons CL, Mangen M-JJ, Plass D, Havelaar AH, Brooke RJ, Kramarz P, Peterson KL, Stuurman AL, Cassini A, Fèvre EM, Kretzschmar MEE. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. *BMC Public Health* 2014;14(1):147.

162. Lee J-S, Carabali M, Lim JK, Herrera VM, Park I-Y, Villar L, Farlow A. Early warning signal for dengue outbreaks and identification of high risk areas for dengue fever in Colombia using climate and non-climate datasets. *BMC Infectious Diseases* 2017;17(1):480.

163. Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology* 2013;7(Supplement C):39-55.

164. Blangiardo M, Cameletti M. Bayesian regression and hierarchical models. *Spatial and Spatio-temporal Bayesian Models with R-INLA* John Wiley & Sons, Ltd, 2015;127-172.

165. Sophia R-H, Anders S. *Multilevel and Longitudinal Modeling Using Stata*. Third Edition ed. College Station, Texas: Stata Press, 2012.

166. Riebler A, Sørbye SH, Simpson D, Rue H. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research* 2016;25(4):1145-1165.

167. Banerjee S, Carlin B, Gelfand A. *Hierarchical modeling and analyisis for spatial data*. Boca Raton: Hall/CRC Biostatistics Series, 2015.

168. Congdon P. Hierarchical models for related units. *Applied Bayesian Modelling* John Wiley & Sons, Ltd, 2014;34-96.

169. Blangiardo M, Cameletti M. Bayesian computing. *Spatial and Spatio-temporal Bayesian Models with R-INLA* John Wiley & Sons, Ltd, 2015;75-126.

170. Congdon P. *Applied Bayesian modelling*. West Sussex: Wiley, 2014.

171. Martínez-Bello D, López-Quílez A, Torres-Prieto A. Bayesian dynamic modeling of time series of dengue disease case counts. *PLoS Negl Trop Dis* 2017;**11**.

172. Lee D. A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spat Spatio Temporal Epidemiol* 2011;**2**.

173. Rue H, Riebler A, Sørbye SH, Illian JB, Simpson DP, Lindgren FK. Bayesian Computing with INLA: A Review. *Annual Review of Statistics and Its Application* 2017;**4**(1):395-421.

174. Lindgren F, Rue H. Bayesian spatial modelling with R-INLA. *Journal of Statistical Software* 2015;**63**(19):1-25.

175. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991;**43**(1):1-20.

176. Besag J. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society: Series B (Methodological)* 1974;**36**(2):192-225.

177. Villar LA, Rojas DP, Besada-Lombana S, Sarti E. Epidemiological trends of dengue disease in Colombia (2000–2011): a systematic review. *PLoS Negl Trop Dis* 2015;**9**.

178. Blangiardo M, Cameletti M. Spatio-temporal models. *Spatial and Spatio-temporal Bayesian Models with R-INLA* John Wiley & Sons, Ltd, 2015;235-258.

179. Baddeley A, Rubak E, Turner R. *Spatial point patterns: methodology and applications with R* CRC Press, 2015.

180. Illian JB, Martino S, Sorbye S, Gallego-Fernandez JB, Zunzunegui M, Esquivias MP, Travis JMJ. Fitting complex ecological point process models with Integrated Nested Laplace Approximation. *Methods in Ecology and Evolution* 2013;**4**:305-315.

181. Simpson D, Illian JB, Lindgren F, Sorbye S, Rue H. Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *arXiv* 2017:1-26.

182. Muff S, Riebler A, Held L, Rue H, Saner P. Bayesian analysis of measurement error models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2015;**64**(2):231-252.

183. Küchenhoff H, Mwalili SM, Lesaffre E. A General Method for Dealing with Misclassification in Regression: The Misclassification SIMEX. *Biometrics* 2006;**62**(1):85-96.

184. Lederer W, Küchenhoff H, Lederer MW, by Küchenhoff M. Package 'simex'. 2019.

185. Lederer W, Küchenhoff H. A short Introduction to the SIMEX and MCSIMEX. *The Newsletter of the R Project Volume 6/4, October 2006* 2006;**6**:26.

186. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. *J Health Soc Behav* 2010;**51 Suppl**:S28-40.

187. Rosa Dias P, O'Donnell O. Health and inequality. Research on economic inequality, 1049-2585 ; 21. Bingley, U.K.: Emerald, 2013.

188. World Health Organization (WHO). Health Equity Assessment Toolkit (HEAT): Software for exploring and comparing health inequalities in countries. Built-in database edition. Version 3.1. *Technical Notes*. Geneva, Switzerland: WHO, 2019;22.

189. Konings P, Harper S, Lynch J, Hosseinpoor AR, Berkvens D, Lorant V, Geckova A, Speybroeck N. Analysis of socioeconomic health inequalities using the concentration index. *International Journal of Public Health* 2010;**55**(1):71-74.

190. Harper S, Lynch J. Measuring inequalities in health. In: Oakes J, Kaufman JS, eds. *Methods in social epidemiology*. San Francisco: Jossey-Bass, 2006.

191. Harper SAM, King NB, Meersman SC, Reichman ME, Breen N, Lynch J. Implicit Value Judgments in the Measurement of Health Inequalities. *Milbank Quarterly* 2010;88(1):4-29.

192. Erreygers G. Correcting the Concentration Index. *Journal of Health Economics* 2009;28(2):504-515.

193. Erreygers G, Kessels R. Regression-Based Decompositions of Rank-Dependent Indicators of Socioeconomic Inequality of Health. In: Kessels R, ed. *Health and Inequality*. Research on Economic Inequality. Vol. 21 Emerald Group Publishing Limited, 2013;227-259.

194. Erreygers G, Van Ourti T. Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: A recipe for good practice. *Journal of Health Economics* 2011;30(4):685-694.

195. Heckley G, Gerdtham U-G, Kjellsson G. A general method for decomposing the causes of socioeconomic inequality in health. *Journal of Health Economics* 2016;48:89-106.

196. Wagstaff A. The concentration index of a binary outcome revisited. *Health Economics* 2011;20(10):1155-1160.

197. O'Donnell O, O'Neill S, Van Ourti T, Walsh B. conindex: Estimation of concentration indices. *The Stata journal* 2016;16(1):112-138.

198. IBGE, Instituto Brasileiro de Geografia e Estatística. Atlas do Censo Demografico Brasileiro 2010. . In: SIDRA, ed. Vol. Banco de Tabelas Estadisticas. Rio de Janeiro, Brazil: Ministerio do Planejamento, Orçamento e Gestao., 2010.

199. Departamento Nacional de Estadistica de Colombia (DANE). Estimaciones de población 1985 - 2005 y proyecciones de población 2005 - 2020 total municipal por área. *Departamento Nacional de Estadistica de Colombia* DANE, 2015.

200. Departamento Administrativo Nacional de Estadística (DANE). Estimaciones de población 1985–2005 y proyecciones de población 2005–2020 total municipal por área. . Vol. Vital Statistics. Bogota, Colombia: DANE, 2016.

201. Martínez-Bello DA, López-Quílez A, Torres Prieto A. Relative risk estimation of dengue disease at small spatial scale. *International Journal of Health Geographics* 2017;16(1):31.

202. Vincenti-Gonzalez MF, Grillet M-E, Velasco-Salas ZI, Lizarazo EF, Amarista MA, Sierra GM, Comach G, Tami A. Spatial Analysis of Dengue Seroprevalence and Modeling of Transmission Risk Factors in a Dengue Hyperendemic City of Venezuela. *PLOS Neglected Tropical Diseases* 2017;**11**(1):e0005317.

203. Leta S, Beyene TJ, De Clercq EM, Amenu K, Kraemer MUG, Revie CW. Global risk mapping for major diseases transmitted by Aedes aegypti and Aedes albopictus. *International Journal of Infectious Diseases* 2018;**67**:25-35.

204. Desjardins MR, Casas I, Victoria AM, Carbonell D, Dávalos DM, Delmelle EM. Knowledge, attitudes, and practices regarding dengue, chikungunya, and Zika in Cali, Colombia. *Health & Place* 2020;**63**:102339.

205. Reyes-Castro PA, Harris RB, Brown HE, Christopherson GL, Ernst KC. Spatio-temporal and neighborhood characteristics of two dengue outbreaks in two arid cities of Mexico. *Acta Tropica* 2017;**167**(Supplement C):174-182.

206. Houghton N, Bascolo E, del Riego A. Socioeconomic inequalities in access barriers to seeking health services in four Latin American countries. *Rev Panam Salud Publica* 2020;**44**(e11):9.

207. Mol MPG, Queiroz JTM, Gomes J, Heller L. Gestão adequada de resíduos sólidos como fator de proteção na ocorrência da dengue [Adequate solid waste management as a protection factor against dengue cases]. *Rev Panam Salud Publica* 2020;**44**(e22):9.

208. Chacón-Duque JC, Adhikari K, Avendaño E, Campo O, Ramirez R, Rojas W, Ruiz-Linares A, Restrepo BN, Bedoya G. African genetic ancestry is associated with a protective effect on Dengue severity in colombian populations. *Infection, Genetics and Evolution* 2014;**27**:89-95.

209. Collins PH. Intersectionality's Definitional Dilemmas. *Annual Review of Sociology* 2015;**41**(1):1-20.

210. Jackson JW. Explaining intersectionality through description, counterfactual thinking, and mediation analysis. *Social Psychiatry and Psychiatric Epidemiology* 2017;**52**(7):785-793.

211. Jackson JW, Williams DR, VanderWeele TJ. Disparities at the intersection of marginalized groups. *Social Psychiatry and Psychiatric Epidemiology* 2016;**51**(10):1349-1359.

212. Agudelo-Suárez AA, Martínez-Herrera E, Posada-López A, Rocha-Buelvas A. Ethnicity and Health in Colombia: What Do Self-Perceived Health Indicators Tell Us? *Ethnicity & Disease* 2016;**26**(2):147-156.

213. Bonham VL, Green ED, Pérez-Stable EJ. Examining how race, ethnicity, and ancestry data are used in biomedical research. *JAMA* 2018.

214. Palacio Chaverra A. The color of child survival in Colombia, 1955–2005. *Ethnicity & Health* 2018;**23**(2):207-220.

215. Runge-Ranzinger S, McCall PJ, Kroeger A, Horstick O. Dengue disease surveillance: an updated systematic literature review. *Trop Med Int Health* 2014;**19**.

216. Waggoner JJ, Gresh L, Mohamed-Hadley A, Balmaseda A, Soda KJ, Abeynayake J, Sahoo MK, Liu Y, Kuan G, Harris E, Pinsky BA. Characterization of Dengue Virus Infections Among Febrile Children Clinically Diagnosed With a Non-Dengue Illness, Managua, Nicaragua. *J Infect Dis* 2017;**215**(12):1816-1823.

217. World Health Organization (WHO). Global Surveillance for human infection with coronavirus disease (COVID-19). *Interim Guidance*. CC BY-NC-SA 3.0 IGO ed. Geneva, Switzerland: World Health Organization, 2020.

218. Jackson JW, VanderWeele TJ. Intersectional decomposition analysis with differential exposure, effects, and construct. *Social Science & Medicine* 2019;**226**:254-259.

219. Jackson JW, VanderWeele TJ. Decomposition Analysis to Identify Intervention Targets for Reducing Disparities. *Epidemiology* 2018;**29**(6):825-835.

220. Sinning M, Hahn M, Bauer TK. The Blinder-Oaxaca decomposition for nonlinear regression models. *The Stata Journal* 2008;**8**(4):480-492.

221. Elder TE, Goddeeris JH, Haider SJ. Unexplained gaps and Oaxaca–Blinder decompositions. *Labour Economics* 2010;**17**(1):284-290.

222. Ahmed F, Ahmed Ne, Pissarides C, Stiglitz J. Why inequality could spread COVID-19. *The Lancet Public Health* 2020;**5**(5):e240.

223. Chowkwanyun M, Reed AL. Racial Health Disparities and Covid-19 — Caution and Context. *New England Journal of Medicine* 2020.

224. Dorn Av, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. *The Lancet* 2020;**395**(10232):1243-1244.

225. Raifman MA, Raifman JR. Disparities in the Population at Risk of Severe Illness From COVID-19 by Race/Ethnicity and Income. *American journal of preventive medicine* 2020:S0749-3797(20)30155-0.