Epidemiology of the 2014-2017 Zika and chikungunya epidemics in Colombia

Kelly Anne Charniga

Department of Infectious Disease Epidemiology
School of Public Health
Imperial College London

Thesis submitted for the degree of Doctor of Philosophy

August 2021

Declaration of originality

I declare that this thesis is my own work. Plans of analyses were formulated with support from my supervisors, Professor Christl Donnelly, Dr. Pierre Nouvellet, and Dr. Zulma Cucunubá. Marcela Mercado, Dr. Diana Walteros, Dr. Franklyn Prieto, and Dr. Martha Ospina were involved in sharing data from Colombia's Instituto Nacional de Salud and editing two manuscripts related to the thesis. I analyzed the data and discussed the results with my supervisors.

Copyright declaration

Abstract

The introduction of Zika virus (ZIKV) and chikungunya virus (CHIKV) into the Americas in the mid-2010s triggered large epidemics across the region fueled by high population-level susceptibility. Colombia was one of the most affected countries, reporting approximately 413,000 cases of chikungunya fever (CF) and 106,000 cases of Zika virus disease (ZVD) in only three years. In this thesis, mathematical models were used to estimate key epidemiological parameters from cases of CF and ZVD that were reported to Colombia's national population-based surveillance system on a weekly basis from 2014-2017.

Both epidemics spread widely throughout the country. Out of 32 departments, all reported ZVD cases, while 31 departments reported CF cases. Females of child-bearing age comprised a large proportion of reported ZVD cases, most likely due to increased reporting linked to the risk of congenital birth defects associated with ZIKV infection during pregnancy. Of the 418 reported cases of ZIKV-associated neurological complications in the country, most were diagnosed with Guillain-Barré syndrome.

The estimated reporting rate of CHIKV was higher than that of ZIKV according to models based on the renewal equation. This result was expected due to the higher rate of asymptomatic ZIKV infection compared to that of CHIKV. Basic reproduction number estimates at the department level were similar for CHIKV compared to ZIKV. Reassuringly, estimates of the time-varying reproduction number from the parametric model were in good agreement with those obtained from the software EpiEstim.

ZIKV infection attack rates, reporting rates of ZVD, and the risk of ZIKV-associated neurological complications were estimated for 28 Colombian capital cities by incorporating multiple data types into a Bayesian hierarchical model. ZIKV infection attack rates varied considerably across cities. The overall estimated reporting rate for ZVD was similar to that estimated previously. Results also showed a low estimated risk of ZIKV-associated neurological complications. Important differences in estimated ZVD reporting rates and the risk of ZIKV-associated neurological complications between sex and age group were found for some cities.

Finally, gravity models, Stouffer's rank models, and radiation models were used to investigate the spatial and temporal invasion dynamics of CHIKV and ZIKV. Both geographic distance and travel time between cities were evaluated. Invasion risk was best captured by a gravity model which accounted for geographic distance and intermediate levels of density dependence. Results also showed that Stouffer's rank model with geographic distance performed well. Short-distance transmission played an important role in spatial spread, and a few long-distance transmission events were identified. Jointly fitted models highlighted similarities between the epidemics. However, ZIKV spread faster than CHIKV.

With new interventions on the horizon, including vaccines and novel methods of vector control, as well as the emergence of new arboviral diseases, having robust estimates of epidemiological parameters will be important for informing surveillance and preparedness for future epidemics.

Resumen

La aparición del virus de Zika y el virus de chikungunya en América a mediados de la década de 2010 desencadenó grandes epidemias a lo largo de la región alimentado por altos niveles de susceptibilidad en la población. Colombia fue uno de los países más afectados, reportando aproximadamente 413,000 casos de fiebre de chikungunya y 106,000 casos de enfermedad de Zika en solo tres años. En esta tesis, modelos matemáticos fueron usados para estimar los parámetros epidemiológicos claves en casos de fiebre de chikungunya y enfermedad de Zika que fueron reportados al sistema nacional de vigilancia poblacional de Colombia semanalmente desde 2014-2017.

Ambas epidemias se propagaron ampliamente a través del país. De los 32 departamentos, todos reportaron los casos de enfermedad de Zika, mientras 31 departamentos reportaron los casos de fiebre de chikungunya. Las mujeres en edad reproductiva constituyeron una proporción grande de los casos reportados de enfermedad de Zika, probablemente debido al reportaje elevado relacionado con el riesgo de los defectos congénitos asociados con la infección de Zika durante el periodo de embarazo. De los 418 casos reportados con complicaciones neurológicas asociadas al virus de Zika en el país, la mayoría fue diagnosticado con el síndrome de Guillain-Barré.

La tasa de reporte estimado del virus de chikungunya fue más alta que la de Zika según unos modelos basados en la ecuación de renovación. Este resultado fue esperado debido a una mayor tasa de infección asintomática del virus de Zika en comparación con la de chikungunya. Las estimaciones del número de reproducción básico al nivel del departamento para el virus de chikungunya fueron parecidos comparado con las de Zika. Tranquilizadoramente, las estimaciones del número de reproducción variable en el tiempo obtenidas del modelo paramétrico estuvieron acorde con las obtenidas del software EpiEstim.

Las tasas de ataque de infección del virus de Zika, las tasas de reporte de enfermedad de Zika y el riesgo de complicaciones neurológicas asociadas con el virus de Zika fueron estimadas para 28 de las ciudades capitales de Colombia incorporando varios tipos de datos en un modelo Bayesiano jerárquico. Las tasas de ataque de infección del virus de Zika variaron considerablemente a través de las ciudades. La tasa de reporte total por

enfermedad de Zika fue parecida a la estimación previa. Los resultados también demostraron un riesgo bajo de las complicaciones neurológicas asociadas con el virus de Zika. Diferencias importantes en las tasas estimadas de reporte de enfermedad de Zika y el riesgo de las complicaciones neurológicas asociadas con el virus de Zika entre el sexo y grupo de edad fueron descubiertos en algunas ciudades.

Finalmente, modelos de gravitación, modelos de oportunidades interpuestas de Stouffer y modelos de radiación fueron usados para investigar las dinámicas de invasión a través del espacio y tiempo de los virus de chikungunya y Zika. Tanto la distancia geográfica como el tiempo para viajar entre las ciudades fueron evaluados. El riesgo de invasión fue capturado de mejor manera por un modelo de gravitación que incluyó la distancia geográfica y los niveles intermedios de la dependencia de densidad. Los resultados además mostraron que el modelo de oportunidades interpuestas con la distancia geográfica funcionó bien. La transmisión de corta distancia jugó un rol importante en la propagación espacial y algunos eventos de transmisión de larga distancia fueron identificados. Los modelos ajustados conjuntamente destacaron las similitudes entre las epidemias. Sin embargo, el virus de Zika se propagó más rápido que el virus de chikungunya.

Con nuevas intervenciones en el horizonte, incluyendo vacunas y nuevos métodos para el control de vectores, además de la aparición de nuevas enfermedades por arbovirus, tener estimaciones robustas de los parámetros epidemiológicos será importante para mejorar la vigilancia y preparación frente a epidemias futuras.

Acknowledgements

I would like to thank my supervisors, Professor Christl Donnelly, Dr. Pierre Nouvellet and Dr. Zulma Cucunubá, for their patience and support throughout my PhD, especially in 2020-2021 while they were responding to the COVID-19 pandemic. I am grateful to have had the opportunity to learn from such brilliant and caring scientists for over three and a half years.

For technical help during my PhD, I thank Dr. Pierre Nouvellet and Dr. Zulma Cucunubá. I would also like to thank Dr. Wes Hinsley and Dr. Rich Fitzjohn for their help with R and the cluster.

Thanks to Team Donnelly, the arbovirus interest group, and my officemates in VA5 for their invaluable peer-support and friendship.

This work was made possible by a collaboration with Colombia's Instituto Nacional de Salud. In particular, I would like to thank Marcela Mercado and Dr. Diana Walteros for sharing data and helping me understand the context of my research.

Thanks also to the healthcare professionals across Colombia involved in reporting cases of chikungunya fever and Zika virus disease to the national public health surveillance system.

I also thank Imperial College London President's PhD Scholarship scheme for generously covering the costs of tuition and living expenses in London as well as travel to Colombia and international conferences. Thanks to the UK Medical Research Council and the UK Foreign, Commonwealth & Development Office for funding the MRC Centre for Global Infectious Disease Analysis.

Thanks to Elmer Enrique. Te amo.

Finally, thanks to Cathy, Jackie, and my parents for supporting me in all of my academic and athletic endeavors over the years.

Relevant publications

1. Charniga K, Cucunubá ZM, Walteros DM, Mercado M, Prieto F, Ospina M, et al. Descriptive analysis of surveillance data for Zika virus disease and Zika virus-associated neurological complications in Colombia, 2015-2017. PLoS ONE. 16(6):e0252236.
2. Charniga K, Cucunubá ZM, Mercado M, Prieto F, Ospina M, Nouvellet P, et al. Spatial and temporal invasion dynamics of the 2014-2017 Zika and chikungunya epidemics in Colombia. PLoS Comp Biol. 2021; 17(7):e1009174.

Presentations

1. Oral presentation: 'Spatial and temporal invasion dynamics of the 2014-2017 Zika and chikungunya epidemics in Colombia.' (2020) Brazil-UK Centre for Arbovirus Discovery, Diagnostics, Genomics and Epidemiology meeting. Virtual.
2. Poster presentation: 'Spatial and temporal spread of Zika and chikungunya viruses in Colombia, a gravity-model based approach.' (2019) Epidemics[7]. Charleston, South Carolina USA.
3. Oral presentation: 'La propagación de los virus de Zika y chikungunya en Colombia.' (2019) Outbreak Analysis and Modeling for Public Health Course. Bogotá, Colombia.
4. Oral presentation: 'Spatial and temporal spread of Zika and chikungunya viruses in Colombia.' (2019) Imperial College President's PhD Scholars' Symposium. London, UK.
5. Poster presentation: 'Spatial and temporal spread of Zika and chikungunya viruses in Colombia, a gravity-model based approach.' (2018) Annual Meeting of the American Society of Tropical Medicine and Hygiene. New Orleans, Louisiana USA.
6. Oral presentation: 'Spatial and temporal spread of Zika and chikungunya viruses in Colombia, a gravity-model based approach.' (2018) Colombian Research in the UK 5th Annual Summit. Cambridge, UK.
7. Poster presentation: 'Spatial and temporal spread of Zika and chikungunya viruses in Colombia, a gravity-model based approach.' (2018) Imperial College Graduate School PhD Summer Showcase. London, UK.

Table of contents

Declaration of originality ... 2
Copyright declaration ... 3
Abstract ... 4
Resumen ... 6
Acknowledgements ... 8
Relevant publications ... 9
Table of contents ... 10
List of figures ... 15
List of tables ... 20
List of abbreviations and acronyms ... 23
Chapter 1 : Introduction ... 24
1 Background ... 24
1.1 Zika ... 24
1.2 Chikungunya ... 27
1.3 Colombia ... 30
1.4 Public health surveillance ... 33
1.5 Models in infectious disease epidemiology ... 37
1.6 Model fitting ... 38
2 Motivation ... 43
3 Data and ethics ... 44
3.1 ZIKV, CHIKV, and DENV surveillance datasets ... 44
3.2 ZIKV-associated neurological complications dataset ... 46
4 Objectives ... 47
Chapter 2 : Descriptive analysis of surveillance data for Zika virus disease and Zika virus-associated neurological complications in Colombia ... 49
Abstract ... 49
1 Introduction ... 49
1.1 Background ... 49
1.2 Aims ... 51
2 Data ... 51
2.1 Epidemiological data ... 51
2.2 Demographic data ... 52
3 Methods ... 52

4 Results 53
4.1 ZVD 53
4.2 ZIKV-associated neurological complications 62
5 Discussion 76
5.1 ZIKV 76
5.2 ZIKV-associated neurological complications 77
5.3 CHIKV 79
5.4 Conclusions and limitations 80
Chapter 3 : Impact of climactic and socioeconomic factors on reporting rates and basic reproduction numbers of Zika and chikungunya viruses in Colombia 81
Abstract 81
1 Introduction 81
1.1 Reproduction numbers 81
1.2 Reporting rates 88
1.3 Weather and arbovirus transmission 93
1.4 Aims 96
2 Data 97
2.1 Epidemiological data 97
2.2 Demographic data 97
2.3 Socioeconomic data 97
2.4 Weather data 98
3 Methods 99
3.1 Processing weather data 99
3.2 Inclusion criteria and level of analysis determination 101
3.3 Weekly time-varying reproduction numbers from EpiEstim 103
3.4 Non-parametric models of arbovirus transmission 104
3.5 Parametric models of arbovirus transmission 106
3.6 Model estimation and computing 111
3.7 Validation of parametric model fit and parameter fitting procedure 112
3.8 Comparing parameter estimates across departments 113
4 Results 113
4.1 Weather data 113
4.2 Epidemiological data 116
4.3 Weekly time-varying reproduction numbers from EpiEstim 117
4.4 Non-parametric models of arbovirus transmission 119
4.5 Fitting GAMs to the residuals of the linear regression models 124

4.6 Fitting Poisson models of arbovirus transmission 129
4.7 Fitting negative binomial models of arbovirus transmission 140
4.8 Validation of parameter fitting procedure 151
4.9 Sensitivity analysis of outlier thresholds 155
4.10 MCMC testing 160
5 Discussion 166
5.1 Non-parametric models of arbovirus transmission 167
5.2 Parametric models of arbovirus transmission 168
5.3 Conclusions and limitations 170
Chapter 4 : Estimating Zika virus attack rates and risk of Zika virus-associated neurological complications in Colombian capital cities with a Bayesian hierarchical model 173
Abstract 173
1 Introduction 174
1.1 Background 174
1.2 Aims 175
2 Data 176
2.1 Epidemiological data 176
2.2 Serological data 177
2.3 Demographic data 177
3 Methods 178
3.1 Bayesian hierarchical model 178
3.2 Expected number of excess neurological complications reported per 10,000 reported cases of ZVD 181
3.3 Model estimation and computing 181
3.4 Sensitivity analysis 182
4 Results 182
4.1 Bayesian hierarchical model 182
4.2 Bayesian hierarchical model by sex 189
4.3 Bayesian hierarchical model by age group 194
4.4 Model convergence and diagnostics 200
5 Discussion 208
Chapter 5 : Spatial and temporal invasion dynamics of the 2014-2017 Zika and chikungunya epidemics in Colombia 214
Abstract 214
1 Introduction 214
1.1 Spatiotemporal epidemiology 214

1.2 Human movements and disease spread 215
1.3 Spatiotemporal analyses of ZIKV and CHIKV 220
1.4 Spatiotemporal analyses of ZIKV and CHIKV in Colombia 221
1.5 Aims 223
2 Data 224
2.1 Epidemiological data 224
2.2 Demographic data 226
2.3 Distance metrics 226
2.4 Data for analysis of invasion risk factors 227
3 Methods 228
3.1 Invasion weeks 228
3.2 Elevation 229
3.3 Potential sources of the epidemics 230
3.4 Long-distance transmission events 230
3.5 Spatial interaction models 231
3.6 Model estimation and computing 234
3.7 Validation of gravity model fit and sensitivity analyses 235
3.8 Validation of parameter fitting procedure 236
3.9 Risk factors of invasion 236
4 Results 237
4.1 Characteristics of invaded cities 237
4.2 Spatiotemporal patterns in invasion weeks 239
4.3 Geographic origin of epidemics 245
4.4 Long-distance transmission events 246
4.5 Models of spread 250
4.6 Validation of model fit and parameter fitting procedure 259
4.7 Sensitivity analyses on spatial transmission model 269
4.8 MCMC testing 277
4.9 Risk factors of invasion 285
5 Discussion 290
5.1 Comparing alternative spatial interaction models 291
5.2 Gravity models 291
5.3 Joint models 294
5.4 Elevation 294
5.5 Conclusions and limitations 294

Chapter 6 : Discussion 299
1 Summary of findings 299
2 Future work and limitations 300
3 Implications of research 305
4 Challenges 306
5 Conclusions 309
References 311
Appendix S1: Additional plots comparing EpiEstim R_ts and model R_ts 332
Appendix S2: MCMC testing for best-fitting Poisson models 338
Appendix S3: Data for sex and age models, additional sensitivity analyses, and Stan code for chapter 4 345
Appendix S4: Week of invasion 353
CHIKV 353
ZIKV 366
Appendix S5: Probability distribution for week of invasion by city 378
CHIKV 378
ZIKV 385

List of figures

Figure 1.1 A female blood-fed *Aedes aegypti* mosquito. 25
Figure 1.2 Microcephaly and measuring head circumference in newborns. 27
Figure 1.3 An *Aedes albopictus* mosquito. 28
Figure 1.4 Predicted habitat suitability of *Aedes aegypti* and *Aedes albopictus* 30
Figure 1.5 Terrain maps of Colombia. 31
Figure 1.6 Flow chart of a traditional surveillance system. 36
Figure 1.7 Susceptible-infected-recovered (SIR) model. 38
Figure 1.8 Weekly reported cases of CF, DF, and ZVD in Colombia, January 2010-June 2017. 46
Figure 2.1 Population of Colombia by age group in 2016. 54
Figure 2.2 Frequency of ZVD cases by age group and sex in Colombia. 55
Figure 2.3 Observed attack rate of ZVD per 100,000 population by age group and sex in Colombia. 56
Figure 2.4 Frequency of female ZVD cases by age group and pregnancy status in Colombia. 57
Figure 2.5 ZVD over time by sex and pregnancy status in Colombia. 58
Figure 2.6 ZVD cases over time by department for 16 departments. 60
Figure 2.7 Number of ZVD cases over time by department for 16 departments (continued). 61
Figure 2.8 Number of cases of CF and ZVD over time in Colombia. 62
Figure 2.9 Frequency of ZIKV-associated neurological complications by age group in Colombia. 64
Figure 2.10 Observed attack rate of ZIKV-associated neurological complications per 100,000 population by age group in Colombia. 64
Figure 2.11 Frequency of ZIKV-associated neurological complications by age group in Colombia. 65
Figure 2.12 Observed attack rate of ZIKV-associated neurological complications per 1,000 cases of ZVD by age group and sex in Colombia. 66
Figure 2.13 Number of cases of ZVD and ZIKV-associated neurological complications over time in Colombia. 68
Figure 2.14 Number of ZIKV-associated neurological complications cases by sex and week of symptom onset in Colombia. 69
Figure 2.15 Final diagnosis for cases of ZIKV-associated neurological complications completed by a neurologist. 70
Figure 2.16 Maps of ZIKV-associated neurological complications. 72
Figure 2.17 Map of the number of ZIKV-associated neurological complications cases by city. 74
Figure 2.18 Maps of ZVD. 75
Figure 3.1 Map of 30 weather stations in Colombia. 100
Figure 3.2 Boxplots of incidence divided by infectivity in each department for all weeks for CHIKV and ZIKV. 110
Figure 3.3 Weekly time series of population weighted mean temperature in °C by department. 114
Figure 3.4 Weekly time series of population weighted cumulative precipitation in mm by department. 115

Figure 3.5 Histograms showing *(A)* the number of R_t estimates and *(B)* the number of weeks where estimated $R_t > 1$ by department. 119
Figure 3.6 R_t estimates versus cumulative incidence divided by population size of each department. 120
Figure 3.7 R_t estimates versus cumulative incidence divided by population size by department for CHIKV. 122
Figure 3.8 R_t estimates versus cumulative incidence divided by population size by department for ZIKV. 123
Figure 3.9 Smooth effect plot of mean weekly temperature in °C averaged over three weeks followed by a six-week lag prior to case reporting from best-fitting GAM for CHIKV. 125
Figure 3.10 Diagnostics of best-fitting GAM for CHIKV. 126
Figure 3.11 Smooth effect plots from best-fitting GAM for ZIKV. 128
Figure 3.12 Diagnostics of the best-fitting GAM for ZIKV. 129
Figure 3.13 R_t as a function of temperature and rainfall from Poisson models with weather covariates. 135
Figure 3.14 Heatmap showing predicted R_t as a function of temperature and rainfall from Poisson models with weather covariates. 136
Figure 3.15 Comparing median estimates of R_t from the best-fitting Poisson model for CHIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines). 138
Figure 3.16 Comparing median estimates of R_t from the best-fitting Poisson model for ZIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines). 139
Figure 3.17 Comparing median estimates of R_t from the best-fitting negative binomial model for CHIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines). 142
Figure 3.18 Comparing median estimates of R_t from the best-fitting negative binomial model for ZIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines). 143
Figure 3.19 Comparison of the posterior densities of estimated R_0s for CHIKV and ZIKV by department from the best-fitting negative binomial models. 146
Figure 3.20 Modeled and observed attack rates of CHIKV by department. 148
Figure 3.21 Modeled and observed attack rates of ZIKV by department. 149
Figure 3.22 Hexagon map of modeled attack rates of CHIKV and ZIKV. 150
Figure 3.23 Histograms of the posterior distribution of parameters obtained from a single dataset simulated from a Poisson model allowing for different R_0s across departments. .. 153
Figure 3.24 Histograms of the posterior distribution of parameters obtained from a single dataset simulated from a negative binomial model allowing for different R_0s across departments 154
Figure 3.25 Comparing median estimates of R_t from the best-fitting Poisson model for CHIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines) when no threshold is used to prevent outliers in the distribution of incidence divided by infectivity from contributing to the likelihood. 156
Figure 3.26 Comparing median estimates of R_t from the best-fitting Poisson model for ZIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines) when no threshold is used to prevent outliers in the distribution of incidence divided by infectivity from contributing to the likelihood. 157
Figure 3.27 Comparing median estimates of R_t from the best-fitting negative binomial model for CHIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines) when no threshold is used to prevent outliers in the distribution of incidence divided by infectivity from contributing to the likelihood. 158

Figure 3.28 Comparing median estimates of R_t from the best-fitting negative binomial model for ZIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines) when no threshold is used to prevent outliers in the distribution of incidence divided by infectivity from contributing to the likelihood. 159
Figure 3.29 Histograms of the posterior distributions of the best-fitting negative binomial model for CHIKV. 161
Figure 3.30 Histograms of the posterior distributions of the best-fitting negative binomial model for ZIKV 162
Figure 3.31 MCMC traces for the CHIKV model. 164
Figure 3.32 MCMC traces for the ZIKV model. 165
Figure 4.1 Number of reported cases of ZIKV-associated neurological complications (NC) and ZVD on a linear and a log-log scale for 28 capital cities. 183
Figure 4.2 Estimated ZIKV infection attack rates, ZVD reporting rates, and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections. 184
Figure 4.3 Reported cases of ZIKV-associated neurological complications (NC) per 10,000 reported cases of ZVD. 185
Figure 4.4 Effect of removing data on estimated ZIKV infection attack rates, ZVD reporting rates, and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections. 188
Figure 4.5 Estimated ZVD reporting rate and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by sex. 190
Figure 4.6 Comparison of the posterior densities of estimated ZVD reporting rate by sex for each city. 192
Figure 4.7 Comparison of the posterior densities of estimated number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by sex for each city. 193
Figure 4.8 Estimated ZVD reporting rate and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by age group 195
Figure 4.9 Comparison of the posterior densities of estimated ZVD reporting rate by age group for each city. 197
Figure 4.10 Comparison of the posterior densities of estimated number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by age group for each city. 198
Figure 4.11 Sensitivity of the prior distributions for ZIKV infection attack rates across age groups on estimated ZIKV infection attack rates, ZVD reporting rates, and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by age group. 200
Figure 4.12 Violin plots of the posterior distribution of ZIKV infection attack rate for all cities. 201
Figure 4.13 Violin plots of the posterior distribution of ZVD reporting rate for all cities. 202
Figure 4.14 Violin plots of the posterior distribution of the number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections for all cities. 203
Figure 4.15 MCMC traces for the ZIKV infection attack rate for all cities. 205
Figure 4.16 MCMC traces for the ZVD reporting rate for all cities. 206
Figure 4.17 MCMC traces for the number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections for all cities. 207

Figure 5.1 Epidemiological curves of CF and ZVD cases in Colombia by department, 2014-2017. 225
Figure 5.2 Map of predicted travel time to the nearest city in minutes. 227
Figure 5.3 Example of algorithm used to estimate week of invasion using the generation time method. 229
Figure 5.4 Comparison of distance metrics for invaded cities. 238
Figure 5.5 City elevation. 239
Figure 5.6 Heatmaps showing the spatiotemporal spread of CHIKV and ZIKV in Colombia. 240
Figure 5.7 Invasion weeks for a random sample of cities from the ZIKV dataset. 242
Figure 5.8 Geographic patterns of invasion weeks in studied cities in Colombia based on first reported cases. 243
Figure 5.9 Comparison of estimated invasion weeks using a method based on the first reported cases in each city (x-axis) and a method based on the generation time distribution of each infection (generation time method, y-axis). 244
Figure 5.10 Comparison of estimated invasion weeks using a method based on the generation time distribution (generation time method, x-axis) and a piecewise spline method (Charu method, as in [271], y-axis). 245
Figure 5.11 Correlations between city invasion weeks and geographic distance from first invaded cities for CHIKV and ZIKV. 246
Figure 5.12 Long-distance transmission events. 248
Figure 5.13 Best-fitting distance kernels for CHIKV (red) and ZIKV (blue). 262
Figure 5.14 Probability distribution of invasion weeks. 263
Figure 5.15 Probability distribution of first reported cases by department for CHIKV. 264
Figure 5.16 Probability distribution of first reported cases by department for ZIKV. 265
Figure 5.17 Probability distribution of invasion week for a random sample of cities for CHIKV. 266
Figure 5.18 Epidemic invasion simulations. 267
Figure 5.19 Epidemic invasion simulations (best-fitting Stouffer's rank models). 269
Figure 5.20 Epidemic simulations of the best-fitting gravity models showing the sensitivity of the thresholds used to determine invasion. 271
Figure 5.21 Epidemic invasion simulations produced from models with all cities. 272
Figure 5.22 Probability distribution of estimated invasion week by generation time method (colored lines). 273
Figure 5.23 Epidemic invasion simulations (generation time method). 274
Figure 5.24 Parameter estimates for the CHIKV model fitted using estimated invasion week by generation time method. 275
Figure 5.25 Parameter estimates for the ZIKV model fitted using estimated invasion week by generation time method. 276
Figure 5.26 Comparison of the distance kernel obtained when running the CHIKV gravity model from week 12 versus the entire dataset. 277
Figure 5.27 Posterior distributions and correlation of parameters for best-fitting CHIKV model. 279
Figure 5.28 Posterior distributions and correlation of parameters for best-fitting ZIKV model. 279
Figure 5.29 Three chains run using different start values for the best-fitting CHIKV gravity model. 281

Figure 5.30 Three chains run using different start values for the best-fitting ZIKV gravity model. 282
Figure 5.31 Histograms of the posterior distributions of the best-fitting CHIKV gravity model. 283
Figure 5.32 Histograms of the posterior distributions of the best-fitting ZIKV gravity model. 283
Figure 5.33 MCMC traces for the CHIKV model. 284
Figure 5.34 MCMC traces for the ZIKV model. 285
Figure 5.35 Distribution of invasion week by dengue risk level. 290
Figure 6.1 Comparison of infection attack rates from chapters 3 and 4 in departments and capital cities, respectively. 302

List of tables

Table 2.1 Number of ZVD cases by sex in Colombia. 53
Table 2.2 Observed attack rate and 95% confidence intervals of ZIKV-associated neurological complications per 1,000 ZVD cases by department. 73
Table 3.1 R_0 estimates for CHIKV and ZIKV in Colombia from the literature. 88
Table 3.2 Reporting rates derived from multisite seroprevalence study of ZIKV and CHIKV in Colombia [191]. 91
Table 3.3 Estimates of reporting rates from modeling and community-based studies for CF and ZVD in Colombia from the literature. 93
Table 3.4 Mean within- and between-department standard deviation (sd) of temperature and rainfall for CHIKV and ZIKV. 103
Table 3.5 Estimates of the mean and standard deviation of the generation time distribution (GTD) and serial interval distribution (SID) for CHIKV and ZIKV from the literature. 104
Table 3.6 Prior distributions for parameters in the parametric model based on the renewal equation. 112
Table 3.7 Cumulative incidence of suspected and laboratory-confirmed cases of CF and ZVD in Colombia, 2014-2017. 117
Table 3.8 AIC values of fitted GAMs for CHIKV. 124
Table 3.9 AIC values of fitted GAMs for ZIKV. 127
Table 3.10 Testing the effect of mean weekly temperature in the weeks prior to case reporting on CHIKV and ZIKV transmission with Poisson models. 130
Table 3.11 Testing the effect of cumulative weekly rainfall in the weeks prior to case reporting on CHIKV and ZIKV transmission with Poisson models. 131
Table 3.12 Testing whether two standard deviations instead of one better describe the effect of temperature on CHIKV and ZIKV transmission in the weeks prior to case reporting with Poisson models. 131
Table 3.13 Effect of socioeconomic factors on CHIKV and ZIKV transmission using Poisson models. 132
Table 3.14 Estimated R_0s and reporting rate of CHIKV from Poisson models with weather covariates and multiple R_0s. 133
Table 3.15 Estimated R_0s and reporting rate of ZIKV from Poisson models with weather covariates and multiple R_0s. 134
Table 3.16 Effect of socioeconomic factors on CHIKV and ZIKV transmission using negative binomial models. 140
Table 3.17 Estimated R_0s and reporting rate of CHIKV and ZIKV from negative binomial models with multiple R_0s. 141
Table 3.18 Estimated R_0 values of CHIKV and ZIKV for each department from the best-fitting negative binomial models. 145
Table 3.19 True values and median parameter estimates obtained from a single dataset simulated from the Poisson model with weather covariates as well as observed population sizes for 29 departments, observed weather data, and the generation time distribution for CHIKV. 151
Table 3.20 True values and median parameter estimates obtained from a single dataset simulated from either the Poisson model with multiple R_0s or the negative binomial model with multiple R_0s. 152

Table 3.21 Gelman-Rubin statistic for each of the best-fitting negative binomial models (after removing the burn-in). 160
Table 3.22 Acceptance percentages for parameters of the best-fitting negative binomial models for CHIKV and ZIKV (after removing the burn-in). 166
Table 3.23 Effective sample sizes from one chain for each of the best-fitting negative binomial models (after removing the burn-in). 166
Table 4.1 Data sources for chapter 4. 176
Table 4.2 Epidemiological and demographic data for 28 Colombian capital cities. 178
Table 4.3 Estimated ZIKV infection attack rates, ZVD reporting rates, and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by city. 187
Table 4.4 Estimated hyperprior distributions (per 10,000). 187
Table 4.5 Overall estimated ZVD reporting rate and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by sex. 189
Table 4.6 Estimated hyperprior distributions (shown per 10,000) by sex and overall ZIKV infection attack rate for both males and females. 189
Table 4.7 Comparison of the posterior probabilities (PP) of estimated ZVD reporting rate and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by sex for each city. 191
Table 4.8 Overall estimated ZVD reporting rate and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by age group 194
Table 4.9 Estimated hyperprior distributions (shown per 10,000) by age group and ZIKV infection attack rate for all ages. 194
Table 4.10 Comparison of the posterior probabilities (PP) of estimated ZVD reporting rate and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by age group for each city. 196
Table 5.1 Population sizes of invaded cities. 237
Table 5.2 Epidemiological characteristics of CHIKV and ZIKV epidemics in Colombia. 242
Table 5.3 Summary statistics of the d-D distributions showing that ZIKV and CHIKV exhibited similar patterns of transmission. 247
Table 5.4 Recipient and potential source cities of long-distance transmission events of CHIKV. 249
Table 5.5 Recipient and potential source cities of long-distance transmission events of ZIKV. 249
Table 5.6 Comparison of alternative models of CHIKV and ZIKV spread in Colombia. 252
Table 5.7 Parameter estimates for six models of CHIKV for 337 cities using geographic distance. 254
Table 5.8 Parameter estimates for six models of CHIKV for 337 cities using travel time between cities. 254
Table 5.9 Parameter estimates for six models of ZIKV for 287 cities using geographic distance. 255
Table 5.10 Parameter estimates for six models of ZIKV for 287 cities using travel time between cities. 255
Table 5.11 Comparison of individual versus joint models of CHIKV and ZIKV spread in Colombia. 257
Table 5.12 Parameter estimates for eight models of CHIKV in Colombia for 338 cities. 260
Table 5.13 Parameter estimates for eight models of ZIKV in Colombia for 288 cities. 261

Table 5.14 Comparison of parameter estimates from observed data versus simulated data. 268
Table 5.15 Comparison of parameter estimates from gravity models fitted to different numbers of cities using thresholds of 10, 20, and 30 cumulative reported cases. 270
Table 5.16 Gelman-Rubin statistic for each of the best-fitting gravity models (after removing the burn-in). 280
Table 5.17 Acceptance percentages for parameters of the best-fitting CHIKV and ZIKV gravity models..... 285
Table 5.18 Univariate analysis of risk factors of CHIKV invasion..... 286
Table 5.19 Univariate analysis of risk factors of ZIKV invasion. 287
Table 5.20 Best-fitting logistic regression model of CHIKV invasion. 289
Table 5.21 Best-fitting logistic regression model of ZIKV invasion..... 289

List of abbreviations and acronyms

AIC	Akaike information criterion
CF	chikungunya fever
CHIKV	chikungunya virus
CI	confidence interval
CrI	credible interval
CZS	congenital Zika syndrome
DANE	Departamento Administrativo Nacional de Estadística (National Administrative Department of Statistics)
DENV	dengue virus
DF	dengue fever
DIC	deviance information criterion
ECSA	East Central and South African
edf	estimated degrees of freedom
ELISA	enzyme-linked immunosorbent assays
GAM	generalized additive model
GBS	Guillain-Barré syndrome
HIV	human immunodeficiency virus
ICD	International Classification of Diseases
IgG	immunoglobulin G
IgM	immunoglobulin M
INS	Instituto Nacional de Salud (National Institute of Health)
IRR	incidence rate ratio
MAYV	Mayaro virus
MCMC	Markov chain Monte Carlo
MOH	Ministerio de Salud y Protección Social (Ministry of Health and Social Protection)
NC	neurological complications
NOAA	National Oceanic and Atmospheric Administration
OR	odds ratio
PAHO	Pan American Health Organization
RIPS	Registros Individuales de Prestación de Servicios de Salud (Individual Records of Health Services Provision)
RR	risk ratio
RT-PCR	reverse transcriptase polymerase chain reaction
SGSSS	Sistema General de Seguridad Social en Salud (System of Comprehensive Social Security in Health)
SIG-OT	Sistema de Información Geográfica para la Planeación y el Ordenamiento Territorial (Geographic Information System for Territorial Planning and Order)
SIR	susceptible-infected-recovered
UK	United Kingdom
USA	United States of America
WHO	World Health Organization
ZIKV	Zika virus
ZVD	Zika virus disease

Chapter 1: Introduction

1 Background

1.1 Zika

Zika virus disease (ZVD) is an emerging infectious disease caused by Zika virus (ZIKV), a single-stranded RNA virus that belongs to the genus *Flavivirus* [1]. Flaviviruses are mainly spread by arthropods, especially ticks and mosquitoes; hence, they are also called "arboviruses." The clinical presentation of acute flavivirus infection in humans is typically mild but can be life-threatening: hemorrhagic fever, shock, encephalitis, paralysis, congenital defects, and liver failure have been reported [1]. These complications can result in death or long-term disability in survivors. Other flaviviruses include yellow fever virus, dengue virus (DENV), West Nile virus, Powassan virus, and Japanese encephalitis virus [1]. The geographical range of flaviviruses is wide and covers most of the tropics. In fact, about half of the world's population is estimated to live in areas at risk of DENV transmission [2].

ZIKV is transmitted to humans according to two ecological cycles. The bite of infected *Aedes* spp. mosquitoes, particularly *Ae. aegypti*, is the main route of ZIKV transmission in urban epidemics (Figure 1.1) [3]. Human-to-human transmission can occur through sex and transfusion of infected blood products as well as from mother to child during or after pregnancy [4-6]. In Africa, an enzootic cycle occurs in which the virus circulates between non-human primates and sylvatic (forest-dwelling) mosquitoes with occasional spillover into humans [7].

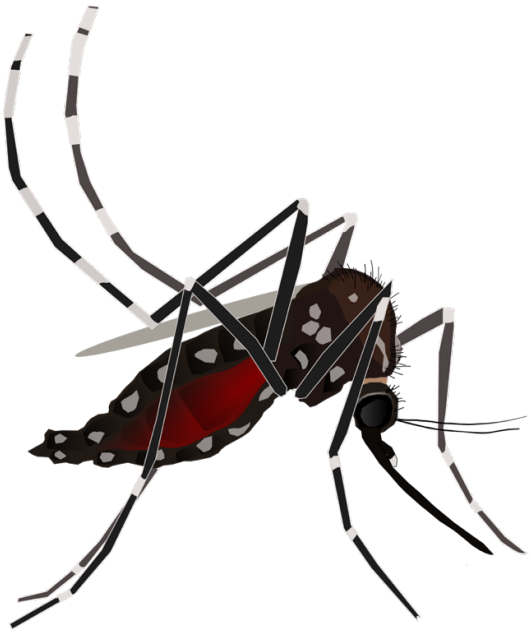


Figure 1.1 A female blood-fed *Aedes aegypti* mosquito. Reproduced from [8].

Symptoms of ZVD include fever, rash, joint pain, muscle aches, conjunctivitis, and headache. Severe cases and deaths are rarely reported [9], and only about 20%-25% of persons infected by ZIKV show symptoms [10]. Clinical diagnosis of ZIKV infection can only be made reliably in the absence of other circulating arboviruses [11]. In some regions, ZIKV co-circulates with both chikungunya virus (CHIKV) and DENV, which are transmitted by the same vectors and cause similar symptoms [12]. For these reasons, outbreaks of ZIKV can be difficult to detect.

Laboratory confirmation of ZIKV infection can take the form of viral culture, molecular assays, or serological assays. Once considered the gold standard, viral culture is no longer routinely used for diagnostic purposes due to long turnaround times and the need for specialized equipment and skilled laboratory staff [13].

Molecular assays for viral RNA, such as reverse transcriptase polymerase chain reaction (RT-PCR), can be performed during the acute phase of ZIKV infection and up to two weeks after symptom onset [14]. Different types of bodily fluids can be tested using molecular methods including serum, whole blood, urine, cerebrospinal fluid, and amniotic fluid. Unfortunately, PCR-based tests are not always available in resource-limited settings due to costly equipment and lack of trained personnel.

Serological assays test for the presence of ZIKV specific immunoglobulin (Ig) M or IgG antibodies and are performed on serum, whole blood, or plasma. IgM antibodies can be detected with enzyme-linked immunosorbent assays (ELISAs) approximately one week to two months after ZIKV infection [14]. Testing for IgG antibodies, which rise soon after IgM and can persist for years to decades, can reveal ZIKV infections that occurred in the past. The gold standard for this test is the plaque reduction neutralization test [14]. However, the interpretation of ELISAs and plaque reduction neutralization tests are not always straightforward, especially in patients with prior exposure to DENV, due to cross-reactivity between antibodies elicited by different flavivirus infections or by vaccination [14].

No approved medical countermeasures exist to prevent or treat ZIKV infection [12], but several vaccine candidates are being evaluated in pre-clinical through phase 2 studies [15]. ZIKV was first isolated from rhesus monkeys in the Zika forest in Uganda in 1947 [3]. In 1952, the first cases of ZVD in humans were discovered in Uganda and present-day Tanzania. Sporadic cases were reported in humans from the 1950s through the end of the twentieth century in Africa and Asia [3]. In 2007, the first major outbreak of ZIKV occurred on Yap Island, Federated States of Micronesia [16]. Five years later, genetic sequences of ZIKV strains isolated from various countries were analyzed. The resulting phylogenetic trees uncovered two main lineages (African and Asian) [17]. Between 2013 and 2014, a neurological condition known as Guillain-Barré syndrome (GBS) was linked to ZIKV during an outbreak in French Polynesia [3]. In May 2015, the first cases of ZVD were reported in Brazil. However, genetic analyses have suggested that the virus may have been introduced as early as 2013 and that the Asian lineage was responsible [18].

In October 2015, Brazil reported an association between ZIKV infection during pregnancy and microcephaly [3]. Microcephaly is a birth defect characterized by head size that is smaller than expected based on age and sex (Figure 1.2). Head size is related to underlying brain size in infants diagnosed with microcephaly. The health consequences of small head size may include seizures, issues with vision or hearing, and developmental disabilities [19]. Cases of microcephaly were also identified retrospectively following the outbreak in French Polynesia [20]. Microcephaly is now recognized as a feature of congenital Zika syndrome (CZS), a pattern of birth defects found among fetuses and newborns of ZIKV-infected mothers [19].

Figure 1.2 Microcephaly and measuring head circumference in newborns. Reproduced from [19].

From Brazil, ZIKV spread widely throughout Latin America and the Caribbean, and in February 2016, the World Health Organization (WHO) declared the cluster of microcephaly and other neurological complications a Public Health Emergency of International Concern [21]. By January 2018, 583,451 confirmed and suspected cases of ZVD had been reported in the Americas, of which 223,477 were confirmed cases [22]. However, these numbers are likely underreported due in part to the high rate of asymptomatic infections, and modeling studies have suggested that ZIKV continued to spread in the region until the number of susceptible individuals declined sufficiently for herd immunity to be reached [23].

As of February 2020, 91 countries and territories were listed as having current or previous ZIKV transmission in all WHO regions [24].

1.2 Chikungunya

CHIKV is a single-stranded RNA virus that is spread by several species of mosquitoes [25]. The virus causes chikungunya fever (CF) and belongs to the genus *Alphavirus* which also includes Eastern equine encephalitis virus, Western equine encephalitis virus, Madariaga virus, Ross River virus, Sindbis virus, Mayaro virus (MAYV), and O'nyong-nyong virus. As their names imply, some of these viruses cause encephalitis in humans which can be fatal [26]. Most commonly, alphaviruses manifest as a febrile disease accompanied by severe pain in one or more joints. Although alphaviruses have been isolated from every continent

except Antarctica, individual viruses in this genus tend to have a more limited geographical distribution [26].

Similar to ZIKV, CHIKV is maintained in enzootic and epidemic transmission cycles. In Asia and the Americas, *Ae. aegypti* and *Ae. albopictus* are the principal vectors in the epidemic cycle (Figure 1.3). Enzootic CHIKV is present in Africa and is not well understood. Transmission is likely maintained by several species of mosquitoes, including *Ae. africanus* and *Ae. furcifer* [27]. Non-human primates may act as amplification hosts [28] with other animals such as rodents, bats, birds, and reptiles serving as reservoir hosts [29-32].

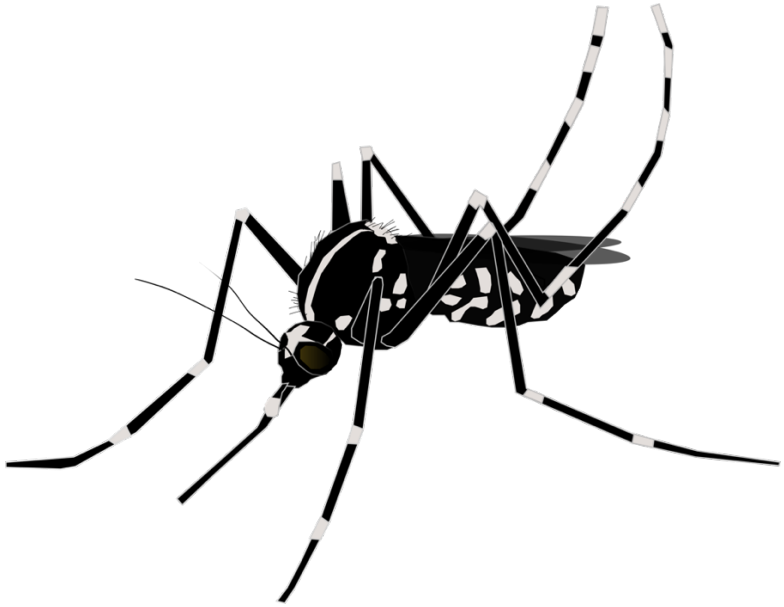


Figure 1.3 An *Aedes albopictus* mosquito. Reproduced from [8].

Unlike ZIKV, which is named after a place, chikungunya is a descriptive word. From the Makonde language, “chikungunya” can be roughly translated as the “disease that bends up the joints” [33]. The name refers to the clinical presentation of CF, which includes sudden onset of fever followed by crippling joint pain, headache, muscle aches, and rash.

Although symptoms of the acute infection typically subside in a week, some patients progress to chronic joint pain which can last for weeks or months [25]. Previously healthy individuals have experienced cardiovascular disorders, such as arrythmias, myocarditis, and myocardial infarction, as well as neurological disorders, including encephalitis and meningoencephalitis, following CHIKV infection [34]. These complications are rare. Nevertheless, young children and the elderly, especially those with underlying medical

conditions, are at high risk of developing severe disease [25]. In contrast to ZIKV, 75%-97% of persons infected by CHIKV exhibit symptoms [35].

CHIKV infection can be confirmed through viral culture, molecular assays, or serological assays, and a combination of the latter two approaches is most commonly used. As with ZIKV, the interpretation of serological assays for CHIKV is problematic due to cross-reactivity with other alphaviruses, including MAYV and O'nyong-nyong virus [36].

Although there are no licensed drugs to treat or prevent CHIKV infection [37], several vaccine candidates are currently under investigation [38]. As of April 2020, three CHIKV vaccines had completed phase 2 clinical trials [39]. In September 2020, Valneva became the first company to initiate a phase 3 clinical trial of a CHIKV vaccine [40]. Recruitment of over 4,000 adults across the United States of America (USA) was completed in April 2021 [41]. Positive results from the trial would be expected to support the licensure of the live-attenuated, single dose vaccine. The target population will include travelers, military personnel, and people living in endemic regions [40].

At least four lineages of CHIKV have been proposed in the literature: Indian Ocean, East Central and South African (ECSA), Asian Urban, and West African. In 2019, a phylogenetic analysis incorporating new virus isolates suggested that the ECSA lineage should be further divided into Eastern African, South American, Middle African, and African/Asian lineages and that the Asian Urban lineage should be further divided into Asian Urban and American lineages [42].

The first CHIKV outbreak may have occurred as early as the seventeenth century in Cairo, Egypt [43]. Due to non-specific symptoms and lack of laboratory-based diagnostic tools, many historical CHIKV outbreaks, including the 1827-1828 outbreaks in the West Indies and southern USA, were classified under the umbrella of dengue-like syndromes [43, 44].

In 1952, CHIKV was first isolated during an outbreak in present-day Tanzania [25]. Over the next several decades, it caused large epidemics throughout Africa and Asia as well as islands in the Indian Ocean [45]. In 2006, India experienced an outbreak of CHIKV with nearly 1.4 million cases; although no attributable deaths were officially reported, one study estimated nearly 3,000 excess deaths in the city of Ahmedabad alone [46]. CHIKV outbreaks have also occurred in Europe, including a 2007 outbreak in Italy with 205 cases and one death [47].

CHIKV returned to the Caribbean in December 2013 with local transmission first reported in St. Martin. The virus belonged to the Asian lineage and was likely imported from Southeast Asia or Oceania [27]. Within a year over one million cases were reported in the region, along with severe cases and deaths [37]. By the end of 2017, 2,225,014 suspected cases of CF had been reported in the Americas, including 338,963 confirmed cases [48].

The expansion of CHIKV into new territories over the last few decades has been attributed to adaptive mutations in the virus that improved transmission by *Ae. albopictus* mosquitoes [25]. According to a recent modeling study, 215 countries have potentially suitable habitat for *Ae. aegypti* and/or *Ae. albopictus* mosquitoes (Figure 1.4) [49]. The range of these mosquitoes is increasing, and this trend is predicted to continue through at least 2050 as a result of urbanization and climate change [50].

As of October 2020, CHIKV cases had been reported in 115 countries and territories in all WHO regions [51].

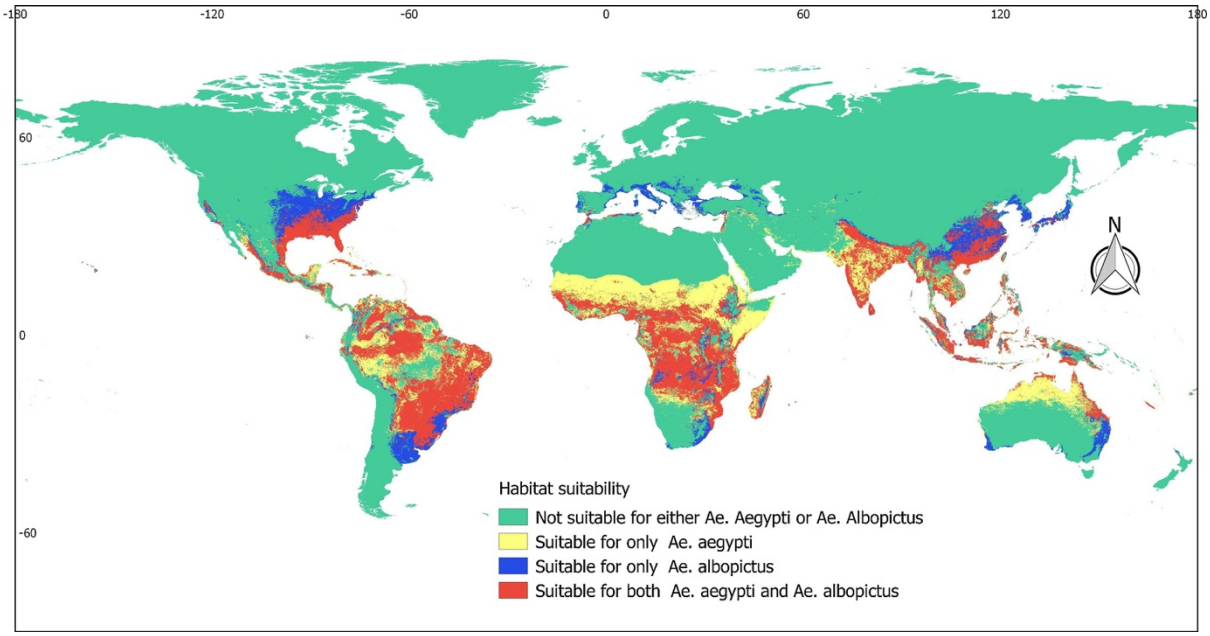


Figure 1.4 Predicted habitat suitability of *Aedes aegypti* and *Aedes albopictus*. Reproduced from [49].

1.3 Colombia

Colombia is located in the northwest corner of South America, where it borders Ecuador, Peru, Panama, Brazil, and Venezuela (Figure 1.5) [52]. The climate is tropical along the coast

and cooler at high elevations. The Andes Mountains pass through the southwest corner and cross diagonally through the center of the country. The elevation of Colombia varies considerably, from sea level at the Pacific Ocean to 5,730 meters above sea level at Pico Cristobal Colon [52]. About half of the country's population of nearly 48 million lives in areas of high elevation [53]. Seventy-seven percent of Colombians reside in urban areas, and the three largest cities are Bogotá (the capital), Medellín, and Cali [52]. The country is organized into 32 departments (administrative level 1) and 1,122 municipalities (or cities, administrative level 2) [54]. The Archipelago of San Andrés, Providencia and Santa Catalina, commonly referred to as San Andrés and Providencia, is the only department that is not attached to the mainland. The islands are located in the Caribbean Sea, 645 km northwest of Colombia.

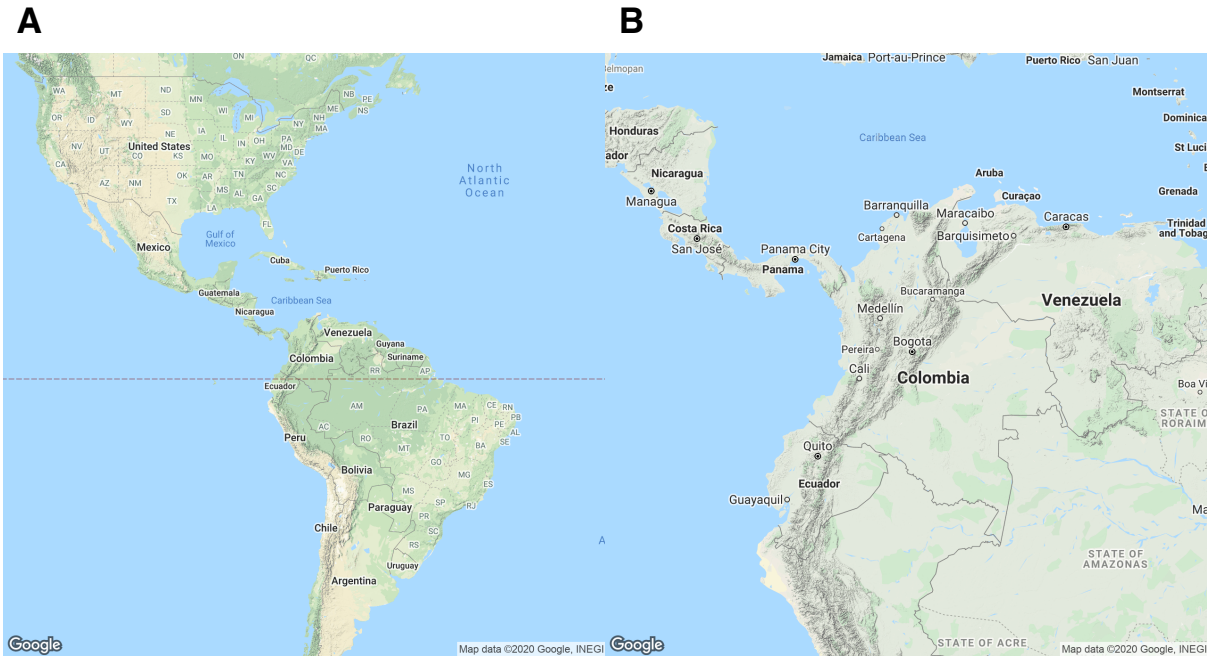


Figure 1.5 Terrain maps of Colombia. *(A)* Colombia's location in the Americas. *(B)* A close-up map showing the major cities. Both maps were made in R using ggmap (version 3.0.0) with data from Google [55].

Civil war and armed conflict have been ongoing in Colombia for over 50 years, resulting in at least 260,000 deaths, tens of thousands of kidnappings, and high levels of population displacement [56]. In 2016, the government signed a historic peace deal with the country's largest armed group, the Revolutionary Armed Forces of Colombia, formally ending the

conflict. Despite the agreement, other armed groups such as the National Liberation Army continue to operate in the country [57]. Natural disasters, including earthquakes, landslides, and floods, have also contributed to internal displacement. At the end of 2019, an estimated 5.6 million Colombians were internally displaced [57], with women, children, adolescents under 14, and ethnic minorities being the groups most affected [58].

Since 1991, healthcare in Colombia has been a fundamental right that is protected by the constitution [59]. The country's healthcare system, known as the System of Comprehensive Social Security in Health (SGSSS), is composed of two main parts, the contributory regime and the subsidized regime. The contributory regime applies to workers, whereas the subsidized regime applies to those who cannot afford to pay [58]. Special regimes apply for other groups, such as teachers and the military. Membership in the SGSSS is mandatory, and over half of Colombians are subsidized by the government. By 2016, healthcare coverage in Colombia reached about 96% [59]. Healthcare delivery is performed by institutional health service providers, who can be either public or private [58]. These providers are contracted by the insurers.

Economic inequality is rife in Colombia. The Gini index, which is based on primary household survey data, measures income inequality among individuals or households within countries. It ranges from 0 to 100 with a value of 0 meaning income is distributed equally among the population and 100 meaning one person earns all of the income in the country. According to the World Bank, Colombia's estimated Gini index of 51.3 in 2019 was the 12th highest (most unequal) out of 164 countries [60]. Income inequality is associated with worse health outcomes at the population level, and evidence suggests that the relationship may even meet epidemiological criteria for causality [61].

There is a high risk of major infectious diseases in Colombia, including bacterial diarrhea from food or water as well as vector-borne diseases [52]. Each year, Colombia endures DENV epidemics caused by one or more of the four known viral serotypes. Malaria transmission is also recorded annually, with epidemic cycles of two to seven years [58]. In addition to infectious diseases, Colombia is experiencing a "double burden" of non-communicable diseases due to demographic changes that have occurred over the last several decades. Declines in fertility and population growth rates, as well as increases in the

average life expectancy at birth, have begun to shift long-term patterns in morbidity and mortality [62]. This shift, whereby chronic diseases such as cardiovascular disease and cancer replace infectious diseases as the primary causes of death and disability, is also known as the epidemiological transition and is occurring in many low- and middle-income countries across the developing world.

1.4 Public health surveillance

Public health surveillance is defined as "the ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those responsible for prevention and control" [63]. The goal of these activities is to provide actionable information to guide public health decisions. Disease prevention, program planning and management, health promotion, quality improvement, and resource allocation are all areas that can be informed by public health surveillance [64].

The practice of gathering and utilizing morbidity and mortality data to improve population health has existed for hundreds of years in high-income countries [65]. One of the earliest examples of surveillance was for plague in the seventeenth century. The number of burials and causes of death for the City of London and surrounding areas were collected, analyzed, and published in the weekly Bills of Mortality to track the course of the epidemic [65]. In the nineteenth century, surveillance systems were put in place to monitor smallpox, influenza, and cholera in the United Kingdom (UK) [66]. The adoption of the International Classification of Diseases (ICD) around the turn of the twentieth century represented a major development in public health surveillance. ICD established an international standard for disease reporting, allowing disease trends to be compared across time and between countries [67].

WHO's legally-binding International Health Regulations require 196 countries to maintain national systems for public health surveillance and response [68]. The systematic monitoring of infectious diseases is especially important for events that may constitute Public Health Emergencies of International Concern, such as the ZIKV epidemic in the Americas and the COVID-19 pandemic. For some countries, the establishment and maintenance of surveillance systems can be accomplished through legislation [63, 69].

Ideally, surveillance activities would also be coordinated at the regional level through intergovernmental agreements and the WHO's regional offices [70].

There are two main types of public health surveillance, passive and active [63]. Passive surveillance relies on healthcare providers to regularly report notifiable conditions to public health authorities. A notifiable infectious disease is one that must be reported in a timely manner in order to prevent and control the disease. Examples include anthrax, botulism, human immunodeficiency virus (HIV), rabies, and Ebola [71]. Passive surveillance is inexpensive but may result in incomplete data [72]. In contrast, active surveillance involves public health authorities contacting healthcare providers to request reports for specific health conditions. This approach is more costly than passive surveillance but ensures more complete reporting [63]. In addition to surveillance systems for human diseases, there are also surveillance systems for plant and animal diseases [73].

Figure 1.6 shows a flow chart of a traditional surveillance system for a health condition in humans [74]. The process starts when a person experiences a health event of interest and seeks medical attention. Healthcare providers then use clinical symptoms and/or laboratory tests to confirm a diagnosis. They may also ask the patient about risk factors, such as international travel or exposure to sick people or animals [69]. Next, healthcare providers are responsible for reporting the case to the surveillance system and notifying the appropriate health jurisdiction. The lowest level of jurisdiction is typically notified first; however, there may be exceptions for diseases that pose a risk to national security [75, 76]. The data users are in charge of managing, cleaning, and storing data, as well as analyzing and interpreting it. Maintaining patient confidentiality is essential at this step and throughout the reporting process. Patient data can be protected by assigning unique ID numbers to each case and only sharing anonymized or aggregated data for legitimate research purposes [77]. Finally, information from the data users and reporting bodies feeds back to the general public and policymakers.

Data generated from surveillance systems are used to describe the distribution of health conditions in a population and are analyzed in terms of person, place, and time [65]. Across time, long-term trends, cyclic trends, seasonal trends, and epidemic occurrences should be considered [65]. The analysis of surveillance data by demographic characteristics ("person")

can help identify risk groups. While datasets categorized by sex and age are most common, other variables, such as socioeconomic status, race/ethnicity, occupation, risk factors, and hospitalization, may also be available [65]. Geographical analysis can identify areas where disease is increasing or decreasing. In doing so, the use of rates is imperative to adjust for the effects of population density on disease incidence [65].

Surveillance data are often presented in tables, graphs, charts, and maps. Statistical tests may be used to determine whether trends are significant [72]. In addition to descriptive epidemiological methods, more complex mathematical and statistical models are increasingly employed. These new methods are possible in part due to the availability of new data streams and tools, including electronic health records, medical claims data, and digital disease surveillance [66, 78].

Ideally, surveillance systems are representative of the population, timely, flexible, useful, and cost effective [63, 65]. They should be evaluated on a regular basis to ensure their objectives are being met and sources of error and delays are minimized [65].

In Colombia, the national population-based surveillance system, known as Sivigila, is operated by the Instituto Nacional de Salud (National Institute of Health, INS). Individual notifications of events of public health interest, including conditions with infectious, non-communicable, and environmental etiologies, have been reported to Sivigila since 2007 [79].

As mentioned in section 1.3, insurers contract with specific healthcare institutions in the Colombian health system. As a result, individuals may have to travel long distances for treatment. This means that Colombians living in rural areas with limited options for transportation may seek healthcare less frequently compared to those living in large, urban cities, potentially biasing the surveillance data.

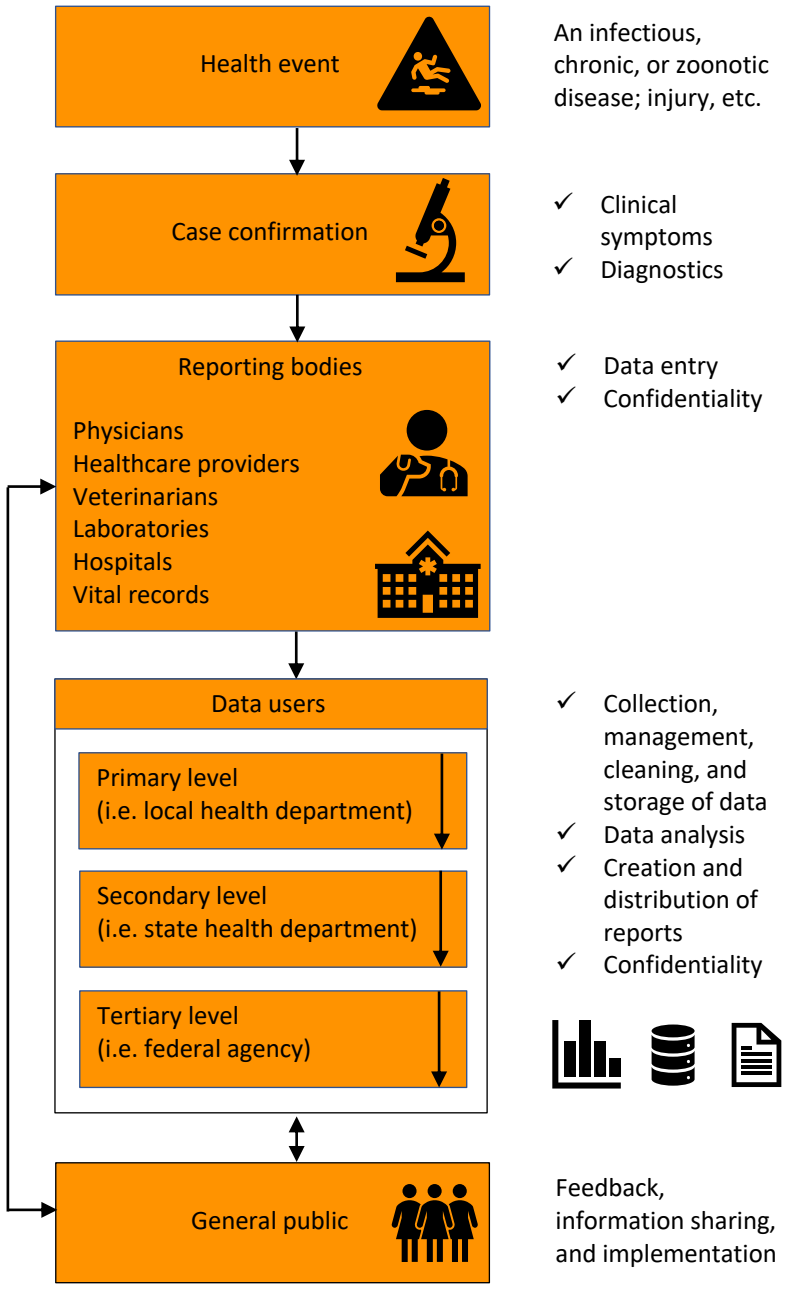


Figure 1.6 Flow chart of a traditional surveillance system. Adapted from [74].

1.5 Models in infectious disease epidemiology

Models are mathematical tools used to describe the behavior of systems or phenomena. We rely on information from models for many aspects of everyday life, including weather forecasts and trip or commute planning. Models are also used to predict the outcome of elections and stock market prices as well as study linguistics [80] and music [81]. Mathematical models are expressed using mathematical concepts or language, which allows the objects under study to be quantified and facilitates rigorous analysis [82].

Models range in complexity from simple "toy" models to highly complex models. The most complicated models may be developed over a period of years by teams of experts. The level of complexity for a particular problem depends on three main factors: (i) the precision required, (ii) available data, and (iii) how quickly the results are needed [83]. There is an adage that all models are wrong, but some are useful. Indeed, building a model involves tradeoffs between accuracy, transparency, and flexibility [83]. Accuracy refers to a model's ability to reproduce observed data and predict future dynamics. Transparency means to what extent the model components can be understood, and flexibility describes how easily the model can be adapted to different situations.

Prediction and understanding are two main purposes of models. While accuracy is considered important for predictive models, transparency is paramount for increasing understanding [83]. Predictive models can be used to identify epidemics early and guide policy decisions. Models can also explain how an infectious disease spreads in a population as well as reveal gaps in knowledge. Although all models have limitations, good models are only as complicated as they need to be and ideally can be fully parameterized from available data [83].

Several different model types are often encountered in infectious disease epidemiology. In general, they can be classified into stochastic and deterministic. Stochastic models capture the randomness associated with infectious processes, and a different result is obtained each time the model is run. In contrast, deterministic models represent the expected behavior on average and always produce the same result [82]. Deterministic models are more popular in the literature due to their simplicity and minimal computational requirements, yet

stochastic models are better suited for studying outbreaks and transmission of infectious diseases in small populations [82].

Examples of models used in infectious disease epidemiology include simple epidemic models, agent-based models, network models, spatial models, and multi-pathogen/multi-host models [83]. All of these models must account for changes in the infection status of individuals in the population but differ in how the population is represented and how individuals interact within the population [82]. In compartmental models, such as the susceptible-infected-recovered (SIR) model, the population is categorized by infection status, and the numbers in each group are tracked over time (Figure 1.7). Agent-based models, conversely, represent each person as an individual and can incorporate individual-level characteristics and behaviors [82].

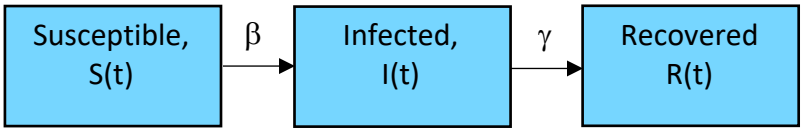


Figure 1.7 Susceptible-infected-recovered (SIR) model. The population is divided into three compartments based on whether individuals are susceptible to infection, are infected and infectious, or have recovered from infection and are therefore immune. Individuals become infected at the rate of β and move from the box on the left-hand side to the middle box. Infected individuals recover at a rate determined by γ, moving from the middle box to the box on the right-hand side.

1.6 Model fitting

1.6.1 Classical and Bayesian inference

Statistical inference is a framework which can be used to test beliefs about the world against data [84]. Beliefs are represented by models of probability. The models are probabilistic due to incomplete understanding of a system's complexity [84].

There are two main approaches to statistical inference, classical (or "frequentist") and Bayesian. In classical statistics, events occur according to probabilities. These probabilities are seen as the long-term frequencies with which a particular result would be obtained in an infinite number of identical experiments [84]. In Bayesian statistics, on the other hand,

probabilities are used to convey subjective beliefs. Intuitively, these beliefs can be revised when new data become available [84]. The goal of both classical and Bayesian statistics is to estimate parameters, or unknown quantities of interest.

In some situations, classical and Bayesian analyses can produce nearly identical results. However, there are some problems that are more suited to one approach over the other. Classical inference involves summarizing data [85]. It should produce estimates that are correct on average (unbiased) with the true parameter values covered by confidence intervals (CIs) 95% of the time. Weaknesses of this approach become apparent when applied to small studies and data that are indirect or highly variable [85].

Bayesian inference is capable of much more than simply summarizing data. Some claim that this approach gives the best predictions about future outcomes and about the results of future experiments [84]. Compared to confidence intervals, credible intervals (CrI) are easier to interpret. Bayesian statistics may also be more appropriate for fitting complex models. A controversial aspect of Bayesian statistics is that prior information must be specified [85]. Bayesian statistics will be used throughout this thesis.

1.6.2 Bayesian statistics

Bayesian inference begins with a probabilistic description of beliefs, otherwise known as a prior distribution. After the data are collected, the prior distribution and the data are combined in a model. The result is called the posterior distribution, which is used to test hypotheses. Parameters are estimated during this process [84].

Bayes' rule provides the basis for updating beliefs about a parameter or model structure in the context of new data. It is used to estimate a probability distribution for parameters after the sample of data is observed [84]. The rule is named for Thomas Bayes, an English clergyman who developed a mathematical theory about cause and effect in the eighteenth century. Bayes never published these ideas, but after he died, the work was rediscovered, corrected, and published by the Welsh minister Richard Price. Additional contributions from French mathematician Pierre Simon Laplace led to the modern version of Bayes' rule that is used today [84].

Bayes' rule is written as,

$$p(\theta|data) = \frac{p(data|\theta) \times p(\theta)}{p(data)},$$

where θ represents the parameters and p is a probability distribution [84]. In the numerator on the right-hand side of the equation, $p(data|\theta)$ is the likelihood, which is the probability of generating the data if the parameters in the model were equal to θ. The term next to the likelihood in the numerator is $p(\theta)$, the prior distribution of θ. It is a probability distribution which expresses the beliefs about the model parameters before the data are considered. The term $p(data)$ in the denominator is the probability of obtaining the data if a particular model and prior distribution are assumed. Finally, $p(\theta|data)$ on the left-hand side of the equation is the posterior probability distribution, which is the probability of obtaining the model parameters given the data. The posterior distribution is used for predictions and model testing [84].

1.6.3 MCMC

Markov chain Monte Carlo (MCMC) methods consist of algorithms that are used to estimate parameters of complex models by sampling from a probability distribution. Several different MCMC algorithms have been developed, the most popular of which include Gibbs, Metropolis-Hastings, and Hamiltonian [84]. Although MCMC is most often associated with Bayesian statistics, it can also be used in classical statistics [86]. "Monte Carlo," as in the famed Monegasque casino [84], refers to the randomness of the algorithms. Markov chains are named after Russian mathematician Andrey Markov, who researched stochastic processes in the late 1800s [84].

The role of the Markov chain in MCMC is to generate the random samples [86]. It does this by exploring all of the possible parameter values via a directed random walk through the parameter space. The random walk is "directed" because some values are more likely to be chosen than others [86]. Parameter values are sampled in a way that is proportional to their probability, which is contingent on the data and, if the analysis is Bayesian, the prior distribution. If there is more statistical support for a proposed parameter, it will be "accepted." Otherwise, it will be "rejected" and the current value in the chain will be carried through to the next iteration. The posterior distributions of the parameters are the end result of this process [86].

To initiate an MCMC, the user must specify (i) a starting point for each of the parameters, (ii) the length of the Markov chains (number of iterations), (iii) proposal distributions for the parameters, and (iv) standard deviations for the proposal distributions (the size of the jump between the current and proposed parameter values).

The starting points are usually chosen at random. Consequently, they should not be used to summarize the posterior distribution and are discarded, along with the samples at the beginning of the Markov chain. This period is called the "burn-in" and occurs during the initial phases of parameter space exploration. Both the number of iterations and the burn-in are often determined through trial and error as there are no fixed rules [86].

The distribution for the likelihood function defines the parameters in the model and should reflect the model assumptions. Some typical likelihood distributions include Bernoulli, binomial, Poisson, negative-binomial, and gamma [84].

In a similar way, prior distributions describe how the parameters behave in the likelihood function. Whether the parameters are constrained or unconstrained must be considered when choosing a prior distribution. An unconstrained parameter can be any real number, whereas a constrained parameter may be limited to only non-negative values or between certain bounds, such as a proportion [84].

Examples of typical prior distributions include uniform, normal, Student-*t*, Cauchy, and gamma. A uniform distribution is considered uninformative because it assumes all possible values of the parameter are equally likely. This means that the shape of the posterior distribution is completely determined by the likelihood [84].

After performing MCMC, the user needs to assess model convergence, which is achieved when the posterior distribution converges at its final distribution. To this end, the following visual diagnostics are routinely evaluated: trace plots, correlation plots, and density plots.

Trace plots show the magnitude of the posterior samples on the y-axis for each iteration of the MCMC procedure (x-axis). When the chain finds the stationary distribution of samples, subsequent samples will seem to be randomly drawn from around the same height of the y-axis [86]. The result looks like a fuzzy caterpillar.

There are two types of correlation plots. Autocorrelation plots show dependence in the chain of samples (when the current value in the chain is correlated with the previous value). Lower autocorrelation is preferred because it means that the Markov chain is closer to generating independent samples [84]. The second type of correlation plot shows the correlation between pairs of parameter values. Ideally, the points on these scatterplots should be randomly distributed, and there should not be high correlation or any strange patterns.

Density plots show the distribution of the sampled parameter values. They resemble smoothed histograms. As with correlation plots, strange shapes can be a sign of problems with model convergence. Density plots can also help visualize the amount of uncertainty around the parameter values; narrow plots are indicative of less uncertainty, whereas wide plots represent greater uncertainty.

Another typical check for model convergence involves running multiple chains from different starting points [86]. If the chains do not arrive at the same posterior distribution, then the model did not converge.

The acceptance rate, which is the number of times that a parameter is accepted over the total number of iterations, should also be calculated for each parameter. Numerical studies have shown that asymptotic acceptance rates of about 0.44 and 0.23 lead to optimal convergence for one-dimensional and multi-dimensional models, respectively [87, 88]. The target acceptance rate for a parameter can be achieved by tinkering with the standard deviation of the proposal distribution.

Once model diagnostics are completed, the fit of the model to the data should be evaluated. The process of checking model fit is also known as posterior predictive checks in Bayesian statistics [84]. Posterior predictive checks involve using the posterior distribution to generate samples from the posterior predictive distribution, which is defined as the probability distribution over possible values of future data. The posterior predictive distribution is used to generate simulated data samples, which are compared to the observed data. Graphical visualizations are typically used for the comparisons. If key aspects of the observed data are captured by the simulated data, then the model is a good fit [84].

In addition to posterior predictive checks, there are several ways to choose between different Bayesian models, including the Akaike information criterion (AIC), deviance information criterion (DIC), Widely Applicable Information Criterion, and leave-one-out cross-validation [84].

Finally, all models rely on assumptions which should be checked through sensitivity analyses. Sensitivity analyses evaluate whether different assumptions change a model's conclusions in meaningful ways. They are particularly important when data are few or there is uncertainty about the choice of model [84]. Common methods of conducting sensitivity analyses include repeating the analysis with different prior distributions and considering different classes of likelihoods.

2 Motivation

Colombia was one of the countries most affected by the ZIKV and CHIKV epidemics in the Americas [22, 48]. Between June 2014 and July 2016, Colombia reported 412,915 suspected and confirmed CF cases (INS data). This number included a total of 85 deaths by the end of 2017 [48], but deaths were likely underreported. Although excess mortality due to CHIKV has not yet been estimated for Colombia [89], studies in Puerto Rico and northeastern Brazil estimated excess deaths as 42 and 60 times greater, respectively, than deaths identified through official surveillance systems during the recent epidemics [90, 91].

There is increasing evidence of long-term health effects following infection with CHIKV. One study found about one-fourth of 485 patients with serologically confirmed CHIKV infection and joint pain in Colombia continued to experience joint pain after 20 months of follow-up [92]. Medical care, loss of productivity, and absenteeism from work and school contributed to substantial economic costs. One study estimated the cost per CF case in Colombia at $152.90 (interquartile range $101.00 - $539.60) with higher costs for pediatric patients compared to adults. The total cost of the CHIKV epidemic in Colombia was estimated at about $67 million [93]. Notably, impacts on tourism were not factored into the analysis, and therefore the actual cost of the epidemic could have been much higher.

From August 2015 to June 2017, Colombia reported 106,033 suspected and confirmed ZVD cases (INS data). By early 2018, 248 cases of confirmed congenital syndrome associated with

ZIKV infection had been reported to the Pan American Health Organization (PAHO) [22]. In addition, 418 cases of neurological complications among suspected or confirmed ZVD cases had been identified (INS data).

The costs associated with these complications are staggering. A United Nations Development Program report estimated that the lifetime cost per case of ZIKV-related microcephaly in Colombia was $690,000, and the lifetime cost per case of GBS was estimated at $176,000 [94]. Estimated short-term costs associated with diagnosing and treating patients, lost productivity, declines in tourism, and annualized costs of microcephaly and GBS ranged from about $456 million in the baseline scenario to about $1.4 billion in the high (worst-case) scenario for three years of the epidemic in Colombia [94].

3 Data and ethics

This thesis relies on four datasets from Colombia's INS which were shared through a Memorandum of Understanding with Imperial College London. The technical and ethical endorsement of the study was provided by the Comité de Ética y de Metodologías de Investigación of the INS (project number 35-2017).

3.1 ZIKV, CHIKV, and DENV surveillance datasets

Three of the datasets include anonymized line lists on ZVD, CF, and dengue fever (DF) suspected and laboratory-confirmed cases reported to Sivigila between 2010 and 2017. Event codes included 895 for suspected or confirmed ZVD, 725 for neurological syndrome probably associated with ZIKV, 549 for extreme maternal morbidity associated with ZIKV, 910 for suspected or confirmed CF, 210 for DF, 220 for severe dengue, and 580 for death due to DENV. Suspected cases included individuals who were reported to Sivigila with symptoms of ZVD, CF, or DF. Laboratory-confirmed cases were patients who had clinical symptoms and tested positive for ZIKV, CHIKV, or DENV on RT-PCR assay [95]. Information related to public health events in Colombia is generated from the local levels by health service providers. There are about 14,000 institutional, municipal, departmental, or national reporting bodies in the country. In 2014, CF was added to the list of notifiable conditions,

and ZVD was added the following year. Each week these data are aggregated and published [95].

The location used in this thesis corresponds to the location of likely infection, which was decided by the clinician who reported the case. The location of likely infection is preferred over residence because it accounts for human movements to areas with higher risk of arbovirus transmission. Information on 106,033 ZVD, 412,915 CF, and 647,665 DF cases was available at administrative level 1 (departments) after removing cases that had Bogotá recorded as the location of likely infection (extremely unlikely, due to high altitude and low temperature) or missing.

Data were further aggregated by week based on either date of symptom onset or date of notification. There were 365 weeks of DENV data, from the week ending January 3, 2010 to that ending December 25, 2016. For CHIKV, data encompassed a 110-week interval, from the week ending June 7, 2014 to that ending July 9, 2016, and for ZIKV, a 97-week interval, from the week ending August 15, 2015 to that ending June 17, 2017. Figure 1.8 shows the epidemiological curves for ZVD, CF, and DF at the country level.

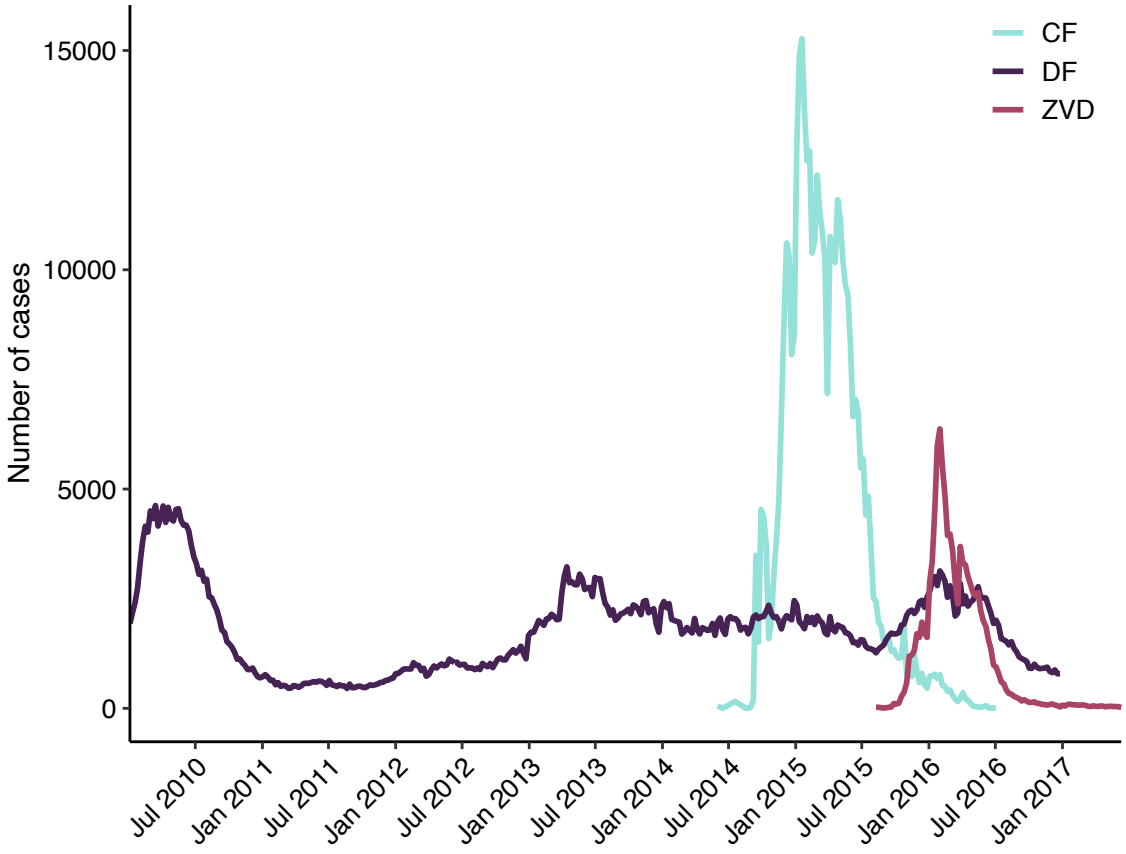


Figure 1.8 Weekly reported cases of CF, DF, and ZVD in Colombia, January 2010-June 2017.

3.2 ZIKV-associated neurological complications dataset

The fourth dataset consists of an anonymized line list on 418 patients with neurological complications and recent history of febrile illness compatible with ZVD. It includes those with GBS as well as similar conditions such as myelitis and meningoencephalitis but excludes cases with microcephaly and other congenital defects. Four cases in this dataset were laboratory-confirmed for ZIKV infection by RT-PCR. Although CHIKV has also been associated with neurological complications [34], a comparable dataset for patients with neurological complications and history of CF from Colombia was not available.

The number of cases with neurological complications here is a few hundred smaller than previously published data from Colombia [96, 97]. Throughout the epidemic, cases of neurological complications associated with previous ZIKV infection were reported in the INS Weekly Epidemiological Bulletin. This information was made publicly available with the caveat that cumulative case numbers could change following a verification process [97].

Medical records of patients with neurological complications were reviewed using case definitions from the Brighton Collaboration Working Group for GBS, myelitis, encephalitis, and acute disseminated encephalomyelitis [98, 99]. With the goal of improving comparability of vaccine safety data, the Brighton Collaboration developed standard case definitions and guidelines for neurologic adverse events following immunization. The criteria could be applied to a range of settings, including across geographical regions as well as different levels of healthcare quality and access. Case definitions are organized according to three levels of diagnostic certainty, from Level 1 (most certain) to Level 3 (least certain) [100]. Patients that did not meet Brighton case definition criteria 1-3 were removed from the dataset.

Both date of symptom onset of neurological complications and date of notification were available for all cases in the neurological complications dataset. Dates corresponding to symptom onset spanned 122 weeks, from the week ending July 4, 2015 to that ending October 28, 2017 (epidemiological week 26 of 2015 to epidemiological week 43 of 2017). Notification dates spanned 108 weeks, from the week ending October 17, 2015 to that ending November 4, 2017 (epidemiological week 41 of 2015 to epidemiological week 44 of 2017). Data were aggregated by week.

4 Objectives

The aim of this thesis is to improve understanding of the ZIKV and CHIKV epidemics in Colombia using surveillance data. This thesis is divided into six chapters which are ordered by increasing methodological complexity and spatial resolution. Following this introductory chapter,

1. Chapter 2: biases in surveillance of ZVD were exposed using line list data and a dataset on ZIKV-associated neurological complications.

2. Chapter 3: reporting rates and reproduction numbers of ZIKV and CHIKV were analyzed.

3. Chapter 4: a Bayesian hierarchical model was used to estimate ZIKV attack rates and risk of ZIKV-associated neurological complications in capital cities as well as quantify biases in the ZIKV surveillance data.

4. Chapter 5: a suite of spatial interaction models was used to study the spatial and temporal invasion dynamics of the ZIKV and CHIKV epidemics.
5. Chapter 6: a discussion chapter summarizing the key findings of the thesis and directions for future research.

Chapter 2: Descriptive analysis of surveillance data for Zika virus disease and Zika virus-associated neurological complications in Colombia

Work in this chapter formed the basis of a manuscript that has been published in *PLOS ONE* [101].

Abstract

In this chapter, a descriptive analysis was performed on approximately 106,000 suspected and laboratory-confirmed cases of ZVD that were reported during the 2015-2017 epidemic in Colombia. A dataset containing patients with neurological complications and recent febrile illness compatible with ZVD was also analyzed. Females had higher observed attack rates of ZVD than males. Compared to the general population, cases were more likely to be reported in young adults (20 to 39 years of age). The observed attack rate of ZVD in pregnant females was estimated at 3,120 reported cases per 100,000 population (95% CI: 3,077-3,164), which was considerably higher than the observed attack rate in both males and non-pregnant females. ZVD cases were reported in all 32 departments. Four-hundred and eighteen patients suffered from ZIKV-associated neurological complications, of which 85% were diagnosed with GBS. The median age of ZVD cases with neurological complications was 12 years older than that of ZVD cases without neurological complications. ZIKV-associated neurological complications increased with age, and the highest observed attack rate was reported among individuals aged 75 and older. Even though neurological complications and deaths due to ZIKV were rare in this epidemic, better risk communication is needed for people living in or traveling to ZIKV-affected areas.

1 Introduction

1.1 Background

From 1952, when cases of ZVD were first reported in humans, until about the last decade, ZIKV was thought to cause only mild illness [3]. Following major epidemics in Micronesia, French Polynesia, the Caribbean, and Latin America in the 2010s, it became clear that a small proportion of individuals infected with ZIKV experience serious disease.

GBS is an autoimmune condition with a global annual incidence estimated at 1.1 to 1.8 cases per 100,000 population [102]. It is typically preceded by a viral or bacterial infection, especially *Campylobacter jejuni* and is the most common cause of non-poliovirus acute flaccid paralysis globally [103]. On rare occasions, GBS has also been associated with certain vaccines, including vaccines for rabies, tetanus, and influenza [103]. In 2013-2014, an unusual number of GBS cases were detected in French Polynesia during the largest documented ZIKV outbreak at that time [104]. Since then, evidence of an association between ZIKV infection and GBS has continued to increase [105, 106].

Symptoms of GBS include tingling, numbness, or pain in the limbs as well as limb weakness. Most patients with GBS require hospitalization and some require intensive care and ventilatory support [107]. Between 3-10% of GBS patients die [108]. Although most patients fully recover, some may experience long-term morbidity, including depression and disability [109, 110]. Treatment for acute GBS involves the administration of intravenous immunoglobulin and plasma exchange [110].

Research suggests that the risk of GBS tends to be higher for males than females and increases with age [103]. According to a meta-analysis that used population-based studies of GBS in North America and Europe, the risk ratio (RR) for males versus females was estimated at 1.78 (95% CI: 1.36-2.33). The study also found that GBS incidence increased 20% for each 10-year increase in age [103].

There is evidence of seasonal variation in GBS with most published studies indicating higher incidence during winter (January to March) compared to the other three seasons [111]. However, heterogeneity between regions has been noted. According to a meta-analysis, greater incidence in winter was found for Western countries (incidence rate ratio, IRR=1.28, 95% CI: 1.11-1.48), the Far East (IRR=1.20, 95% CI: 1.00-1.44), and the Middle East (IRR=1.12, 95% CI: 0.89-1.42), while lower incidence was found for the Indian subcontinent (IRR=0.86, 95% CI: 0.66-1.13) and Latin America (IRR=0.75, 95% CI: 0.46-1.24)[1] [111]. This result could be due to regional differences in the seasonality of infections that trigger GBS.

[1]Season was defined by the reporting study, or where monthly data were reported, winter: January-March, spring: April-June, summer: July-September, autumn: October-December. Seasons were inverted for countries in the southern hemisphere.

In addition to GBS, ZIKV infection during pregnancy has been associated with CZS in fetuses and newborns. CZS is characterized by microcephaly, decreased brain tissue, eye damage, limited range of motion in the joints, and excessive muscle tone that restricts movement [19]. Most newborns with prenatal exposure to ZIKV do not develop clinical signs of CZS [112]. However, cohort studies have shown that children without birth defects who were exposed to ZIKV in utero can still experience neurological problems and developmental delays during the first two years of life [113, 114]. Infants and children can also become infected with ZIKV during the postnatal period through mosquito bites and possibly breast milk; however, few studies have evaluated postnatal ZIKV infection prospectively [115].

In Colombia, surveillance for ZVD began in August 2015. By December 2015, GBS cases and other neuroinflammatory disorders began to rise in the country [106]. From the end of January to mid-November 2016, the number of reported microcephaly cases in Colombia increased fourfold compared to the same time period in 2015 [116].

1.2 Aims

The aim of this chapter is to describe epidemiological trends of ZVD and ZIKV-associated neurological complications in Colombia. Sex, age, temporal, and geographic trends among reported ZVD cases were investigated. Observed attack rates, RRs, and tests for statistical significance were estimated for high-risk groups.

Understanding risk factors for neurological complications could inform prevention efforts and improve interpretation of ZIKV surveillance data.

2 Data

2.1 Epidemiological data

Two main datasets were used in this chapter which were both described in chapter 1: the ZIKV line list from Sivigila and the dataset on ZIKV-associated neurological complications. As mentioned previously, dates for the ZIKV line list ranged from the week ending August 15, 2015 to that ending June 17, 2017, which correspond to epidemiological week 32 of 2015 and epidemiological week 24 of 2017. Similarly, for the neurological complications dataset, dates corresponding to symptom onset spanned 122 weeks, from the week ending July 4, 2015 to that ending October 28, 2017 (epidemiological week 26 of 2015 to epidemiological

week 43 of 2017), and notification dates spanned 108 weeks, from the week ending October 17, 2015 to that ending November 4, 2017 (epidemiological week 41 of 2015 to epidemiological week 44 of 2017). The dataset for CF did not have detailed information about both sex and age group of cases, and as mentioned in chapter 1, a dataset on CHIKV-associated neurological complications was not available. Therefore, ZIKV is the main focus of this chapter.

2.2 Demographic data

Population projections derived from the 2005 Census were obtained for 2016 from DANE, Colombia's National Administrative Department of Statistics.

3 Methods

Observed attack rates rather than infection attack rates were estimated in this chapter. The observed attack rate is the number of reported cases divided by the population. In contrast, the infection attack rate is the number of infections divided by the population. Observed attack rates of ZVD were estimated using DANE population projections for 2016 as the denominator. Unless otherwise noted, the observed attack rates of ZIKV-associated neurological complications were estimated using reported cases of ZVD as the denominator. For observed attack rates by geographic location, the total population was included for each reporting area regardless of altitude (except for the capital city of Bogotá, whose population was excluded from the department of Cundinamarca). Confidence intervals were calculated using the binomial exact function in the R package epitools (version 0.5-10.1). RRs and 95% confidence intervals were calculated using the riskratio function in the R package fmsb (version 0.7.0).

In order to estimate the observed attack rate of ZVD by pregnancy status, it was necessary to estimate the number of pregnant females and non-pregnant females in Colombia. In 2017, the Colombian Ministry of Health and Social Protection (MOH) estimated the annual number of pregnant females for 2017-2019. The mean estimate of 822,396 for 2017 was multiplied by ¾ (females are only pregnant for ¾ of the year on average) to obtain the number of pregnant females in the population at any time (616,797). To obtain the number

of females who were not pregnant at any time, this number was subtracted from the projected total number of females in the population in 2016 (24,678,673, from DANE).

4 Results

4.1 ZVD

4.1.1 Sex and age trends

More ZVD cases were reported in females than males each year from 2015-2017 (Table 2.1). Overall, females represented two-thirds (66.2%) of reported cases, which differs significantly from the general population of Colombia (50.6% female [exact binomial test, $p < 0.001$]). The RR of ZVD in females was nearly two times higher than in males (1.91, 95% CI: 1.88-1.93).

The median age of ZVD cases in Colombia was 29 years (range 0 to >100 years). Nearly half (49.0%) of cases were reported in individuals between the ages of 20 and 39. Compared to the general population, the number of cases reported in this age range was significantly different than expected (31.2% [exact binomial test, $p < 0.001$]).

Figure 2.1 shows the age distribution of Colombia by sex. In contrast, the age distribution of ZVD cases for each sex is shown in Figure 2.2. The observed attack rate of ZVD was significantly higher for females compared to males across all age groups except 0-4 years and those 80 years and over (Figure 2.3). Among those 15 to 29 years of age, the risk of ZVD was about three times higher in females compared to males. The highest RR of female to male cases was observed in the 20 to 24-year age group (3.16, 95% CI: 3.04-3.28).

Table 2.1 Number of ZVD cases by sex in Colombia. Epidemiological week 32 of 2015 – epidemiological week 24 of 2017.

	Female		Male		Total	
	N	%	N	%	N	%
2015	8,940	63.7	5,085	36.3	14,025	13.2%
2016	60,494	66.7	30,153	33.3	90,647	85.5%
2017	738	54.2	623	45.8	1,361	1.3%
Total	70,172	66.2	35,861	33.8	106,033	100%

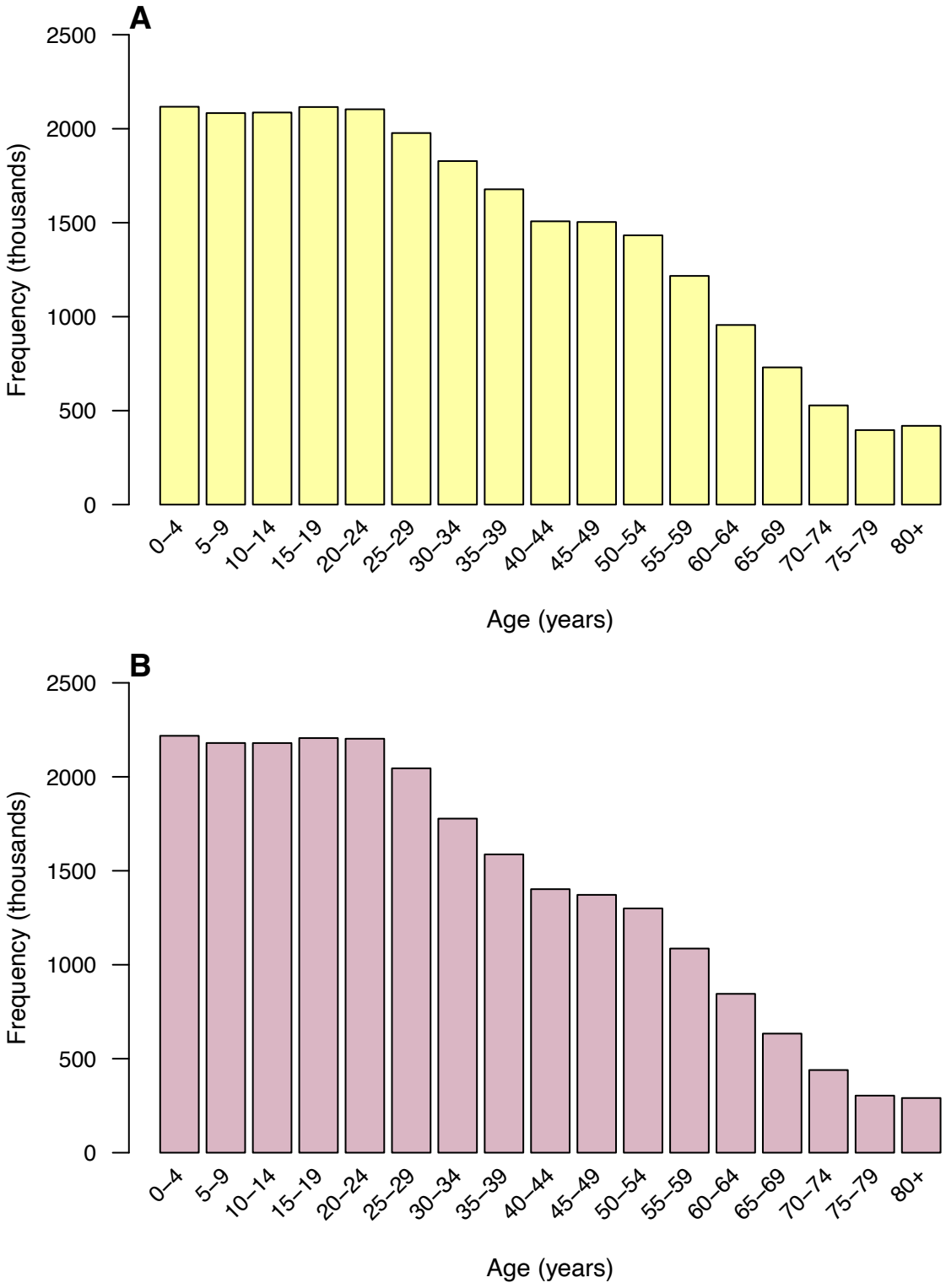


Figure 2.1 Population of Colombia by age group in 2016. *(A)* Females and *(B)* males.

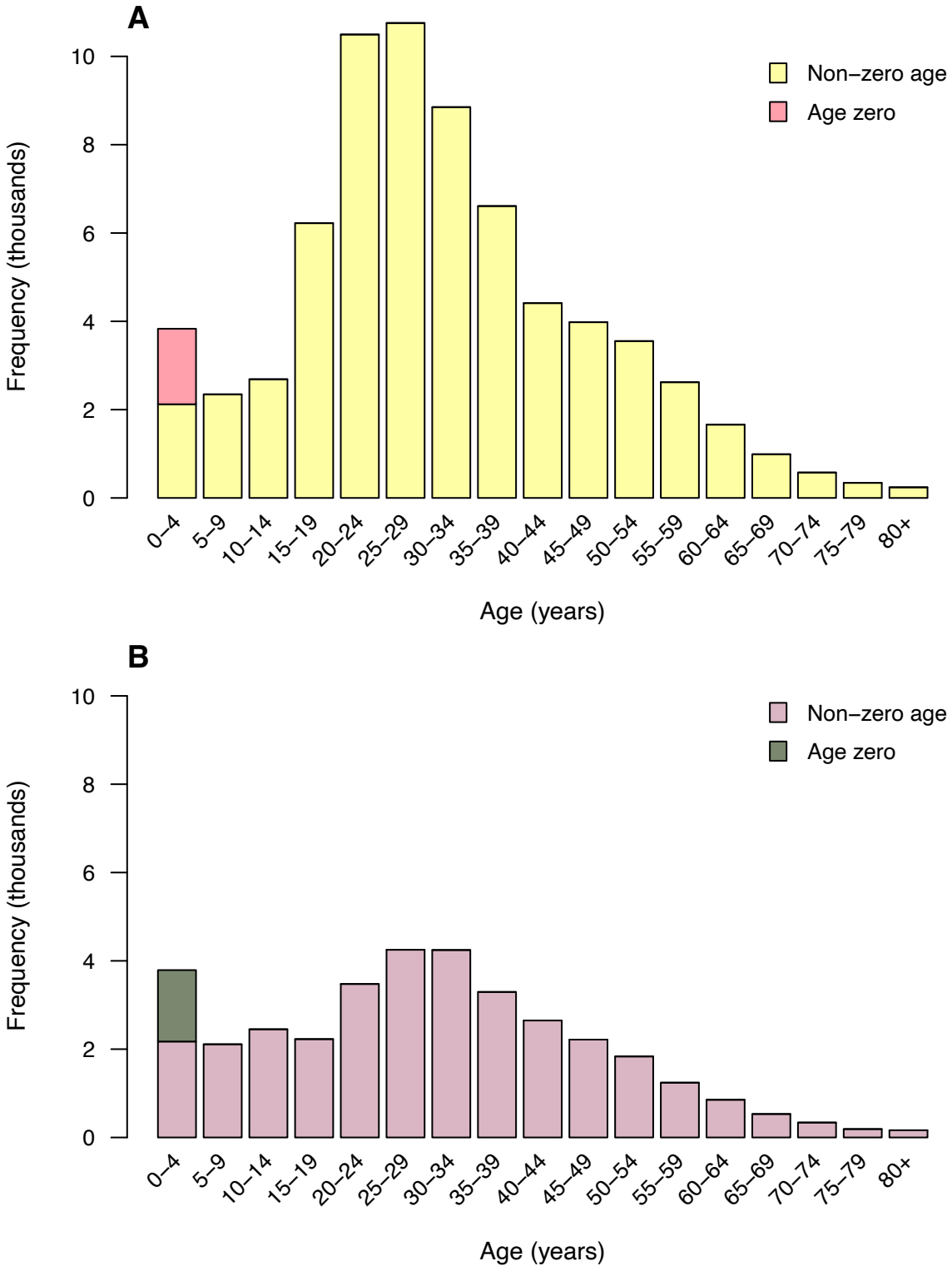


Figure 2.2 Frequency of ZVD cases by age group and sex in Colombia. *(A)* Females and *(B)* males. Epidemiological week 32 of 2015 – epidemiological week 24 of 2017.

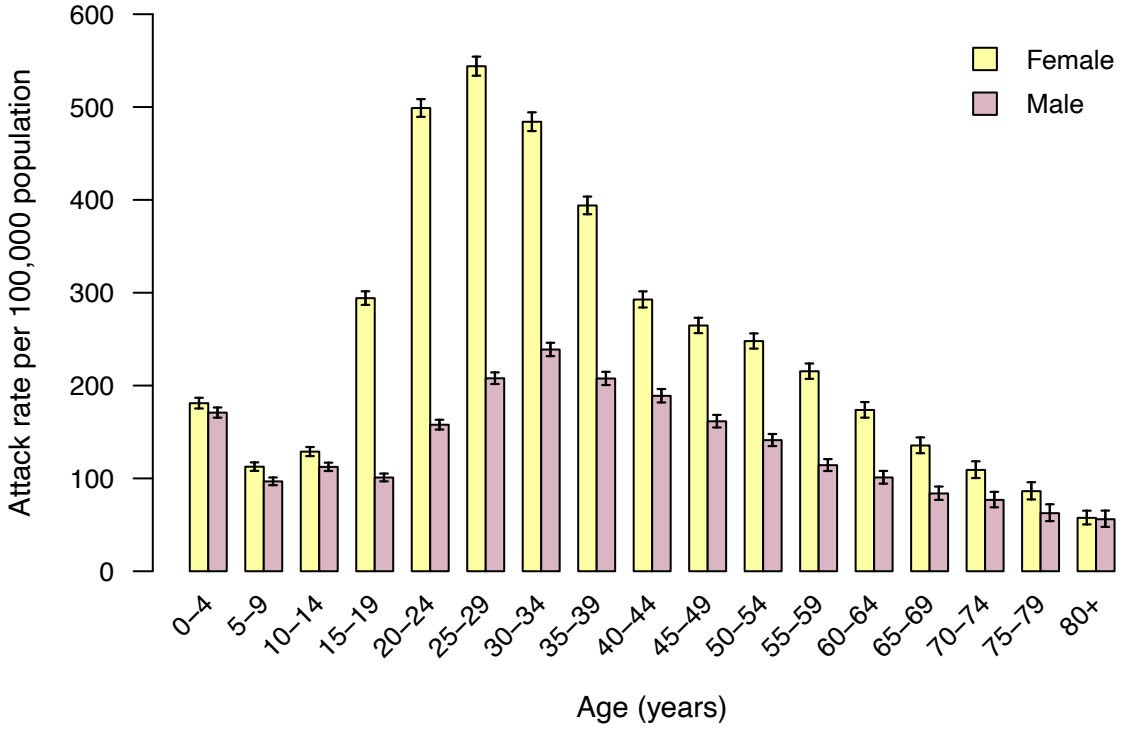


Figure 2.3 Observed attack rate of ZVD per 100,000 population by age group and sex in Colombia. Epidemiological week 32 of 2015 – epidemiological week 24 of 2017.

4.1.2 Pregnancy

A total of 19,243 ZVD cases was reported in pregnant females. Figure 2.4 shows the age distribution of female ZVD cases by pregnancy status. The discrepancy in males and females described in the previous section is largely explained by enhanced surveillance in pregnant females. The epidemiological curve of ZVD cases by sex and pregnancy status is shown in Figure 2.5. The timing of the peak of the epidemic was similar across groups. However, pregnant females had much higher risk compared to males and non-pregnant females. The observed attack rate of ZVD in pregnant females was 3,120 reported cases per 100,000 population (95% CI: 3,077-3,164). In contrast, the observed attack rate in non-pregnant females was 205 cases per 100,000 population (95% CI: 203-207), and the observed attack rate in males was 149 cases per 100,000 population (95% CI: 147-151). The differences in observed attack rates of ZVD by sex and pregnancy status were significantly different.

In Figure 2.5, Christmas/New Year and Easter holidays are indicated by vertical lines. For males and non-pregnant females, the number of reported cases decreased before New

Year's Day (epidemiological week 52 of 2015) and Easter (epidemiological week 12 of 2016) and then sharply increased. This trend is less noticeable for pregnant females.

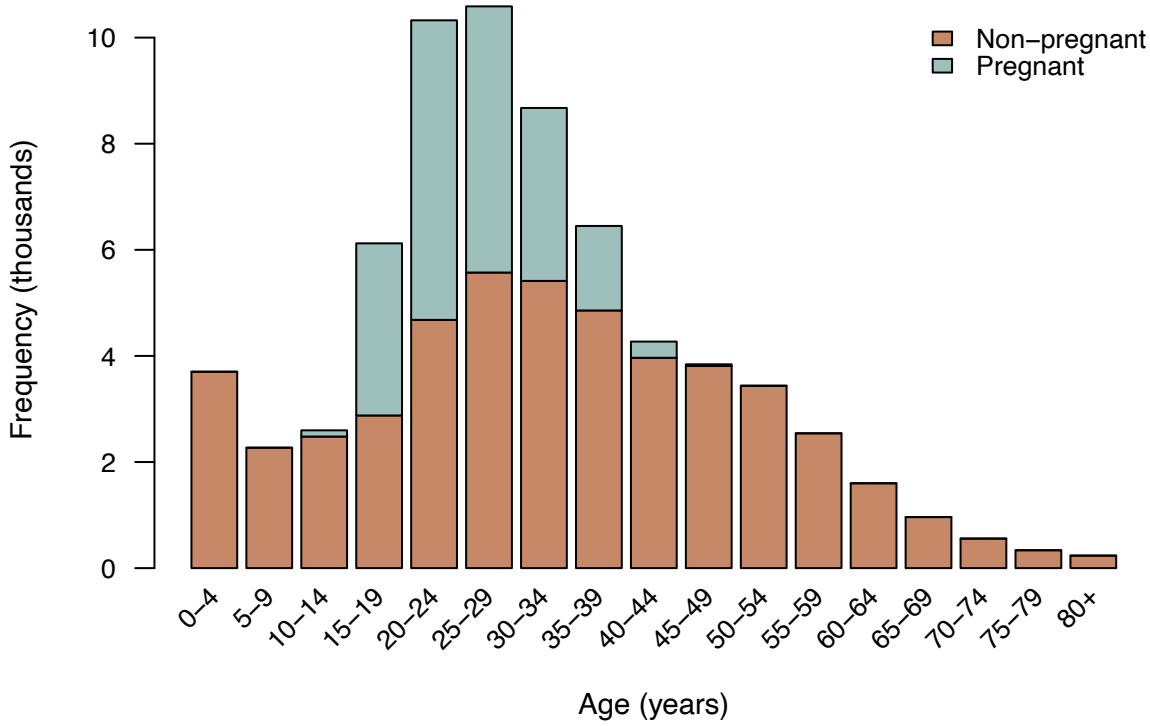


Figure 2.4 Frequency of female ZVD cases by age group and pregnancy status in Colombia. Epidemiological week 32 of 2015 – epidemiological week 24 of 2017.

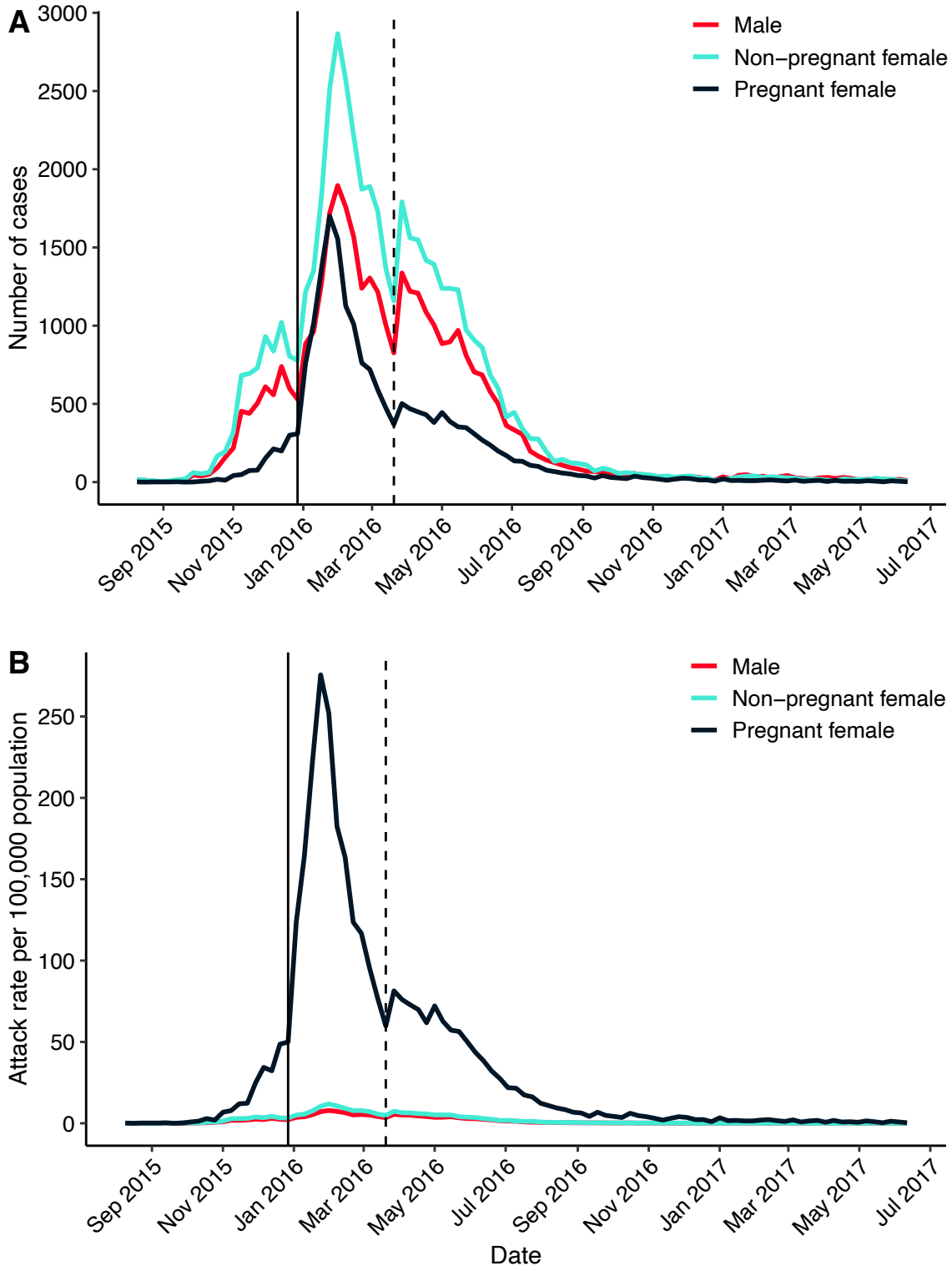


Figure 2.5 ZVD over time by sex and pregnancy status in Colombia. *(A)* Number of cases and *(B)* observed attack rate per 100,000 population. N = 104,397[2]. Dotted line marks the week preceding Easter 2016 (epidemiological week 12). Solid line marks the week preceding New Year's Day in 2016 (epidemiological week 52 of 2015). Epidemiological week 32 of 2015 – epidemiological week 24 of 2017.

[2] 1,636 females had missing pregnancy status.

4.1.3 Temporal trends

Epidemiological curves of ZVD cases are shown for all departments in Figures 2.6-2.7. Most departments had one epidemic peak between about January and April of 2016 but some appeared to have two (Bolívar and Cundinamarca). Chocó, Guainía, Nariño, Vaupés, and Vichada had irregular time series due to small numbers of reported cases.

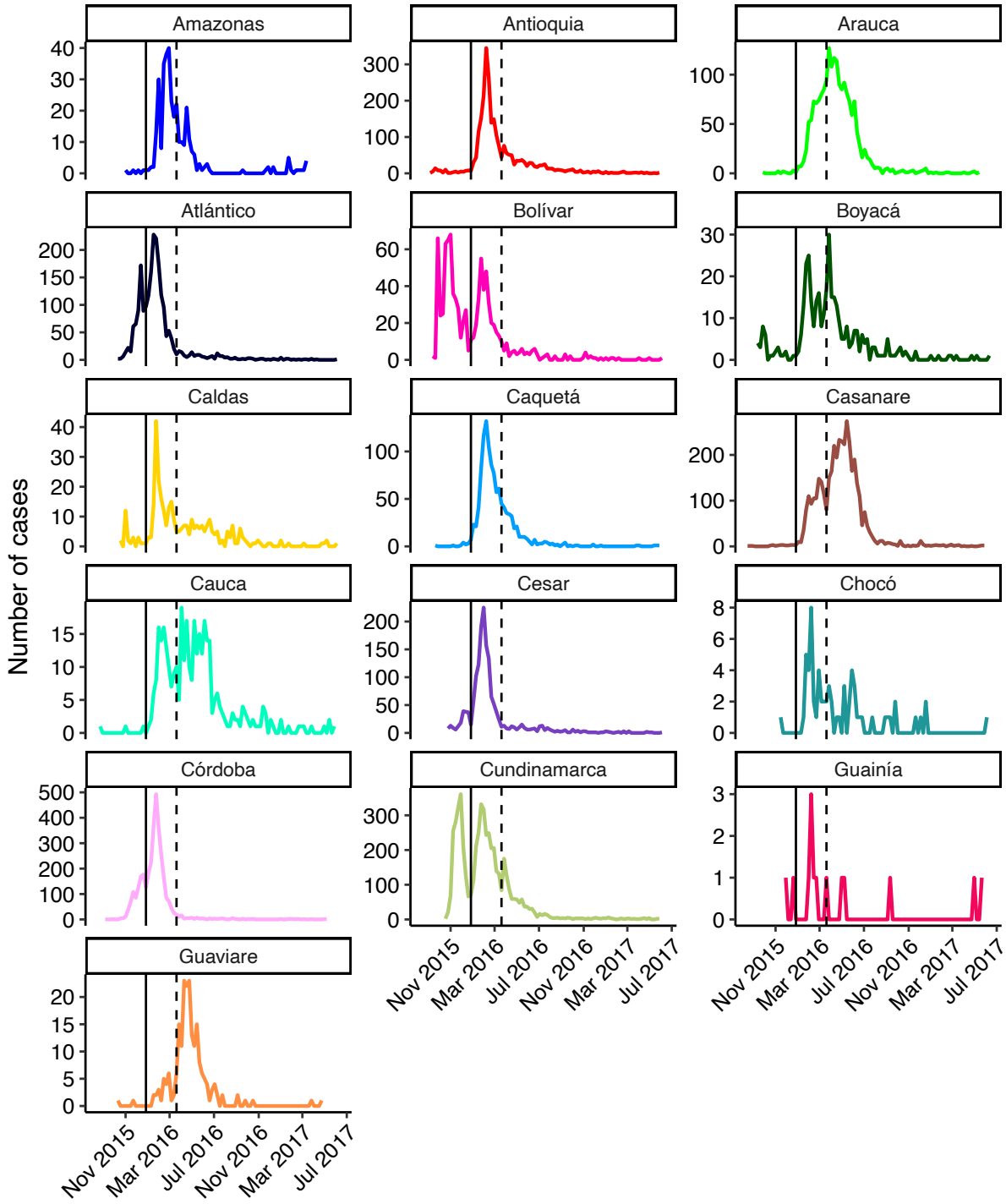


Figure 2.6 ZVD cases over time by department for 16 departments. The remainder are in Figure 2.7. Dotted lines mark the week preceding Easter 2016 (epidemiological week 12). Solid lines mark the week preceding New Year's Day in 2016 (epidemiological week 52). Y-axes are different, and x-axes are the same. Epidemiological week 32 of 2015 – epidemiological week 24 of 2017.

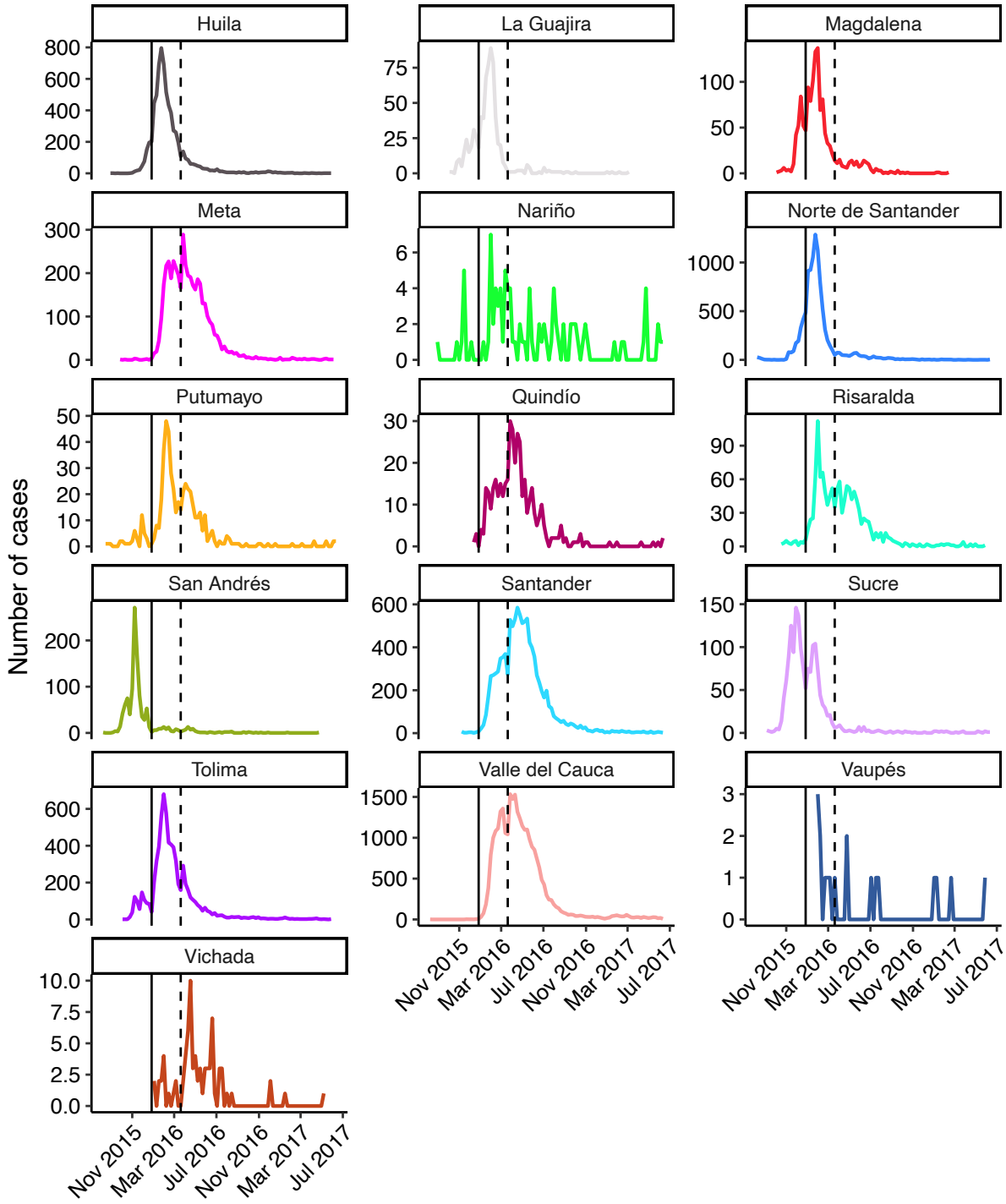


Figure 2.7 Number of ZVD cases over time by department for 16 departments (continued). The remainder are in Figure 2.6. Dotted lines mark the week preceding Easter 2016 (epidemiological week 12). Solid lines mark the week preceding New Year's Day in 2016 (epidemiological week 52). Y-axes are different, and x-axes are the same. Epidemiological week 32 of 2015 – epidemiological week 24 of 2017.

Figure 2.8 shows the epidemiological curves for ZVD and CF cases. Similar to the trend for ZVD, there is a decrease in reported CF cases in the week prior to both New Year's Day and Easter in 2015 followed by a sharp increase.

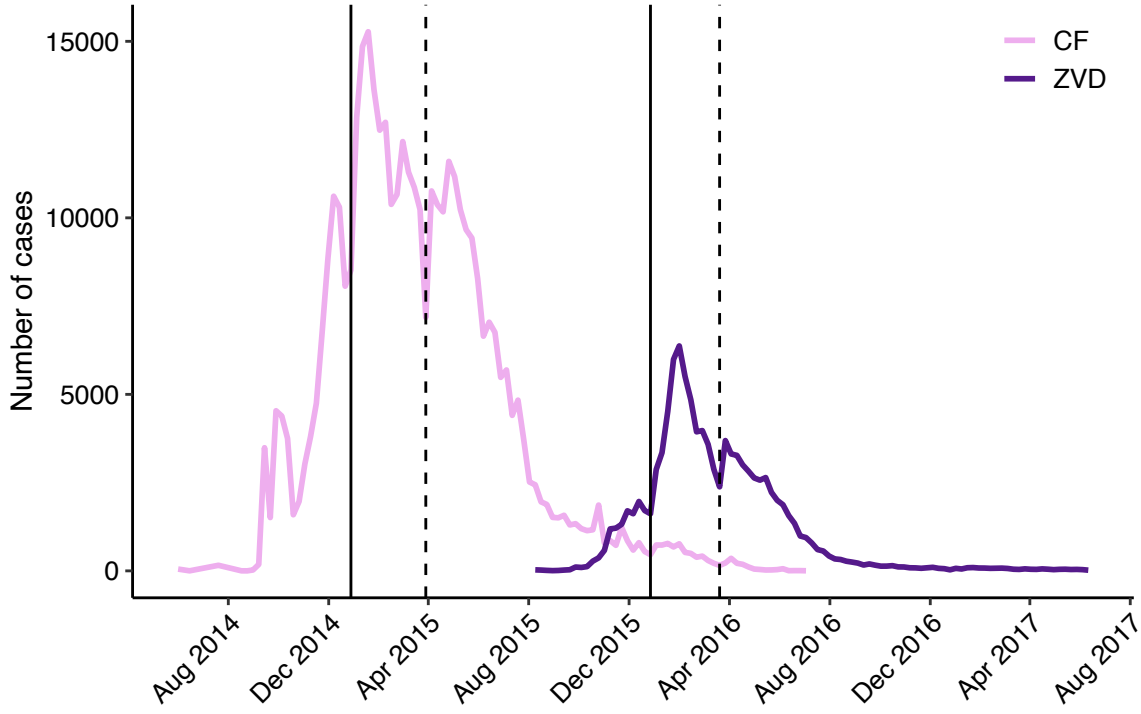


Figure 2.8 Number of cases of CF and ZVD over time in Colombia. Dotted lines mark the weeks preceding Easter in 2015 (epidemiological week 13) and 2016 (epidemiological week 12). Solid lines mark the weeks preceding New Year's Day in 2015 (epidemiological week 53 of 2014) and 2016 (epidemiological week 52 of 2015). Epidemiological week 23 of 2014 – epidemiological week 24 of 2017.

4.2 ZIKV-associated neurological complications

4.2.1 Sex and age trends

The median age of ZVD cases with neurological complications was 41 years (range 0-93). The highest number of neurological complications was reported in middle-aged groups from 35-49 (Figure 2.9), and the observed attack rate of neurological complications was highest in those at least 75 years of age with a high degree of uncertainty (Figure 2.10).

The distributions of cases of neurological complications by age and sex are shown in Figure 2.11. Two-hundred and forty-four cases (58%) were male. In the general population, the risk

of neurological complications was 44% higher in males than females (RR=1.44, 95% CI: 1.18-1.75). Among reported ZVD cases, the risk of neurological complications was 174% higher in males compared to females (RR=2.74, 95% CI: 2.26-3.33).

For both sexes, the greatest number of complications was reported in middle-aged groups. Observed attack rates of neurological complications per 1,000 cases of ZVD by age and sex are shown in Figure 2.12. For both sexes, observed attack rates of neurological complications increased with age. Males aged 75 and older had the highest observed attack rate but also the most uncertainty, which is reflected in the wide confidence intervals.

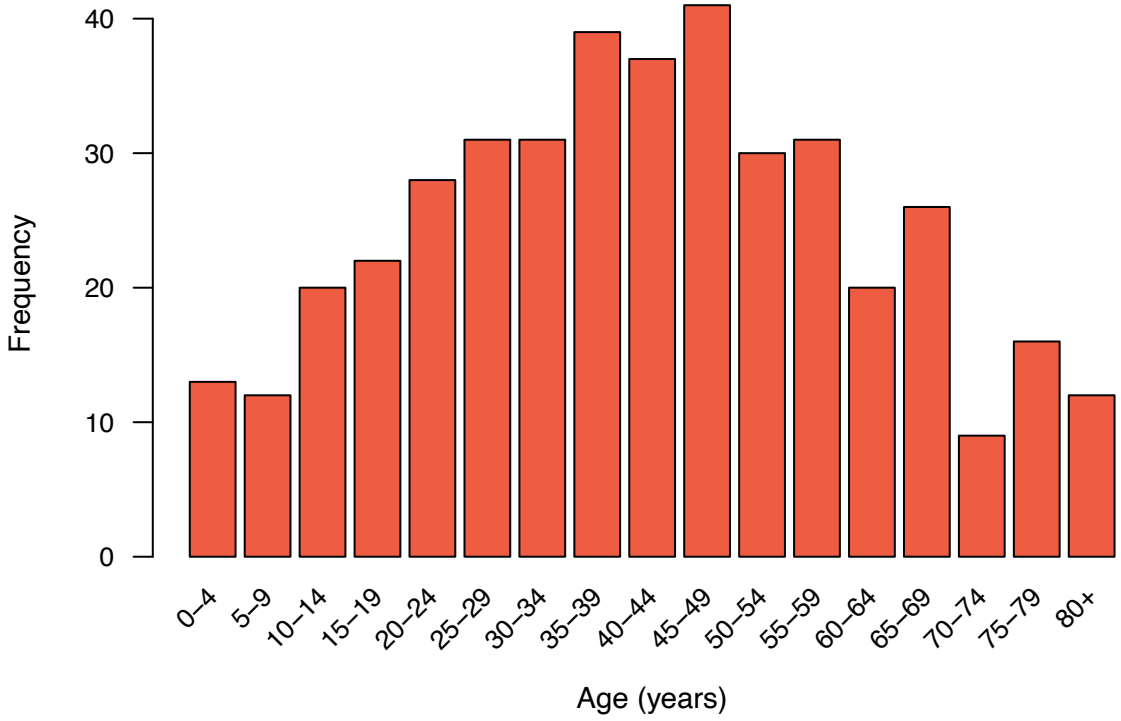


Figure 2.9 Frequency of ZIKV-associated neurological complications by age group in Colombia.

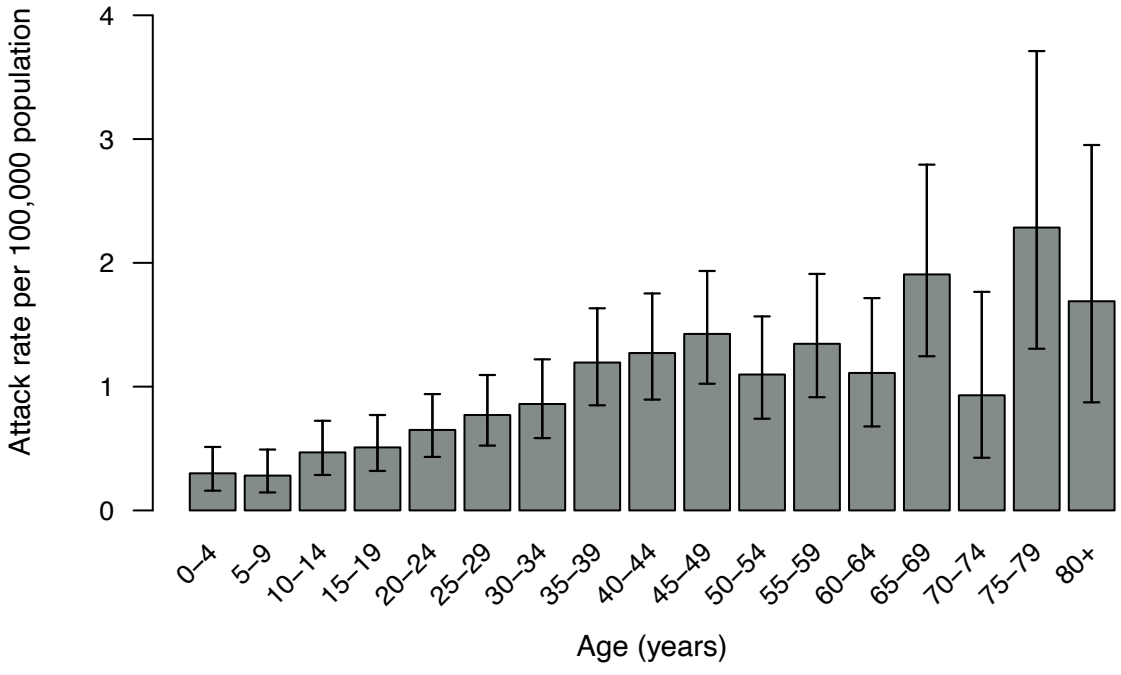


Figure 2.10 Observed attack rate of ZIKV-associated neurological complications per 100,000 population by age group in Colombia.

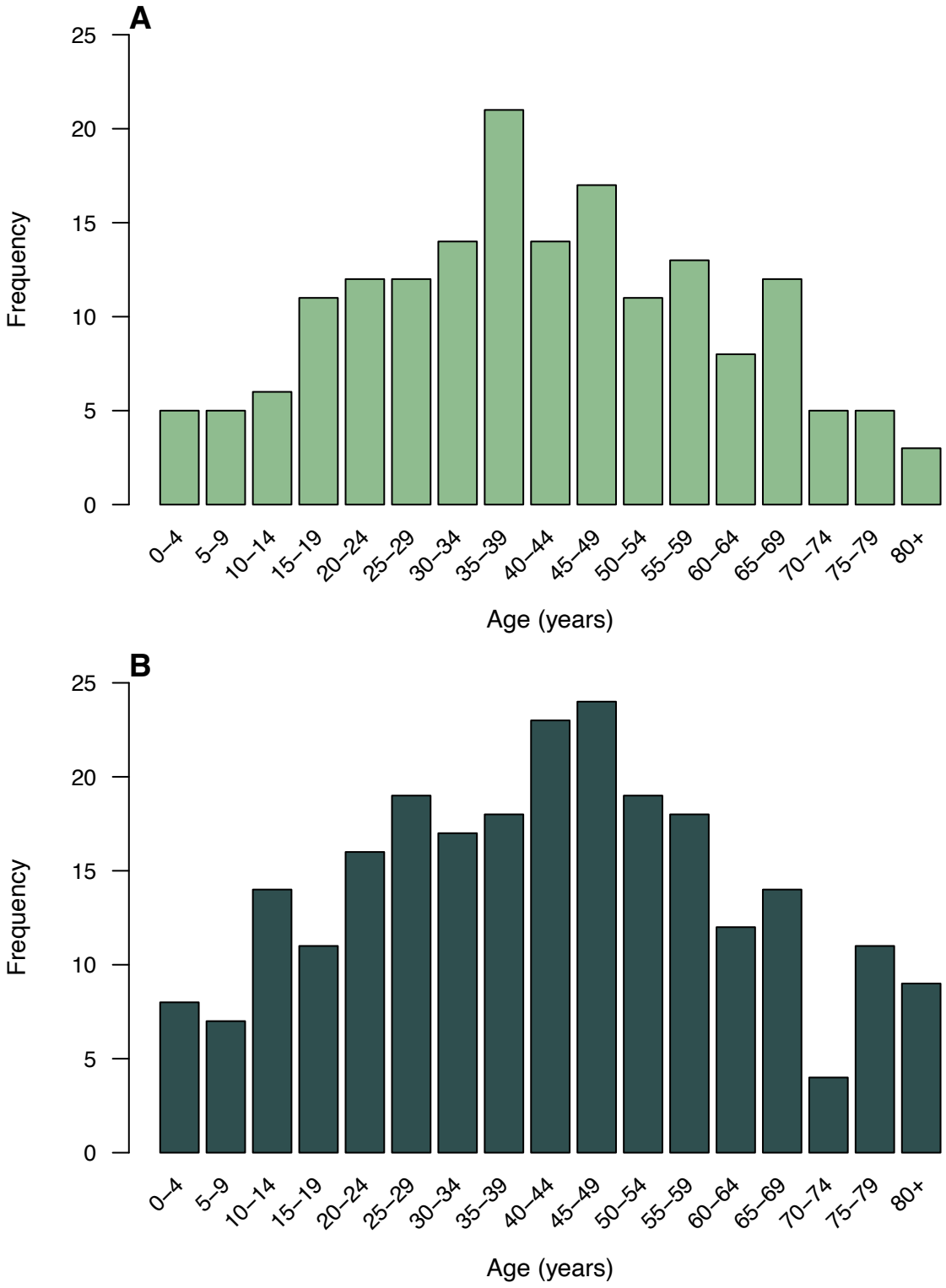


Figure 2.11 Frequency of ZIKV-associated neurological complications by age group in Colombia. *(A)* Females and *(B)* males.

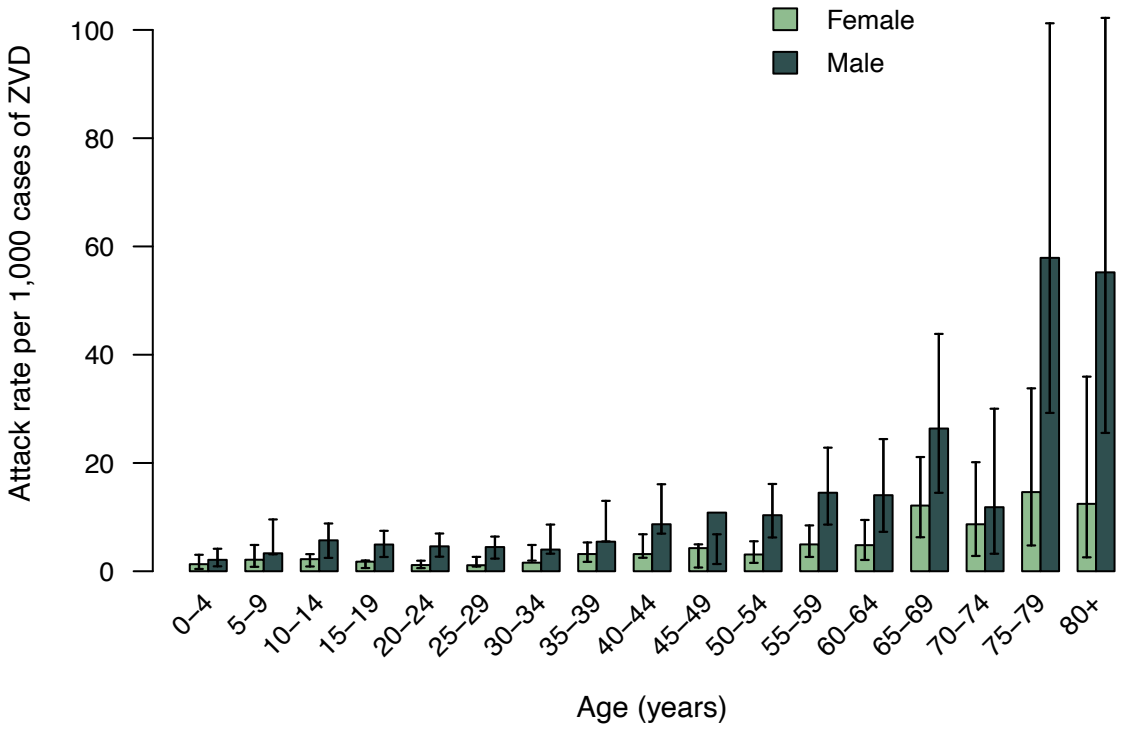


Figure 2.12 Observed attack rate of ZIKV-associated neurological complications per 1,000 cases of ZVD by age group and sex in Colombia.

4.2.2 Temporal trends

The distributions of reported ZVD cases and cases of ZIKV-associated neurological complications over time are similar (Figure 2.13). There are noticeable differences, however, in the time series of ZIKV-associated neurological complications when date of notification is considered rather than date of symptom onset. The week with the highest number of neurological complications according to date of symptom onset was the same week in which the most ZVD cases were reported (30 cases during the week ending on February 6, 2016). In contrast, the most neurological complications according to date of notification were reported two weeks earlier, in the week ending on January 23, 2016 (33 cases).

Similar to the ZIKV dataset, which did not have date of symptom onset for all cases, date of notification (reporting) seems to have decreased during Christmas/New Year and Easter holidays.

Although the distribution of neurological complications by sex did not vary across time, more cases were consistently reported in males than in females throughout the epidemic (Figure 2.14).

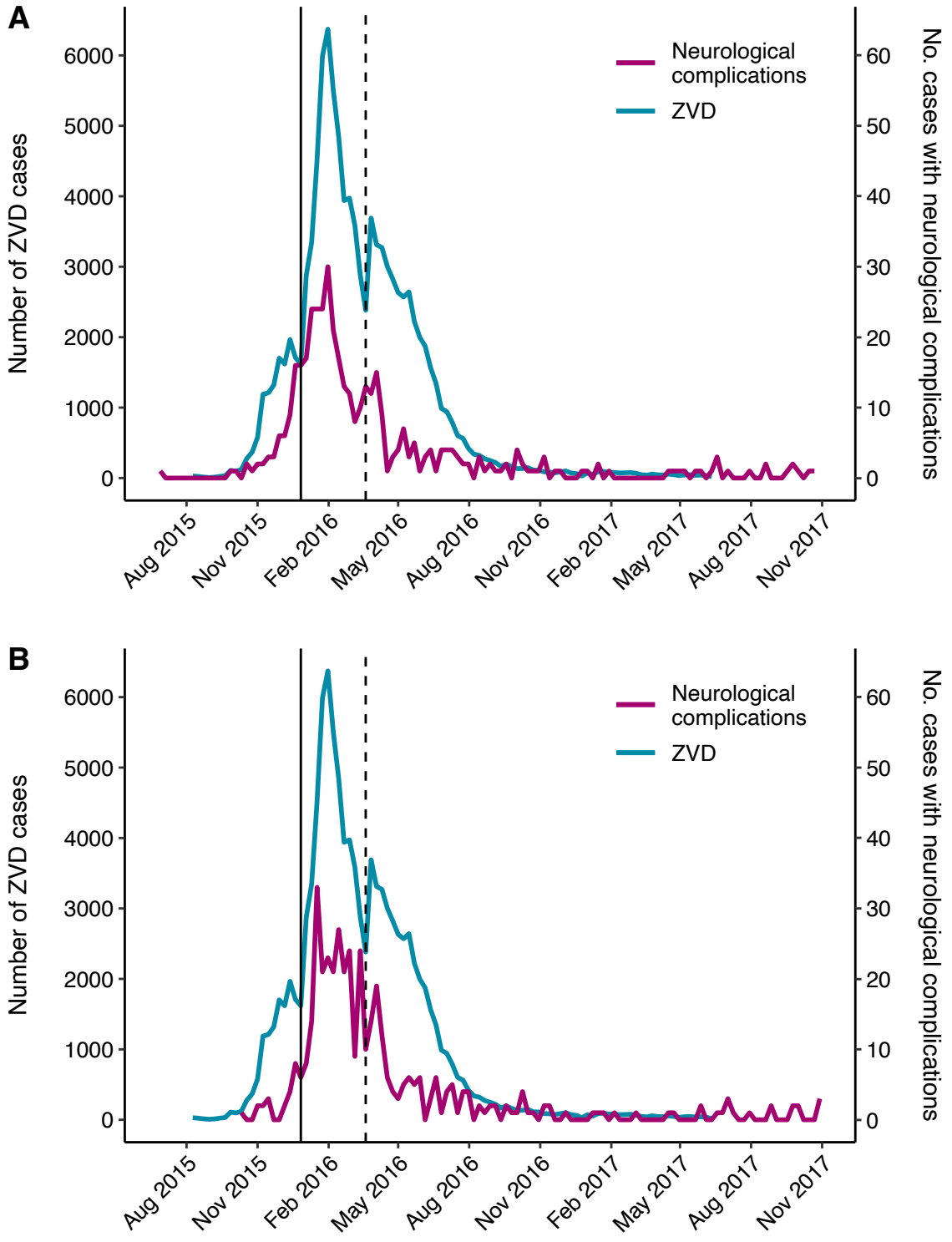


Figure 2.13 Number of cases of ZVD and ZIKV-associated neurological complications over time in Colombia. The dotted line marks the week preceding Easter 2016 (epidemiological week 12), and the solid line marks the week preceding New Year's Day in 2016 (epidemiological week 52 of 2015). For cases of ZIKV-associated neurological complications, *(A)* uses dates of symptom onset, whereas *(B)* uses dates of notification. The blue line is the same in both *(A)* and *(B)* as symptom onset was not available for all ZVD cases.

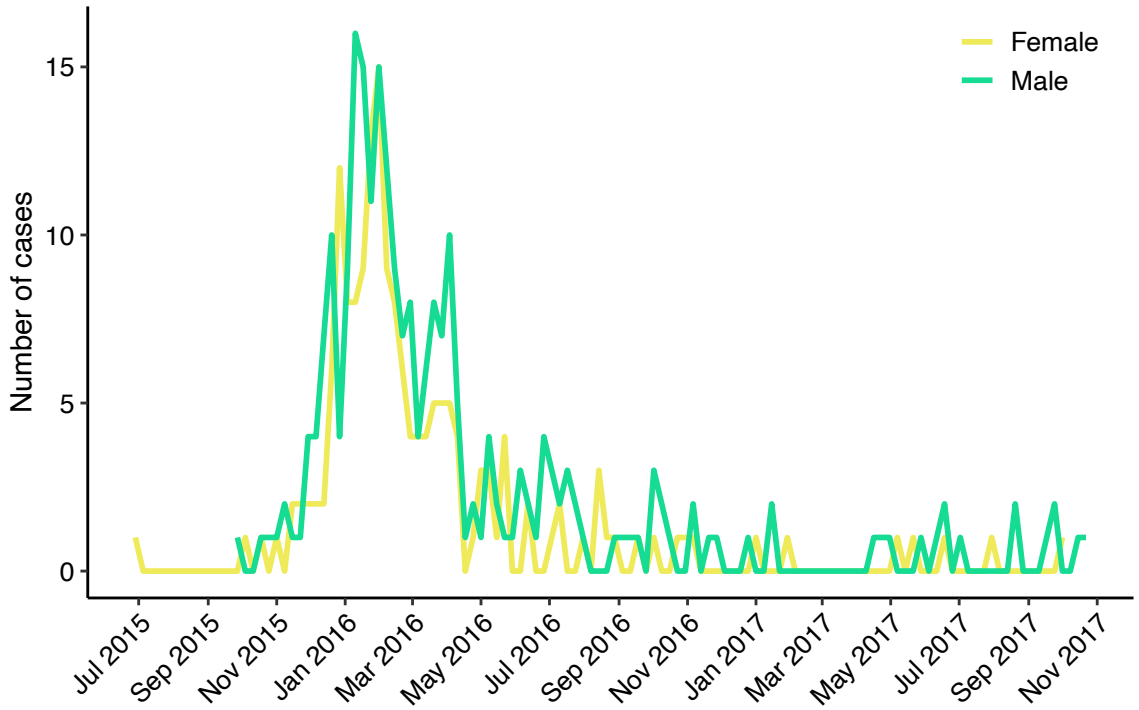


Figure 2.14 Number of ZIKV-associated neurological complications cases by sex and week of symptom onset in Colombia. Epidemiological week 26 of 2015 – epidemiological week 43 of 2017.

4.2.3 Diagnosis and final condition

The majority of ZIKV-associated neurological complications cases had a diagnosis of GBS (85.0%). Encephalitis and myelitis were the second and third most common diagnoses, respectively (Figure 2.15).

Thirty-five cases out of 418 (8.4%) resulted in death.

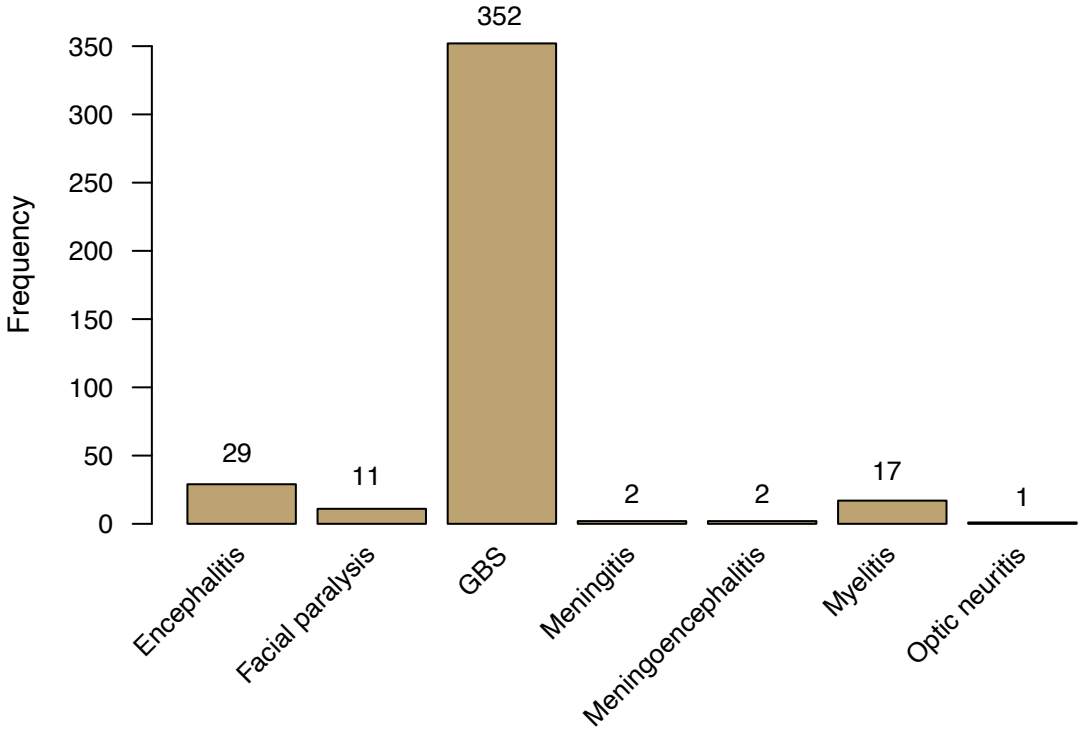


Figure 2.15 Final diagnosis for cases of ZIKV-associated neurological complications completed by a neurologist.

4.2.4 Geographic distribution

Twenty-eight out of 32 departments in Colombia were reported as the location of likely infection for at least one case of ZIKV-associated neurological complications (Figure 2.16). The highest number of cases was reported in Atlántico followed by Norte de Santander and Valle del Cauca with 104, 61, and 44 cases, respectively. The highest observed attack rate of ZIKV-associated neurological complications per 1,000 cases of ZVD was reported in Nariño followed by Chocó and Atlántico (Table 2.2 and Figure 2.16). However, 24 departments had 10 cases or fewer, and the estimated attack rates should therefore be interpreted with caution.

At the city level, Barranquilla was reported as the location of likely infection for the highest number of ZIKV-associated neurological complications with 80, followed by Cúcuta with 44 and Cali with 23. Cases were spread out geographically and tended to cluster in large cities (Figure 2.17).

In contrast, the departments with the highest number of reported ZVD cases included Valle del Cauca with 27,712 (26%), followed by Santander with 10,374 (10%) and Norte de Santander with 10,361 (10%). The highest observed attack rate of ZVD per 100,000 population was reported in San Andrés and Providencia (1,489), followed by Casanare (1,087) and Norte de Santander (758) (Figure 2.18).

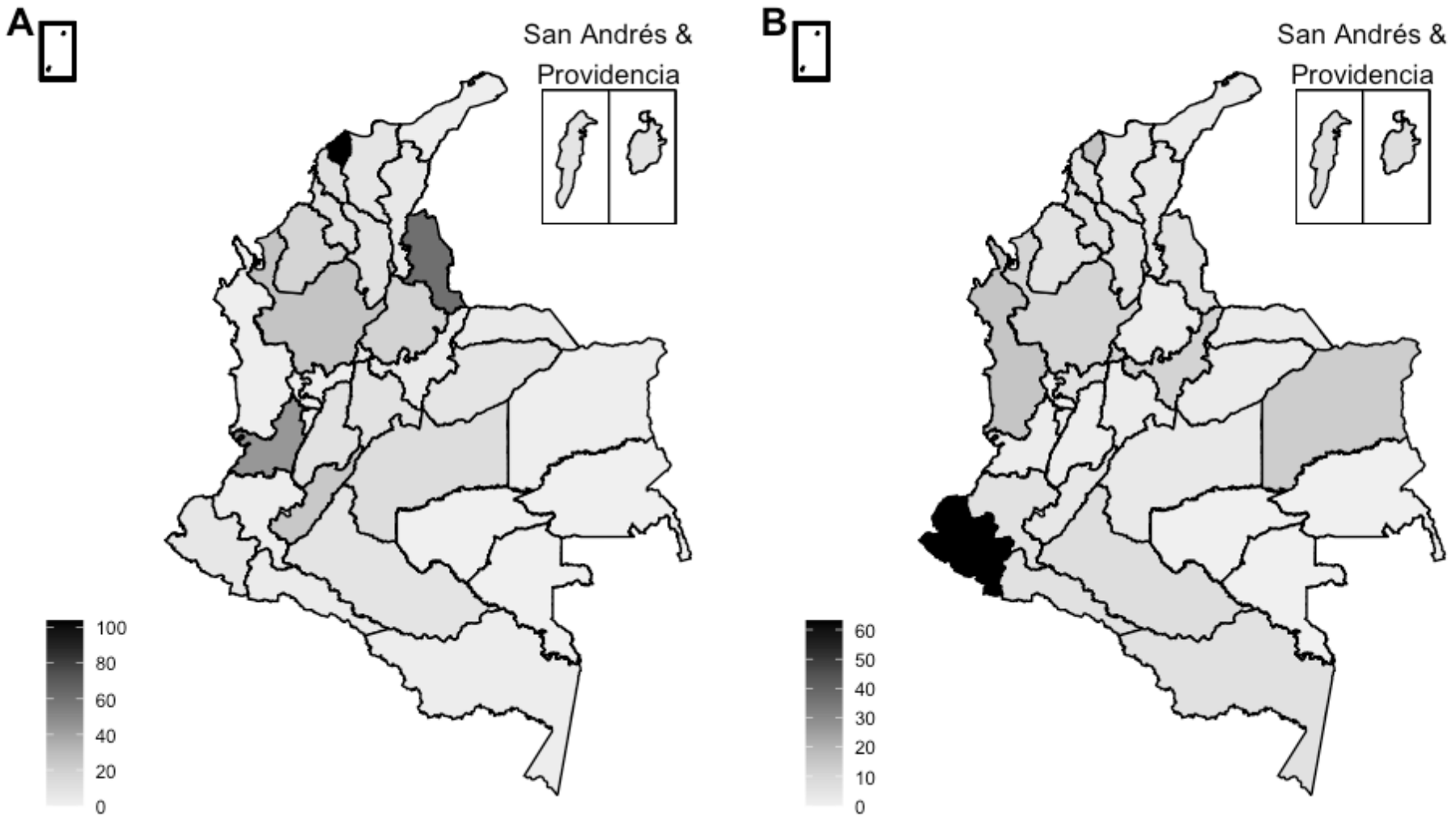


Figure 2.16 Maps of ZIKV-associated neurological complications. *(A)* Map of the number of cases with ZIKV-associated neurological complications by department. *(B)* Map of the observed attack rate of ZIKV-associated neurological complications per 1,000 ZVD cases by department. N = 406 observations with non-missing location at department (administrative 1) level. Maps produced from SIG-OT shapefiles [117].

Table 2.2 Observed attack rate and 95% confidence intervals of ZIKV-associated neurological complications per 1,000 ZVD cases by department.

Department	Observed attack rate of ZIKV-associated neurological complications per 1,000 ZVD cases	95% CI	
Nariño	63.2	23.5	132.4
Chocó	15.4	0.4	82.8
Atlántico	15.3	12.6	18.6
Vichada	12.8	0.3	69.4
Boyacá	10.9	3.0	27.7
Antioquia	10.3	6.7	15.1
Caquetá	7.0	3.0	13.8
San Andrés & Providencia	7.0	3.0	13.7
Sucre	6.1	2.9	11.2
Caldas	6.0	0.7	21.5
Putumayo	6.0	1.2	17.4
Norte de Santander	5.9	4.5	7.6
Amazonas	5.8	0.7	21.0
Cauca	5.6	0.7	20.2
Bolívar	5.2	2.5	9.6
Córdoba	5.1	3.0	8.2
Huila	3.4	2.2	5.1
Cesar	3.1	1.0	7.3
Meta	2.8	1.4	4.8
Quindío	2.5	0.1	13.6
Magdalena	2.2	0.9	4.5
Arauca	2.1	0.6	5.5
Casanare	1.8	0.7	3.7
Santander	1.7	1.0	2.7
Cundinamarca	1.7	0.8	3.2
Valle del Cauca	1.6	1.2	2.1
La Guajira	1.4	0.0	7.9
Tolima	1.3	0.6	2.4
Guainía	0.0	0.0	231.6
Vaupés	0.0	0.0	185.3
Guaviare	0.0	0.0	17.6
Risaralda	0.0	0.0	2.8

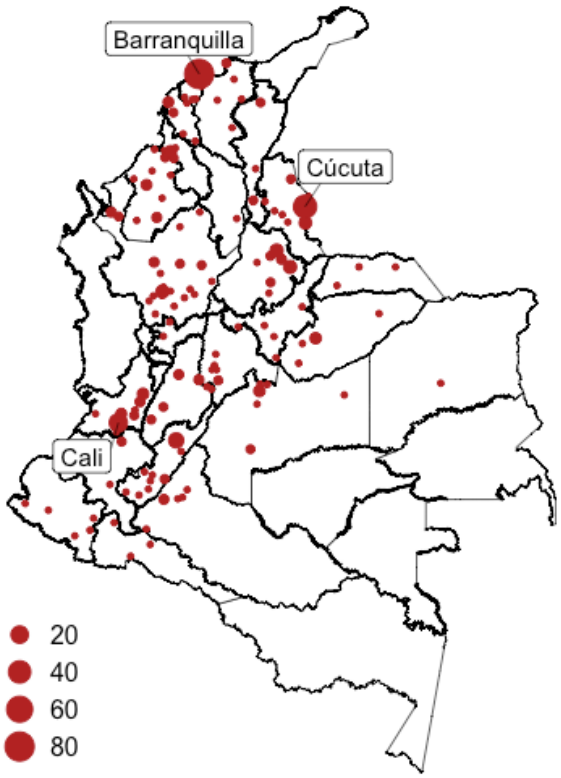


Figure 2.17 Map of the number of ZIKV-associated neurological complications cases by city. N = 398 observations with non-missing location at city (administrative 2) level. The island municipalities of San Andrés and Providencia are not shown (San Andrés reported 8 cases, whereas Providencia reported 0). Map produced from SIG-OT shapefiles [117].

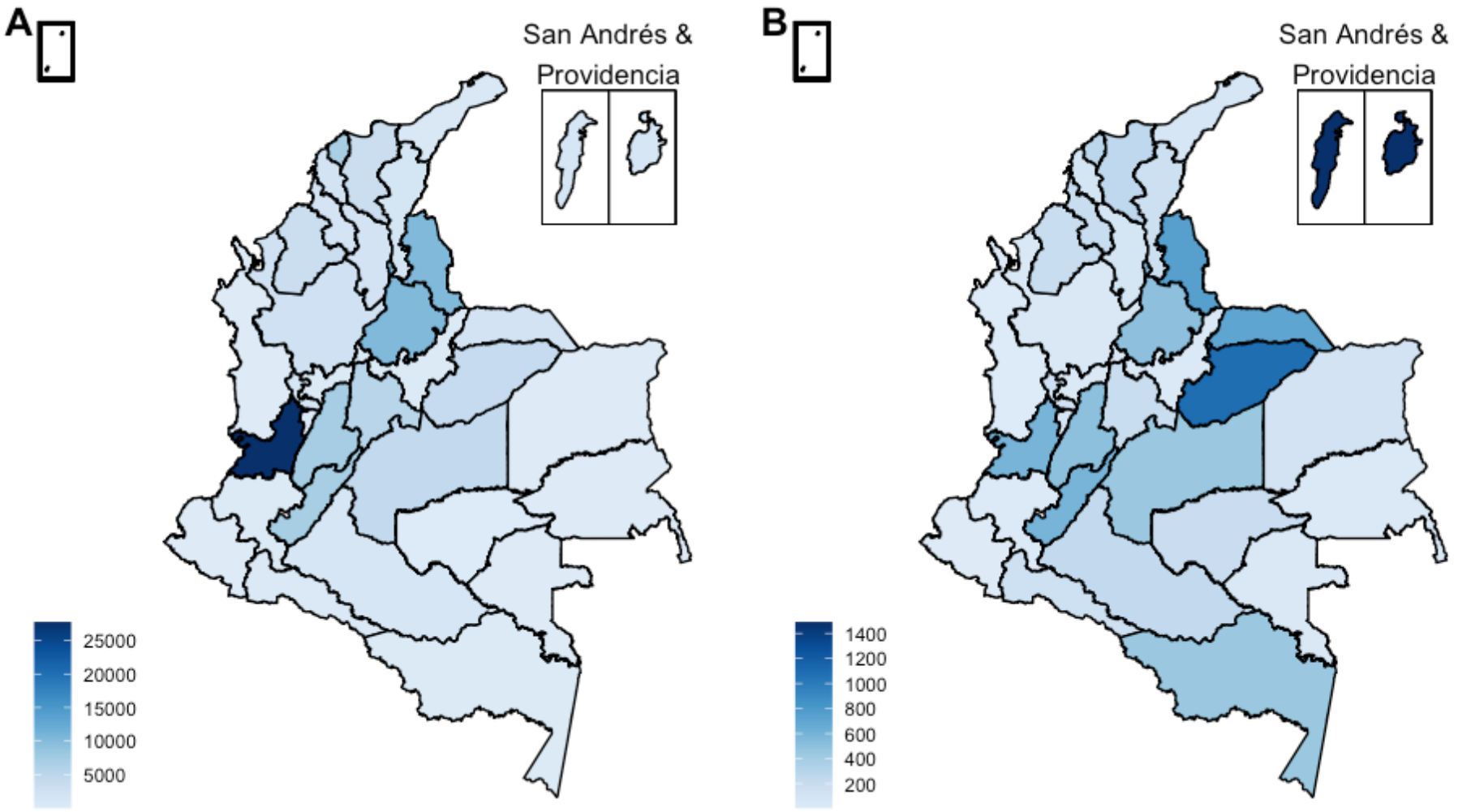


Figure 2.18 Maps of ZVD. *(A)* Map of the number of cases and *(B)* map of the observed attack rate per 100,000 population. N = 106,033 observations. Maps produced from SIG-OT shapefiles [117].

5 Discussion

This work builds on a preliminary report of ZVD cases in Colombia that also used data from the national population-based surveillance system [95]. The earlier report was published just after the peak of the epidemic and included 65,726 cases reported between August 9, 2015 and April 2, 2016. This analysis adds an additional 40,307 cases reported until mid-June 2017 and data on severe cases with neurological complications, including GBS.

5.1 ZIKV

Compared to the general population of Colombia, ZVD cases were more likely to be reported in individuals in their 20s and 30s. Several factors can affect the age distribution of cases found through epidemic surveillance, including age-related variation in susceptibility, reporting bias, pre-existing immunity, the age distribution of the population, and level of exposure to infection [118].

Seroprevalence studies, which test for antibodies indicative of past infection, can be used to assess age-related variation in susceptibility. Although at least one such study found a positive association between ZIKV infection and age, several studies have found no significant association [16, 119-123].

Based on the timings and origins of ZIKV arriving in the Americas from Southeast Asia and the Pacific [18], no immunological protection for ZIKV was assumed at the population level prior to this epidemic. If there was pre-existing immunity, lower infection rates would have been expected in older age groups that had been exposed in the past assuming long-lasting immunity.

Risk factors for ZIKV infection are poorly understood [120]. On Yap Island, Duffy et al. found no behavioral risk factors for ZIKV infection [16]. In contrast, Lozier et al. found higher prevalence of ZIKV among those who reported being bitten by mosquitoes at home in bivariate analyses, but the association was not statistically significant in multivariable analysis [119].

In this study, females had higher observed attack rates of reported ZVD than males. Case ascertainment was likely higher in females of child-bearing age due to concerns about birth

defects. However, reporting bias does not explain the elevated risk in females versus males between the ages of 45 and 79. This result could be explained by differences in susceptibility or exposure. If females in this age range spend more time at home than their male counterparts, they might experience higher exposure to *Ae. aegypti* mosquitoes, which tend to live in and around people's homes in urban areas [124, 125]. The higher observed attack rates of ZVD in females are consistent with ZIKV epidemics in other locations such as Brazil, Puerto Rico, and Yap Island [16, 119, 126]. Spending more time at home might also increase the risk of exposure to mosquito bites and therefore ZIKV infection in young children, which could explain the higher incidence of ZVD in the youngest age group (0-4 years) compared to older children in this study.

Reporting of ZVD cases decreased during Christmas/New Year and Easter holidays in 2015-2016. This trend could be seen at the national and subnational level as well as during the CHIKV epidemic. However, the pattern was less discernable for reports of ZVD in pregnant females and in those with neurological complications. Changes in the number of patient consultations for general practice services on and immediately after public holidays have been observed in several countries, including the UK. This has been called the "public holiday effect" [127]. In Colombia, where 93% of the population identifies as Christian, surveillance for notifiable diseases is likely impacted by religious holidays [52].

5.2 ZIKV-associated neurological complications

Despite higher observed attack rates of ZVD cases in females compared to males, higher observed attack rates of neurological complications were observed in males, a finding which is consistent with GBS epidemiology and other studies [128, 129].

The median age of ZVD cases with neurological complications was 12 years older than that of ZVD cases. Observed attack rates of ZIKV-associated neurological complications increased with age, and the highest observed attack rate was reported among individuals aged 75 and older. The positive association between age and neurological complications here is also consistent with GBS epidemiology [128].

An unexpected finding from this analysis was that neurological complications peaked before ZVD cases. According to a 2016 study of 71 patients in Puerto Rico, the median time between onset of ZIKV infection symptoms and GBS was 7 days (range 0-21 days) [105].

Similarly, a 2015-2016 study reported that the median time between onset of ZIKV infection symptoms and GBS symptoms was 7 days (interquartile range, 3-10 days) for 66 patients from six Colombian hospitals [106]. Given these statistics, neurological complications would be expected to peak about one week after ZVD cases; however, the number of reported ZIKV-associated neurological complications in this study was small, and a formal correlation analysis was not undertaken. As part of future work, an Auto Regressive Integrated Moving Average model could be used to determine the reliability of the spatiotemporal association between ZVD and neurological complications [130]. The relationship could also be explored using another data source such as the Individual Records of Health Services Provision (RIPS). RIPS is a national registry of health interventions that was created in 2000. It contains basic information such as age, sex, and medical diagnosis for patients treated by public and private providers within the Colombian healthcare system [131]. Although RIPS is used most commonly for billing services, it can also be used for public health surveillance [132]. The number of patients hospitalized with any neurological disease, including GBS, could be obtained for 2015-2017 and compared to the trends presented here.

The most common diagnosis among cases in this dataset was GBS. However, six other neurological conditions were also documented, including encephalitis, myelitis, facial paralysis, meningitis, meningoencephalitis, and optic neuritis. Although some studies have focused exclusively ZIKV-associated GBS [105, 106, 133], others have considered a wider range of neurological conditions linked to recent ZIKV infection [134, 135]. Case reports have described ZIKV-associated myelitis [136], encephalitis [137, 138], meningoencephalitis [139], acute disseminated encephalomyelitis [140], Miller-Fisher syndrome [141], and myasthenia gravis [142]. Some of these reports involved fatalities, young people, and previously healthy individuals. In addition to patients with ZIKV-associated neurological complications and CZS, studies using human and animal models have accumulated broader evidence that ZIKV is neurotrophic. The virus targets neuronal cell types, including neural progenitor cells and mature neurons, as well as the brain [143]. ZIKV infection of the central nervous system has been found in both young and adult animals such as mice and non-human primates [143].

Reports of neurological complications associated with ZVD were reported in nearly every department and tended to cluster in large cities with better access to healthcare. This

pattern reflects the widespread dissemination of ZIKV throughout Colombia. The city of Barranquilla had the highest number of reported neurological complications. While there was a large ZIKV epidemic in Barranquilla, the city was also subjected to more intensive surveillance for ZIKV-associated neurological complications compared to other cities [144].

There is some agreement between locations with the highest number of ZVD cases and locations with the highest number of ZIKV-associated neurological complications. However, the locations with the highest observed attack rates of ZIKV per 100,000 population and locations with the highest observed attack rates of neurological complications per 1,000 ZVD cases are discordant. This mismatch could be due to randomness associated with reporting small numbers of rare events or differences in reporting mild versus severe ZVD cases across the country.

5.3 CHIKV

Although it was not possible to examine the epidemiological trends of CHIKV in as much detail as ZIKV in this analysis, a recent study in Barranquilla, Colombia found that 64.5% of 1,160 patients who were clinically diagnosed with CHIKV infection during weeks 36 to 52 of 2014 were female [145]. A statistically significant difference in the distribution of cases by sex was found using chi-square test ($p < 0.0001$). Results also showed a statistically significant difference across age groups: 11.1% of patients were children under the age of 15, 28.7% were 15-29 years of age, 46.1% were 30-59 years of age, and 14.1% were 60 years of age or older (chi-square test, $p < 0.0001$) [145]. The study also identified important differences in the patterns of symptoms exhibited by female patients compared to male patients with females showing a wider array of symptoms [145]. Based on their findings, the authors suggested that females should be considered an at-risk population for CHIKV infection.

Data regarding CHIKV-associated neurological complications in Colombia are scarce. A 2018 systematic review identified four case studies describing a total of only five cases, three of which were perinatally acquired infections [146]. One neonate with congenital CF in Neiva was diagnosed with meningoencephalitis [147], along with two newborns in Sincelejo [148]. There was also a report of a three-week old infant who developed encephalitis following

CHIKV infection in Santander [149] and a report of a 77-year-old woman in Sincelejo with laboratory-confirmed CHIKV infection who was later diagnosed with GBS [150].

5.4 Conclusions and limitations

A strength of this analysis is the quality of the datasets. The ZIKV dataset encompasses the entire duration of the epidemic in Colombia, and all individuals in the neurological complications dataset were checked against standardized case definitions. Limitations include lack of detailed clinical information and lack of laboratory confirmation for most ZVD cases. This report does not include neurological complications in newborns. However, a recently published report found that out of 5,673 pregnancies with laboratory-confirmed ZVD in Colombia, 2% of infants or fetuses had neurological or eye complications [151]. Another recent study from Colombia found that nine out of 60 children (15%) with laboratory-confirmed ZIKV infection at ages 1-12 months had adverse outcomes on neurologic, hearing, or eye examinations at 20-30 months of age. Six of the remaining 47 children (12.8%) had an alert score in the hearing-language domain [152].

Another limitation of this analysis is the reliability of the demographic data as the population projections for 2016 were based on the 2005 Census. Inaccuracy in these data would affect the denominators of the observed attack rates as well as the risk ratios in this chapter. As a result, some departments or groups of individuals with certain combinations of sex and age may have had higher or lower risk of disease than what was estimated here. Other data sources were not considered but could have included WorldPop [153] or DANE's retrospective projections for 2016 based on the 2018 Census [154].

Neurological complications and deaths due to ZIKV were rare in this epidemic. However, more awareness about these risks is needed for people living in or traveling to ZIKV-affected areas. While GBS is relatively easy for non-neurologists to identify, variants such as Miller-Fisher syndrome may not be [141]. Future research should investigate long-term patient outcomes as well as the pathophysiology of these conditions, which can improve treatment strategies [137]. To fully understand the burden of ZIKV, surveillance should encompass a broader spectrum of neurological symptoms of ZVD beyond GBS and microcephaly. Surveillance should also focus on young children, considering the neurotropism of the virus and its effects on postnatal development.

Chapter 3: Impact of climactic and socioeconomic factors on reporting rates and basic reproduction numbers of Zika and chikungunya viruses in Colombia

Abstract

Reporting rates (ρs) and basic reproduction numbers (R_0s) are key epidemiological parameters. In this chapter, both were estimated for the CHIKV and ZIKV epidemics at the department level in Colombia from surveillance data using parametric and non-parametric models based on the renewal equation. Rough approximations of ρ and R_0 were obtained non-parametrically by fitting linear regression models to the relationship of the time-varying reproduction number R_t and cumulative incidence divided by population size for each department. The estimated ρs for CF and ZVD were different (0.044, 95% CI: 0.032-0.056 and 0.015, 95% CI: 0.012-0.018 respectively), but the estimated R_0s were comparable (1.71, 95% CI: 1.54-1.88 and 1.69, 95% CI: 1.59-1.78 respectively). Temperature and socioeconomic factors, including the percentage of households with overcrowded conditions and the percentage of households with inadequate exterior walls, were identified as potentially important covariates in arbovirus transmission using generalized additive models. For both viruses, the best-fitting parametric model allowed for different R_0s across departments and showed significant evidence for overdispersion in incidence (i.e. a negative binomial likelihood was preferred over Poisson). The estimated ρ for CF was higher than that for ZVD (0.045, 95% CrI: 0.042-0.049 and 0.016, 95% CrI: 0.015-0.017, respectively). Estimates of R_0 across departments showed some heterogeneity, ranging from 0.96-2.93 for CHIKV and 0.98-5.87 for ZIKV. From models with a Poisson likelihood, weather covariates improved the fit, i.e. transmission of ZIKV was highest at a higher mean temperature and lower cumulative rainfall compared to CHIKV. Socioeconomic factors appeared uncorrelated with arbovirus transmission in the parametric models.

1 Introduction

1.1 Reproduction numbers

The reproduction number, R, is an epidemiological parameter used to quantify the average number of secondary infections resulting from one typically infected individual [155]. The

value of R is positive and unitless as well as a threshold: once established, if $R > 1$, then an epidemic will continue, whereas if $R < 1$, then an epidemic will eventually end [156]. Estimates of R for the same pathogen tend to vary by geographic location and over time due to differences in contact patterns, demographic rates, population immunity, and other factors [83].

There are different names for R depending on the context in which it is estimated. The basic reproduction number, R_0, applies to situations in which a pathogen is introduced into an entirely susceptible and infinite population [157]. In the presence of population-level immunity, this parameter is considered an effective reproduction number, R_e [158]. Both R_0 and R_e provide information about a pathogen's transmission potential [158]. The instantaneous, or time-varying, reproduction number, R_t, is obtained when R is monitored over the course of an epidemic [159]. Once an epidemic is underway, R_t can be reduced through public health interventions, i.e. social distancing for respiratory diseases such as COVID-19 [160] and enhanced vector control combined with community engagement for mosquito-borne diseases such as ZVD [161]. R_t can also be reduced by the exhaustion of susceptible individuals in the population resulting from either widespread transmission or vaccination efforts [162].

Reproduction numbers can be estimated using a variety of mathematical and statistical methods [163-168]. Examples include exponential growth methods, branching processes, and compartmental models. Data type and availability play important roles in the choice of method. In general, more complex methods can be used with increasing amounts of data. The types of data that can be used to estimate R include chains of transmission ("who infected whom"), cluster size of cases, and epidemic time series [169].

The renewal equation, derived from branching processes, is an example of a method that can be used to estimate R from time series data [170]. In this model, past infections give rise to future infections according to a Poisson process. The link between renewal processes and modeling epidemics comes from work on compartmental models, the Euler-Lokta equation in ecology, and age-dependent branching processes [171]. In epidemiology, the standard equation is:

$$I(t) = \sum_{0}^{\infty} \beta(t,\tau) I(t-\tau) d\tau \quad (3.1)$$

where I is the mean incidence at time t and $\beta(t,\tau)$ denotes transmissibility, which is a function of calendar time t and time since infection τ. Here, the number of newly infected individuals is proportional to the number of past and current cases times their infectiousness [170]. The renewal equation has been used to estimate *R* in a variety of contexts, including the 1918 influenza pandemic, the ZIKV epidemic in the Americas, and the HIV epidemic in Europe [172-174].

EpiEstim is a method based on the renewal equation [159]. It was used to assess changes in transmissibility of Ebola virus during the West Africa epidemic and is currently being used during the ongoing COVID-19 pandemic [175, 176]. Although EpiEstim was originally developed to estimate R_t, it can also be used to estimate R_0. This method models transmission as a Poisson process with the mean case incidence, μ_t, equal to $R_t \sum_{s=1}^{t} I_{t-s} w_s$ where I_t is the observed incidence at time t and w_s is a probability distribution characterizing the generation time distribution (the average time between the time of infection in a primary case and the time of infection in a secondary case infected by the primary case) [159]. The likelihood of the incidence I_t follows:

$$P(I_t | I_0, \ldots, I_{t-1}, w, R_t) = \frac{(R_t \Lambda_t)^{I_t} e^{-R_t \Lambda_t}}{I_t\,!} \quad (3.2)$$

with $\Lambda_t = \sum_{s=1}^{t} I_{t-s} w_s$. R_t can be estimated in a Bayesian framework with a user-specified prior distribution (the default in EpiEstim is a gamma prior distribution with mean 5 and standard deviation 5) [159]. In practice, the distribution of the serial interval (time between onset of symptoms in the primary case and onset of symptoms in their secondary cases) may be used to approximate w_s [159]. Also, a time window rather than a single time step is often used in order to decrease variability and increase precision of the R_t estimates [159]. Software to implement EpiEstim has been developed with versions available in R and Microsoft Excel.

1.1.1 Estimated reproduction numbers for CHIKV in Colombia

Peña-García and Christofferson used the White and Pagano maximum likelihood method to estimate R_0 for CHIKV in 85 Colombian cities [177]. These cities, which accounted for just over 65% of all reported cases in the country, were selected based on having an apparent epidemic peak and reporting more than 150 cases. A serial interval of one or two weeks was assumed. Estimates ranged from 1.11-9.53, and 76% of cities had estimated R_0 values between 1-2 [177].

1.1.2 Estimated reproduction numbers for ZIKV in Colombia

Compared to CHIKV, there are many more estimates of reproduction numbers in the literature for the ZIKV epidemic in Colombia. The studies employed a variety of methods, and while some studies focused on specific cities, most studies estimated countrywide reproduction numbers.

Rojas et al. used maximum likelihood methods to fit a chain-binomial model to daily incidence data from the Colombian cities of Girardot and San Andrés [178]. Data included probable and laboratory-confirmed cases reported to Sivigila between September 2015 and January 2016. R_0 was defined as the median effective reproduction number during the growth phase of the epidemic after adjusting for early reporting delays. Due to uncertainty in the natural history of ZIKV in mosquitoes and humans, three sensitivity settings (short, medium, and long) were considered for the incubation period in humans, infectious period in humans, and infectious period in mosquitoes, resulting in a range of serial intervals. A mean serial interval of 22 days was assumed. The estimated R_0 was 1.41 (95% CI: 1.15-1.74) in San Andrés and 4.61 (95% CI: 4.11-5.16) in Girardot [178].

Towers et al. estimated R_0 in the city of Barranquilla by analyzing 359 clinically identified cases reported during the exponential growth phase of the ZIKV epidemic (until the end of November 2015) [179]. After obtaining the rate of exponential rise in cases, they fitted a mathematical model with compartments for Susceptible, Exposed, Infected, and Recovered humans and well as Susceptible, Exposed, and Infected mosquitoes (SEIR/SEI model). An estimated average R_0 of 3.8 (95% CI: 2.4-5.6) was obtained. They also estimated that the fraction of R_0 due to human-to-human sexual transmission was 0.23 (95% CI: 0.01-0.47) [179].

Ospina et al. estimated R_0 for 20 cities in the department of Antioquia by fitting an SIR model with implicit vector dynamics [180]. Epidemic parameters were estimated by fitting the expression for recovered individuals per unit time to the daily cumulative number of suspected and laboratory-confirmed cases of ZVD according to symptom onset date relative to the index case. The number of cases in each city ranged from 17 to 347 from January 1 to April 11, 2016. R_0 was estimated to be greater than 1 for 15 cities and less than 1 for five cities. The median R_0 estimate across all cities was 1.12 [180].

Chowell et al. estimated R_0 for the department of Antioquia using daily counts of suspected ZVD cases by date of symptom onset reported from January 2016 to April 2016 [174]. They applied the renewal equation to incident cases simulated from a generalized-growth model. The generalized-growth model describes the initial growth phase of an epidemic; it has two parameters, the growth rate, *r*, and a "deceleration of growth" parameter, *p*, which allows the model to capture sub-exponential growth patterns [181]. Assuming a gamma distributed generation time with mean 14 days and standard deviation 2 days, they estimated an R_0 of 10.3 (95% CI: 8.3-12.4) after 14 days of the epidemic and 2.2 (95% CI: 1.9-2.8) after 28 days of the epidemic [174].

Hsieh used the Richards growth model to estimate the R_0 of ZIKV for 13 countries and territories, including Colombia [182]. The Richards model is a modified version of the logistic growth model. It has three parameters: the growth rate, *r*, the size of the epidemic, *K*, and a parameter, *a*, which captures how much the epidemic dynamics differ from the S-shape of the classic logistic growth model [183]. Data for the 11 countries in the Americas were obtained from PAHO and consisted of weekly laboratory-confirmed cases reported until epidemiological week 18 of 2016. The estimated mean R_0 for Colombia was 1.75 (95% CI: 1.34-2.16) when the model was fitted to data from epidemiological weeks 32-43 of 2015 and 1.79 (1.29-2.30) when fitted to data from epidemiological week 49 of 2015 to epidemiological week 16 of 2016 [182].

Nishiura et al. used maximum likelihood estimation and early exponential growth rate methods to estimate a countrywide R_0 of ZIKV for Colombia [184]. ZVD cases were first reported in epidemiological week 32 of 2015. They fitted their model to laboratory-confirmed cases reported from epidemiological week 35 of the epidemic after which

exponential growth in the number of cases was observed. Assuming that the exponential growth continued for 3, 4, and 5 weeks from week 35, the following estimates of R_0 were obtained: 3.9 (95% CI: 2.4-5.7), 6.6 (95% CI: 5.5-7.7), and 3.0 (95% CI: 2.5-3.6), respectively [184].

Rocklöv et al. used a temperature-driven vectorial capacity model to estimate R_0 and assess the risks of ZIKV introduction and spread in Europe [185]. Vectorial capacity quantifies the relationship between arthropod vectors and their hosts and is a function of vector competence, vector lifespan, and the extrinsic incubation period. In mosquitoes, the extrinsic incubation period is the time it takes for a mosquito to become infectious after ingesting a pathogen. The model was validated with surveillance data from the epidemics in the Americas. "Observed" R_0 values were estimated from weekly case count data during the initial phases of the epidemics using the exponential growth rate method [165]. Cases were assumed to be Poisson distributed with a serial interval distribution of mean 16 days and standard deviation 3 days. Estimates at the subnational level were aggregated to national averages. Surveillance data for Colombia included 1,593 confirmed cases in 15 subnational administrative regions from January 1, 2016 to March 13, 2016. The nationwide average R_0 was estimated to be 3.2 [185]. The predicted R_0 from the best-fitting vectorial capacity model (model 1) had mean 3.0 and 2.7 for *Ae. aegypti* and *Ae. albopictus* mosquitoes, respectively.

Majumder et al. used the Incidence Decay and Exponential Adjustment model to estimate the R_0 of ZIKV in Colombia from both digital disease surveillance data and traditional (INS) surveillance data [186]. After reported cases of ZVD were obtained from HealthMap, comprising nongovernmental media alerts between October 16, 2015 and April 16, 2016, Google search data were used to smooth the curve of cumulative reported cases over time. Nonlinear optimization was used to minimize the sum of squared differences between observed and modeled cumulative incidence curves [186]. Across 14 serial interval lengths ranging from 10 to 23 days, the authors estimated a mean R_0 of 2.56 (range 1.42-3.83) using the smoothed HealthMap data and a mean R_0 of 4.82 (range 2.34-8.32) using the INS data. They also estimated R_{obs}, the observed transmission given existing interventions, which is similar to R_t. Results showed a mean R_{obs} of 1.80 (range 1.42-2.30) and 2.34 (range 1.60-3.31) for the HealthMap data and INS data, respectively [186].

Sasmal et al. estimated R_0 by fitting five different compartmental models of ZIKV transmission, including both vector-borne and sexual routes of transmission, to the number of reported suspected and laboratory-confirmed cases of ZVD in pregnant women in Colombia [187]. Data were obtained from PAHO and ranged from epidemiological week 42 of 2015 to epidemiological week 33 of 2016. The best-fitting model stratified each gender into high- and low-risk groups according to sexual behavior. The estimated R_0 for this model had mean 1.89 (95% CI: 1.21-2.13) [187]. They also estimated that human-to-human sexual transmission contributed 15.36% (95% CI: 12.83-17.14) to R_0 on average.

O'Reilly et al. used a spatiotemporal dynamic transmission model for ZIKV infection in 90 cities within 35 Latin American and Caribbean countries to estimate $R_0(t)$, the time-varying reproduction number in a completely susceptible population. Five Colombian cities were included in their analysis: Barranquilla, Bucaramanga, Cali, Cartagena, and Medellín. A deterministic meta-population model was used to model ZIKV transmission between cities with migration between cities represented by a gravity model. They fitted the model to summary statistics related to the timing of peak incidence rather than weekly incidence of ZIKV cases. For Colombia, their results estimated that there were 314 days (95% CrI: 311-315) when $R_0(t) > 1$ with an average $R_0(t)$ during a typical year of 1.94 (95% CrI: 1.87-2.01).

A summary of the R_0 estimates for CHIKV and ZIKV in Colombia from the literature can be found in Table 3.1.

Table 3.1 R_0 estimates for CHIKV and ZIKV in Colombia from the literature.

Ref	Virus	Location	Study year	Method	R_0 estimates (95% CI)
[177]	CHIKV	85 cities	2014-2016	White and Pagano	Range: 1.11-9.53
[178]	ZIKV	Girardot	2015-2016	Maximum likelihood to fit chain-binomial model	4.61 (4.11-5.16)
[178]	ZIKV	San Andrés	2015-2016	Maximum likelihood to fit chain-binomial model	1.41 (1.15-1.74)
[179]	ZIKV	Barranquilla	2015	SEIR/SEI model	3.8 (2.4-5.6)
[180]	ZIKV	20 cities in Antioquia	2016	SIR model with vector dynamics	1.12
[174]	ZIKV	Antioquia	2016	Generalized-growth model	Range: 2.2-10.3
[182]	ZIKV	National	2015-2016	Richards model	Range: 1.75-1.79
[184]	ZIKV	National	2015	Maximum likelihood estimation and early exponential growth rate model	Range: 3.0-6.6
[185]	ZIKV	National	2016	Exponential growth rate method	3.2
[186]	ZIKV	National	2015-2016	Incidence Decay and Exponential Adjustment model	Range: 2.56-4.82
[187]	ZIKV	National	2015-2016	Compartmental models	1.89 (1.21-2.13)

1.2 Reporting rates

In addition to reproduction numbers, the reporting rate (ρ), or the proportion of infections that are ultimately reported as cases of the disease, is of considerable interest during and after an epidemic. As mentioned in chapter 1, surveillance systems do not capture all cases which leads to uncertainty in the "true" incidence of disease. Cases can be missed due to (i) under ascertainment of infections at the community level and (ii) underreporting of infections at the healthcare level [188]. Under-ascertained infections occur in individuals who do not seek healthcare, whereas underreported infections occur in individuals that do seek healthcare but whose health event is not reported. Factors influencing whether an individual seeks healthcare include symptom severity and health literacy as well as culture, religion, and cost [188]. Underreporting can occur when infections are undiagnosed or misdiagnosed and also when infections are correctly diagnosed but not appropriately reported [188]. As surveillance data are used to guide resource allocation and estimate epidemiological parameters, it is important to understand how accurately they reflect the

true burden of disease in populations and adjust observed incidence with multiplication factors as needed. Several methods and study designs can be used to estimate under ascertainment and underreporting in surveillance systems, including community-based studies, serological surveys, returning traveler studies, capture-recapture studies, and modeling [188].

1.2.1 Estimated reporting rates for CF in Colombia

Reporting rates for the CHIKV epidemic in Colombia have been estimated from community-based studies and serological surveys.

A retrospective, community-based study was performed by the INS in the city of Girardot, Cundinamarca to estimate reporting rates of CF between November 2014 and May 2015 [189]. The study design involved community- and institutional-based active case-finding among inhabitants of the city's urban areas. At the community level, surveys were administered to all households in blocks that were selected by simple random sampling. At the institutional level, all RIPS were reviewed as well as the number of cases reported to the national surveillance system (Sivigila). The case definition was based on clinical diagnosis and self-reported symptoms (fever, joint pain, and rash during the study period) [189]. There were 8,788 cases of CF reported in Girardot during the study period. An infection attack rate of 0.648 (95% CI: 0.631-0.664) was estimated from a sample size of 3,380 survey participants. Considering the urban population size (101,610), the investigators estimated that 87.1% (95% CI: 86.8-87.3) of symptomatic CHIKV infections in the city were not reported to the surveillance system (ρ = 0.129, 95% CI: 0.127-0.132), of which 36.1% (95% CI: 34.1-38.1) could be explained by individuals not seeking health services and 24.8% (95% CI: 23.7-25.8) could be explained by underreporting. The remaining 26.2% of underreporting was not explained in the study. Among those surveyed, the most frequently cited reason for not seeking care was self-medication (55%) followed by the collapse of the healthcare system (28%) [189]. A limitation of this study is the case definition, which could have increased the number of perceived cases in the community (i.e., if the symptoms were caused by a different virus) and would have excluded asymptomatic cases.

Another community-based study was conducted in the city of El Espinal, Tolima using a similar study design as in Girardot for the period from October 2014 to June 2015 [190].

Sivigila was notified of 3,794 cases in El Espinal during the study period. The estimated infection attack rate was 0.670 (95% CI: 0.666-0.674) among those surveyed (N=5,774). Considering the urban population size (58,367), the number of affected individuals was estimated at 39,106 (95% CI: 38,872-39,339) with a symptomatic reporting rate of 0.097 (95% CI: 0.096-0.098). Of 3,872 cases that were identified through active case-finding in the community, 46.6% (95% CI: 45.5-48.5) did not seek health services and 94.7% (95% CI: 93.9-95.3) were not registered in either RIPS or Sivigila databases. As in Girardot, the main motivation for not seeking care among survey participants was self-medication (53.5%). At the institutional level, of 3,052 patients that received a diagnosis of CF at San Rafael Hospital, 44.8% (95% CI: 43.6-46.0) were not reported to Sivigila.

Nouvellet et al. conducted a household-based seroprevalence study in four cities located in different transmission areas from October to December of 2016 [191]. Using multistage probabilistic sampling, they obtained a total sample size in excess of 2,400 participants between 2-45 years of age. Past infection by CHIKV, ZIKV, and DENV was determined by testing for IgG antibodies using a multiplex recombinant antigen-based microsphere immunoassay. The estimated post-epidemic seroprevalence from the study can be found in Table 3.2. The last column in the table shows the reporting rates which were estimated from the study results. Estimated reporting rates for CHIKV ranged from <0.001 in Medellín to 0.099 in Neiva.

Table 3.2 Reporting rates derived from multisite seroprevalence study of ZIKV and CHIKV in Colombia [191].

Virus	City	Population	Reported suspected and laboratory-confirmed cases	Raw attack rate (cases/ pop.)	Estimated post-epidemic seroprevalence and 95% CI	Multiplication factor (Seroprevalence/ raw attack rate)	Reporting rate (1/ multiplication factor)
CHIKV	Cúcuta	656,380	26,512	0.040	0.723 (0.687-0.758)	17.9	0.056
	Medellín	2,486,723	17	<0.001	0.071 (0.052-0.094)	10,386	<0.001
	Neiva	344,026	19,751	0.057	0.580 (0.539-0.619)	10.1	0.099
	Sincelejo	279,031	12,342	0.044	0.606 (0.566-0.644)	13.7	0.073
ZIKV	Cúcuta	656,380	6,485	0.010	0.479 (0.440-0.519)	48.5	0.021
	Medellín	2,486,723	549	<0.001	0.067 (0.048-0.090)	303	0.003
	Neiva	344,026	3,409	0.010	0.578 (0.538-0.618)	58.3	0.017
	Sincelejo	279,031	856	0.003	0.659 (0.620-0.696)	215	0.005

1.2.2 Estimated reporting rates for ZVD in Colombia

Reporting rates for the ZIKV epidemic in Colombia have been estimated from modeling studies, community-based studies, and serological surveys.

The O'Reilly et al. modeling study from section 1.1.2 also estimated country-specific reporting rates for ZIKV. Results for Colombia estimated a median ρ of 0.017 (95% CrI: 0.013-0.025) [23].

Mier-y-Teran-Romero et al. used a Bayesian inference model to estimate the probability of ZIKV infection, the proportion of ZIKV infections that are reported as suspected/laboratory-confirmed ZVD cases, and the risk of ZIKV-associated GBS for nine locations in the Americas plus Yap Island and French Polynesia [133]. The data consisted of the cumulative number of reported GBS cases and suspected ZVD cases during the outbreaks as well as seroprevalence data from Yap Island and French Polynesia. The probabilities of interest were related to the observed data through a binomial sampling process. Results for Colombia showed a probability of ZIKV infection with mean 0.09 (95% CrI: 0.03-0.23) and a ρ with mean 0.03 (95% CrI: 0.01-0.07) [133].

Moore et al. used multiple publicly available data sources within a Bayesian hierarchical model framework to estimate national and subnational reporting rates, the fraction of symptomatic infections, and subnational infection attack rates for 15 Latin American countries and territories [192]. The primary analysis only included locations with subnational data available for at least one of the following data types: suspected and laboratory-confirmed ZVD cases in pregnant woman and in the total population, ZIKV-associated GBS cases, and cases of CZS. For Colombia, the estimated infection attack rate was 0.19 (95% CrI: 0.15-0.23) resulting in 9,302,116 (95% CrI: 7,133,364-11,360,866) infections [192]. The estimated probability that a symptomatic ZIKV infection is reported as a suspected and laboratory-confirmed ZVD case had an overall mean of 0.036 (95% CrI: 0.018-0.070) and 0.004 (95% CrI: 0.002-0.007), respectively, and at the department level, the probability of reporting a suspected case ranged from mean <0.001 (95% CrI: <0.001-<0.001) in Bogotá to 0.145 (95% CrI: 0.061-0.304) in Cundinamarca[3]. Compared to the other locations analyzed, Colombia was found to have the greatest variation in reporting probabilities within administrative units for suspected cases with 70.2% of the variance explained by between-administrative unit variance [192]. The probability of reporting a confirmed case varied from mean <0.001 (95% CrI: <0.001-0.001) in Bogotá to 0.011 (95% CrI: 0.004-0.024) in Boyacá.

A community-based study was carried out in Girardot, Cundinamarca to estimate ZVD reporting rates between October 2015 and May 2016 [193]. Similar methodology was used as in the CHIKV studies previously conducted in Girardot and El Espinal [189, 190]. A total of 1,256 cases were reported to Sivigila during the study period. At the community level, the estimated infection attack rate was 0.152 (95% CI: 0.142-0.161) among 5,542 survey respondents. Accounting for the urban population of the city (102,225 in 2015), an estimated 91.9% (95% CI: 91.4-92.4) of symptomatic ZIKV infections were not reported to the surveillance system (ρ = 0.081, 95% CI: 0.076-0.086). Nearly half (49.1%, 95% CI: 45.7-52.5) of 845 individuals who met the case definition (presence of fever, red eyes, headache,

[3] The estimates in both this sentence and the last sentence of the paragraph were not reported in the paper by Moore et al. Rather, they were obtained from the posterior samples of the model, which were kindly provided by Dr. Sean Moore in personal communication.

musculoskeletal pain, light sensitivity, itching, arthralgia, or rash during the study period) did not seek health services. Similar to the results of the CHIKV studies, the primary reason given for not seeking healthcare was self-medication (40.0%) [193]. At the institutional level, 83.3% (95% CI: 80.1-86.2) of 594 patient records identified through RIPS that met the case definition for ZVD were not reported to Sivigila. In particular, 81.7% (95% CI: 78.1-85.0) of 520 patients that received a diagnosis of ZVD were not reported to Sivigila, along with 94.4% (95% CI: 86.2-98.4) of 71 patients that received a diagnosis other than ZVD but met the case definition [193].

As discussed in section 1.2.1, the seroprevalence study by Nouvellet et al. also included ZIKV, and the reporting rates estimated from their results are presented in Table 3.2. Estimated reporting rates for ZVD ranged from 0.003 in Medellín to 0.021 in Cúcuta [191]. A summary of reporting rates for CF and ZVD in Colombia from other studies in the literature can be found in Table 3.3.

Table 3.3 Estimates of reporting rates from modeling and community-based studies for CF and ZVD in Colombia from the literature.

Ref	Disease	Location	Study year	Study type	ρ estimates (95% CI or 95% CrI)
[190]	CF	El Espinal	2014-2015	Community-based	0.097 (0.096-0.098)*
[189]	CF	Girardot	2014-2015	Community-based	0.129 (0.127-0.132)*
[193]	ZVD	Girardot	2015-2016	Community-based	0.081 (0.076-0.086)*
[192]	ZVD	All departments	2015-2018	Modeling	Range: <0.001-0.145**
[192]	ZVD	National	2015-2018	Modeling	0.036 (0.018-0.070)**
[23]	ZVD	National	2015-2017	Modeling	0.017 (0.013-0.025)
[133]	ZVD	National	2015-2016	Modeling	0.03 (0.01-0.07)

*Symptomatic reporting rate.
**Probability that a symptomatic ZIKV infection is reported as a suspected ZVD case.

1.3 Weather and arbovirus transmission

Weather is an important driver of arbovirus transmission due to its impacts on mosquito ecology. For example, temperature has been shown to affect mosquito survival, larvae development, and adult feeding behavior [194]. In general, the effect of temperature on arbovirus transmission is unimodal: transmission is optimized at a particular temperature and tends to decline at extreme temperatures [195]. Similarly, while mosquitoes need enough rainfall for breeding and larvae development, excessive amounts of rainfall can

wash away breeding sites, causing larvae mortality [196, 197]. Several study designs have been used to assess the relationships between weather and arbovirus transmission, including experimental studies in laboratories and in the field as well as mathematical modeling approaches. Additionally, some studies have focused on the influence of weather on the life history traits of the mosquito vectors.

Brady et al. used generalized additive models (GAMs) to study survival of adult *Ae. aegypti* and *Ae. albopictus* at different temperatures under laboratory conditions and in the field [198]. GAMs are non-parametric models that apply smoothing functions to explanatory variables. In this way, they are capable of capturing unknown and non-linear effects of covariates. Data from 351 published studies on adult *Ae. aegypti* and *Ae. albopictus* survival experiments in the laboratory were used to build models for each species across a range of temperatures. These models were then adapted to estimate the effects of temperature on survival of *Ae. aegypti* and *Ae. albopictus* in the field using data from 59 experiments. They found that although *Ae. albopictus* tended to survive longer than *Ae. aegypti* in the laboratory and in the field, *Ae. aegypti* could withstand a wider range of temperatures, including lower temperatures [198]. A limitation of the study was that few field experiments were conducted at extreme temperatures. Also, few experiments on mosquito mortality in the field were available.

Riou et al. used a hierarchical time-dependent SIR model to jointly analyze CHIKV and ZIKV transmission in French Polynesia and the French West Indies [199]. Their model incorporated a disease-specific serial interval and accounted for effects of virus, location, and weather. The model was fitted to weekly case data spanning the entire duration of the outbreaks with the exception of the ZIKV outbreaks in the French West Indies which were still ongoing at the time of the analysis. Results showed that mean temperatures in the range of 22-29°C did not impact transmission. However, a 1 cm increase in average weekly rainfall led to a 10% decrease in transmission after one to two weeks; after four to six weeks, transmission increased by about 20-40% [199].

Mordecai et al. developed mechanistic transmission models to study the effects of mean temperature on transmission of DENV, CHIKV, and ZIKV by *Ae. aegypti* and *Ae. albopictus* mosquitoes [195]. The models incorporated empirical estimates of the unimodal effects of

temperature on mosquito and pathogen characteristics, such as survival, development, and biting rates, and were validated by fitting country-level human case incidence data from 2014-2016 for all countries in the Americas. Results showed that transmission of the three viruses was possible between 18-34°C and optimal between 26-29°C [195]. According to the study, "the thermal response curve for *Ae. albopictus* is shifted towards lower temperatures than *Ae. aegypti*, so that *Ae. albopictus* transmission is better suited to cooler environments." A limitation of the study was that thermal response data for CHIKV and ZIKV were not available.

Tesla et al. used mechanistic R_0 models parameterized with data from laboratory experiments with *Ae. aegypti* to estimate the effects of temperature on ZIKV transmission [200]. Vector competence, extrinsic incubation period, and mosquito survival were assessed at eight different temperatures. Strong, unimodal effects of temperature were observed for each of the life history traits considered. Thermal response data were then fed into a previously developed temperature-dependent model. Results showed that the optimal temperature for ZIKV transmission was 29°C and ranged from 22.7°C to 34.7°C. Notably, the authors found that the predicted thermal minimum for ZIKV transmission is 5°C warmer than that of DENV [200]. This means that models that assume ZIKV behaves similarly to DENV may over-predict environmentally suitable areas at risk of ZIKV.

Kakarla et al. used a mathematical model to study the effects of temperature on CHIKV transmission in India [201]. Their model was a temperature-dependent dynamical transmission model based on R_0. R_0 as a function of temperature was related to the relative density of mosquitoes per human, the biting rate, vector to host infection probability, the mortality rate of the vector, the extrinsic incubation period, and viremia. Equations for the temperature-dependent parameters were obtained from the literature; however, most of them were actually specific to DENV rather than CHIKV. They used a dataset consisting of station observation-based global land monthly mean surface air temperature from 1948 to 2016 as well as gridded monthly rainfall data for the same period. Results showed that the optimal temperature for peak CHIKV transmission was 29°C. Temperatures greater than 24°C and less than 34°C were also favorable for transmission (mean $R_0 > 1$) [201].

Harris et al. used a time-varying SIR model to study how climate factors may have influenced the emergence and intensity of the ZIKV epidemic in Latin America [202]. Epidemiological data included weekly suspected and laboratory-confirmed cases of ZVD between November 2015 and November 2017 for 127 provinces across six countries. Weather data, including daily mean relative humidity and total rainfall as well as mean, minimum, and maximum temperatures, were used to calculate climate metrics with time lags relevant to ZIKV spread via *Ae. aegypti*. They found that the force of infection for ZIKV was highest in provinces with temperatures of 23.6°C (95% CI: 22.2-25.5) [202]. Important predictors of ZIKV presence were temperature, temperature range, and rainfall. ZIKV intensity and burden were best explained by rainfall, relative humidity, and a nonlinear effect of temperature. However, climate factors were not strong predictors of ZIKV epidemic dynamics across weeks.

Chien et al. used a Bayesian structured additive regressive modeling approach to assess the risk of ZIKV across Colombian departments according to weather factors [203]. Weekly reported cases of ZVD were obtained from the INS Epidemiological Bulletins for week 39 of 2015 until the last week of 2017. The study considered the effects of temperature, dew point temperature, relative humidity, sea-level pressure, wind speed, and total rainfall. A zero-inflated Poisson model was chosen to account for excess zeros in the data. The model included spatial interaction terms with weather factors as well as a geospatial function intended to capture spatial variation of ZIKV that could not be explained by weather factors. Results showed that the best-fitting model according to DIC included average temperature and total rainfall. An increase in rainfall of 1 inch was associated with an increase in the logarithm of relative risk of ZIKV of no more than 1.66 (95% CrI: 1.09-2.15), and an increase of 1°F of mean temperature was associated with no more than 0.79 (95% CrI: 0.12-1.22) increase in the logarithm of relative risk of ZIKV [203].

1.4 Aims

The aim of this chapter is to analyze the drivers of CHIKV and ZIKV transmission in Colombia using two approaches based on the renewal equation. Disease incidence was modeled parametrically and non-parametrically. The proportion of infections that were reported as cases to the surveillance system was estimated as well as reproduction numbers for each

department. The effects of temperature, rainfall, and socioeconomic status were also considered.

2 Data

2.1 Epidemiological data

This chapter utilized the CHIKV and ZIKV surveillance datasets from Sivigila that were described in chapter 1. Specifically, suspected and laboratory-confirmed cases with missing information on city (administrative level 2) location were included in this analysis, resulting in 412,915 CF cases and 106,033 ZVD cases. As previously stated, dates for the CHIKV dataset span 110 weeks, from the week ending June 7, 2014 to that ending July 9, 2016, while dates for the ZIKV dataset span 97 weeks from the week ending August 15, 2015 to that ending June 17, 2017. Also, the location of cases refers to location of likely infection, which is determined by the clinician who reported the case.

2.2 Demographic data

As before, population projections for 2016 were obtained from DANE.

2.3 Socioeconomic data

Multidimensional poverty data for the year 2018 were downloaded from DANE at the department level. These data come from the Encuesta Nacional de Calidad de Vida (National Quality of Life Survey), which collects data on Colombians' living and housing conditions. Three variables that might be relevant to arbovirus transmission, including the percentage of households in each department with overcrowded conditions, inadequate exterior walls, and inadequate floors, were obtained. Overcrowding was defined in terms of the number of people sleeping per room excluding the kitchen, bathroom, and garage and including the living room and dining room. A house was considered overcrowded if there were three or more people sleeping per room in an urban area and at least four people per room in a rural area. In an urban area, exterior walls made from unfinished wood, boards, planks, bamboo or other vegetation, zinc, fabric, and cardboard were defined as inadequate, along with damaged walls or no walls. In a rural area, walls constructed of bamboo or other vegetation, zinc, fabric, and cardboard were considered inadequate as well as damaged walls or no walls. Inadequate floors were those that were made of dirt in both urban and rural areas.

2.4 Weather data

Weekly mean temperature data for Colombia from January 1, 2014 to October 1, 2016 were downloaded from Dryad Digital Repository [204]. These data were weighted by the population and aggregated at the department and city levels. However, ZVD cases were reported through mid-2017. To obtain temperature data after October 1, 2016, daily meteorological station readings were downloaded from the National Oceanic and Atmospheric Administration (NOAA)'s Climate Data Online [205]. This website contains past weather and climate data that are publicly and freely available to download.

PERSIANN-Cloud Classification System (PERSIANN-CCS) satellite precipitation data for Colombia were downloaded at daily time steps from the Center for Hydrometeorology and Remote Sensing (CHRS) Data Portal from January 1, 2014 to December 31, 2017 [206]. This source of meteorological data was chosen over NOAA's Climate Data Online based on the work of Siraj et al. [207]. They found large spatial variability in the NOAA data, and as a result, they obtained substantially different estimates of precipitation from spatial models compared to the observed values of the weather stations. Instead, they used satellite-based data from NOAA's Center for Satellite Applications and Research. However, a third data source was sought for this thesis as the resulting Siraj et al. datasets contained one outlier for the city of San Andrés (915 mm on the week ending July 23, 2016) and two outliers for the department of San Andrés and Providencia (771 mm and 823 mm on the weeks ending on August 30, 2014 and July 23, 2016, respectively) that could not be explained by extreme weather events such as hurricanes.

WorldPop Project data were downloaded from [204] as the weighting variable for the spatial aggregation of weather covariates. These data consist of 2015 estimates of the number of people per pixel with national totals adjusted to match the United Nations Population Division estimates. They were subset and resampled by Siraj et al. to match other climate variables considered in their database (~93 m resolution) [207].

3 Methods

3.1 Processing weather data

3.1.1 Mean temperature

The data processing workflow from Siraj et al. [207] was used to generate the mean temperature at the department and city level of Colombia from January 1, 2016 through January 31, 2017. In general, spatial models (kriging) were used on readings from meteorological stations in Colombia. A model is needed to interpolate the data over the parts of the country that do not have stations. This particular model with altitude and secondary temperature data as covariates was selected over (i) kriging without covariates and (ii) non-parametric surface fitting with thin-plate splines with or without covariates, based on leave-one-out cross validation. After kriging, the data were rasterized. Daily gridded data were generated from the raster files, which were then aggregated by week and multiplied by the population.

Siraj et al. used the first version of NOAA's Climate Data Online tool, the Legacy Climate Data Online [208]. From the Global Summary of the Day data product, they extracted the minimum daily temperature, maximum daily temperature, mean daily temperature, and relative humidity from 30 stations between January 1, 2016 and December 31, 2016. The dataset containing mean temperature that they downloaded for 2016 was obtained from the GitHub repository linked to their paper (https://github.com/asiraj-nd/zika-colombia) as well as the Global 30 Arc-Second Elevation dataset, the WorldClim dataset, and NOAA's Climate Prediction Center surface air temperature dataset. Here, the Daily Summaries data product from the current version of NOAA's Climate Data Online tool was selected to obtain 2017 station readings [205]. Five stations were missing from the new data. However, these stations had relatively few observations in the data originally downloaded by Siraj et al.; the five stations had between seven and 362 observations over three years (1,096 days) compared to the other 25 stations, which had between 1,078 and 1,096 observations. Figure 3.1 shows the locations of the 30 weather stations in Colombia.

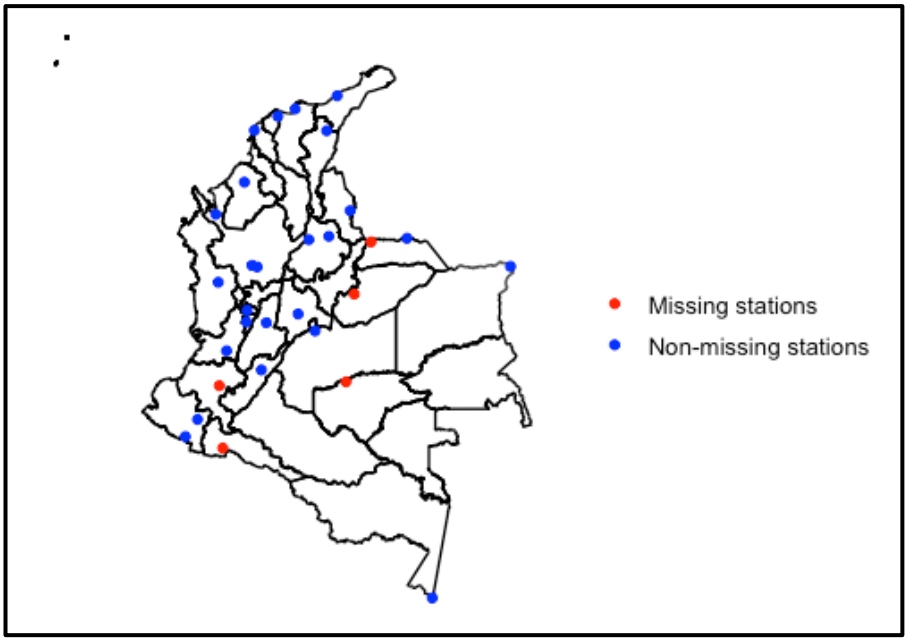


Figure 3.1 Map of 30 weather stations in Colombia. Data from all stations were obtained by Siraj et al. from the National Oceanic and Atmospheric Organization's Legacy Climate Data Online tool between 2014 and 2016. Blue points are stations that also appear in the newly downloaded data for 2017, while red points are those that were missing. The map was created from SIG-OT shapefiles [117].

The 2_KRG_predict_tmean.R file in the above GitHub repository was used to perform the kriging on the mean temperature data for 2016 and 2017, resulting in one .bil file for each day. These files were read back into R as rasters. The daily data were aggregated at weekly time steps by taking the average of each consecutive seven raster layers. Next, the WorldPop layer was subset and resampled to match the spatial extent and resolution of the temperature layers (4.65 km x 4.65 km resolution). The weekly temperature layers were then multiplied by the WorldPop layer as in [207]. The resulting weekly layers and the WorldPop layer were exported to Python (version 3.8.2).

In Python, the QgsZonalStatistics.Sum function from the qgis.analysis module was used to spatially aggregate the temperature data. This function calculates the sum of the raster values for a polygon and appends the results as attributes. For the city level, shapefiles from Colombia's Sistema de Información Geográfica para la Planeación y el Ordenamiento Territorial (SIG-OT) were used for the year 2018 [117]. To be consistent with the data from Siraj et al., in which the capital of Bogotá is separate from the department of Cundinamarca, department level shapefiles were used from the Humanitarian Data Exchange [209] (the only shapefiles currently available from SIG-OT combine the two). Spatial aggregation by

summing was performed on the weekly temperature layers that had been multiplied by the WorldPop layer as well as the WorldPop layer. The resulting shapefiles were re-imported into R, where the aggregated mean temperature values were divided by the aggregated population values for each spatial scale.

The new 2016-2017 population weighted weekly time series of mean temperature were compared to the 2014-October 2016 data from Siraj et al. There was good agreement for 2016, and seasonal trends were consistent in the 2017 data. The final temperature datasets consist of the mean temperature from Siraj et al. for 2014-October 2016 and the newly processed data for the remaining two months of 2016 and 2017.

3.1.2 Precipitation

Daily satellite precipitation data were read into R as rasters. Missing values were re-coded from -99 to 0. As with the temperature data, the spatial extent and resolution of the WorldPop layer was subset and resampled to match the precipitation data (4 km x 4 km resolution). Next, the daily data were aggregated at weekly time steps by taking the sum of each consecutive seven raster layers, resulting in cumulative precipitation for each week (if the missing values are not re-coded, then negative rainfall can be obtained for some locations and weeks as a result of this step). These layers were multiplied by the WorldPop layer and were exported to Python.

Spatial aggregation of the precipitation data was performed with the same shapefiles and Python code as the temperature data. The final precipitation datasets consist of only the population weighted weekly time series generated from the PERSIANN-CCS data. The new data overlapped well with the corresponding time series from Siraj et al., except for the aforementioned outliers.

3.2 Inclusion criteria and level of analysis determination

In order to decide whether to perform the analysis at the city or department level, it was necessary to develop inclusion criteria for geographic location. Departments that reported at least 50 cases of CF or 50 cases of ZVD were considered. The use of this cut-off resulted in 29 departments for CHIKV (Amazonas, Guainía, and Vaupés were dropped) and 30 departments for ZIKV (Guainía and Vaupés were dropped). At the city level, cities that

reported fewer than 20 cases of CF or cities with fewer than six weeks of observations were dropped, leaving 321 cities for CHIKV. Similarly, cities that reported fewer than 30 cases of ZVD or cities with fewer than six weeks of observations total were dropped, leaving 288 cities for ZIKV. Time series with more weekly cases and fewer gaps are preferred as more precise estimates of the reproduction numbers can be obtained from them [164].

As the impact of weather on arbovirus transmission was of primary interest, the between- and within-department standard deviations of weather covariates were calculated over relevant time periods for CHIKV and ZIKV. Following Harris et al.'s work on ZIKV transmission in Latin America, temperature was defined as mean weekly temperature in °C averaged over three weeks followed by a time lag of six weeks prior to case reporting. They defined rainfall as cumulative weekly rainfall in mm summed over six weeks followed by a three-week lag using the rationale that a larger window for rainfall compared to temperature could better account for the effect of water accumulation over time [202]. For within-department standard deviation, the variance was calculated first: temperature (or rainfall) at the department level was subtracted from temperature (or rainfall) at the city level and squared for each week. The results were added together and divided by the number of observations (number of cities within each department times number of weeks) minus 1. Similarly, to calculate between-department variance, temperature (or rainfall) at the national level was subtracted from temperature (or rainfall) at the department level, squared, summed together, and divided by the number of observations minus 1. Temperature and rainfall at the national level was first computed as the mean and sum, respectively, across all departments on a weekly basis before taking the three-week average, or six-week sum, and accounting for the time lags. For CHIKV and ZIKV, the mean between-department standard deviation across departments was greater than the mean within-department standard deviation for both temperature and rainfall, making the department level the preferred spatial scale (Table 3.4).

Table 3.4 Mean within- and between-department standard deviation (sd) of temperature and rainfall for CHIKV and ZIKV.

Virus	Temperature (°C)	Rainfall (mm)
CHIKV	Mean within-department sd: 2.65 Mean between-department sd: 3.39	Mean within-department sd: 120 Mean between-department sd: 8,218
ZIKV	Mean within-department sd: 2.27 Mean between-department sd: 3.39	Mean within-department sd: 96 Mean between-department sd: 8,819

3.3 Weekly time-varying reproduction numbers from EpiEstim

The EpiEstim package in R was used to estimate weekly R_ts for the CHIKV and ZIKV epidemics at the department level over 5-week sliding windows [164]. Time windows of 2, 3, 4, 5, 6, 7, and 8 weeks were considered, and a sliding window of 5 weeks was chosen for both viruses, which was also used by Ferguson et al. for the ZIKV epidemic in the Americas [210]. The "parametric_si" method was selected in the function estimate_R() using estimates of the generation time distributions obtained from the literature (Table 3.5), and the default prior distribution for R_t with mean 5 and standard deviation 5 was used. R_t estimates were removed if the mean was greater than 100 or highly uncertain (coefficient of variation greater than 0.5).

The generation time distribution for an arbovirus consists of two main components: the human-to-mosquito generation time distribution and the mosquito-to-human generation time distribution [211]. There are few estimates of the generation time distribution for CHIKV and ZIKV in the literature, partly due to the types of data sources required. Data to estimate the generation time distribution include viral load data in humans, human case reports with time of exposure and symptom onset (typically for travelers), mosquito daily mortality rate, and the number of mosquitoes testing positive each day after being experimentally infected with the virus in the laboratory [211]. The serial interval distribution can be estimated using similar data sources or, alternatively, as the difference in symptom onset between isolated pairs of confirmed cases [212]. The serial interval distribution can also be estimated as a function of temperature, as temperature affects parameters related to the mosquitoes, such as the mosquito mortality rate [199]. Table 3.5 includes estimates of the generation time distribution and serial interval distribution of CHIKV and ZIKV from the literature.

Table 3.5 Estimates of the mean and standard deviation of the generation time distribution (GTD) and serial interval distribution (SID) for CHIKV and ZIKV from the literature. All values have units in days. The estimates used in this thesis are in bold.

Virus	GTD or SID	Mean	Standard deviation	Reference
CHIKV	**GTD**	**14.0**	**6.2**	**[213]**
CHIKV	SID	23	6	[214]
CHIKV	SID	Range: 10.5-18.9	Not reported	[199]
CHIKV	SID	11.2	4.2	[215]
ZIKV	**GTD**	**20.0**	**7.4**	**[210]**
ZIKV	SID	7.4 (95% CI: 4.6-10.2)	Not reported	[212]
ZIKV	SID	Range: 10-23	Not reported	[216]
ZIKV	SID	Range: 15.4-32.9	Not reported	[199]
ZIKV	SID	17.5	4.9	[215]

3.4 Non-parametric models of arbovirus transmission

For each epidemic, linear regression models were fitted to the relationship of the median R_ts obtained from EpiEstim and the cumulative incidence at the end of each sliding window divided by the population size of the department. In theory,

$$R_t = R_0 \times \left(1 - \frac{sum(I)}{N}\right) \tag{3.3}$$

where I is the incidence of disease, and N is the population size. However, not all cases were observed in the epidemics. Therefore, equation (3.3) becomes

$$R_t = R_0 \times \left(1 - \frac{\frac{sum(I_{obs})}{N}}{\rho}\right) \tag{3.4}$$

where I_{obs} is the observed cases, and ρ is the proportion of infections that are reported as cases to the surveillance system (reporting rate). The first week of cases was not included in the calculation of cumulative incidence as this week is not used by EpiEstim to estimate R_t. When plotting R_t against the observed attack rate ($\frac{sum(I_{obs})}{N}$), the y-intercept of the linear regression models can be interpreted as a rough approximation of R_0 across all departments (equation (3.5)), and the ratio of the estimated R_0 by the absolute value of the slope is a rough approximation of ρ (equation (3.6)):

$$R_0 \cong intercept \tag{3.5}$$

$$\rho \cong \frac{R_0}{abs(slope)} \tag{3.6}$$

Linear regression was performed using the lm function in R. The 95% confidence interval for the y-intercept was obtained by using the confint function on the model output. The 95% confidence interval for the ratio of the y-intercept and minus the slope was obtained by using the Delta Method [217]. Because these quantities are not independent, the point estimate was calculated as the absolute value of

$$E\left(\frac{X}{Y}\right) \equiv E\big(f(X,Y)\big) \approx \frac{\mu_X}{\mu_Y} - \frac{\mathrm{Cov}(X,Y)}{(\mu_Y)^2} + \frac{\mathrm{Var}(Y)\mu_X}{(\mu_Y)^3} \tag{3.7}$$

where X and Y are random variables representing the y-intercept and slope respectively, Cov is the covariance of the variables, Var is the variance, and μ is the mean. The variance-covariance matrix was obtained by using the vcov function. To calculate the variance of the ratio, equation (3.8) was used,

$$\mathrm{Var}\left(\frac{X}{Y}\right) = \frac{(\mu_X)^2}{(\mu_Y)^2}\left[\frac{\mathrm{Var}(X)}{(\mu_X)^2} - 2\frac{\mathrm{Cov}(X,Y)}{\mu_X\mu_Y} + \frac{\mathrm{Var}(Y)}{(\mu_Y)^2}\right] \tag{3.8}$$

The 95% confidence interval was then calculated as the point estimate $\pm\ 1.96 * \sqrt{\mathrm{Var}(\frac{X}{Y})}$.

The residuals from the linear regression models were then plotted against covariates thought to be associated with arbovirus emergence and spread, including mean temperature and cumulative rainfall as well as the percentage of households in each department with overcrowded conditions, inadequate materials for exterior walls, and inadequate floors. Time lags were considered for temperature and rainfall due to the delay between changes in weather and resulting changes in transmission [202]. The delay is typically assumed to vary between one and two months [199, 202]. Temperature and rainfall were defined as in section 3.2. GAMs were fitted to the plots of the residuals versus covariates using the R package mgcv (version 1.8-28) [218]. Smoothness estimation was done by restricted maximum likelihood by selecting “REML” for the method. Smooth terms for each predictor were added to the models one at a time and kept if they were significant at the 0.05 level. The estimated degrees of freedom (edf) describe the curviness of the lines; higher values of edf are associated with more bends, while edfs close to 1 approximate linear terms. The basis dimensions (k) for smooth terms were adjusted as needed. k sets the

upper limit on the degrees of freedom associated with a smooth term. A low k value is indicated by a low p value in the model diagnostics output. If k is too small, it can force oversmoothing [219]. AIC was used to compare models. AIC is a method of determining a model's predictive accuracy. Lower values are preferred, and a difference of more than two is considered important [220].

3.5 Parametric models of arbovirus transmission

In the parametric model approach, the renewal equation was used to model the number of reported cases of CF and ZVD separately according to a Poisson distribution:

$$I_{t,i} \sim P\left(R_{t,i}\sum_{s=0}^{t}\left[I_{s,i}w_{t-s}\right]\right) \tag{3.9}$$

where $I_{t,i}$ is the number of reported cases in location *i* at time *t*, $R_{t,i}$ is the time-varying reproduction number in location *i* at time *t*, w is the generation time distribution (see section 3.3). The methods underlying the models described in this section are the same as those used by EpiEstim with the difference that R_t is parametrically constrained. Following Riou et al. [199], a negative binomial model was also considered to account for overdispersion in the data (as shown by equation (3.10)),

$$I_{t,i} \sim NegBin\left(R_{t,i}\sum_{s=0}^{t}\left[I_{s,i}w_{t-s}\right], \phi\right) \tag{3.10}$$

where ϕ is the overdispersion parameter. Overdispersion is caused by variation in the number of secondary cases resulting from each case of CF or ZVD [221]. As ϕ increases (> 5), the discrete probability distribution of the negative binomial model converges on a Poisson distribution [176]. R_t is influenced by (i) saturation due to the rise in cumulative cases and decline in the number of susceptible individuals, (ii) environmental conditions, namely temperature and rainfall, and (iii) socioeconomic factors. Saturation, denoted by α, is represented by equation (3.11),

$$\alpha_{i,t} = 1 - \frac{\sum_{s=0}^{t} I_{s,i}}{N_i\,\rho} \tag{3.11}$$

where ρ is the reporting rate and N_i is the population size at location *i*. Simple gaussian (symmetric or asymmetric) functions were considered for the dependence of transmissibility on rainfall and temperature. The effects of temperature and rainfall on transmission (β) are represented by equation (3.12),

$$\beta_{i,t} = \beta_{i,t}^{T}\beta_{i,t}^{R} \tag{3.12}$$

$\beta_{i,t}^{T}$ or $\beta_{i,t}^{R}$ is equal to $\beta_{i,t}^{X}$:

$$\beta_{i,t}^{X} = \exp\left(-\frac{\left(X_{i,t} - X^{best}\right)^2}{(2 * \sigma^2)}\right) \tag{3.13}$$

where X^{best} is the optimal weather condition (either temperature or rainfall) for transmission, σ_X is the standard deviation associated with this optimal condition, and $f\left(X_{i,t};\ \mu = X^{best}, \sigma_X\right)$ is the density of the normal distribution at condition $X_{i,t}$. At the optimal temperature (or rainfall) for pathogen transmission, β = 1. At less optimal weather conditions, 0 $\leq \beta <$ 1. For models with a single standard deviation for temperature or rainfall, transmission declines symmetrically above and below the optimal weather condition. Models that included two standard deviations for temperature were also considered in which transmission declines asymmetrically around the optimal temperature.

$$\begin{aligned}
\beta_{i,t}^{X(temp)} &= \exp\left(-\frac{\left(X_{i,t}^{(temp)} - X^{best(temp)}\right)^2}{\left(2*\sigma_{temp1}^2\right)}\right) \text{ if } X_{i,t}^{(temp)} < X^{best(temp)} \\
\beta_{i,t}^{X(temp)} &= \exp\left(-\frac{\left(X_{i,t}^{(temp)} - X^{best(temp)}\right)^2}{\left(2*\sigma_{temp2}^2\right)}\right) \text{ if } X_{i,t}^{(temp)} \geq X^{best(temp)}
\end{aligned} \tag{3.14}$$

Unlike weather covariates, which are believed to have a unimodal relationship with arbovirus transmission (see section 1.3), there is evidence that worse housing conditions are associated with greater risk of arbovirus infection [222, 223]. Consequently, socioeconomic status covariates represented by γ were considered in the following way:

$$\gamma_i = e^{\sum_i^K b_i(X_i - \bar{X})} \tag{3.15}$$

with $\bar{X}$ denoting the mean value of X across the dataset. This expression assumes a log-linear effect of the covariates on transmissibility. Unlike β, γ is not limited to 0 and 1, but it must be positive, and γ = 1 when $X_i = \bar{X}$. The final model for R_t is shown by equation (3.16):

$$R_{t,i} = R_0 \, \alpha_{i,t} \, \beta_{i,t} \, \gamma_i \tag{3.16}$$

The mean of the Poisson and negative binomial models equals transmissibility multiplied by infectiousness (equations (3.9) and (3.10)). In other words, the number of newly reported cases is modeled as the number of past and current cases weighted by their infectiousness (see section 1.1).

For the Poisson models, the following parameters were estimated: R_0, ρ, $X^{best(temp)}$, $X^{best(rain)}$, standard deviation of temperature (σ_{temp}), standard deviation of rainfall (σ_{rain}), $b_{i(inadequate\ walls)}$, and $b_{i(overcrowding)}$. In models with two standard deviations for temperature, σ_{temp1} is the standard deviation of temperatures below the optimal temperature, and σ_{temp2} is the standard deviation of temperatures that are greater than or equal to the optimum. Different definitions of and time lags associated with temperature and rainfall were tested separately in addition to those used in the covariate exploration. For temperature, mean weekly temperature in °C averaged over i) five weeks, ii) four weeks, and iii) three weeks as well as mean weekly temperature averaged over three weeks followed by a iv) three-week lag and v) four-week lag, all prior to case reporting, were modeled. For rainfall, cumulative weekly rainfall in mm summed over six weeks followed by a two-week lag prior to case reporting was tested. After the best-fitting definition of each weather covariate was identified, parameters related to temperature and rainfall were estimated together in the same models. For the negative binomial models, R_0, ρ, and ϕ were estimated along with the weather and socioeconomic status covariates. Both Poisson and negative binomial models that estimated a different R_0 for each department were also considered.

The CHIKV data for some departments, particularly Antioquia, Caldas, and Nariño, suffer from substantial censoring, because the epidemics were underway for some time before cases were reported [170]. This issue can pose problems for estimating reproduction numbers. Another department, Chocó, had a jagged epidemic curve, which can cause similar problems. To address these issues and smooth the reproduction number estimates,

a threshold was used to prevent outliers in the distribution of incidence divided by infectivity from contributing to the likelihood. This ratio is given by equation (3.17),

$$incidence - to - infectivity = \frac{I_{t,i}}{\sum_{s=0}^{t}\left[I_{s,i}w_{t-s}\right]} \tag{3.17}$$

Weeks in which the incidence-to-infectivity ratio was above 20 were ignored. This threshold was determined by examining the distribution of the ratios (Figure 3.2). Ninety-nine percent of the values for both viruses were below 55; 98% of the values were below 20, and 97% were under 15. Two sensitivity analyses were performed. First, model fits of the best-fitting Poisson and negative binomial models were assessed with and without the thresholds. Secondly, model fits of the overall best-fitting models were assessed using thresholds of 15 and 55.

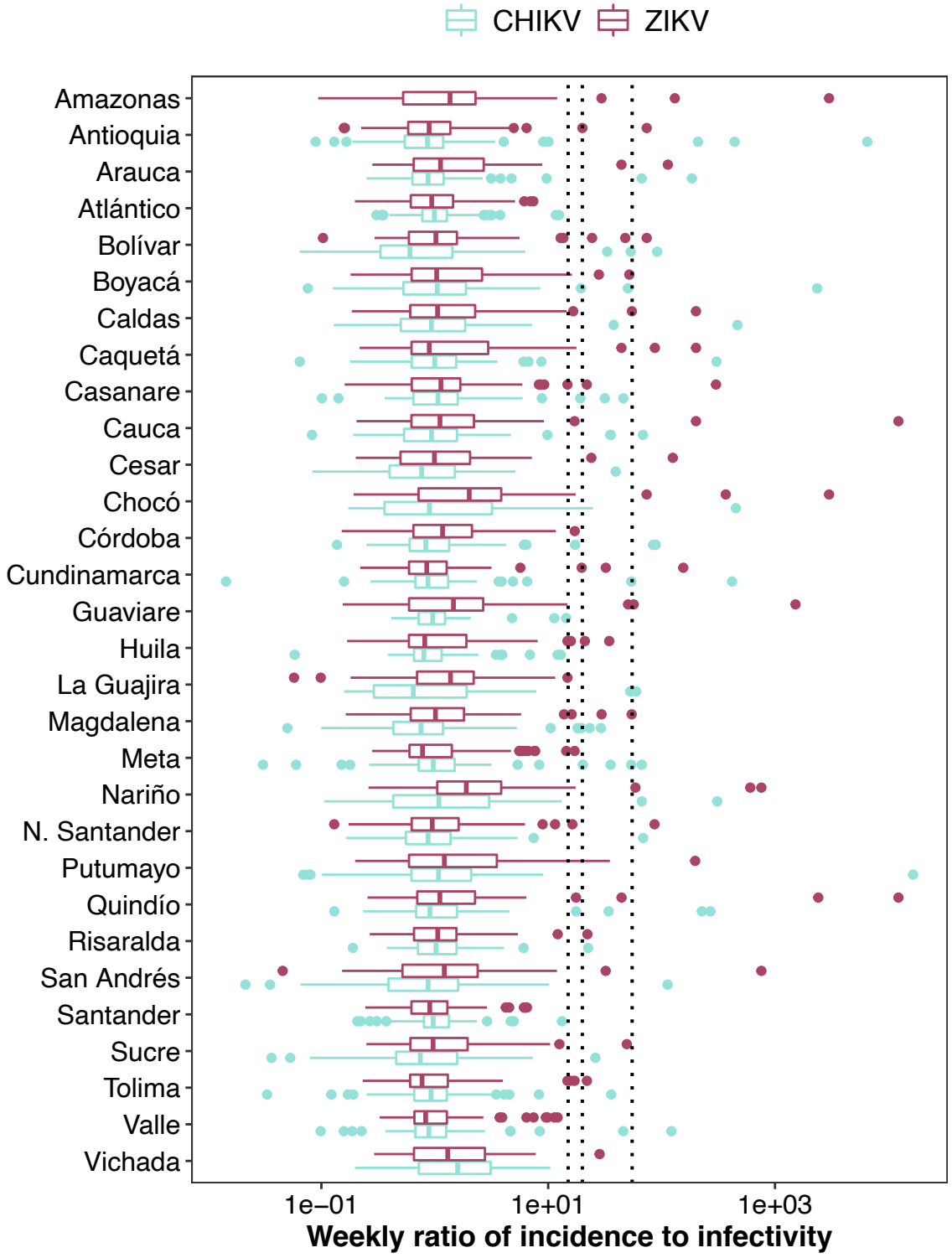


Figure 3.2 Boxplots of incidence divided by infectivity in each department for all weeks for CHIKV and ZIKV. The x-axis is plotted on a $\log_{10}$ scale. The black dotted lines indicate the thresholds of 15, 20, and 55. Weeks in which the ratio was above 20 (to the right of the middle black line) did not contribute to the likelihood in the main analysis.

3.6 Model estimation and computing

Metropolis-Hastings MCMC was used to estimate the model parameters for the parametric approach described in section 3.5 [224, 225]. Parameter values were sampled from a log normal distribution, and the Metropolis accept-reject rule was corrected for the asymmetry of the proposal distribution. Parameters were updated one at a time. Each model was run three times with different starting values, and chains were visually checked for convergence after 100,000 iterations with a burn in of 0.2 times the length of the chains (iterations times number of parameters). After removing the burn-in period, median parameter estimates and 95% credible intervals were calculated from the posterior distributions.

The coda package (version 0.19-4) in R was used to calculate the Gelman-Rubin statistic for each best-fitting model. This statistic assesses model convergence by comparing the variance between- versus within-MCMC chains. Lack of convergence is indicated by values above one [86, 226]. The coda package was also used to calculate the effective sample size using the effectiveSize function. The effective sample size is the sample size of the parameters adjusted for autocorrelation [226]. Dependence in MCMC samples tends to be more important for complex models; as more parameters are estimated, larger sample sizes are required to approximate the posterior distribution. For these models, an effective sample size of less than 10% of the actual sample size is not uncommon, but even this level can indicate problems with a particular model [84].

Uniform prior distributions were used for all parameters (Table 3.6). The range of the prior distributions for $X^{best(temp)}$ and $X^{best(rain)}$ were informed by the range of the data and mosquito biology. The lower limits on the prior distributions for ρ were calculated for each virus as the maximum of the cumulative incidence in department *i* divided by the population size of department *i* for all departments (the maximum observed attack rate). Values of ρ lower than the limits would result in negative values of α for some locations which means that the number of infections exceeds the population size for a given ρ. The maximum observed attack rate of CHIKV was observed in Casanare with about 0.037. For ZIKV, San Andrés and Providencia had the highest observed attack rate with about 0.015 (Table 3.6).

Model comparison was performed with DIC. Lower values of DIC indicate better model fit, and a difference of about 5 is important [227]. DIC was calculated using the medians of the

posterior distributions of the parameters due to non-normality of the likelihood. Although model comparison statistics other than DIC were not examined, AIC, Widely Applicable Information Criterion, and leave-one-out information criterion could have been used in conjunction with DIC or as an alternative. DIC was emphasized as it is easy to calculate when using MCMC sampling and widely used.

All analyses were performed in R version 4.0.3.

Table 3.6 Prior distributions for parameters in the parametric model based on the renewal equation.

Parameter	Prior distribution	Range
R_0	Uniform	(0, 10)*
ρ (reporting rate)	Uniform	CHIKV: (0.037, 1)** ZIKV: (0.015, 1)**
$X^{best(temp)}$	Uniform	(0, 50)
σ_{temp1} (< optimum)	Uniform	(0, 50)
σ_{temp2} (≥ optimum)	Uniform	(0, 50)
$X^{best(rain)}$	Uniform	(0, 1000)
σ_{rain}	Uniform	(0, 1000)
$b_{i(inadequate\ walls)}$	Uniform	(0, 100)
$b_{i(overcrowding)}$	Uniform	(0, 100)
ϕ	Uniform	(0, 10)

*For the negative binomial models with multiple R_0s, the upper limits on the prior distributions were increased to 15 to account for overdispersion in the data.
**Lower limit rounded to nearest thousandth.

3.7 Validation of parametric model fit and parameter fitting procedure

Parametric model fits were validated by comparing the R_ts for each department and virus with the R_ts obtained from EpiEstim. One R_t matrix was calculated by the model for each of 100,000 iterations. After removing a burn-in of 20,000, the remaining matrices were thinned by taking every 80th matrix (keeping 800 total). The median values of the thinned matrices were plotted with the median R_t estimates from EpiEstim. EpiEstim R_ts were plotted in the center of the 5-week sliding window used to compute each estimate.

As an additional check, the infection attack rates for CHIKV and ZIKV were estimated for each department and compared to available seroprevalence estimates using the estimate for ρ from the best-fitting models and the cumulative incidence. As the seroprevalence estimates from Nouvellet et al. [191] correspond to the city level, the seroprevalence data

were compared to the estimated infection attack rate for the department in which the city is located.

The parameter fitting procedure was validated by creating one simulated dataset for each of two Poisson models (the Poisson model with weather covariates and the Poisson model with multiple R_0s and rainfall). A third simulated dataset was created from a negative binomial model with multiple R_0s. The simulated datasets used observed population sizes for 29 departments, observed weather data (for the Poisson models), and the generation time distribution for CHIKV. The inference procedure was re-run on the simulated datasets to check that the true parameter values could be recovered.

3.8 Comparing parameter estimates across departments

The posterior probability that the R_0 value for CHIKV was greater than the R_0 value for ZIKV (and vice versa) was estimated for each department. Each best-fitting model was run for 500,000 iterations. After removing the burn-in (0.2 times the number of parameters times the number of iterations), the chains were thinned by saving every 100th value. For CHIKV, this resulted in 124,000 posterior samples for each parameter. There were 128,000 posterior samples for ZIKV. The last 100,000 samples of each R_0 were taken, and as above, the proportion of times that the estimated R_0 was higher for CHIKV than ZIKV (and vice versa) was determined.

The same method was used to compare ρ for CHIKV and ZIKV as well as the modeled attack rates.

4 Results

4.1 Weather data

Figures 3.3-3.4 show the weekly time series of population weighted mean temperature and cumulative precipitation, respectively, aggregated at the department level. While temperature is relatively constant throughout the year in most departments, there are strong seasonal trends in rainfall.

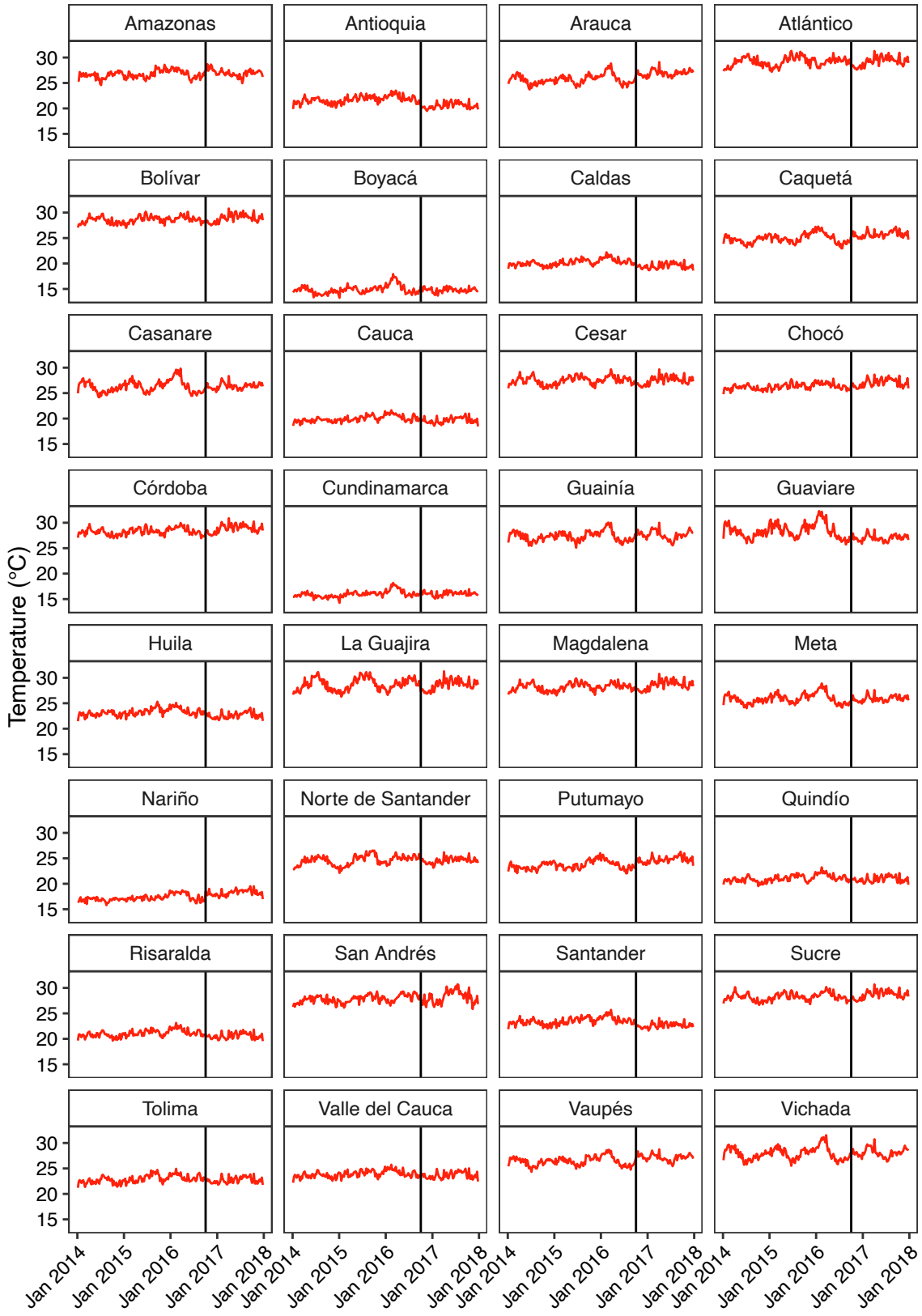


Figure 3.3 Weekly time series of population weighted mean temperature in °C by department. Data before the vertical black lines are from [207]. Bogotá (not shown) was considered separately from Cundinamarca.

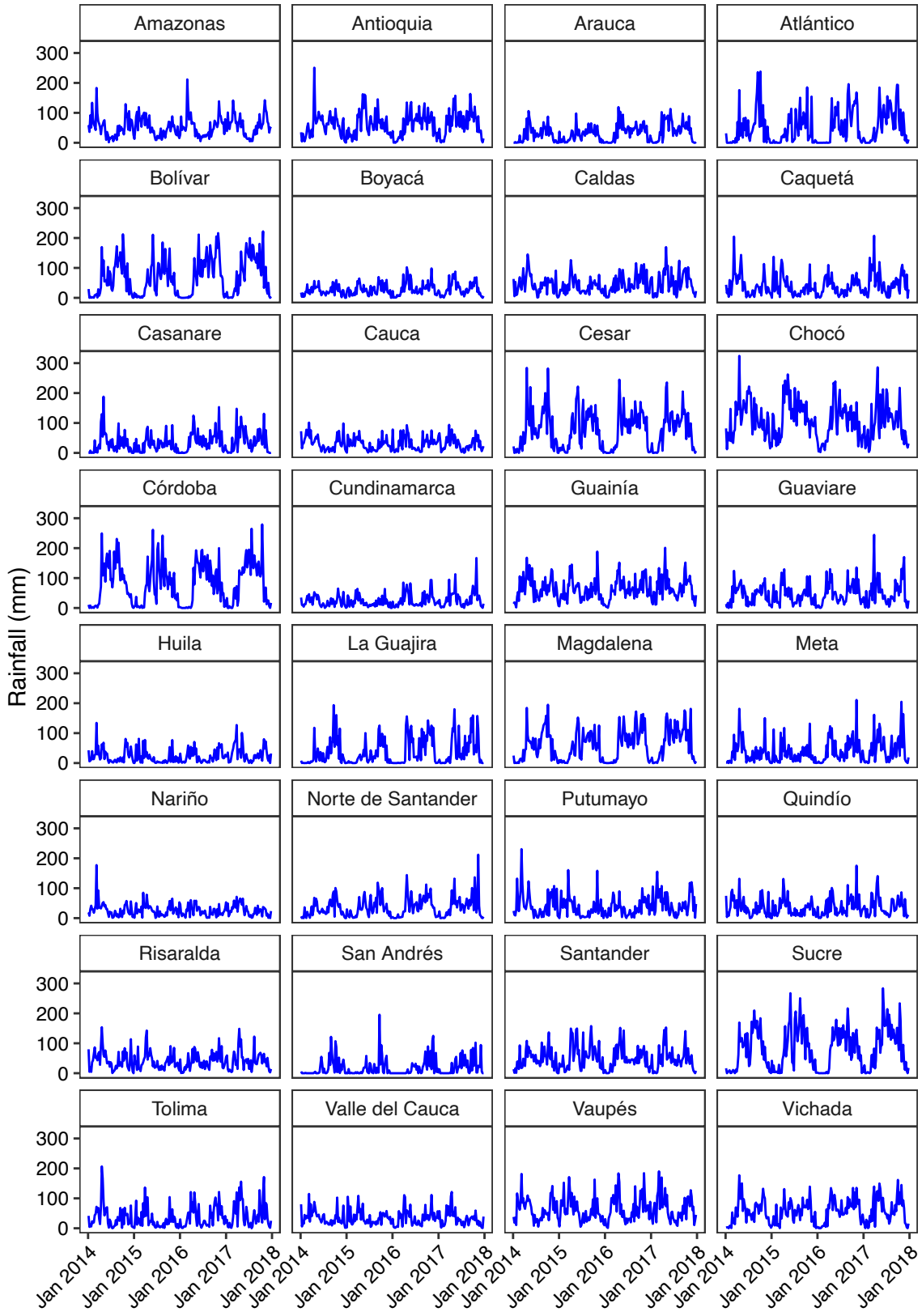


Figure 3.4 Weekly time series of population weighted cumulative precipitation in mm by department. Bogotá (not shown) was considered separately from Cundinamarca.

4.2 Epidemiological data

Table 3.7 shows the total number of suspected and laboratory-confirmed cases of CF and ZVD reported to Sivigila from June 2014-June 2017. Most departments did not vary greatly in the proportion of cases reported across diseases. Valle del Cauca reported more cases than any other department (27% and 26% for CF and ZVD, respectively). As mentioned in chapter 1, cases that had Bogotá recorded as the location of likely infection were considered extremely unlikely and were removed from the data.

Table 3.7 Cumulative incidence of suspected and laboratory-confirmed cases of CF and ZVD in Colombia, 2014-2017.

		CF cases		ZVD cases	
Department	**Population in 2016**	**Number**	**%**	**Number**	**%**
Amazonas	77,088	10	<1%	342	<1%
Antioquia	6,534,764	13,911	3%	2,523	2%
Arauca	265,190	5,335	1%	1,875	2%
Atlántico	2,489,709	12,381	3%	6,778	6%
Bogotá	7,980,001	0	0%	0	0%
Bolívar	2,122,021	22,810	6%	1,914	2%
Boyacá	1,278,061	398	<1%	366	<1%
Caldas	989,942	2,886	1%	334	<1%
Caquetá	483,834	6,404	2%	1,142	1%
Casanare	362,698	13,366	3%	3,942	4%
Cauca	1,391,889	3,253	1%	355	<1%
Cesar	1,041,203	3,813	1%	1,605	2%
Chocó	505,046	479	<1%	65	<1%
Córdoba	1,736,218	16,932	4%	3,327	3%
Cundinamarca	2,721,368	17,464	4%	5,272	5%
Guainía	42,123	10	<1%	14	<1%
Guaviare	112,621	1,867	<1%	208	<1%
Huila	1,168,910	29,493	7%	6,966	7%
La Guajira	985,498	10,464	3%	699	<1%
Magdalena	1,272,278	9,935	2%	3,232	3%
Meta	979,683	18,020	4%	4,323	4%
Nariño	1,766,008	2,179	1%	95	<1%
Norte de Santander	1,367,716	29,137	7%	10,361	10%
Putumayo	349,537	894	<1%	502	<1%
Quindío	568,473	4,265	1%	406	<1%
Risaralda	957,250	5,266	1%	1,296	1%
San Andrés and Providencia	77,101	1,482	<1%	1,148	1%
Santander	2,071,044	10,900	3%	10,374	10%
Sucre	859,909	20,636	5%	1,639	2%
Tolima	1,412,230	38,901	9%	7,122	7%
Valle del Cauca	4,660,438	109,935	27%	27,712	26%
Vaupés	44,079	0	0%	18	<1%
Vichada	73,702	89	<1%	78	<1%
Total	48,747,632	412,915	100%	106,033	100%

4.3 Weekly time-varying reproduction numbers from EpiEstim

A total of 1,938 weekly estimates of R_t were obtained from EpiEstim for CHIKV, while 1,977 were obtained for ZIKV. The number of R_t estimates for each virus is shown in Figure 3.5 as well as the number of weeks during which estimated $R_t > 1$. For CHIKV, Valle del Cauca and

Córdoba had the greatest number of R_t estimates with 87 each. For ZIKV, Norte de Santander had the greatest number of R_t estimates with 90, followed by Atlántico with 87. Meta and Santander had the greatest number of weeks where estimated $R_t > 1$ for CHIKV with 39 each. For ZIKV, Casanare had the greatest number of weeks where estimated $R_t > 1$ with 45.

The number of weeks with R_t estimates varied by department because the duration of the epidemic varied in each location; for each department, R_t was estimated from the week when cases were first reported until the week when cases were no longer reported. Departments could also lose R_t estimates for certain weeks if the estimates were highly uncertain (see section 3.3).

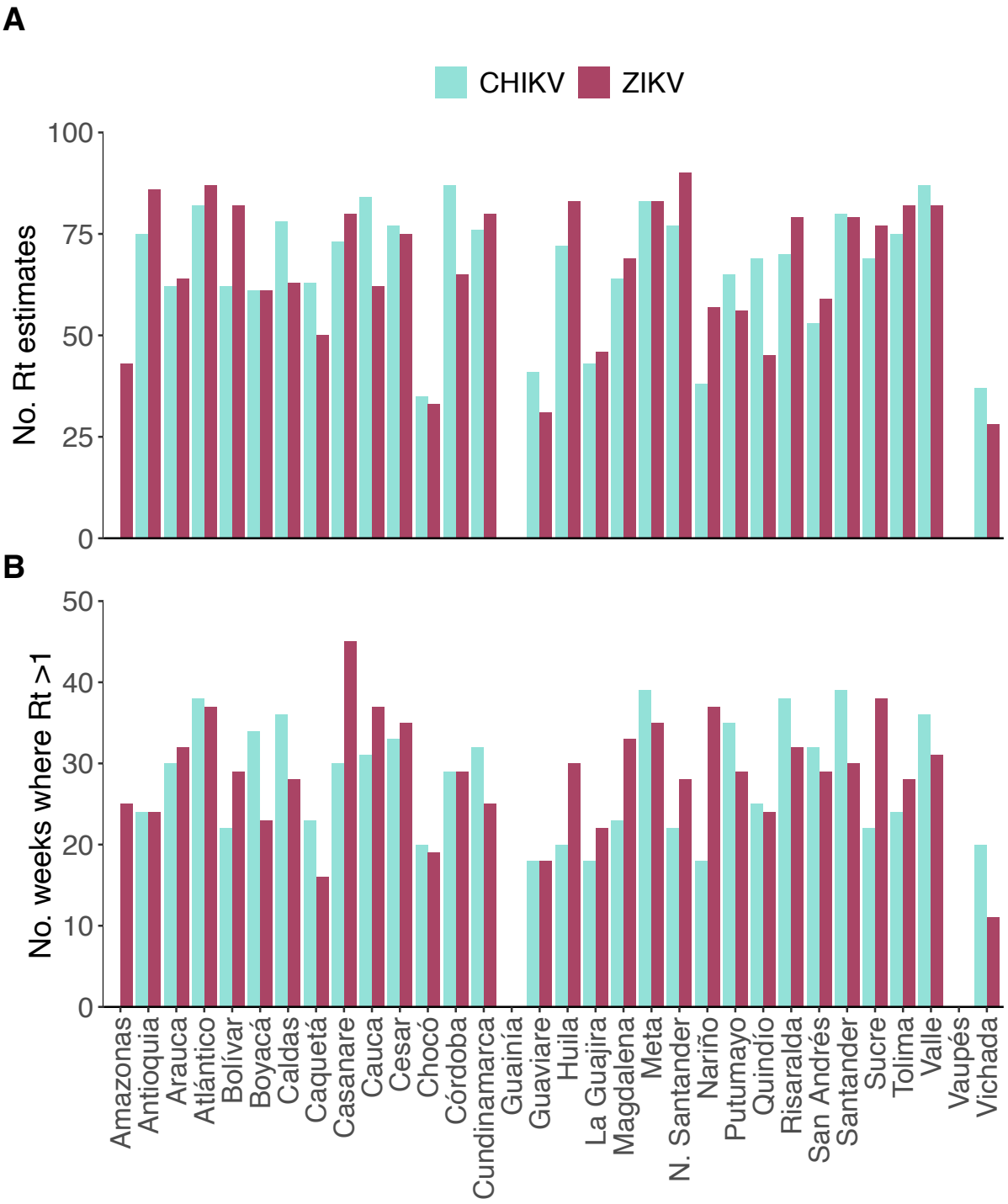


Figure 3.5 Histograms showing *(A)* the number of R_t estimates and *(B)* the number of weeks where estimated R_t > 1 by department.

4.4 Non-parametric models of arbovirus transmission

Figure 3.6 shows the plots of the EpiEstim R_t estimates versus the cumulative incidence of reported cases divided by the population of each department with fitted linear regression

lines. The y-intercept from the linear regression model, which is a rough approximation of the overall R_0, was 1.71 (95% CI: 1.54-1.88) for CHIKV and 1.69 (95% CI: 1.59-1.78) for ZIKV. The slope was -39.8 (95% CI: -53.2 - -26.4) for CHIKV and -113.0 (95% CI: -135.8 - -90.2) for ZIKV. For CHIKV, the corresponding estimate of the reporting rate was 0.044 (95% CI: 0.032-0.056), and for ZIKV, it was 0.015 (95% CI: 0.012-0.018).

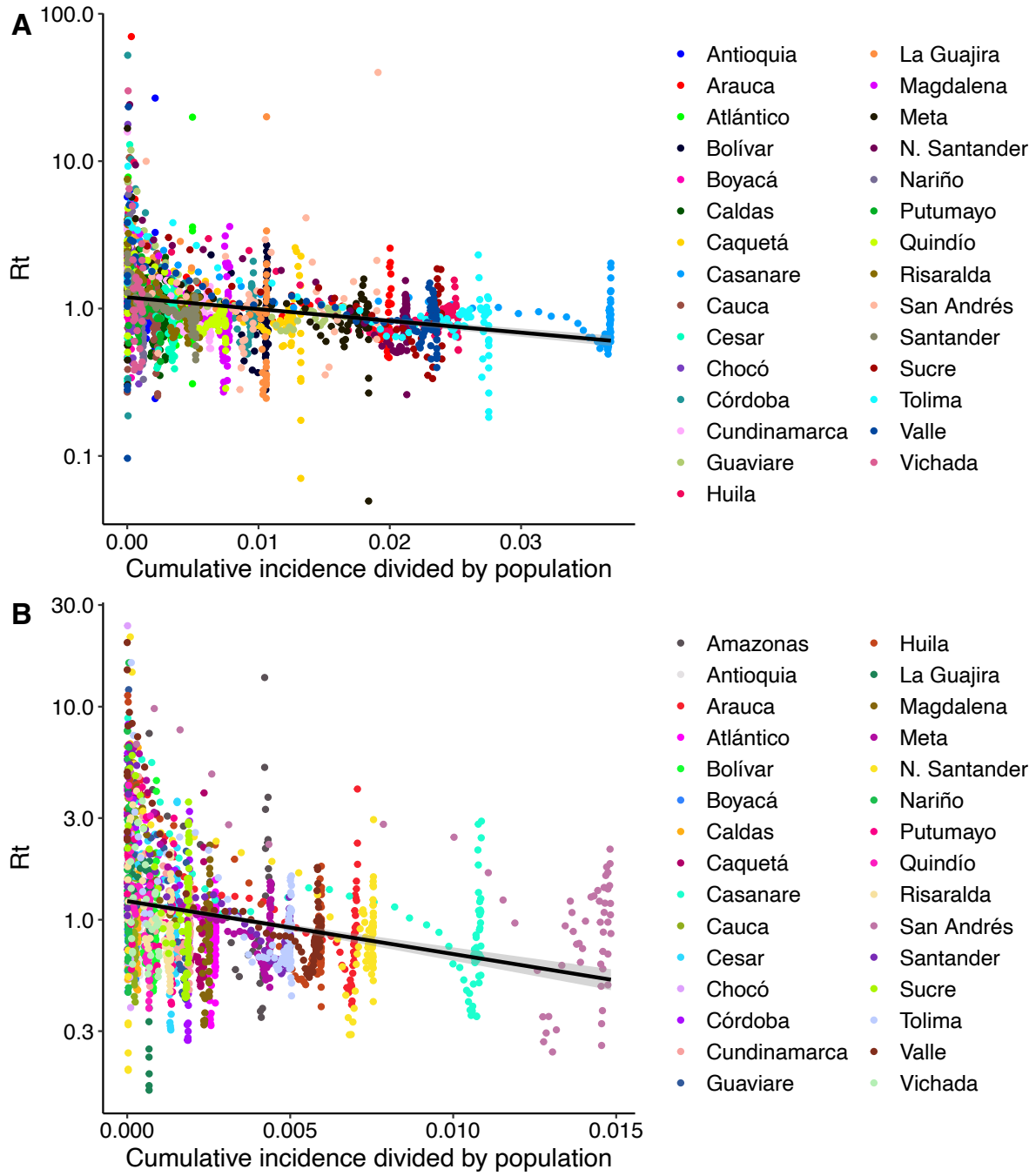


Figure 3.6 R_t estimates versus cumulative incidence divided by population size of each department. *(A)* CHIKV and *(B)* ZIKV. The black line is the fitted linear regression line, and the shaded line represents the 95% CI. The y-axis is plotted on a $\log_{10}$ scale to better visualize trends.

Linear regression models were also fitted to the relationship between the EpiEstim R_ts and cumulative incidence divided by population separately for each department (Figures 3.7-3.8). The estimated R_0s across departments ranged from 0.13 to 8.07 for CHIKV and from 1.59 to 8.22 for ZIKV, while the estimated ρs across departments ranged from 0.001 to 0.053 for CHIKV and from <0.001 to 0.017 for ZIKV. The uncertainty of these estimates was high, as indicated by the 95% confidence intervals in Figures 3.7-3.8.

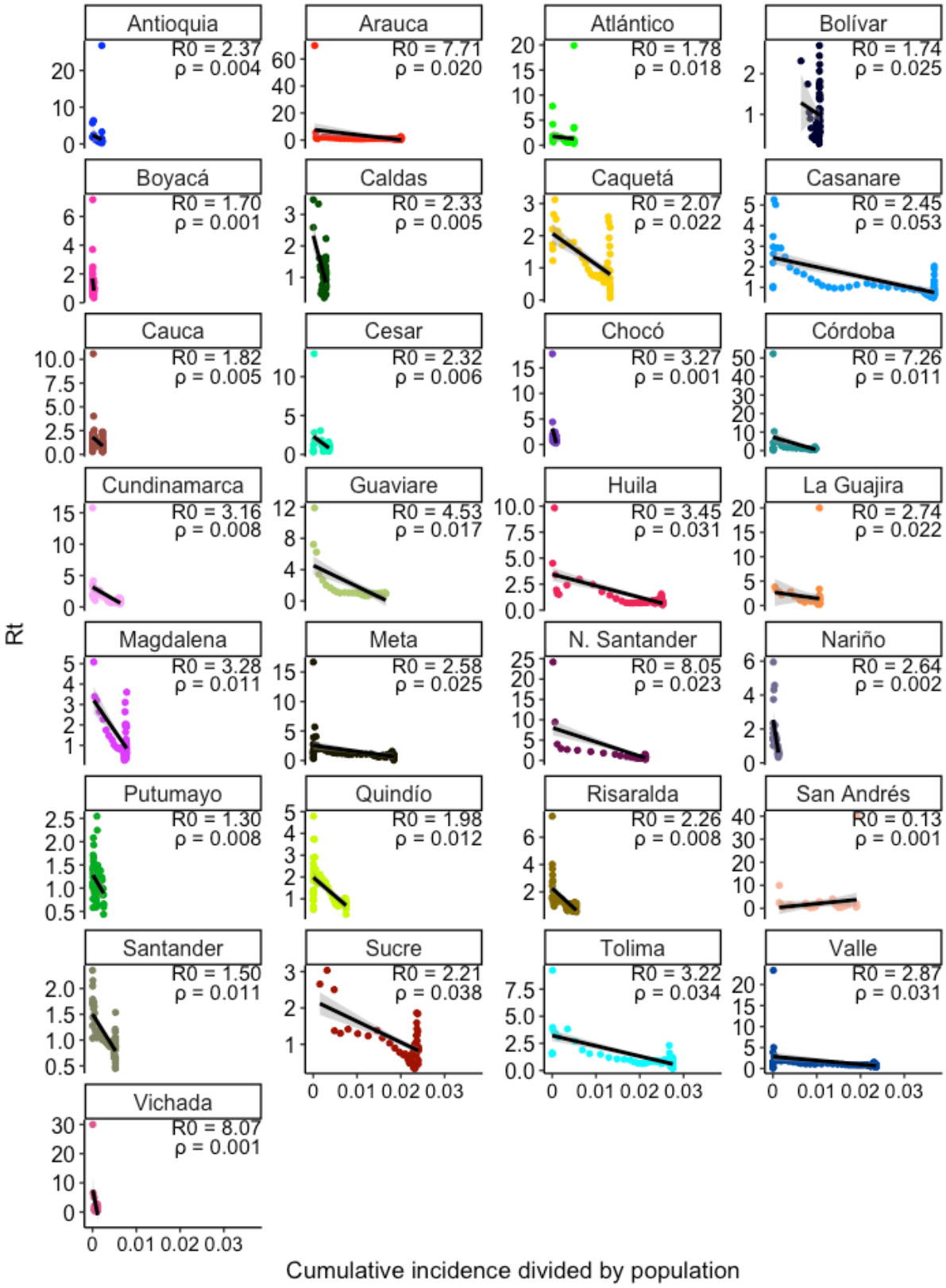


Figure 3.7 R_t estimates versus cumulative incidence divided by population size by department for CHIKV. The black line is the fitted linear regression line, and the shaded line represents the 95% CI. The estimated R_0 and ρ are shown in the upper right corner of each plot.

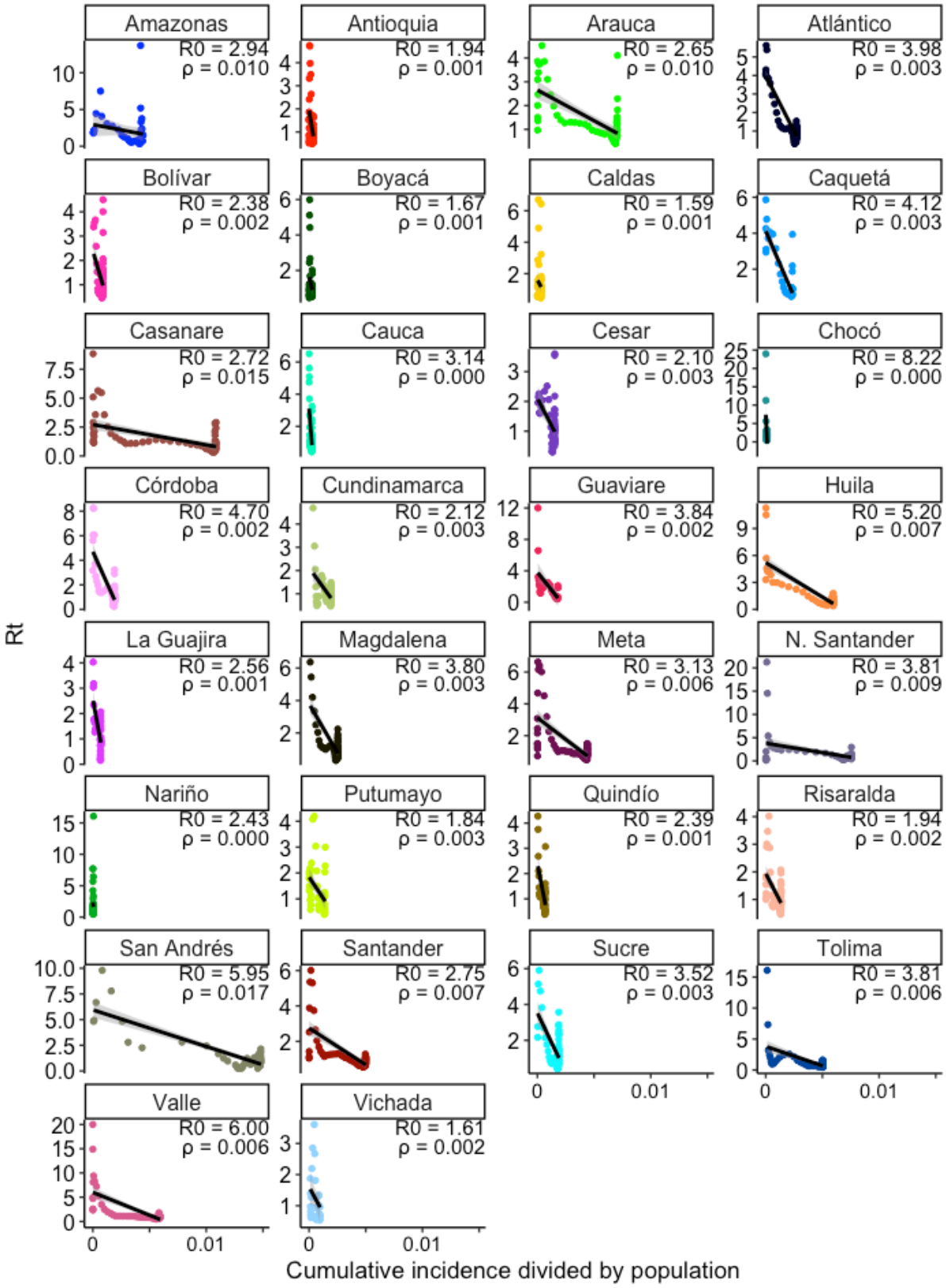


Figure 3.8 R_t estimates versus cumulative incidence divided by population size by department for ZIKV. The black line is the fitted linear regression line, and the shaded line represents the 95% CI. The estimated R_0 and ρ are shown in the upper right corner of each plot.

4.5 Fitting GAMs to the residuals of the linear regression models

GAMs were fitted to the residuals of the linear regression models which were fitted to all departments. The best-fitting GAM to the residuals of the linear regression models versus covariates for CHIKV included mean weekly temperature averaged over three weeks followed by a six-week lag prior to case reporting. Although this model did not have the lowest AIC overall, the differences in AIC between this model and the others were all less than two (Table 3.8). Therefore, the simplest model is preferred. The smooth effect plot is shown in Figure 3.9. The edf of the smooth term for temperature was 1.00 in the best-fitting CHIKV model, and the effect of temperature was significant (p = 0.003). The model reached full convergence after seven iterations. The k value was not too low for the smooth term (p = 0.09). Model diagnostics are shown in Figure 3.10. They show that the model does not fit well, likely due to several outliers. The Q-Q plot in the top-left is curved rather than forming a straight line; the histogram of the residuals is skewed toward zero rather than forming a bell curve; while many of the residual values appear to cluster around zero, there are some that are much higher; and many of the points in the plot of response versus fitted values do not lie along the line y=x. The AIC values for all of the fitted GAMs for CHIKV can be found in Table 3.8.

Table 3.8 AIC values of fitted GAMs for CHIKV. Best-fitting model in bold.

Model	Degrees of freedom	AIC
Temperature	**3.0**	**9,385.0**
Temperature and rainfall	7.9	9,384.8
Temperature and overcrowding	4.0	9,386.7
Temperature and inadequate floors	4.0	9,386.7
Temperature and inadequate exterior walls	4.0	9,383.8

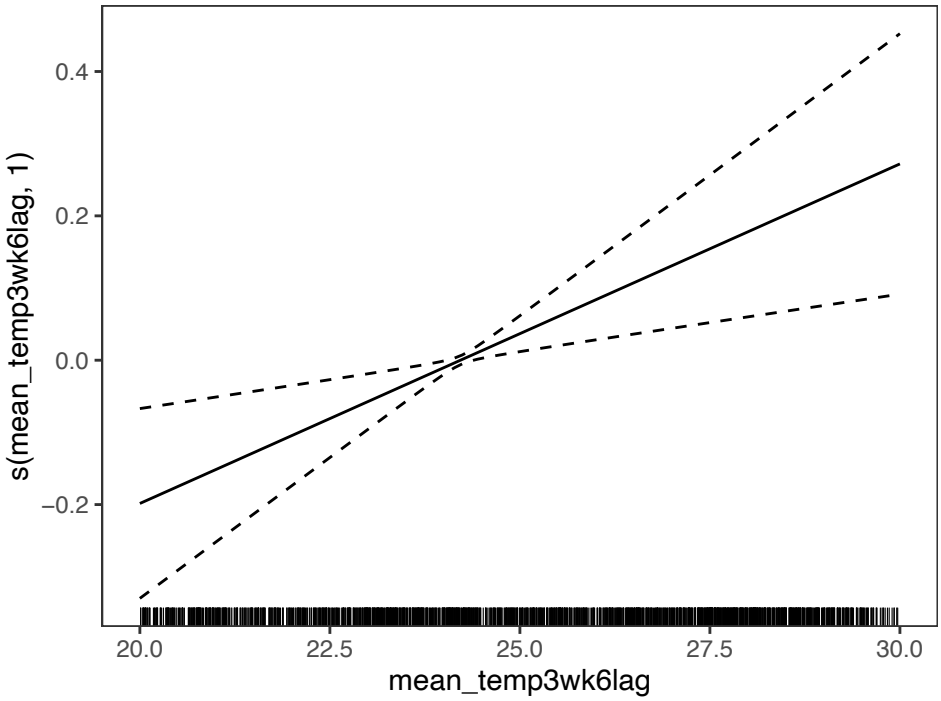


Figure 3.9 Smooth effect plot of mean weekly temperature in °C averaged over three weeks followed by a six-week lag prior to case reporting from best-fitting GAM for CHIKV. The x-axis limits were set to show the range of the most biologically plausible values. The estimated degrees of freedom (edf) of the smooth term can be found on the y-axis label. Edf values around 1 approximate linear terms.

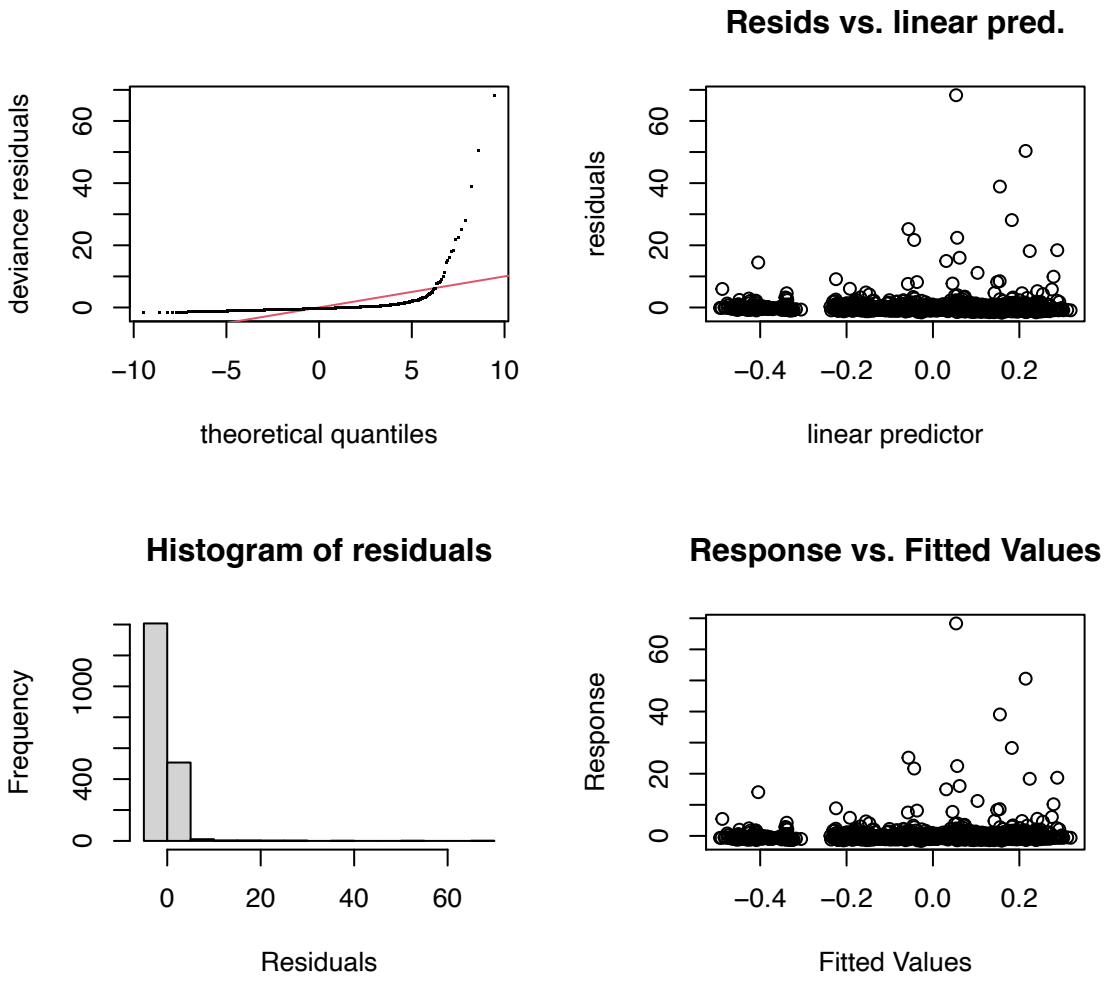


Figure 3.10 Diagnostics of best-fitting GAM for CHIKV. A Q-Q plot is shown on the top-left. It compares the model residuals to a normal distribution. The residuals of a model that has a good fit will be close to a straight line, as indicated by the red line. The bottom left plot is a histogram of the residuals, which should resemble a symmetrical bell curve. The residual values are plotted in the top-right plot and should be equally distributed around zero. On the bottom-right, a plot of response versus fitted values is shown. The points lie along the line y=x in a perfect model.

The best-fitting GAM for ZIKV included mean weekly temperature averaged over three weeks followed by a six-week lag prior to case reporting, the percentage of households with overcrowded conditions, and the percentage of households with inadequate exterior walls. Model diagnostics indicated that the default k value was too low for all three covariates. Model AIC improved when k was increased from nine to 25; however, the p values were still significant for overcrowding and inadequate exterior walls, likely a result of too few data points (each department only had one observation and the model was trying to bend the curve for each data point). Therefore, the best-fitting model was considered the one in

which the default values of k were used for overcrowding and inadequate exterior walls and a k value of 25 was used for temperature (Table 3.9). Smooth effect plots are shown in Figure 3.11. The edf of the smooth term in the model with the lowest AIC was 16.7 for temperature, 1.01 for overcrowding, and 3.00 for inadequate exterior walls. All variables were significant (p value <0.0001 for both temperature and overcrowding and 0.04 for inadequate exterior walls). The model reached full convergence after six iterations. Model diagnostics are shown in Figure 3.12. The model fit for ZIKV is better than that for CHIKV; the Q-Q plot is closer to a straight line, though it is still curved; the histogram of the residuals is still skewed but is closer to a bell curve; the plot of the residuals versus the linear predictors has more residual values clustered around zero; and the points in the plot of response versus fitted values lie closer to the ideal of y=x. The AIC values for all of the fitted GAMs for ZIKV can be found in Table 3.9.

Table 3.9 AIC values of fitted GAMs for ZIKV. Best-fitting model in bold.

Model	Degrees of freedom	AIC
Temperature	9.6	7,390.9
Temperature and rainfall	14.9	7,388.1
Temperature and overcrowding	10.7	7,369.7
Temperature, overcrowding, and inadequate floors	13.5	7,370.8
Temperature, overcrowding, and inadequate exterior walls	14.4	7,367.3
Temperature (higher k), overcrowding, and inadequate exterior walls	**25.1**	**7,324.0**

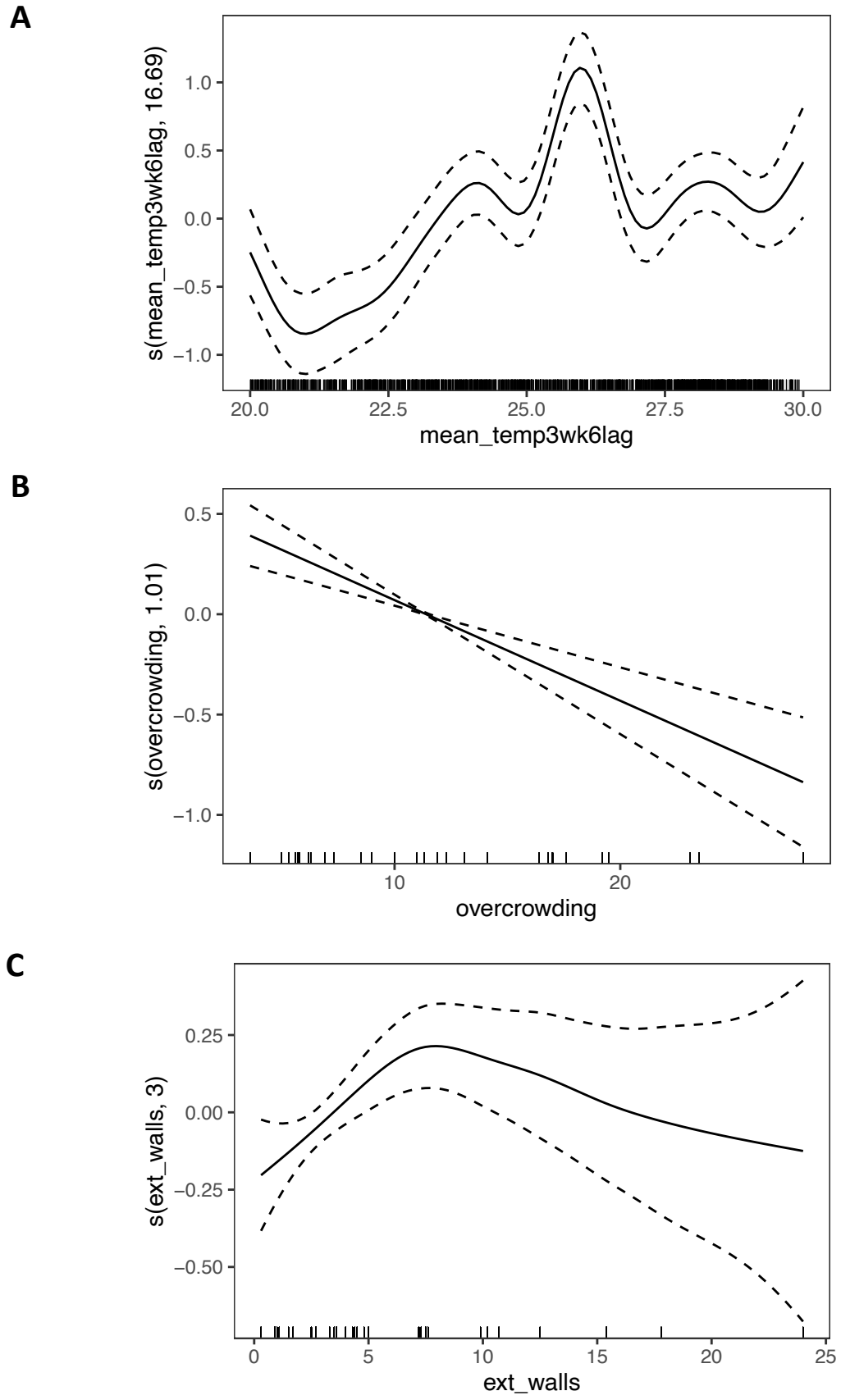


Figure 3.11 Smooth effect plots from best-fitting GAM for ZIKV. *(A)* mean weekly temperature in °C averaged over three weeks followed by a six-week lag prior to case reporting, *(B)* percentage of households with overcrowded conditions, and *(C)* percentage of households with inadequate exterior walls**.** The x-axis limits for *(A)* were set to show the range of the most biologically plausible values. The estimated degrees of freedom (edf) of the smooth term can be found on the y-axis label. Edf values around 1 approximate linear terms.

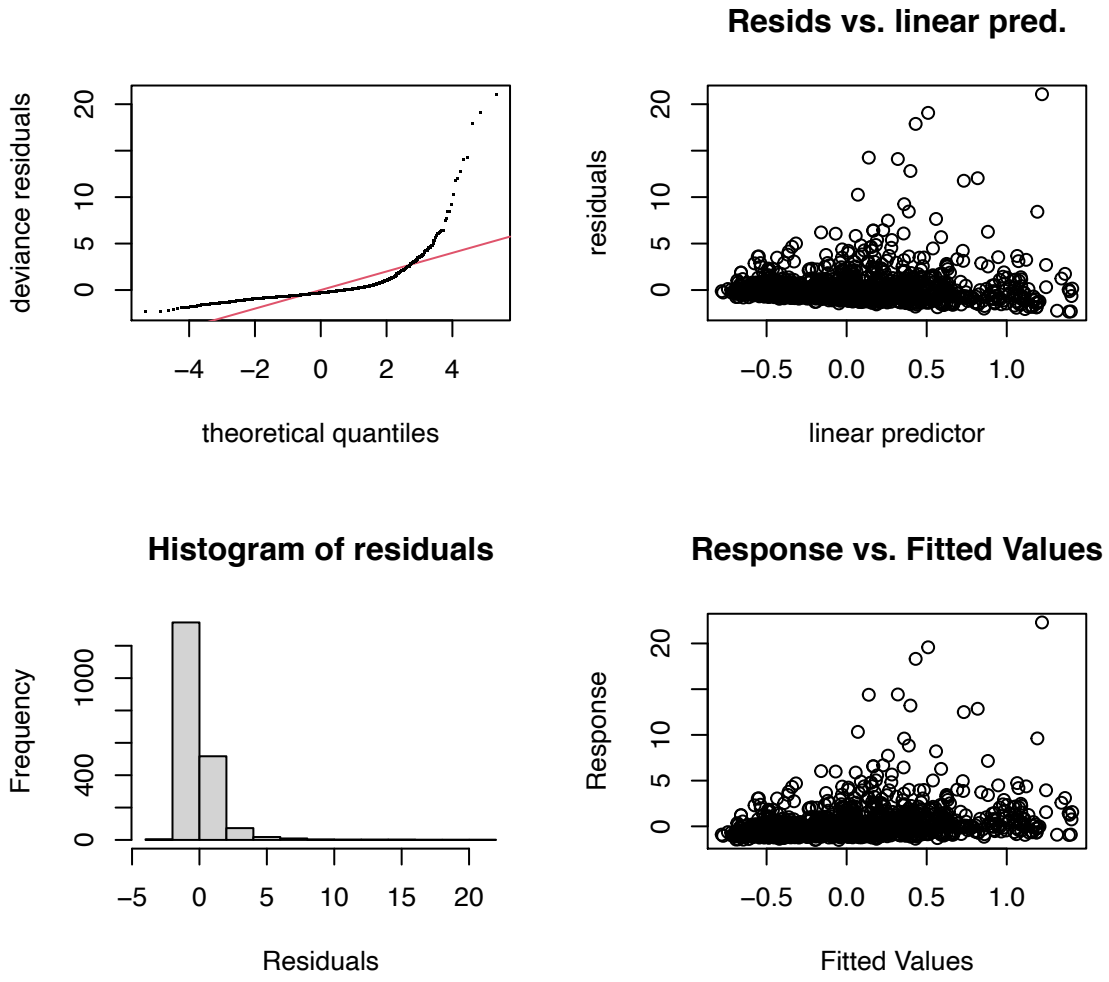


Figure 3.12 Diagnostics of the best-fitting GAM for ZIKV. A Q-Q plot is shown on the top-left. It compares the model residuals to a normal distribution. The residuals of a model that has a good fit will be close to a straight line, as indicated by the red line. The bottom left plot is a histogram of the residuals, which should resemble a symmetrical bell curve. The residual values are plotted in the top-right plot and should be equally distributed around zero. On the bottom-right, a plot of response versus fitted values is shown. The points lie along the line y=x in a perfect model.

4.6 Fitting Poisson models of arbovirus transmission

The effects of temperature on CHIKV and ZIKV transmission using Poisson models are shown in Table 3.10. For CHIKV, the model with the lowest DIC defined temperature as the mean weekly temperature averaged over three weeks followed by a four-week lag prior to case reporting. In contrast, the model with the lowest DIC for ZIKV defined temperature as the mean weekly temperature averaged over three weeks with no lag prior to case reporting.

Table 3.10 Testing the effect of mean weekly temperature in the weeks prior to case reporting on CHIKV and ZIKV transmission with Poisson models. Best-fitting model in bold for each virus.

Virus	Model parameters	LogL	DIC	# Params	Eff params
CHIKV	ρ, R_0	-62,201	124,405	2	1.38
	ρ, R_0, $X^{best(temp)}$ averaged over 5 weeks, σ_{temp}	-61,273	122,553	4	3.37
	ρ, R_0, $X^{best(temp)}$ averaged over 4 weeks, σ_{temp}	-61,289	122,585	4	3.31
	ρ, R_0, $X^{best(temp)}$ averaged over 3 weeks, σ_{temp}	-61,307	122,620	4	3.36
	ρ, R_0, $X^{best(temp)}$ averaged over 3 weeks followed by 3-week lag, σ_{temp}	-61,160	122,327	4	3.35
	ρ, R_0, $X^{best(temp)}$ averaged over 3 weeks followed by 4-week lag, σ_{temp}	**-61,152**	**122,310**	**4**	**3.26**
	ρ, R_0, $X^{best(temp)}$ averaged over 3 weeks followed by 6-week lag, σ_{temp}	-61,275	122,558	4	3.31
ZIKV	ρ, R_0	-22,238	44,479	2	1.35
	ρ, R_0, $X^{best(temp)}$ averaged over 5 weeks, σ_{temp}	-22,097	44,200	4	3.29
	ρ, R_0, $X^{best(temp)}$ averaged over 4 weeks, σ_{temp}	-22,078	44,162	4	3.33
	ρ, R_0, $X^{best(temp)}$ averaged over 3 weeks, σ_{temp}	**-22,061**	**44,129**	**4**	**3.34**
	ρ, R_0, $X^{best(temp)}$ averaged over 3 weeks followed by 3-week lag, σ_{temp}	-22,145	44,298	4	3.33
	ρ, R_0, $X^{best(temp)}$ averaged over 3 weeks followed by 4-week lag, σ_{temp}	-22,155	44,316	4	3.36
	ρ, R_0, $X^{best(temp)}$ averaged over 3 weeks followed by 6-week lag, σ_{temp}	-22,151	44,308	4	3.38

LogL: log likelihood, Params: parameters, Eff params: effective number of parameters, $X^{best(temp)}$: optimal temperature for transmission, σ_{temp}: standard deviation of $X^{best(temp)}$

Similarly, Table 3.11 shows the results of testing different definitions of rainfall with Poisson models. The rainfall model with the lowest DIC for CHIKV used the definition of cumulative weekly rainfall summed over a six-week period followed by a two-week lag prior to case reporting. The lowest DIC model for ZIKV was the one that defined rainfall as the cumulative weekly rainfall summed over a six-week period followed by a three-week lag prior to case reporting.

Table 3.11 Testing the effect of cumulative weekly rainfall in the weeks prior to case reporting on CHIKV and ZIKV transmission with Poisson models. Best-fitting model in bold for each virus.

Virus	Model parameters	LogL	DIC	# Params	Eff params
CHIKV	ρ, R_0	-62,201	124,405	2	1.38
	ρ, R_0, $X^{best(rain)}$ summed over 6 weeks followed by 2-week lag, σ_{rain}	**-60,086**	**120,178**	**4**	**3.31**
	ρ, R_0, $X^{best(rain)}$summed over 6 weeks followed by 3-week lag, σ_{rain}	-60,561	121,129	4	3.29
ZIKV	ρ, R_0	-22,238	44,479	2	1.35
	ρ, R_0, $X^{best(rain)}$ summed over 6 weeks followed by 2-week lag, σ_{rain}	-21,965	43,936	4	3.29
	ρ, R_0, $X^{best(rain)}$summed over 6 weeks followed by 3-week lag, σ_{rain}	**-21,818**	**43,642**	**4**	**3.25**

LogL: log likelihood, Params: parameters, Eff params: effective number of parameters, $X^{best(rain)}$: optimal rainfall for transmission, σ_{rain}: standard deviation of $X^{best(rain)}$

Due to large estimated standard deviations for the temperature function, models with two standard deviations were also tested. Table 3.12 shows that two standard deviations are preferred over one for the temperature function in both CHIKV and ZIKV Poisson models that also estimate reporting rate ρ and a single R_0.

Table 3.12 Testing whether two standard deviations instead of one better describe the effect of temperature on CHIKV and ZIKV transmission in the weeks prior to case reporting with Poisson models. "Best temperature" uses the best-fitting definition of temperature for each virus from Table 3.10. For both CHIKV and ZIKV, the best-fitting model included two standard deviations for the effect of temperature.

Virus	Model parameters	LogL	DIC	# Params	Eff params
CHIKV	ρ, R_0, $X^{best(temp)}$, σ_{temp1}	-61,152	122,310	4	3.26
	ρ, R_0, $X^{best(temp)}$, σ_{temp1}, σ_{temp2}	-60,944	121,897	5	4.31
ZIKV	ρ, R_0, $X^{best(temp)}$, σ_{temp1}	-22,061	44,129	4	3.34
	ρ, R_0, $X^{best(temp)}$, σ_{temp1}, σ_{temp2}	-21,952	43,913	5	4.25

LogL: log likelihood, Params: parameters, Eff params: effective number of parameters, $X^{best(temp)}$: optimal temperature for transmission, σ_{temp1}: standard deviation associated with temperatures below $X^{best(temp)}$, σ_{temp2}: standard deviation associated with temperatures greater than or equal to $X^{best(temp)}$

Results from Poisson models with socioeconomic covariates can be found in Table 3.13. The percentage of households with inadequate exterior walls as well as those with overcrowded conditions appeared uncorrelated with CHIKV and ZIKV transmission dynamics as indicated by the low effective number of parameters relative to the actual number of parameters in each model. DIC values were also similar across models. Although the ZIKV model with inadequate exterior walls had a change in DIC of 7, the effect size of the parameter was extremely small (0.004, 95% CrI: 0.001-0.006), resulting in a multiplication factor (γ_i) ranging from 0.98 to 1.07 across departments at the median estimate.

Table 3.13 Effect of socioeconomic factors on CHIKV and ZIKV transmission using Poisson models.

Virus	Model parameters	LogL	DIC	# Params	Eff params
CHIKV	ρ, R_0	-62,201	124,405	2	1.38
	ρ, R_0, inadequate exterior walls	-62,202	124,407	3	1.96
	ρ, R_0, overcrowding	-62,202	124,407	3	1.98
ZIKV	ρ, R_0	-22,238	44,479	2	1.35
	ρ, R_0, inadequate exterior walls	-22,234	44,472	3	2.38
	ρ, R_0, overcrowding	-22,239	44,482	3	1.94

LogL: log likelihood, Params: parameters, Eff params: effective number of parameters

Tables 3.14-3.15 show results of Poisson models with weather covariates as well as multiple R_0s for CHIKV and ZIKV. Figures 3.13-3.14 show the relationship between predicted R_t and weather using results from the Poisson model with weather covariates. The predicted R_t for CHIKV appears to be above the threshold of 1 for a larger range of temperature and rainfall combinations compared to ZIKV.

Table 3.14 Estimated R_0s and reporting rate of CHIKV from Poisson models with weather covariates and multiple R_0s. Posterior median and 95% credible interval presented for each parameter. Best-fitting model in bold.

	Poisson	Poisson with weather covariates	Poisson with multiple R_0s	**Poisson with multiple R_0s and rainfall**
DIC	124,405	119,056	113,042	**108,106**
Number of parameters	2	7	30	**32**
ρ (reporting rate)	0.051 (0.050-0.052)	0.049 (0.048-0.049)	0.041 (0.040-0.041)	**0.041 (0.041-0.042)**
R_0	1.23 (1.23-1.24)	1.49 (1.48-1.50)	Range: 0.66-2.02	**Range: 0.92-2.89**
$X^{best(temp)}$ (°C)*		25.3 (25.1-25.5)		
σ_{temp1} (°C)**		15.1 (14.4-15.7)		
σ_{temp2} (°C)**		5.9 (5.6-6.3)		
$X^{best(rain)}$ (mm)***		477 (470-484)		**747 (712-788)**
σ_{rain} (mm)		562 (549-576)		**701 (671-734)**

*The best-fitting temperature $X^{best(temp)}$ uses the mean weekly temperature averaged over three weeks followed by a four-week lag prior to case reporting.

** σ_{temp1}: standard deviation associated with temperatures below $X^{best(temp)}$, σ_{temp2}: standard deviation associated with temperatures greater than or equal to $X^{best(temp)}$.

***The best-fitting rainfall $X^{best(rain)}$ uses cumulative weekly rainfall summed over six weeks followed by a two-week lag prior to case reporting.

Table 3.15 Estimated R_0s and reporting rate of ZIKV from Poisson models with weather covariates and multiple R_0s. Posterior median and 95% credible interval presented for each parameter. Best-fitting model in bold.

	Poisson	Poisson with weather covariates	Poisson with multiple R_0s	**Poisson with multiple R_0s and rainfall**
DIC	44,479	43,090	42,234	**41,471**
Number of parameters	2	7	31	**33**
ρ (reporting rate)	0.015 (0.015-0.015)	0.015 (0.015-0.015)	0.015 (0.015-0.015)	**0.015 (0.015-0.015)**
R_0	1.25 (1.24-1.25)	1.47 (1.46-1.49)	Range: 0.89-3.06	**Range: 1.05-3.16**
$X^{best(temp)}$ (°C)*		27.2 (27.0-27.3)		
σ_{temp1} (°C)**		14.0 (13.3-14.8)		
σ_{temp2} (°C)**		3.5 (3.2-3.9)		
$X^{best(rain)}$ (mm)***		310 (302-319)		**345 (333-359)**
σ_{rain} (mm)		434 (418-452)		**467 (445-492)**

*The best-fitting temperature $X^{best(temp)}$ uses the mean weekly temperature averaged over three weeks prior to case reporting.

** σ_{temp1}: standard deviation associated with temperatures below $X^{best(temp)}$, σ_{temp2}: standard deviation associated with temperatures greater than or equal to $X^{best(temp)}$.

***The best-fitting rainfall $X^{best(rain)}$ uses cumulative weekly rainfall summed over six weeks followed by a three-week lag prior to case reporting.

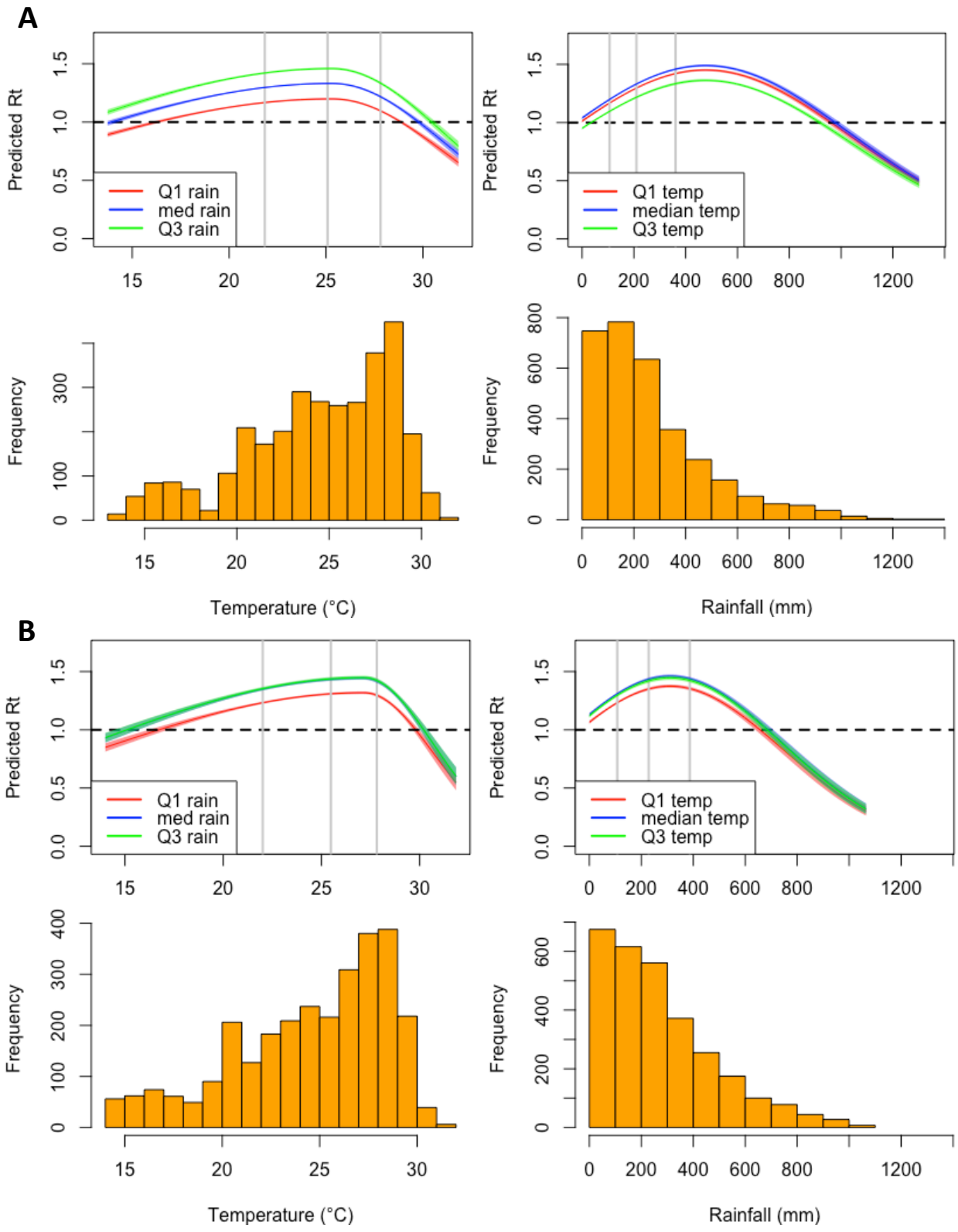


Figure 3.13 R_t as a function of temperature and rainfall from Poisson models with weather covariates. *(A)* CHIKV and *(B)* ZIKV. The orange bars show the distribution of the data, and the gray lines show the interquartile range of the temperature and rainfall data. The horizontal dashed line shows the threshold $R_t = 1$. Predictions were made by simulating the model with 5,000 parameter sets randomly sampled from the posterior distribution. The blue line is underneath the green line in *(B)*.

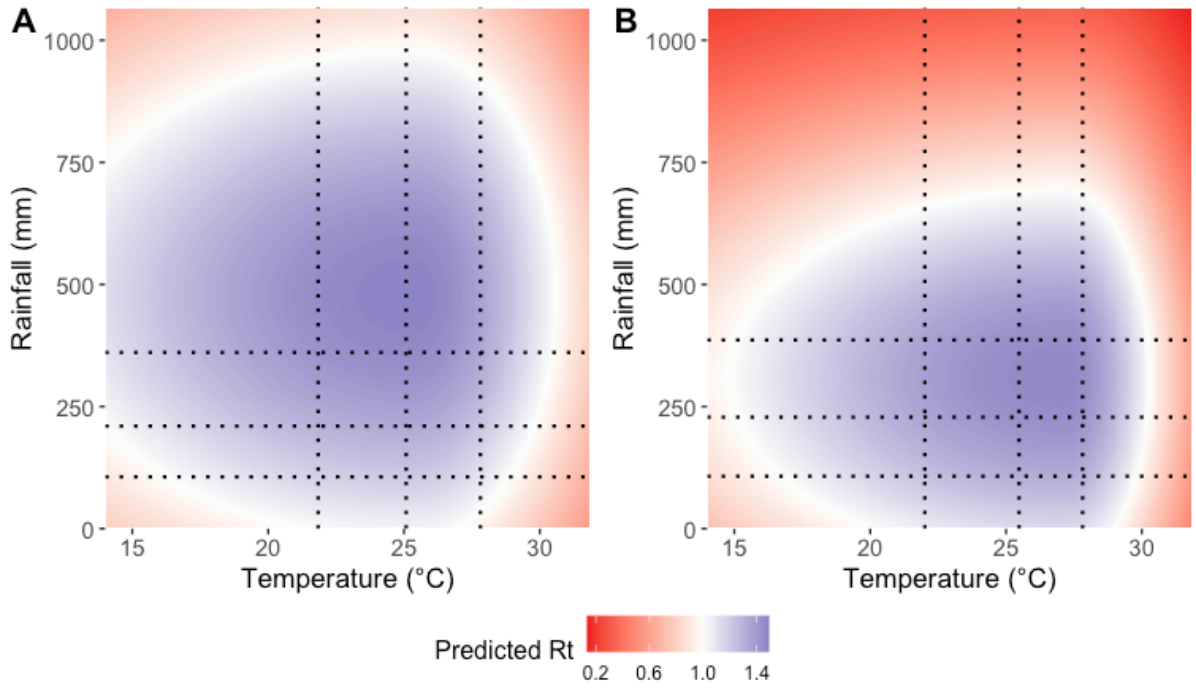


Figure 3.14 Heatmap showing predicted R_t as a function of temperature and rainfall from Poisson models with weather covariates. *(A)* CHIKV and *(B)* ZIKV. The vertical and horizontal dotted lines show the interquartile range of the temperature and rainfall data, respectively. Predictions were made by simulating the model with 1,000 parameter sets randomly sampled from the posterior distribution and averaging across the simulations.

The best-fitting Poisson model for both viruses is the one that estimates ρ, one R_0 for each department, and rainfall with one standard deviation (Tables 3.14-3.15). The estimate for ρ was 0.041 (95% CrI: 0.041-0.042) for CHIKV compared to 0.015 (95% CrI: 0.015-0.015) for ZIKV. Estimates of R_0 ranged from 0.92-2.89 for CHIKV and 1.05-3.16 for ZIKV. Optimal conditions for CHIKV transmission occurred when the cumulative weekly rainfall summed over six weeks was 747 mm (95% CrI: 712-788) followed by a two-week lag prior to case reporting with a standard deviation of 701 (95% CrI: 671-734). Similarly, ZIKV transmission was optimal when the cumulative weekly rainfall summed over six weeks was 345 mm (333-359) followed by a three-week lag prior to case reporting with a standard deviation of 467 (95% CrI: 445-492). For both CHIKV and ZIKV, the parameters for rainfall vary considerably between the Poisson with weather covariates model and the Poisson model with multiple R_0s and rainfall. Estimates of the rainfall parameter and its standard deviation were both higher in the latter model, suggesting that transmission of CHIKV and ZIKV is optimized at higher levels of cumulative rainfall regardless of temperature and at lower levels of cumulative rainfall when temperature is considered. For Poisson models with multiple R_0s

and temperature, neither the CHIKV model nor the ZIKV model converged after 100,000 iterations. When the ZIKV model was run for 200,000 iterations, the model converged, but the prior distribution was returned for $X^{best(temp)}$. Similarly, running the CHIKV model for 300,000 iterations was necessary for it to reach convergence; however, the prior distribution was returned for $X^{best(temp)}$, and the mixing of the chains was slow, especially for $X^{best(temp)}$ and σ_{temp}.

Figures 3.15-3.16 show how the estimated R_ts from the best-fitting Poisson models compare to the R_ts obtained from EpiEstim. The relationship is positive and statistically significant for both viruses, but the 95% credible intervals of the Poisson model estimates are very narrow.

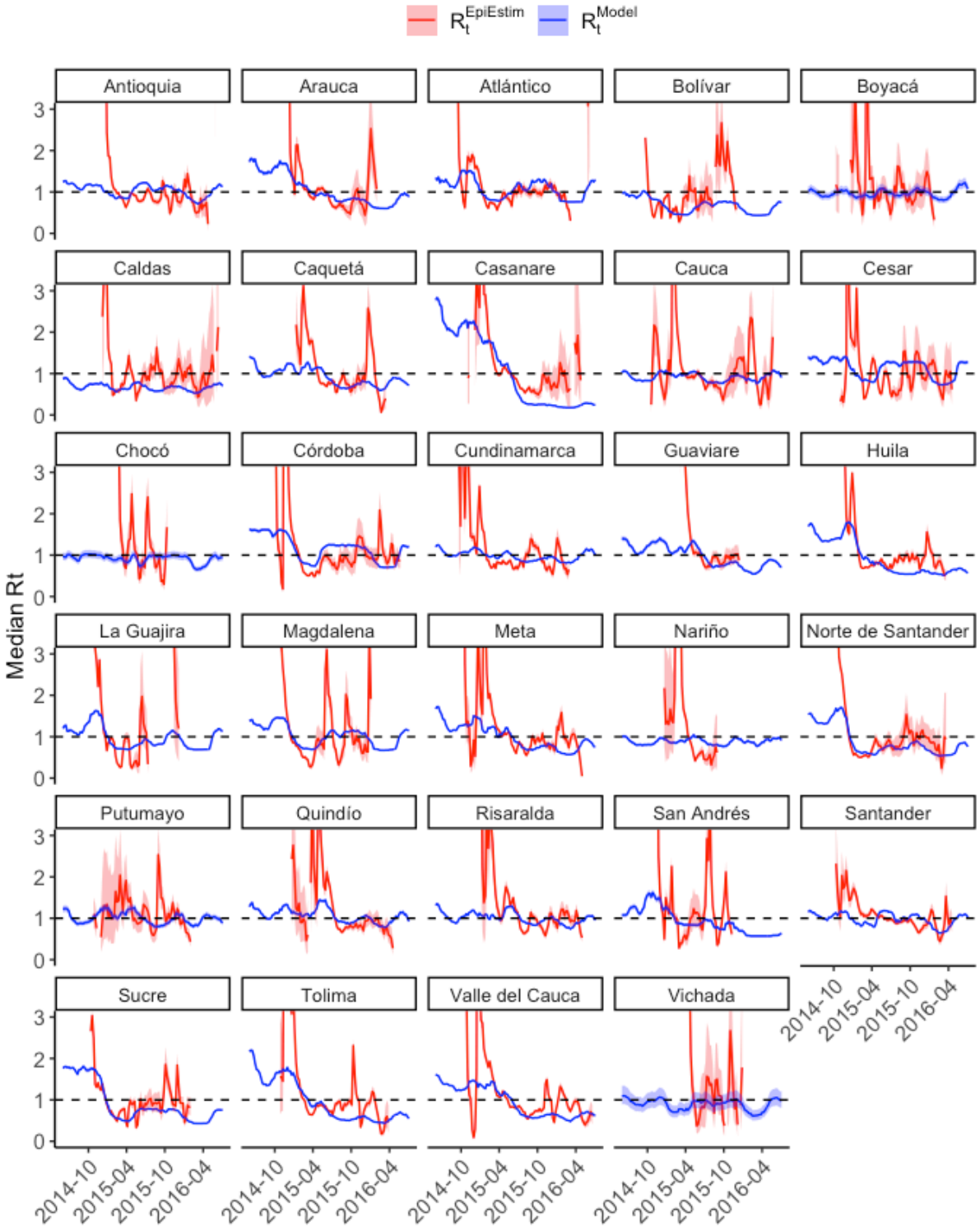


Figure 3.15 Comparing median estimates of R_t from the best-fitting Poisson model for CHIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines). EpiEstim R_ts are plotted in the center of the 5-week window used to compute each estimate. Shaded areas represent 95% CrI. There is a positive statistically significant correlation of 0.21 (Pearson's correlation coefficient, 95% CI: 0.16-0.25, $p < 0.0001$).

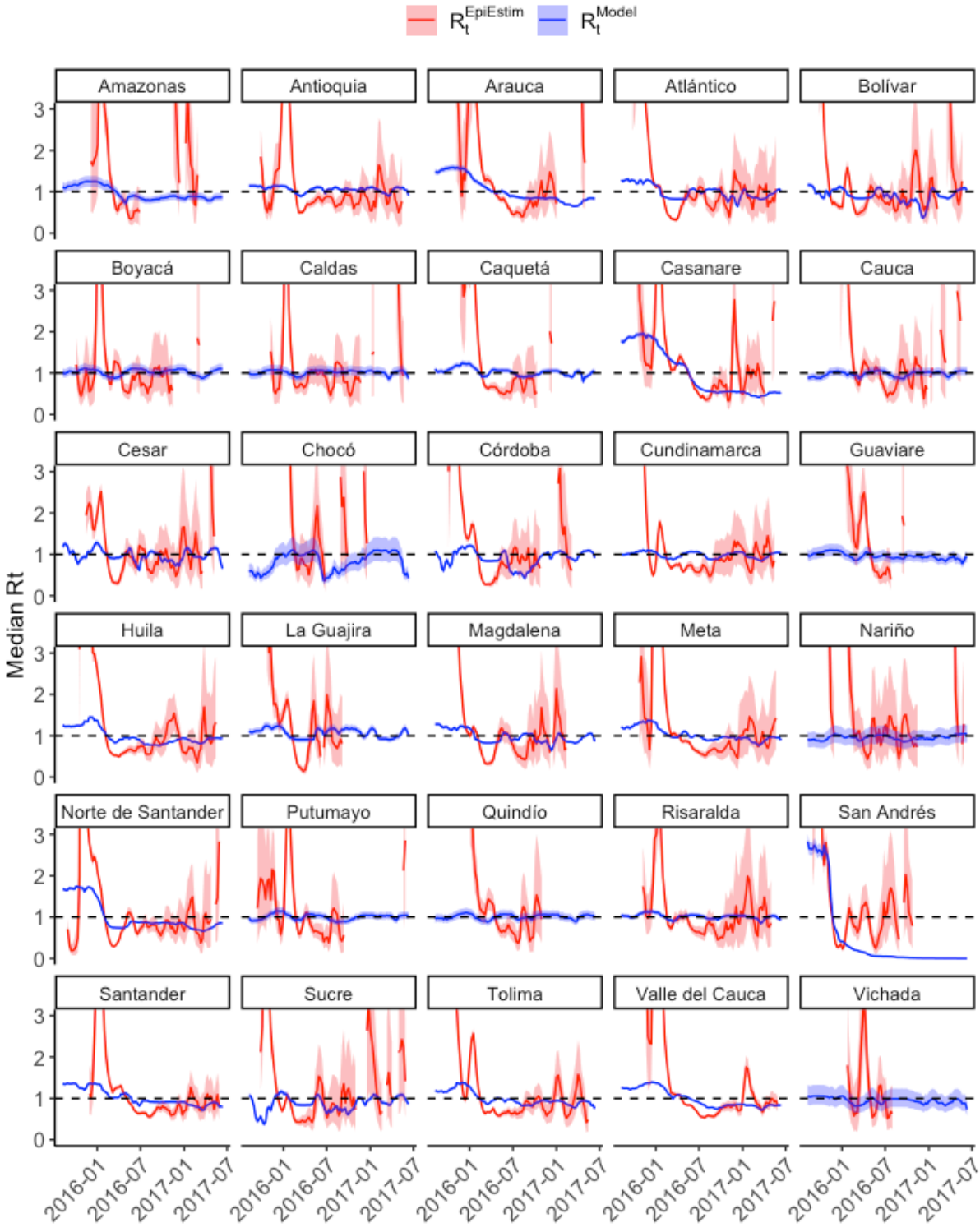


Figure 3.16 Comparing median estimates of R_t from the best-fitting Poisson model for ZIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines). EpiEstim R_ts are plotted in the center of the 5-week window used to compute each estimate. Shaded areas represent 95% CrI. There is a positive statistically significant correlation of 0.32 (Pearson's correlation coefficient, 95% CI: 0.28-0.36, $p < 0.0001$).

4.7 Fitting negative binomial models of arbovirus transmission

Table 3.16 shows the results of fitting negative binomial models with and without socioeconomic covariates. The model DIC does not change in a meaningful way when the percentage of households with inadequate exterior walls and those with overcrowded conditions are added to the models with ρ, a single R_0, and overdispersion, ϕ. Negative binomial models with ρ, a single R_0, ϕ, and weather covariates (temperature plus its standard deviation or rainfall plus its standard deviation) converged, but the prior distribution was returned for at least one weather parameter in each model.

Table 3.16 Effect of socioeconomic factors on CHIKV and ZIKV transmission using negative binomial models.

Virus	Model parameters	LogL	DIC	# Params	Eff params
CHIKV	ρ, R_0, ϕ	-9,341	18,688	3	2.41
	ρ, R_0, ϕ, inadequate exterior walls	-9,341	18,689	4	3.08
	ρ, R_0, ϕ, overcrowding	-9,341	18,688	4	2.98
ZIKV	ρ, R_0, ϕ	-7,056	14,117	3	2.38
	ρ, R_0, ϕ, inadequate exterior walls	-7,057	14,119	4	2.89
	ρ, R_0, ϕ, overcrowding	-7,057	14,119	4	2.97

LogL: log likelihood, Params: parameters, Eff params: effective number of parameters

Table 3.17 shows the results of fitting negative binomial models to estimate ρ and R_0s for CHIKV and ZIKV. For both viruses, the model with the lowest DIC was the negative binomial model with multiple R_0s. The estimated ρ for CHIKV was 0.045 (95% CrI: 0.042-0.049), and it was 0.016 (95% CrI: 0.015-0.017) for ZIKV. Estimated R_0s ranged from 0.96-2.93 (median 1.30) for CHIKV and 0.98-5.87 (median 1.33) for ZIKV. ϕ was lower for CHIKV compared to ZIKV (1.44, 95% CrI: 1.34-1.55 for CHIKV versus 1.80, 95% CrI: 1.66-1.95 for ZIKV). Low estimates for this parameter suggest high overdispersion in the data. This means that a small proportion of cases are responsible for causing a greater proportion of secondary cases. Similar to the models with a single R_0 and weather covariates, negative binomial models with multiple R_0s and temperature or rainfall converged, but the prior distribution for one of the weather parameters was returned in each model.

Table 3.17 Estimated R_0s and reporting rate of CHIKV and ZIKV from negative binomial models with multiple R_0s. Posterior median and 95% credible interval presented for each parameter. Best-fitting models in bold.

	CHIKV		ZIKV	
Model	Negative binomial	**Negative binomial with multiple R_0s**	Negative binomial	**Negative binomial with multiple R_0s**
DIC	18,688	**18,645**	14,117	**14,040**
Number of parameters	3	**31**	3	**32**
Effective parameters	2.41	**30.28**	2.38	**31.39**
ρ (reporting rate)	0.068 (0.059-0.081)	**0.045 (0.042-0.049)**	0.021 (0.019-0.023)	**0.016 (0.015-0.017)**
R_0	1.28 (1.22-1.34)	**Range: 0.96-2.93**	1.37 (1.31-1.43)	**Range: 0.98-5.87**
ϕ (overdispersion)	1.38 (1.28-1.48)	**1.44 (1.34-1.55)**	1.68 (1.55-1.82)	**1.80 (1.66-1.95)**

Figures 3.17-3.18 show the estimated R_ts from the best-fitting negative binomial models versus the R_ts obtained from EpiEstim. The model fits are good. A version of Figure 3.18 with a different y-axis that shows the full 95% credible interval of the model estimate for San Andrés and Providencia can be found in Appendix S1.

For CHIKV, the correlation between the EpiEstim R_ts and the model R_ts was lower for the best-fitting negative binomial model compared to the best-fitting Poisson model (0.16 versus 0.21 respectively). However, the corresponding correlations for ZIKV were similar (0.31 versus 0.32). While the Poisson models capture the general trends of the EpiEstim R_ts better than the negative binomial models, the negative binomial models better characterize the uncertainty in the R_t estimates. This increased uncertainty is statistically favored by the DIC. Additionally, the low estimates for ϕ from the negative binomial models suggest that Poisson models may not be appropriate due to overdispersion in the data.

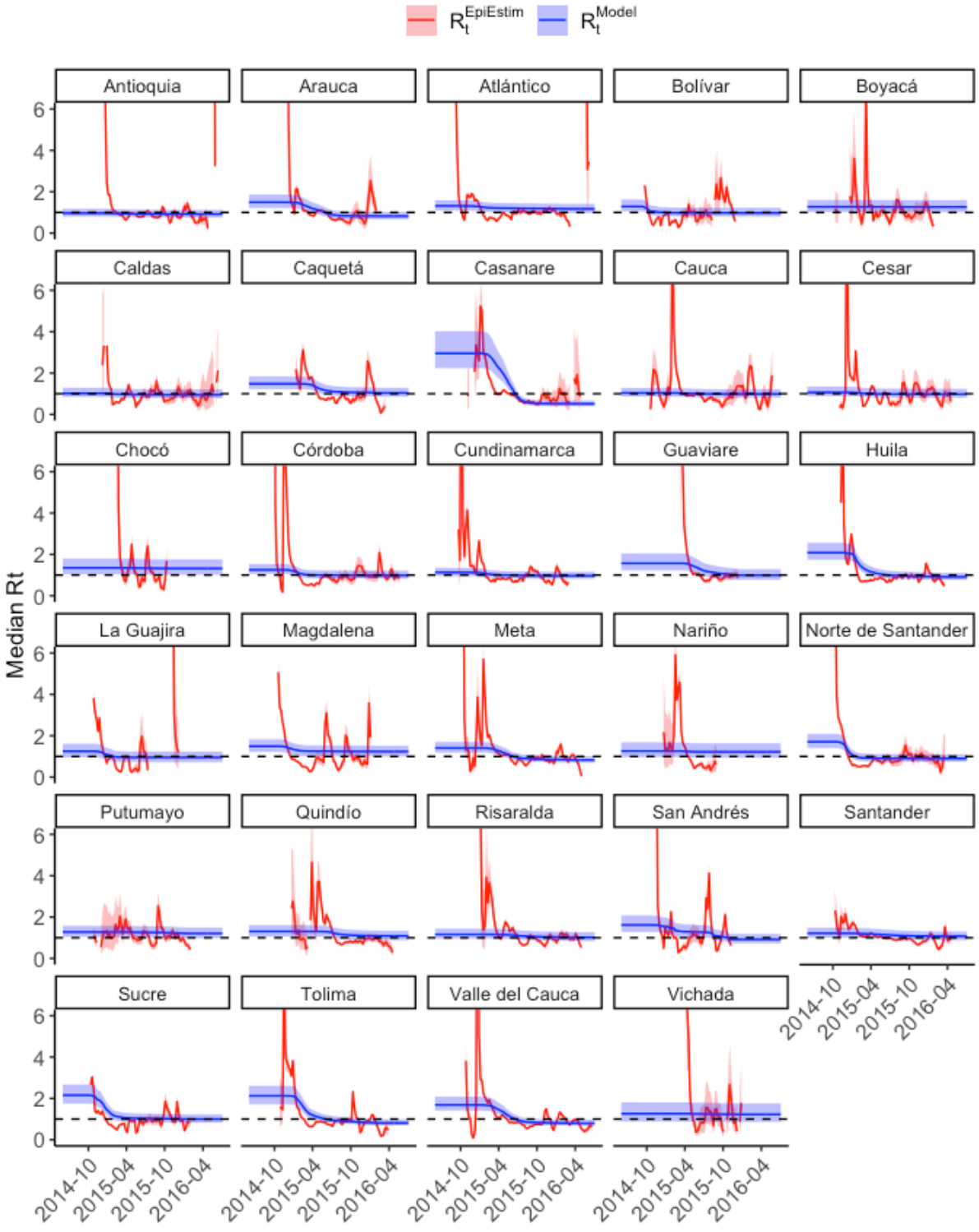


Figure 3.17 Comparing median estimates of R_t from the best-fitting negative binomial model for CHIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines). EpiEstim R_ts are plotted in the center of the 5-week window used to compute each estimate. Shaded areas represent 95% CrI. There is a positive statistically significant correlation of 0.16 (Pearson's correlation coefficient, 95% CI: 0.12-0.21, $p < 0.0001$).

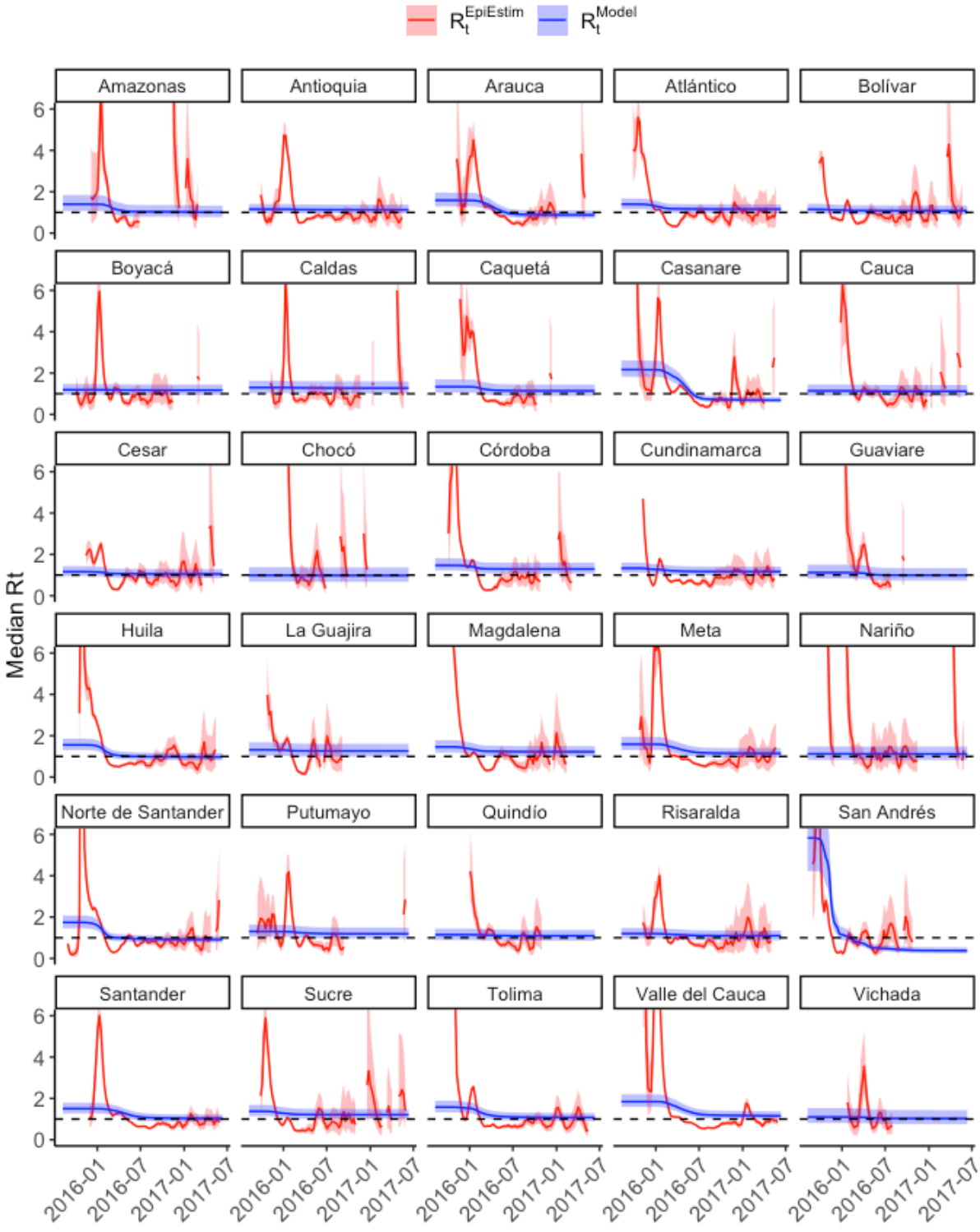


Figure 3.18 Comparing median estimates of R_t from the best-fitting negative binomial model for ZIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines). EpiEstim R_ts are plotted in the center of the 5-week window used to compute each estimate. Shaded areas represent 95% CrI. There is a positive statistically significant correlation of 0.31 (Pearson's correlation coefficient, 95% CI: 0.27-0.35, $p < 0.0001$).

Table 3.18 shows the estimated R_0 values for CHIKV and ZIKV from the best-fitting negative binomial models across departments. Out of 29 departments that were modeled for both CHIKV and ZIKV, the estimated R_0s for only four departments had substantially different posterior probabilities that the R_0 for one virus was higher than that for the other virus. Of these, all but one department (San Andrés and Providencia) had a higher estimated R_0 for CHIKV compared to ZIKV. A plot comparing the posterior densities of the estimated R_0s for CHIKV and ZIKV by department can be found in Figure 3.19.

Table 3.18 Estimated R_0 values of CHIKV and ZIKV for each department from the best-fitting negative binomial models. Median posterior and 95% CrI shown.

Department	R_0 CHIKV	R_0 ZIKV	Posterior probability that $R_{0,CHIKV} > R_{0,ZIKV}$	Posterior probability that $R_{0,CHIKV} < R_{0,ZIKV}$
Amazonas		1.41 (1.07-1.85)		
Antioquia	0.96 (0.79-1.16)	1.16 (0.98-1.39)	0.08	0.92
Arauca	1.49 (1.20-1.88)	1.57 (1.27-1.96)	0.37	0.63
Atlántico	1.31 (1.09-1.59)	1.39 (1.17-1.67)	0.32	0.68
Bolívar	1.28 (1.04-1.61)	1.15 (0.95-1.40)	0.76	0.24
Boyacá	1.26 (1.00-1.61)	1.20 (0.96-1.50)	0.61	0.39
Caldas	1.02 (0.83-1.27)	1.30 (1.04-1.63)	0.06	0.94
Caquetá	1.48 (1.20-1.84)	1.33 (1.06-1.69)	0.74	0.26
Casanare	2.93 (2.21-3.96)	2.18 (1.79-2.66)	0.95	0.05
Cauca	1.04 (0.85-1.28)	1.14 (0.91-1.44)	0.28	0.72
Cesar	1.08 (0.89-1.32)	1.16 (0.96-1.42)	0.30	0.70
Chocó	1.35 (1.02-1.79)	0.98 (0.68-1.42)	0.91	0.09
Córdoba	1.25 (1.05-1.52)	1.47 (1.20-1.82)	0.13	0.87
Cundinamarca	1.14 (0.95-1.37)	1.33 (1.12-1.60)	0.11	0.89
Guaviare	1.57 (1.23-2.06)	1.13 (0.85-1.52)	0.95	0.05
Huila	2.10 (1.74-2.57)	1.56 (1.31-1.89)	0.99	0.01
La Guajira	1.25 (0.97-1.63)	1.32 (1.03-1.71)	0.37	0.63
Magdalena	1.49 (1.20-1.86)	1.44 (1.18-1.78)	0.58	0.42
Meta	1.41 (1.18-1.72)	1.58 (1.31-1.92)	0.21	0.79
Nariño	1.25 (0.95-1.68)	1.13 (0.84-1.50)	0.70	0.30
Norte de Santander	1.71 (1.40-2.11)	1.75 (1.47-2.10)	0.44	0.56
Putumayo	1.28 (1.03-1.60)	1.31 (1.05-1.65)	0.43	0.57
Quindío	1.30 (1.05-1.61)	1.14 (0.89-1.46)	0.78	0.22
Risaralda	1.16 (0.96-1.43)	1.20 (0.99-1.47)	0.41	0.59
San Andrés & Providencia	1.61 (1.30-2.04)	5.87 (4.18-8.31)	<0.0001	>0.9999
Santander	1.22 (1.02-1.46)	1.50 (1.27-1.80)	0.05	0.95
Sucre	2.15 (1.76-2.65)	1.38 (1.12-1.70)	0.999	0.001
Tolima	2.11 (1.72-2.60)	1.57 (1.32-1.88)	0.98	0.02
Valle del Cauca	1.69 (1.40-2.04)	1.86 (1.57-2.20)	0.23	0.77
Vichada	1.26 (0.88-1.80)	1.09 (0.78-1.54)	0.71	0.29

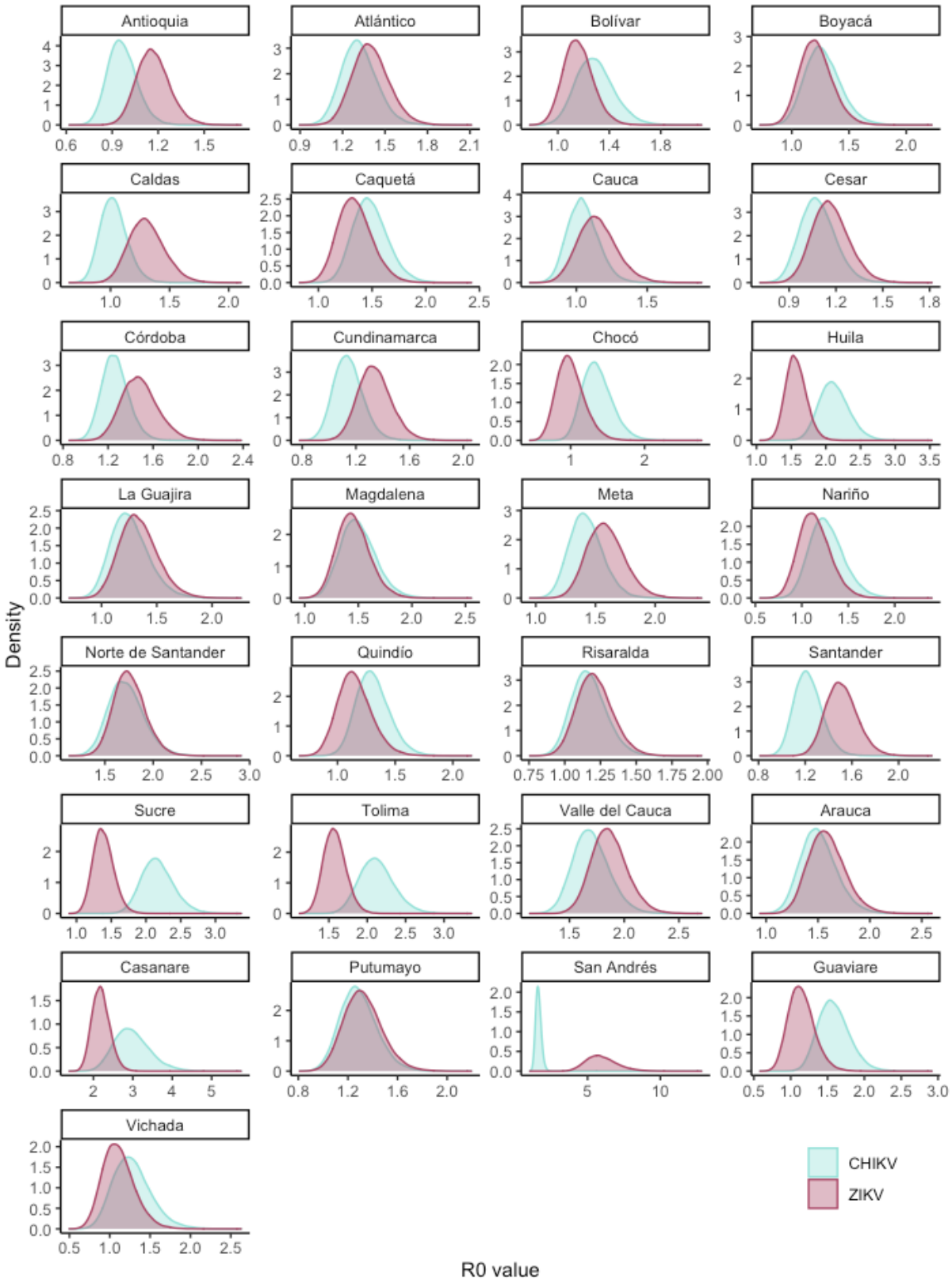


Figure 3.19 Comparison of the posterior densities of estimated R_0s for CHIKV and ZIKV by department from the best-fitting negative binomial models.

Using the incidence of disease and the estimates of ρ from the best-fitting negative binomial models, the estimated infection attack rate for each department and week were calculated (Figures 3.20-3.21). For CHIKV, estimated infection attack rates after 110 weeks ranged from 0.01 in Boyacá to 0.82 in Casanare with a median of 0.17. Estimated infection attack rates of ZIKV after 97 weeks ranged from 0.003 in Nariño to 0.93 in San Andrés and Providencia with a median of 0.12.

For departments with available post-epidemic seroprevalence estimates for their capital cities [191], estimated infection attack rates for CHIKV were 0.05, 0.47, 0.56, and 0.53 for Antioquia, Norte de Santander, Huila, and Sucre, respectively. For ZIKV, the estimates were 0.02, 0.47, 0.37, and 0.12, respectively. Compared to the seroprevalence estimates from Nouvellet et al. [191], the estimated infection attack rate of CHIKV for the department of Huila was within the 95% confidence interval for its capital city of Neiva, while the estimates for Norte de Santander, Huila, and Sucre were outside that for their capital cities (all lower). For ZIKV, only the estimated infection attack rate for Norte de Santander was within the 95% confidence interval for the estimated seroprevalence of its capital city of Cúcuta, while estimates for Antioquia, Huila, and Sucre were all lower than their respective capital cities.

A hexagon map comparing the modeled attack rates for CHIKV and ZIKV is shown in Figure 3.22. Of the 29 departments with estimates for both viruses, only two had similar posterior probabilities that the infection attack rate for one virus was higher than the other. The posterior probability that the infection attack rate of CHIKV was higher than that of ZIKV was 0.65 for Arauca and 0.54 for Norte de Santander. Similarly, the posterior probability that the infection attack rate of ZIKV was higher than that of CHIKV was 0.35 for Arauca and 0.46 for Norte de Santander. The posterior probability that the infection attack rate of CHIKV was higher than that of ZIKV was 0.98 for Magdalena (0.02 that ZIKV was higher than CHIKV), while the remaining departments had posterior probabilities of either <0.0001 or >0.9999.

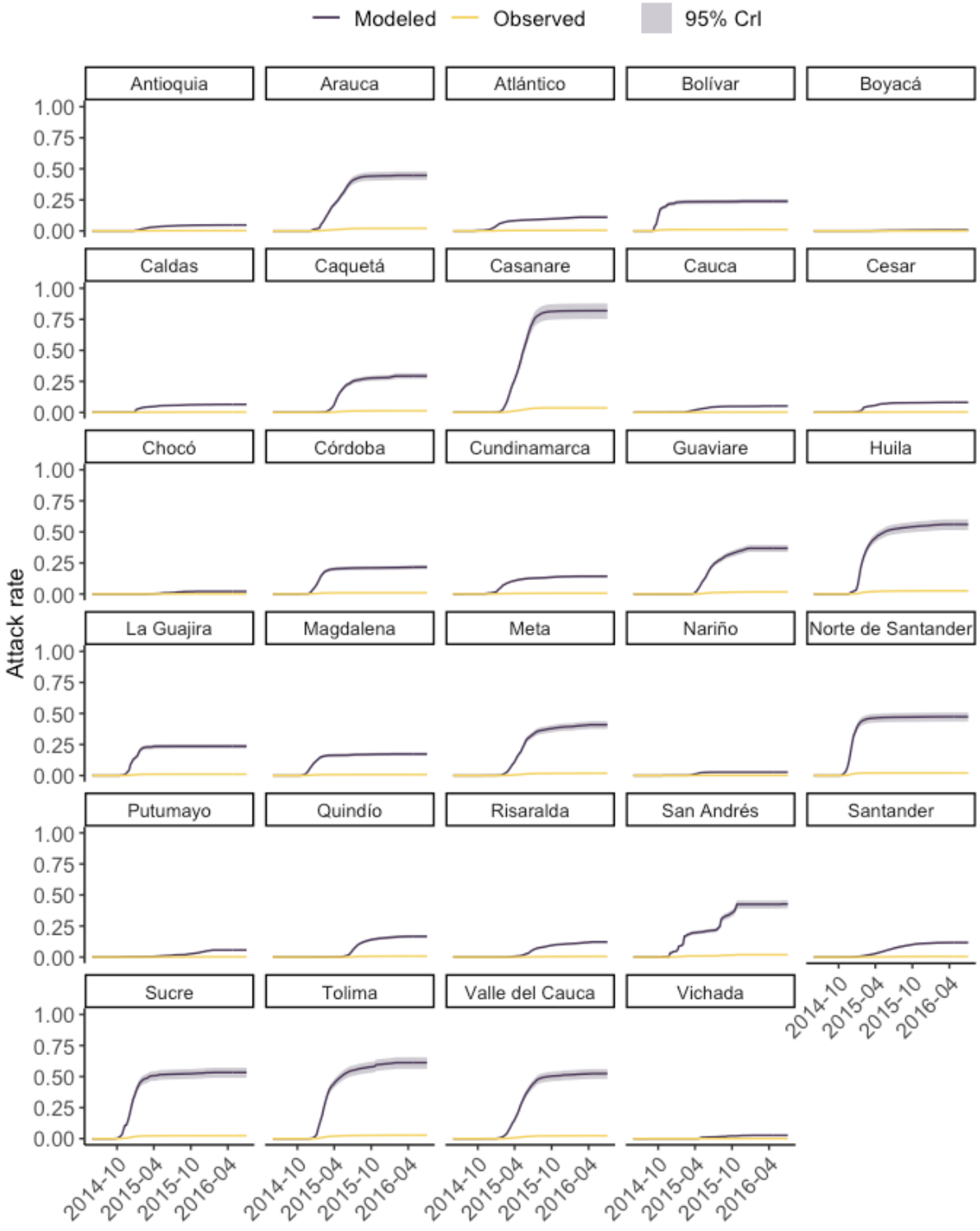


Figure 3.20 Modeled and observed attack rates of CHIKV by department. Yellow lines show the observed attack rates, while purple lines show the estimated infection attack rates, which were obtained by adjusting the observed attack rates by the estimated reporting rate. The reporting rate was estimated from the best-fitting negative binomial model. Purple shading represents the 95% CrI associated with this estimate.

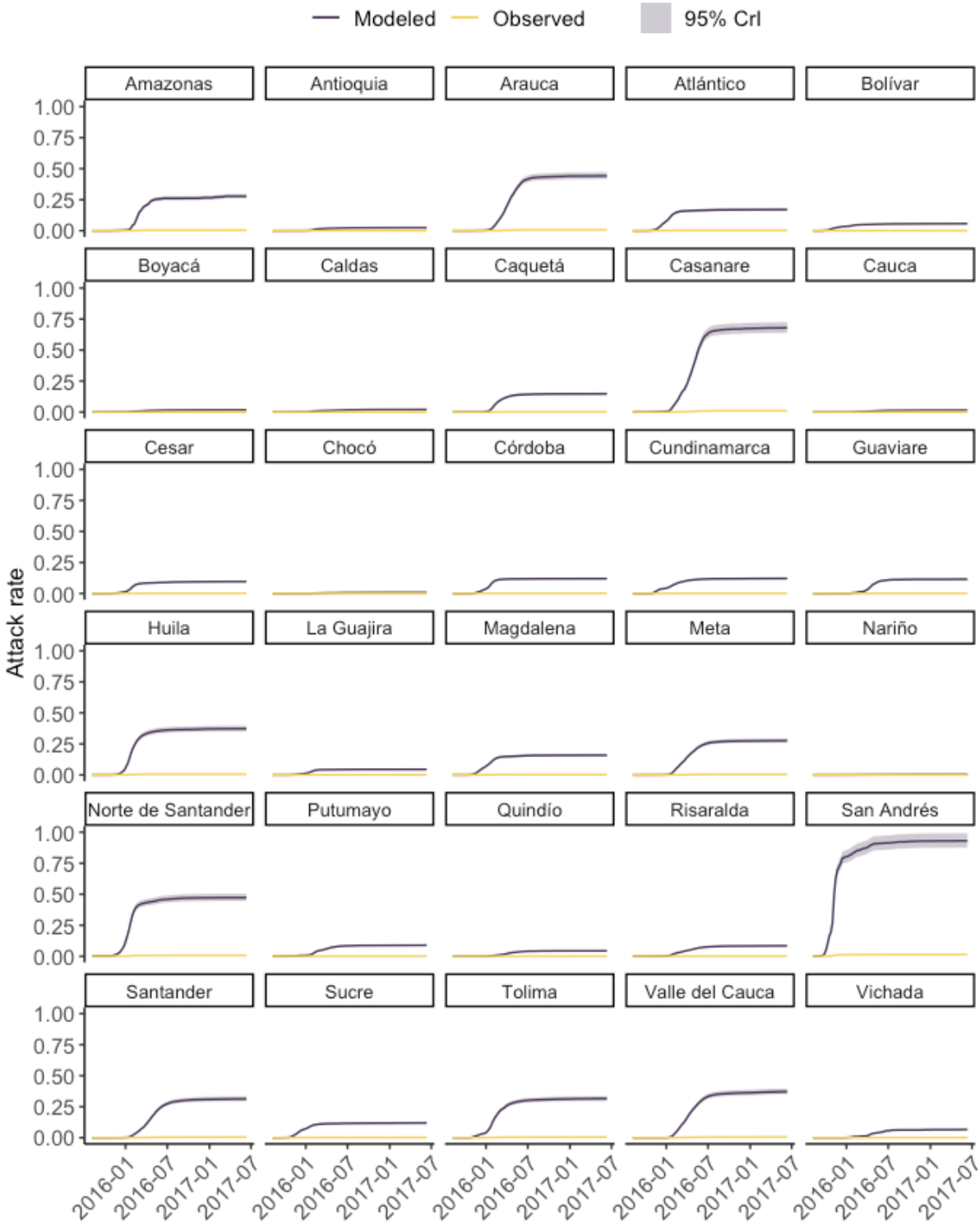


Figure 3.21 Modeled and observed attack rates of ZIKV by department. Yellow lines show the observed attack rates, while purple lines show the estimated infection attack rates, which were obtained by adjusting the observed attack rates by the estimated reporting rate. The reporting rate was estimated from the best-fitting negative binomial model. Purple shading represents the 95% CrI associated with this estimate.

Figure 3.22 Hexagon map of modeled attack rates of CHIKV and ZIKV. Estimated infection attack rates (IARs) were obtained by adjusting the observed attack rates by the estimated reporting rate. The reporting rates were estimated from the best-fitting negative binomial models. The color gradient represents the mean IAR across viruses. Note that the island department of San Andrés and Providencia is attached to the upper left side of the map.

4.8 Validation of parameter fitting procedure

Accurate parameter estimates were recovered from the Poisson model with weather covariates fitted to a simulated dataset created by simulating the epidemic with observed population sizes of 29 departments, observed weather data, and the generation time distribution for CHIKV (Table 3.19). Accurate parameter estimates were also recovered from the Poisson model with multiple R_0s (Figure 3.23 and Table 3.20) as well as from the negative binomial model with multiple R_0s (Figure 3.24 and Table 3.20), each fitted to one simulated dataset. It is unclear why the models with multiple R_0s are not able to recover all parameter estimates used to generate the data. It does not appear to be due to small sample sizes.

Table 3.19 True values and median parameter estimates obtained from a single dataset simulated from the Poisson model with weather covariates as well as observed population sizes for 29 departments, observed weather data, and the generation time distribution for CHIKV.

Parameter	True values	Estimated values from simulated data (95% CrI)
ρ (reporting rate)	0.05	0.050 (0.050-0.051)
R_0	1.50	1.50 (1.49-1.51)
$X^{best(temp)}$ (°C)*	26.0	25.9 (25.4-26.3)
σ_{temp1} (°C)**	15.0	14.9 (13.2-16.8)
σ_{temp2} (°C)**	8.0	8.0 (6.9-9.5)
$X^{best(rain)}$ (mm)***	480	483 (478-488)
σ_{rain} (mm)	500	507 (499-515)

*Mean weekly temperature averaged over three weeks followed by a three-week lag prior to case reporting.
** σ_{temp1}: standard deviation associated with temperatures below $X^{best(temp)}$, σ_{temp2}: standard deviation associated with temperatures greater than or equal to $X^{best(temp)}$.
***Cumulative weekly rainfall summed over six weeks followed by a two-week lag prior to case reporting.

Table 3.20 True values and median parameter estimates obtained from a single dataset simulated from either the Poisson model with multiple R_0s or the negative binomial model with multiple R_0s. Data were simulated using the observed population sizes for 29 departments and the generation time distribution for CHIKV.

	Poisson with multiple R_0s		Negative binomial with multiple R_0s	
	True values	**Estimated values from simulated data (95% CrI)**	**True values**	**Estimated values from simulated data (95% CrI)**
ρ (reporting rate)	0.04	0.04 (0.04-0.04)	0.03	0.03 (0.03-0.03)
R_0, range	1.07-3.45	0.85-3.44	1.11-4.66	1.10-9.98
ϕ (overdispersion)			1.5	1.45 (1.33-1.58)

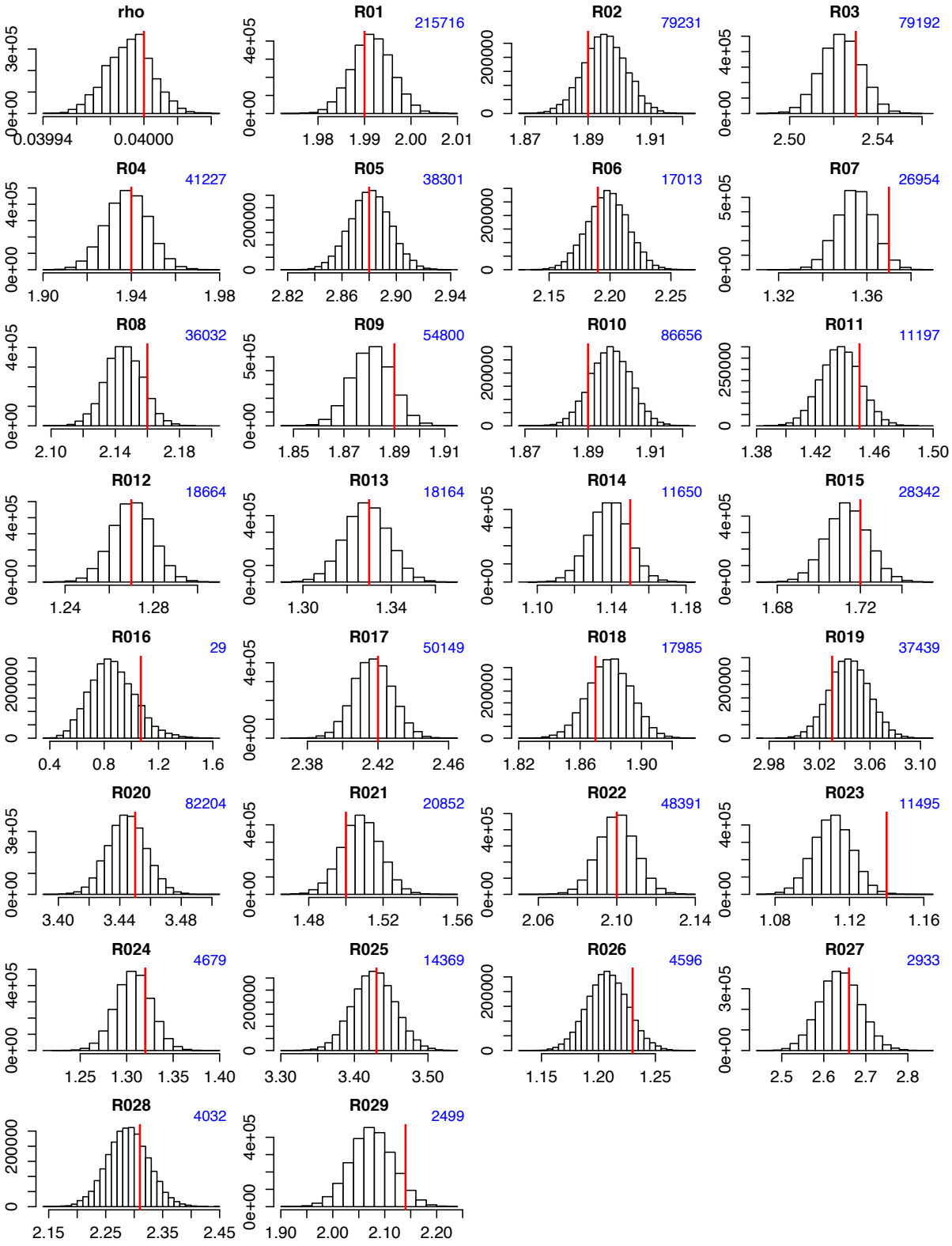


Figure 3.23 Histograms of the posterior distribution of parameters obtained from a single dataset simulated from a Poisson model allowing for different R_0s across departments. Observed population sizes for 29 departments and the generation time distribution for CHIKV were used to simulate the data. In each plot, the red line is the true value used to simulate the data. The blue text in the upper right corner of each plot is the simulated number of cases.

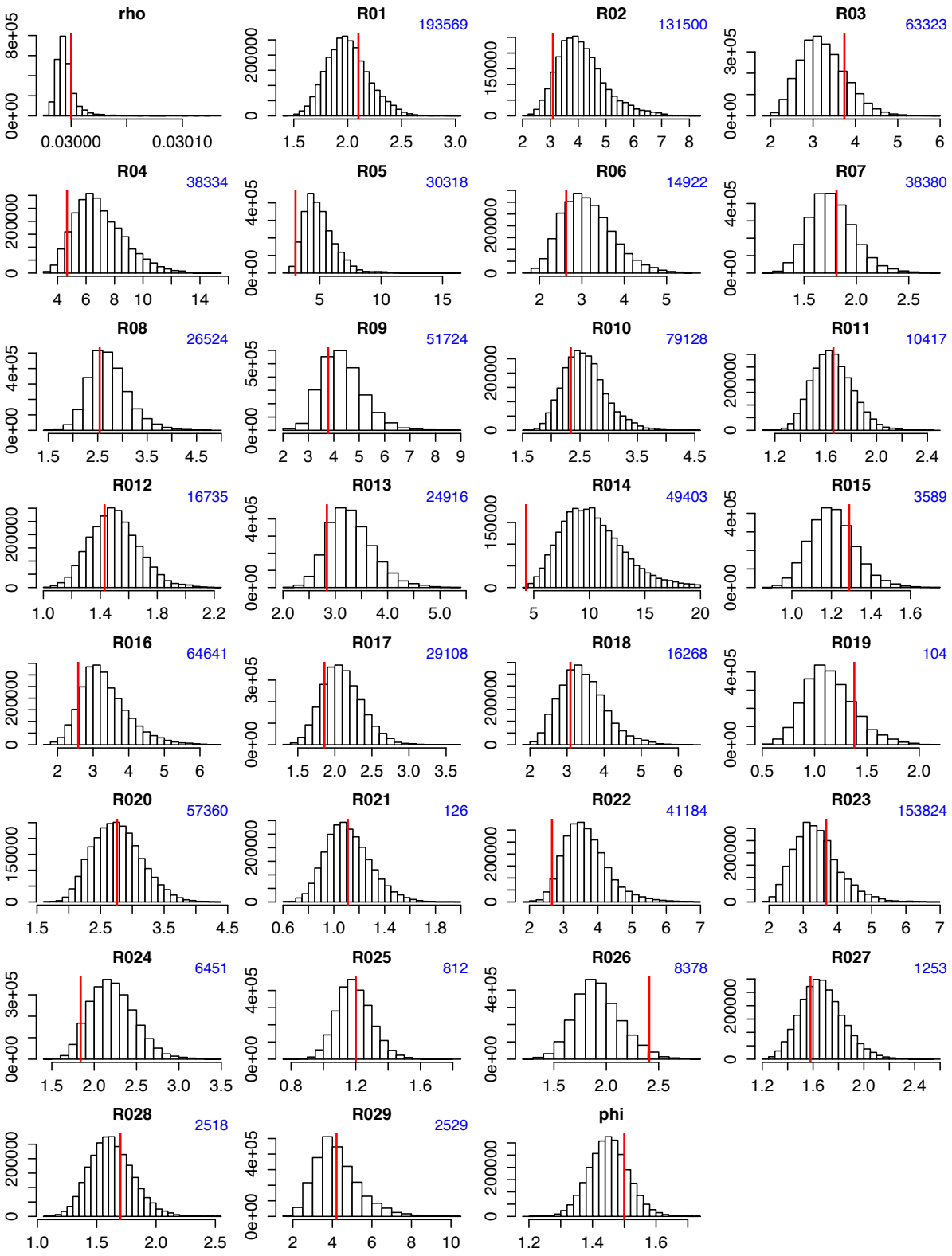


Figure 3.24 Histograms of the posterior distribution of parameters obtained from a single dataset simulated from a negative binomial model allowing for different R_0s across departments.
Observed population sizes for 29 departments and the generation time distribution for CHIKV were used to simulate the data. In each plot, the red line is the true value used to simulate the data. The blue text in the upper right corner of each plot is the simulated number of cases.

4.9 Sensitivity analysis of outlier thresholds

Figures 3.25-3.26 show estimated R_ts from the best-fitting Poisson models when no threshold is used to exclude outliers in the distribution of incidence divided by infectivity from contributing to the likelihood. These plots are nearly identical to Figures 3.15-3.16. Figures 3.27-3.28 show estimated R_ts from the best-fitting negative binomial models that did not use thresholds to ignore the outliers. The plots for CHIKV in Figure 3.27 appear noticeably different compared to those in Figure 3.17. In particular, Antioquia, Caldas, Chocó, and Nariño all have much higher R_ts that consistently remain above 1 throughout the epidemic. After excluding the outliers from fitting, the correlation between the EpiEstim R_ts and the R_ts from CHIKV's best-fitting negative binomial model increased from 0.06 to 0.16. In contrast, the plots for ZIKV in Figure 3.28 closely resemble those shown in Figure 3.18. Exceptions include Bolívar and Cundinamarca; in Figure 3.28, the model R_ts for these departments do not decrease below the threshold of 1 at the end of the epidemic in contrast with those in Figure 3.18.

Plots of the estimated R_ts from the best-fitting negative binomial models using thresholds for the incidence-to-infectivity ratio of 15 and 55 can be found in Appendix S1. For CHIKV, 55 was too high as the model R_ts for some departments were above 1 at the end of the epidemic. R_t estimates from models that used thresholds of 15 and 20 respectively are nearly indistinguishable, so the more conservative threshold of 20, which includes 98% of the data, was preferred. Similar to CHIKV, the ZIKV model that used a threshold of 15 had similar results compared to the model that used a threshold of 20. With a threshold of 55, model estimates were slightly worse, with a higher R_t initially estimated for San Andrés and Providencia. Also, Bolívar and Cundinamarca have slightly higher final estimated R_ts, which are above 1.

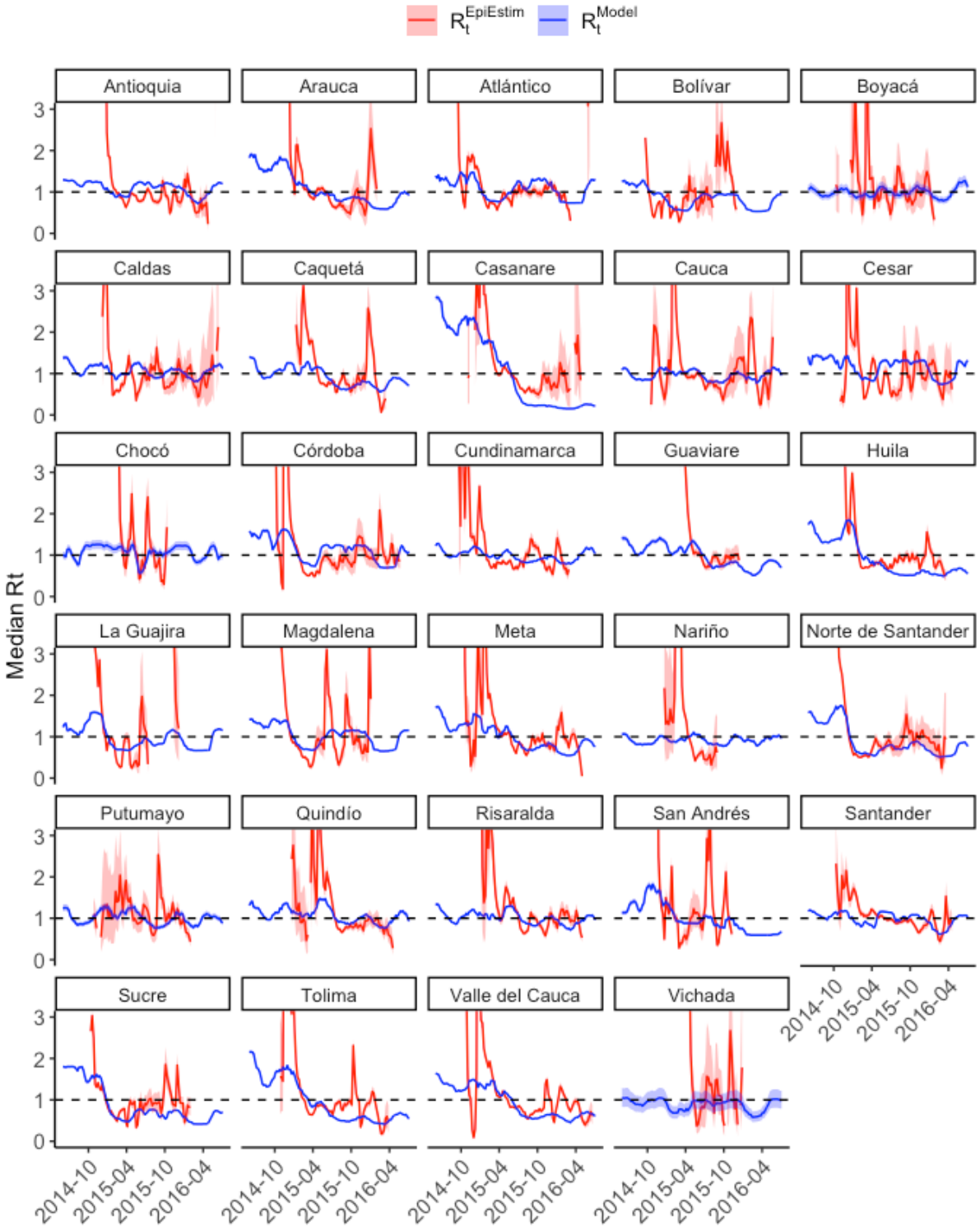


Figure 3.25 Comparing median estimates of R_t from the best-fitting Poisson model for CHIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines) when no threshold is used to prevent outliers in the distribution of incidence divided by infectivity from contributing to the likelihood. EpiEstim R_ts are plotted in the center of the 5-week window used to compute each estimate. Shaded areas represent 95% CrI. There is a positive statistically significant correlation of 0.20 (Pearson's correlation coefficient, 95% CI: 0.16-0.24, $p < 0.0001$).

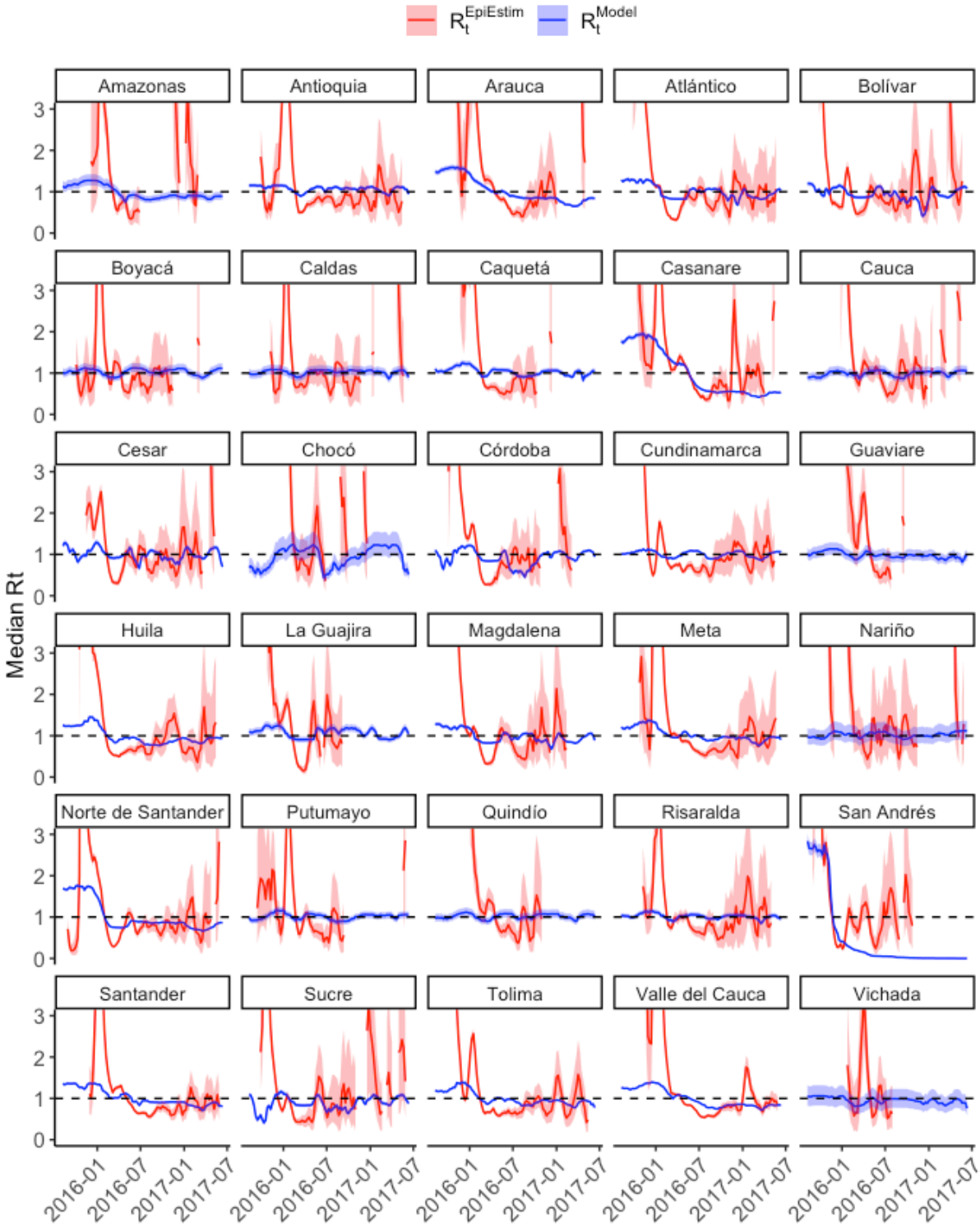


Figure 3.26 Comparing median estimates of R_t from the best-fitting Poisson model for ZIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines) when no threshold is used to prevent outliers in the distribution of incidence divided by infectivity from contributing to the likelihood. EpiEstim R_ts are plotted in the center of the 5-week window used to compute each estimate. Shaded areas represent 95% CrI. There is a positive statistically significant correlation of 0.33 (Pearson's correlation coefficient, 95% CI: 0.29-0.37, $p < 0.0001$).

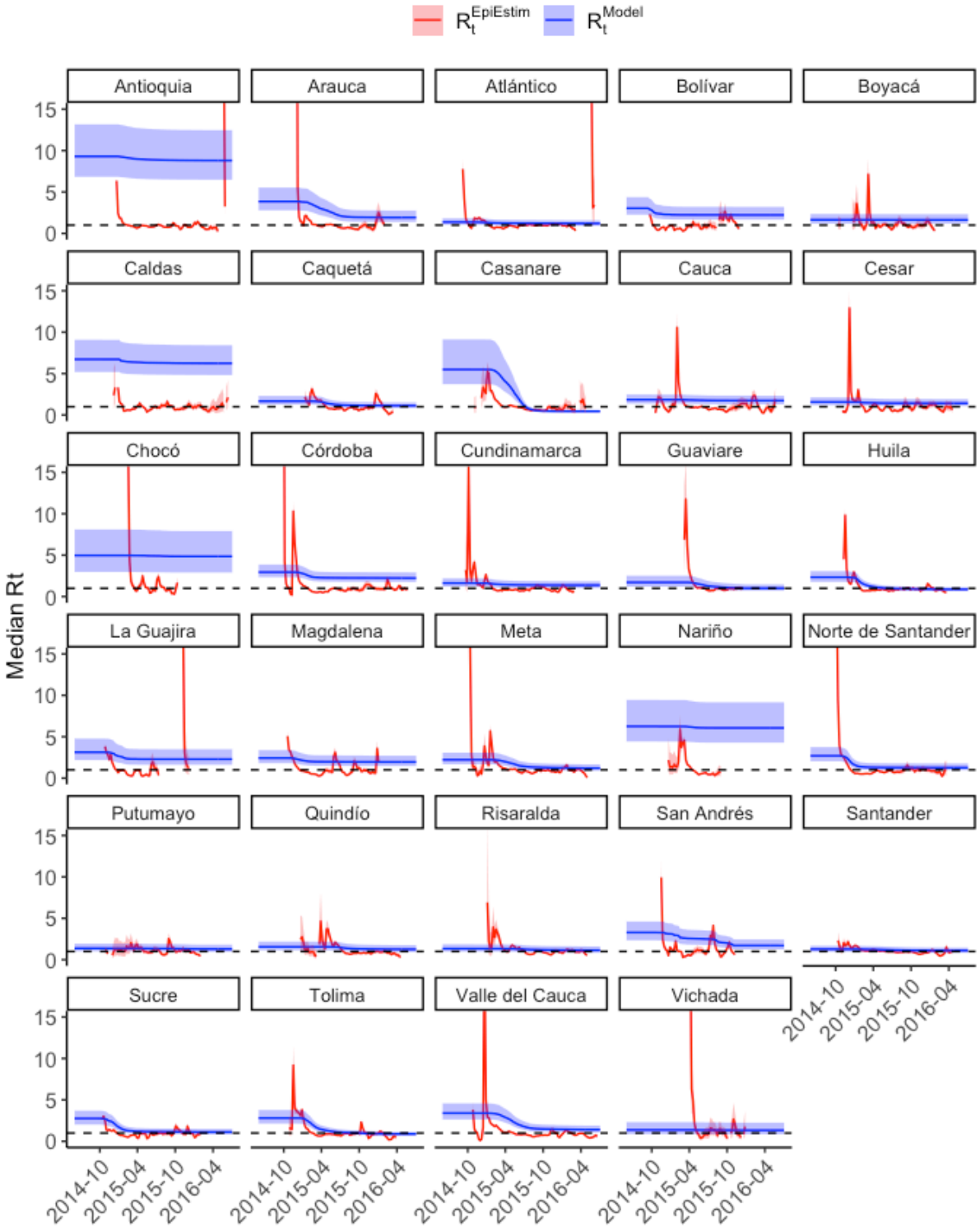


Figure 3.27 Comparing median estimates of R_t from the best-fitting negative binomial model for CHIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines) when no threshold is used to prevent outliers in the distribution of incidence divided by infectivity from contributing to the likelihood. EpiEstim R_ts are plotted in the center of the 5-week window used to compute each estimate. Shaded areas represent 95% CrI. There is a positive statistically significant correlation of 0.06 (Pearson's correlation coefficient, 95% CI: 0.01-0.10, p = 0.01).

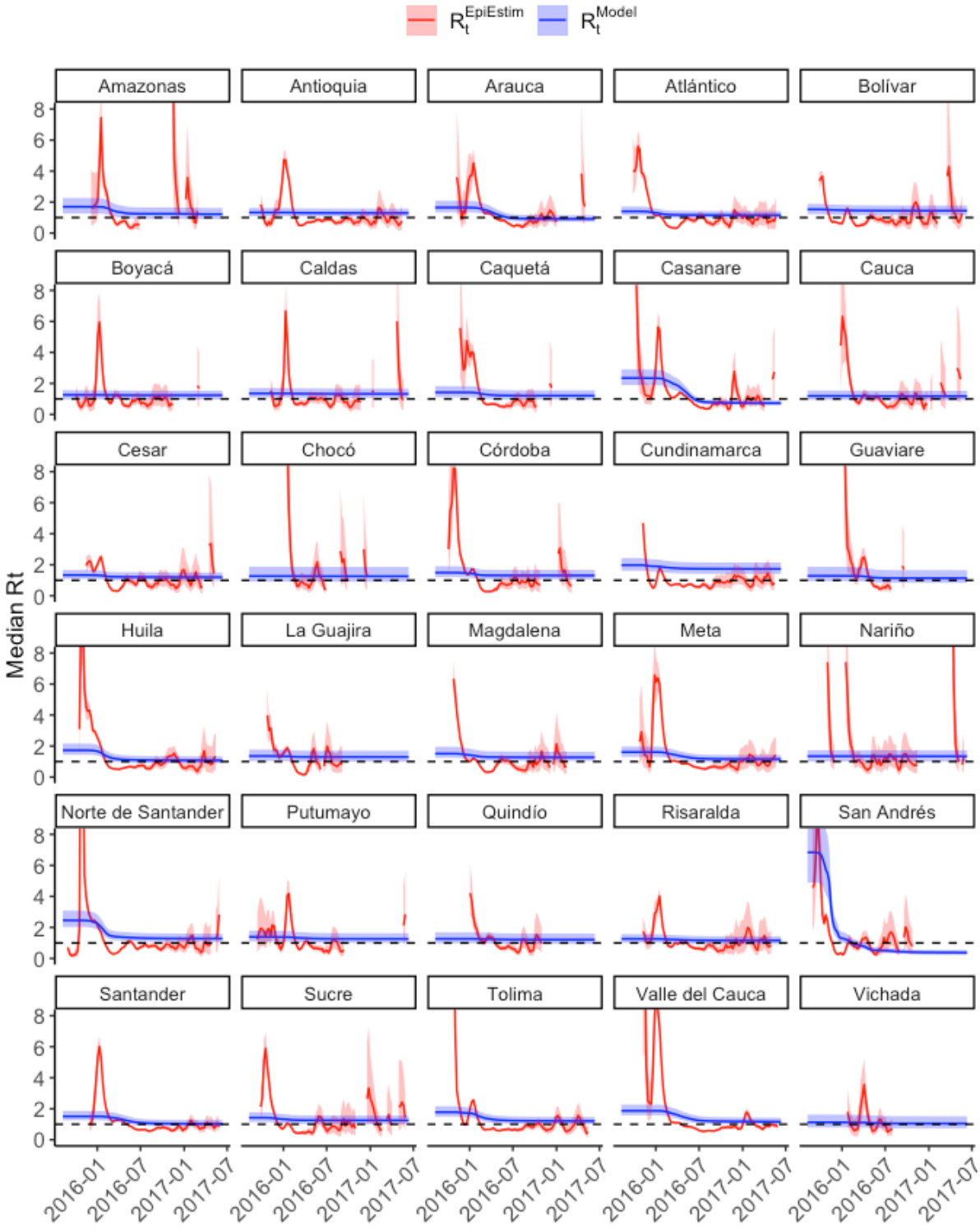


Figure 3.28 Comparing median estimates of R_t from the best-fitting negative binomial model for ZIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines) when no threshold is used to prevent outliers in the distribution of incidence divided by infectivity from contributing to the likelihood. EpiEstim R_ts are plotted in the center of the 5-week window used to compute each estimate. Shaded areas represent 95% CrI. There is a positive statistically significant correlation of 0.29 (Pearson's correlation coefficient, 95% CI: 0.25-0.33, $p < 0.0001$).

4.10 MCMC testing

The diagnostics in this section correspond to each individual virus' best-fitting negative binomial model from Table 3.17. Model diagnostics for the best-fitting Poisson models can be found in Appendix S2.

4.10.1 Convergence diagnostics

The models were run from three different starting points to ascertain convergence. Table 3.21 shows the Gelman-Rubin statistic for each of the best-fitting negative binomial models after removing the burn-in. All point estimates and 95% CI equal 1, suggesting model convergence. Figures 3.29-3.30 show the posterior distributions of one MCMC chain for each parameter after removing the burn-in. All the distributions are close to normal, suggesting that the chains converged.

Table 3.21 Gelman-Rubin statistic for each of the best-fitting negative binomial models (after removing the burn-in).

	CHIKV		ZIKV	
Parameter	**Point estimate**	**Upper CI**	**Point estimate**	**Upper CI**
ρ (reporting rate)	1	1	1	1
R_0 (all)	1	1	1	1
ϕ (overdispersion)	1	1	1	1

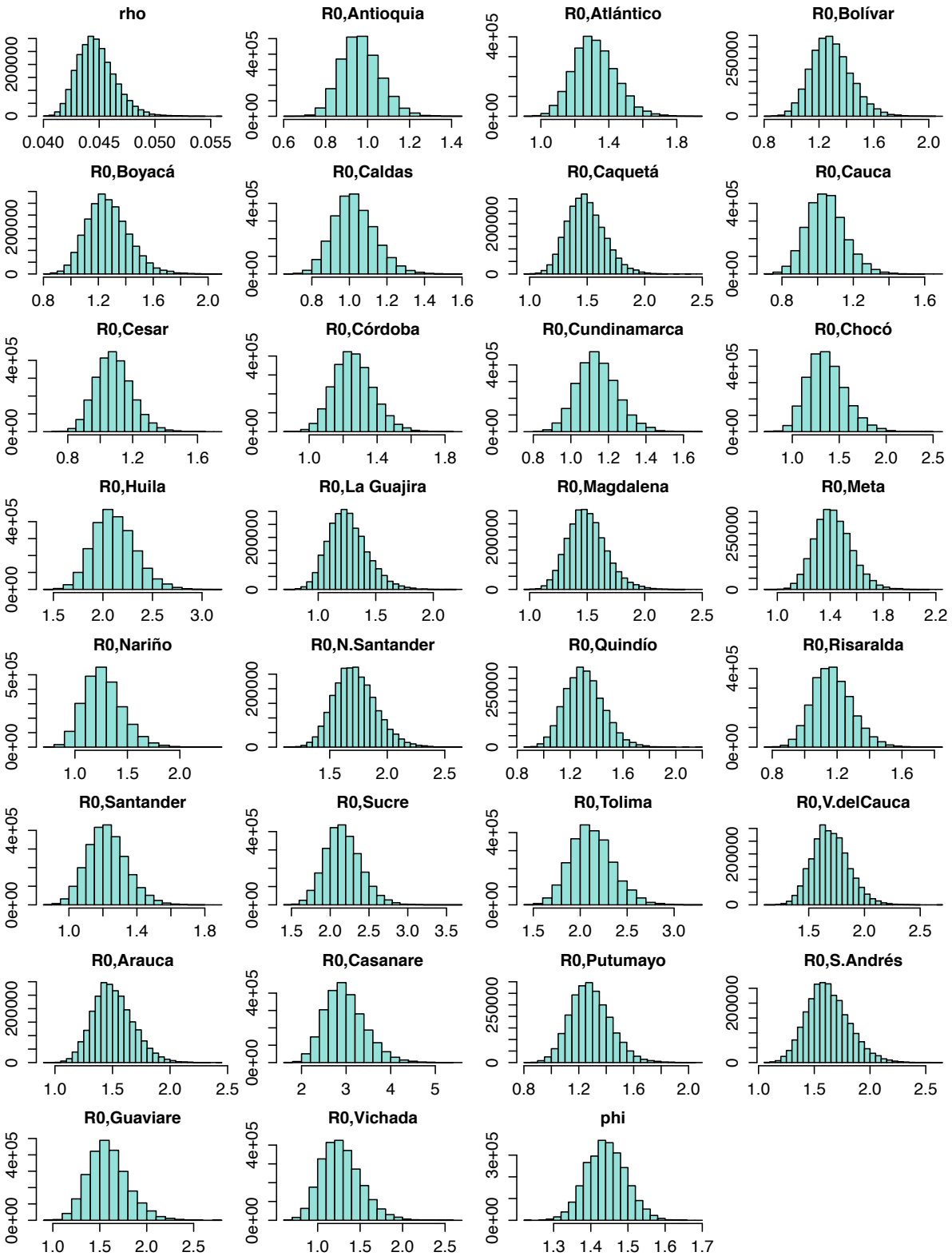


Figure 3.29 Histograms of the posterior distributions of the best-fitting negative binomial model for CHIKV.

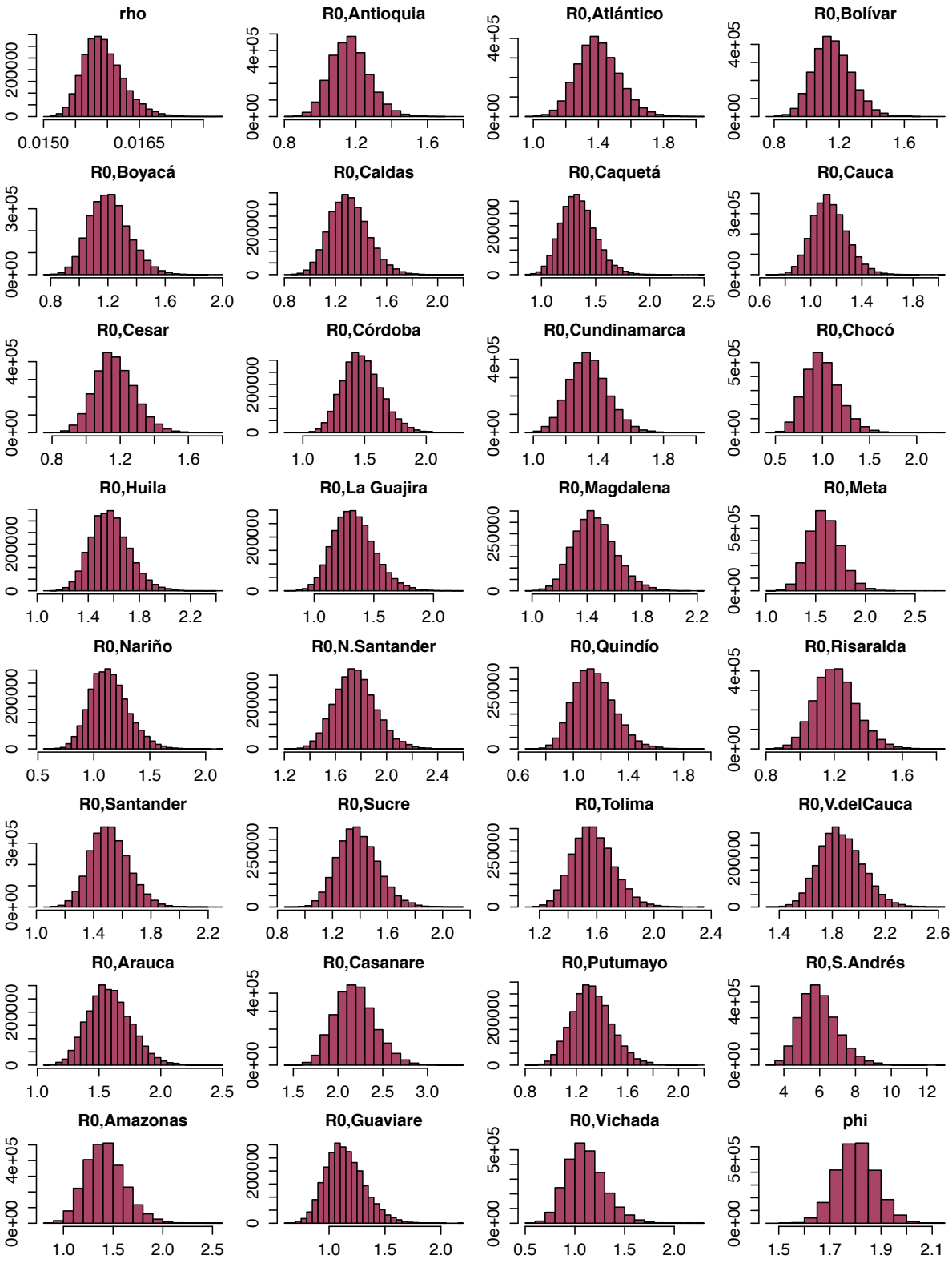


Figure 3.30 Histograms of the posterior distributions of the best-fitting negative binomial model for ZIKV.

4.10.2 Traces

Figures 3.31-3.32 show the MCMC traces for three chains of the CHIKV and ZIKV models, respectively. Mixing is good for all parameters based on visual assessment.

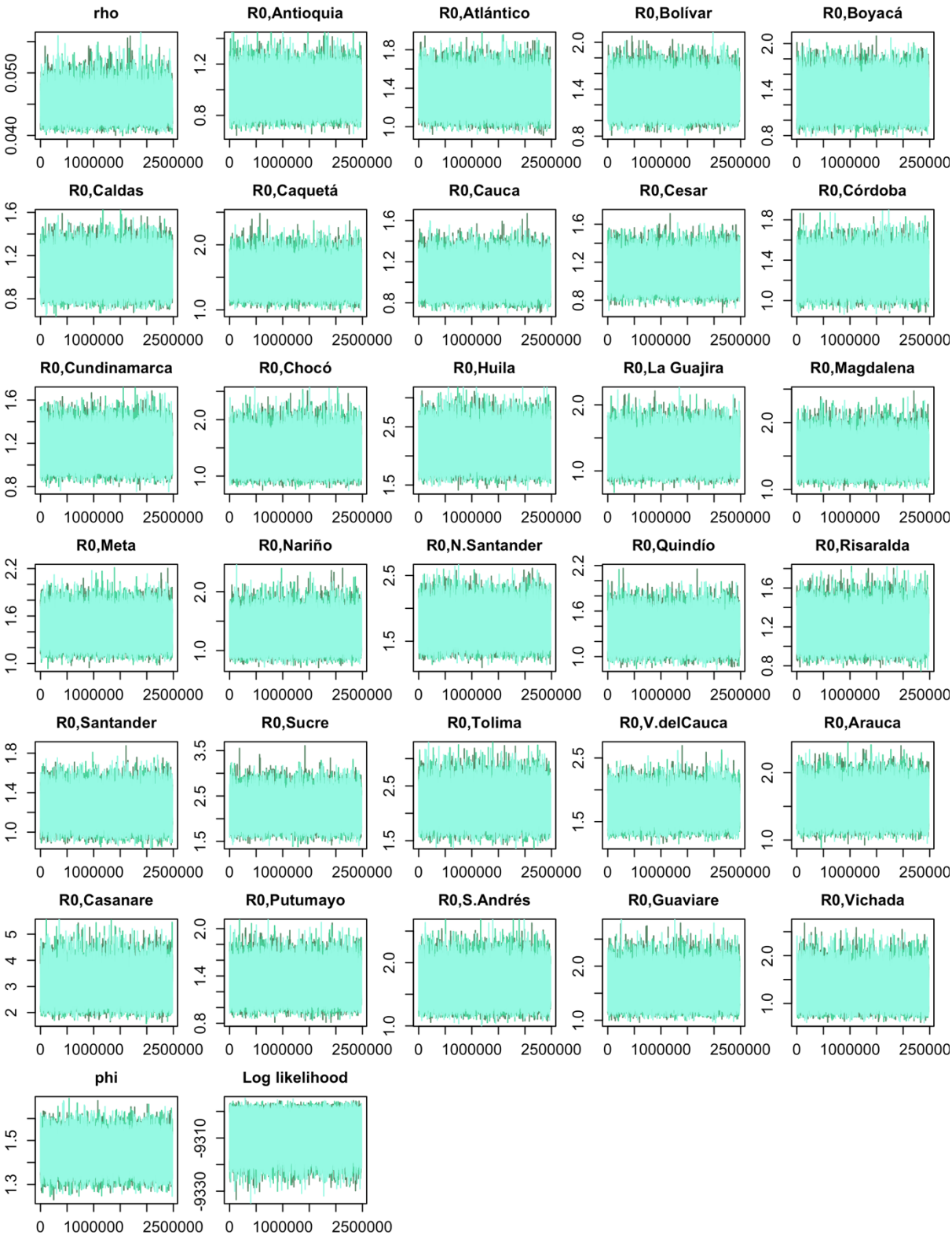


Figure 3.31 MCMC traces for the CHIKV model. Three chains run using different start values are shown.

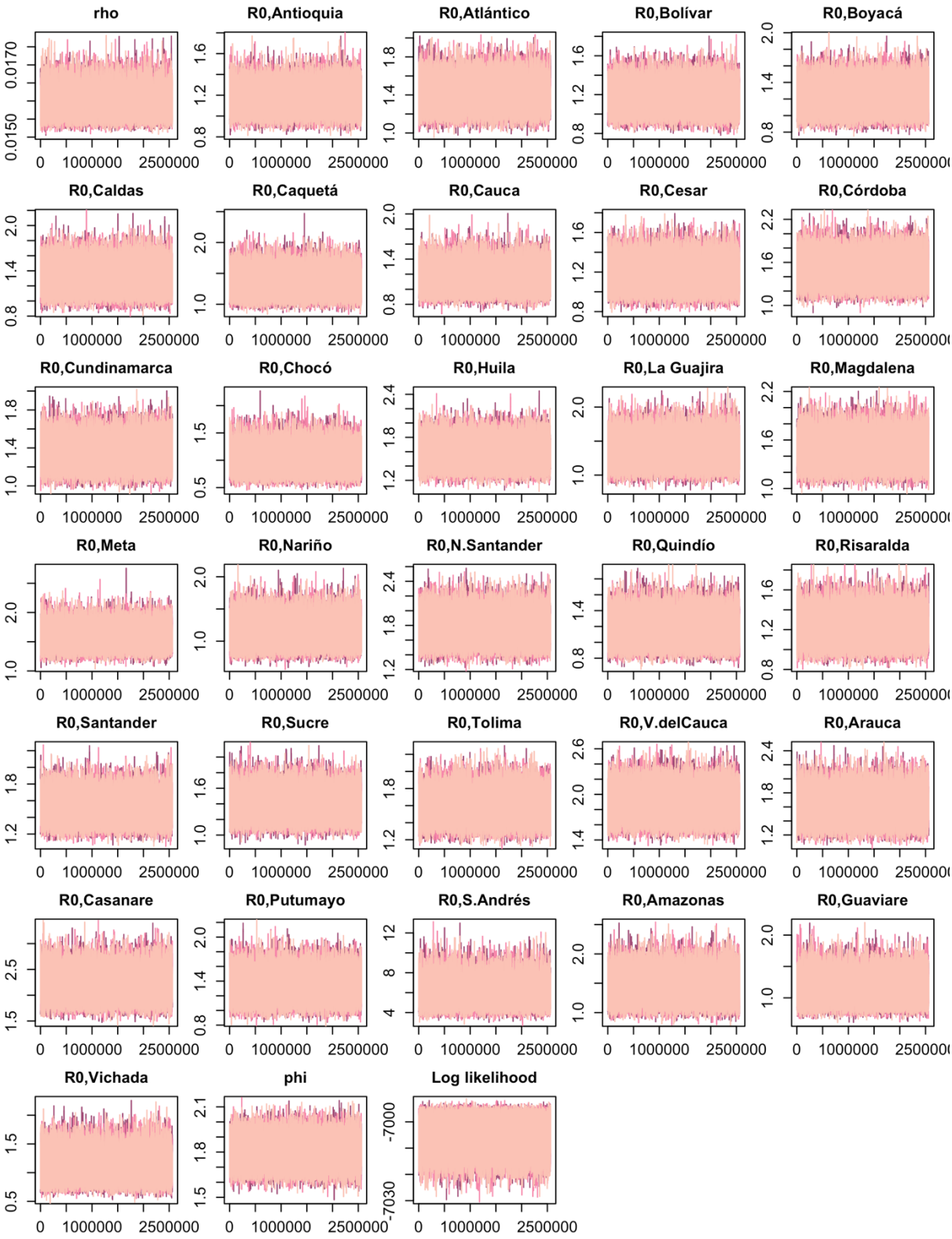


Figure 3.32 MCMC traces for the ZIKV model. Three chains run using different start values are shown.

4.10.3 Acceptance rate and effective sample size

Table 3.22 shows the acceptance rate of parameters for the CHIKV and ZIKV models. Both models have good acceptance rates. Table 3.23 shows the calculation of the effective sample size for each parameter after removing the burn-in. All parameters have good effective sample sizes (most are at least 10% of the total number of iterations).

Table 3.22 Acceptance percentages for parameters of the best-fitting negative binomial models for CHIKV and ZIKV (after removing the burn-in).

Parameter	CHIKV	ZIKV
ρ (reporting rate)	34.2	32.4
Department-specific R_0, range	27.2-35.1	25.9-37.4
ϕ (overdispersion)	18.0	24.5

Table 3.23 Effective sample sizes from one chain for each of the best-fitting negative binomial models (after removing the burn-in).

Parameter	CHIKV	ZIKV
ρ (reporting rate)	4,660	7,537
Department-specific *R0*, median (range)	13,794 (6,182-15,071)	18,070 (8,038-21,078)
ϕ (overdispersion)	8,597	14,422

5 Discussion

In this chapter, reporting rates and reproduction numbers from the ZIKV and CHIKV epidemics in Colombia were estimated with non-parametric and parametric models based on the renewal equation. Both approaches incorporated the effects of weather and socioeconomic status and were conducted at the department level.

The largest proportions of CF and ZVD cases during the epidemics were reported in Valle del Cauca. This department reported 27% of CF cases and 26% of ZVD cases despite making up less than 10% of the country's population. According to a systematic review, only 13% of DF cases in Colombia between 2000-2011 were reported in the Pacific Coastal region with Valle del Cauca accounting for most of those cases [228]. Cali, the capital of Valle del Cauca, is considered hyperendemic for DENV [229] which could explain the discrepancy. High levels of herd immunity in areas that are mesoendemic or hyperendemic for DENV would be

expected to decrease transmission even during years with major national outbreaks, such as 2010. Another explanation could be that Valle del Cauca had a higher reporting rate of CF and ZVD compared to other departments.

5.1 Non-parametric models of arbovirus transmission

As the non-parametric approach relied on median R_ts from EpiEstim, uncertainty surrounding R_t was not accounted for in the initial estimates of R_0 and the reporting rate ρ.

From the linear regression models of the R_t estimates versus the cumulative incidence of reported cases divided by the population of each department, rough approximations of the R_0 across all departments were obtained from the estimated y-intercepts. An estimated R_0 of 1.71 (95% CI: 1.54-1.88) for CHIKV is consistent with results from Peña-García and Christofferson who found that 76% out of 85 Colombian cities in their analysis had estimated R_0 values for CHIKV between 1-2 [177]. An estimated R_0 of 1.69 (95% CI: 1.59-1.78) for ZIKV was similar to some estimates in the literature (1.41, 95% CI: 1.15-1.74 in San Andrés [178] and 1.89, 95% CI: 1.21-2.13 nationally [187]) but lower than several others (for example, 4.61, 95% CI: 4.11-5.16 in Girardot [178]; 3.8, 95% CI: 2.4-5.6 in Barranquilla [179]; 10.3, 95% CI: 8.3-12.4 and 2.2, 95% CI: 1.9-2.8 in the department of Antioquia, depending on assumptions [174]). Depending on model assumptions, the estimate here was similar or lower than estimates from two studies [182, 186]. Only one study had an estimated R_0 lower than that presented here, with a median of 1.12 across 20 cities in the department of Antioquia [180]. Again, differences could be related to different spatial and temporal scales used in the analyses.

Rough approximations of the national ρ of CHIKV and ZIKV infections were also obtained from the linear regression models with 0.044 (95% CI: 0.032-0.056) for CHIKV and 0.015 (95% CI: 0.012-0.018) for ZIKV. A lower ρ for ZIKV would be expected due to a much higher rate of asymptomatic infection compared to CHIKV [10, 35]. Riou et al. found that only 19% of ZVD cases were reported compared to 40% of all CF cases in French Polynesia and the French West Indies [199]. In Colombia, two community-based studies estimated a symptomatic reporting rate for CHIKV more than two times higher than the estimated ρ in this study (0.129, 95% CI: 0.127-0.132 in Girardot and 0.097, 95% CI: 0.096-0.098 in El Tolima). This finding is expected as symptomatic reporting rates do not account for

asymptomatic infections, which comprise 3%-25% of all CHIKV infections [35]. The CHIKV estimate here was consistent with the results from Nouvellet et al.'s study which reported seroprevalence estimates that implied ρ between <0.001 in Medellín and 0.099 in Neiva with mean 0.057 [191]. The estimated ρ for ZIKV in this study was lower than that reported by O'Reilly et al. and Mier-y-Teran-Romero et al. (0.017, 95% CrI: 0.013-0.025 and 0.03, 95% CrI: 0.01-0.07, respectively) [23, 133]. As expected, it was also lower than the symptomatic reporting rates estimated by Martínez Duran et al. and Moore et al. (0.081, 95% CI: 0.076-0.086 and 0.036, 95% CrI: 0.018-0.070, respectively) [192, 193]. The estimate here was similar to that implied by the seroprevalence estimates reported by Nouvellet et al., ranging from 0.003 in Medellín to 0.021 in Cúcuta (mean 0.012) [191].

When GAMs were fitted to the residuals of the linear regression models, the best-fitting model for CHIKV only included temperature. In addition to temperature, the percentage of households with overcrowded conditions and the percentage of households with inadequate exterior walls were significant covariates in the best-fitting ZIKV model. A unimodal effect of temperature was seen for ZIKV but not CHIKV with edfs of 16.7 and 1.00, respectively.

5.2 Parametric models of arbovirus transmission

The best-fitting Poisson models that estimated ρ, R_0, and a function for temperature used different definitions of temperature for CHIKV and ZIKV (four-week lag for CHIKV versus no lag for ZIKV). Similarly, models that estimated ρ, R_0, and a function for rainfall also defined rainfall slightly differently for CHIKV compared to ZIKV (two-week lag for CHIKV versus three-week lag for ZIKV). Given that urban epidemics of CHIKV and ZIKV are typically associated with different species of mosquitoes (*Ae. albopictus* and *Ae. aegypti*, respectively) [25, 230], differences could be attributed to different life history traits of the vectors. The plots of predicted R_t as a function of temperature and rainfall from the Poisson models with weather covariates showed that $R_t > 1$ for a larger range of temperature and rainfall combinations for CHIKV compared to ZIKV. In contrast, Brady et al. found that *Ae. aegypti* could withstand a greater range of temperatures, including lower temperatures, compared to *Ae. albopictus* [198].

From the overall best-fitting models (negative binomial model with multiple R_0s), department-specific R_0s ranged from 0.96-2.93 for CHIKV and 0.98-5.87 for ZIKV with median 1.30 and 1.33, respectively. Compared to the single R_0s estimated for each virus in the non-parametric approach, the medians of the R_0s are lower. The estimated ρ was 0.045 (95% CrI: 0.042-0.049) for CHIKV compared to 0.016 (95% CrI: 0.015-0.017) for ZIKV. Although the point estimates for both viruses are slightly higher than those from the non-parametric approach, they are consistent with a higher ρ for CHIKV compared to ZIKV.

While some of the departments' estimated infection attack rates were within the 95% confidence interval of the seroprevalence estimates for their corresponding cities, others were mostly lower. This result could be related to the different spatial scales considered in the studies. A range of infection attack rates between cities would be expected due to within-department heterogeneity in elevation and climate [52], factors that are associated with arbovirus transmission [203, 231]. Higher estimated infection attack rates in some departments are consistent with other studies in the Americas that have reported high seroprevalence following CHIKV and ZIKV epidemics [232]. At the same time, lower estimated infection attack rates could be attributed to more people living in areas at high elevation with negligible risk of arbovirus transmission, such as Antioquia.

As expected, 25 out of 29 departments had similar estimated R_0s across viruses. Riou et al. found no significant difference in the transmissibility of CHIKV and ZIKV in French Polynesia and the French West Indies (relative transmission of ZIKV compared to CHIKV of 1.04, 95% CrI: 0.97-1.13) [199]. Funk et al. also found that the R_0s for ZIKV and DENV were similar when estimated in the same location (7.6, 95% CrI: 4.8-14 and 11, 95% CrI: 8.0-16 for ZIKV and DENV respectively on the Yap Main Island) [233]. Future work could involve jointly fitting a model that estimates a single R_0 across viruses for each department.

The finding that models with multiple R_0s fitted better than models with a single R_0 was expected because this parameter is context specific. In particular, the contact rate between humans and vectors influences R_0 [234] and likely varies across Colombian departments due to differences in factors such as altitude, population density, and seasonality. The fact that weather parameters became more difficult to estimate in models with multiple R_0s was also

expected as between-department differences in these parameters are considerably greater than within-department differences.

The negative binomial models were more sensitive to the thresholds for outliers in the distribution of incidence divided by infectivity than the Poisson models. This finding is likely due to the estimation of overdispersion. The Poisson models assume no overdispersion (the mean equals the variance). For CHIKV, the estimate for ϕ from the negative binomial model with multiple R_0s more than doubles when the threshold is used to remove the outliers from 0.62 (95% CrI: 0.58-0.66) to 1.44 (95% CrI: 1.34-1.55), suggesting less overdispersion. The estimated ϕ for ZIKV similarly increases from 1.44 (95% CrI: 1.33-1.56) to 1.80 (95% CrI: 1.66-1.95) with the threshold. Without the thresholds, the model must allow a few cases to give rise to many secondary cases to account for such outliers. Consequently, the estimates for R_0 were higher, and the estimates for ϕ were lower.

5.3 Conclusions and limitations

Two approaches for exploring predictors of R_0s and ρs were used in this chapter. As the methods for both approaches are based on the renewal equation, similar estimates were expected and obtained for both CHIKV and ZIKV. Given that uncertainty in the R_t estimates was not taken into account in the EpiEstim approach, the parametric model approach is likely more accurate.

One limitation of this analysis was that it was not possible to estimate both department-specific ρs and department-specific R_0s in the parametric models as there was not enough power. Consequently, the same ρ was assumed for all departments. Although both parameters were estimated for each department from linear regression models, uncertainty was high. Moore et al. found a moderate amount of variability in the reporting rate of symptomatic ZIKV infection across Colombian departments. Their estimates were as low as <0.001 (95% CrI: <0.001-<0.001) in Bogotá and as high as 0.145 (95% CrI: 0.061-0.304) in Cundinamarca compared to an overall estimate of 0.036 (95% CrI: 0.018-0.070) for Colombia. Reassuringly, the estimated ρ from the best-fitting negative binomial models here were similar to the observed reporting rates from the four Colombian cities with available seroprevalence estimates, which had mean 0.057 for CHIKV and 0.012 for ZIKV [191].

In the parametric models, the reporting rate does not affect the estimate of R_t in the first week; however, in subsequent weeks, a higher reporting rate would lead to higher estimates of R_t over time because $R_t \sim 1 - \frac{1}{\rho}$. Similarly, regional variation in reporting would lead to lower or higher R_t estimates depending on whether reporting was lower or higher, respectively, in each department.

The parametric model assumption of a single ρ across departments also carries over to the estimation of the infection attack rate from the number of reported cases. An additional assumption involved in estimating the infection attack rate (and R_t) is that the reporting rate did not change over time. The reporting rate could have changed during the epidemics if, for example, hospital capacity was reached or public health policies changed. Moreover, reporting rates could have increased as more healthcare providers became aware of the new diseases.

Missing early CHIKV incidence data posed problems for estimating R_0s and R_ts. Model fits were improved by excluding outliers in the ratio of incidence and infectivity. Without these exclusions, the model had no way of smoothing over the data unlike EpiEstim, which can do so with longer user-specified window lengths. Another way to address the missing infection dates problem is to characterize a distribution of the time from infection to being reported as a case. Using the symptom onset of cases and the incubation period distribution, the incidence time series can be back-calculated [170]. However, the resulting time series may be over-smoothed compared to the observed time series, and unfortunately, symptom onset was not known for all CF cases. The method also does not solve the problem of missing case data.

This analysis was conducted at the department level rather than at the city level to better understand the effects of weather on CHIKV and ZIKV transmission as well as avoid issues related to estimating parameters from irregular time series with low case numbers. However, using data aggregated at the department level may mask important spatial variation in disease dynamics. Using a variation of the Disease Transmission Kernel (DTK)-Dengue model, Moore et al. modeled CHIKV transmission at three different spatial scales across Colombia. They fitted versions of the model to department- and national-level weekly case report data and found the models performed better at finer spatial scales [235].

Future research could involve conducting the analysis at the city level and comparing the results across spatial scales.

Chapter 4: Estimating Zika virus attack rates and risk of Zika virus-associated neurological complications in Colombian capital cities with a Bayesian hierarchical model

Abstract

Reporting rates as well as biases in ZIKV surveillance data were explored in previous chapters. Here, multiple data sources were combined to improve estimates of ZIKV infection attack rates, reporting rates of ZVD, and the risk of ZIKV-associated neurological complications in Colombia. ZVD surveillance data were combined with post-epidemic seroprevalence data and a dataset on ZIKV-associated neurological complications in a Bayesian hierarchical model for 28 capital cities. Models were also fitted by sex and by two age groups. Substantial heterogeneity was observed for the ZIKV infection attack rates across cities, ranging from 0.03 (95% CrI: 0.00-0.10) in Quibdó to 0.80 (95% CrI: 0.56-0.99) in San Andrés. The overall estimated infection attack rate for ZIKV across the 28 cities was 0.38 (95% CrI: 0.17-0.92). The estimated reporting rate for ZVD was 0.013 (95% CrI: 0.004-0.024), and 0.51 (0.17-0.92) cases of ZIKV-associated neurological complications were estimated to be reported per 10,000 ZIKV infections. When the same ZIKV infection attack rate was assumed across sex, females were more likely to be reported as ZVD cases to the surveillance system and less likely to be reported as ZVD cases with neurological complications compared to males. Similarly, when the same ZIKV infection attack rate was assumed across age group, younger individuals were more likely to be reported as ZVD cases to the surveillance system and less likely to be reported as ZVD cases with neurological complications compared to older individuals. Important differences in these estimates were also found for some cities. These results highlight how additional data sources can be utilized to overcome biases in surveillance data and estimate key epidemiological parameters.

1 Introduction

1.1 Background

The epidemiology of GBS in Latin America and the Caribbean is not well understood [236]. A recent systematic review on the incidence of GBS in the region identified only 10 papers with primary data from 1980 to 2014. An additional 21 papers related to the ZIKV epidemic were found between 2015 and 2018 [236]. Data were not pooled to estimate the annual incidence rate of GBS in Latin America due to substantial heterogeneity across studies. Only one study estimated background rates of GBS in Colombia. Although the estimate was not provided in the original study, the study authors reported in personal communication an annual incidence rate of 0.95 (95% CI: 0.73-1.22) per 100,000 persons [236]. They also estimated that GBS diagnoses more than doubled during the peak of the ZIKV epidemic (IRR 2.29, 95% CI: 1.69-3.14) compared to baseline rates [236]. Five studies from Colombia estimated the incidence rate of GBS during the ZIKV epidemic. Those estimates ranged from 0.31 (95% CI: 0.23-0.41) per 100,000 per year among children aged one month to 18 years nationally to 7.63 (95% CI: 6.16-9.35) per 100,000 per year among both children and adults in the city of Barranquilla [236].

As mentioned in chapter 1, Colombia's INS collated data on ZIKV-associated neurological complications reported during the ZIKV epidemic. Unlike the ZIKV surveillance data, which was biased toward pregnant females and females of child-bearing age, data on ZIKV-associated neurological complications were expected to be more reliable due to the severity of symptoms and the fact that cases' medical records were reviewed against standardized case definitions. These data could be used in combination with other data, such as seroprevalence data and the number of CZS cases, to estimate key epidemiological parameters that are difficult to quantify, such as the ZIKV infection attack rate and the risk of ZIKV-associated neurological complications.

As mentioned in the introduction to chapter 3, Mier-y-Teran-Romero et al. used publicly available data on suspected ZVD and GBS cases for 11 countries or territories as well as post-epidemic serosurveys from two locations (Yap Island and French Polynesia) to estimate the probability of ZIKV infection (infection attack rate), the proportion of ZIKV infections that are reported as suspected ZVD cases (reporting rate), and the risk of ZIKV-associated

GBS [133]. This study used cumulative data on reported GBS cases and ZVD cases aggregated at the country level. Due to limited data, associations between GBS risk and other factors such as age and sex were not explored. Also, the results were sensitive to omitting data from Yap Island and French Polynesia, the only two locations for which serological data were available.

A 2020 study by Moore et al. was also introduced in chapter 3. Similar to Mier-y-Teran-Romero et al., they also used a variety of data types to estimate reporting rates, ZIKV infection attack rates, the probability of reporting symptomatic ZIKV infections with GBS, and the probability that a ZIKV infection during pregnancy results in a reported microcephaly case [192]. A limitation of their study was that their infection attack rate estimates did not agree with estimates from seroprevalence studies for some locations such as Bahia, Brazil. One possible explanation for this result is that their estimate was for the state level, while the seroprevalence data corresponded with a particular city within the state. Other estimates from the study should also be approached with caution. For example, the estimated reporting probabilities for ZIKV-associated microcephaly were much lower than the risk of CZS during pregnancy from other published studies [192]. However, the main aim of their analysis was to estimate the total number of ZIKV infections rather than individual epidemiological parameters.

1.2 Aims

In chapter 2, the analysis of the ZIKV line list data showed that compared to males, females had higher risk of being reported as a ZVD case but lower risk of being reported as a ZVD case with neurological complications. In chapter 3, rough estimates of the reporting rate of ZVD cases were obtained. Following from those chapters, the aim of this chapter is to use a Bayesian hierarchical model to estimate ZIKV infection attack rates, reporting rates of ZVD, and the risk of neurological complications following ZIKV infection in capital cities of Colombia. A secondary aim is to assess how these estimates vary by age and sex. As mentioned in previous chapters, neurological complications from ZIKV infection are severe and costly [107]. Having reliable estimates of the number of expected cases could improve resource allocation for future outbreaks.

2 Data

Three different datasets were used for this chapter. They are summarized in Table 4.1.

Table 4.1 Data sources for chapter 4.

Dataset	Definition	Level	Data type	Data availability	Time period
ZVD	Suspected and laboratory-confirmed	Line list	Weekly case counts	All cities	2015-2017
ZIKV	Age-stratified IgG seroprevalence	5-year age groups (range 2-45)	Cross-sectional post-epidemic seroprevalence	4 cities	Dec. 2016
ZIKV-associated neurological complications	Suspected	Line list	Weekly case counts	All cities	2015-2017

2.1 Epidemiological data

This chapter utilized the same Sivigila surveillance dataset as in previous chapters. A full description of these data can be found in chapter 1. Suspected and laboratory-confirmed cases with missing information on city location were excluded, resulting in 105,152 ZVD cases. As seroprevalence data were only available for four capital cities, the analysis was restricted to capital cities at risk of arbovirus transmission. A report by Colombia's MOH on the patterns of dengue endemicity in the country was used to determine which cities were at risk of arbovirus transmission. Criteria for level of dengue endemicity included trends of reported cases over time, number of circulating serotypes, age range of cases, and the presence of dengue hemorrhagic fever (severe dengue) from 2008 to 2013 [229]. Based on the classification of Tunja, Bogotá, Pasto, and Manizales as not endemic, they were excluded from the present analysis. All four cities have elevations between 2,160-2,810 m. There were 28 cities remaining for the analysis, with 54,737 cumulative ZVD cases (Table 4.2). As stated previously, the location of cases refers to location of likely infection, which is determined by the clinician who reported the case.

A dataset consisting of an anonymized line list on 418 patients with neurological complications and recent history of febrile illness compatible with ZVD was also used. A full description can be found in chapter 1. Briefly, medical records of patients with neurological

complications were reviewed using case definitions from the Brighton Collaboration Working Group for GBS, myelitis, encephalitis, and acute disseminated encephalomyelitis [98, 99]. Patients that did not meet Brighton case definition criteria 1-3 were removed from the dataset. After removing imported cases and cases with unknown department location, 406 patients remained for analysis. After further restricting the data to capital cities at risk of arbovirus transmission, 212 patients remained (Table 4.2). As mentioned in chapters 1 and 2, an equivalent dataset was not available for CHIKV.

2.2 Serological data

As discussed in the introduction to chapter 3, a household-based multisite seroprevalence study of arboviruses was conducted by Nouvellet et al. in the Colombian cities of Cúcuta, Medellín, Neiva, and Sincelejo between October and December 2016 [191]. Although the seroprevalence estimates from the study were sex-specific, no differences were found between males and females, and sex was not significant in a statistical analysis of risk factors (personal communication). No relationship between age and exposure to CHIKV or ZIKV was found. Table 4.2 shows key information about the four cities, including the estimated post-epidemic seroprevalence of ZIKV, as well as information about the other 24 capital cities.

2.3 Demographic data

As in previous chapters, population projections for 2016 based on the 2005 Census were obtained from DANE, including breakdowns by sex and age group. The population sizes of the 28 cities considered comprise 28% of Colombia's total population of about 49 million in 2016 (Table 4.2).

Table 4.2 Epidemiological and demographic data for 28 Colombian capital cities.

City	Department	Population in 2016	Reported cases of ZIKV-associated NC*	Reported suspected and laboratory-confirmed cases of ZVD	Estimated post-epidemic seroprevalence and 95% CI
Arauca	Arauca	89,712	1	788	
Armenia	Quindío	298,199	0	189	
Barranquilla	Atlántico	1,223,616	80	4,665	
Bucaramanga	Santander	528,269	8	4,322	
Cali	Valle del Cauca	2,394,925	23	16,279	
Cartagena	Bolívar	1,013,389	4	1,021	
Cúcuta	Norte de Santander	656,380	44	6,485	0.479 (0.440-0.519)
Florencia	Caquetá	175,407	3	663	
Ibagué	Tolima	558,805	3	4,076	
Inírida	Guainía	19,983	0	12	
Leticia	Amazonas	41,639	0	278	
Medellín	Antioquia	2,486,723	8	549	0.067 (0.048-0.090)
Mitú	Vaupés	31,861	0	17	
Mocoa	Putumayo	42,882	1	57	
Montería	Córdoba	447,668	4	1,785	
Neiva	Huila	344,026	13	3,409	0.578 (0.538-0.618)
Pereira	Risaralda	472,000	0	463	
Popayán	Cauca	280,054	0	51	
Puerto Carreño	Vichada	16,000	0	17	
Quibdó	Chocó	115,907	0	14	
Riohacha	La Guajira	268,712	0	279	
San Andrés	San Andrés & Providencia	71,946	0	1,109	
San José del Guaviare	Guaviare	65,611	0	154	
Santa Marta	Magdalena	491,535	2	1,913	
Sincelejo	Sucre	279,031	6	856	0.659 (0.620-0.696)
Valledupar	Cesar	463,219	2	788	
Villavicencio	Meta	495,227	5	2,377	
Yopal	Casanare	142,979	5	2,121	

*Neurological complications

3 Methods

3.1 Bayesian hierarchical model

Following Mier-y-Teran-Romero et al., a Bayesian hierarchical model was used to estimate the following probabilities from observed data via a binomial sampling process: (i) the

probability of ZIKV infection p_{Zl}, or infection attack rate, (ii) the probability of reporting a case of ZVD per ZIKV infection p_{Sl}, or reporting rate, and (iii) the probability of reporting a case of ZVD with neurological complications p_{CZl}, or reporting rate of ZIKV-associated neurological complications, where l denotes location (city) [133]. The probabilities of interest were estimated for the four cities with seroprevalence data as well as 24 other capital cities. Overall (non-location specific) estimates of the probabilities were also produced.

The total number of ZIKV infections Z_l and the number of infections that go on to be reported as either suspected or laboratory-confirmed cases S_l in each city l are binomially distributed as

$$Z_l \sim \text{Bin}(p_{Zl}, N_l),$$
$$S_l \sim \text{Bin}(p_{Sl}, Z_l),$$

where N_l is the population size of each city. For cities with seroprevalence data, the prior distributions for p_{Zl} were

$$p_{Zl} \sim \text{Beta}(a_{Zl}, b_{Zl})$$

where the method of moments was used to determine a_{Zl} and b_{Zl}. Means and variances corresponding to a uniform random variable with the possible range of infection attack rates were selected using the optim function in R. The range of infection attack rates was defined by the 95% CI of the post-epidemic seroprevalence (Table 4.2). In this way, the seroprevalence data informed the prior distributions while still permitting values outside the observed ranges. The following prior distribution were used (rounded to the nearest tenth),

$p_{Z,Cúcuta} \sim \text{Beta}(294.2, 320.0)$ $\quad$ $p_{Z,Neiva} \sim \text{Beta}(337.6,\ 246.5)$

$p_{Z,Medellín} \sim \text{Beta}(36.9, 513.2)$ $\quad$ $p_{Z,Sincelejo} \sim \text{Beta}(393.5,\ 203.6)$

For cities without serological data and the overall probability of ZIKV infection (p_Z), a_{Zl} and b_{Zl} were both set to 1 which is equivalent to a uniform distribution between 0 and 1. In other words, the proportion of the population infected by ZIKV was considered unknown and was allowed to vary between 0 and 100%. The model was also run using a Beta(2,2)

prior distribution for p_{Zl} in cities with no seroprevalence data. This prior distribution lightly constrains the estimates as 50% of the expected values are between 0.33 and 0.67.

The risks of developing symptoms and neurological complications following ZIKV infection were assumed to be similar across cities; however, reporting rates were expected to differ. The model accounted for these differences by assigning hyperprior distributions to p_{Sl} and p_{CZl}:

$$p_{Sl} \sim \text{Unif}(p_{S\,min}, p_{S\,max}), \qquad p_{CZl} \sim \text{Unif}(p_{CZ\,min}, p_{CZ\,max}),$$

$$p_{S\,min} \sim \text{Unif}(0,1), \qquad p_{CZ\,min} \sim \text{Unif}(0,1),$$

$$p_{S\,max} \sim \text{Unif}(p_{S\,min}, 1), \qquad p_{CZ\,max} \sim \text{Unif}(p_{CZ\,min}, 1),$$

The bounds of the hyperprior distributions ($p_{S\,min}$, $p_{S\,max}$, $p_{CZ\,min}$, and $p_{CZ\,max}$) are independent of city location and were estimated by the framework. They were also used to estimate the overall risk of being reported as a ZVD case (p_S) and a ZVD case with neurological complications (p_{CZ}), respectively.

$$S \sim \text{Bin}(p_S * p_Z, N)$$

$$C \sim \text{Bin}(p_{CZ} * p_Z, N)$$

where C is the total number of ZIKV-associated neurological complications that were reported during the epidemic. Finally, the following equation estimated the total number of neurological complications due to ZIKV infection in each city:

$$C_l \sim \text{Bin}(p_{CZl} * p_{Zl}, N_l)$$

There were 91 total parameters in the model (87 excluding the hyperprior distributions).

After performing the analysis on all data from the 28 cities, the data were classified by sex and age group (data can be found in Appendix S3). Age was dichotomized into 0 to 39 years and $\geq$ 40 years (the median age of patients with ZIKV- associated neurological complications was 41 years). Although the seroprevalence estimates for Cúcuta, Neiva, Medellín, and Sincelejo were sex-specific, the study by Nouvellet et al. found no differences between males and females (personal communication). Consequently, the model here was fitted to both sexes simultaneously assuming the same infection attack rate for each sex. The model was fitted by age group in the same way. There are 153 parameters in each of these models.

The posterior probability that each parameter estimate for males was greater than that for females (and vice versa) was estimated for each city. Similarly, the posterior probabilities were estimated for each age group. After removing the burn-in and merging all four MCMC chains, the proportion of times that the estimated parameter was higher for one sex (or age group) than the other (and vice versa) was determined.

As the seroprevalence study only included participants between ages 2-45 years, a sensitivity analysis was performed by re-fitting the model separately for each age group. For the younger age group, the prior distributions for p_{Zl} in Cúcuta, Medellín, Neiva, and Sincelejo came from the post-epidemic seroprevalence as described above, and Beta(1,1) prior distributions were used for the remaining cities. For the older age group, Beta(1,1) prior distributions for p_{Zl} were used for all cities.

3.2 Expected number of excess neurological complications reported per 10,000 reported cases of ZVD

Using the posterior samples of p_{CZ} and p_S, the number of ZIKV-associated neurological complications reported per 10,000 reported ZVD cases was estimated for each city and overall. The four MCMC chains for each parameter were merged after removing the burn-in. Then, the merged samples of p_{CZl} were divided by those of p_{Sl} for each city, and the result was multiplied by 10,000. For the overall estimate, the samples of each parameter were pooled across all cities before the division step.

3.3 Model estimation and computing

Parameters were estimated in Stan using the R package rstan (version 2.21.2) [237]. Stan is based on the No-U-Turn Sampler which is a type of Hamiltonian MCMC. Hamiltonian MCMC is capable of exploring the posterior distribution of parameters more efficiently compared to other MCMC algorithms [84]. Four Markov chains were run from different starting points for 2,000 iterations with a burn-in of 1,000. Maximum treedepth was increased from 10 to 15 to improve efficiency of the sampler. Convergence of the chains was checked visually. R-hat values and effective sample sizes were also checked for each parameter. R-hat compares the between- and within-chain estimates for model parameters, and values around 1 indicate that chains have mixed well [238]. It is related to the Gelman-Rubin statistic, which was calculated in the previous chapter. The effective sample size estimates

the number of independent samples after accounting for dependence in the MCMC chains [84]. Higher autocorrelation in the chains leads to lower effective sample sizes. Larger effective sample sizes are preferred, and when running four chains, a total effective sample size of at least 400 is recommended [238]. Stan code can be found in Appendix S3. All analyses were conducted in R version 4.0.3.

3.4 Sensitivity analysis

Following [133], the sensitivity of model results to data from different cities was explored. Parameters were re-estimated after removing data one at a time from the cities with available seroprevalence data (Cúcuta, Medellín, Neiva, and Sincelejo) as well as Barranquilla, which was subjected to more intensive surveillance of ZIKV-associated neurological complications compared to other cities [144]. Parameters were also re-estimated after removing all four cities with seroprevalence data from the model.

4 Results

4.1 Bayesian hierarchical model

Eleven out of 28 capital cities reported zero cases of ZIKV-associated neurological complications: Armenia, Inírida, Leticia, Mitú, Pereira, Popayán, Puerto Carreño, Quibdó, Riohacha, San Andrés, and San José del Guaviare. Barranquilla reported the highest number of cases with 80. As expected, cities that reported more ZVD cases also tended to report more cases of ZIKV-associated neurological complications (Figure 4.1). Pearson's correlation coefficient for the relationship is 0.51 (95% CI: 0.17-0.74, p = 0.006).

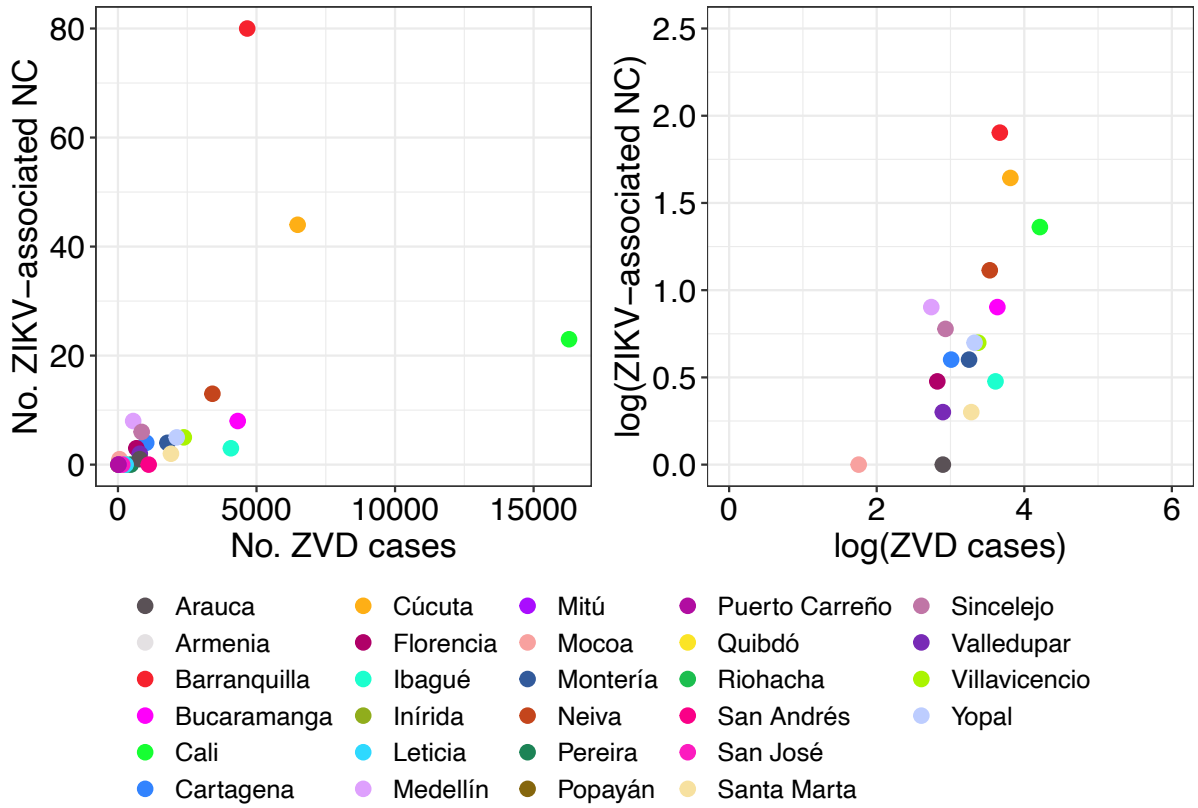


Figure 4.1 Number of reported cases of ZIKV-associated neurological complications (NC) and ZVD on a linear and a log-log scale for 28 capital cities. There is a positive statistically significant correlation of 0.51 (Pearson's correlation coefficient, 95% CI: 0.17-0.74, p = 0.006) on the linear scale and 0.70 (95% CI: 0.32-0.88, p = 0.002) on the $\log_{10}$ scale.

The ZIKV infection attack rate, reporting rate of ZVD, and reporting rate of ZIKV-associated neurological complications were estimated for each city and overall (Figure 4.2). There was substantial heterogeneity in estimates across cities, especially for infection attack rates. The overall estimate for the ZIKV infection attack rate had mean 0.38 (95% CrI: 0.17-0.92). Both the reporting rate of ZVD and the risk of reporting a case of ZIKV-associated neurological complications per 10,000 ZIKV infections were low (mean 0.013, 95% CrI: 0.004-0.024 and mean 0.51, 95% CrI: 0.17-0.92 respectively). Overall, 54 (95% CrI: 5-210) cases of ZIKV-associated neurological complications were expected to be reported for every 10,000 reported cases of ZVD on average. Figure 4.3 shows the mean and credible intervals of this estimate for all cities as well as the point estimates and 95% binomial confidence intervals calculated from the raw data. The point estimate falls outside of the 95% credible interval for 12 cities, 11 of which reported zero cases of ZIKV-associated neurological complications.

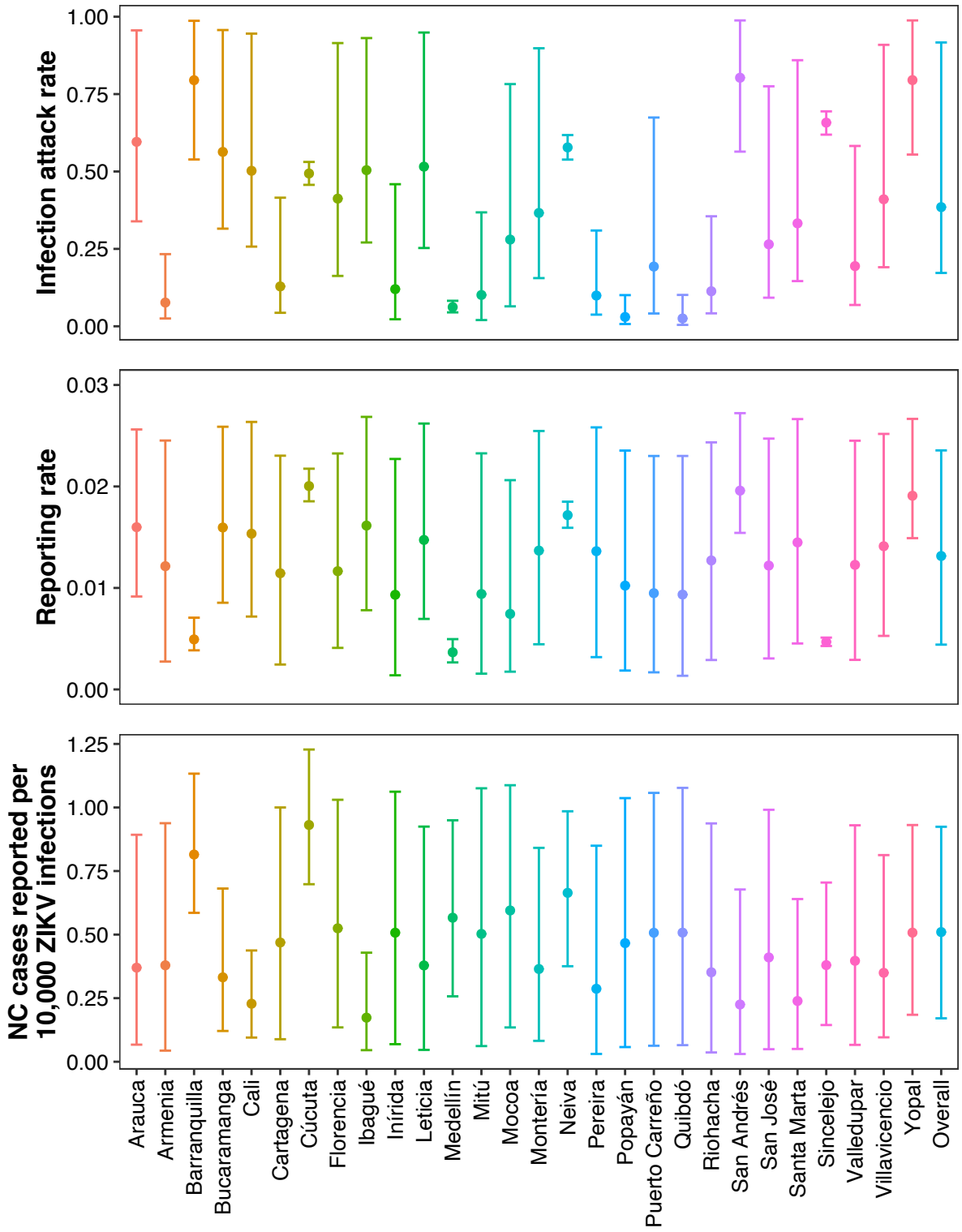


Figure 4.2 Estimated ZIKV infection attack rates, ZVD reporting rates, and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections. Seroprevalence data were incorporated into the analysis for Cúcuta, Neiva, Medellín, and Sincelejo. Posterior mean (points) and 95% credible interval (error bars) are shown for each city and overall.

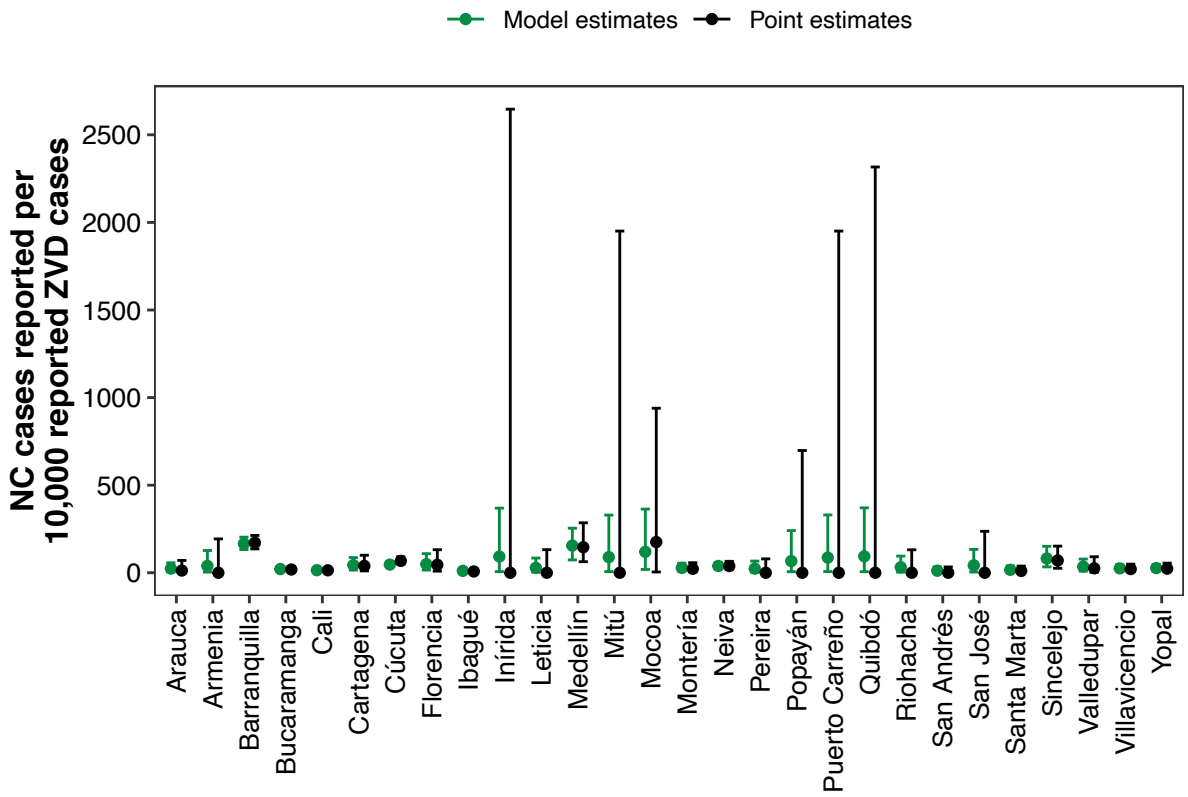


Figure 4.3 Reported cases of ZIKV-associated neurological complications (NC) per 10,000 reported cases of ZVD. Posterior mean (green points) and 95% credible interval (green error bars) are shown for each city. Black points correspond to the point estimates calculated from the raw data, while the black error bars correspond to the 95% binomial confidence intervals.

Across cities, mean estimates for the infection attack rate ranged from 0.03 in Quibdó and Popayán to 0.82 in San Andrés (Table 4.3). Mean estimates for the reporting rate ranged from 0.004 in Medellín to 0.020 in Cúcuta. The mean estimated number of ZIKV-associated neurological complications reported per 10,000 ZIKV infections was lowest in Ibagué (0.17) and highest in Cúcuta (0.93). Estimates of the hyperprior distributions for p_S and p_{CZ} are shown in Table 4.4.

Removing data from Barranquilla and each of the four cities with seroprevalence data did not have any major impacts on parameter estimates (Figure 4.4). Also, the model has power to estimate the ZIKV infection attack rates even without serological data. When all the cities with serological data are removed from the model (Appendix S3), the point estimates for ZIKV infection attack rates for the remaining cities are slightly higher, and the estimates for the reporting rates of ZVD are slightly lower. However, the credible intervals are similar for both parameters across cities. The estimates for the number of neurological complications

reported per 10,000 ZIKV infections are also lower without the serological data, and the upper limits of credible intervals are lower.

Parameter estimates from the model fitted with Beta(2,2) prior distributions for the ZIKV infection attack rates in cities without seroprevalence data can be found in Appendix S3. Compared to the model that used Beta(1,1) prior distributions, the estimated ZIKV infection attack rates are shifted slightly toward 0.5, and the overall estimate across all cities is similar.

Table 4.3 Estimated ZIKV infection attack rates, ZVD reporting rates, and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by city. Mean posterior and 95% CrI are presented.

City	Estimated ZIKV infection attack rate	Estimated ZVD reporting rate	Estimated ZIKV-associated NC cases reported per 10,000 ZIKV infections
Arauca	0.60 (0.34-0.96)	0.016 (0.009-0.026)	0.37 (0.07-0.89)
Armenia	0.08 (0.03-0.23)	0.012 (0.003-0.025)	0.38 (0.04-0.94)
Barranquilla	0.79 (0.54-0.99)	0.005 (0.004-0.007)	0.81 (0.59-1.13)
Bucaramanga	0.56 (0.32-0.96)	0.016 (0.009-0.026)	0.33 (0.12-0.68)
Cali	0.50 (0.26-0.95)	0.015 (0.007-0.026)	0.23 (0.10-0.44)
Cartagena	0.13 (0.04-0.42)	0.011 (0.002-0.023)	0.47 (0.09-1.00)
Cúcuta	0.49 (0.46-0.53)	0.020 (0.019-0.022)	0.93 (0.70-1.23)
Florencia	0.41 (0.16-0.91)	0.012 (0.004-0.023)	0.52 (0.14-1.03)
Ibagué	0.50 (0.27-0.93)	0.016 (0.008-0.027)	0.17 (0.05-0.43)
Inírida	0.12 (0.02-0.46)	0.009 (0.001-0.023)	0.51 (0.07-1.06)
Leticia	0.52 (0.25-0.95)	0.015 (0.007-0.026)	0.38 (0.05-0.92)
Medellín	0.06 (0.04-0.08)	0.004 (0.003-0.005)	0.57 (0.26-0.95)
Mitú	0.10 (0.02-0.37)	0.009 (0.002-0.023)	0.50 (0.06-1.08)
Mocoa	0.28 (0.06-0.78)	0.007 (0.002-0.021)	0.59 (0.13-1.09)
Montería	0.37 (0.16-0.90)	0.014 (0.004-0.025)	0.36 (0.08-0.84)
Neiva	0.58 (0.54-0.62)	0.017 (0.016-0.018)	0.66 (0.38-0.99)
Pereira	0.10 (0.04-0.31)	0.014 (0.003-0.026)	0.29 (0.03-0.85)
Popayán	0.03 (0.01-0.10)	0.010 (0.002-0.024)	0.47 (0.06-1.04)
Puerto Carreño	0.19 (0.04-0.67)	0.009 (0.002-0.023)	0.51 (0.06-1.06)
Quibdó	0.03 (0.00-0.10)	0.009 (0.001-0.023)	0.51 (0.07-1.08)
Riohacha	0.11 (0.04-0.36)	0.013 (0.003-0.024)	0.35 (0.04-0.94)
San Andrés	0.80 (0.56-0.99)	0.020 (0.015-0.027)	0.23 (0.03-0.68)
San José del Guaviare	0.26 (0.09-0.78)	0.012 (0.003-0.025)	0.41 (0.05-0.99)
Santa Marta	0.33 (0.15-0.86)	0.014 (0.005-0.027)	0.24 (0.05-0.64)
Sincelejo	0.66 (0.62-0.69)	0.005 (0.004-0.005)	0.38 (0.14-0.70)
Valledupar	0.19 (0.07-0.58)	0.012 (0.003-0.025)	0.40 (0.07-0.93)
Villavicencio	0.41 (0.19-0.91)	0.014 (0.005-0.025)	0.35 (0.10-0.81)
Yopal	0.80 (0.55-0.99)	0.019 (0.015-0.027)	0.51 (0.18-0.93)

Table 4.4 Estimated hyperprior distributions (per 10,000). Mean posterior and 95% CrI are presented.

Parameter	Estimates
$p_{S\,min}$	0.0022 (0.0002-0.0040)
$p_{S\,max}$	0.024 (0.020-0.032)
$p_{CZ\,min}$	0.058 (0.002-0.171)
$p_{CZ\,max}$	1.00 (0.75-1.34)

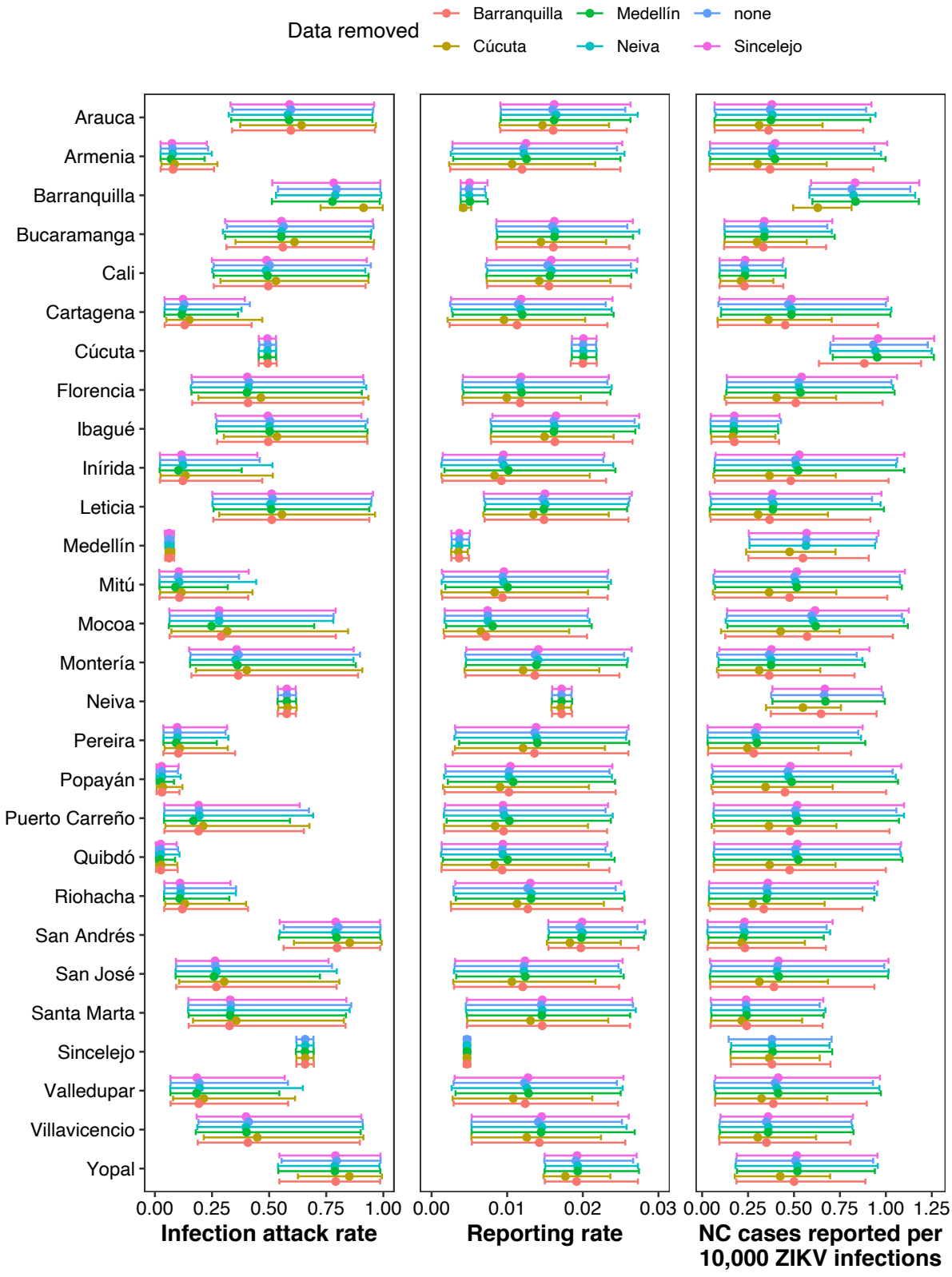


Figure 4.4 Effect of removing data on estimated ZIKV infection attack rates, ZVD reporting rates, and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections. Posterior mean (points) and 95% credible interval (error bars) are shown for each city and overall.

4.2 Bayesian hierarchical model by sex

Overall parameter estimates for the ZVD reporting rate and the number of ZIKV-associated neurological complications cases reported per 10,000 ZIKV infections by sex can be found in Table 4.5. When the same ZIKV infection attack rate is assumed for males and females, females were more likely to be reported as ZVD cases compared to males but less likely to be reported as cases with neurological complications. The estimates for all cities are shown in Figure 4.5. A comparison of the posterior probabilities by sex for each city is shown in Table 4.7 as well as in Figures 4.6-4.7. Important differences in the posterior probabilities for ZVD reporting rates were identified in nearly all cities, but there were no major differences in the risk of ZIKV-associated neurological complications. The estimated hyperprior distributions and the overall ZIKV infection attack rate are shown in Table 4.6.

Table 4.5 Overall estimated ZVD reporting rate and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by sex. Mean posterior and 95% CrI are presented. The posterior probabilities (PP) are shown in the final two columns.

Parameter	Males	Females	PP males > females	PP males < females
Estimated ZVD reporting rate	0.011 (0.005-0.016)	0.020 (0.009-0.029)	<0.0001	>0.9999
Estimated number of ZIKV-associated NC cases reported per 10,000 ZIKV infections	0.69 (0.30-1.01)	0.53 (0.23-0.81)	0.97	0.03

Table 4.6 Estimated hyperprior distributions (shown per 10,000) by sex and overall ZIKV infection attack rate for both males and females. Mean posterior and 95% CrI are presented.

Parameter	Males	Females
Estimated ZIKV infection attack rate	0.28 (0.18-0.59)	
$p_{S\,min}$	0.002 (0.001-0.003)	0.004 (0.001-0.006)
$p_{S\,max}$	0.016 (0.013-0.021)	0.030 (0.026-0.038)
$p_{CZ\,min}$	0.15 (0.01-0.37)	0.062 (0.002-0.177)
$p_{CZ\,max}$	1.00 (0.73-1.35)	0.91 (0.66-1.28)

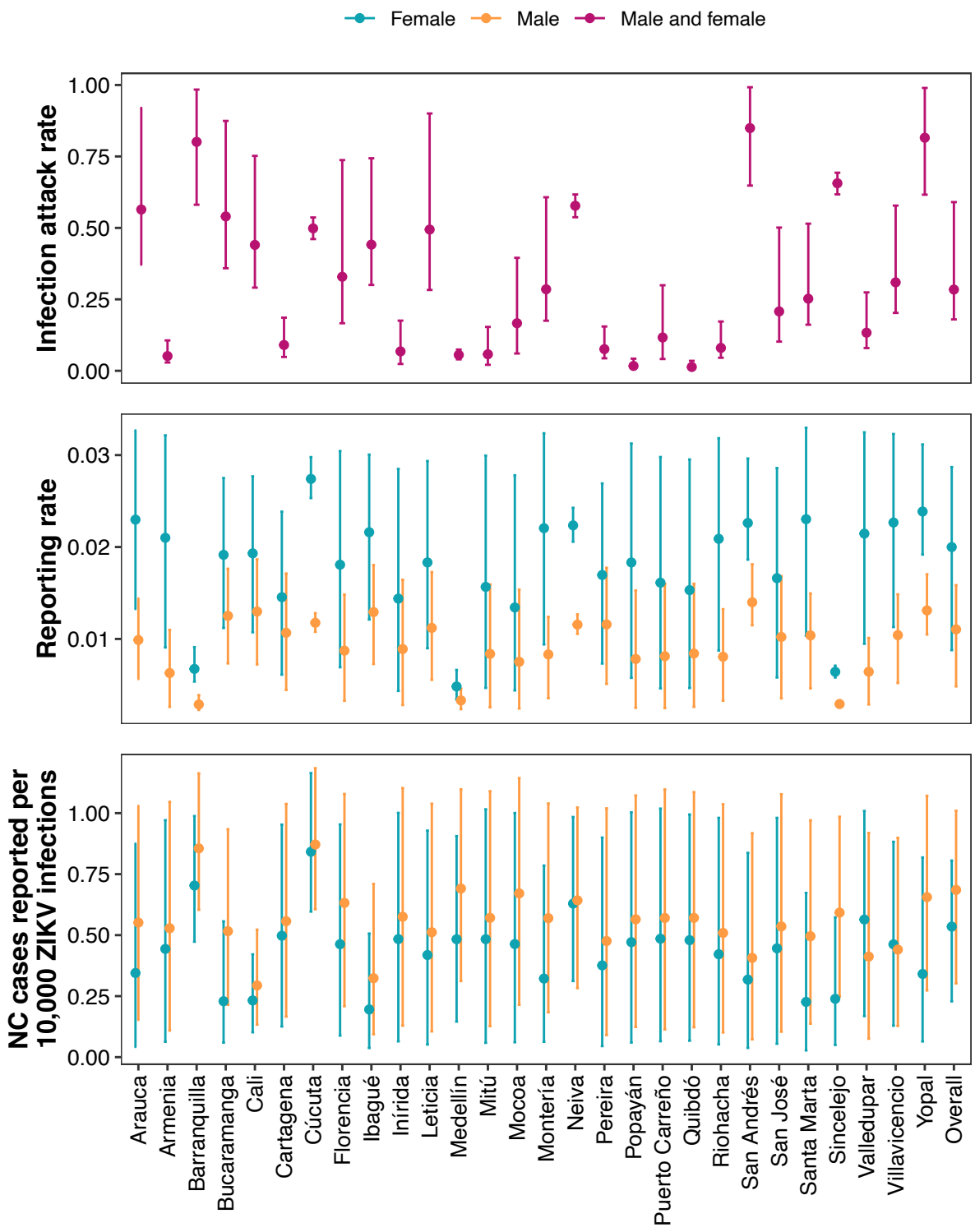


Figure 4.5 Estimated ZVD reporting rate and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by sex. The same ZIKV infection attack rate was assumed for males and females. Seroprevalence data were incorporated into the analysis for Cúcuta, Neiva, Medellín, and Sincelejo. Posterior mean (points) and 95% credible interval (error bars) are shown for each city and overall.

Table 4.7 Comparison of the posterior probabilities (PP) of estimated ZVD reporting rate and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by sex for each city.

City	Estimated ZVD reporting rate		Estimated ZIKV-associated NC cases reported per 10,000 ZIKV infections	
	PP males > females	PP males < females	PP males > females	PP males < females
Arauca	<0.0001	>0.9999	0.74	0.26
Armenia	<0.0001	>0.9999	0.60	0.40
Barranquilla	<0.0001	>0.9999	0.84	0.16
Bucaramanga	<0.0001	>0.9999	0.93	0.07
Cali	<0.0001	>0.9999	0.74	0.26
Cartagena	<0.0001	>0.9999	0.57	0.43
Cúcuta	<0.0001	>0.9999	0.57	0.43
Florencia	<0.0001	>0.9999	0.69	0.31
Ibagué	<0.0001	>0.9999	0.77	0.23
Inírida	0.17	0.83	0.59	0.41
Leticia	<0.0001	>0.9999	0.60	0.40
Medellín	<0.0001	>0.9999	0.77	0.23
Mitú	0.07	0.93	0.59	0.41
Mocoa	0.01	0.99	0.71	0.29
Montería	<0.0001	>0.9999	0.81	0.19
Neiva	<0.0001	>0.9999	0.52	0.48
Pereira	<0.0001	>0.9999	0.63	0.37
Popayán	0.001	0.999	0.60	0.40
Puerto Carreño	0.05	0.95	0.58	0.42
Quibdó	0.10	0.90	0.60	0.40
Riohacha	<0.0001	>0.9999	0.60	0.40
San Andrés	<0.0001	>0.9999	0.63	0.37
San José del Guaviare	0.0005	0.9995	0.60	0.40
Santa Marta	<0.0001	>0.9999	0.85	0.15
Sincelejo	<0.0001	>0.9999	0.93	0.07
Valledupar	<0.0001	>0.9999	0.30	0.70
Villavicencio	<0.0001	>0.9999	0.46	0.54
Yopal	<0.0001	>0.9999	0.86	0.14

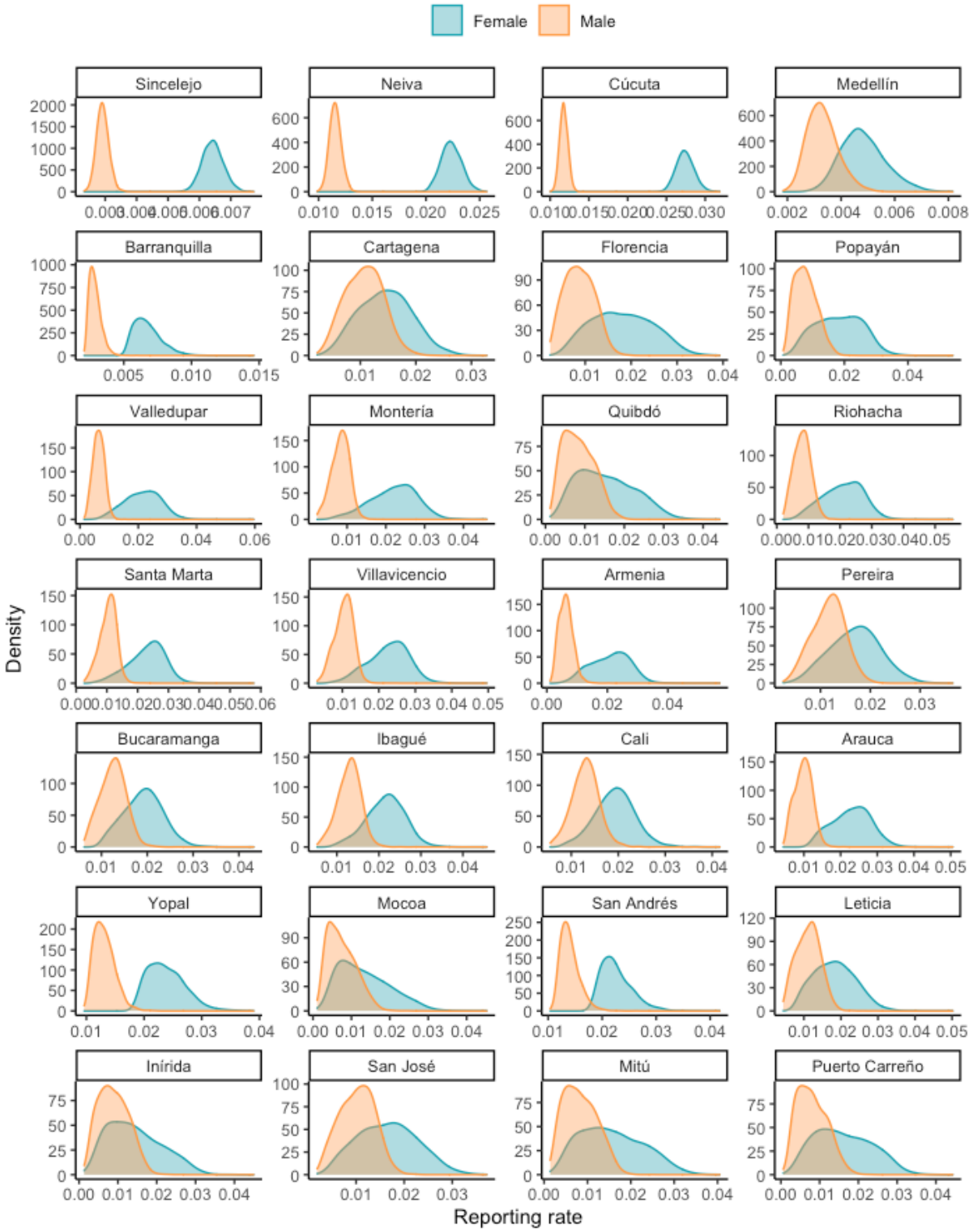


Figure 4.6 Comparison of the posterior densities of estimated ZVD reporting rate by sex for each city.

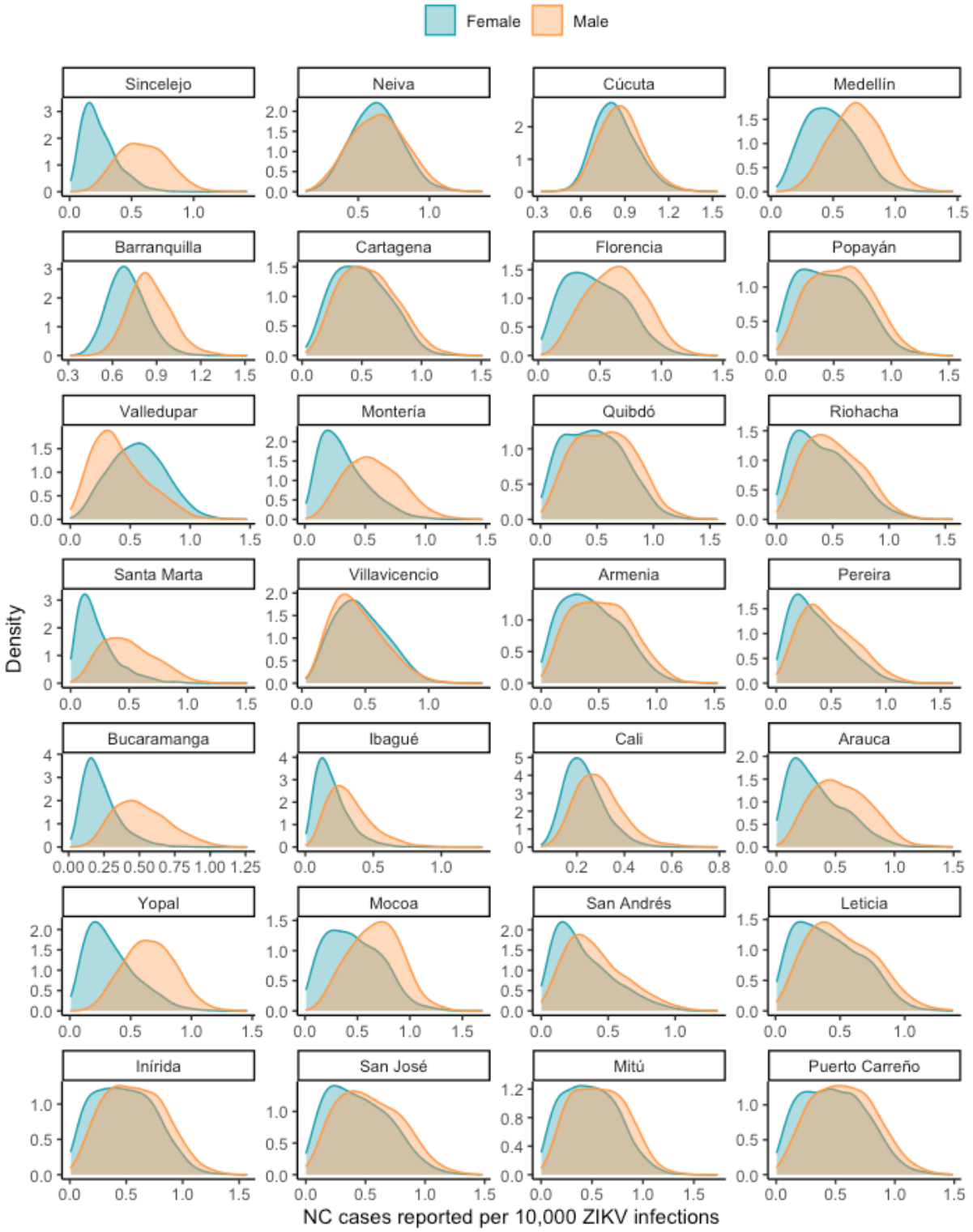


Figure 4.7 Comparison of the posterior densities of estimated number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by sex for each city.

4.3 Bayesian hierarchical model by age group

Overall parameter estimates for the ZVD reporting rate and the number of ZIKV-associated neurological complications cases reported per 10,000 ZIKV infections by age group can be found in Table 4.8. When the same ZIKV infection attack rate is assumed for all ages, younger individuals were more likely to be reported as ZVD cases compared to older individuals but less likely to be reported as cases with neurological complications. The estimates for all cities are shown in Figure 4.8. A comparison of the posterior probabilities by age group for each city is shown in Table 4.10 as well as in Figures 4.9-4.10. Important differences in the posterior probabilities for ZVD reporting rates were identified in several cities, but there were only a couple of cities with major differences in the risk of ZIKV-associated neurological complications. The estimated hyperprior distributions and the overall ZIKV infection attack rate are shown in Table 4.9.

Table 4.8 Overall estimated ZVD reporting rate and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by age group. Mean posterior and 95% CrI are presented. The posterior probabilities (PP) are shown in the final two columns.

Parameter	0-39 years	40 or more years	PP 0-39 > 40 or more	PP 0-39 < 40 or more
Estimated ZVD reporting rate	0.018 (0.008-0.026)	0.012 (0.005-0.017)	>0.9999	<0.0001
Estimated number of ZIKV-associated NC cases reported per 10,000 ZIKV infections	0.44 (0.19-0.66)	0.93 (0.40-1.39)	<0.0001	>0.9999

Table 4.9 Estimated hyperprior distributions (shown per 10,000) by age group and ZIKV infection attack rate for all ages. Mean posterior and 95% CrI are presented.

Parameter	0-39 years	40 or more years
Estimated ZIKV infection attack rate	0.28 (0.18-0.59)	
$p_{S\,min}$	0.004 (0.001-0.005)	0.0017 (0.0004-0.0028)
$p_{S\,max}$	0.025 (0.022-0.032)	0.018 (0.015-0.024)
$p_{CZ\,min}$	0.065 (0.002-0.189)	0.17 (0.01-0.41)
$p_{CZ\,max}$	0.69 (0.50-0.95)	1.50 (1.08-2.10)

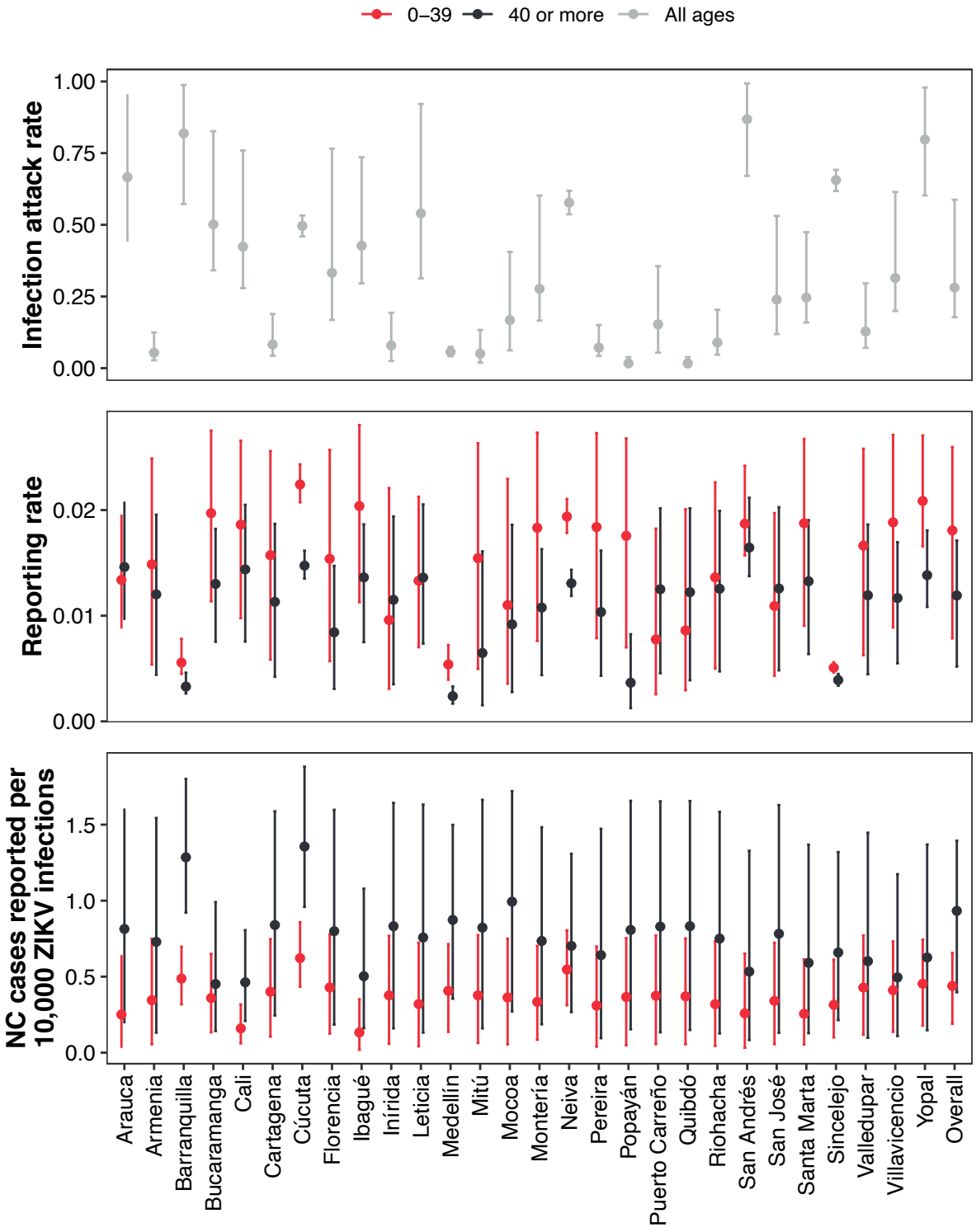


Figure 4.8 Estimated ZVD reporting rate and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by age group. The same ZIKV infection attack rate was assumed for all ages. Seroprevalence data were incorporated into the analysis for Cúcuta, Neiva, Medellín, and Sincelejo. Posterior mean (points) and 95% credible interval (error bars) are shown for each city and overall.

Table 4.10 Comparison of the posterior probabilities (PP) of estimated ZVD reporting rate and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by age group for each city.

City	Estimated ZVD reporting rate		Estimated ZIKV-associated NC cases reported per 10,000 ZIKV infections	
	PP 0-39 > 40 or more	PP 0-39 < 40 or more	PP 0-39 > 40 or more	PP 0-39 < 40 or more
Arauca	0.12	0.88	0.08	0.92
Armenia	0.92	0.08	0.20	0.80
Barranquilla	>0.9999	<0.0001	<0.0001	>0.9999
Bucaramanga	>0.9999	<0.0001	0.37	0.63
Cali	>0.9999	<0.0001	0.002	0.998
Cartagena	>0.9999	<0.0001	0.12	0.88
Cúcuta	>0.9999	<0.0001	0.0002	0.9998
Florencia	>0.9999	<0.0001	0.18	0.82
Ibagué	>0.9999	<0.0001	0.03	0.97
Inírida	0.33	0.67	0.17	0.83
Leticia	0.41	0.59	0.18	0.82
Medellín	>0.9999	<0.0001	0.07	0.93
Mitú	0.96	0.04	0.16	0.84
Mocoa	0.74	0.26	0.07	0.93
Montería	>0.9999	<0.0001	0.13	0.87
Neiva	>0.9999	<0.0001	0.33	0.67
Pereira	>0.9999	<0.0001	0.22	0.78
Popayán	>0.9999	<0.0001	0.17	0.83
Puerto Carreño	0.12	0.88	0.18	0.82
Quibdó	0.21	0.79	0.16	0.84
Riohacha	0.70	0.30	0.18	0.82
San Andrés	0.98	0.02	0.23	0.77
San José del Guaviare	0.19	0.81	0.16	0.84
Santa Marta	>0.9999	<0.0001	0.15	0.85
Sincelejo	0.9998	0.0002	0.13	0.87
Valledupar	>0.9999	<0.0001	0.37	0.63
Villavicencio	>0.9999	<0.0001	0.43	0.57
Yopal	>0.9999	<0.0001	0.35	0.65

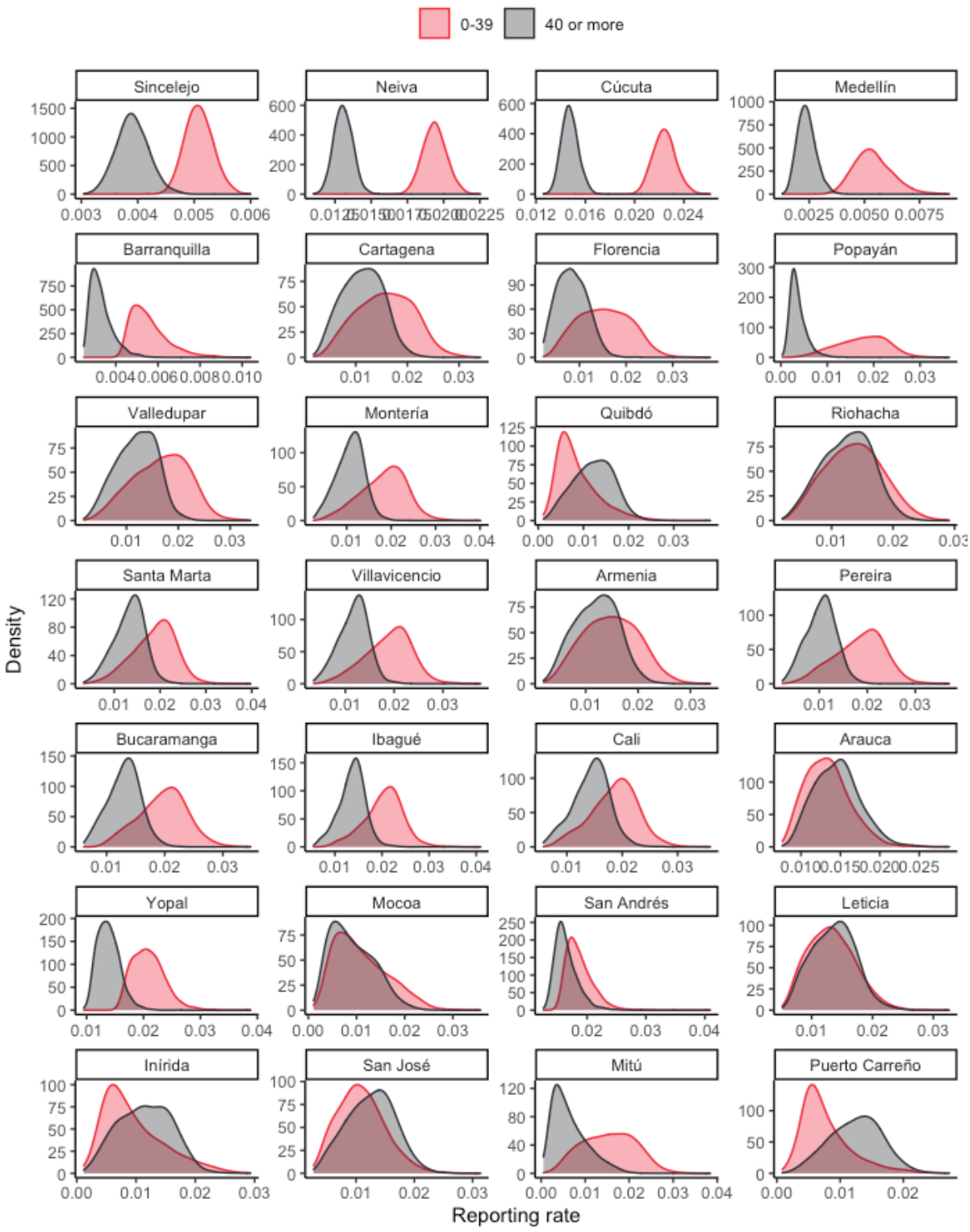


Figure 4.9 Comparison of the posterior densities of estimated ZVD reporting rate by age group for each city.

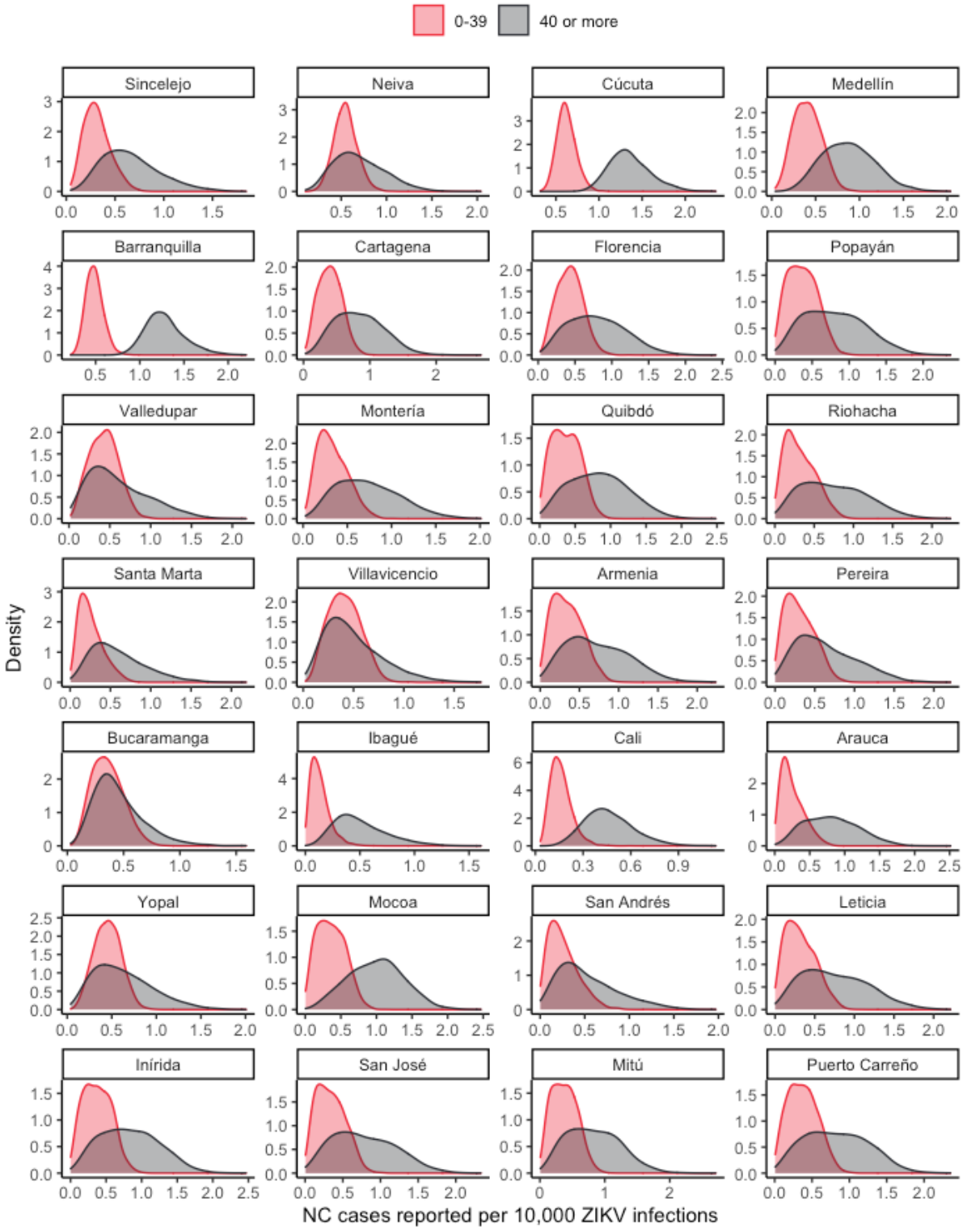


Figure 4.10 Comparison of the posterior densities of estimated number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by age group for each city.

When a Beta(1,1) prior distribution was used for the ZIKV infection attack rate in the older age group in Cúcuta, Medellín, Neiva, and Sincelejo, the uncertainty of that parameter increased substantially for those cities and also affected the reporting rates to some degree (Figure 4.11).

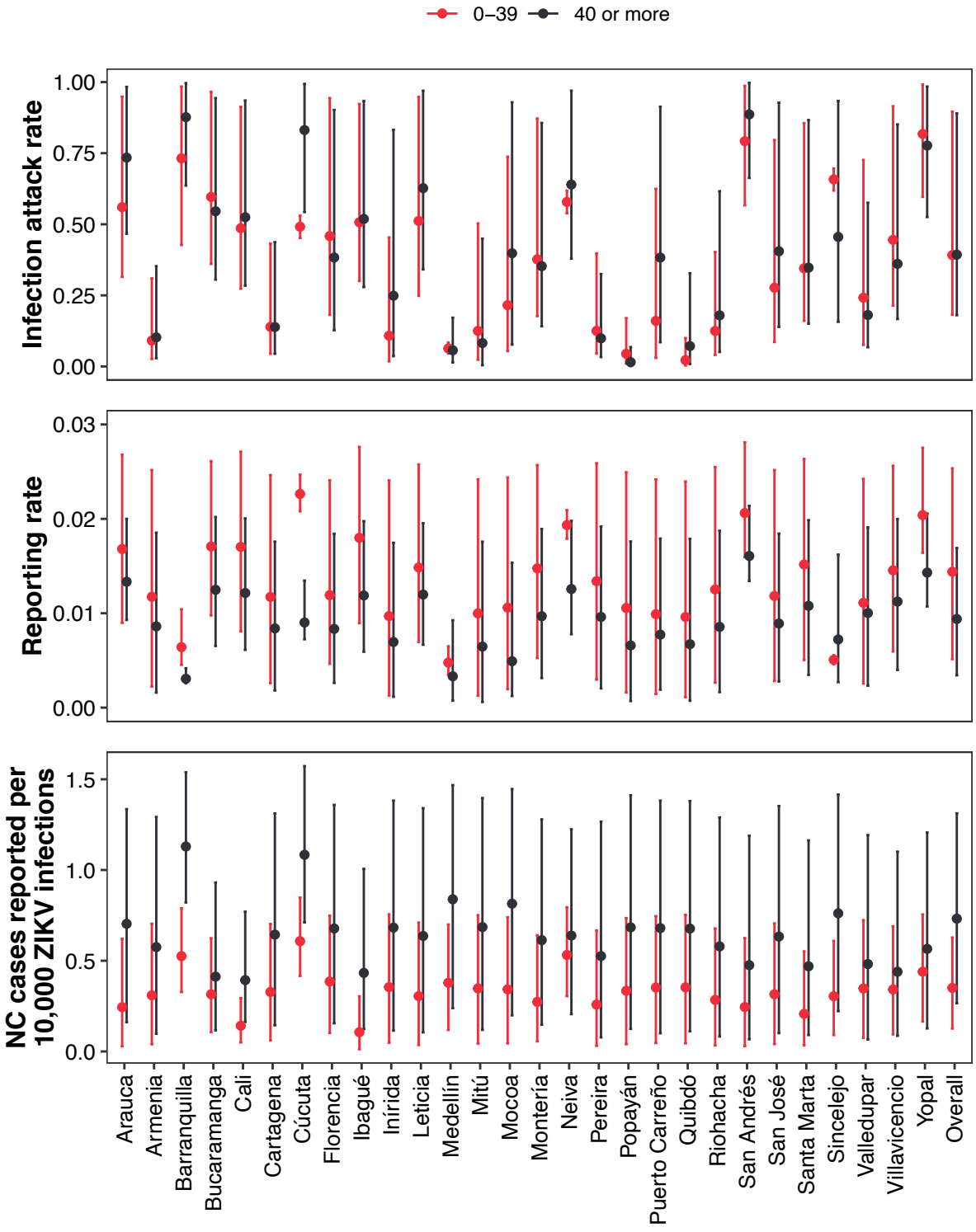


Figure 4.11 Sensitivity of the prior distributions for ZIKV infection attack rates across age groups on estimated ZIKV infection attack rates, ZVD reporting rates, and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections by age group. In Cúcuta, Medellín, Neiva, and Sincelejo, different prior distributions were used for individuals between 0 and 39 years and those 40 years or more. Posterior mean (points) and 95% credible interval (error bars) are shown for each city and overall.

4.4 Model convergence and diagnostics

Model convergence and diagnostics in this section correspond to the first model which includes all the data. Model diagnostics were also checked for the sex and age models. Although some divergent transitions occurred after the warm-up period for the combined age model with 153 parameters, the warning message was eliminated by increasing adapt_delta (the target acceptance rate) from the default of 0.8 to 0.95. No other issues with these models were identified.

4.4.1 Distributions

Figures 4.12-4.14 show violin plots of the posterior distributions of the parameters after removing the burn-in and merging all four chains. Violin plots are a type of density plot; the peak of the plot has the most support from the data and the prior distribution. Although density plots are not a formal way of determining model convergence, unusual shapes can indicate poor convergence [86].

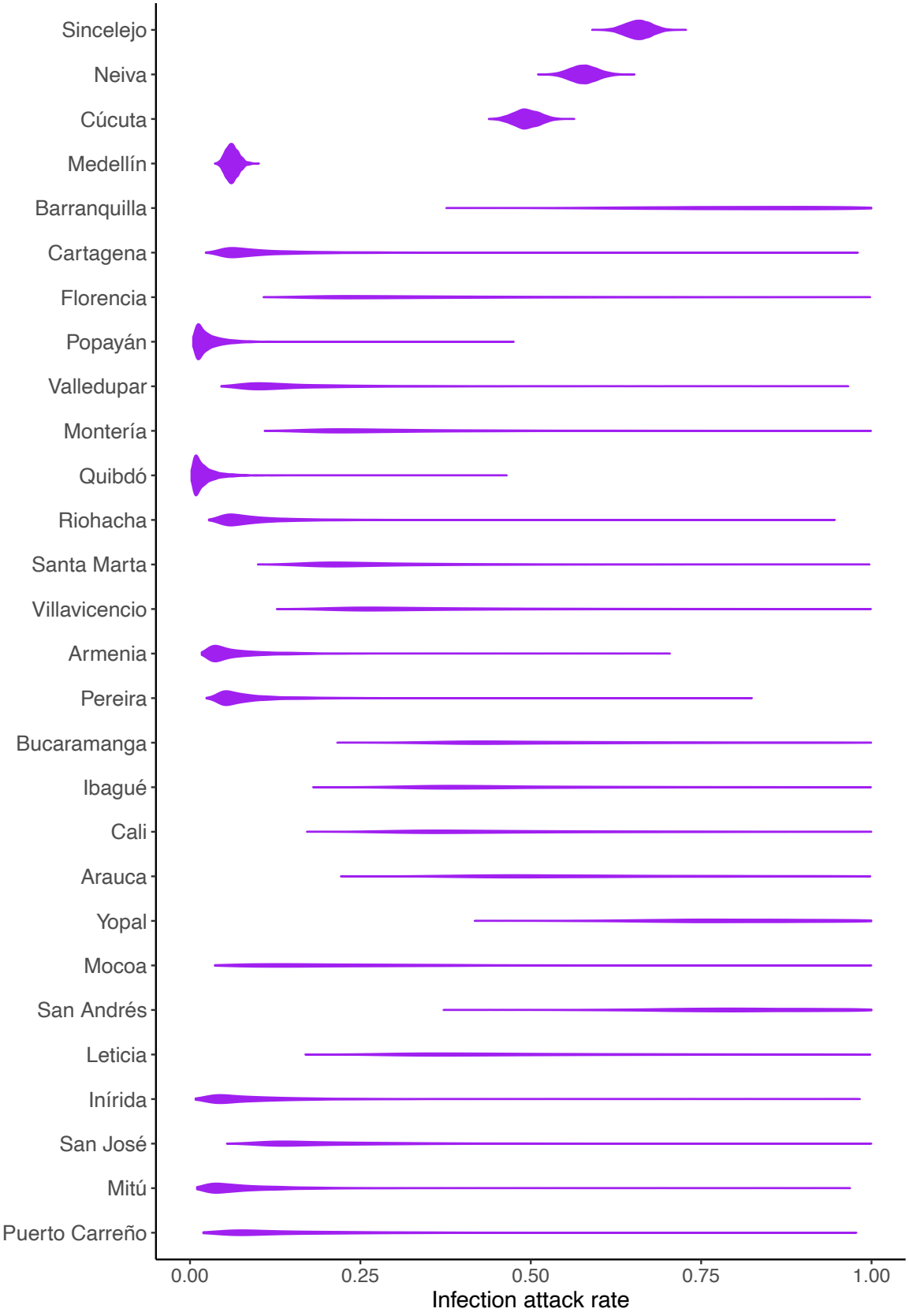


Figure 4.12 Violin plots of the posterior distribution of ZIKV infection attack rate for all cities. All four chains were merged after removing the burn-in.

Figure 4.13 Violin plots of the posterior distribution of ZVD reporting rate for all cities. All four chains were merged after removing the burn-in.

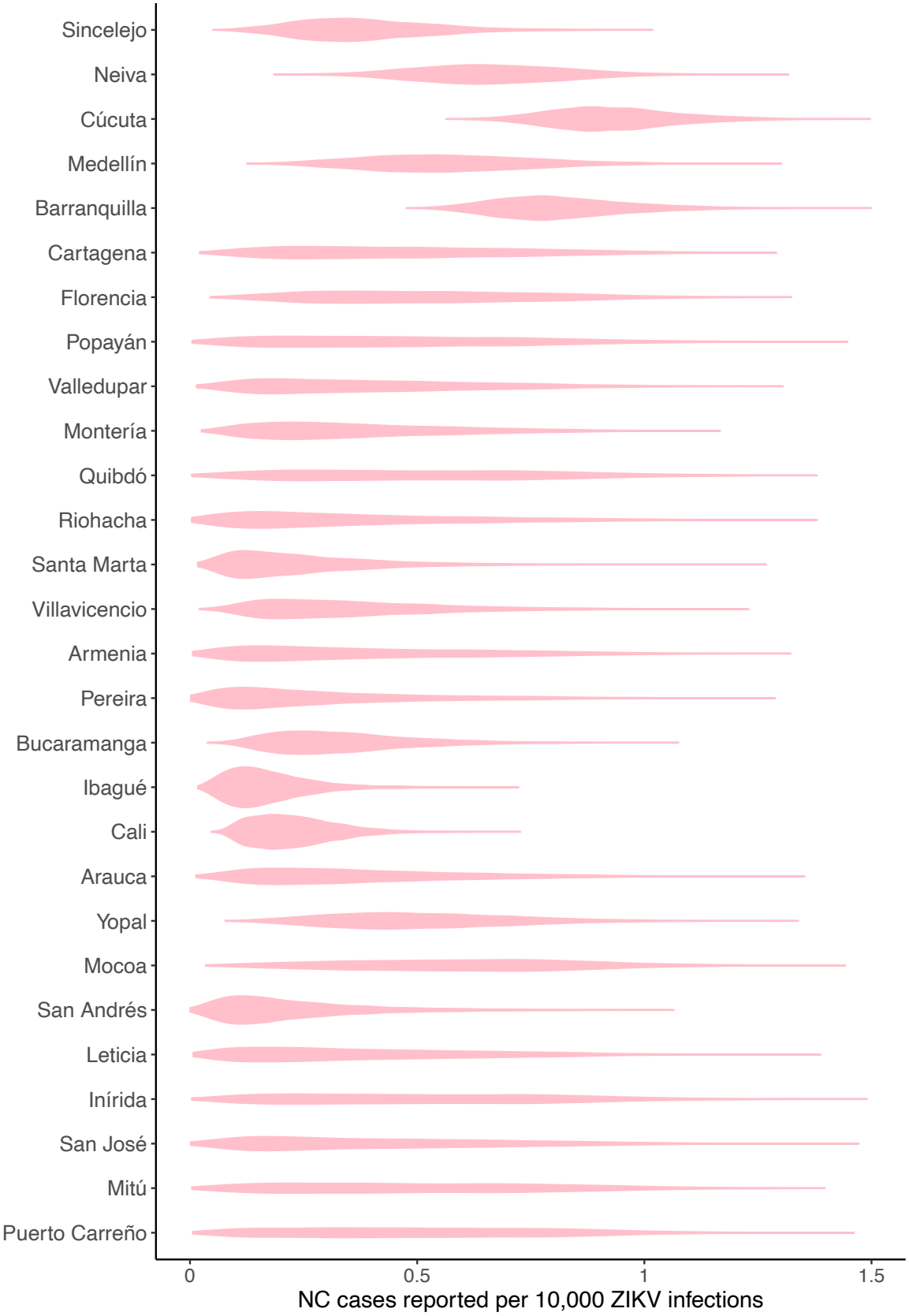


Figure 4.14 Violin plots of the posterior distribution of the number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections for all cities. All four chains were merged after removing the burn-in.

4.4.2 Convergence diagnostics

The R-hat values for all parameters were very close to 1, suggesting model convergence. All parameters also had good effective sample sizes, and no warnings about divergent transitions were produced.

4.4.3 Traces

Figures 4.15-4.17 show the MCMC traces for four chains after removing the burn-in period. Mixing is good for all parameters based on visual assessment.

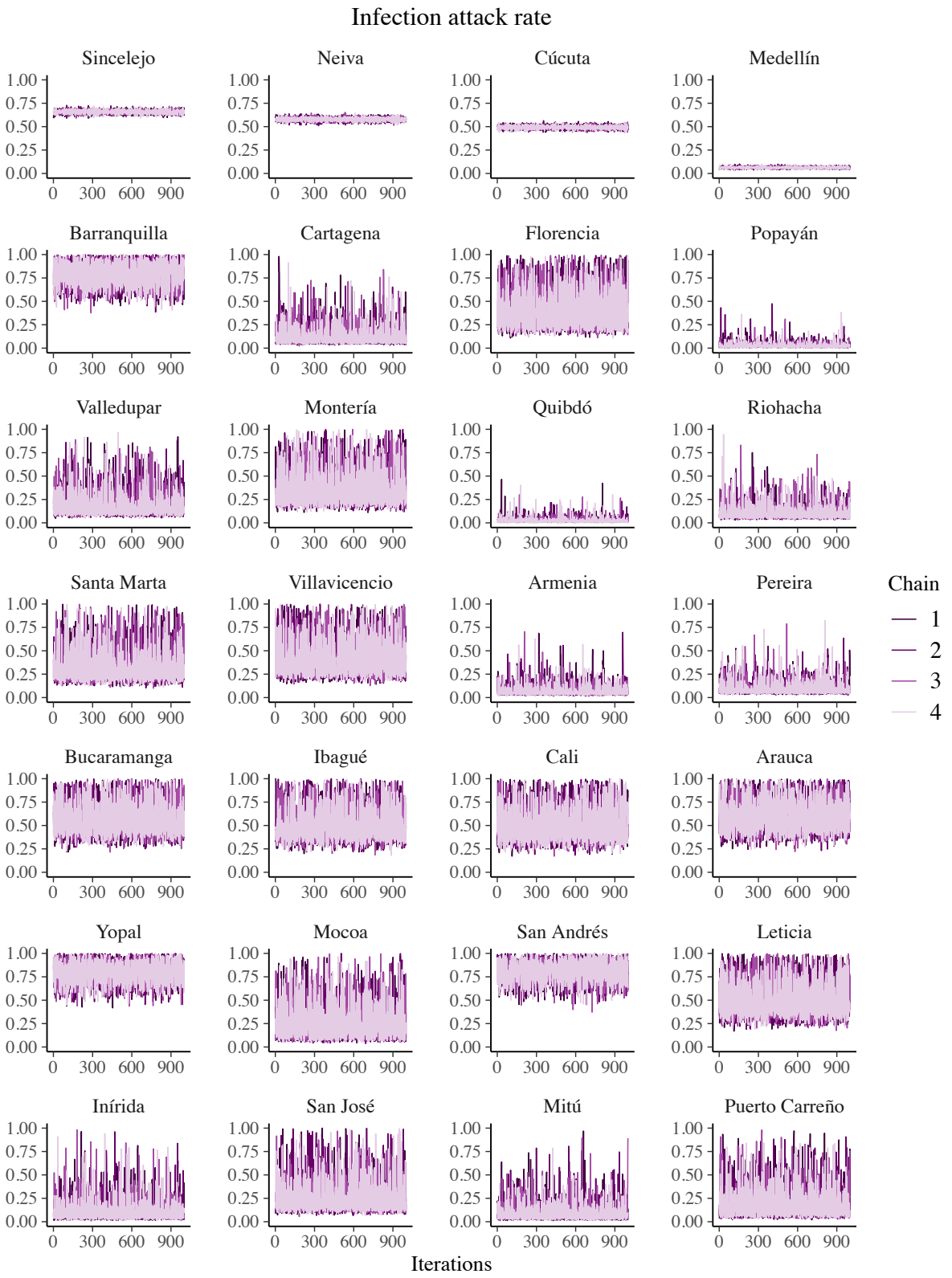


Figure 4.15 MCMC traces for the ZIKV infection attack rate for all cities.

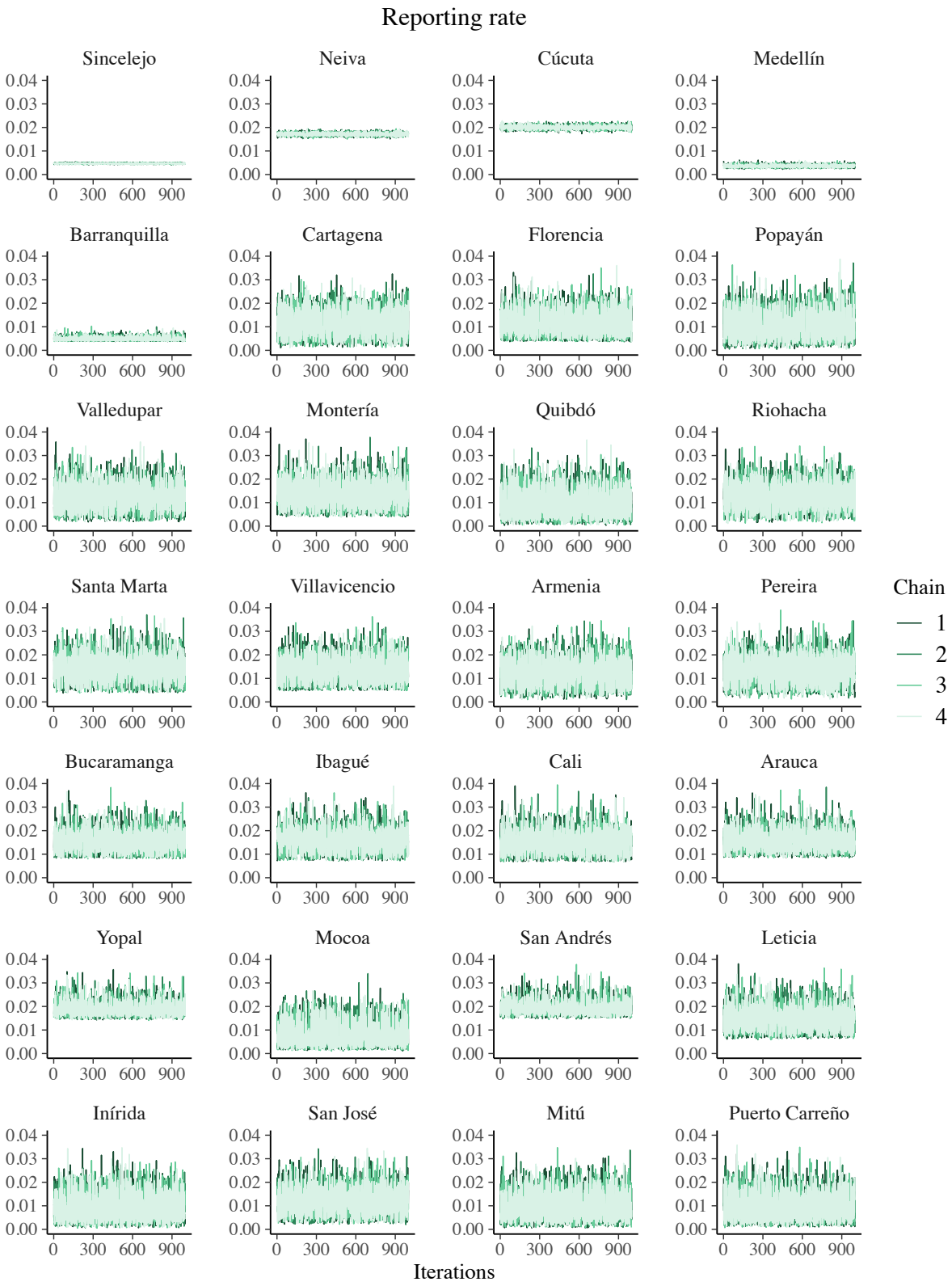


Figure 4.16 MCMC traces for the ZVD reporting rate for all cities.

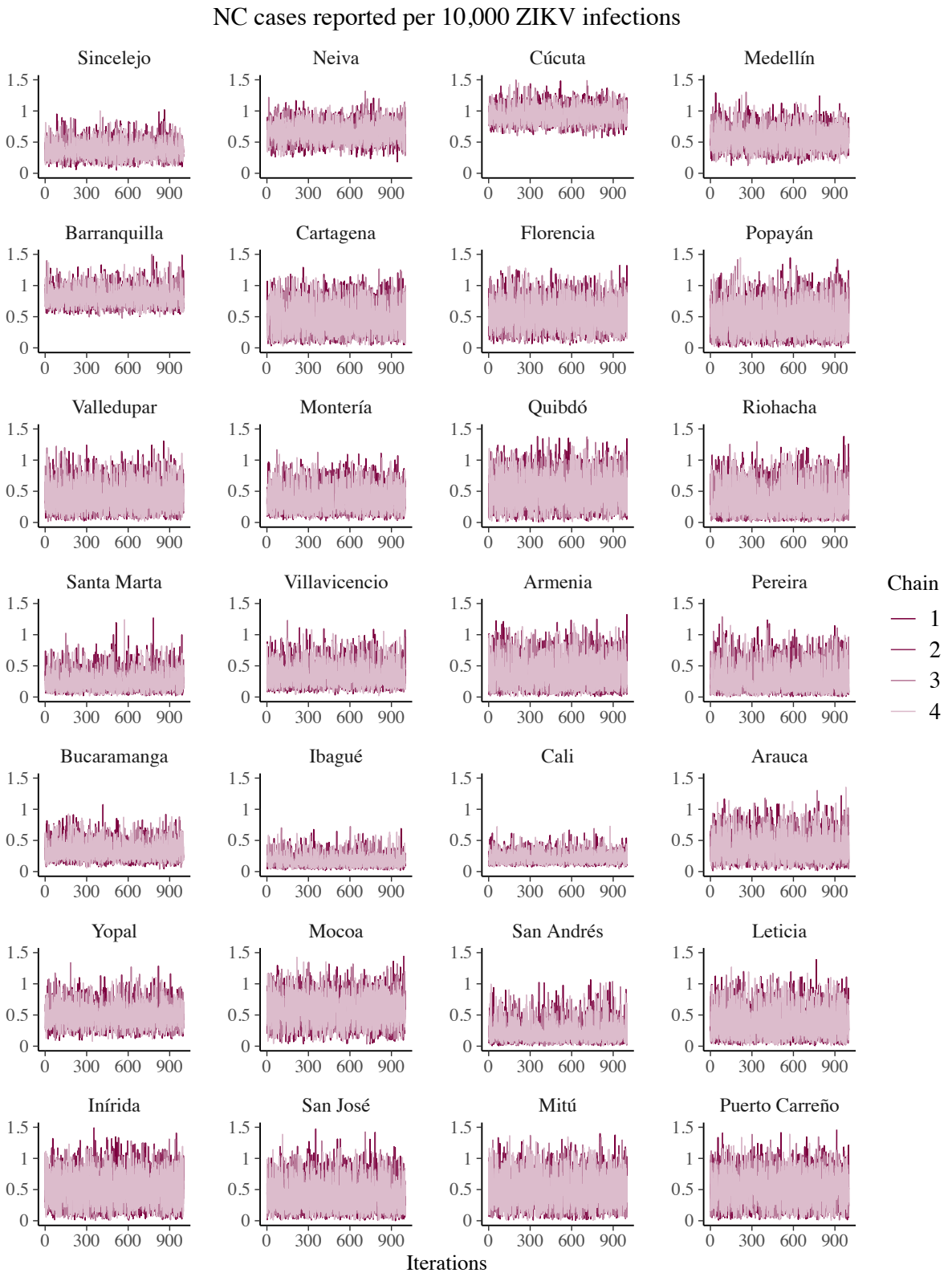


Figure 4.17 MCMC traces for the number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections for all cities.

5 Discussion

In this chapter, estimates of ZIKV infection attack rates, ZVD reporting rates, and risk of ZIKV-associated neurological complications were refined for Colombia using surveillance data stratified by age and sex, seroprevalence data from four cities, and a dataset on ZIKV-associated neurological complications. Estimated ZIKV infection attack rates were heterogeneous across cities, and differences in reporting ZVD cases and ZVD cases with neurological complications were quantified for each sex and two age groups.

The overall estimate of the ZIKV infection attack rate for the 28 capital cities analyzed was 0.38 (95% CrI: 0.17-0.92), which is higher than that reported by both Mier-y-Teran-Romero et al. (0.09, 95% CrI: 0.03-0.23) and Moore et al. (0.19, 95% CrI: 0.15-0.23) [133, 192]. Moore et al. used a Beta(1,2) prior distribution to lightly constrain the infection attack rate; however, their result from using a Beta(1,1) prior distribution was more similar (0.26, 95% CrI: 0.21-0.31) to the one presented here. Differences in the estimates could be attributed to the different spatial scales considered: while this study focused on capital cities, Moore et al. used department-level data and Mier-y-Teran-Romero et al. used national-level data. Moreover, all 28 cities in this study were at risk of arbovirus transmission and reported ZVD cases. In contrast, many locations which were included in the other two studies were not at risk and did not report ZVD cases, including Bogotá, with a population of about eight million. Higher overall infection attack rates would be expected from an analysis that only included at-risk locations.

The credible intervals for the overall estimate of the infection attack rate were wider compared to those reported by both Mier-y-Teran-Romero et al. and Moore et al. and likely reflect heterogeneity in infection attack rates across cities. The post-epidemic seroprevalence estimates, which informed the prior distributions for four cities in this study, ranged from about 0.07 in Medellín to between 0.48-0.66 in Cúcuta, Neiva, and Sincelejo. There is evidence that high infection attack rates resulting in herd immunity brought an end to the ZIKV epidemic in the Americas [23]. Seroprevalence studies conducted in other large cities in Latin America have also reported high infection attack rates such as 0.46 (95% CI: 0.44-0.48) in Managua, Nicaragua [239] and 0.73 (95% CI: 0.70-0.76) in Salvador, Brazil [120]. However, infection attack rates were not uniform even at small spatial scales, leaving

pockets of susceptible populations [120, 232]. This heterogeneity may not have been captured as well by Mier-y-Teran-Romero et al. and Moore et al., who used coarser surveillance data and did not incorporate seroprevalence data from Colombia.

The estimated reporting rate of ZVD across cities was 0.013 (95% CrI: 0.004-0.024). This estimate is similar to the reporting rate which was obtained in chapter 3 from the best-fitting negative binomial models fitted to department-level data (0.016, 95% CrI: 0.015-0.017). In contrast, Mier-y-Teran-Romero et al. estimated a reporting rate over two times larger (0.03, 95% CrI: 0.01-0.07) [133]. Moore et al. also obtained a larger result for this parameter. They estimated both the probability that a symptomatic ZIKV infection is reported as a suspected ZVD case and the probability that a symptomatic ZIKV infection is reported as a laboratory-confirmed ZVD case [192]. The probability that a symptomatic ZIKV infection is reported as a suspected or confirmed ZVD case had mean 0.040 (95% CrI: 0.019-0.077)[4]. Again, the difference in estimates could be explained by the use of seroprevalence data in this study to estimate infection attack rates. More precise estimates of reporting rates would be expected from better estimates of the infection attack rates, which could explain why the credible intervals here are narrower than those reported by the other studies.

The overall estimate of the number of ZIKV-associated neurological complications reported per 10,000 ZIKV infections was 0.51 (95% CrI: 0.17-0.92). This estimate is lower than that reported by Mier-y-Teran-Romero et al. (2.0, 95% CrI: 0.6-4.6) [133] and Moore et al. (2.9 GBS cases per 10,000 symptomatic ZIKV infections, 95% CrI: 1.4-5.5[5]) [192] for Colombia. One possible reason for the discrepancy is that three different datasets for ZIKV-associated neurological complications were used by the studies. Mier-y-Teran-Romero et al. used a total of 677 cumulative cases, which were reported in the INS Weekly Epidemiological Bulletin at the end of 2016, and Moore et al. used 773 cases which were reported to PAHO. Although these were the only case numbers available at the time, they were nevertheless

[4] This estimate was not reported in the paper. It was obtained from the posterior samples of the MCMC. After removing the burn-in, the four chains of each parameter were merged. Then, the posterior samples for the relevant overall probabilities for Colombia were summed.

[5] This estimate was also not reported in the paper and was obtained from the posterior samples of the MCMC for the overall probability for Colombia as above.

subject to misclassification. Here, 418 cases remained following a verification process. Another contributing factor is that higher infection attack rates were found in this study compared to those reported by Mier-y-Teran-Romero et al. and Moore et al. Higher estimated infection attack rates would lead to lower estimates of the risk of neurological complications following ZIKV infection.

The model estimated that there are 54 (95% CrI: 5-210) reported cases of ZIKV-associated neurological complications for every 10,000 reported cases of ZVD on average. This estimate is less than half that from Mier-y-Teran-Romero et al.'s study (111, 95% CrI: 0-567), but the two findings are not inconsistent according to the credible intervals. Their estimate incorporated baseline levels of GBS for all locations except Colombia and Puerto Rico. They assumed that GBS cases occur at a mean rate of 1.1 cases per 100,000 population per year [133]. Thus, the interpretation of their estimate more closely aligns with the total expected number of reported GBS cases from all causes for every 10,000 reported ZVD cases, while the estimate here can be interpreted as excess cases of neurological complications due to ZIKV per 10,000 reported ZVD cases.

Interestingly, the point estimates from the raw data (calculated as reported cases of ZIKV-associated neurological complications divided by ZVD cases) for 12 out of 28 cities in this study fell outside of their 95% credible interval. Eleven of those cities reported zero cases of ZIKV-associated neurological complications. Based on the number of reported ZVD cases in those cities, the model predicted there would be reported ZVD cases with neurological complications. It is possible that ZVD cases with neurological complications were not reported because they did not occur due to chance. Thirteen out of seventeen cities that did report ZVD cases with neurological complications reported fewer than 10, making it a rare event. Another possible explanation is that severe cases of ZVD were under-ascertained because of barriers to healthcare access, particularly in rural areas. A study on the availability and distribution of medical specialists offering high and medium complexity services in Colombia estimated that the country had only one neurologist and one neurosurgeon for every 100,000 population in 2011 [240]. Together, these indicators are lower than the global median of the total neurological workforce (comprising the total number of adult neurologists, neurosurgeons, and pediatric neurologists), which was estimated by the WHO at 3.1 per 100,000 population from data collected between 2014 and

2015 [241]. Among World Bank income groups, high-income countries reported the largest number with a median of 7.1 per 100,000 population, which were followed by upper-middle-income countries (including Colombia) with a median of 3.1 [241]. In Colombia, medical specialists tend to be concentrated in the capital cities. Despite this fact, capital cities located in predominately rural departments struggle with lack of specialists. According to one hospital manager in Leticia, a major limitation faced by specialists is work-related travel. For every three weeks that a specialist works, another specialist must agree to replace them for one week while they rest [240].

The models for sex and age group quantified the biases in the ZIKV surveillance data that were explored in chapter 2. Important differences in overall ZVD reporting rates and the risk of being reported as a ZVD case with neurological complications were found by comparing the estimated posterior probabilities. Overall, female cases of ZVD were more likely than male cases to be reported to the surveillance system but less likely to be reported as cases with neurological complications, assuming the same infection attack rates. These findings are consistent with the results in chapter 2, which showed that more ZVD cases were reported in women of child-bearing age than expected based on the age distribution of the population and males have higher risk of neurological complications. The same trend in reporting rates was found for most cities, but no important differences in the risk of neurological complications by sex were found at the city level (possibly due to lack of power). Overall, younger individuals had higher ZVD reporting rates than older individuals assuming the same infection attack rates, while older individuals were more likely to be reported as cases with neurological complications. These findings are also expected based on the results in chapter 2 and GBS epidemiology. Similar results were obtained at the city level for reporting rates by age, but few cities had important differences in the risk of neurological complications, which, again, could be related to small sample sizes and lack of power. Interestingly, the cities with important differences in estimates by sex and age group tended to include those with more available data, including seroprevalence data and reported cases of ZIKV-associated neurological complications.

ZVD is not the only disease for which multiple data types have been employed to overcome biases or gaps in surveillance data. For example, Watson et al. used community-uploaded obituary certificates to validate a mathematical model of COVID-19 transmission dynamics

in Damascus, Syria [242]. The model, which was fitted to reported COVID-19 deaths, estimated that only 1.25% of deaths (sensitivity range 1%-3%) from COVID-19 were reported in Damascus between July 2020 and August 2020. The alternative data source confirmed substantial under ascertainment of mortality over that time period [242]. Another example comes from the Global Polio Eradication Initiative, which began in 1988 after the WHO declared poliovirus a target for eradication [243]. Surveillance of acute flaccid paralysis (AFP) is the primary means by which poliovirus is monitored globally. However, as only one in 200-1,000 individuals infected by poliovirus becomes paralyzed, the majority of infections are not detected by AFP surveillance, leaving gaps. Consequently, both environmental sampling of sewage and genetic sequencing of polioviruses have been employed to improve the identification of poliovirus outbreaks, understanding of their spread, and determination of the appropriate vaccination response [243].

A limitation of this analysis is that seroprevalence data were only available for four cities. Cities without these data had much wider credible intervals, especially for ZIKV infection attack rates and ZVD reporting rates. Although removing data from the four cities did not seem to greatly affect parameter estimates for the remaining cities, the estimated infection attack rates for some cities were nonetheless surprising. For example, Barranquilla and Cartagena are located just 100 km apart along Colombia's Caribbean coast and have similar altitudes. Yet, the estimated infection attack rate for Barranquilla was much higher than Cartagena (0.79, 95% CrI 0.54-0.99 versus 0.13, 95% CrI 0.04-0.42). Estimated infection attack rates for Pereira and Riohacha were also lower than expected, given that both cities are considered hyperendemic for DENV [229]. Seroprevalence data from these cities would help improve these estimates.

As mentioned in chapter 2, inaccuracy in population projections from DANE would also affect the results in this chapter and in the previous one, namely impacting the estimated reporting rates. For example, if a particular city or department had a smaller population size than the projected population size, then the observed reporting rate would appear to be smaller. Given that a single reporting rate was estimated across all departments in chapter 3, it is unclear whether ρ would have been over- or under-estimated.

All available data from Colombia, including 50,415 additional cases of ZVD and 194 additional cases of ZIKV-associated neurological complications, were not used in this analysis due to restricting to capital cities. Also, because reporting rates are expected to be higher in capital cities compared to non-capital cities and departments, the findings here may not be generalizable to the rest of the country. A strength of this analysis is the fine spatial scale of the available data and the fact that the neurological complications dataset was checked against standardized case definitions. Future research should investigate ZIKV surveillance biases in other countries; however, comparable datasets may not be available. Some countries such as Ecuador reported very few cases of ZIKV-associated neurological complications to PAHO [244].

In conclusion, the need for additional data sources to overcome biases in surveillance data was highlighted in this chapter. Differences in ZIKV infection attack rates and reporting of ZVD cases were observed across capital cities and across sex and age groups. The risk of ZIKV-associated neurological complications was also estimated. These severe ZVD cases may have been under ascertained in rural cities, where greater incentives are needed to attract and retain medical specialists.

Chapter 5: Spatial and temporal invasion dynamics of the 2014-2017 Zika and chikungunya epidemics in Colombia

Work in this chapter formed the basis of a manuscript that has been published in *PLOS Computational Biology* [245].

Abstract

Understanding the spatial and temporal dynamics of Zika virus (ZIKV) and chikungunya virus (CHIKV) at the subnational level is key to informing surveillance and preparedness for future epidemics. In this chapter, surveillance data were used to analyze transmission between cities using a suite of (i) gravity models, (ii) Stouffer's rank models, and (iii) radiation models with two types of distance metrics, geographic distance and travel time between cities. Invasion risk was best captured by a gravity model when accounting for geographic distance and intermediate levels of density dependence; Stouffer's rank model with geographic distance performed similarly well. Although a few long-distance invasion events occurred at the beginning of the epidemics, an estimated distance power of 1.7 (95% CrI: 1.5-2.0) from the gravity models suggests that spatial spread was primarily driven by short-distance transmission. Similarities between the epidemics were highlighted by jointly fitted models, which were preferred over individual models when the transmission intensity was allowed to vary across arboviruses. However, ZIKV spread considerably faster than CHIKV.

1 Introduction

1.1 Spatiotemporal epidemiology

Spatiotemporal epidemiology is the study of the distribution and determinants of health-related states or events across time and geographic space. Since at least the 1800s, maps have been used to study the causes of infectious disease outbreaks. Early notable examples include yellow fever in the southern USA and cholera in London, England [246]. Since then, computers, modern statistics, and geographic information systems have greatly increased researchers' capacity to investigate factors involved in the geographic variation of disease, infectious disease transmission, and control strategies [247].

Methods for analyzing spatiotemporal data tend to fall into one of two categories, spatial statistical modeling and spatial transmission dynamic (mathematical) modeling [248]. Spatial statistical methods involve finding relationships between space-time patterns in infectious diseases and aspects of the host or environment, creating maps of incidence or prevalence, and identifying hotspots or clusters of disease [248]. In contrast, mathematical models are helpful for understanding the mechanisms underlying disease transmission and may be used to envisage alternative scenarios of potential epidemic trajectories, describe and forecast outbreaks, and evaluate interventions [248].

Over the last two decades, several infectious disease epidemics have been analyzed using spatiotemporal methods. In 2001 during the foot-and-mouth disease epidemic in the UK, mathematical models accounting for spatial contact patterns between farms were used to estimate the impact of interventions, including culling, vaccination, and movement restrictions [249]. Spatial transmission models of the 2014-2016 Ebola virus epidemic in West Africa were used to compare the effectiveness of local versus long-range interventions such as quarantine and border closures [250]. Spatiotemporal models are currently being used in conjunction with mobility data to estimate transmission intensity of COVID-19 and the impact of social distancing measures on the pandemic [251].

1.2 Human movements and disease spread

As recently as three or four generations ago, it was not uncommon for individuals to live their entire lives within a few tens of kilometers from where they were born [252]. Over the last two centuries human mobility is estimated to have increased by over 1,000-fold in western countries [253]. The rise in mobility has been associated with elevated risk of infectious disease spread. Before the twentieth century, travel via walking, horseback, and ships was implicated in spreading devastating epidemics of plague, cholera, and smallpox. The movement of these and other diseases had profound effects on the course of human history [254]. In the twentieth century, commercial aviation dramatically decreased the amount of time needed to travel from one place to another; people, and they diseases they carry, began to move with unprecedented speed. Consequently, air travel has played an important role in the spread of influenza, HIV, and novel coronaviruses over the last few decades [254]. In 2003, the year in which the SARS epidemic occurred, individuals took

about 1.7 billion journeys by plane. By 2018, that number more than doubled to 4.2 billion [255]. Today, infections introduced by travelers continue to pose risks ranging from sporadic cases of infectious diseases, such as monkeypox introduced into the UK [256], to epidemics and pandemics that can lead to new areas of endemicity [257].

The study of human movements is relevant to several areas of research and is carried out by geographers, ecologists, engineers, epidemiologists, and others. The types of movements under investigation can include trips to work, tourism, and migration [258]. Economic and social activities are key drivers of human mobility. Movement patterns are based on a population's transportation network, which can include road and rail routes as well as air and sea corridors [258]. Sources of mobility data include census data and surveys, currency tracking, mobile phone records, GPS on cars and mobile phones, and location-based social network data [259]. Understanding human movements can improve traffic forecasting, urban planning, and epidemic modeling [259].

Population movements can be modeled as flows between distinct locations that are connected by a network. A spatial kernel defines how the different subpopulations interact. In infectious disease epidemiology, a spatial transmission kernel is the probability distribution of distances between the location of a donor and recipient of an infection [260]. Factors that influence the shape of the spatial kernel include host behavior, mode of transmission, and environment [260]. For vector-borne diseases, human movements affect exposure to vectors and consequently the transmission of pathogens. There is considerable evidence that human movement is responsible for the spread of DENV across large spatial scales. Outbreaks in several countries including Australia, China, and Japan have been associated with the arrival of DENV-infected travelers from endemic regions [261]. At a finer spatial scale, Stoddard et al. used a case-control study with contact tracing to study human-mediated DENV dispersal between households over two transmission seasons in Iquitos, Peru. They found that households visited by DENV-infected individuals had higher infection risk and transmission rates than those visited by uninfected individuals [262]. In other words, movement between a person's residence and the households of family and friends was an important driver of DENV spread in the community.

1.2.1 Gravity models

Gravity models describe movement from one location to another based on population size and distance [263]. The idea that humans are drawn together by a social gravitational force comes from sociology and dates back to at least the 1850s. Drawing from Newton's law of gravitation and observations of human movement, Carey stated that "Man, the molecule of society,… tends of necessity to gravitate towards his fellow-man… Gravitation is here, as everywhere else in the material world, in the direct ratio of the mass, and in the inverse one of the distance" [264]. Nearly one-hundred years later, this concept was formalized in a mathematical equation. In the 1940s, Zipf described the movement of people between population centers in the USA. He proposed that the number of people that travel between cities 1 and 2 is proportional to P_1*P_2/d, where P is the population size of each city and d is the distance between them [265]. This formula reflects how larger population centers exert a greater "pull" on people than smaller ones with a penalty for distance traveled. Zipf tested his hypothesis with data on the circulation of newspapers, telephone calls, and bus passenger movements [265].

Traditionally gravity models have been used to study the flow of goods and services in spatially distributed populations [263]. Over the last few decades, gravity models have been applied to biological systems, including the transmission of infectious diseases between regions, due to their simplicity and ability to capture several aspects of epidemic dynamics. One of the first studies to apply gravity models to infectious disease dynamics was Xia et al. [266]. They used a time series SIR model within a gravity model metapopulation framework to investigate the spread of measles outbreaks from 1944-1967 in England and Wales. Their model was able to reproduce key spatiotemporal features of measles dynamics during the pre-vaccination era, including case rates, cyclic and seasonal patterns, as well as epidemic extinction rates [266]. Gravity models have since been used to study the spread of a wide variety of human infectious diseases, including cholera [267, 268], dengue [269], Ebola [250, 270], influenza [271-274], and yellow fever [275], among others [276].

Although early applications of gravity models in the field of ecology assumed an inverse relationship between connectivity of locations and distance (a spatial kernel of 1/distance) [277], this assumption likely does not hold for human travel. More complex relationships

have been used through modifications to the gravity model, most commonly by the estimation of a distance exponent. This allows the model to estimate the rate at which movement decreases with distance [278]. Some models also include an "offset," an additional parameter that limits the kernel at short distances [279]. Other distance kernels have been used in the literature as well as different functional forms. For example, Viboud et al. used different kernels above and below a distance threshold of 119 km in their study of influenza in the USA [274]. The distance between locations can be the Euclidean (straight-line) distance or proxies of human mobility, including work commutes and air travel [271, 274].

Additional parameters, such as those allowing for spatial interaction, can be added to the basic gravity model. The connection between locations such as cities can be described by the extent of density dependence, which ranges from density dependent to density independent [272]. If connection is density dependent, the total connectivity of a city increases along with its number of neighbors according to the sizes of those cities and distances between them. If, however, connection is density independent, a city's number of neighbors does not affect the total connectivity between that city and its neighbors. For the spread of infectious diseases, density-dependent transmission means that cities with many neighbors will experience greater total infection pressure compared to cities with few neighbors, and density-independent transmission means that cities with many neighbors will experience the same total infection pressure as more isolated cities [272]. While past studies tended to assume density dependence [266, 274] or density independence [267, 280], more recent studies have estimated the level of density dependence [271-273]. Gravity models that estimate density dependence are also known as Fotheringham's competing destinations model [281].

Gravity models can be used to infer population mobility even when mobility data are unavailable [272]. Lack of movement data is common for many low- and middle-income countries, including Colombia [282]. Once gravity models are validated, they can predict changes in connectivity as populations increase or decrease, or as migration changes due to violence, economic crises, and natural disasters. This ability to make predictions is a clear advantage over movement surveys, which only provide information about a particular point in time [272].

1.2.2 Alternative models

Despite their growing popularity, gravity models have clear limitations, such as analytical inconsistencies [283]. Bjørnstad et al. recently recommended that more than one class of models should be explored when attempting to predict the spatial spread of infectious diseases [284]. Alternative models for understanding disease spread have been proposed, including Stouffer's rank model and the radiation model. Stouffer's rank model, also known as the law of intervening opportunities, states that the number of people traveling a particular distance is proportional to the number of opportunities at that distance and inversely proportional to the number of opportunities along the way [285]. It was first suggested in 1940 by Samuel Stouffer, an American sociologist. The equation is

$$\frac{\Delta y}{\Delta s} = \frac{a}{x}\frac{\Delta x}{\Delta s} \tag{5.1}$$

where Δy is the number of persons moving from an origin to a circular band of width Δs. Distance can be measured in units of space, time, or cost. x is the number of intervening opportunities, or the total number of opportunities between the origin and distance s. Δx is the number of opportunities within the band of width Δs, and a is a constant. Opportunities must be clearly defined and can vary depending on the study. Examples include jobs for people of a particular profession, such as nurses, or universities to which recent high school graduates could apply. Stouffer tested his theory empirically using data on residential mobility of families in Cleveland, Ohio [285].

Similar to Stouffer's rank model, the radiation model also accounts for higher-order interactions among population centers. It was proposed as an alternative to the gravity model by Simini et al. in 2012 [283] and can be used to predict commuting and mobility fluxes. According to the radiation model, the average flux T_{ij} from location *i* to location *j* is

$$\langle T_{i,j} \rangle = T_i \frac{N_i N_j}{\left(N_i + s_{i,j}\right)\left(N_i + N_j + s_{i,j}\right)} \tag{5.2}$$

where N_i and N_j are the population sizes of locations *i* and *j*, respectively, which are separated from each other by distance $r_{i,j}$. $s_{i,j}$ is the population in the circle of radius $r_{i,j}$ centered at *i* (the source and destination population sizes are not included). $T_i = \sum_{j \neq i} T_{i,j}$ is the number of commuters that start their trip from location *i*. Therefore, $T_i = N_i(N_c/N)$,

where N_c denotes the number of commuters and N is the population size of the country. The radiation model is "parameter-free" and has been used to study epidemics such as ZIKV [23], Ebola virus [286], and cholera [284].

1.3 Spatiotemporal analyses of ZIKV and CHIKV

Since the 2010s, several studies have been conducted on the spatiotemporal spread of ZIKV and CHIKV.

1.3.1 Mathematical models

In 2015, Cauchemez et al. used a simple distance-based model to describe the spread of CHIKV across 40 countries and territories in the Caribbean. They estimated that the risk of transmission between areas was inversely proportional to distance [214]. Air passenger flow was a weak predictor of transmission. Roche et al. analyzed the spatiotemporal dynamics of CHIKV on the island of Martinique. They found that mosquito abundance and human behavior, as measured by textual analysis of posts on the social media site Twitter, best explained the spread of disease [287]. Chadsuthi et al.'s best-fitting metapopulation model for the spatial spread of CHIKV in Thailand included driving distance between districts, human movement represented by a gravity model, rubber plantation area, and three long-distance translocation events [288].

Using an individual-based, stochastic, and spatial epidemic model, Zhang et al. assessed the spatiotemporal dynamics of ZIKV for several countries in the Americas. They found that the epidemic moved slowly across the continent and was mainly limited by seasonality in the virus' transmission [53]. Gardner et al. estimated risk factors of ZIKV spread resulting in local mosquito-borne transmission in the Americas using a stochastic-dynamic epidemic model. Mosquito abundance, incidence rate at the origin region, and human population density were identified as risk factors. Air passenger flows played less of a role, and the most important factor was an inverse relationship with regional gross domestic product per capita, a proxy for vector control availability [289]. O'Reilly et al. used a deterministic metapopulation model to project ZIKV incidence for 2018 in 90 large cities within 35 countries in the Americas. They also estimated key transmission parameters and country-specific disease reporting rates, as mentioned in chapter 3. They found that migration between cities was better captured by a gravity model than a model based on flight data

[23]. Prem et al. estimated an infection tree for the 2016 ZIKV outbreak in Singapore. They estimated 64.2% of infections occurred at workplaces and several individuals may have spread ZIKV over long distances [290].

1.3.2 Statistical models

Rossi et al. used boosted regression trees to analyze CHIKV outbreak data from 1959 to 2009 in 76 countries in the Indian Ocean region. By studying pairs of countries that experienced outbreaks in the same year, they determined that CHIKV outbreaks were more likely to co-occur in countries that were near each other and had high population densities [291]. Nsoesie et al. performed spatiotemporal clustering analyses on data from the 2013-2015 CHIKV outbreak in Dominica. Densely populated areas were found to have statistically significant clustering of cases [292]. Lizarazo et al. studied the 2014 CHIKV outbreak in Carabobo state, Venezuela. Using trend surface analysis and kriging, they estimated that the virus spread radially over a distance of 9.4 km at an average velocity of 82.9 m/day [293]. They also used the Knox method to identify disease clusters. The area with the most clusters was described as densely populated with lower socioeconomic status of residents and crowded living conditions [293].

Bisanzio et al. used historical data on DENV to infer ZIKV and CHIKV introduction and spread in Merida, Mexico. Clustering analyses by census tract identified statistically significant overlap in the spatiotemporal distribution of the three viruses (Kendall's $W > 0.63$, $p < 0.01$) [294]. Dalvi and Braga studied the spatial diffusion patterns of CHIKV, DENV, and ZIKV during the 2015-2016 epidemics in Rio de Janeiro, Brazil. The results of a suite of statistical analyses indicated that all three viruses followed an expansion diffusion pattern, meaning the diseases spread outward from the place of origin with decreasing intensity [295].

1.4 Spatiotemporal analyses of ZIKV and CHIKV in Colombia

Given the size of the ZIKV and CHIKV epidemics in Colombia, some spatiotemporal studies have focused specifically on this country.

1.4.1 Mathematical models

Moore et al. fitted agent-based models to weekly case reports of CF at three different spatial scales (national, departmental, and municipal) in Colombia. They showed that model

fits improved when the model was calibrated at finer spatial scales, highlighting the role of heterogeneity in environmental and demographic characteristics on epidemic dynamics [235].

1.4.2 Statistical models

Rees et al. used two statistical models to examine the spread of ZIKV across cities in Colombia [296]. Using a logistic regression model, they estimated the probability of a city reporting at least one ZIKV case during the epidemic. They found that the probability of reporting ZIKV increased with warmer mean study period nighttime temperatures, higher connectivity between cities, a higher proportion of neighboring cities reporting ZIKV, and time. Factors that decreased probability of reporting ZVD cases included higher total study period precipitation and higher poverty (proportion of the population with Unsatisfied Basic Needs), especially in areas with low connectivity [296]. Maps based on this model showed increased risk of ZIKV in the northern and central regions of the country.

Using an accelerated failure time survival model, Rees et al. also estimated the time to first reported case in week *t* for each city as an indicator of invasion. Connectivity between cities and proportion of neighbors reporting ZVD cases increased time to invasion, whereas mean elevation, total weekly precipitation, poverty, distance to nearest city reporting ZIKV, and the interaction between poverty and connectivity decreased time to invasion. When precipitation was assessed on a weekly scale, they found an immediate rather than a time-lagged effect on invasion [296]. An important limitation of this analysis is that weekly data at the city level were not available before January 9, 2016. To obtain cases counts prior to January 9, they assumed cases doubled weekly since the start of the epidemic. As a result, their study period began on October 24, 2015, which is still 10 weeks after the first cases were reported in the country [296].

Perkins et al. used an algorithm to classify proportional cumulative incidence curves for ZIKV at the city, department, and national levels. They also simulated data from a stochastic transmission model to assess the effects of environmental variables on how incidence curves were classified. They found that temporal incidence varied by city but were unable to evaluate the role of spatial interaction [297]. The study period ranged from August 2015 to September 2016, and missing data for 2015 was imputed.

Flórez-Lozano et al. used a Bayesian hierarchical model to study the spatial distribution of ZIKV risk in Colombia [298]. The central and southern parts of the country were found to have higher risk; during the peak of the epidemic, the risk in these regions was up to four times higher compared to the initial phase [298]. This study was limited to about 41,000 suspected and laboratory-confirmed cases aggregated at the department level.

Martínez-Bello et al. used the integrated nested Laplace approximation to model risk of DENV and ZIKV in the department of Santander and its capital city, Bucaramanga [299]. Their best-fitting department model indicated that the risk of DENV or ZIKV in one city was associated with the risk in the same city in the preceding week. At the city level, the best-fitting model indicated that the risk of DENV or ZIKV in one census sector was associated with both its neighboring census sectors in the same week and in the previous week. High risk cities and census sectors were identified and mapped. The study period ranged from October 2015 to December 2016 [299].

McHale et al. studied the spatiotemporal heterogeneity in the ZIKV and CHIKV epidemics and the potential for clustering in the city of Barranquilla between 2014 and 2016 [300]. Hotspots for both viruses were identified using Moran's I statistic and local indicators of spatial association. Multivariate conditional autoregressive models were used to identify risk factors for case incidences associated with living in each neighborhood [300]. They found higher socioeconomic stratum and proximity of neighborhood to major roads significantly increased the risk of ZIKV case incidences. None of the explored risk factors were significant to explain CHIKV case incidences. Typically, low socioeconomic status is associated with increased risk of arboviral infections. The authors note that their findings could be related to higher reporting rates in wealthier communities with better access to healthcare [300].

1.5 Aims

With few exceptions [294, 300], the spatiotemporal spread of ZIKV [53, 289, 297] and CHIKV [214, 287, 293] in the Americas has been studied separately. However, the viruses share common vectors and were both introduced into a new region with immunologically naïve populations. An integrated study of these diseases in the same locations may help uncover similarities and differences between the two. Previous analyses of the ZIKV and CHIKV

epidemics in Colombia have relied on partial INS surveillance datasets [296-298]. The aim of this chapter was to analyze transmission between cities in Colombia by fitting a suite of spatial interaction models, including variations of the gravity model, Stouffer's rank model, and radiation model, using the complete surveillance datasets. The invasion dynamics of both viruses were examined as well as the extent to which inter-city transmission depended on distance, population sizes of invaded and susceptible cities, and the infectivity of ZIKV and CHIKV.

2 Data

2.1 Epidemiological data

The ZIKV and CHIKV datasets that were described in chapter 1 were used in this chapter. After removing cases with missing administrative level 2 location, 105,152 ZVD and 411,789 CF cases remained for analysis.

Figure 5.1 shows the epidemiological curves for CF and ZVD cases at the department level. Most departments had some overlap in reported CF cases and ZVD cases. Although none of the departments had a peak in the incidence of both diseases at the same time, only eight weeks separated the peaks of the CHIKV and ZIKV epidemics in Putumayo.

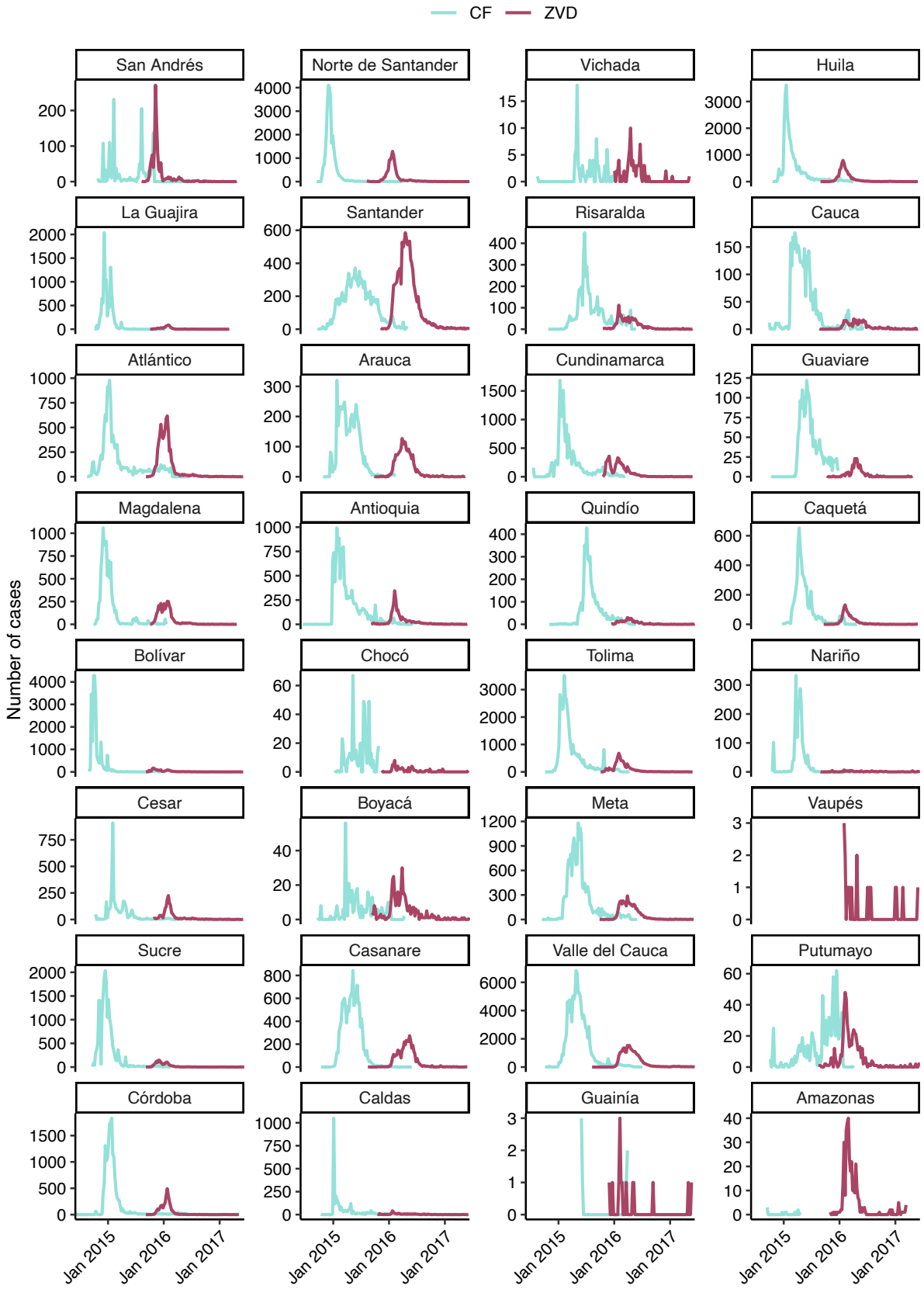


Figure 5.1 Epidemiological curves of CF and ZVD cases in Colombia by department, 2014-2017. Departments are ordered from North to South down the columns. Y-axes are different for each plot.

2.2 Demographic data

As in previous chapters, population projections derived from the 2005 Census were obtained for 2016 from DANE. Population sizes were re-scaled by dividing by 10,000. For cities included in the main analysis, population sizes ranged from 1,670 to about 2.4 million.

2.3 Distance metrics

Data on two different distance metrics were obtained, (i) geographic distance and (ii) a proxy of human mobility. Latitudes and longitudes corresponding to the geographic center of each city were provided by collaborators at Colombia's INS. For geographic distance, the geodesic distances between cities were calculated using the Vincenty inverse formula for ellipsoids using the gdist function in the R package Imap (version 1.32) [301].

Data on accessibility to cities in Colombia for 2015 were obtained from the Malaria Atlas Project [302]. The downloaded friction surface consists of average land-based travel speed, and units are in minutes to travel 1 meter. The costDistance function in the R package gdistance (version 1.2-2) was used to calculate the time to travel between cities in the country [303]. The islands of San Andrés and Providencia were not included in the dataset because they cannot be reached by land. In order to compare models across distance metrics, these two cities were dropped from preliminary analyses. Puerto Colombia in the department of Atlántico was also missing from this dataset but is not an island. Because this city shares a large border with Barranquilla, the same values were used for both locations, and the time to travel between them was assumed to be 0 minutes.

Figure 5.2 is a map of travel times for the northern part of South America. In Colombia, the most accessible cities can be found in the Andean region and on the Caribbean coast. The least accessible parts of the country are located in the Amazon region in the south, where travel between cities can take up to several days.

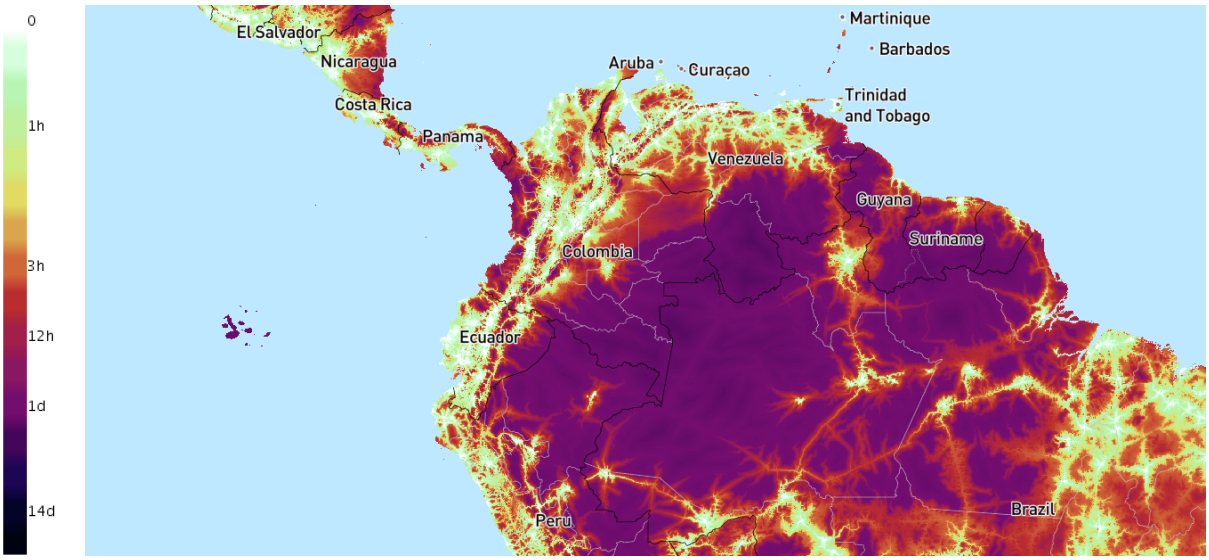


Figure 5.2 Map of predicted travel time to the nearest city in minutes. Shown for the northern part of South America. Reproduced from [302].

2.4 Data for analysis of invasion risk factors

2.4.1 Epidemiological data

The DENV dataset described in chapter 1 was also used to analyze risk factors of CHIKV and ZIKV invasion.

2.4.2 Elevation

SRTM 90m Digital Elevation Data for Colombia were downloaded from the CGIAR-CSI GeoPortal. The altitude of each city in meters was extracted according to its latitude and longitude [304]. To better approximate human risk of vector-borne disease, coordinates corresponding to the population weighted centroids were used for most cities (n = 1,047). Due to missing data, geographic center was used for some locations (n = 75).

2.4.3 Weather data

Weather data were described in chapter 3. They consist of population weighted weekly mean temperature and population weighted weekly cumulative rainfall at the city level.

2.4.4 Socioeconomic data

Multidimensional poverty data for Colombia were downloaded from DANE at the city level for the year 2018 [305]. The source of information for the calculation of multidimensional

poverty at the department level is the Encuesta Nacional de Calidad de Vida (Quality of Life Survey). The annual Quality of Life Survey gathers information about Colombians' living and housing conditions. DANE used information collected from the 2018 Censo Nacional de Población y Vivienda (the Census) to approximate multidimensional poverty at the city level. The calculations are based on households that were effectively censused. The definitions of each variable can be found in chapter 3.

3 Methods

The methods in this chapter involve performing a descriptive analysis and fitting spatial interaction models. Both rely on the estimated invasion week in each administrative level 2 unit, denoted here as "cities." After the virus is introduced into a city, the invasion week represents the time in which an epidemic begins, allowing the city to spread the virus to other cities. In the descriptive analysis, the geographic origin of the epidemics will be estimated, long-distance transmission events will be identified, and the elevation of invaded and uninvaded cities will be compared. For the spatial interaction models, four types of models will be considered as well as two distance metrics. The results of a sensitivity analysis and validation of the parameter fitting procedure will also be presented. Finally, risk factors of CHIKV and ZIKV invasion will be determined.

3.1 Invasion weeks

3.1.1 Determination of invasion week using first reported cases

Whether a city was invaded by either ZIKV or CHIKV was determined by the number of reported cases in each city. For ZIKV, cities that reported at least 30 cases of ZVD were considered to have been "invaded." For CHIKV, the cut-off for CF cases was set at 20. Cities with case counts below these thresholds were not considered in the primary analysis. Invasion was defined as the week before cases were first reported in each city. A latent period of one week was assumed, after which the city is considered infectious and can spread the infection to other cities.

3.1.2 Generation time method for determining invasion week

To test the sensitivity of the models to the choice of invasion weeks, an alternative method was explored using weekly time series of CF and ZVD cases. For each disease and city

separately, the first week of maximum incidence was identified. If zero cases were reported in the week just before (t_{max-1}) and in the preceding two or three weeks, for CHIKV and ZIKV respectively, then this was the week of invasion. If cases were reported during this time period, the rule stipulated counting back one week at a time until the condition was met. The two and three weeks correspond to each infection's generation time [210, 213]. If zero cases are reported during this time period, then there is no evidence that transmission is occurring, and the disease is not yet established in that place assuming complete reporting.

Figure 5.3 shows an example of this algorithm for determining invasion week in each city.

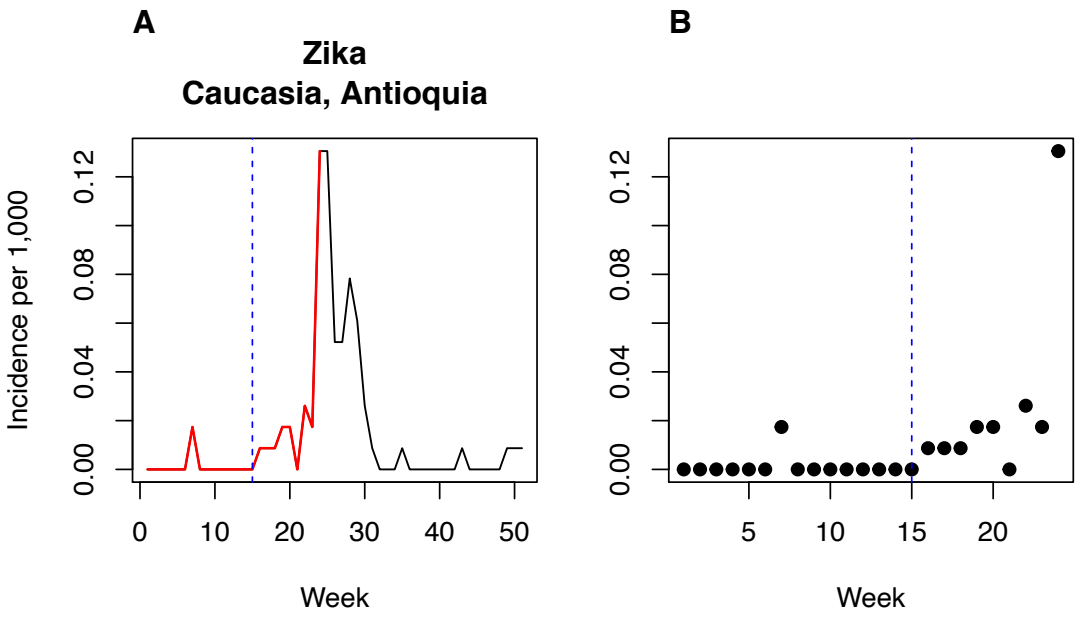


Figure 5.3 Example of algorithm used to estimate week of invasion using the generation time method. *(A)* The time series for Caucasia in the department of Antioquia during the 2015-2017 ZIKV epidemic. In this figure, week 1 corresponds to the week ending on August 15, 2015, and week 51 corresponds to the week ending on July 30, 2016. The algorithm identifies the point of maximum incidence in the time series and counts backward one week at a time until there are no reported cases. If there are no cases in this week or the prior two or three weeks depending on the infection's generation time, then this is the week of invasion. If not, the algorithm continues to go back in time until the condition is met. The part of the line in red is the period used to determine the onset of invasion, and the blue dashed line is the estimated invasion week. *(B)* The same time series as in *(A)* is shown until the point of maximum incidence (week ending on January 23, 2016). The estimated week of invasion is week 15 rather than week 21 because cases were reported in weeks 16-20.

3.2 Elevation

The elevation of cities that were invaded versus cities that escaped invasion were compared for (i) CHIKV, (ii) ZIKV, and (iii) CHIKV, ZIKV, or DENV.

Invasion for CHIKV and ZIKV was defined as above. Invasion by DENV was defined as cities with at least 98 reported cases between 2010 and 2016. Ninety-eight was used as a cut-off because it is the median number of cases reported among cities that reported at least one case during this time period.

The Wilcoxon rank sum test was used to assess the difference in elevation between invaded cities and uninvaded cities.

3.3 Potential sources of the epidemics

The Colombian cities where the epidemics most likely began were identified. The method is based on the concept that epidemics spread radially from the origin. In other words, the relationship between invasion week and geographic distance from the source is linear [271]. The first 10% of invaded cities were considered as potential origins for the epidemics assuming a single introduction of each virus into Colombia. Pearson's correlation coefficient for the relationship between the city's invasion week and its geographic distance to the origin was calculated for each potential origin. The city with the highest correlation coefficient was identified as the most likely source.

3.4 Long-distance transmission events

The number and location of long-distance transmission events of ZIKV and CHIKV were identified using the invasion week in each city. The method detects outliers in the distribution of pairwise distances between newly invaded cities and the set of infectious cities at the previous time step [271]. The set of cities in the network is C. At time t_j, C is divided into the set of invaded cities, which are capable of spreading infection, I_{tj} and the set of susceptible cities, which can become invaded, S_{tj}:

$$I_{t_j} = \{k: t_k < t_j\} \tag{5.3}$$

$$S_{t_j} = \{k: t_k \geq t_j\} \tag{5.4}$$

where t_j is the timing of invasion of city j and t_k is the timing of infectiousness (invasion week plus one week) in each of k cities. For city j, the minimum distance between city j and invaded cities in I_{tj} was calculated as d_j, the most likely route of invasion when the spatial dynamics are driven by distance.

$$d_j = \min_k d_{jk} \tag{5.5}$$

for $k \in I_{tj.}$ D_j, the minimum distance between city j and any other city in the network, was also calculated:

$$D_j = \min_i d_{ji} \tag{5.6}$$

for $i \in C$. If the process were entirely spatial, cities would usually be infected by their neighbors. Thus, the distance to the nearest city is approximated by $d_j - D_j \approx 0$. For each city, $d_j - D_j$ was calculated; those included in the 99th percentile of the distribution of $d_j - D_j$ were considered long-distance transmission events.

3.5 Spatial interaction models

Four main types of spatial interaction models were considered: (i) the gravity model, (ii) the competing destinations model, (iii) Stouffer's rank model, and (iv) the radiation model. For each model type, both geographic distance and travel time between cities were considered.

Model parameters were initially estimated independently for each virus. From the four best-fitting models with a common structure, joint models were run assuming the same parameters across CHIKV and ZIKV. Joint models in which some parameters were allowed to vary across arboviruses were also explored. From the first approach, the best-fitting model for each virus was obtained. From the second approach, the best-fitting joint model highlighting the commonalities in the spatiotemporal dynamics across CHIKV and ZIKV was obtained.

3.5.1 Gravity models

Gravity models were fitted to analyze transmission of CHIKV and ZIKV between cities that met the thresholds for reported cases.

For each virus separately, N cities have an invasion week, t_i, which was defined as one week prior to the first report of cases. Cities also have population size, P_i, which is assumed to be constant over time and weekly case counts weighted by the generation time distribution, $c_{i,t}$. The geographic distance in km (or travel time in minutes) between invaded city i and susceptible city j is d_{ij}. For geographic distance, the geodesic distance on an ellipsoid was

used. This distance is the shortest path between two points accounting for the curvature of the Earth.

At each time point, a city can be either "susceptible," "latently infected," or "infectious." Once cities are invaded, they are latently infected for one week and then infectious. The model assumes no additional cases are imported from abroad after external seeding into Colombia occurs. This implies that if a city was invaded in week t_i, only cities within the country that were infectious in the previous week could have spread the disease to that city. Transmission parameters were assumed to remain constant over time.

As in Eggo et al.[272], the force of infection, λ, represents the hazard of infection from an invaded city to a susceptible city. At time *t*, the force of infection from city *i* to city *j* can be defined as:

$$\lambda_{i\to j,t} = \beta c_{i,t}^{\phi} P_j^{\mu} \frac{\frac{P_i^{\upsilon}}{d_{ij}^{\gamma}}}{\left(\Sigma_{k,k\neq j} \frac{P_k^{\upsilon}}{d_{kj}^{\gamma}}\right)^{\varepsilon}} \quad (5.7)$$

Exponents ν and μ are for population sizes of city *i* and *j*, respectively. The distance between cities is d_{ij} and γ is the power parameter. β describes transmission intensity. ϕ captures the relationship between infectivity of a city and its weekly case count weighted by the generation time distribution. Weighting was performed in R using the overall_infectivity function in the EpiEstim package (version 1.1-2) [164]. This function calculates the overall infectivity due to previously infected individuals from an incidence time series. The overall infectivity λ_t at time *t* equals the sum of the previously infected individuals (from incidence I) weighted by their infectivity at time *t* (from the discrete serial interval distribution w_k). In other words, $\lambda_t = \sum_{k=1}^{t-1} I_{t-k} w_k$. Values for the mean and standard deviation of the generation time distributions used for the weighting are the same as in chapter 3 (Table 3.5) with mean 14 days and standard deviation 6.2 days for CHIKV [213] and mean 20 days and standard deviation 7.4 days for ZIKV [210]. A value of ϕ = 1 indicates that the infectiousness of a city at time t_i is proportional to the number of cases reported in that city weighted by the generation time distribution at time t_i. When ϕ = 0, infectiousness does not depend on the number of reported cases in the source city. Values of ϕ between 0 and 1 lead to

infectiousness profiles that vary according to weekly case counts. Parameter ε characterizes the density dependence of the connection between a susceptible city and all invaded cities. When ε = 0, transmission scales linearly with population density, and the formulation above reduces to a simple density-dependent model. When ε = 1, transmission does not depend on population density, and the formulation above reduces to a density-independent model. When ε is estimated, the model is equivalent to Fotheringham's competing destinations model [281]. The total force of infection on city *j* at time *t* is defined by:

$$\lambda_{j,t} = \sum_{i \neq j}^{i} \lambda_{i \to j,t} I_{ij,t} \tag{5.8}$$

where $I_{ij,t} = \begin{cases} 1, \text{if } i = \text{Infectious and } j = \text{Susceptible} \\ 0, \text{otherwise} \end{cases}$

The probability that a susceptible city *j* is invaded at time t_j is

$$P(t_j) = exp\left(- \sum_{\tau=0}^{t_j - 1} \lambda_{j,\tau} \right) \left(1 - exp\left(-\lambda_{j,t_j} \right) \right) \tag{5.9}$$

The first part of equation (5.9) is the probability that a city escaped invasion from *t* = 0 until just before t_j. The second part is the probability that the city was invaded at t_j given that it was susceptible until that week. The conditional log likelihood is summed over all susceptible cities:

$$l = \sum_{j} \ln \left(P(t_j) \right) \tag{5.10}$$

For models that were fitted to all 1,122 cities in Colombia, the probability that a susceptible city *j* is invaded at time t_j was the same as above. The probability that a city escaped invasion was

$$P(t_j) = exp\left(- \sum_{\tau=0}^{T} \lambda_{j,\tau} \right) \tag{5.11}$$

where the force of infection for city *j* is summed over the entire duration of the epidemic (from week 0 to T).

Null models that only included β were fitted first. Additional parameters were added to test for a spatial effect in transmission, the role of population size of invaded and susceptible cities, and infectivity. Except for β, which is always estimated by MCMC, parameters can be fixed at 0, at 1 (not γ), or estimated by MCMC.

3.5.2 Stouffer's rank model

Following [284], population size was used as a proxy for "opportunities." The force of infection from city *i* to city *j* at time *t* is

$$\lambda_{i\to j,t} = \beta c_{i,t}^{\phi} P_j^{\mu} \left(\frac{P_i}{\sum_{k\in\Omega(j,i)} P_k} \right)^{v} \tag{5.12}$$

where $k \in \Omega(i,j)$ is the group of cities that are closer to susceptible city *j* than invaded city *i*: $\Omega(j,i) = \{k : 0 < d(j,k) \leq d(j,i)\}$. A variant of this model in which city *j* is included among the intervening opportunities was also considered. In "Stouffer's rank variant model," $\Omega(j,i) = \{k : 0 \leq d(j,k) \leq d(j,i)\}$ which allows within-city opportunities to decrease spatial coupling.

3.5.3 Radiation model

The version of the radiation model used here is shown in equation (5.13)

$$\lambda_{i\to j,t} = \beta c_{i,t}^{\phi} P_i \frac{P_i P_j}{\left(P_i + \sum_{k\in\Omega(i,j)} P_k\right)\left(P_j + P_i + \sum_{k\in\Omega(i,j)} P_k\right)} \tag{5.13}$$

where again two variants are considered, one in which city *j* is excluded ("radiation") from the set $\Omega(j,i)$ and one in which it is included ("radiation variant"). This model lacks a parameter for the spatial component.

3.6 Model estimation and computing

As in chapter 3, Metropolis-Hastings MCMC sampling was used to investigate the posterior distributions of parameters [224, 225]. A log normal distribution was used as a proposal distribution. As this distribution is asymmetric and only allows positive parameter values, the Metropolis accept-reject rule was corrected for asymmetric jumping. Parameters were

updated one at a time. Uniform prior distributions were used for all parameters. Three chains were run for each model with different starting values. Chains were visually checked for convergence after 100,000 iterations with a burn-in of 0.2 times the length of the chains (iterations times number of parameters). The coda package (version 0.19-4) in R was used to calculate the Gelman-Rubin statistic for each best-fitting model to check convergence [86, 226]. Median parameter estimates and 95% credible intervals were calculated from the posterior distributions after excluding the burn-in.

DIC was used to compare models. Lower values of DIC are preferred, and a difference of about 5 is important [227]. DIC was calculated using the medians of the posterior distributions of the parameters due to non-normality of the likelihood.

All analyses were performed in R version 3.5.1. Aggregated data and code are available on GitHub (https://github.com/kcharniga/zika_chik_invasion).

3.7 Validation of gravity model fit and sensitivity analyses

The probability distribution of invasion week was calculated for each city based on the observed start of invasion in other cities up to that time. This calculation was performed for each virus using the median parameter estimates from the posterior distribution. The probability distributions were compared with the observed invasion weeks.

Simulations were also performed to check model fit. The ZIKV and CHIKV epidemics in Colombia were simulated using 1,000 parameter sets sampled from the posterior distributions. For each set, the epidemic started in the city (or cities) invaded in the first week. A random variable was chosen from a uniform distribution between 0 and 1 for each city in each week. If the probability of t_j was higher than the random variable, the city became invaded. Once invaded, the observed weekly case counts weighted by the generation time distribution were used to model that city's infectiousness over time. Epidemic simulations were also used to test the sensitivity of the cut-offs used to determine invasion criteria for each virus and were performed for the best-fitting Stouffer's rank models.

Gravity models were also fitted to (i) first reported cases and all 1,122 cities in Colombia and (ii) estimated invasion week using the generation time method rather than the method based on first reported cases as sensitivity analyses.

3.8 Validation of parameter fitting procedure

Using the framework described in the previous section, the fitting procedure for the model parameters was validated by simulating one dataset for each virus with the median parameter estimates obtained from the best-fitting models. The analysis was re-run on each simulated dataset to check that the fitted parameter estimates could be recovered.

3.9 Risk factors of invasion

Logistic regression models were used to determine risk factors for invasion by CHIKV and ZIKV. The outcome was defined as a city reporting at least 20 cases of CF and at least 30 cases of ZVD for each respective model. Predictors included population size, elevation, dengue risk, temperature, rainfall, and mean travel time as well as the percentage of households in each city with overcrowding, inadequate exterior walls, and inadequate flooring. Dengue risk was categorized into four levels as follows: cities located at or below 1800 m of elevation that reported any cases of DF between 2010-2016 were considered “at risk” of dengue. The natural logarithm of the cumulative number of cases over this period was taken and divided into tertiles (1-3 with 3 being the highest). All other cities were assigned values of 0. Mean temperature for each city was obtained by taking the mean of the weekly population-weighted weekly time series of mean temperature over the study period, defined as the time during which cases were being reported in the country (110 weeks for CHIKV and 97 weeks for ZIKV). Similarly, mean rainfall was calculated for each city as the mean of the population-weighted weekly time series of cumulative precipitation for each respective study period. As a sensitivity analysis, the mean of the weather covariates from the week cases were first reported until the peak of each epidemic (34 weeks for CHIKV and 26 weeks for ZIKV) was also considered. Mean travel time for each city was defined as the average time to travel from that city to all other cities, excluding the islands of San Andrés and Providencia.

The predictors were first explored in a univariate analysis. Significance of difference between invaded and uninvaded cities was tested by chi-square tests for categorical

variables and Wilcoxon rank-sum test for continuous variables, none of which were normally distributed. P values < 0.05 were considered statistically significant. A forward stepwise approach was then used to build each logistic regression model: predictors were added to the model one at a time and only kept if they were significant at the 0.05 level. The units of rainfall and elevation were changed to 10 mm and 100 m, respectively, to improve the interpretation of the odds ratios (ORs) which were computed by exponentiating model coefficients. Models testing the effect of mean travel time were fitted to 1,120 cities for which data were non-missing. The Hosmer and Lemeshow goodness-of-fit test was applied to the best-fitting models using 10 groups.

Linear regression was also performed to assess the relationship between dengue risk and week of invasion for CHIKV and ZIKV.

4 Results

4.1 Characteristics of invaded cities

Invasion criteria was met by 338 cities for CHIKV and 288 cities for ZIKV out of 1,122 total cities. Table 5.1 shows a summary of the population sizes of the invaded cities. The minimum population sizes are the same, but the quartiles, median, mean, and maximum are larger for ZIKV compared to CHIKV.

Table 5.1 Population sizes of invaded cities. Summary statistics are given for cities invaded by either CHIKV or ZIKV.

Virus	Min	1st quartile	Median	Mean	3rd quartile	Max
CHIKV	1,670	13,578	26,192	72,081	55,805	2,394,925
ZIKV	1,670	14,322	27,786	88,591	60,829	2,486,723

Figure 5.4 compares the distance metrics for invaded cities with non-missing travel times data. The histograms show the geographic distance in km and the travel time in minutes between Barranquilla and all other invaded cities. There is one outlier in the travel times between ZIKV-invaded cities that is not shown; while the median travel time from Barranquilla to other invaded cities was about 580 minutes (9 hours), the estimated travel time to Leticia in the Amazonas department was 10,293 minutes (172 hours or about 7 days).

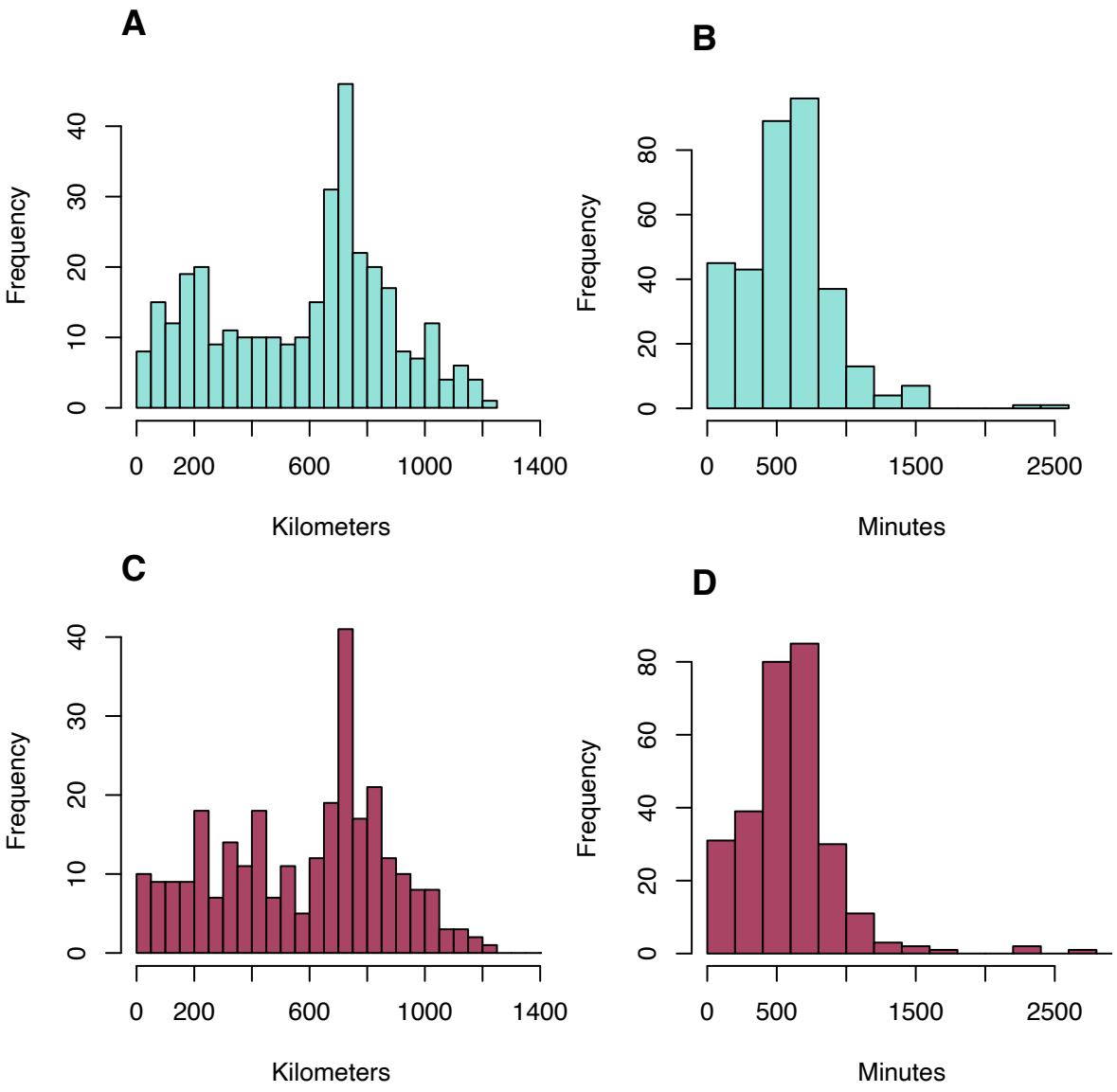


Figure 5.4 Comparison of distance metrics for invaded cities. *(A)* and *(B)* show the geographic distance in kilometers and travel time in minutes, respectively, between the first city invaded by CHIKV and all other invaded cities. *(C)* and *(D)* show the same for ZIKV. One outlier for Leticia is not shown in *(D)*.

In addition to cities invaded by either ZIKV or CHIKV, 525 cities met invasion criteria for DENV. The number of cities invaded by any of the three arboviruses (n=591) is consistent with a 2013 report by Colombia's MOH, which classified 56 cities as hyperendemic for DENV and 575 cities as mesoendemic [229]. As mentioned in chapter 4, the MOH criteria for level of DENV endemicity included trends of reported cases over time, number of circulating serotypes, age range of cases, and the presence of dengue hemorrhagic fever (severe dengue). Invaded cities were typically located below 2,000 m elevation (Figure 5.5). The

difference in elevation between invaded cities and uninvaded cities was statistically significant for each of the three sets of comparisons ($p < 0.0001$).

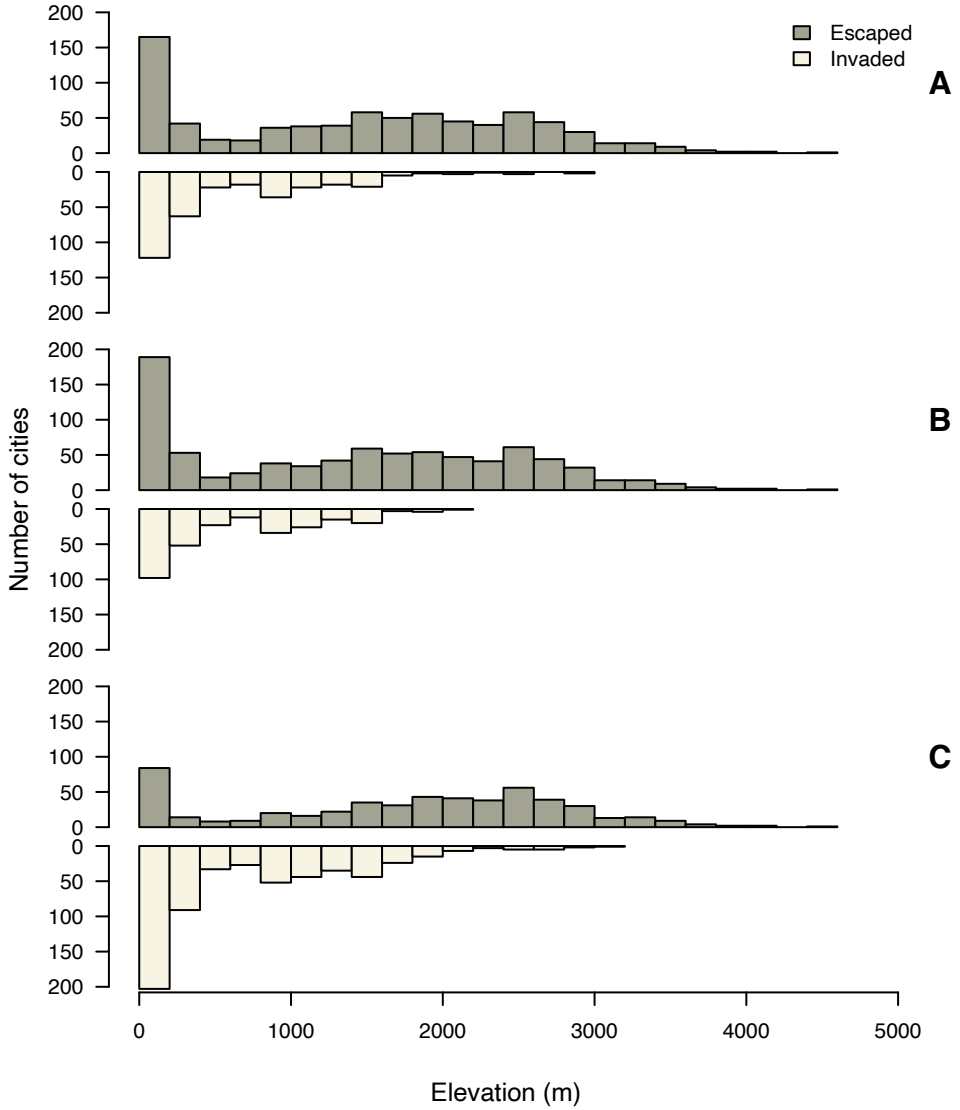


Figure 5.5 City elevation. Comparison of elevation in meters between cities that were invaded versus cities that escaped invasion for *(A)* CHIKV, *(B)* ZIKV, and *(C)* CHIKV, ZIKV, or DENV.

4.2 Spatiotemporal patterns in invasion weeks

4.2.1 Heatmaps

Figure 5.6 shows the spread of reported CF and ZVD cases during the epidemics.

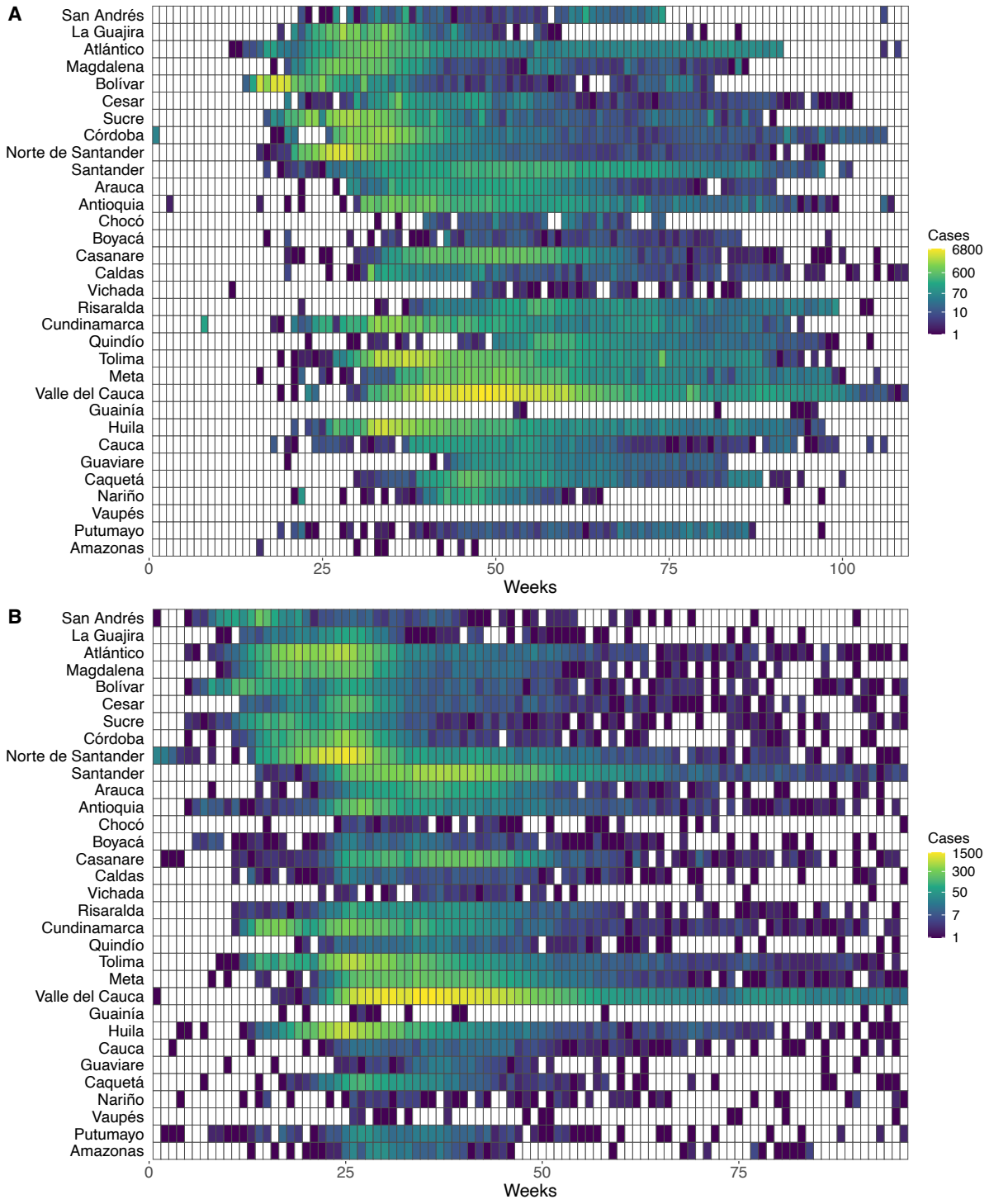


Figure 5.6 Heatmaps showing the spatiotemporal spread of CHIKV and ZIKV in Colombia. Population-weighted centroids were used to rank departments in order from North to South. Colors across rows represent the number of *(A)* CF and *(B)* ZVD cases for each department. Weeks are plotted on the x-axis starting from the first week cases were reported to the last week cases were reported. Dates for CHIKV *(A)* range from the week ending June 7, 2014 to that ending July 9, 2016, and dates for ZIKV *(B)* range from the week ending August 15, 2015 to that ending June 17, 2017. White rectangles are weeks with zero reported cases.

4.2.2 Invasion weeks

Invasion weeks for a random sample of cities from the ZIKV dataset are shown in Figure 5.7. The results for all cities can be found in Appendix S4. Invasion weeks ranged from the week ending May 31, 2014 to that ending September 19, 2015 for CHIKV and from the week ending August 8, 2015 to that ending March 26, 2016 for ZIKV. The time for the diseases to invade 50% of cities ever affected was shorter for ZIKV compared to CHIKV; while invasion weeks for ZIKV tended to cluster within five months (from September 2015 to January 2016), 90% of invasion weeks for CHIKV clustered within nine months (between September 2014 and May 2015) (Table 5.2). Two-hundred and five cities experienced epidemics of both CHIKV and ZIKV. For these cities, invasion weeks for both viruses were significantly positively correlated (Pearson's correlation coefficient 0.45, $p < 0.0001$). Both epidemics were first recorded in northern Colombia and spread from there. Early foci of disease also appeared in the central parts of the country (Figure 5.8).

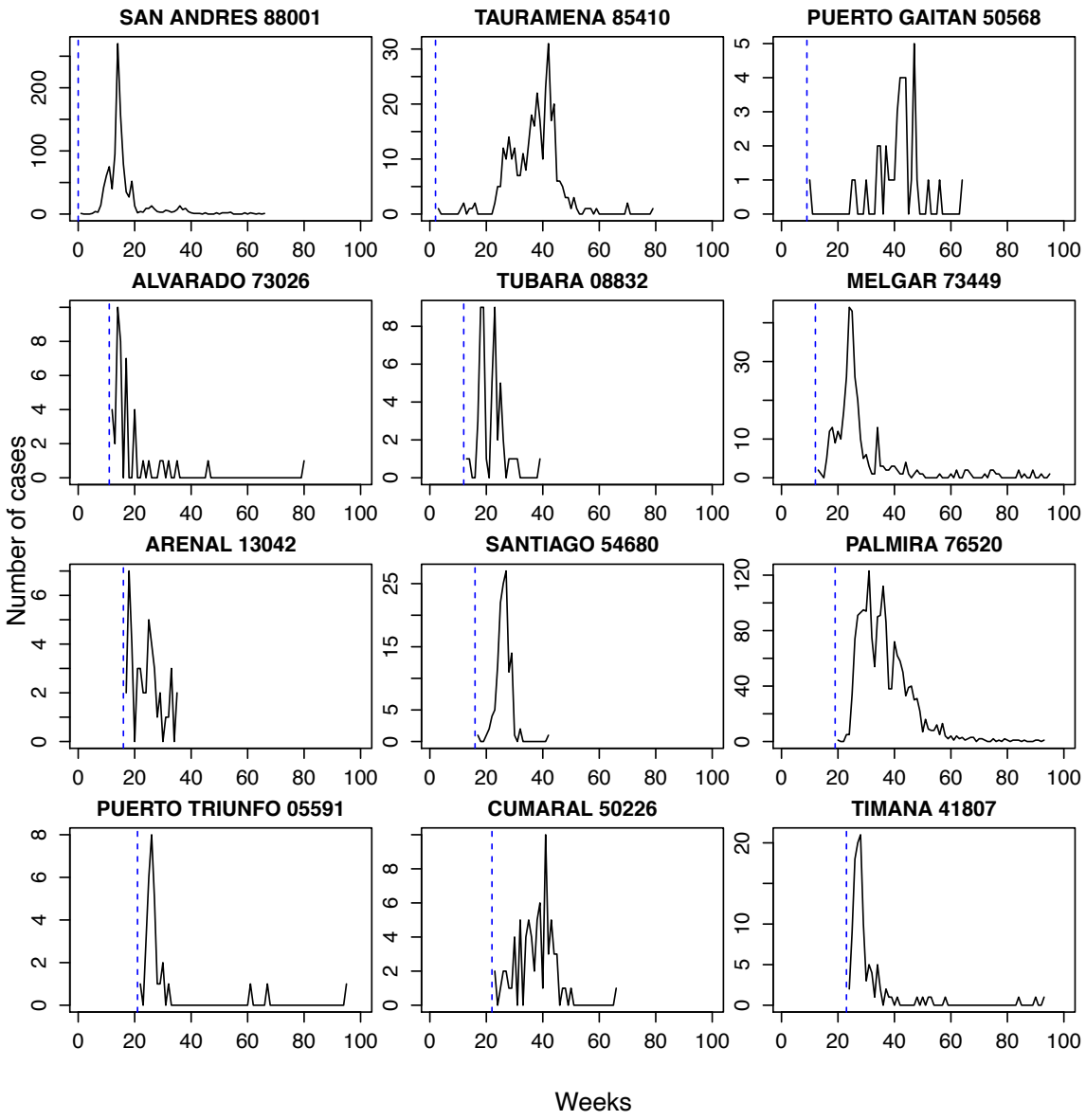


Figure 5.7 Invasion weeks for a random sample of cities from the ZIKV dataset. Invasion week, based on the first week of reported cases, is shown by the dashed blue line.

Table 5.2 Epidemiological characteristics of CHIKV and ZIKV epidemics in Colombia.

Virus	Cities (#)	Time for spread to 50%* (weeks)	Time for spread to 100%** (weeks)	Calendar time for 90% of spread***	Long-distance transmission events**** (#)
CHIKV	338	31	68	Sept. 2014-May 2015 (35 weeks)	4
ZIKV	288	16	33	Sept. 2015-Jan. 2016 (21 weeks)	3

*Time for 50% of cities to be invaded.
**Time from the first city to be invaded to the last city to be invaded.
***Calendar time for 90% of cities to be invaded (5th percentile to 95th percentile).
****More than 344.4 km for CHIKV and more than 321.2 km for ZIKV.

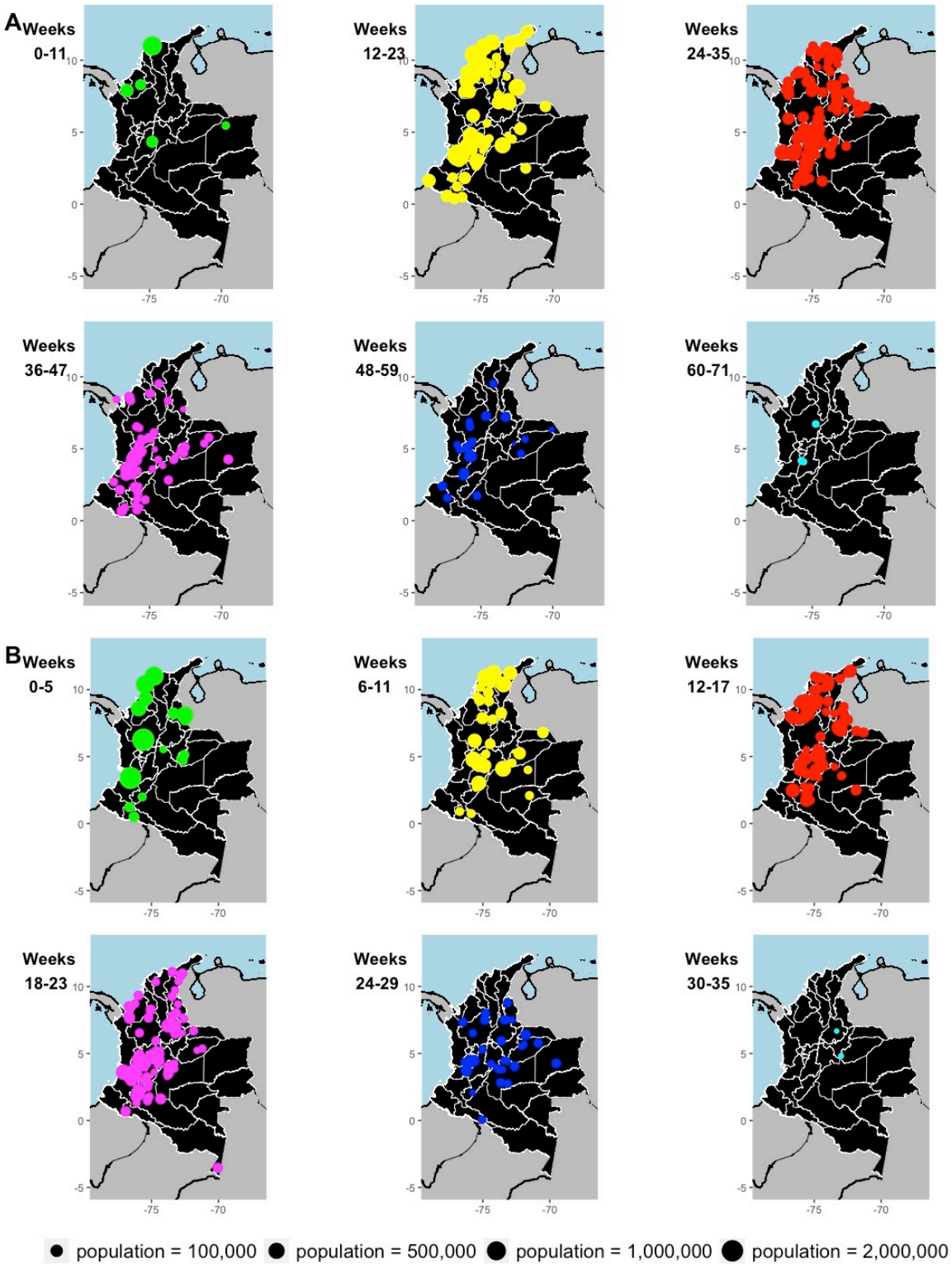


Figure 5.8 Geographic patterns of invasion weeks in studied cities in Colombia based on first reported cases. By 12-week period for (*A)* CHIKV and 6-week period for *(B)* ZIKV. Each circle represents a city. The size of the circle is proportional to population size. Each panel shows only cities newly invaded during the time period indicated in the upper left-hand-corner. The island of San Andrés is not shown but was invaded by CHIKV in week 21 and by ZIKV in week 0. Maps produced from GADM version 2.0.

4.2.3 Comparison of invasion week methods

Figure 5.9 shows the correlation between the generation time method of determining invasion week and the method based on first reported cases.

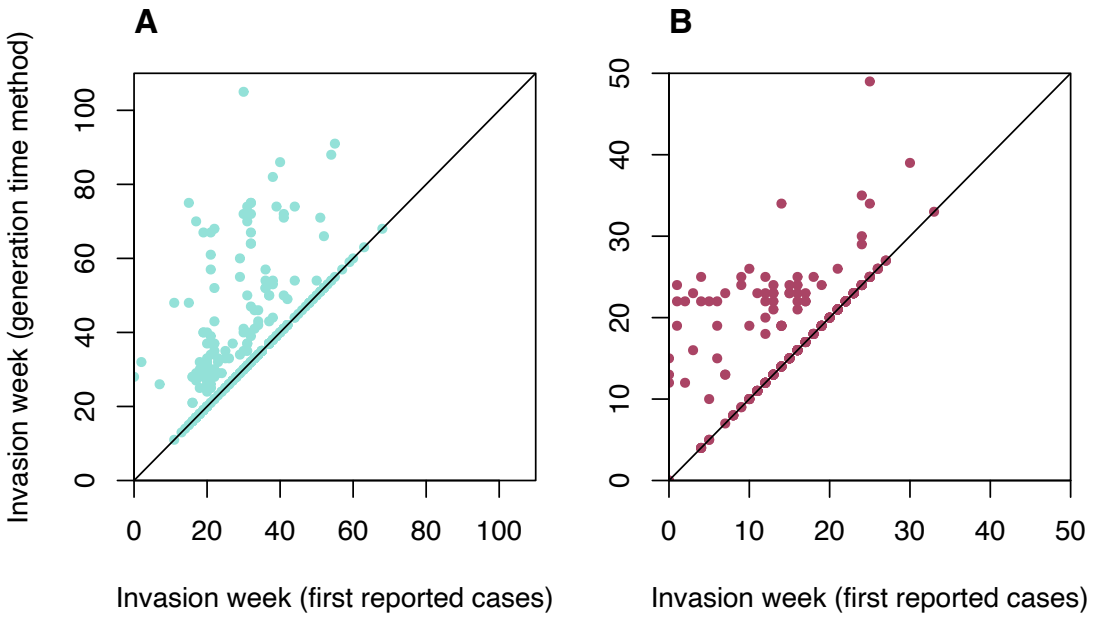


Figure 5.9 Comparison of estimated invasion weeks using a method based on the first reported cases in each city (x-axis) and a method based on the generation time distribution of each infection (generation time method, y-axis). *(A)* CHIKV and *(B)* ZIKV. The black line is y = x. The two methods show good agreement (CHIKV: r = 0.60, ZIKV: r = 0.68).

Methods for estimating invasion weeks have been developed previously to study influenza epidemics [272, 273]. These methods try to determine when influenza illnesses exceed baseline disease levels due to seasonal influenza. However, as there were no baseline levels of CF or ZVD in Colombia prior to 2014, even small numbers of cases are potentially interesting. Many of the epidemic curves at the city level do not resemble typical "bell curves" with reported cases rising and falling repeatedly. The proposed methods here are preferred over existing methods because invasion weeks more frequently come before large increases in cases. Additionally, vector biology is considered by including the generation time.

Figure 5.10 compares invasion weeks estimated using the generation time method and a method from Charu et al. using linear piecewise splines [271]. There is very good agreement between the two.

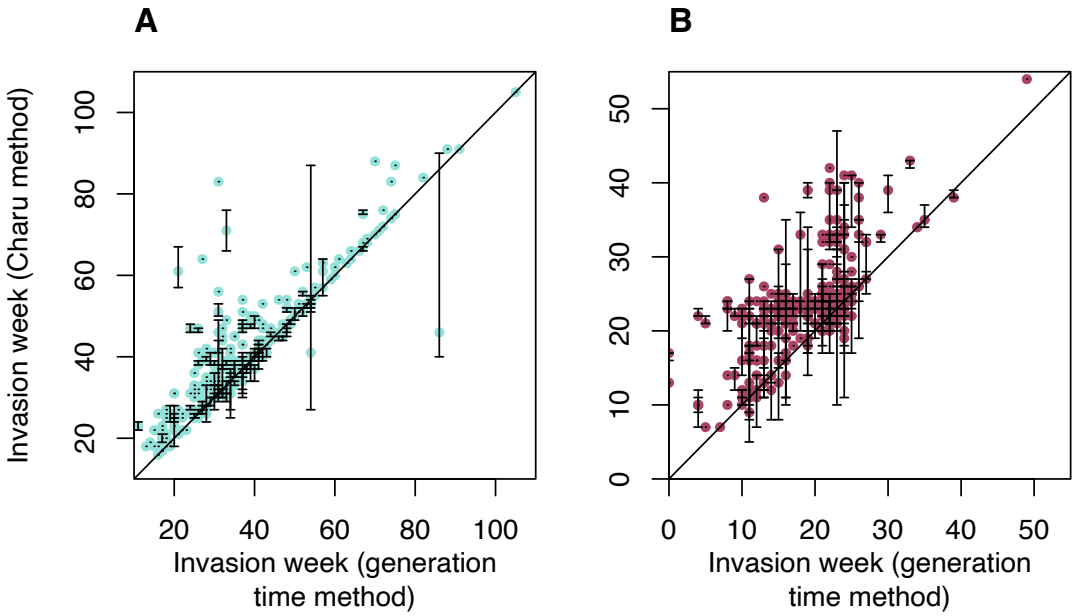


Figure 5.10 Comparison of estimated invasion weeks using a method based on the generation time distribution (generation time method, x-axis) and a piecewise spline method (Charu method, as in [271]**, y-axis).** *(A)* CHIKV and *(B)* ZIKV. 95% confidence intervals are shown for the Charu method only. For some cities, only the point estimate for t_j fell within the 95% confidence interval; this is shown by a lack of vertical bar. The diagonal line is y = x. The two methods show very good agreement (CHIKV: r = 0.90, ZIKV: r = 0.70).

4.3 Geographic origin of epidemics

The first city to report cases of CF in Colombia was Planeta Rica, Córdoba. Cases of ZVD were first reported in the country simultaneously by five cities: (i) Cali, Valle del Cauca, (ii) San Andrés, San Andrés and Providencia, (iii) Cúcuta, Norte de Santander, (iv) El Zulia, Norte de Santander, and (v) Puerto Santander, Norte de Santander. Assuming a linear relationship between invasion week and distance from the source of the epidemic, the estimated origin of both epidemics was Barranquilla, Colombia's fourth most populated city located on the Caribbean coast (Figure 5.11). According to the line list data, Barranquilla was among the first five cities to report cases of CF, first reporting cases in week 12 (invaded in week 11). The city was also among the first 18 cities to report cases of ZVD, first reporting cases in week 5 (invaded in week 4).

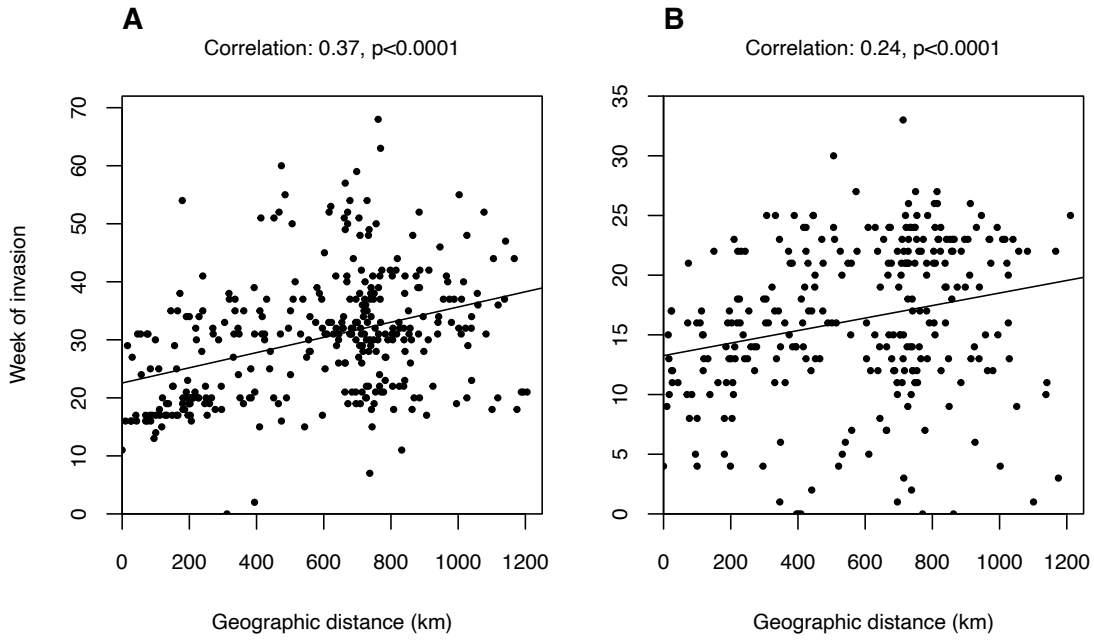


Figure 5.11 Correlations between city invasion weeks and geographic distance from first invaded cities for CHIKV and ZIKV. Week of invasion for each invaded city is shown on the y-axis for both plots. These weeks are plotted against *(A)* the geographic distance from the most likely origin of CHIKV in Colombia, Barranquilla and *(B)* the geographic distance from the most likely origin of ZIKV in Colombia, also Barranquilla. Pearson's correlation coefficients and significance are shown above each plot.

4.4 Long-distance transmission events

The minimum distance between a newly invaded city and cities invaded earlier was used to create a distribution (d-D), and cities invaded in the 99th percentile were classified as being invaded via long-distance events. Figure 5.12 shows the distribution of (d-D) for CHIKV and ZIKV graphically, and Table 5.3 shows summary statistics of this distribution.

Four long-distance transmission events were identified for CHIKV and three were identified for ZIKV. Cities on the receiving end of CHIKV included (i) Girardot, Cundinamarca, (ii) La Primavera, Vichada, (iii) Mocoa, Putumayo, and (iv) Puerto Asís, Putumayo. For ZIKV, cities included (i) Barranquilla, Atlántico, (ii) Tauramena, Casanare, and (iii) Cartagena, Bolívar. Three of these seven cities are department capitals. All of these events occurred early in the epidemics, within the first 15% of invaded cities. Long-distance transmission events for CHIKV occurred at distances of 366, 402, 431, and 475 km compared to a mean distance of 25.7 km. Long-distance transmission events for ZIKV occurred at distances of 322, 336, and

346 km compared to a mean distance of 25.6 km. Potential sources of the affected cities can be found in Tables 5.4-5.5.

The methods for estimating the epidemic origin and long-distance transmission events are independent of one another. For instance, Barranquilla is estimated as both the epidemic origin and a long-distance transmission event of ZIKV.

Table 5.3 Summary statistics of the d-D distributions showing that ZIKV and CHIKV exhibited similar patterns of transmission. The first six columns have units in km. The seventh column is the total sample size, and the last two columns contain the number of long-distance transmission events for two distance thresholds.

(d-D) Summary Statistics									
Virus	**Min**	**1st quartile**	**Median**	**Mean**	**3rd quartile**	**Max**	**# cities**	**N>99th%**	**N>97.5th%**
CHIKV	0	0	5.00	25.72	27.00	475	337*	4	9
ZIKV	0	0	3.50	25.60	23.00	346	280**	3	7

*Cities invaded in week 0 were excluded.
**Cities invaded in weeks 0 and 1 were excluded as week 1 was the first week in which any cities were considered infectious.

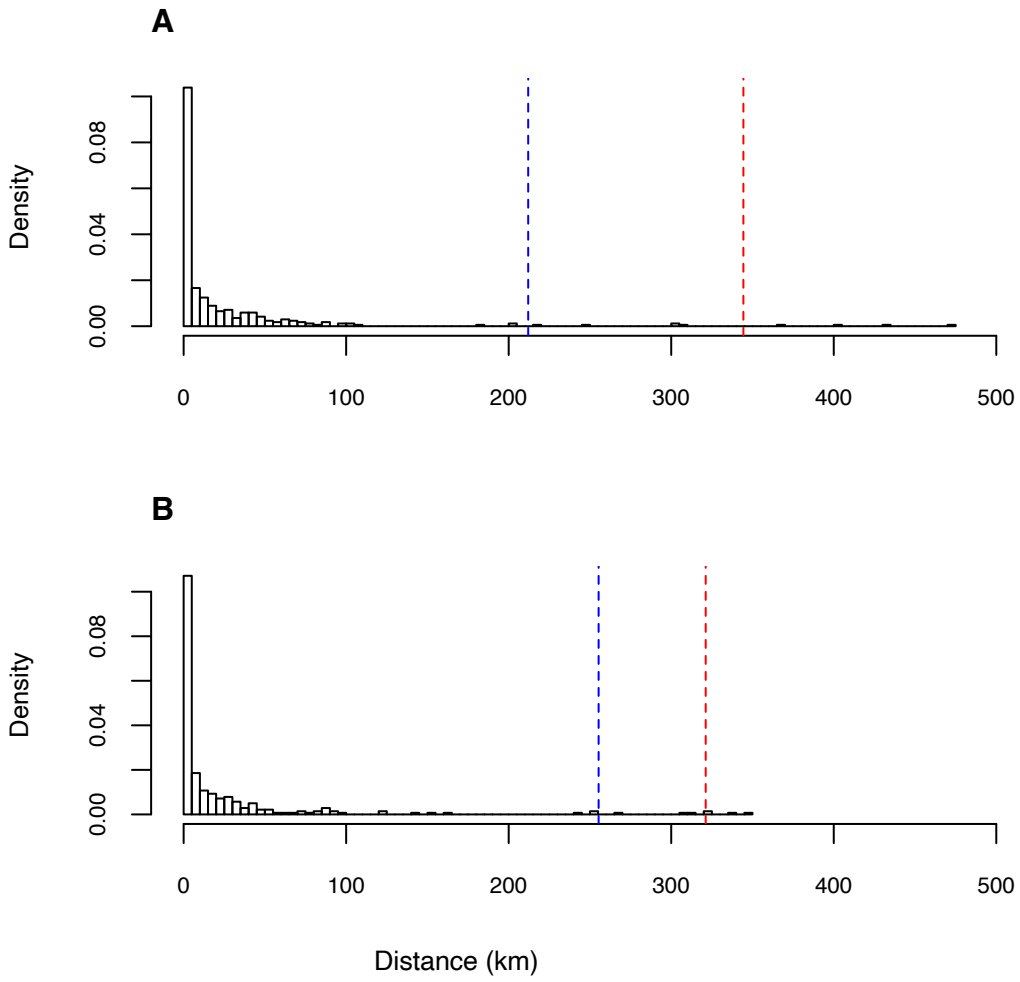


Figure 5.12 Long-distance transmission events. The distribution of d-D for *(A)* CHIKV and *(B)* ZIKV in this study. The dashed blue lines are plotted at the 97.5th percentile (corresponding to 212.0 km and 255.3 km for CHIKV and ZIKV, respectively) and the dashed red lines are plotted at the 99th percentile (corresponding to 344.4 km and 321.2 km for CHIKV and ZIKV, respectively). Long-distance transmission events were defined as invasions that occurred in cities included in the 99th percentile of this distribution.

Table 5.4 Recipient and potential source cities of long-distance transmission events of CHIKV.

Recipient cities	Potential source cities
Girardot, Cundinamarca	Apartadó, Antioquia Planeta Rica, Córdoba
La Primavera, Vichada	Apartadó, Antioquia Planeta Rica, Córdoba Girardot, Cundinamarca
Mocoa, Putumayo and Puerto Asís, Putumayo	Apartadó, Antioquia Envigado, Antioquia Barranquilla, Atlántico Baranoa, Atlántico Campo de la Cruz, Atlántico Sabanalarga, Atlántico Santo Tomás, Atlántico Soledad, Atlántico Suan, Atlántico Turbaco, Bolívar Cartagena, Bolívar San Juan Nepomuceno, Bolívar Santa Rosa, Bolívar Planeta Rica, Córdoba Girardot, Cundinamarca Barranca de Upía, Meta Cúcuta, Norte de Santander Floridablanca, Santander Corozal, Sucre La Primavera, Vichada

Table 5.5 Recipient and potential source cities of long-distance transmission events of ZIKV.

Recipient cities	Potential source cities
Tauramena, Casanare	Cúcuta, Norte de Santander El Zulia, Norte de Santander Puerto Santander, Norte de Santander San Andrés, San Andrés and Providencia Cali, Valle del Cauca
Barranquilla, Atlántico and Cartagena, Bolívar	Aguazul, Casanare Tauramena, Casanare Cúcuta, Norte de Santander El Zulia, Norte de Santander Puerto Santander, Norte de Santander Ocaña, Norte de Santander Villa del Rosario, Norte de Santander Mocoa, Putumayo San Andrés, San Andrés and Providencia Cali, Valle del Cauca

4.5 Models of spread

4.5.1 Models fitted independently to each virus

Models were initially fitted to cities that had available data on both distance metrics (337 and 287 cities for CHIKV and ZIKV, respectively). The best-fitting CHIKV model was Stouffer's rank model with geographic distance (Table 5.6). The next best-fitting models were Stouffer's rank model and a version of the gravity model that incorporates spatial interaction, also known as the competing destinations model, both fitted to travel time between cities. The change in DIC among the first three models was not meaningful (≤ 4). The fourth best-fitting model was a gravity model (competing destinations version) incorporating geographic distance, with a change in DIC of 6.5 compared to the best-fitting model. Although some models fitted to travel time between cities had lower DIC values than the same model type fitted to geographic distance, the difference was only meaningful for the radiation variant model. The radiation and radiation variant models performed the least well.

In contrast to CHIKV, the best-fitting model for ZIKV was the competing destinations version of the gravity model with geographic distance (Table 5.6). This model was followed by Stouffer's rank model and Stouffer's rank variant model, both incorporating geographic distance. Within model types, ZIKV models fitted to geographic distance were preferred over those fitted to travel time between cities. As with CHIKV, the models that performed least well were the radiation and radiation variant models. Other versions of the gravity models can be found in Tables 5.7-5.10.

For both epidemics, the best-fitting gravity model (based on the lowest DIC) included the following parameters: a distance power, power for invaded city population size, density dependence, infectivity, and transmission intensity. In both instances, the population size of the susceptible city appeared uncorrelated with the invasion dynamics. The estimated distance power, γ, for each model was 1.68 (95% CrI: 1.44-1.90) for CHIKV and 1.74 (95% CrI: 1.50-1.97) for ZIKV. Thus, the possibility that the relationship with distance was the same for both viruses cannot be excluded. Both models also estimated intermediate levels of density dependence.

In contrast, invasion risk was associated with the population size of the susceptible city in both Stouffer's rank models. Estimates of the infectivity parameter were similar to those obtained from the gravity models. Although the estimates of transmission intensity were lower in the Stouffer's rank models, ZIKV still had a higher estimate compared to CHIKV. Estimates of the effect of invaded city population size were stronger; however, because this parameter additionally captures spatial interaction, the interpretation is different compared to the gravity models.

Table 5.6 Comparison of alternative models of CHIKV and ZIKV spread in Colombia. Posterior median and 95% credible interval presented for each parameter. Models are ordered by sum of DIC and were fitted separately to 337 cities for CHIKV and 287 cities for ZIKV.

Virus	Model type*	Distance type**	DIC	Sum of DIC	γ (distance power)	μ (susceptible population)***	ν (invaded population)	ε (spatial interaction)	ϕ (infectivity)	β (intensity)
CHIKV	G	GD	2329.4	4044.9	1.68 (1.44-1.90)	0	0.65 (0.53-0.76)	0.83 (0.68-0.99)	0.35 (0.25-0.46)	0.24 (0.14-0.39)
ZIKV	G	GD	1715.5		1.74 (1.50-1.97)	0	0.55 (0.41-0.69)	0.67 (0.50-0.83)	0.27 (0.13-0.40)	1.11 (0.68-1.81)
CHIKV	S	GD	2322.9	4047.3		0.48 (0.37-0.58)	1.18 (1.01-1.36)		0.32 (0.24-0.42)	0.009 (0.005-0.015)
ZIKV	S	GD	1724.4			0.43 (0.31-0.55)	1.37 (1.12-1.63)		0.53 (0.44-0.63)	0.021 (0.013-0.032)
CHIKV	S	TT	2325.6	4054.8		0.47 (0.36-0.58)	1.16 (0.99-1.35)		0.31 (0.24-0.40)	0.008 (0.005-0.013)
ZIKV	S	TT	1729.2			0.42 (0.30-0.54)	1.44 (1.15-1.79)		0.48 (0.38-0.58)	0.023 (0.013-0.037)
CHIKV	G	TT	2326.9	4061.3	1.97 (1.69-2.25)	0	0.57 (0.45-0.68)	0.79 (0.70-0.87)	0.39 (0.28-0.52)	0.36 (0.19-0.65)
ZIKV	G	TT	1734.4		2.06 (1.73-2.40)	0	0.46 (0.33-0.59)	0.81 (0.71-0.90)	0.23 (0.11-0.36)	1.20 (0.72-1.97)
CHIKV	SV	GD	2333.5	4062.0		0.51 (0.40-0.61)	1.17 (0.99-1.35)		0.33 (0.25-0.41)	0.009 (0.005-0.014)
ZIKV	SV	GD	1728.5			0.46 (0.34-0.57)	1.36 (1.10-1.66)		0.54 (0.44-0.64)	0.022 (0.013-0.035)
CHIKV	SV	TT	2333.6	4070.4		0.51 (0.40-0.62)	1.15 (0.98-1.35)		0.33 (0.25-0.42)	0.008 (0.004-0.014)
ZIKV	SV	TT	1736.8			0.45 (0.33-0.57)	1.41 (1.11-1.71)		0.49 (0.40-0.59)	0.022 (0.013-0.035)
CHIKV	RV	GD	2427.3	4236.2					0.29 (0.24-0.36)	0.034 (0.025-0.043)
ZIKV	RV	GD	1808.9						0.65 (0.56-0.73)	0.034 (0.026-0.044)
CHIKV	RV	TT	2421.7	4240.0					0.30 (0.24-0.36)	0.029 (0.021-0.037)

ZIKV	RV	TT	1818.3						0.61 (0.52-0.69)	0.032 (0.025-0.040)
CHIKV	R	GD	2432.9	4246.9					0.30 (0.24-0.36)	0.030 (0.023-0.038)
ZIKV	R	GD	1814.0						0.65 (0.56-0.74)	0.032 (0.024-0.040)
CHIKV	R	TT	2431.3	4253.6					0.30 (0.24-0.37)	0.025 (0.019-0.033)
ZIKV	R	TT	1822.3						0.61 (0.52-0.70)	0.029 (0.023-0.036)

*G: gravity (competing destinations), S: Stouffer's rank, SV: Stouffer's rank variant, R: radiation, RV: radiation variant

**GD: geographic distance, TT: travel time between cities

***When μ is set to 0, this means that cities with large populations have the same risk of being invaded as cities with small populations.

Table 5.7 Parameter estimates for six models of CHIKV for 337 cities using geographic distance. Posterior median and 95% credible interval presented for each parameter. Bold indicates the best-fitting model. Travel time data were only available for 337 out of 338 cities. To compare across distance metrics, 337 cities were also used for geographic distance models.

	Distance-only model	Density-dependent population model	Density-independent population model	Estimated density dependence population model	Full (infectivity) model	**Infectivity model with μ set to 0***
DIC	2511.7	2440.7	2397.7	2396.9	2330.2	**2329.4**
γ (distance power)	1.05 (0.87-1.21)	1.02 (0.81-1.19)	1.46 (1.22-1.69)	1.49 (1.25-1.73)	1.68 (1.43-1.91)	**1.68 (1.44-1.90)**
μ (susceptible population)	0	0.081 (0.004-0.328)	0.037 (0.001-0.164)	0.034 (0.002-0.146)	0.060 (0.002-0.201)	**0**
ν (invaded population)	0	0.45 (0.34-0.55)	0.53 (0.42-0.64)	0.53 (0.41-0.64)	0.64 (0.52-0.76)	**0.65 (0.53-0.76)**
ε (spatial interaction)	0	0	1	0.85 (0.66-1.03)	0.83 (0.68-0.98)	**0.83 (0.68-0.99)**
ϕ (infectivity)	0	0	0	0	0.35 (0.25-0.47)	**0.35 (0.25-0.46)**
β (intensity)	0.16 (0.07-0.36)	0.080 (0.023-0.207)	0.26 (0.20-0.31)	0.31 (0.22-0.44)	0.21 (0.11-0.36)	**0.24 (0.14-0.39)**

*When μ is set to 0, this means that cities with large populations have the same risk of being invaded as cities with small populations.

Table 5.8 Parameter estimates for six models of CHIKV for 337 cities using travel time between cities. Posterior median and 95% credible interval presented for each parameter. Bold indicates the best-fitting model. Travel time data were only available for 337 out of 338 cities.

	Distance-only model	Density-dependent population model	Density-independent population model	Estimated density dependence population model	Full (infectivity) model	**Infectivity model with μ set to 0***
DIC	2518.9	2459.4	2403.5	2395.9	2328.3	**2326.9**
γ (distance power)	0.89 (0.71-1.06)	0.72 (0.53-0.90)	1.66 (1.37-1.96)	1.74 (1.45-2.03)	1.95 (1.66-2.24)	**1.97 (1.69-2.25)**
μ (susceptible population)	0	0.22 (0.01-0.50)	0.042 (0.002-0.187)	0.027 (0.001-0.127)	0.054 (0.002-0.203)	**0**
ν (invaded population)	0	0.40 (0.30-0.51)	0.49 (0.38-0.60)	0.48 (0.37-0.59)	0.56 (0.45-0.67)	**0.57 (0.45-0.68)**
ε (spatial interaction)	0	0	1	0.85 (0.75-0.94)	0.79 (0.69-0.87)	**0.79 (0.70-0.87)**
ϕ (infectivity)	0	0	0	0	0.39 (0.28-0.53)	**0.39 (0.28-0.52)**
β (intensity)	0.11 (0.04-0.28)	0.019 (0.005-0.067)	0.24 (0.18-0.29)	0.41 (0.26-0.62)	0.30 (0.13-0.59)	**0.36 (0.19-0.65)**

*When μ is set to 0, this means that cities with large populations have the same risk of being invaded as cities with small populations.

Table 5.9 Parameter estimates for six models of ZIKV for 287 cities using geographic distance. Posterior median and 95% credible interval presented for each parameter. Bold indicates the best-fitting model. Travel time data were only available for 287 out of 288 cities. To compare across distance metrics, 287 cities were also used for geographic distance models.

	Distance-only model	Density-dependent population model	Density-independent population model	Estimated density dependence population model	Full (infectivity) model	**Infectivity model with μ set to 0***
DIC	1804.5	1764.7	1749.0	1737.9	1715.5	**1715.5**
γ (distance power)	1.21 (1.01-1.38)	1.28 (1.10-1.44)	1.55 (1.31-1.78)	1.65 (1.43-1.87)	1.74 (1.50-1.97)	**1.74 (1.50-1.97)**
μ (susceptible population)	0	0.021 (0.001-0.093)	0.026 (0.001-0.113)	0.018 (0.001-0.084)	0.053 (0.003-0.210)	**0**
ν (invaded population)	0	0.37 (0.26-0.48)	0.46 (0.33-0.59)	0.47 (0.34-0.60)	0.55 (0.41-0.69)	**0.55 (0.41-0.69)**
ε (spatial interaction)	0	0	1	0.67 (0.48-0.86)	0.68 (0.50-0.86)	**0.67 (0.50-0.83)**
ϕ (infectivity)	0	0	0	0	0.29 (0.16-0.42)	**0.27 (0.13-0.40)**
β (intensity)	0.73 (0.28-1.63)	0.69 (0.28-1.47)	0.40 (0.32-0.48)	0.86 (0.51-1.39)	0.93 (0.53-1.59)	**1.11 (0.68-1.81)**

*When μ is set to 0, this means that cities with large populations have the same risk of being invaded as cities with small populations.

Table 5.10 Parameter estimates for six models of ZIKV for 287 cities using travel time between cities. Posterior median and 95% credible interval presented for each parameter. Bold indicates the best-fitting model. Travel time data were only available for 287 out of 288 cities.

	Distance-only model	Density-dependent population model	Density-independent population model	Estimated density dependence population model	Full (infectivity) model	**Infectivity model with μ set to 0***
DIC	1828.1	1803.8	1764.9	1753.2	1735.8	**1734.4**
γ (distance power)	0.79 (0.52-1.02)	0.73 (0.44-0.97)	1.85 (1.49-2.19)	1.99 (1.66-2.31)	2.05 (1.71-2.39)	**2.06 (1.73-2.40)**
μ (susceptible population)	0	0.034 (0.002-0.133)	0.026 (0.001-0.111)	0.017 (0.001-0.081)	0.080 (0.003-0.279)	**0**
ν (invaded population)	0	0.29 (0.18-0.40)	0.39 (0.27-0.52)	0.41 (0.28-0.54)	0.46 (0.32-0.59)	**0.46 (0.33-0.59)**
ε (spatial interaction)	0	0	1	0.82 (0.73-0.91)	0.82 (0.73-0.90)	**0.81 (0.71-0.90)**
ϕ (infectivity)	0	0	0	0	0.27 (0.13-0.41)	**0.23 (0.11-0.36)**
β (intensity)	0.12 (0.02-0.44)	0.06 (0.01-0.23)	0.36 (0.29-0.44)	0.94 (0.57-1.55)	0.91 (0.42-1.69)	**1.20 (0.72-1.97)**

*When μ is set to 0, this means that cities with large populations have the same risk of being invaded as cities with small populations.

4.5.2 Models fitted jointly to arboviruses

The four model variants selected for the joint analysis included the gravity model and Stouffer's rank model, each with either geographic distance or travel time between cities. The individual and joint models for the four best-fitting model variants are shown in Table 5.11. The difference in the individual models' sum of DIC between the two best-fitting variants was only 2.4. For each model variant considered, the joint model which assumed the same parameters across arboviruses had a higher DIC (i.e. worse fit) than the sum of the individual models' DIC. The fit of the joint models improved when the parameter for transmission intensity was allowed to vary across arboviruses. When both transmission intensity and infectivity parameters were allowed to vary across arboviruses, the DIC values of the joint models were only 1-2 units away from the individual models' summed DIC. Overall, the most parsimonious model with the lowest DIC was the joint gravity model with geographic distance and two parameters for transmission intensity (Table 5.11).

Table 5.11 Comparison of individual versus joint models of CHIKV and ZIKV spread in Colombia. Posterior median and 95% credible interval presented for each parameter. Models were fitted to 337 cities for CHIKV and 287 cities for ZIKV. For the joint models that estimate different parameters across arboviruses, parameters with subscript a refer to CHIKV, while parameters with subscript b refer to ZIKV.

Virus	Model type*	Distance type**	DIC	Sum of DIC	γ (distance power)	μ (susceptible population)***	ν (invaded population)	ε (spatial interaction)	ϕ_a (infectivity)	ϕ_b (infectivity)	β_a (intensity)	β_b (intensity)
CHIKV	G	GD	2329.4	4044.9	1.68 (1.44-1.90)	0	0.65 (0.53-0.76)	0.83 (0.68-0.99)	0.35 (0.25-0.46)		0.24 (0.14-0.39)	
ZIKV	G	GD	1715.5		1.74 (1.50-1.97)	0	0.55 (0.41-0.69)	0.67 (0.50-0.83)	0.27 (0.13-0.40)		1.11 (0.68-1.81)	
Joint	G	GD		4129.0	1.69 (1.54-1.84)	0	0.55 (0.47-0.64)	0.81 (0.71-0.92)	0.12 (0.07-0.16)		0.52 (0.39-0.70)	
Joint	G	GD		4042.6	1.72 (1.56-1.89)	0	0.61 (0.52-0.70)	0.76 (0.65-0.87)	0.32 (0.24-0.40)		0.31 (0.21-0.46)	0.90 (0.65-1.28)
Joint	G	GD		4043.3	1.71 (1.56-1.87)	0	0.60 (0.51-0.69)	0.76 (0.65-0.86)	0.35 (0.25-0.45)	0.27 (0.14-0.40)	0.28 (0.18-0.44)	0.91 (0.66-1.24)
CHIKV	S	GD	2322.9	4047.3		0.48 (0.37-0.58)	1.18 (1.01-1.36)		0.32 (0.24-0.42)		0.009 (0.005-0.015)	
ZIKV	S	GD	1724.4			0.43 (0.31-0.55)	1.37 (1.12-1.63)		0.53 (0.44-0.63)		0.021 (0.013-0.032)	
Joint	S	GD		4143.5		0.43 (0.36-0.51)	1.31 (1.15-1.46)		0.21 (0.17-0.25)		0.022 (0.016-0.028)	
Joint	S	GD		4055.5		0.44 (0.36-0.51)	1.18 (1.03-1.33)		0.41 (0.34-0.49)		0.006 (0.004-0.010)	0.020 (0.014-0.028)
Joint	S	GD		4046.3		0.45 (0.38-0.53)	1.25 (1.09-1.40)		0.31 (0.23-0.40)	0.54 (0.44-0.63)	0.011 (0.006-0.017)	0.017 (0.012-0.024)
CHIKV	S	TT	2325.6	4054.8		0.47 (0.36-0.58)	1.16 (0.99-1.35)		0.31 (0.24-0.40)		0.008 (0.005-0.013)	
ZIKV	S	TT	1729.2			0.42 (0.30-0.54)	1.44 (1.15-1.79)		0.48 (0.38-0.58)		0.023 (0.013-0.037)	
Joint	S	TT		4145.1		0.43 (0.35-0.50)	1.32 (1.16-1.47)		0.20 (0.16-0.24)		0.020 (0.015-0.026)	
Joint	S	TT		4060.7		0.44 (0.36-0.51)	1.18 (1.05-1.33)		0.39 (0.32-0.46)		0.006 (0.004-0.010)	0.019 (0.013-0.026)
Joint	S	TT		4054.8		0.44 (0.37-0.52)	1.23 (1.08-1.39)		0.30 (0.23-0.39)	0.49 (0.39-0.59)	0.010 (0.006-0.015)	0.016 (0.011-0.023)

CHIKV	G	TT	2326.9	4061.3	1.97 (1.69-2.25)	0	0.57 (0.45-0.68)	0.79 (0.70-0.87)	0.39 (0.28-0.52)		0.36 (0.19-0.65)	
ZIKV	G	TT	1734.4		2.06 (1.73-2.40)	0	0.46 (0.33-0.59)	0.81 (0.71-0.90)	0.23 (0.11-0.36)		1.20 (0.72-1.97)	
Joint	G	TT		4141.9	2.02 (1.80-2.21)	0	0.49 (0.40-0.57)	0.84 (0.77-0.90)	0.13 (0.08-0.17)		0.72 (0.52-1.01)	
Joint	G	TT		4059.2	2.02 (1.82-2.24)	0	0.52 (0.44-0.61)	0.80 (0.75-0.86)	0.32 (0.25-0.40)		0.43 (0.30-0.63)	1.20 (0.86-1.67)
Joint	G	TT		4058.8	2.01 (1.76-2.21)	0	0.52 (0.43-0.61)	0.80 (0.74-0.86)	0.38 (0.27-0.49)	0.25 (0.12-0.37)	0.35 (0.20-0.58)	1.22 (0.86-1.75)

*G: gravity (competing destinations), S: Stouffer's rank

**GD: geographic distance, TT: travel time between cities

***When μ is set to 0, this means that cities with large populations have the same risk of being invaded as cities with small populations.

4.6 Validation of model fit and parameter fitting procedure

4.6.1 Gravity models

Model validation was performed for each individual virus' best-fitting gravity model with geographic distance. As geographic distance data were available for all cities, models were fitted to 338 and 288 cities for CHIKV and ZIKV, respectively. The parameter values for these models can be found in Tables 5.12-5.13, and the distance kernels are shown in Figure 5.13. A version of the best-fitting models incorporating per capita infectivity rather than infectivity were also tested, but the DICs were higher (indicating worse fit) for both CHIKV and ZIKV (ΔDIC of 9.5 for CHIKV and ΔDIC of 4 for ZIKV).

Table 5.12 Parameter estimates for eight models of CHIKV in Colombia for 338 cities. The first seven models are variations of the gravity model. Posterior median and 95% credible interval presented for each parameter. Bold indicates the model variant used in the validations.

	Distance-only model	Density-dependent population model	Density-independent population model	Estimated density dependence population model	Full (infectivity) model	Per capita infectivity model with μ set to 0*	**Infectivity model with μ set to 0***	Stouffer's rank model
DIC	2521.0	2448.4	2403.0	2402.6	2336.1	2344.6	**2335.1**	2328.3
γ (distance power)	1.03 (0.83-1.21)	1.02 (0.84-1.17)	1.46 (1.22-1.70)	1.49 (1.25-1.72)	1.68 (1.43-1.90)	1.68 (1.47-1.88)	**1.68 (1.44-1.90)**	
μ (susceptible population)	0	0.076 (0.003-0.286)	0.036 (0.001-0.163)	0.030 (0.001-0.141)	0.061 (0.002-0.207)	0	**0**	0.48 (0.37-0.58)
ν (invaded population)	0	0.45 (0.34-0.55)	0.54 (0.42-0.65)	0.53 (0.41-0.64)	0.64 (0.52-0.76)	0.67 (0.55-0.78)	**0.65 (0.53-0.76)**	1.19 (1.01-1.36)
ε (spatial interaction)	0	0	1	0.86 (0.68-1.05)	0.84 (0.69-1.00)	0.83 (0.68-0.98)	**0.83 (0.69-0.98)**	
ϕ (infectivity)	0	0	0	0	0.34 (0.25-0.46)	0.29 (0.20-0.41)	**0.35 (0.25-0.48)**	0.32 (0.24-0.41)
β (intensity)	0.15 (0.05-0.37)	0.08 (0.03-0.19)	0.26 (0.20-0.31)	0.30 (0.22-0.43)	0.21 (0.11-0.35)	0.45 (0.31-0.68)	**0.24 (0.13-0.39)**	0.009 (0.005-0.015)

*When μ is set to 0, this means that cities with large populations have the same risk of being invaded as cities with small populations.

Table 5.13 Parameter estimates for eight models of ZIKV in Colombia for 288 cities. The first seven models are variations of the gravity model. Posterior median and 95% credible interval presented for each parameter. Bold indicates the model variant used in the validations

	Distance-only model	Density-dependent population model	Density-independent population model	Estimated density dependence population model	Full (infectivity) model	Per capita infectivity model with μ set to 0*	**Infectivity model with μ set to 0***	Stouffer's rank model
DIC	1803.8	1764.4	1748.7	1737.6	1716.2	1719.3	**1715.3**	1724.4
γ (distance power)	1.19 (0.96-1.36)	1.27 (1.08-1.42)	1.55 (1.31-1.78)	1.66 (1.43-1.89)	1.75 (1.49-1.99)	1.71 (1.48-1.92)	**1.74 (1.51-1.96)**	
μ (susceptible population)	0	0.021 (0.001-0.094)	0.027 (0.001-0.113)	0.020 (0.001-0.086)	0.091 (0.005-0.264)	0	**0**	0.43 (0.32-0.55)
ν (invaded population)	0	0.37 (0.26-0.48)	0.46 (0.3- 0.59)	0.47 (0.34-0.60)	0.55 (0.40-0.68)	0.54 (0.41-0.68)	**0.55 (0.41-0.69)**	1.37 (1.15-1.65)
ε (spatial interaction)	0	0	1	0.67 (0.47-0.84)	0.69 (0.52-0.84)	0.64 (0.46-0.81)	**0.68 (0.50-0.84)**	
ϕ (infectivity)	0	0	0	0	0.30 (0.17-0.44)	0.16 (0.07-0.25)	**0.27 (0.13-0.40)**	0.53 (0.43-0.63)
β (intensity)	0.66 (0.22-1.53)	0.65 (0.25-1.35)	0.40 (0.32-0.48)	0.84 (0.52-1.41)	0.86 (0.42-1.52)	1.48 (0.89-2.60)	**1.10 (0.68-1.77)**	0.021 (0.013-0.033)

*When μ is set to 0, this means that cities with large populations have the same risk of being invaded as cities with small populations.

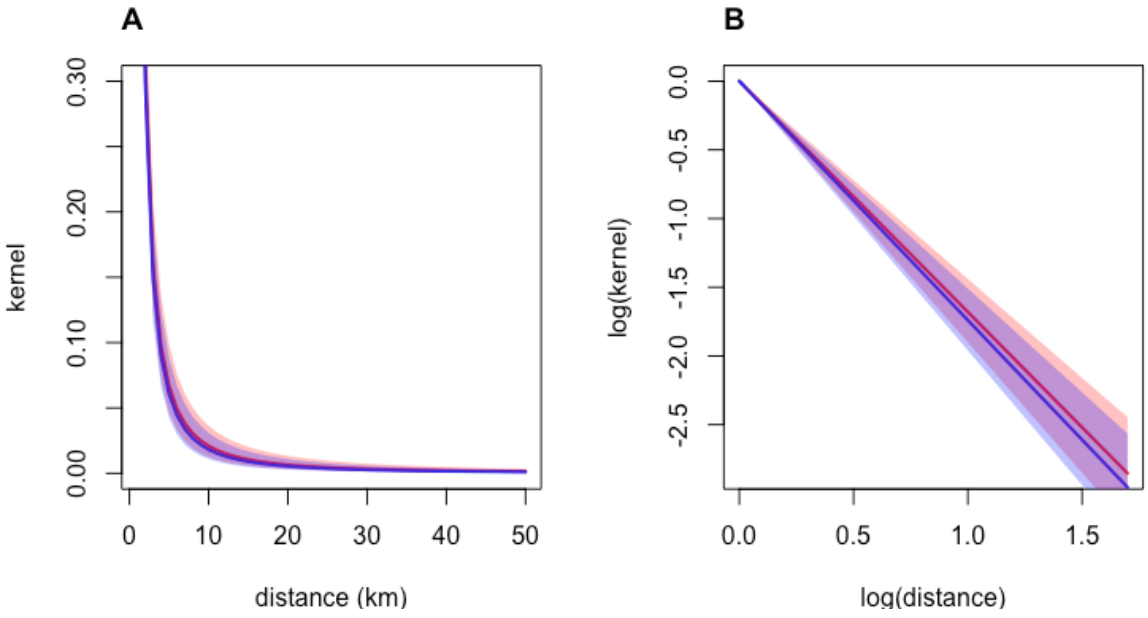


Figure 5.13 Best-fitting distance kernels for CHIKV (red) and ZIKV (blue). The median of the posterior distribution is shown by the darker line, and the 95% credible intervals are represented by the shaded area (purple is where the red and blue shaded areas overlap). *(A)* is unlogged and *(B)* is logged ($\log_{10}$).

For each city, the predicted invasion week given the observed invasion weeks in other cities up to that time was evaluated. The best-fitting model for each virus predicted distribution of the local start of epidemics well (Figure 5.14).

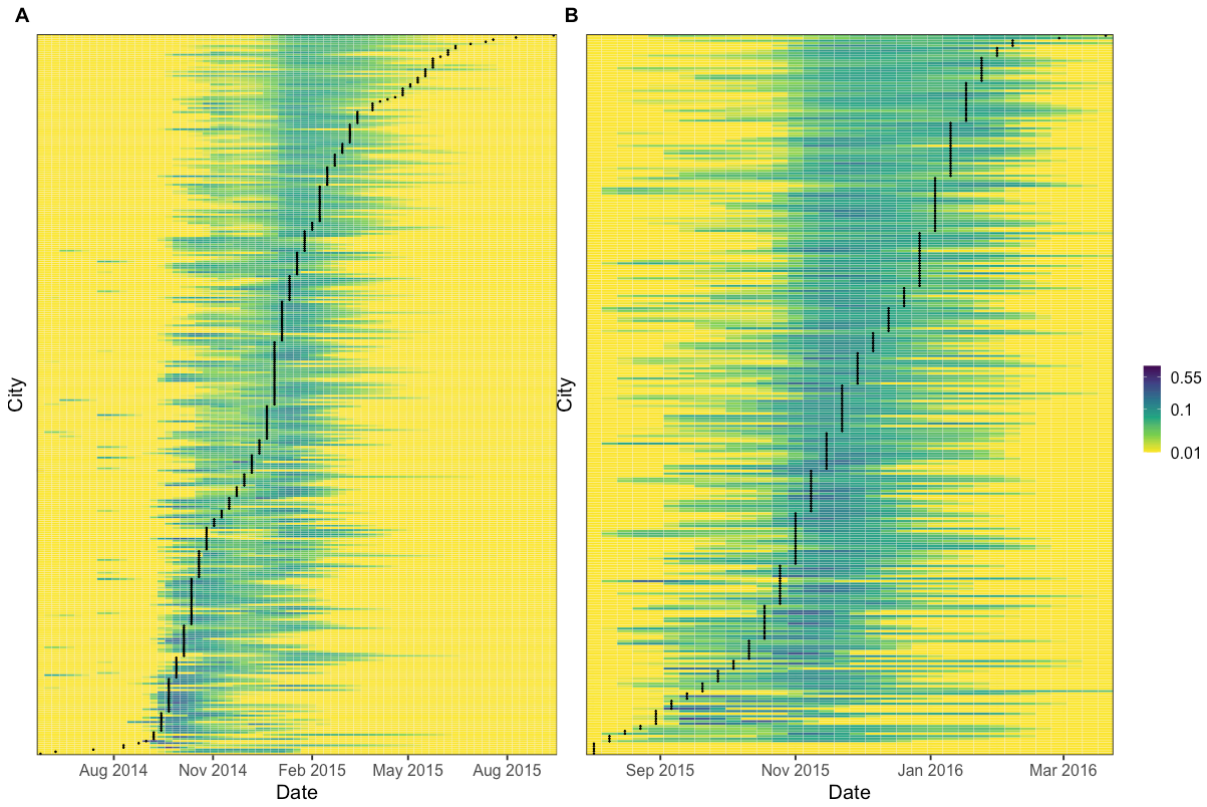


Figure 5.14 Probability distribution of invasion weeks. The above panels show the estimated probability distributions of invasion week for each city (colored lines) for *(A)* CHIKV and *(B)* ZIKV based on the observed start of invasion in other cities up to that time. The calculations were performed using the median parameter estimates from the posterior distributions of the best-fitting models for CHIKV and ZIKV. The black lines show the observed invasion week based on the first reported cases in each city. Values plotted as 0.01 represent probabilities of 0.01 or less.

Excluding cities that were invaded in week 0, 304 out of 337 cities (90% of cities, 95% CI: 87-93%) lie within the 95% interval of their expected distribution for CHIKV, and for ZIKV, 268 out of 283 cities (95% of cities, 95% CI: 91-97%) lie within the 95% interval of their expected distribution. Cities that fell outside of these intervals tended to be invaded at the beginning or the end of the epidemics. Whereas the best-fitting ZIKV model captured the shape of the observed invasion week distribution well, the best-fitting CHIKV model did not capture the shape well at the end of the epidemic. Cities invaded late in the CHIKV epidemic (week 53 or later) had smaller population sizes and fewer cases compared to cities invaded earlier (up to week 52) (Wilcoxon rank sum tests for population size: W = 885, p = 0.004 and for cumulative case numbers: W = 900, p = 0.005). However, these 11 late-invaded cities represent a small proportion of all invaded cities (3%). Figures 5.15-5.16 show the probability distribution of invasion week by department.

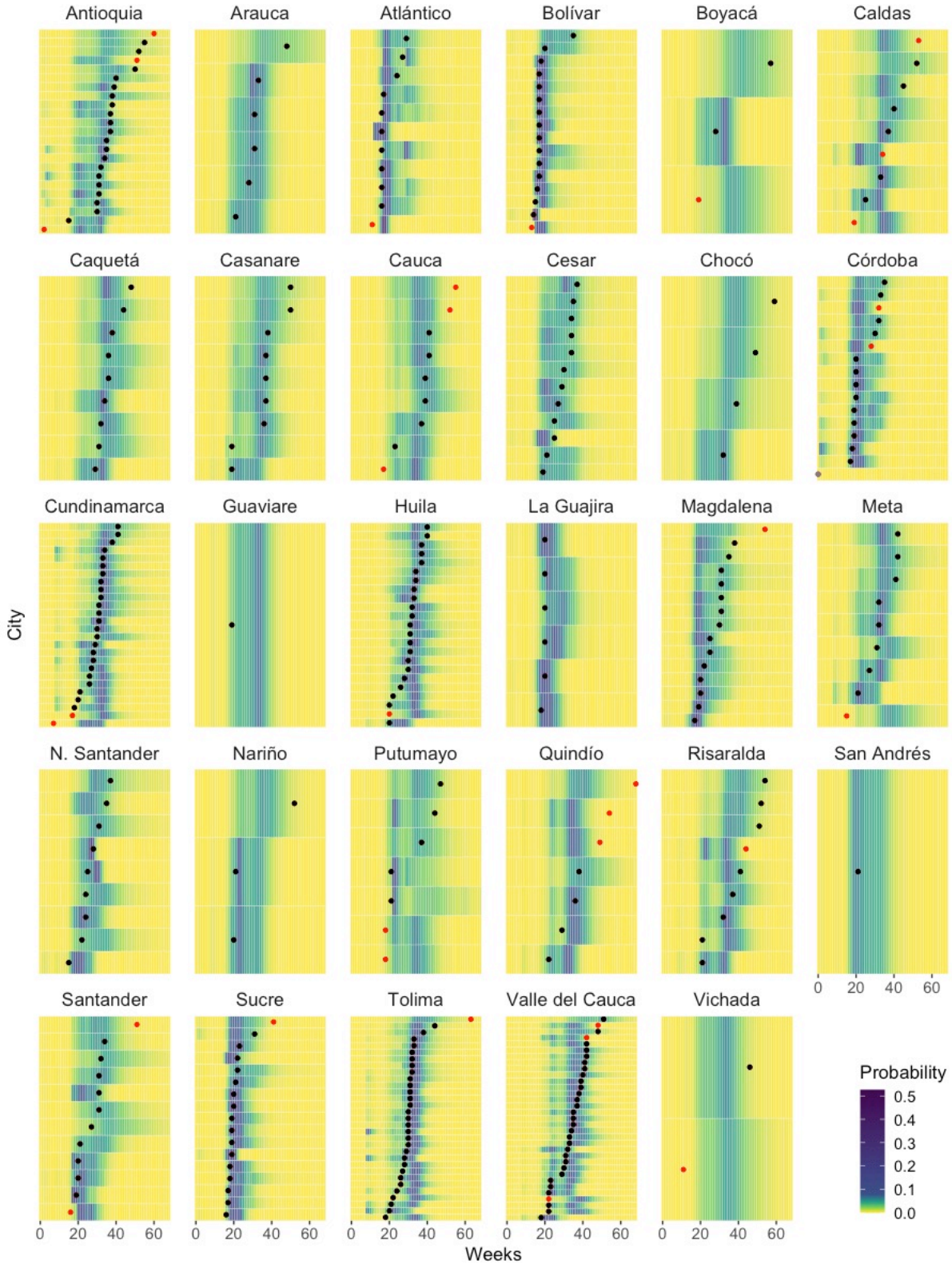


Figure 5.15 Probability distribution of first reported cases by department for CHIKV. Black circles are cities that fall within the 95% interval of their expected distribution, and red circles fall outside this interval. The gray circle in the department of Córdoba represents Planeta Rica, the city that was invaded in week 0.

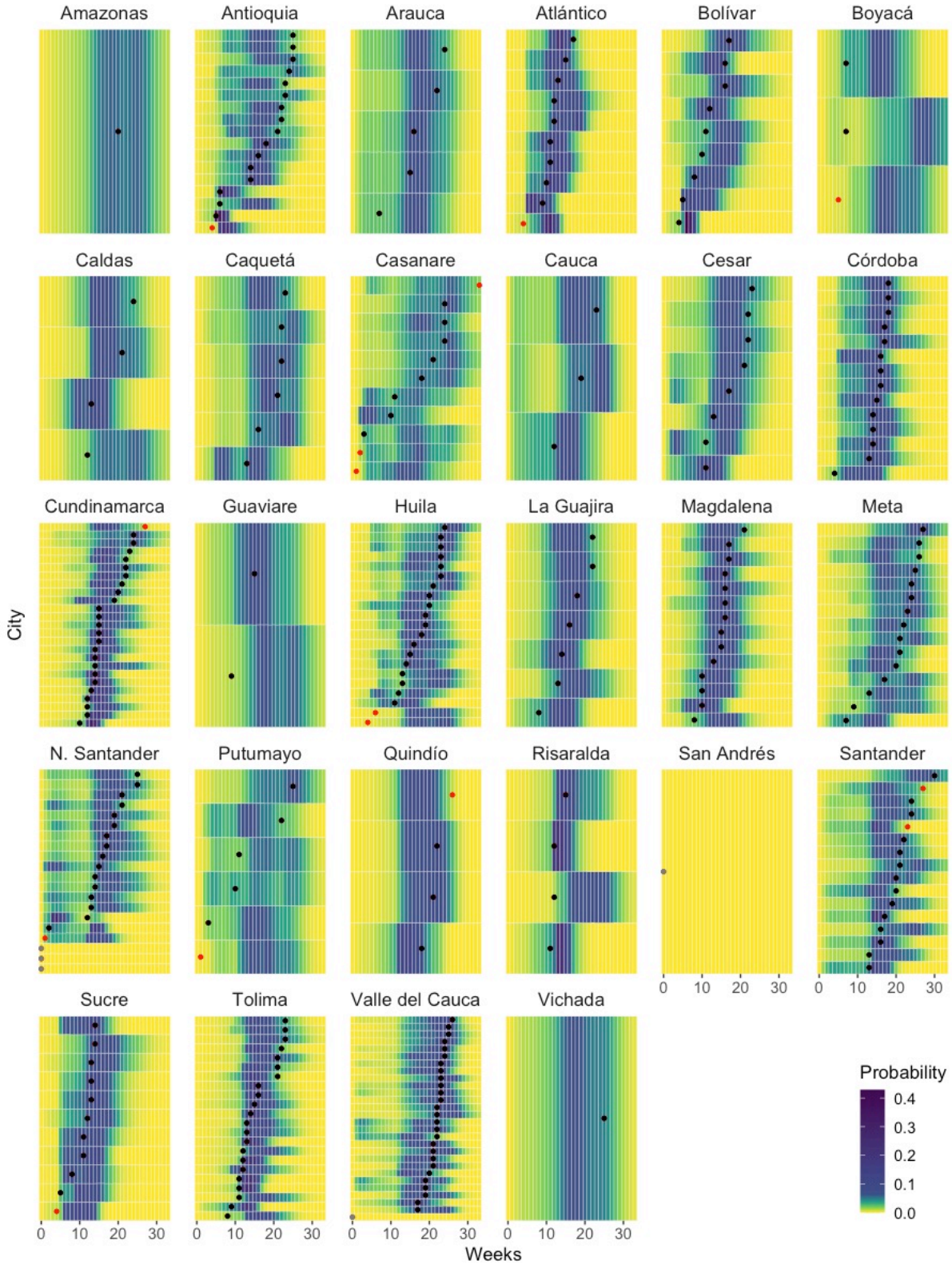


Figure 5.16 Probability distribution of first reported cases by department for ZIKV. Black circles are cities that fall within the 95% interval of their expected distribution, and red circles fall outside this interval. Gray circles in the departments of San Andrés and Providencia, Valle del Cauca, and Norte de Santander represent cities that were invaded in week 0.

The probability distribution for a random sample of cities for the CHIKV dataset is displayed in Figure 5.17. The gray bars are the expected distribution of invasion week and sum to 1. The blue vertical lines demark 95% of this distribution, and the red point below each plot is the observed invasion week. The number above each plot is the city code. The distributions for all cities for both CHIKV and ZIKV can be found in Appendix S5.

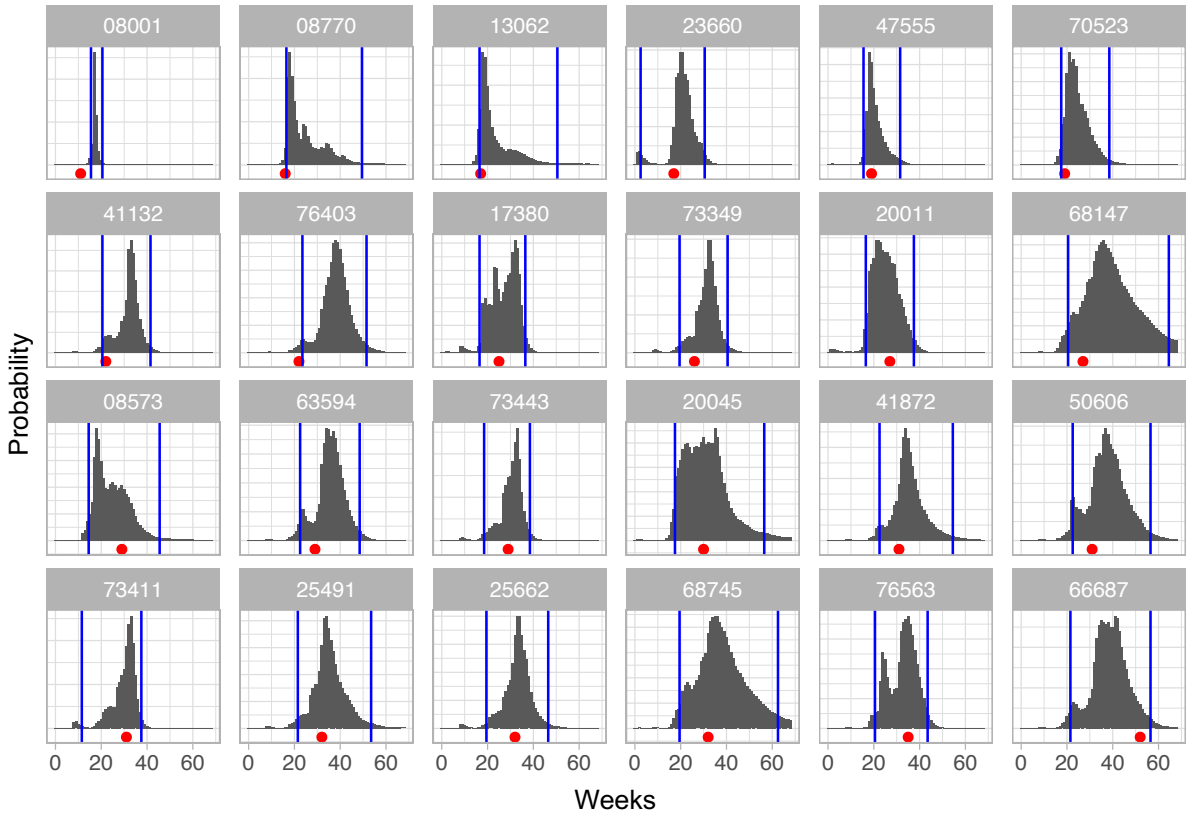


Figure 5.17 Probability distribution of invasion week for a random sample of cities for CHIKV. The probability is shown in gray, 95% of the distribution is within the blue lines, and the observed invasion week is represented by the red points. Y-axes differ between plots. Results for all cities are presented in Appendix S5.

Simulated epidemics from the best-fitting model for each virus were consistent with the observed epidemic in terms of the number of invaded cities over time (Figure 5.18). For CHIKV, half of the cities were invaded by week 31 of the epidemic, while 1,000 simulations predicted half of the cities invaded by week 34.4 on average (min. 17, mode 27, max. 66).

For ZIKV, half of the cities were invaded by week 16, while simulations predicted 15.7 on average (min. 11, mode 15, max. 23).

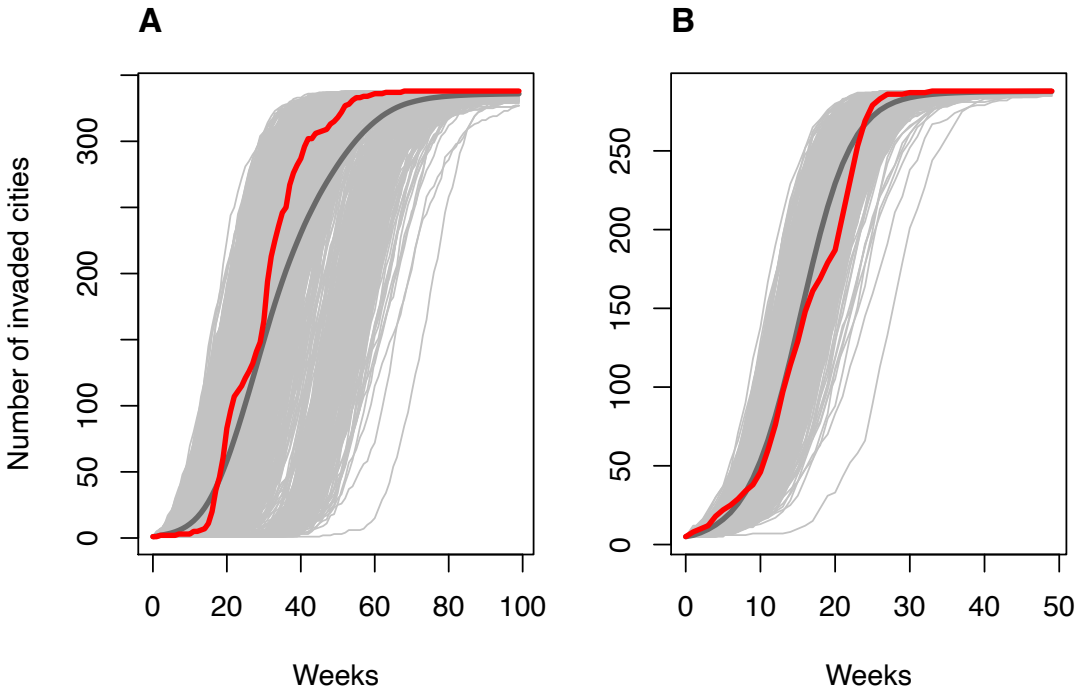


Figure 5.18 Epidemic invasion simulations. Simulated invasion week (as week of first reported cases) for *(A)* CHIKV and *(B)* ZIKV from the best-fitting models. Simulated epidemics are shown in light gray. The dark gray lines are the average across the 1,000 simulations. The red lines show the observed number of cities that first reported cases in each week.

For each virus, fitted parameter estimates were recovered from a model fitted to a single simulated dataset created by simulating the epidemic from the median parameter estimates (Table 5.14).

Table 5.14 Comparison of parameter estimates from observed data versus simulated data.

	CHIKV		ZIKV	
Parameters	**Estimates from observed data**	**Estimates from simulated data**	**Estimates from observed data**	**Estimates from simulated data**
γ (distance power)	1.68 (1.44-1.90)	1.61 (1.40-1.81)	1.74 (1.51-1.96)	1.80 (1.58-2.03)
μ (susceptible population)	0	0	0	0
ν (invaded population)	0.65 (0.53-0.76)	0.76 (0.65-0.87)	0.55 (0.41-0.69)	0.47 (0.35-0.59)
ε (spatial interaction)	0.83 (0.69-0.98)	0.84 (0.68-1.00)	0.68 (0.50-0.84)	0.71 (0.54-0.88)
ϕ (infectivity)	0.35 (0.25-0.48)	0.31 (0.17-0.49)	0.27 (0.13-0.40)	0.29 (0.15-0.42)
β (intensity)	0.24 (0.13-0.39)	0.29 (0.14-0.50)	1.10 (0.68-1.77)	1.09 (0.64-1.79)

4.6.2 Stouffer's rank models

Epidemic simulations were also performed for the best-fitting Stouffer's rank models presented in Tables 5.12-5.13. Even though this model variant had a lower DIC than the best-fitting gravity model for CHIKV, the epidemic simulations were worse overall (Figure 5.19). In contrast, simulations for ZIKV were comparable across model types.

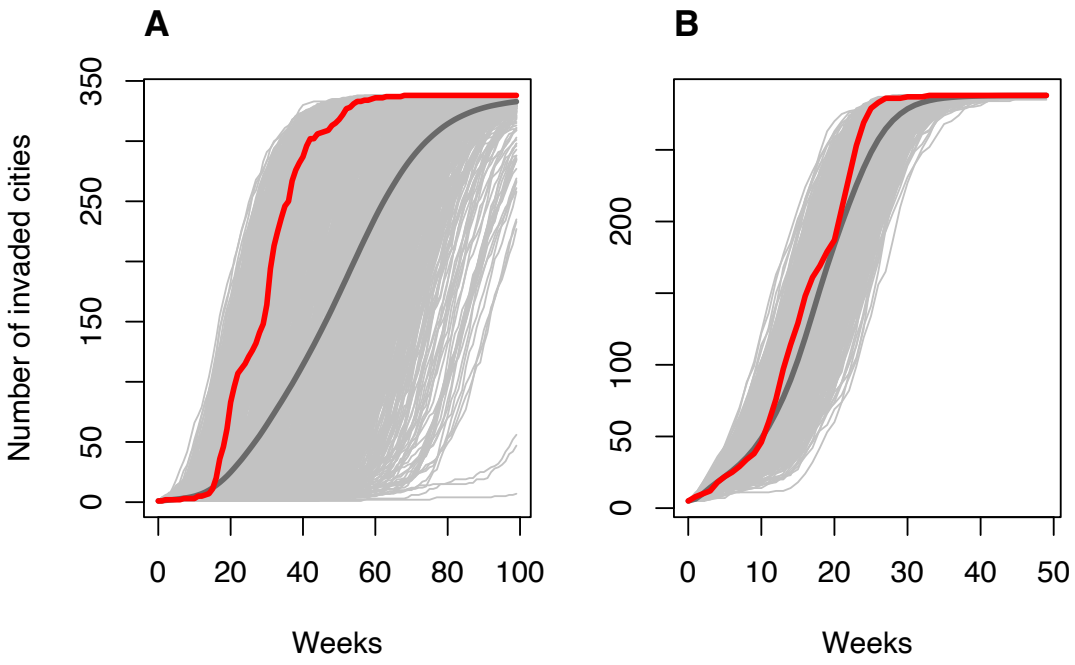


Figure 5.19 Epidemic invasion simulations (best-fitting Stouffer's rank models). Results correspond to the models presented in Tables 5.12-5.13. Simulated invasion for *(A)* CHIKV and *(B)* ZIKV from the models using week of first reported cases. Simulated epidemics are shown in light gray. The dark gray lines are the average across the 1,000 simulations. The red lines are the observed incidence curves.

4.7 Sensitivity analyses on spatial transmission model

4.7.1 Thresholds for cumulative reported cases

Table 5.15 shows the estimated parameters of gravity models fitted to different numbers of cities using thresholds of 10, 20, and 30 cumulative reported cases. In each case, the best-fitting model was the infectivity model with the μ parameter set to 0. All credible intervals overlap, indicating the model results are robust to the choice of threshold. Figure 5.20 shows the results of the corresponding epidemic simulations.

Table 5.15 Comparison of parameter estimates from gravity models fitted to different numbers of cities using thresholds of 10, 20, and 30 cumulative reported cases. In each case, the model variant is the infectivity model with μ set to 0. Columns in bold correspond to main results presented in Tables 5.12-5.13.

	CHIKV			ZIKV		
	Threshold of 30 reported cases	**Threshold of 20 reported cases**	Threshold of 10 reported cases	**Threshold of 30 reported cases**	Threshold of 20 reported cases	Threshold of 10 reported cases
Number of cities	317	**338**	379	**288**	360	485
γ (distance power)	1.67 (1.43-1.90)	**1.68 (1.44-1.90)**	1.64 (1.41-1.85)	**1.74 (1.51-1.96)**	1.85 (1.64-2.08)	1.93 (1.73-2.12)
μ (susceptible population)	0	**0**	0	**0**	0	0
ν (invaded population)	0.64 (0.52-0.75)	**0.65 (0.53-0.76)**	0.57 (0.46-0.67)	**0.55 (0.41-0.69)**	0.60 (0.49-0.72)	0.47 (0.38-0.56)
ε (spatial interaction)	0.86 (0.72-1.02)	**0.83 (0.69-0.98)**	0.75 (0.59-0.89)	**0.68 (0.50-0.84)**	0.70 (0.56-0.83)	0.70 (0.57-0.80)
ϕ (infectivity)	0.32 (0.22-0.42)	**0.35 (0.25-0.48)**	0.35 (0.25-0.49)	**0.27 (0.13-0.40)**	0.32 (0.20-0.44)	0.27 (0.16-0.37)
β (intensity)	0.25 (0.16-0.40)	**0.24 (0.13-0.39)**	0.25 (0.13-0.42)	**1.10 (0.68-1.77)**	1.20 (0.78-1.88)	1.20 (0.83-1.72)

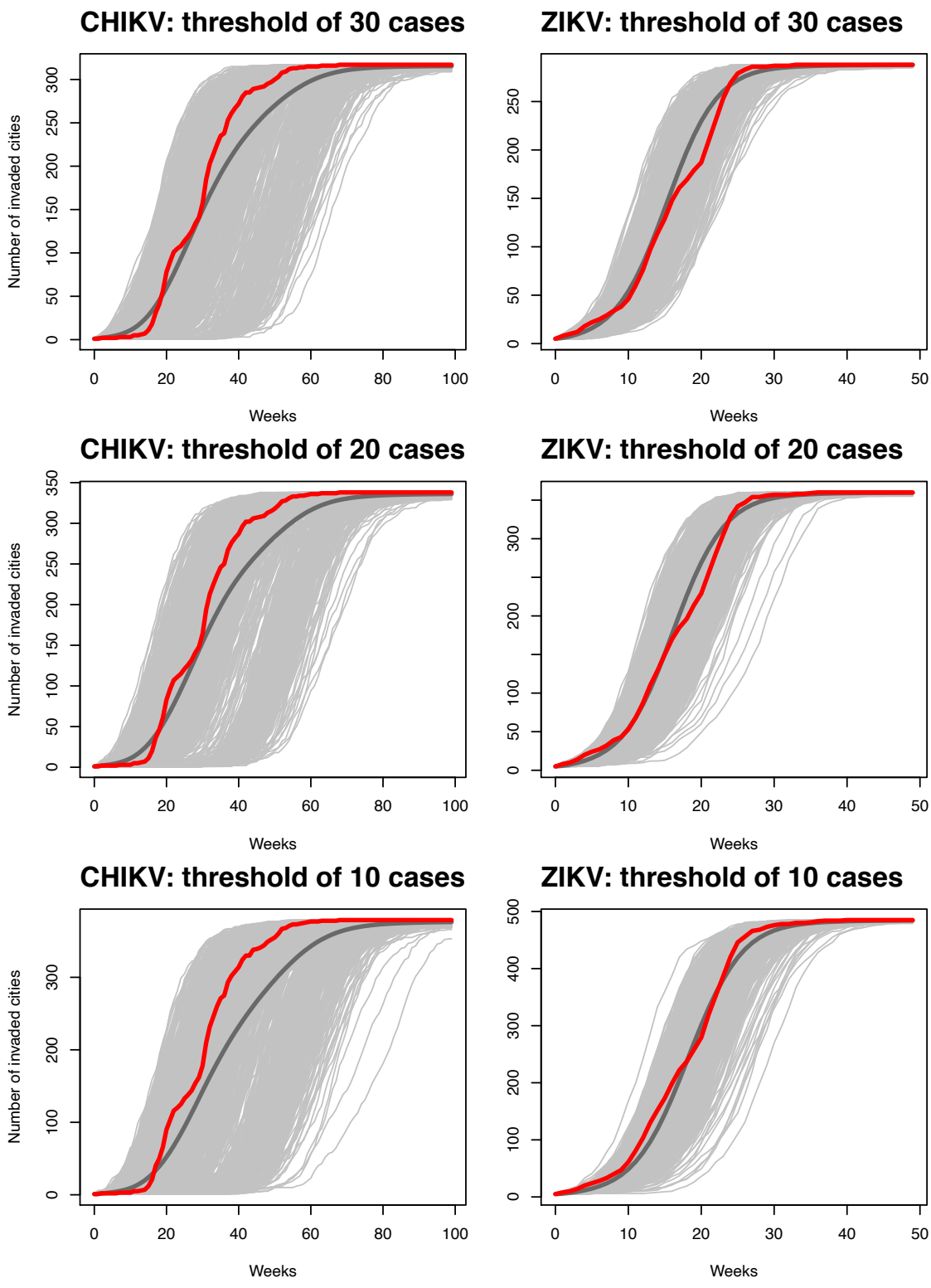


Figure 5.20 Epidemic simulations of the best-fitting gravity models showing the sensitivity of the thresholds used to determine invasion.

4.7.2 Models fitted to all cities

Figure 5.21 shows epidemic simulations using the first reported cases method of determining invasion week and all 1,122 cities in Colombia. The best-fitting model for CHIKV estimated the same five parameters as above, whereas the best-fitting model for ZIKV estimated all six parameters. As expected, the transmission intensity estimate was much lower for both models, and the epidemic simulations showed a very delayed and prolonged epidemic compared to the observed incidence of invaded cities.

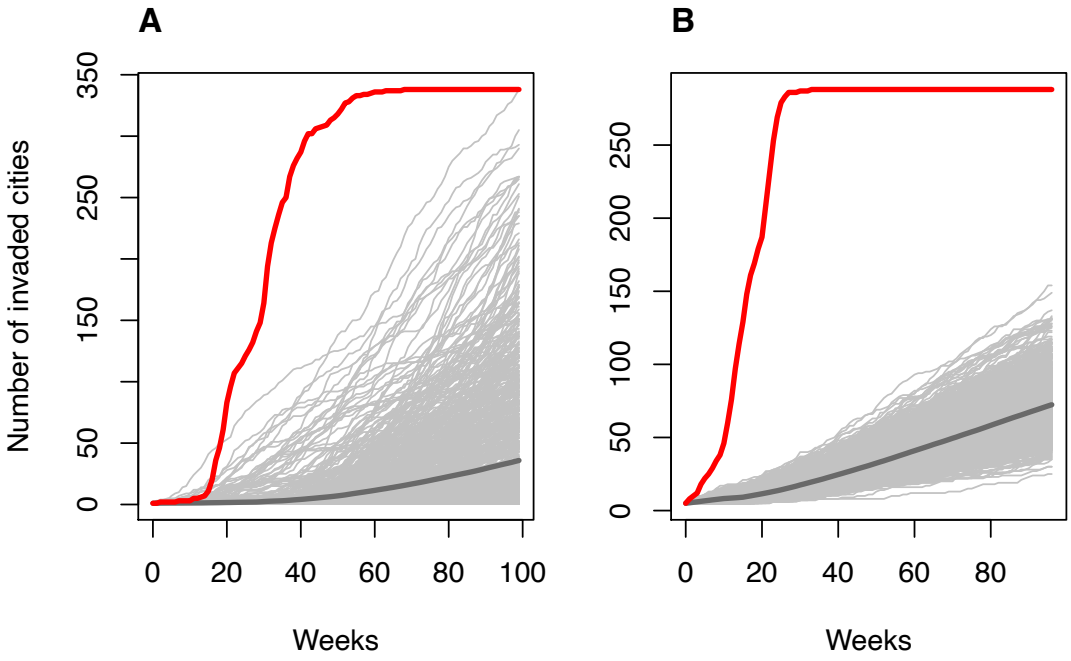


Figure 5.21 Epidemic invasion simulations produced from models with all cities. Simulated invasion for *(A)* CHIKV and *(B)* ZIKV from the models using week of first reported cases and all 1,122 cities in Colombia. Simulated epidemics are shown in light gray. The dark gray lines are the average across the 1,000 simulations. The red lines are the observed incidence curves.

4.7.3 Models fitted to invasion weeks determined from generation time method

Gravity models were also fitted to invasion weeks determined from the generation time method. Figures 5.22-5.23 show the probability distributions of invasion weeks and epidemic simulations, respectively, for these results which used geographic distance. The same number of cities were used as in section 4.6 (338 for CHIKV and 288 for ZIKV). The best-fitting model for CHIKV was the infectivity model (estimating six parameters), whereas the best-fitting ZIKV model was the one that estimated density dependence with the power

for susceptible city population size set to 0 (estimated β, γ, ν, and ε). Parameter estimates (Figures 5.24-5.25) were similar to those presented in Tables 5.12-5.13.

Using these estimates of invasion weeks, the CHIKV model fitted the observed data less well. It is possible that this method for determining invasion weeks is more sensitive to lower reporting at the beginning of the CHIKV epidemic compared to the first reported cases. When the first cases were identified in Colombia in June 2014, CF fever was not yet a notifiable disease. Consequently, some cities reported all of their CF cases retrospectively; one city reported nearly 1,400 cases in a single week. This could have caused the model to underestimate the infection pressure early in the epidemic.

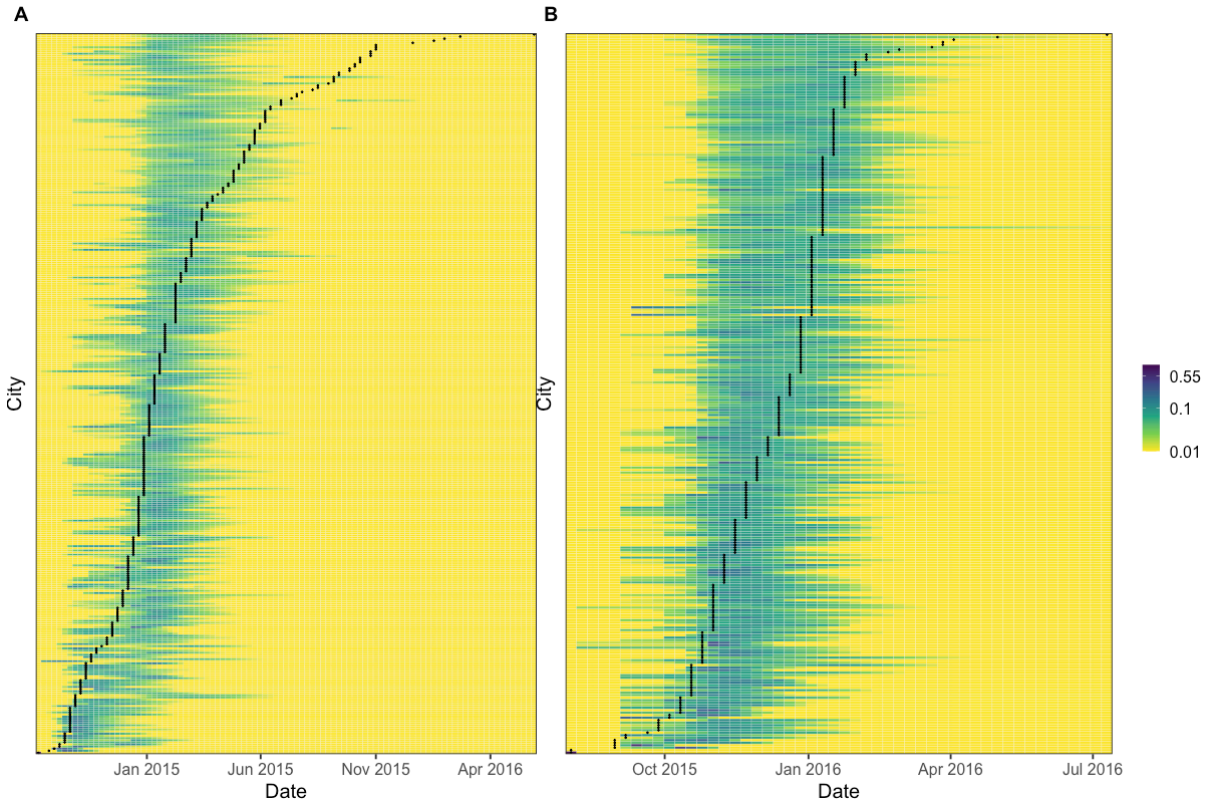


Figure 5.22 Probability distribution of estimated invasion week by generation time method (colored lines). *(A)* CHIKV and *(B)* ZIKV. The calculations were performed using the median parameter estimates from the posterior distributions of the models using estimated invasion week rather than first reported cases. The black lines show the estimated invasion week in each city using a method based on each infection's generation time. Values of 0.01 represent probabilities of 0.01 or less.

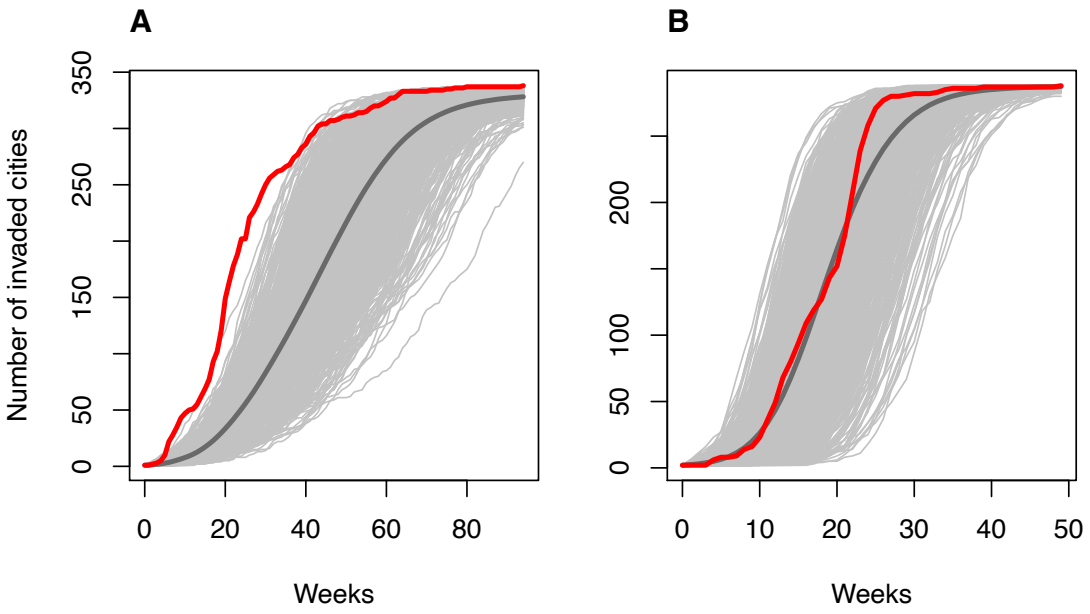


Figure 5.23 Epidemic invasion simulations (generation time method). Simulated invasion for *(A)* CHIKV and *(B)* ZIKV from the models using estimated invasion week by generation time method rather than week of first reported cases. Simulated epidemics are shown in light gray. The dark gray lines are the average across the 1,000 simulations. The red lines are the observed incidence curves.

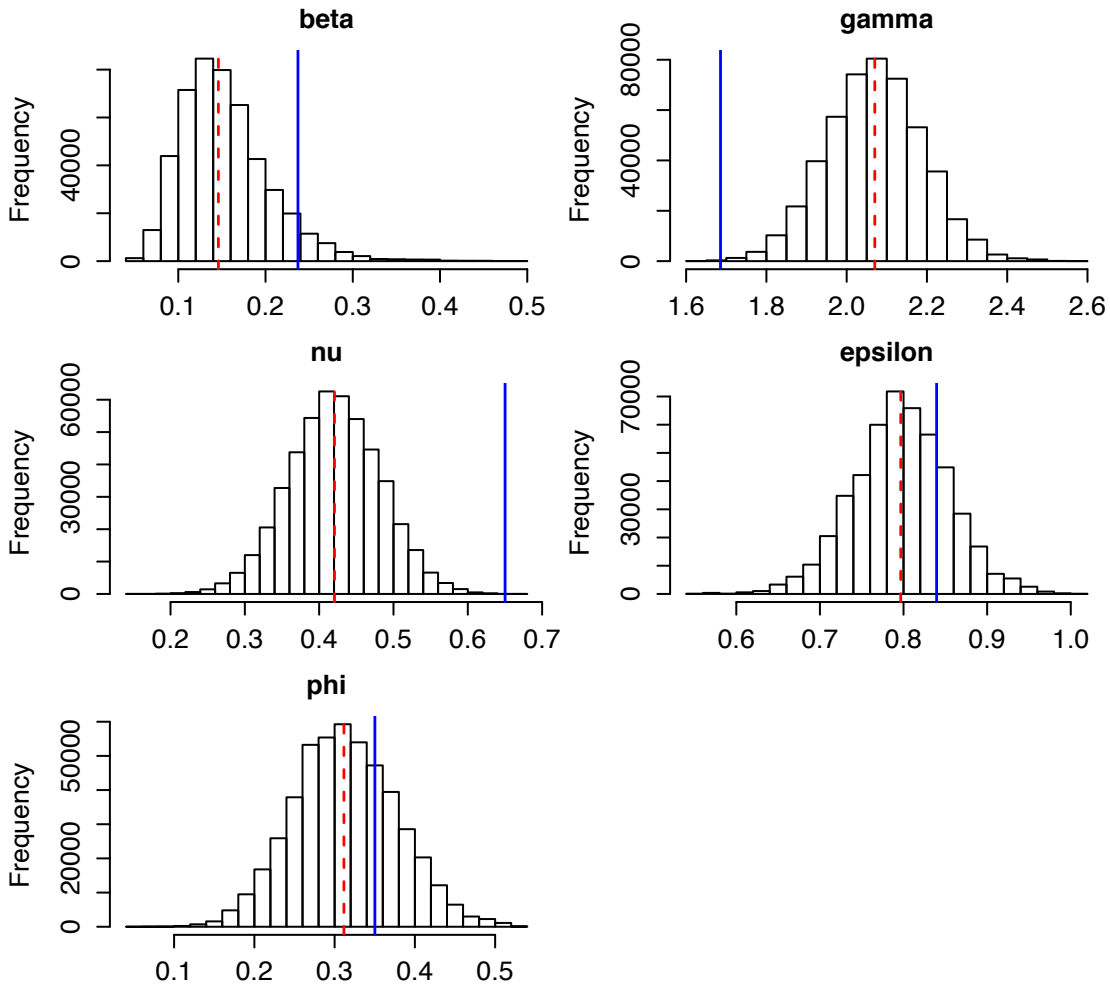


Figure 5.24 Parameter estimates for the CHIKV model fitted using estimated invasion week by generation time method. The dashed red line shows the median of the posterior distribution of each parameter after removing the burn-in period. The blue line shows the median of the posterior distribution from the model fitted using first reported cases as in section 4.6.1. Only parameters that are estimated in both models are shown.

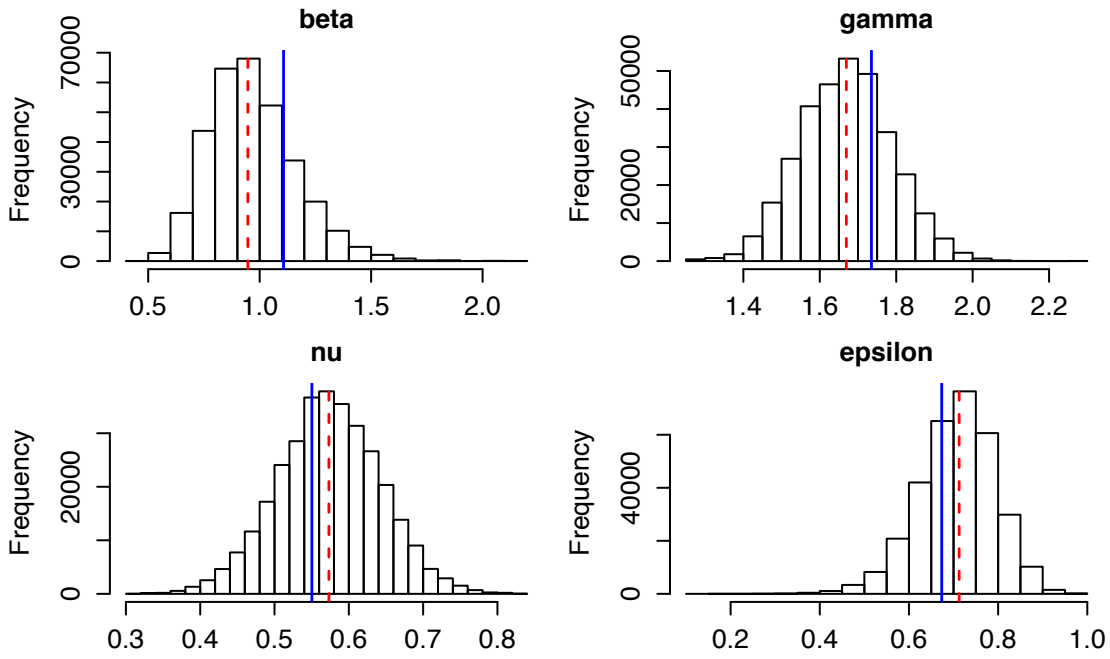


Figure 5.25 Parameter estimates for the ZIKV model fitted using estimated invasion week by generation time method. The dashed red line shows the median of the posterior distribution of each parameter after removing the burn-in period. The blue line shows the median of the posterior distribution from the model fitted using first reported cases as in section 4.6.1. Only parameters that are estimated in both models are shown.

4.7.4 Single-introduction assumption

The single-introduction assumption for CHIKV was relaxed to test the effect on parameter estimates following Eggo et al. [272]. By starting parameter estimation from week 12, five cities were allowed to seed the epidemic rather than one. The parameter estimate of the distance kernel was slightly higher, but the credible intervals largely overlapped, suggesting that this assumption does not greatly affect the model fit (Figure 5.26).

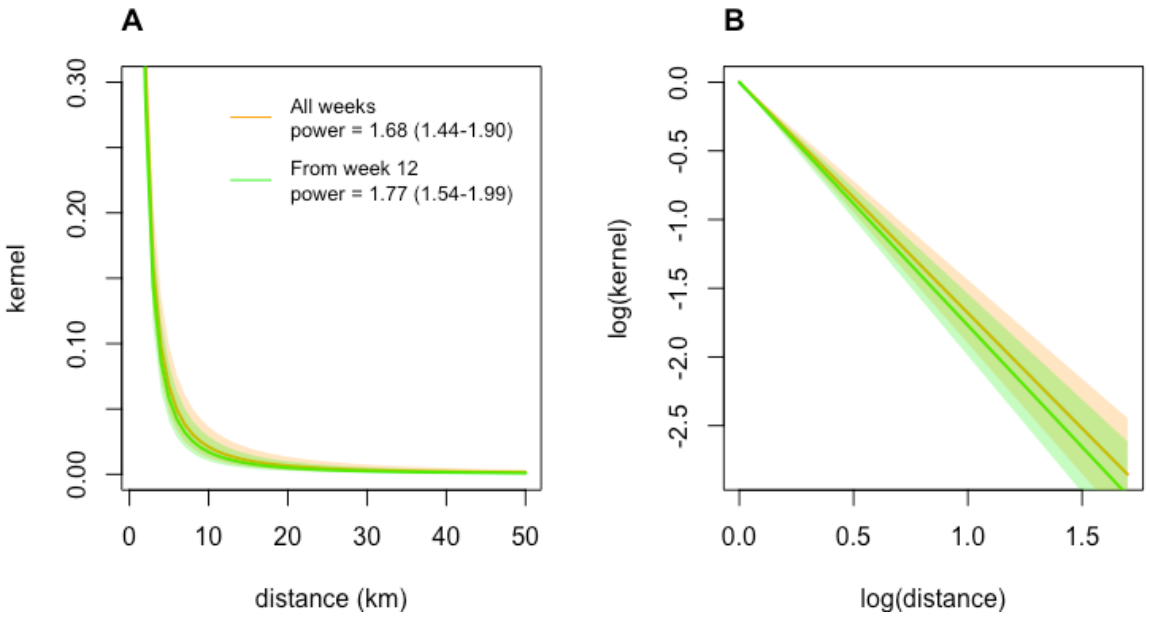


Figure 5.26 Comparison of the distance kernel obtained when running the CHIKV gravity model from week 12 versus the entire dataset. The distance power estimates were similar when parameter estimation started from week 12 (1.77, 95% CrI: 1.54-1.99) compared to week 1 (1.68, 95% CrI: 1.44-1.90).

4.7.5 Sensitivity of outlier in the distribution of travel time between cities for ZIKV

The effect of the outlier in the distribution of travel time between cities for ZIKV was tested by re-fitting the infectivity model with μ set to 0 without the city of Leticia. These models, one with geographic distance and the other with travel time between cities, were each fitted to 286 cities. The model with geographic distance was still preferred over the model with travel time between cities (DIC of 1709.6 versus 1728.1 respectively).

4.8 MCMC testing

The diagnostics in this section correspond to each individual virus' best-fitting gravity model with geographic distance (Tables 5.12-5.13).

4.8.1 Correlation and distributions

Figures 5.27 and 5.28 show the posterior distributions and correlation of parameters for the best-fitting CHIKV and ZIKV models, respectively. Scatterplots of the parameter pairs are on the left part of the figure. Pearson correlation coefficients are displayed on the right, and the posterior distributions are shown on the diagonal. Neither figure appears to show

problems with very high correlation, and the distributions are acceptable.

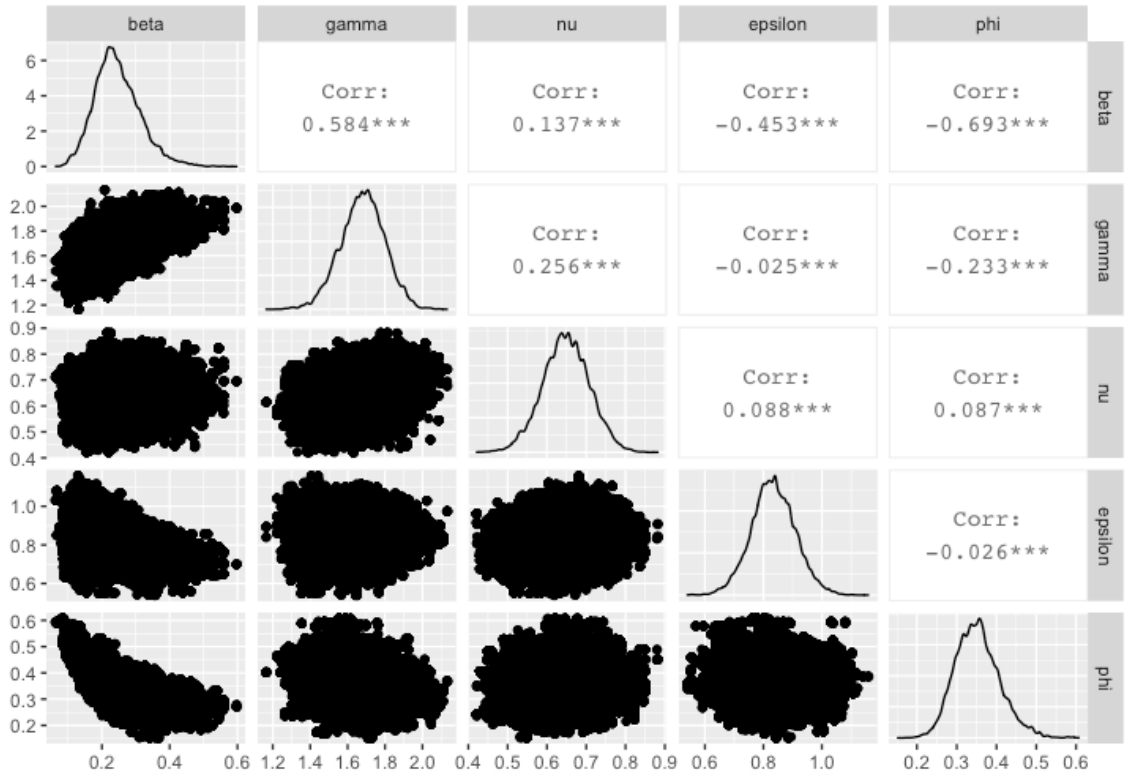


Figure 5.27 Posterior distributions and correlation of parameters for best-fitting CHIKV model. *** indicates statistical significance at the 0.001 level.

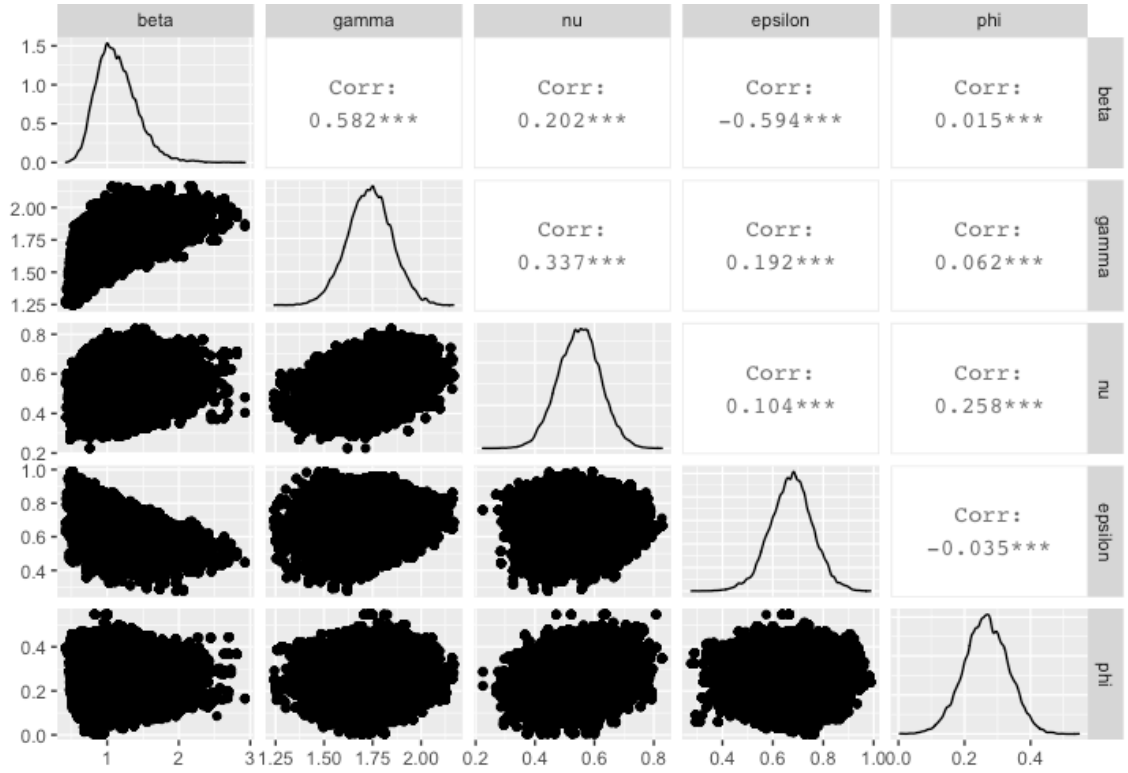


Figure 5.28 Posterior distributions and correlation of parameters for best-fitting ZIKV model. *** indicates statistical significance at the 0.001 level.

4.8.2 Convergence diagnostics

The models were run from three different starting points to ascertain convergence. Table 5.16 shows the Gelman-Rubin statistic for each of the best-fitting gravity models (after removing the burn-in). All point estimates and 95% CI bounds are one or approximately one, suggesting model convergence. Figures 5.29-5.30 show the MCMC traces for each model including the burn-in period. It does not take long for the chains to move toward the same values and overlap with each other. Similarly, Figures 5.31-5.32 show the posterior distributions of the three chains for each parameter after removing the burn-in. The distributions of the parameters are very similar across the three chains. This suggests that the chains converged.

Table 5.16 Gelman-Rubin statistic for each of the best-fitting gravity models (after removing the burn-in).

	CHIKV		ZIKV	
Parameter	**Point estimate**	**Upper CI**	**Point estimate**	**Upper CI**
γ (distance power)	1	1.00	1	1.00
ν (invaded population)	1	1.00	1	1.00
ε (spatial interaction)	1	1.01	1	1.00
ϕ (infectivity)	1	1.00	1	1.00
β (intensity)	1	1.00	1	1.00

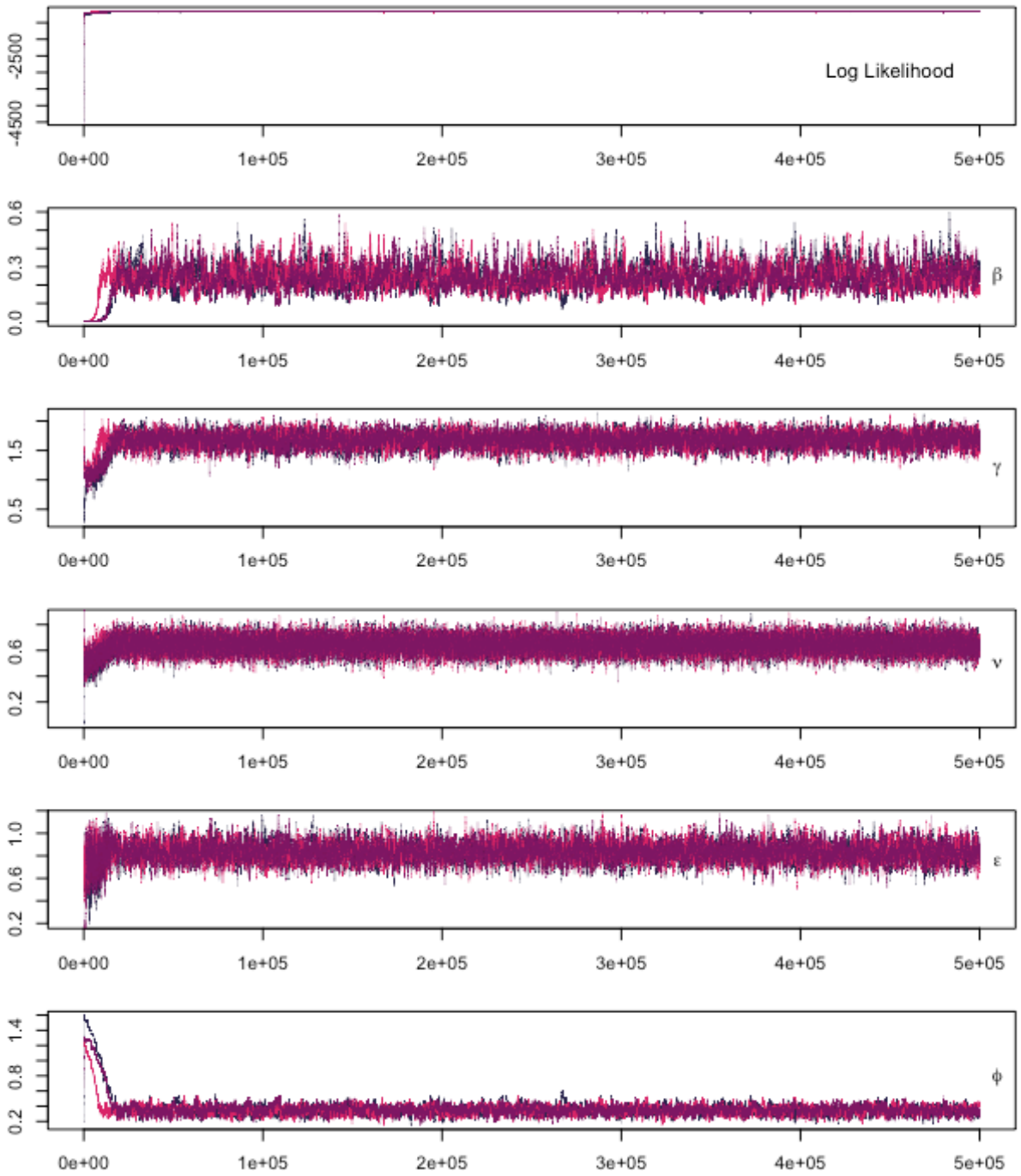


Figure 5.29 Three chains run using different start values for the best-fitting CHIKV gravity model.

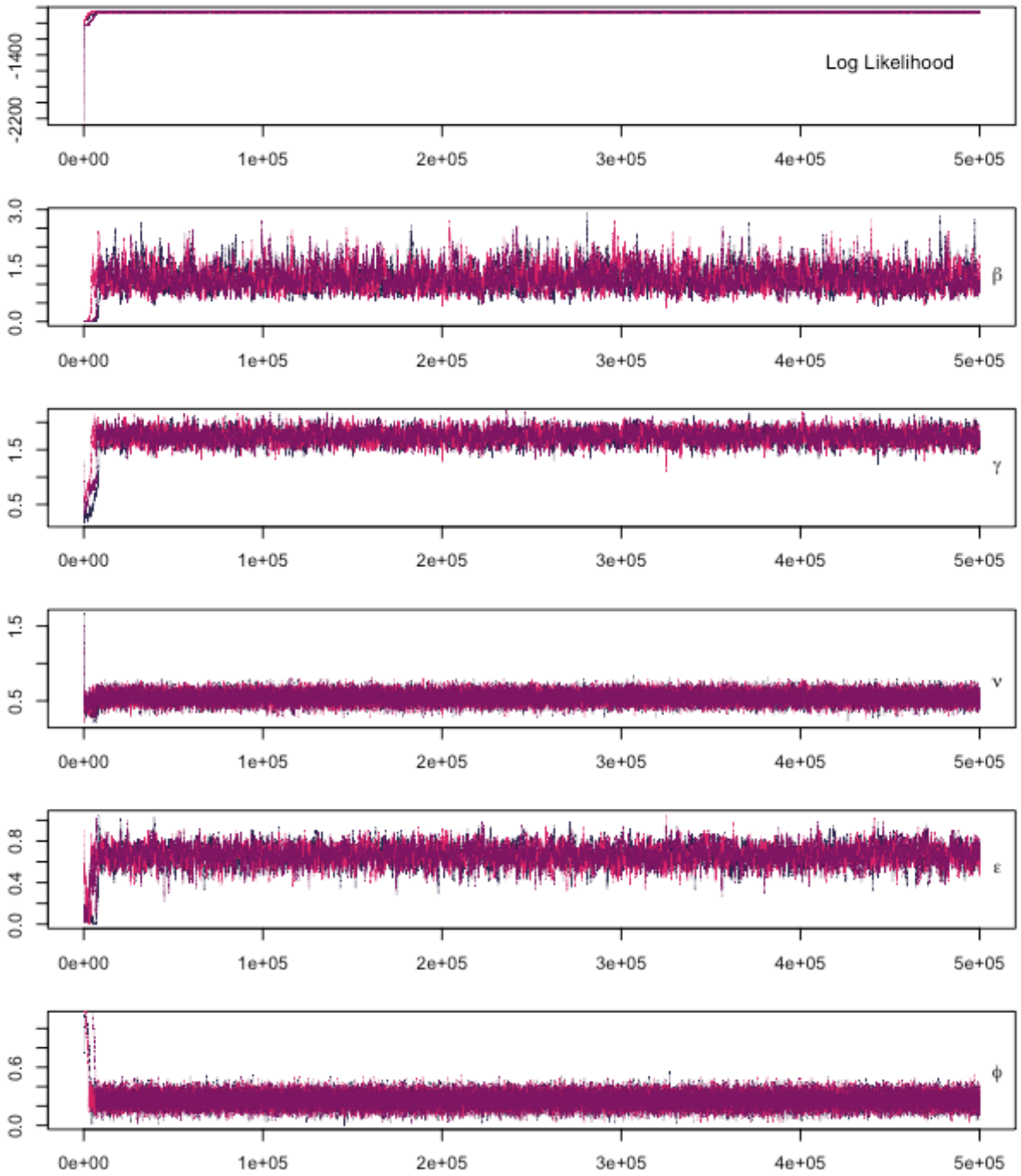


Figure 5.30 Three chains run using different start values for the best-fitting ZIKV gravity model.

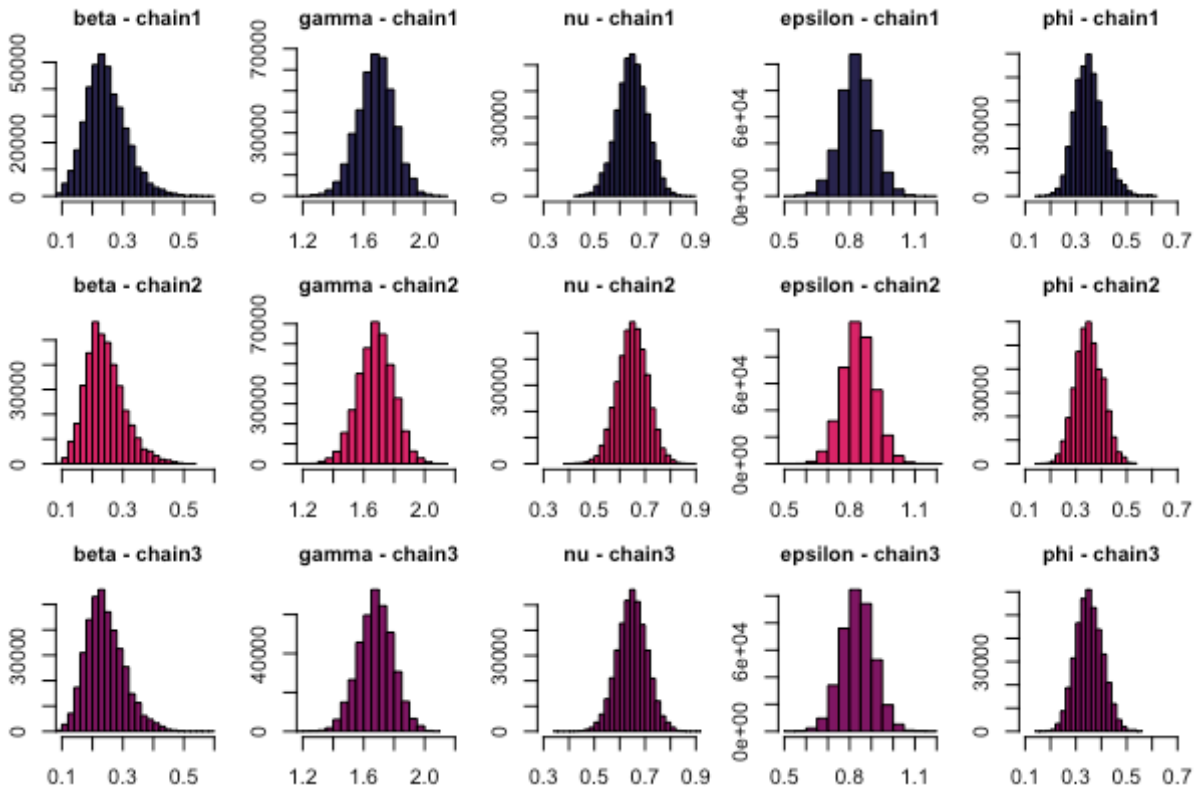


Figure 5.31 Histograms of the posterior distributions of the best-fitting CHIKV gravity model.

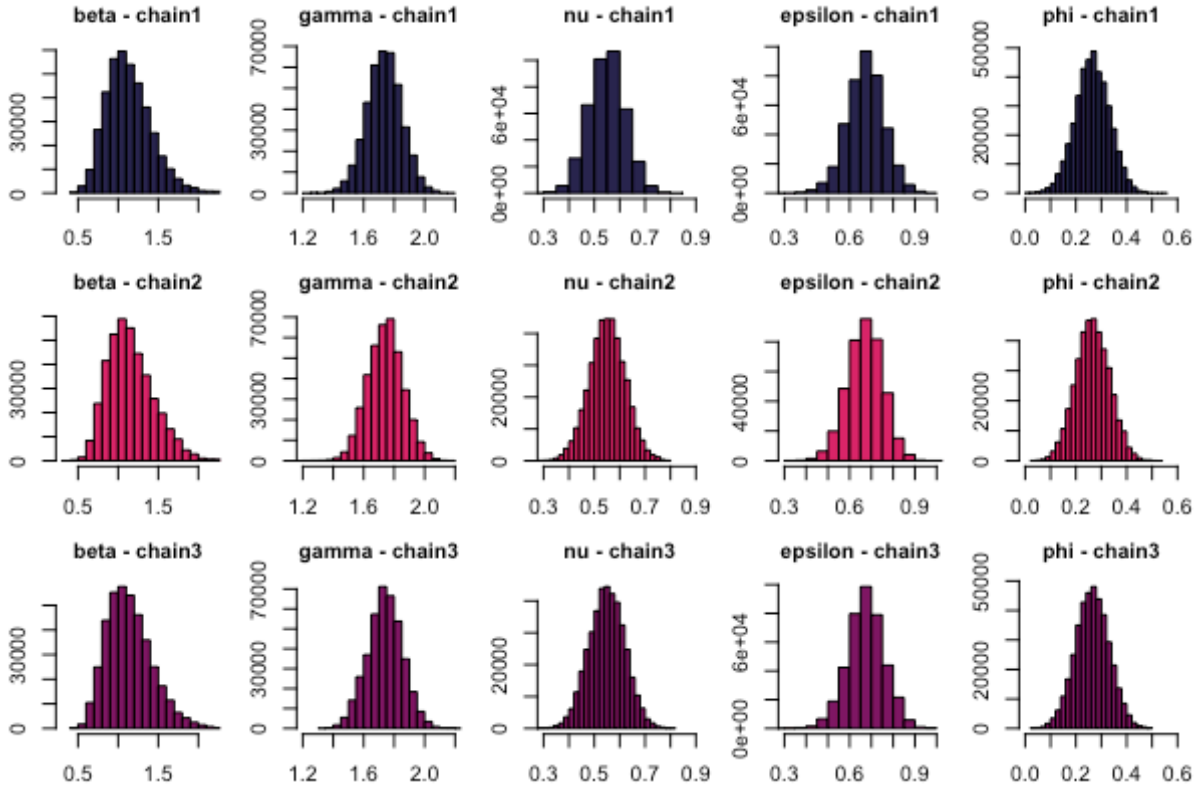


Figure 5.32 Histograms of the posterior distributions of the best-fitting ZIKV gravity model.

4.8.3 Traces

Figures 5.33-5.34 show the MCMC traces for one chain of the CHIKV and ZIKV models, respectively. Mixing is good for all parameters.

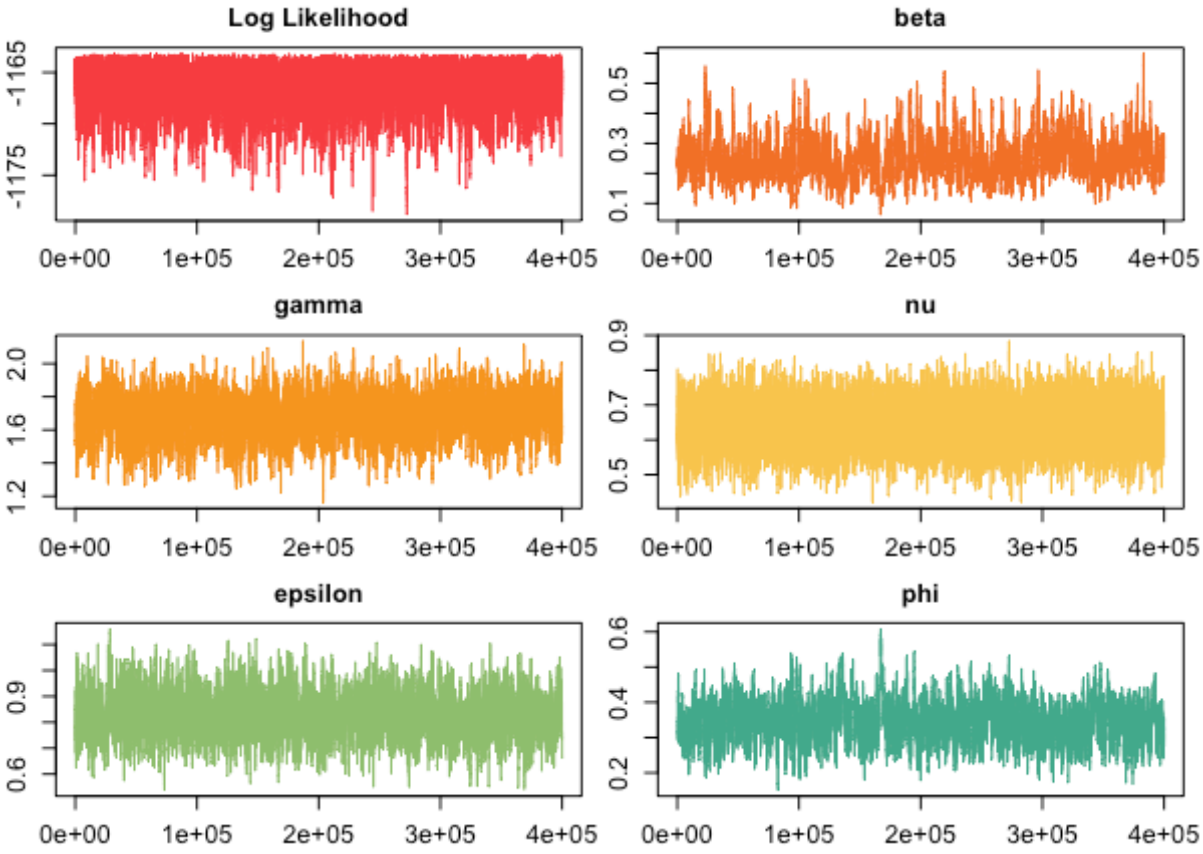


Figure 5.33 MCMC traces for the CHIKV model.

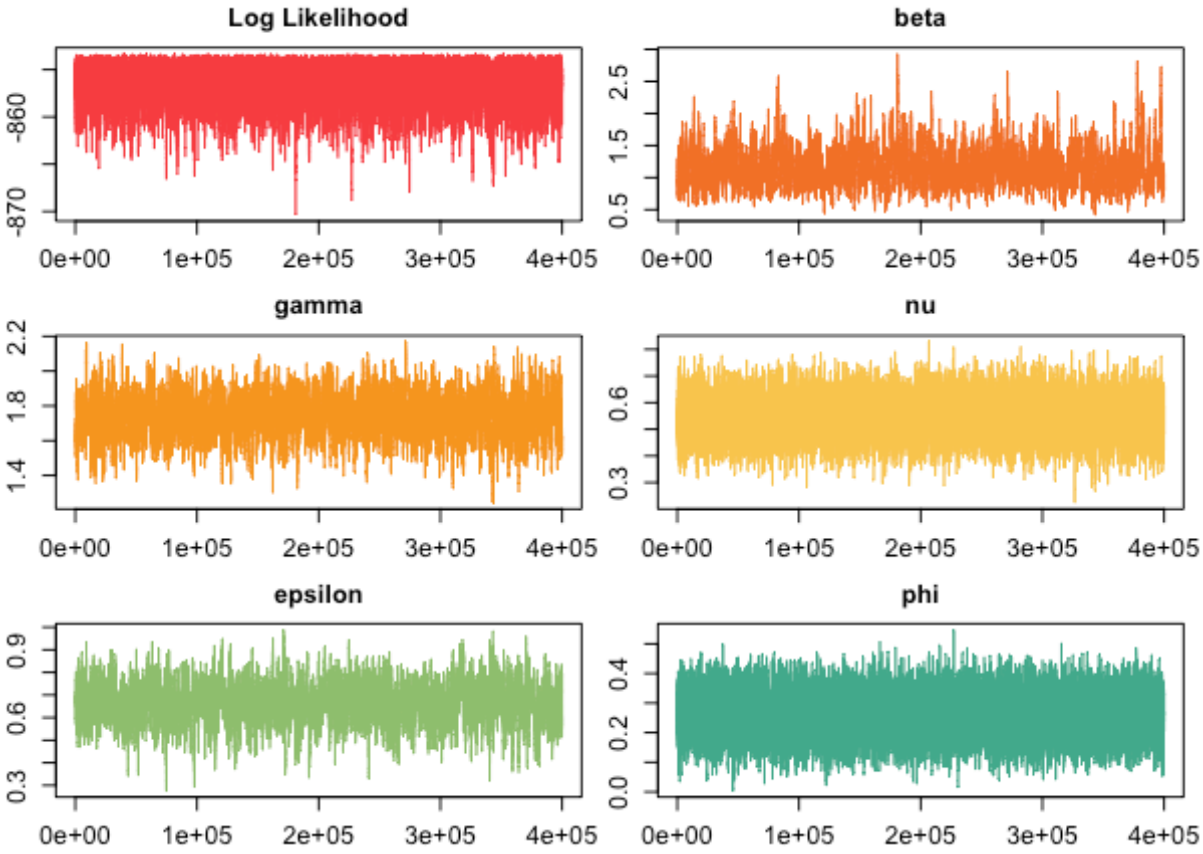


Figure 5.34 MCMC traces for the ZIKV model.

4.8.4 Acceptance rate

Table 5.17 shows the acceptance rate of parameters for the CHIKV and ZIKV models. Both models have good acceptance rates.

Table 5.17 Acceptance percentages for parameters of the best-fitting CHIKV and ZIKV gravity models.

Parameter	CHIKV	ZIKV
γ (distance power)	20.3	25.6
ν (invaded population)	22.4	26.8
ε (spatial interaction)	21.9	22.8
ϕ (infectivity)	19.5	27.3
β (intensity)	22.2	24.1

4.9 Risk factors of invasion

For CHIKV, the following predictors of invasion were significant at the 0.05 level in the

univariate analysis: population size, elevation, mean temperature during the study period, mean temperature up to the epidemic peak, mean rainfall over the study period, mean rainfall up to the epidemic peak, percentage of households with overcrowding, percentage of households with inadequate exterior walls, and risk of dengue (Table 5.18). Except for mean rainfall up to the epidemic peak, the other eight predictors were also significantly associated with invasion by ZIKV. In addition, the percentage of households with inadequate floors was almost significant (p = 0.06) (Table 5.19).

Table 5.18 Univariate analysis of risk factors of CHIKV invasion.

	Invaded cities (N = 338)*		**Uninvaded cities (N = 784)****		
	Mean	**SD**	**Mean**	**SD**	**P value**
Population size, thousands	72	180	31	300	<0.0001
Elevation, m	585	592	1,461	1,045	<0.0001
Mean study period temperature, °C	25.5	3.3	20.5	5.8	<0.0001
Mean temperature up to epidemic peak, °C	25.1	3.3	19.9	5.9	<0.0001
Mean study period rainfall, mm	48.3	27.4	44.4	29.1	0.0003
Mean rainfall up to epidemic peak, mm	49.5	31.6	43.3	31.8	<0.0001
Percentage of households with overcrowding	12.3	7.0	10.7	7.6	<0.0001
Percentage of households with inadequate exterior walls	6.1	7.1	4.8	8.7	<0.0001
Percentage of households with inadequate floors	14.2	15.5	16.3	17.3	0.10
Mean travel time, min.	736	224	935	1,118	0.87
Dengue risk, No. (%)					
0	11 (3.3)		326 (41.6)		<0.0001
1	38 (11.2)		224 (28.6)		
2	105 (31.1)		157 (20.0)		
3	184 (54.4)		77 (9.8)		

*337 cities were included in the row for mean travel time.
**783 cities were included in the row for mean travel time.

Table 5.19 Univariate analysis of risk factors of ZIKV invasion.

	Invaded cities (N = 288)*		Uninvaded cities (N = 834)**		
	Mean	**SD**	**Mean**	**SD**	**P value**
Population size, thousands	89	241	28	278	<0.0001
Elevation, m	577	522	1,411	1,055	<0.0001
Mean study period temperature, °C	25.4	3.2	20.9	5.7	<0.0001
Mean temperature up to epidemic peak, °C	25.7	3.2	21.1	5.9	<0.0001
Mean study period rainfall, mm	50.7	24.2	49.0	26.4	0.04
Mean rainfall up to epidemic peak, mm	40.3	26.4	40.1	26.7	0.36
Percentage of households with overcrowding	12.3	6.8	10.8	7.7	<0.0001
Percentage of households with inadequate exterior walls	5.6	6.0	5.0	8.9	<0.0001
Percentage of households with inadequate floors	13.5	14.5	16.5	17.5	0.06
Mean travel time, min.	769	600	912	1,038	0.86
Dengue risk, No. (%)					
0	5 (1.7)		332 (39.8)		<0.0001
1	21 (7.3)		241 (28.9)		
2	81 (28.1)		181 (21.7)		
3	181 (62.8)		80 (9.6)		

*287 cities were included in the row for mean travel time.
**833 cities were included in the row for mean travel time.

Four variables were included in the best-fitting logistic regression model for CHIKV invasion: mean temperature during the study period, mean rainfall during the study period, dengue risk, and mean travel time. Both rainfall and travel time were protective for invasion. In contrast, temperature and dengue risk were associated with increased odds of invasion. The odds of invasion by CHIKV were 15.5 (95% CI: 7.39-34.84) times higher among cities in the third tertile of dengue risk compared to cities with no risk of dengue adjusting for other variables in the model (Table 5.20). In the model where weather covariates were defined up to the epidemic peak rather than during the study period, the OR for temperature was

about the same (1.25), while the OR for rainfall decreased slightly to 0.86. Four variables were also included in the best-fitting logistic regression model for ZIKV invasion: elevation, mean rainfall during the study period, dengue risk, and the percentage of households with inadequate exterior walls. Elevation, rainfall, and inadequate exterior walls were all protective for invasion, while dengue risk was associated with an increase in the odds of ZIKV invasion. The odds of invasion were 42.3 (95% CI: 16.0-135.4) times higher among cities in the third tertile of dengue risk compared to cities with no risk of dengue adjusting for other variables in the model (Table 5.21). In the model where weather covariates were defined up to the epidemic peak, the OR for rainfall increased slightly to 0.84. There was no evidence of poor fit for either model according to the Hosmer and Lemeshow test (CHIKV: $p = 0.56$, ZIKV: $p = 0.40$).

Table 5.20 Best-fitting logistic regression model of CHIKV invasion. Models were fitted to 1,120 cities because mean travel time between cities was not available for two island cities.

	Estimate	Std. error	Z value	P value	OR (95% CI)
Intercept	-5.75	0.55	-10.52	<0.0001	0.003 (0.001-0.009)
Mean study period temperature, °C	0.22	0.03	7.46	<0.0001	1.24 (1.17-1.32)
Mean study period rainfall, 10 mm	-0.18	0.04	-5.00	<0.0001	0.84 (0.78-0.90)
Dengue tertile 1 (ref 0)	0.58	0.41	1.42	0.16	1.78 (0.82-4.07)
Dengue tertile 2 (ref 0)	1.58	0.40	3.99	<0.0001	4.85 (2.30-10.97)
Dengue tertile 3 (ref 0)	2.74	0.39	6.97	<0.0001	15.50 (7.39-34.84)
Mean travel time, hr	-0.06	0.02	-3.32	0.0009	0.94 (0.91-0.97)

Table 5.21 Best-fitting logistic regression model of ZIKV invasion. Models were fitted to all 1,122 cities in Colombia.

	Estimate	Std. error	Z value	P value	OR (95% CI)
Intercept	-0.96	0.68	-1.42	0.16	0.38 (0.09-1.38)
Elevation, 100 m	-0.11	0.02	-5.58	<0.0001	0.90 (0.86-0.93)
Mean study period rainfall, 10 mm	-0.21	0.04	-4.99	<0.0001	0.81 (0.74-0.88)
Percentage of households with inadequate exterior walls	-0.03	0.01	-2.58	0.01	0.97 (0.94-0.99)
Dengue tertile 1 (ref 0)	0.68	0.57	1.20	0.23	1.98 (0.69-6.61)
Dengue tertile 2 (ref 0)	2.30	0.54	4.27	<0.0001	9.96 (3.73-31.9)
Dengue tertile 3 (ref 0)	3.75	0.54	6.99	<0.0001	42.3 (16.0-135.4)

Results from linear regression models showed that the time to invasion decreased by 3.4 (95% CI: 2.1-4.8) weeks for CHIKV and by 2.3 (95% CI: 1.3-3.4) weeks for ZIKV on average for each one-unit difference in dengue risk level. Figure 5.35 shows the distribution of invasion weeks according to dengue risk level.

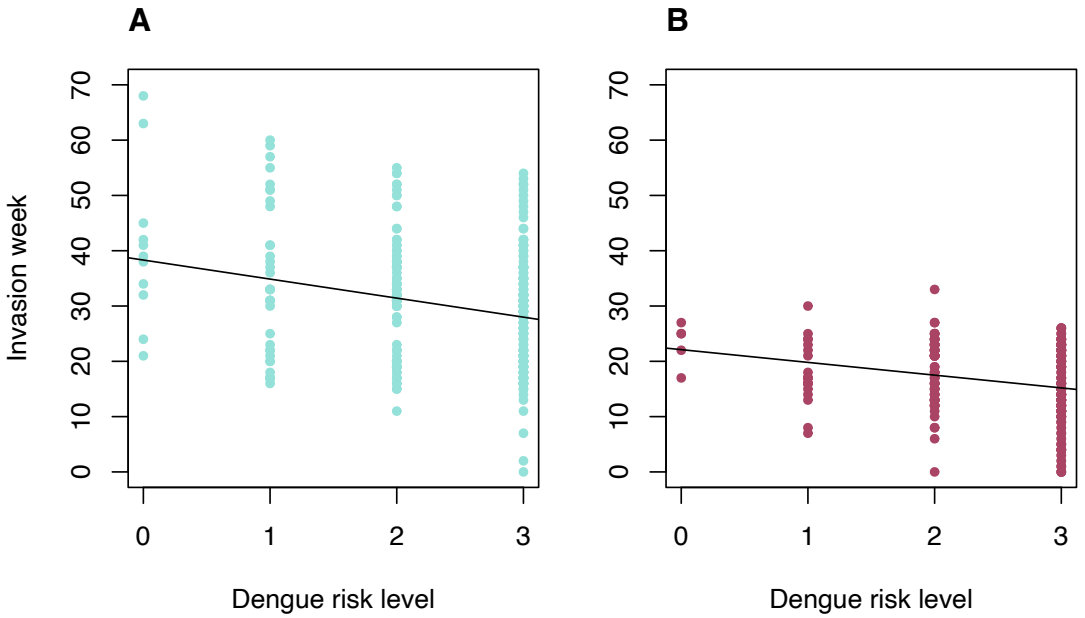


Figure 5.35 Distribution of invasion week by dengue risk level. *(A)* CHIKV and *(B)* ZIKV. The black lines are the fitted linear regression models.

5 Discussion

Similarities and key differences in the spatiotemporal dynamics of the ZIKV and CHIKV epidemics in Colombia were identified using spatial interaction models. Spatial invasion of both epidemics likely began in the north (Caribbean region). From there, the Andes Mountains may have delayed epidemic spread southwards by serving as a natural barrier to human movement. The best-fitting models were different for each virus, and the ZIKV epidemic spread twice as fast as the CHIKV epidemic. Besides having different transmission intensities, parameter estimates obtained for γ, ν, ε, and ϕ, characterizing the effects of distance, population size of the invaded city, density dependence, and infectivity, respectively, for CHIKV and ZIKV were similar and consistent with those obtained in studies of seasonal and pandemic influenza spread [271, 272]. Parameter estimates for the effect of susceptible city population size, μ, and the effect of population sizes of invaded and intervening cities, ν, obtained from Stouffer's rank models were also consistent with those obtained in a study of measles [284]. Cities with high historical DENV transmission had greater odds of being invaded compared to cities with no risk of dengue, and higher levels of dengue risk were associated with decreased time to invasion.

5.1 Comparing alternative spatial interaction models

Across model types, geographic distance was the preferred distance metric to describe spread of ZIKV; in contrast, geographic distance described spread similarly to travel time between cities for CHIKV. Although Viboud et al. found that work commutes better described the spread of seasonal influenza in the USA compared to geographic distance, Charu et al. found that geographic distance outperformed models with work commutes and models with air traffic [271, 274]. Geographic distance was also a better predictor of CHIKV spread in the Caribbean region than air traffic [214].

Using Stouffer's rank models, similar estimates of μ for CHIKV and ZIKV were obtained (0.48, 95% CrI: 0.37-0.58 and 0.43, 95% CrI: 0.32-0.55, respectively). The credible intervals for ν also overlapped (CHIKV: 1.19, 95% CrI: 1.01-1.36, ZIKV: 1.37, 95% CrI: 1.15-1.65). The estimates for ν in this study are similar to those reported by Bjørnstad et al. in their investigation of measles in England and Wales from 1944-1965. However, the estimates of μ here were lower; they reported 1.44 for ν and 0.82 for μ. They also found that Stouffer's rank model performed the best, followed by an extended version of the radiation model and the competing destinations model [284]. Kraemer et al. found that when used together, gravity and radiation models helped explain heterogeneity in the invasion process of Ebola virus in West Africa during the 2014-2016 epidemic [286].

Epidemic simulations of Stouffer's rank models were unable to reproduce the CHIKV epidemic, despite this model variant having a lower DIC than the best-fitting gravity model. Although Stouffer's rank model was better able to capture the beginning of the epidemic, the gravity model performed better in the middle and at the end of the epidemic. A possible reason for this finding is that radiation-type models tend to capture commuting patterns, while gravity models are better suited toward longer-distance movements [286].

5.2 Gravity models

5.2.1 Distance

The estimated power of the effect of geographic distance on spread, γ, was about 1.7, indicating that short-distance interactions were important for transmission. Slightly higher estimates of about 2.0 were obtained for models using travel time between cities. Similar

estimates of γ were expected for CHIKV and ZIKV because they were spread by the same vectors in the same region. Using geolocated genotype and serotype data, Salje et al. found evidence that short-distance interactions were important in transmission of DENV in Bangkok, Thailand with the majority of infection events occurring near the home [306].

A range of estimates of γ can be found in the literature. Gog et al. reported 2.6 (95% CI: 2.3-2.8) for 2009 pandemic influenza in the USA, and Charu et al. reported a median of 2.2 (range 2.1-2.7 with standard deviations between 0.13 and 0.33) across seven influenza seasons in the USA [271, 273]. However, Eggo et al. reported lower values of 1.2 and 0.86 for 1918 pandemic influenza in England and Wales and in the USA, respectively [272]. Differences could be attributed, at least in part, to data being aggregated at different spatial scales; fewer data points in an area will lead to lower estimates of the distance power because locations are farther apart.

5.2.2 Population

Similar values for the estimated power for the effect of invaded city population size, ν, were produced by the best-fitting ZIKV and CHIKV models. This finding suggests that cities with large populations are more likely to spread disease than cities with smaller populations. Gog et al. did not include this parameter, and in Eggo et al., it was not selected in their best-fitting England and Wales model. For both CHIKV and ZIKV, the null hypotheses, stating that the estimated powers (μ) for the effects of susceptible city population size were 0, were accepted. In other words, cities with large populations have the same risk of being invaded as cities with small populations. However, μ did appear to contribute to the fit of the Stouffer's rank models, and therefore the role of susceptible city population size in the spread of the CHIKV and ZIKV epidemics is unclear. Low, but significant, estimates of μ were reported by Gog et al. (0.27, 95% CI: 0.11-0.44) and Eggo et al. (0.40, 95% CrI: 0.25-0.54) for seasonal and pandemic influenza, respectively.

5.2.3 Density

Intermediate levels of density dependence best described transmission (ε for CHIKV 0.83, 95% CrI: 0.69-0.98 and ZIKV 0.68, 95% CrI: 0.50-0.84). This means that the connection between cities somewhat depended on the number and size of neighboring cities. For

influenza in the USA, Gog et al. and Charu et al. reported estimates of ε close to 1 [271, 273]. Eggo et al. reported ε close to 1 for influenza in England and Wales but also found that a density-dependent model ($\varepsilon = 0$) fit the data best for influenza in the USA [272]. Similarly, Salje et al. found that DENV transmission in Bangkok, Thailand was consistent with density-dependent transmission ($\varepsilon=0$) [306]. Differences in estimates could be due to differences in both coverage of datasets and spatial scales considered.

5.2.4 Infectivity

Low, though significant, estimates were obtained for the infectivity parameter, ϕ, from both CHIKV (0.35, 95% CrI: 0.25-0.48) and ZIKV (0.27, 95% CrI: 0.13-0.40) models. This suggests that cities with more reported cases were more infectious than cities with fewer reported cases. Low estimates could be explained by discrepancies between reported case incidence and the true incidence of infection in a city. For instance, if reporting varies over time or by location, surveillance data may not represent actual infection incidence. Eggo et al. reported a similar estimate of 0.24 (95% CrI: 0.03-0.47) for pandemic influenza in England and Wales using mortality rate (representing the fraction of travelers who are infectious) as a proxy for infectiousness [272].

5.2.5 Transmission intensity

Transmission intensity, represented by β, clearly differs between the two viruses. The estimated β for ZIKV is significantly higher than that for CHIKV, reflecting the faster spread of the ZIKV epidemic. Differences in transmission intensity could be related to the 2015-2016 El Niño. El Niño is a weather phenomenon characterized by above average sea temperatures in the equatorial Pacific Ocean. The warm water affects the movement of air and moisture around the globe, and the 2015-2016 El Niño was one of the strongest ever recorded [307]. As discussed in chapter 3, temperature and rainfall affect mosquito development and growth. Caminade et al. studied the impact of El Niño on risk of ZIKV. They found that increased temperatures associated with El Niño created conditions across South America that were favorable for ZIKV transmission in 2015 [307].

5.3 Joint models

Joint models of CHIKV and ZIKV were preferred over models fitted to each virus separately when parameters for transmission intensity and infectivity were allowed to vary across viruses. This finding suggests that some aspects of the spatiotemporal patterns of epidemic arboviruses in Colombia were the same.

5.4 Elevation

Evidence suggests that the distribution of mosquitoes occurs across altitudinal gradients in South America [308]. In Colombia, *Ae. aegypti*, one of the primary vectors of CHIKV and ZIKV, is not usually found at elevations above 2,200 m due to environmental factors, especially temperature [309]. Consequently, cities located at high elevations have less risk of invasion [231]. Elevation was not incorporated into the spatial interaction models in this study. If there were an association between city elevation and week of invasion, there may not have been enough variation in elevation to explain disease spread. The models only included cities that met cut-offs for reported cases, almost all of which are located in low-lying areas (under 2,200 m). Rees et al. estimated the probability of a city reporting at least one ZIKV case during the epidemic in Colombia and dropped mean city elevation from their best-fitting logistic regression model. However, they did find that mean city elevation significantly decreased the time to the first reported ZIKV case using accelerated failure time models, but the effect size was small (the expected time slowed by a factor of 1.18), and they assumed all cities could become infected [296].

5.5 Conclusions and limitations

The results in this study rely on estimates of invasion week in each city. Invasion week was defined as the week before cases were first reported in each city. Although a few reported cases at the beginning of an outbreak may not be enough to sustain chains of transmission resulting in spread to other cities, these cases could signal previously undetected transmission. A genomic epidemiological study found evidence that ZIKV had been circulating undetected in Colombia for five to eight months before the first cases were confirmed in September 2015 [310]. Furthermore, it is possible that cities with better surveillance or healthcare infrastructure could have been the first to report cases in

travelers returning from cities with no prior evidence of transmission because the city of likely infection was modeled rather than city of notification or residence.

The results are robust to uncertainty in invasion weeks. When models were fitted using an alternative definition for invasion week, CHIKV model fits were slightly worse, but ZIKV model fits were very comparable. Importantly, parameter estimates were similar. The results are also robust to the choice of threshold for the number of reported cases. For each virus, gravity model simulations were similar for thresholds of 10, 20, and 30 cases, and the credible intervals of all parameter estimates overlapped.

Cities that did not meet the thresholds for cumulative reported cases were treated as missing in the analysis. Similar approaches have been utilized in the study of seasonal and pandemic influenza [271-273]. For example, Gog et al. included 271 cities for the USA, and Eggo et al. included 47 cities for the USA and 246 cities for England and Wales. Charu et al. included a range from 135 to 306 U.S. cities [271-273]. Here, some unaffected cities were not invaded because they were not at risk (due to environmental factors). Of the cities that were at risk, some were invaded; others appeared to have escaped invasion by chance or other unexamined factors. Among cities that escaped invasion, it is possible that they were invaded but never reported cases. Alternative study designs would be more appropriate for determining why some cities appeared to escape invasion. For example, a mechanistic model of disease transmission accounting for environmental conditions such as temperature and rainfall could be used to ascertain why some cities were invaded and others were not [53]. Also, community-based studies conducted shortly after the epidemic could have assessed whether a city was invaded but did not report cases. In fact, community-based studies were conducted in Colombia following both epidemics but only in cities that reported many cases [189, 190, 193].Community-based studies in rural and economically disadvantaged areas should be prioritized during and in the aftermath of future epidemics to assess surveillance effort and estimate reporting rates.

Ninety-nine percent of CF cases and 95% of ZVD cases were clinically confirmed, rather than laboratory confirmed. This could have led to misclassification, especially considering DENV, CHIKV, and ZIKV were circulating at the same time. Also, asymptomatic infection, mild illnesses, and limited access to healthcare likely led to underreporting. Issues with reporting

and misdiagnoses may have affected the fit of the probability distribution of invasion week for cities invaded near the end of the CHIKV epidemic. Some of these late-invaded cities might have been invaded earlier but failed to report cases to the surveillance system in a timely manner. Another possibility is that cases reported at the end of the CHIKV epidemic were actually misdiagnosed ZVD cases. Oliviera et al. studied the interrelationships between cases of DF, CF, and ZVD in Brazil from 2015-2017. Confirmed cases included all suspected cases reported to the national surveillance system, while discarded cases were defined as suspected cases that met at least one of the following conditions: (i) negative laboratory diagnosis by IgM serology, (ii) laboratory confirmation of another disease, and (iii) clinical and epidemiological compatibility with another disease. Using an autoregressive model, they found that the time series of confirmed and discarded cases of DF significantly affected the time series of confirmed and discarded cases of ZVD, and the other way around. Although confirmed and discarded cases of CF were found to affect the reporting of DF, there was no evidence that the reporting of ZVD or DF affected reporting of CF [311].

Historical dengue transmission in Colombia could have played a role in the spread of CHIKV and ZIKV. It is unlikely that high levels of DENV would have affected susceptibility to CHIKV, an unrelated alphavirus; however, as mentioned in chapter 1, there is some evidence of cross-reactivity among flaviviruses, such as ZIKV and DENV. If pre-existing immunity for DENV increased the risk of symptomatic ZIKV infection, faster recognition of ZIKV in cities that are hyperendemic for DENV would be expected. A cohort study in Managua, Nicaragua found evidence that prior DENV infection was protective for symptomatic ZIKV infection among children (IRR 0.62, 95% CI: 0.44-0.86) adjusting for age, sex, and recent infection with DENV [312]. In contrast, a cohort study in Salvador, Brazil found that individuals with high antibody titers to DENV had less risk of ZIKV infection and symptoms [120]. In this study, high historical levels of DENV in a city were associated with decreased time to invasion for both CHIKV and ZIKV, suggesting that other factors such as environmental suitability of *Aedes* mosquitoes are more important to city invasion than potential impacts of cross-reactive immunity among flaviviruses.

As mentioned in previous chapters, inaccuracy in population projections would also have implications for the results in this chapter, specifically estimates of the population

parameters μ and ν. If more people had migrated to large cities than predicted based on the 2005 Census, then the estimates of μ and ν may have been reduced.

Another limitation is that the model only includes one distance metric at a time. In reality, the spread of ZIKV and CHIKV likely resulted from a combination of air travel, land-based travel, and vector movement. The model also does not account for time-varying changes in reporting, human behavior, or transmission. Nevertheless, these aspects could have changed during the epidemics, especially when the Public Health Emergency of International Concern was declared by the WHO in February 2016 [21].

The model assumes that CHIKV and ZIKV were each introduced into Colombia one time. Two recent genomic studies suggest that this assumption holds which is why background importation rates of CHIKV and ZIKV were not accounted for. Black et al. found evidence of two separate introductions of ZIKV into Colombia; however, the majority of cases were associated with a single introduction [310]. Similarly, Villero-Wolf et al. found evidence of only three introductions of CHIKV in Colombia, suggesting that most cases resulted from transmission within the country, rather than repeated travel-related importations [313]. Also, in the ZIKV dataset used in this study, out of 105,152 cases, 93.2% (97,962) had matching administrative level 2 locations of likely infection and residence, meaning most people were infected where they live. Three-thousand and eighty-six cases (3.3%) resided in Bogotá, and 3,630 cases (3.5%) had different but valid within-country administrative level 2 locations. These cases were likely infected while traveling to other Colombian cities. Only 74 cases (0.07%) had another country listed as their residence, including the USA, Peru, France, Italy, Switzerland, Portugal, and Venezuela. While some of these travelers could have introduced ZIKV into Colombia, the possibility that they were infected while visiting cannot be excluded.

The gravity model formulation used in this study works well retrospectively. Yet more work is needed to understand why some cities appear to escape invasion. Until this issue is resolved, these methods have limited use for real-time forecasting of epidemics.

Future directions for this work include the use of this approach to understand the invasion dynamics of other epidemics. Further research should also focus on quantifying the relative contribution of human versus vector movement on spatial transmission. This would have

broad implications for surveillance and control for other mosquito-borne epidemics such as DENV, MAYV, and yellow fever virus.

Chapter 6: Discussion

The main motivation underlying this thesis has been to compare the CHIKV and ZIKV epidemics in Colombia with a focus on estimating reporting rates (ρs), time-varying reproduction numbers (R_ts), and basic reproduction numbers (R_0s). Additionally, reporting gaps and biases in ZIKV surveillance data were examined and quantified by incorporating seroprevalence data and data on ZIKV-associated neurological complications. Finally, the spatial and temporal dynamics of both CHIKV and ZIKV epidemics were analyzed with several different spatial interaction models.

1 Summary of findings

In chapter 2, epidemiological trends of ZVD and ZIKV-associated neurological complications in Colombia were analyzed using surveillance data. Observed attack rates, risk ratios, and tests for statistical significance were estimated for high-risk groups. Approximately 106,000 suspected and laboratory-confirmed cases of ZVD were reported. High observed attack rates of ZVD were reported in females and young adults. As expected, pregnant females were overrepresented in the data due to increased risk of CZS associated with ZIKV infection during pregnancy. All 32 departments in the country reported at least one case of ZVD. Cases of ZIKV-associated neurological complications were rarer, with only 418 reported cases in 28 departments. GBS was the most common diagnosis among these severe cases.

In chapter 3, ρ and R_0 were estimated for CHIKV and ZIKV from surveillance data. The analysis was conducted at the department level using two approaches based on the renewal equation. Results showed that the estimated ρ for CHIKV was higher than that for ZIKV. However, estimated R_0s were similar across viruses. R_t estimates from parametric models were in good agreement with R_t estimates obtained from the software EpiEstim. Reporting rates estimated from the best-fitting models were also consistent with those observed in a seroprevalence study conducted in four Colombian cities [191].

In chapter 4, ZIKV infection attack rates, reporting rates of ZVD, and the risk of ZIKV-associated neurological complications were estimated for 28 Colombian capital cities using a Bayesian hierarchical model. In addition to the datasets on ZVD and ZIKV-associated neurological complications that were used in previous chapters, published estimates of

post-epidemic seroprevalence [191] were incorporated into the model. ZIKV infection attack rates showed substantial variation across cities. The overall estimated reporting rate for ZVD was similar to that estimated in chapter 3, and the estimated risk of ZIKV-associated neurological complications was low. Important differences in the overall estimated ZVD reporting rates and the risk of ZIKV-associated neurological complications between sex and age group were found, assuming the same ZIKV infection attack rates. Important differences in the hypothesized direction were also found for some cities which tended to have more data.

Finally in chapter 5, the spatial and temporal invasion dynamics of CHIKV and ZIKV were explored using gravity models, Stouffer's rank models, and radiation models. Both geographic distance and travel time between cities were considered. Although invasion risk was best captured by a gravity model which accounted for geographic distance and intermediate levels of density dependence, Stouffer's rank model with geographic distance performed similarly well. Results also showed that short-distance transmission was a main driver of spatial spread following a few long-distance transmission events. Jointly fitted models highlighted similarities between the epidemics. Yet, ZIKV spread faster than CHIKV.

2 Future work and limitations

A major limitation of this work is that control strategies were not incorporated into any of the methods. Public health interventions, including vector control and education campaigns, were implemented during the epidemics in Colombia [314, 315]. For the first seven months of 2016, the MOH recommended that women residing in areas below 2,200 m consider delaying pregnancy, while pregnant women living at higher altitudes were advised to restrict their travel to areas below the cut-off [316, 317]. It would have been interesting to compare the effectiveness of these interventions across the CHIKV and ZIKV epidemics within the same country. For example, in chapter 3, changes in transmissibility of each virus could have been correlated with the timing of specific government recommendations or policy shifts. However, further data would be needed to undertake such an analysis. Once CHIKV and ZIKV vaccines become licensed, the potential impact of these new tools could be evaluated, along with that of novel vector control strategies, such as the use of *Wolbachia*-infected and genetically modified mosquitoes [318].

Another limitation of this thesis is that seroprevalence data were only available for four cities in Colombia. This resulted in large amounts of uncertainty in the estimated infection attack rates in chapter 4; having more data would increase the precision of the estimates. In chapter 3, Valle del Cauca was identified as having the largest burden of CF and ZVD, reporting about one-quarter of all cases. However, relatively low infection attack rates were estimated for both viruses in this department (0.52, 95% CrI: 0.48-0.56 for CHIKV and 0.37, 95% CrI: 0.35-0.40 for ZIKV). In chapter 4, the estimated infection attack rate for ZIKV in Valle del Cauca's capital city of Cali was 0.50 (95% CrI: 0.26-0.95) (a full comparison of the estimated infection attack rates for ZIKV from chapters 3 and 4 at the department and city levels, respectively, can be found in Figure 6.1). Conducting a seroprevalence study in Cali would help understand whether the observed epidemic dynamics in Valle del Cauca were driven by differences in reporting or transmission.

Not all departments reported similar proportions of CF and ZVD cases. For example, Bolívar reported 6% of all CF cases in the country compared to 2% of ZVD cases. On the other hand, only 3% of CF cases were reported in Santander versus 10% of ZVD cases. Relative differences in the burden of these diseases could be related to climactic factors, including the 2015-2016 El Niño. The fact that the vast majority of CF and ZVD cases were clinically confirmed rather than laboratory confirmed complicates the matter and highlights the need for increased testing capacity for arboviruses in Colombia.

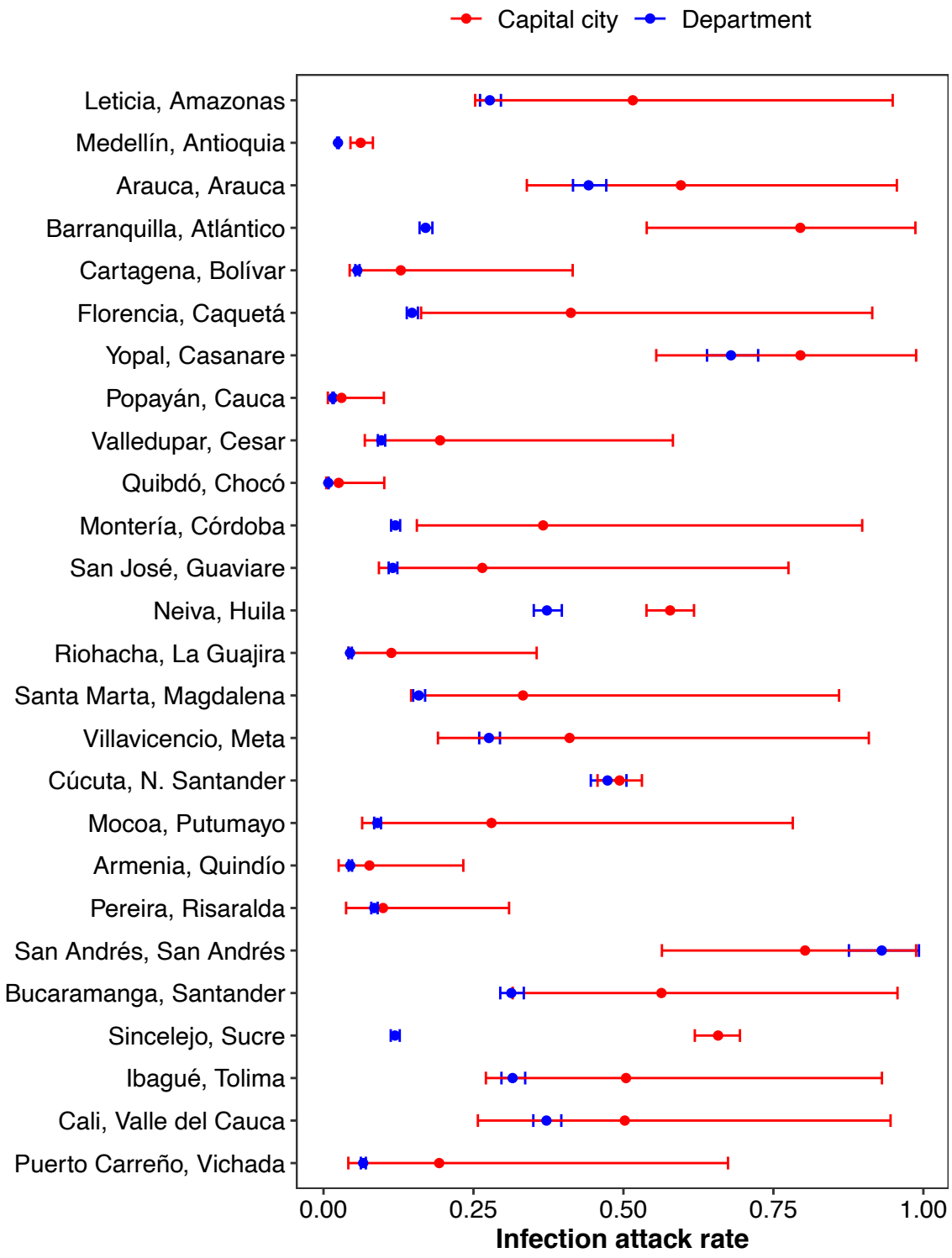


Figure 6.1 Comparison of infection attack rates from chapters 3 and 4 in departments and capital cities, respectively. Mean and 95% CrI are shown for the city estimates, while median and 95% CrI calculated from the estimated reporting rate are shown for the estimates at the department level. Most locations have overlapping credible intervals. Only locations with estimates for both the capital city and department are shown (N = 26).

In chapters 2 and 4, it was not possible to investigate CHIKV-associated neurological complications as line list data were not available. In lieu of a specific dataset, Individual Records of Health Services Provision (RIPS) data could be used to assess trends in neurological disease over time and whether incidence increased in the country during the CHIKV epidemic. The number of patients hospitalized with neurological disease could be compiled for the years prior to, during, and after the CHIKV epidemic. RIPS data can be requested by Colombian citizens through SISPRO [319], an information system managed by the MOH. Statistical tests, such as the Farrington algorithm, could be used to ascertain the significance of any increases in the historical time series. The Farrington algorithm is a regression model that calculates the expected number of cases in a particular week based on the numbers of cases in past weeks [320]. Angelo et al. performed a similar ecological study in Rio de Janeiro, Brazil using data from the national hospitalization database for the years 1997-2017 [321]. They found that the incidence of GBS significantly increased in the city after the introduction of H1N1 influenza in 2009, DENV type 4 in 2013, and ZIKV in 2015-2016. Although not mentioned by the authors, the increase in GBS case incidence in Rio de Janeiro between 2015-2016 also coincided with reports of the first locally-transmitted cases of CF there in late 2015 [322].

In chapter 5, invasion was defined as the week before CF and ZVD cases were first reported in each city. This definition was used as a proxy for the week in which the viruses were first introduced into each city. Although parameter estimates were robust to uncertainty in invasion week using an alternative definition, more accurate timing of disease introduction and subsequent spread could have been achieved with genetic sequence data, especially in locations where the epidemic was not fully observed. Genomic studies have found that ZIKV was circulating in the Americas several months before the first cases were detected and confirmed by PCR [18]. The time between the arrival and discovery of a virus is known as the “surveillance gap” [323]. Unfortunately, few complete genomes of CHIKV and ZIKV are publicly available for Colombia. As of 2019, only 16 complete genomes of CHIKV [313] and 20 of ZIKV [310] had been sequenced from the country. It is unlikely that many more complete genomes from the epidemics will become available in the future due to the degradation of viral RNA resulting from long storage times [310]. Although validating the parameter estimates in chapter 5 with genomic data will likely not be possible, increasing

genomic surveillance could improve detection and response for future epidemics. For example, genomic sequencing is currently being used in numerous countries to track mutations of SARS-CoV-2. As of March 24, 2021, five variants of concern have been identified, some of which are associated with increased transmissibility and more severe illness [324].

Another important area of future research involves studying how the COVID-19 pandemic has impacted arbovirus epidemiology. In 2020-2021, many countries implemented lockdowns, social distancing, and mobility restrictions in response to the pandemic. These measures could have disrupted vector control programs and increased exposure to mosquito bites in the home, while reducing social mixing between households. Surveillance could have been affected through decreases in both healthcare seeking behavior and capacity for laboratory testing [325]. Only a few studies have rigorously examined the impact of lockdowns on arbovirus transmission [326, 327], and the evidence is currently mixed. By late 2020, a higher-than-average incidence rate of DF had been reported in Pakistan, Peru, Singapore, Thailand, and Ecuador. Meanwhile, some countries, including Taiwan, Bhutan, and Sri Lanka, as well as parts of Brazil and Colombia, reported lower-than-average case numbers [325]. The long-term impacts of the COVID-19 pandemic on DENV epidemiology as well as that of other arboviruses are still unknown.

To improve arbovirus surveillance during the COVID-19 pandemic, Colombian health authorities should encourage individuals with arbovirus symptoms to seek healthcare. Patients should be triaged at healthcare facilities so that those with severe symptoms can be treated first and the risk of SARS-CoV-2 infection can be minimized [328]. It is also important to investigate the impact of stay-at-home orders on mosquito control programs throughout the country. One such study, which used a self-administered online survey, was performed in Florida, USA in June 2020 [329]. The potential interruption to mosquito control and surveillance activities in 2020-2021 could increase the risk of future arbovirus threats and epidemics. Another strategy that should be adopted by Colombian health authorities is to extend current seroprevalence studies for SARS-CoV-2 to include arboviruses. The same blood samples that are being used to test individuals for previous infection with SARS-CoV-2 (such as in Montería [330]) could also be tested for a panel of arboviruses, including CHIKV, DENV, and ZIKV.

Future studies could also compare the spatiotemporal dynamics and transmissibility of DENV with those of CHIKV and ZIKV. As mentioned in chapter 3, Funk et al. found that the R_0s for ZIKV and DENV were similar when they were both estimated on the Yap Main Island [233]. Explicitly modeling vector dynamics could be another possible extension to the work presented in this thesis. Compartmental models, such as the SEIR/SEI model (see chapter 3), are often used to capture vector dynamics and could be used to further refine the R_0 estimates in chapter 3. Finally, as mentioned in chapter 1, enzootic transmission of CHIKV and ZIKV occurs in Africa [7, 27]; however, it is unknown whether a similar cycle has been established in the Americas. If these viruses have spilled over into animal reservoirs, epidemics in humans could return to Central and South America sooner than expected due to increased exposure via the animal host(s). Seroprevalence studies of potential hosts should be prioritized.

3 Implications of research

The research presented in this thesis focused on estimating important epidemiological parameters of CHIKV and ZIKV at the subnational level in Colombia. As progress continues toward developing the first vaccines for these viruses as well as deploying novel vector control measures at increasingly larger spatial scales, reliable estimates of these parameters can play an important role in preparedness and resource allocation.

This work has demonstrated that data collected through passive surveillance systems can be used to understand the prevalence and trends of infectious diseases in populations. This information can then be used to better target interventions and control measures. At the same time, the work in chapters 2 and 4 brings attention to potential biases and gaps in these data. If possible, additional data sources should be used to mitigate these issues.

The methods presented here, particularly in chapters 3 and 5, can be used to study other emerging arboviruses in the region, including MAYV. Despite the fact that MAYV case reports and outbreaks have not yet been reported in Colombia [211], a cross-sectional seroprevalence study conducted in 1960 in the department of Santander found seroprevalence of 0.22 among human study participants [331]. A similar study found seroprevalence of 0.19 in individuals living in the Amazonas department in 1966 [332]. By fitting catalytic models to these data, a recent study found that the force-of-infection in

1960 varied across age groups, and in 1966, the force-of-infection was constant across age groups [211]. These results are consistent with epidemic and endemic transmission patterns, respectively; however, they should be interpreted with caution as the sample size in the 1966 study was over two times that of the 1960 study with nearly 400 people. The study also estimated R_0 with mean 1.23 (95% CrI: 1.04-1.66) and 2.10 (95% CrI: 1.71-2.69) for 1960 and 1966 respectively, highlighting the epidemic potential of this virus in the country [211].

4 Challenges

The number of infectious disease outbreaks caused by emerging and re-emerging pathogens has risen over the last few decades, culminating in the current COVID-19 pandemic. The increase has been attributed to climate change, urbanization, global travel, and healthcare worker shortages as mentioned previously in the context of CHIKV and ZIKV. Additionally, increased contact between animals and humans has played a role [333]. Several emerging infectious diseases, such as COVID-19, Ebola, and Nipah, are zoonotic, meaning they originate in animals. CHIKV and ZIKV both circulate in animals and mosquitoes in forested areas in Africa. As humans continue to move into previously uninhabited areas due to population growth or displacement, there will be more opportunities for pathogens to "jump" the species barrier, potentially sparking an epidemic or pandemic [333]. Jones et al.[334] and Allen et al.[335] found that while emerging infectious disease events are mostly detected in developed countries with strong surveillance systems in place, the predicted risk of these events is highest in tropical, developing countries; both research teams identified hotspots in forested areas experiencing land-use changes and where wildlife biodiversity is high [335].

To tackle current and future infectious disease threats, public health surveillance systems should be strengthened around the world. While big data has offered opportunities for high-income countries to integrate novel data streams, such as electronic health data, into traditional surveillance, many low-income countries lack core surveillance capacities [66]. For example, all-cause, excess population mortality has been an important source of data for comparing COVID-19 epidemics across countries; however, low- and middle-income countries may not have sufficient vital registration systems in place to produce real-time

estimates of these data [242]. Improvements are needed in screening and diagnostics as well as laboratory infrastructure. Syndromic surveillance, which uses pre-diagnostic indicator data and statistical algorithms, are less costly compared to diagnosis-based surveillance and could be used to detect outbreaks [336]. Surveillance systems in developing countries could also benefit from the use of a "One Health" approach in which the health of humans, animals, and the environment are considered [337]. Activities that increase the risk of zoonotic disease spillover should be closely monitored. These include mining, logging, and road development [335] as well as bushmeat hunting [338] and certain cultural practices such as harvesting date palm sap [339].

Investment is needed not only in healthcare infrastructure, but also in establishing and strengthening international research collaborations. It is now considered bad practice for researchers in high-income countries to extract data from low- and middle-income countries, analyze the data, and publish the results remotely [340]. This approach is unfair to local researchers and overlooks the valuable knowledge and insights that they can contribute to the field. Though the problem has not been completely eliminated, clear authorship policies from academic journals have helped ensure that collaborators from countries where the research is being conducted are included in publications. For example, the PLOS family of journals has adopted the CRediT (Contributor Roles Taxonomy) system, which consists of 14 categories (including those for "investigation," and "data curation") that describe each authors' contribution to the work [341].

In addition to recognizing the efforts of local teams through co-authorship, capacity building can be used to strengthen research partnerships. For example, researchers in low-resource settings often do not receive formal training in scientific writing [342]. Language barriers pose additional obstacles for non-native English speakers. Some programs, such as the Pre-Publication Support Service, have recently been established to work with authors in low-resource settings to improve their manuscripts for publication in peer-reviewed journals [343]. In terms of technical skills and tools, the R Epidemics Consortium is an example of one organization that has created free analytics tools and training materials for outbreak response, health emergencies, and humanitarian crises using the R software [344]. The R Epidemics Consortium is also actively involved in organizing workshops and short courses on

outbreak analytics and data science, having partnered with the WHO Regional Office for Africa and Colombia's INS in the past [345].

Lack of medical countermeasures to prevent or treat CF and ZVD was a main driver of transmission during the 2014-2017 epidemics in Latin America. As vaccines become available for these diseases, vaccine hesitancy may limit their effectiveness in subsequent outbreaks. Vaccine hesitancy is a phenomenon in which individuals choose to delay or refuse an available vaccine [346]. In 2019, the WHO named vaccine hesitancy as one of the top 10 threats to global health [347]. The issue has been exacerbated during the COVID-19 pandemic; worries about the speed at which the vaccines were developed and tested, potential short- and long-term side effects, and misinformation about the pandemic, especially online, have all been associated with SARS-CoV-2 vaccine hesitancy [346]. In Colombia, vaccine hesitancy has been documented during the pandemic in indigenous groups and university students. Colombia is home to nearly two million indigenous people. Many of them have opted for traditional medicines and isolation to combat COVID-19 citing mistrust of the government and lack of consultation on the vaccine roll-out [348]. Among students at a Colombian university, one study found high levels of SARS-CoV-2 vaccine mistrust (79%). Pursuing a non-health science degree, rural residence, low income, and low pandemic-related perceived stress were all significantly associated with vaccine mistrust in the study [349].

Based on the estimated R_0s in this study, moderate to high vaccination coverage will be needed to achieve herd immunity. Given the challenges associated with vaccine hesitancy, it will be important to closely monitor the impact of vaccination on future outbreaks, particularly for ZIKV. Like congenital rubella syndrome, the risk of severe disease from ZIKV infection is age dependent. As vaccines decrease transmission, the mean age of first infection increases in the population [350]. If R_0 declines but remains above 1, women are more likely to be infected during their childbearing years when the risk of CZS is highest.

In addition to vaccines, increased government transparency and data sharing can also improve the response to future outbreaks. One study uncovered a large unreported outbreak of ZIKV in Cuba in 2017. By integrating surveillance of international travelers, local case reporting, and genomic sequencing of ZIKV from infected travelers, Grubaugh et al.

estimated that 5,707 (interquartile range: 1,071-22,611) ZVD cases were unreported in the country, nearly all of which occurred in 2017 [351]. As discussed in this thesis, many ZVD cases go unreported, and hence the number of ZIKV infections was likely much higher. Prior to this study, PAHO had no record of any cases in Cuba in 2017 or 2018. This hidden outbreak was likely similar in size to known outbreaks that occurred in 2016 in countries with similar population sizes, such as Haiti, Dominican Republic, and Jamaica, and may have been delayed by one year due to a successful vector control program [351]. As updates from PAHO and other international public health organizations are one of the main ways of disseminating information about infectious disease outbreaks, unreported or unrecognized outbreaks have the potential to spread to other countries. Following publication of the paper, a *New York Times* article reported that Cuba did in fact report 1,384 ZVD cases to PAHO in 2017, but due to a technical glitch, the information was not visible on the website [352]. Meanwhile, many travelers to Cuba could have been unwittingly exposed. Four point seven million foreigners visited the island that year during a record year for tourism [352].

5 Conclusions

CHIKV and ZIKV continue to exhibit low levels of transmission in Colombia following the back-to-back epidemics in 2014-2017 [353]. This suggests that after enough susceptible individuals have accumulated in the population due to births and immigration, these viruses have the potential to cause future outbreaks. Hopefully before then, new tools will become available to mitigate the threats posed by these neglected tropical diseases. Having robust estimates of the reporting rates, infection attack rates, and transmissibility associated with CHIKV and ZIKV as well as the patterns of their historical spatiotemporal spread will be important in the design of control strategies. Reporting rates of CF and ZVD could increase during subsequent outbreaks due to increased awareness of the diseases among healthcare workers and the general population. In contrast, infection attack rates and transmissibility would be expected to be lower due to existing population-level immunity. These parameters could also decrease if vaccines and novel vector control interventions are introduced. Climate change may shift the burden of CHIKV and ZIKV outside of the tropics due to rising temperatures which would further reduce future infection attack rates and transmission in Colombia [354]. Given that endemic transmission of CHIKV and ZIKV seems to have been established in the country, the spatiotemporal dynamics of subsequent

outbreaks may not follow the same patterns of spread. However, the introduction of other *Aedes*-borne viruses could follow similar trajectories.

References

1. Pierson T, Diamond M. The continued threat of emerging flaviviruses. Nature Microbiology. 2020;5:796-812.
2. Messina J, Brady O, Golding N, Kraemer M, William Wint G, Ray S, et al. The current and future global distribution and population at risk of dengue. Nature Microbiology. 2019;4:1508-15.
3. Kindhauser MK, Allen T, Frank V, Santhana R, Dye C. Zika: The origin and spread of a mosquito-borne virus. Bulletin of the World Health Organization. 2016;94:675-86.
4. Foy BD, Kobylinski KC, Foy JLC, Blitvich BJ, da Rosa AT, Haddow AD, et al. Probable Non-Vector-borne Transmission of Zika Virus, Colorado, USA. Emerging Infectious Diseases. 2011;17(5):880-2.
5. Musso D, Stramer S, Busch M. Zika virus: a new challenge for blood transfusion. The Lancet. 2016;387(10032):1993-4.
6. U.S. Centers for Disease Control and Prevention. Zika in Infants & Children Atlanta, GA2020 [Available from: https://www.cdc.gov/pregnancy/zika/testing-follow-up/zika-in-infants-children.html.
7. Gregory C, Oduyebo T, Brault A, Brooks J, Chung K, Hills S, et al. Modes of Transmission of Zika Virus. J Infect Dis. 2017;216:S875–S83.
8. Ramírez AL. Mosquitoes. figshare. Collection.; 2018.
9. Paixão ES, Barreto F, Da Glória Teixeira M, Da Conceição N Costa M, Rodrigues LC. History, epidemiology, and clinical manifestations of Zika: A systematic review. American Journal of Public Health. 2016;106(4):606-12.
10. Lessler J, Chaisson LH, Kucirka LM, Bi Q, Grantz K, Salje H, et al. Assessing the global threat from Zika virus. Science. 2016;353(6300):aaf8160-aaf.
11. Fauci A, Morens D. Zika Virus in the Americas — Yet Another Arbovirus Threat. N Engl J Med. 2016;374(7):601-4.
12. Malone RW, Homan J, Callahan MV, Glasspool-Malone J, Damodaran L, Schneider ADB, et al. Zika Virus: Medical Countermeasure Development Challenges. PLoS Neglected Tropical Diseases. 2016;10(3):e0004530.
13. Hodinka R, Kaiser L. Is the Era of Viral Culture Over in the Clinical Microbiology Laboratory? Journal of Clinical Microbiology. 2013;51(1):2-8.
14. Theel E, Hata D. Diagnostic Testing for Zika Virus: a Postoutbreak Update. Journal of Clinical Microbiology. 2018;56(4):e01972-17.
15. Poland GA, Ovsyannikova IG, Kennedy RB. Zika Vaccine Development: Current Status. Mayo Clin Proc. 2019;94:2572-86.
16. Duffy MR, Chen TH, Hancock T, Powers AM, Kool JL, Lanciotti RS. Zika Virus Outbreak on Yap Island, Federated States of Micronesia. N Engl J Med. 2009;360:2536-43.
17. Haddow A, Schuh A, Yasuda C, Kasper M, Heang V, Huy R, et al. Genetic Characterization of Zika Virus Strains: Geographic Expansion of the Asian Lineage. PLoS Neglected Tropical Diseases. 2012;6(2):e1477.
18. Faria RN, Quick J, Morales I, Thézé J, Jesus JG, Giovanetti M, et al. Establishment and Cryptic Transmission of Zika Virus in Brazil and the Americas. Nature. 2017;546(7658):406-10.

19. U.S. Centers for Disease Control and Prevention. Congenital Zika Syndrome & Other Birth Defects Atlanta, GA2020 [Available from: https://www.cdc.gov/pregnancy/zika/testing-follow-up/zika-syndrome-birth-defects.html.
20. Cauchemez S, Besnard M, Bompard P, Dub T, Guillemette-artur P, Eyrolle-guignot D, et al. Association between Zika virus and microcephaly in French Polynesia , 2013 – 15 : a retrospective study. The Lancet. 2016;387:2125-32.
21. WHO Director-General summarizes the outcome of the Emergency Committee regarding clusters of microcephaly and Guillain-Barré syndrome [press release]. 2016.
22. Pan American Health Organization. Zika Cumulative Cases 2018 [Available from: http://www.paho.org/hq/index.php?option=com_content&view=article&id=12390&Itemid=42090&lang=en.
23. O'Reilly KM, Lowe R, Edmunds WJ, Mayaud P, Kucharski A, Eggo RM, et al. Projecting the end of the Zika virus epidemic in Latin America: a modelling analysis. BMC Medicine. 2018;16:180.
24. World Health Organization. Countries and territories with current or previous Zika virus transmission 2019 [Available from: https://www.who.int/emergencies/diseases/zika/countries-with-zika-and-vectors-table.pdf?ua=1.
25. Weaver SC, Lecuit M. Chikungunya Virus and the Global Spread of a Mosquito-Borne Disease. New England Journal of Medicine. 2015;372(13):1231-9.
26. Strauss J, Strauss E. The Alphaviruses: Gene Expression, Replication, and Evolution. Microbiological Reviews. 1994;58(3):491-562.
27. Weaver S, Forrester N. Chikungunya: Evolutionary history and recent epidemic spread. Antiviral Research. 2015;120:32-9.
28. Althouse BM, Guerbois M, Cummings DAT, Diop O, Faye O, Faye A, et al. Role of monkeys in the sylvatic cycle of chikungunya virus in Senegal. Nat Commun. 2018;9:1046.
29. Vourc'h G, Halos L, Desvars A, Boué F, Pascal M, Lecollinet S, et al. Chikungunya antibodies detected in non-human primates and rats in three Indian Ocean islands after the 2006 ChikV outbreak. Vet Res. 2014;45(1):52.
30. Bosco-Lauth A, Nemeth N, Kohler D, Bowen R. Viremia in North American Mammals and Birds after Experimental Infection with Chikungunya Viruses. Am J Trop Med Hyg. 2016;94(3):504-6.
31. Hartwig A, Bosco-Lauth A, Bowen R. Chikungunya virus in non-mammalian species: a possible new reservoir. New Horizons in Translational Medicine. 2015;2(4-5):128.
32. McIntosh B, Paterson H, McGillivray G, De Sousa J. Further studies on the chikungunya outbreak in southern Rhodesia in 1962 I. - Mosquitoes, wild primates and birds in relation to the epidemic. Annals of tropical medicine and parasitology. 1964;58(1):45-51.
33. Ross R. The Newala epidemic. III. The virus: isolation, pathogenic properties and relationship to the epidemic. J Hyg (Lond). 1956;54(2):177-91.
34. Economopoulou A, Dominguez M, Helynck B, Sissoko D, Wichmann O, Quenel P, et al. Atypical Chikungunya virus infections: clinical manifestations, mortality and risk factors for severe disease during the 2005–2006 outbreak on Réunion. Epidemiol Infect. 2009;137:534-41.

35. Staples JE, Breiman RF, Powers AM. Chikungunya Fever: An Epidemiological Review of a Re-Emerging Infectious Disease. Emerging Infections. 2009;49:942-8.
36. Natrajan M, Rojas A, Waggoner J. Beyond Fever and Pain: Diagnostic Methods for Chikungunya Virus. Journal of Clinical Microbiology. 2019;57(6):e00350-19.
37. Yactayo S, Staples JE, Millot V, Cibrelus L, Ramon-Pardo P. Epidemiology of chikungunya in the Americas. Journal of Infectious Diseases. 2016;214(Suppl 5):S441-S5.
38. Erasmus JH, Rossi SL, Weaver SC. Development of Vaccines for Chikungunya Fever. The Journal of Infectious Diseases. 2016;214(suppl 5):S488-S96.
39. Stapleford KA, Mulligan MJ. A New Vaccine for Chikungunya Virus. JAMA 2020;323:1351-2.
40. Valneva Initiates Phase 3 Clinical Study for its Chikungunya Vaccine Candidate VLA1553 [press release]. Saint-Herblain, France September 8, 2020.
41. Valneva Completes Recruitment for Pivotal Phase 3 Trial of Chikungunya Vaccine Candidate and Initiates Antibody Persistence Trial [press release]. Saint-Herblain, France, April 12, 2021.
42. Schneider ADB, Ochsenreiter R, Hostager R, Hofacker I, Janies D, Wolfinger M. Updated Phylogeny of Chikungunya Virus Suggests Lineage-Specific RNA Architecture. Viruses. 2019;11(9):798.
43. Kuno G. A Re-Examination of the History of Etiologic Confusion between Dengue and Chikungunya. PLoS Neglected Tropical Diseases. 2015;9(11):e0004101.
44. Halstead S. Reappearance of Chikungunya, Formerly Called Dengue, in the Americas. Emerging Infectious Diseases. 2015;21(4):557-61.
45. World Health Organization. Chikungunya Fact Sheet 2017 [Available from: http://www.who.int/mediacentre/factsheets/fs327/en/.
46. Mavalankar D, Shastri P, Bandyopadhyay T, Parmar J, Ramani KV. Increased Mortality Rate Associated with Chikungunya Epidemic, Ahmedabad, India. Emerging Infectious Diseases. 2008;14:412-5.
47. Rezza G, Nicoletti R, Angelini R, Romi R, Finarelli AC, Panning M, et al. Infection with chikungunya virus in Italy-an outbreak in a temperate region. Lancet. 2007;370:1840-46.
48. Pan American Health Organization. Chikungunya: Data, Maps and Statistics 2017 [Available from: http://www.paho.org/hq/index.php?option=com_topics&view=readall&cid=5927&Itemid=40931&lang=en.
49. 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.
50. Kraemer MUG, Reiner RC, Brady OJ, Messina JP, Gilbert M, Pigott DM, et al. Past and future spread of the arbovirus vectors Aedes aegypti and Aedes albopictus. Nature Microbiology. 2019;4:854-63.
51. U.S. Centers for Disease Control and Prevention. Where Has Chikungunya Virus Been Found? Atlanta, GA 2020 [Available from: https://www.cdc.gov/chikungunya/geo/index.html.
52. U.S. Central Intelligence Agency. The World Factbook: Colombia 2020 [Available from: https://www.cia.gov/library/publications/the-world-factbook/geos/co.html.

53. Zhang Q, Sun K, Chinazzi M, Pastore y Piontti A, Dean NE, Rojas DP, et al. Spread of Zika virus in the Americas. Proceedings of the National Academy of Sciences. 2017;114(22):E4334-E43.
54. Colombia Departamento Administrativo Nacional de Estadística. Geovisor de Consulta de Codificación de la Divipola 2020 [Available from: https://geoportal.dane.gov.co/geovisores/territorio/consulta-divipola-division-politico-administrativa-de-colombia/.
55. Kahle D, Wickham H. ggmap: Spatial Visualization with ggplot2. The R Journal; 2013.
56. Reardon S. Colombia: after the Violence. Nature. 2018;557:19-24.
57. Internal Displacement Monitoring Centre. Colombia 2020 [Available from: https://www.internal-displacement.org/countries/colombia.
58. Pan American Health Organization. Colombia. 2012.
59. Lamprea E, García J. Closing the Gap between Formal and Material Health Care Coverage in Colombia. 2016 December 5, 2016.
60. The World Bank Group. GINI index (World Bank estimate) 2021 [Available from: https://data.worldbank.org/indicator/SI.POV.GINI?end=2019&most_recent_value_desc=true&start=2019&view=map.
61. Pickett KE, Wilkinson RG. Income Inequality and Health: A Causal Review. Social Science & Medicine. 2015;128:316-26.
62. Frenk J, Frejka T, Bobadilla JL, Stern C, Lozano R, Sepúlveda J, et al. La Transición Epidemiológica en América Latina. Bol of Sainit Panam. 1991;111:485-96.
63. U.S. Centers for Disease Control and Prevention. Introduction to Public Health. In: Public Health 101 Series Atlanta, GA: U.S. Department of Health and Human Services; 2014 [Available from: https://www.cdc.gov/publichealth101/surveillance.html.
64. Groseclose SL, Buckeridge DL. Public Health Surveillance Systems: Recent Advances in Their Use and Evaluation. Annu Rev Public Health. 2017;38:57-79.
65. Declich S, Carter AO. Public health surveillance: historical origins, methods and evaluation. Bulletin of the World Health Organization. 1994;72(2):285-304.
66. Simonsen L, Gog JR, Olson D, Viboud C. Infectious Disease Surveillance in the Big Data Era: Towards Faster and Locally Relevant Systems. The Journal of Infectious Diseases. 2016;214:S380-5.
67. World Health Organization. Classifications: Classification of Diseases (ICD) 2020 [Available from: https://www.who.int/classifications/icd/en/.
68. World Health Organization. International Health Regulations (2005) Third Edition. Geneva; 2016.
69. Public Health England. Notifiable diseases and causative organisms: how to report 2020 [Available from: https://www.gov.uk/guidance/notifiable-diseases-and-causative-organisms-how-to-report.
70. Magnusson R. Advancing the right to health: the vital role of law. Geneva: World Health Organization; 2017.
71. Adams D, Fullerton K, Jajosky R, Sharp P, Onweh D, Schley A, et al. Summary of Notifiable Infectious Diseases and Conditions — United States, 2013. MMWR. 2015;62(53):1-119.
72. U.S. Centers for Disease Control and Prevention. Principles of Epidemiology in Public Health Practice, Third Edition. An Introduction to Applied Epidemiology and

Biostatistics. Lesson 5: Public Health Surveillance Atlanta, GA 2012 [Available from: https://www.cdc.gov/csels/dsepd/ss1978/lesson5/appendixe.html.
73. Walport M, Boyd I. Animal and Plant Health in the UK: Building our Science Capability. UK: UK Government Office for Science; 2014.
74. U.S. Centers for Disease Control and Prevention. Updated Guidelines for Evaluating Public Health Surveillance Systems: Recommendations from the Guidelines Working Group MMWR. 2001;50(RR13):1-35.
75. U.S. Centers for Disease Control and Prevention. Botulism: Information for Health Professionals Atlanta, GA2019 [Available from: https://www.cdc.gov/botulism/health-professional.html.
76. U.S. Centers for Disease Control and Prevention. Bioterrorism Agents/Diseases Atlanta, GA2018 [Available from: https://emergency.cdc.gov/agent/agentlist-category.asp.
77. World Health Organization. Q&A: Ethics in public health surveillance Geneva2017 [Available from: https://www.who.int/news-room/q-a-detail/q-a-ethics-in-public-health-surveillance.
78. Bhatia S, Lassmann B, Cohn E, Carrion M, Kraemer MUG, Herringer M, et al. Using digital surveillance tools for near real-time mapping of the risk of infectious disease spread. npj Digital Medicine. 2021;4:73.
79. Instituto Nacional de Salud. Lineamientos Nacionales 2021 Bogotá, Colombia 2021 [Available from: https://www.ins.gov.co/Direcciones/Vigilancia/Lineamientosydocumentos/Lineamientos%202021.pdf.
80. Nefdt R. The Foundations of Linguistics: Mathematics, Models, and Structures. St Andrews, Scotland: University of St Andrews; 2016.
81. de Poli G. Mathematical Models of the Sound of Music. In: Emmer M, editor. Mathematics and Culture III. Heidelberg, Germany: Springer Verlag; 2012.
82. White P. Mathematical Models in Infectious Disease Epidemiology. Infectious Diseases. 2017:49-53.
83. Keeling M, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton, New Jersey, USA: Princeton University Press; 2008.
84. Lambert B. A Student's Guide to Bayesian Statistics. Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE; 2018.
85. Gelman A, Hill J, Vehtari A. Regression and Other Stories. Cambridge, UK: Cambridge University Press; 2021.
86. Hamra G, MacLehose R, Richardson D. Markov Chain Monte Carlo: an introduction for epidemiologists. International Journal of Epidemiology. 2013;42:627-34.
87. Roberts G, Gelman A, Gilks WR. Weak convergence and optimal scaling of random walk Metropolis algorithms. The Annals of Applied Probability. 1997;7(1):110-20.
88. Rosenthal J. Optimal proposal distributions and adaptive MCMC. In: Brooks S, Gelman A, Jones G, Meng X-L, editors. Handbook of Markov Chain Monte Carlo. Boca Raton, FL: Chapman & Hall/CRC Press; 2011. p. 93-112.
89. Neto ASL, Sousa GS, Nascimento OJ, Castro MC. Chikungunya-attributable deaths: A neglected outcome of a neglected disease. PLoS Neglected Tropical Diseases. 2019;13:e0007575.
90. Freitas ARR, Cavalcanti L, Von Zuben APB, Donalisio MR. Excess Mortality Related to Chikungunya Epidemics in the Context of Co-Circulation of Other Arboviruses in

Brazil. PLoS Curr. 2017;9:ecurrents.outbreaks.14608e586cd321d8d5088652d7a0d884.
91. Freitas ARR, Donalisio MR, Alarcón-Elbal PM. Excess Mortality and Causes Associated with Chikungunya, Puerto Rico, 2014–2015. Emerging Infectious Diseases. 2018;24:2352-5.
92. Chang AY, Encinales L, Porras A, Pacheco N, Reid SP, Martins KAO, et al. Frequency of chronic joint pain following chikungunya virus infection. Arthritis & Rheumatology. 2018;70:578-84.
93. Alvis-Zakzuk N, Díaz-Jiménez D, Castillo-Rodríguez L, Castañeda-Orjuela C, Paternina-Caicedo A, Pinzón-Redondo H, et al. Economic Costs of Chikungunya Virus in Colombia. Value in Health Regional Issues. 2018;17:32-7.
94. Yagnik P, Linou N, Webb D, Blanco U. A Socio-Economic Impact Assessment of the Zika Virus in Latin America and the Caribbean: with a Focus on Brazil, Colombia and Suriname. United Nations Development Programme; 2017.
95. Pacheco O, Beltrán M, Nelson CA, Valencia D, Tolosa N, Farr SL, et al. Zika Virus Disease in Colombia — Preliminary Report. New England Journal of Medicine. 2016.
96. Pan American Health Organization. Zika-Epidemiological Report Colombia. Washington, D.C.; 2017.
97. Instituto Nacional de Salud. Boletín Epidemiológico Semanal. 2016.
98. Sejvar JJ, Kohl KS, Bilynsky R, Blumberg D, Cvetkovich T, Galama J, et al. Encephalitis, myelitis, and acute disseminated encephalomyelitis (ADEM): Case definitions and guidelines for collection, analysis, and presentation of immunization safety data. Vaccine. 2007;25:5771-92.
99. Sejvar JJ, Kohl KS, Gidudu J, Amato A, Bakshi N, Baxter R, et al. Guillain–Barré syndrome and Fisher syndrome: Case definitions and guidelines for collection, analysis, and presentation of immunization safety data. Vaccine. 2011;29:599-612.
100. Kohl KS, Bonhoeffer J, Braun MM, Buettcher M, Chen RT, Duclos P, et al. The development of standardized case definitions and guidelines for adverse events following immunization. Vaccine. 2007;25:5671-4.
101. Charniga K, Cucunubá ZM, Walteros DM, Mercado M, Prieto F, Ospina M, et al. Descriptive analysis of surveillance data for Zika virus disease and Zika virus-associated neurological complications in Colombia, 2015–2017. PLoS ONE. 2021;16(6):e0252236.
102. McGrogan A, Madle GC, Seaman HE, de Vries CS. The Epidemiology of Guillain-Barré Syndrome Worldwide. Neuroepidemiology. 2009;32:150-63.
103. Sejvar JJ, Baughman AL, Wise M, Morgan OW. Population Incidence of Guillain-Barré Syndrome: A Systematic Review and Meta-Analysis. Neuroepidemiology. 2011;36(2):123-33.
104. Cao-Lormeau VM, Blake A, Mons S, Lastère S, Roche C, Vanhomwegen J, et al. Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. Lancet. 2016;387:1531-9.
105. Dirlikov E, Major CG, Medina NA, Lugo-Robles R, Matos D, Muñoz-Jordan J, et al. Clinical Features of Guillain-Barré Syndrome With vs Without Zika Virus Infection, Puerto Rico, 2016. JAMA Neurol. 2018;75(9):1089-97.
106. Parra B, Lizarazo J, Jiménez-Arango JA, Zea-Vera AF, González-Manrique G, Vargas J. Guillain–Barré Syndrome Associated with Zika Virus Infection in Colombia. N Engl J Med. 2016;375(16):1513-23.

107. Gold CA, Josephson A. Anticipating the Challenges of Zika Virus and the Incidence of Guillain-Barré Syndrome. JAMA Neurol. 2016;73(8):905-6.
108. van Doorn PA. Diagnosis, treatment and prognosis of Guillain-Barré syndrome (GBS). Presse Med. 2013;42:e193-e201.
109. Bersano A, Carpo M, Allaria S, Franciotta D, Citterio A, Nobile-Orazio E. Long term disability and social status change after Guillain-Barré syndrome. J Neurol. 2006;253:214-8.
110. Walteros DM, Soares J, Styczynski AR, Abrams JY, Galindo-Buitrago JI, Acosta-Reyes J, et al. Long-term outcomes of Guillain-Barre´ syndrome possibly associated with Zika virus infection. PLoS ONE. 2019;14(8): e0220049.
111. Webb AJS, Brain SAE, Wood R, Rinaldi S, Turner MR. Seasonal variation in Guillain-Barré syndrome: a systematic review, meta-analysis and Oxfordshire cohort study. J Neurol Neurosurg Psychiatry. 2015;86:1196-201.
112. Hoen B, Schaub B, Funk AL, Ardillon V, Boullard M, Cabié A, et al. Pregnancy Outcomes after ZIKV Infection in French Territories in the Americas. N Engl J Med. 2018;378(11):985-94.
113. Bertolli J, Attell JE, Rose C, Moore CA, Melo F, Staples JE, et al. Functional Outcomes among a Cohort of Children in Northeastern Brazil Meeting Criteria for Follow-Up of Congenital Zika Virus Infection. Am J Trop Med Hyg. 2020;102(5):955-63.
114. Mulkey SB, Arroyave-Wessel M, Peyton C, Bulas DI, Fourzali Y, Jiang J, et al. Neurodevelopmental Abnormalities in Children With In Utero Zika Virus Exposure Without Congenital Zika Syndrome. JAMA Pediatr. 2020;174(3):269-76.
115. Raper J, Chahroudi A. Clinical and Preclinical Evidence for Adverse Neurodevelopment after Postnatal Zika Virus Infection. Trop Med Infect Dis. 2021;6(1):10.
116. Cuevas EL, Tong VT, Rozo N, Valencia D, Pacheco O, Gilboa SM, et al. Preliminary Report of Microcephaly Potentially Associated with Zika Virus Infection During Pregnancy — Colombia, January–November 2016. MMWR. 2016;65(49):1409-13.
117. Sistema de información geográfica para la planeación y el ordenamiento territorial. Visor SIG-OT Colombia2018 [Available from: https://sigot.igac.gov.co.
118. Gambhir M, Swerdlow DL, Finelli L, Van Kerkhove MD, Biggerstaff M, Cauchemez S, et al. Multiple Contributory Factors to the Age Distribution of Disease Cases: A Modeling Study in the Context of Influenza A(H3N2v). Clin Infect Dis. 2013;57:S23-S7.
119. Lozier M, Adams L, Febo MF, Torres-Aponte J, Bello-Pagan M, Ryff KR. Incidence of Zika Virus Disease by Age and Sex: Puerto Rico, November 1, 2015-October 20, 2016. MMWR. 2016;65(44):1219-23.
120. Rodriguez-Barraquer I, Costa F, Nascimento EJM, Júnior NN, Castanha PMS, Sacramento GA, et al. Impact of preexisting dengue immunity on Zika virus emergence in a dengue endemic region. Science. 2019;363:607-10.
121. Villarroel PMS, Nurtop E, Pastorino B, Roca Y, Drexler JF, Gallian P, et al. Zika virus epidemiology in Bolivia: A seroprevalence study in volunteer blood donors. PLoS Neglected Tropical Diseases. 2018;12(3):e0006239.
122. Netto EM, Moreira-Soto A, Pedroso C, Höser C, Funk S, Kucharski AJ, et al. High Zika Virus Seroprevalence in Salvador, Northeastern Brazil Limits the Potential for Further Outbreaks. mBio. 2017;8(6):e01390-17.

123. Diarra I, Nurtop E, Sangaré A, Sagara I, Pastorino B, Sacko S, et al. Zika Virus Circulation in Mali. Emerging Infectious Diseases. 2020;26(5).
124. Honório NA, da Costa Silva W, Leite PJ, Gonçalves JM, Lounibos LP, Lourenço-de-Oliveira R. Dispersal of Aedes aegypti and Aedes albopictus (Diptera: Culicidae) in an Urban Endemic Dengue Area in the State of Rio de Janeiro, Brazil. Mem Inst Oswaldo Cruz. 2003;98(2):191-8.
125. Overgaard HJ, Olano VA, Jaramillo JF, Matiz MI, Sarmiento D, Stenström TA, et al. A cross-sectional survey of Aedes aegypti immature abundance in urban and rural household containers in central Colombia. Parasites & Vectors. 2017;10:356.
126. Coelho FC, Durovni B, Saraceni V, Lemos C, Codeco CT, Camargo S. Higher Incidence of Zika in Adult Women than Adult Men in Rio de Janeiro Suggests a Significant Contribution of Sexual Transmission from Men to Women. Int J Infect Dis. 2016;51:128-32.
127. Buckingham-Jeffery E, Morbey R, House T, Elliot AJ, Harcourt S, Smith GE. Correcting for day of the week and public holiday effects: improving a national daily syndromic surveillance service for detecting public health threats. BMC Public Health. 2017;17(477).
128. Esposito S, Longo MR. Guillain–Barré Syndrome. Autoimmunity Reviews. 2017;16(1):96-101.
129. Leonhard SE, Bresani-Salvi CC, Lyra Batista JDL, Cunha S, Jacobs BC, Brito Ferreira ML, et al. Guillain-Barré syndrome related to Zika virus infection: A systematic review and meta-analysis of the clinical and electrophysiological phenotype. PLoS Neglected Tropical Diseases. 2020;14(4):e0008264.
130. Box G, Jenkins G. Time series analysis: Forecasting and control. San Francisco Holden-Day; 1970.
131. Ministerio de Salud de Colombia. Preguntas frecuentes registro individual de atención - RIPS Versión 1 Bogotá 2015 [Available from: https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/OT/FAQ-RIPS.pdf.
132. Martínez Ramos M, Pacheco García O. Utilidad de los Registros Individuales de Prestación de Servicios (RIPS) para la vigilancia en salud pública, Colombia, 2012. Informe Quincenal Epidemiológico Nacional. 2013;18(17):176-92.
133. Mier-y-Teran-Romero L, Delorey MJ, Sejvar JJ, Johansson MA. Guillain–Barré syndrome risk among individuals infected with Zika virus: A multi-country assessment. BMC Medicine. 2018;16:67.
134. da Silva IRF, Frontera JA, de Filippis AMB, do Nascimento OJM. Neurologic Complications Associated With the Zika Virus in Brazilian Adults. JAMA Neurol. 2017;74(10):1190-8.
135. Mehta R, Soares CN, Medialdea-Carrera R, Ellul M, da Silva MTT, Rosala-Hallas A, et al. The spectrum of neurological disease associated with Zika and chikungunya viruses in adults in Rio de Janeiro, Brazil: A case series. PLoS Neglected Tropical Diseases. 2018;12(2):e0006212.
136. Mécharles S, Herrmann C, Poullain P, Tran T, Deschamps N, Mathon G, et al. Acute myelitis due to Zika virus infection. The Lancet. 2016;387(10026):P1481.
137. Rozé B, Najioullah F, Signate A, Apetse K, Brouste Y, Gourgoudou S, et al. Zika virus detection in cerebrospinal fluid from two patients with encephalopathy, Martinique, February 2016. Euro Surveill. 2016;21(16):pii=30205.

138. Soares CN, Brasil P, Medialdea-Carrera R, Sequeira P, de Filippis AMB, Borges VA, et al. Fatal encephalitis associated with Zika virus infection in an adult. J Clin Virol. 2016;83:63-5.
139. Carteaux G, Maquart M, Bedet A, Contou D, Brugières P, Fourati S, et al. Zika Virus Associated with Meningoencephalitis. N Engl J Med. 2016;374:1595-6.
140. Niemeyer B, Niemeyer R, Borges R, Marchiori E. Acute Disseminated Encephalomyelitis Following Zika Virus Infection. European Neurology. 2017;77:45-6.
141. Kassavetis P, Joseph JB, Francois R, Perloff MD, Berkowitz AL. Zika virus–associated Guillain-Barré syndrome variant in Haiti. Neurology. 2016;87(3):336-7.
142. Molko N, Simon O, Guyon D, Biron A, Dupont-Rouzeyrol M, Gourinat A. Zika virus infection and myasthenia gravis: report of 2 cases. Neurology. 2017;88(11):1097-8.
143. Miner J, Diamond M. Zika Virus Pathogenesis and Tissue Tropism. Cell Host & Microbe. 2017;21(2):134-42.
144. Salinas JL, Walteros DM, Styczynski A, Garzón F, Quijada H, Bravo E, et al. Zika virus disease-associated Guillain-Barré syndrome—Barranquilla,Colombia 2015–2016. Journal of the Neurological Sciences. 2017;381:272-7.
145. Vidal O, Acosta-Reyes J, Padilla J, Navarro-Lechuga E, Bravo E, Viasus D, et al. Chikungunya outbreak (2015) in the Colombian Caribbean: Latent classes and gender differences in virus infection. PLoS Neglected Tropical Diseases. 2020;14(6):e0008281.
146. Mehta R, Gerardin P, Antunes de Brito CA, Soares CN, Ferreira MLB, Solomon T. The neurological complications of chikungunya virus: A systematic review. Rev Med Virol. 2018;28(e1978):1-24.
147. Torres J, Falleiros-Arlant L, Dueñas L, Pleitez-Navarrete J, Salgado D, Castillo J. Congenital and perinatal complications of chikungunya fever: a Latin American experience. International Journal of Infectious Diseases. 2016;51:85-8.
148. Villamil-Gomez W, Alba-Silvera L, Menco-Ramos A, Gonzalez-Vergara A, Molinares-Palacios T, Barrios-Corrales M, et al. Congenital Chikungunya Virus Infection in Sincelejo, Colombia: A Case Series. Journal of Tropical Pediatrics. 2015;61:386-92.
149. Alvarado-Socarras J, Ocampo-González M, Vargas-Soler J, Rodriguez-Morales A, Franco-Paredes C. Congenital and Neonatal Chikungunya in Colombia. Journal of the Pediatric Infectious Diseases Society. 2016;5(3):e17-e20.
150. Villamil-Gomez W, Alba-Silvera L, Páez-Castellanos J, Rodriguez-Morales A. Guillain–Barré syndrome after Chikungunya infection: A case in Colombia. Enferm Infecc Microbiol Clin. 2016;34(2):139-44.
151. Ospina ML, Tong VT, Gonzalez M, Valencia D, Mercado M, Gilboa SM, et al. Zika Virus Disease and Pregnancy Outcomes in Colombia. New England Journal of Medicine. 2020;383(6):537-45.
152. Pacheco O, Newton S, Daza M, Cates J, Reales J, Burkel V, et al. Neurodevelopmental findings in children 20-30 months of age with postnatal Zika infection at 1-12 months of age, Colombia, September-November 2017. Paediatr Perinat Epidemiol. 2021;35(1):92-7.
153. WorldPop. Open spatial demographic data and research 2021 [Available from: https://www.worldpop.org.
154. DANE. Proyecciones de población 2021 [Available from: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion.

155. Anderson RM, May RM. Infectious diseases of humans: dynamics and control: Oxford University Press; 1991.
156. Van Kerkhove MD, Bento AI, Mills HL, Ferguson NM, Donnelly CA. A review of epidemiological parameters from Ebola outbreaks to inform early public health decision-making. Nature Scientific Data. 2015;2:150019.
157. Heesterbeek J. A brief history of R0 and a recipe for its calculation. Acta Biotheoretica. 2002;50:189-204.
158. O'Driscoll M, Harry C, Donnelly CA, Cori A, Dorigatti I. A comparative analysis of statistical methods to estimate the reproduction number in emerging epidemics with implications for the current COVID-19 pandemic. Clin Infect Dis. 2020(ciaa1599).
159. Cori A, Ferguson NM, Fraser C, Cauchemez S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. American Journal of Epidemiology. 2013;178(9):1505-12.
160. Ferguson NM, Laydon D, Nedjati-Gilani GL, Imai N, Ainslie K, Baguelin M, et al. Report 9 - Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. London, UK: Imperial College COVID-19 Response Team; 2020.
161. The Singapore Zika Study Group. Outbreak of Zika virus infection in Singapore: an epidemiological, entomological, virological, and clinical analysis. Lancet Infect Dis. 2017;17(8):813-21.
162. Anderson RM, May RM. Vaccination and herd immunity to infectious diseases. Nature. 1985;318:323-9.
163. Bettencourt LMA, Ribeiro RM. Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases. PLoS ONE. 2008;3(5):e2185.
164. Cori A, Cauchemez S, Ferguson NM, Fraser C, Dahlqwist E. R package EpiEstim: Estimate Time Varying Reproduction Numbers from Epidemic Curves. 1.1-2 ed: CRAN; 2013.
165. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proceedings of the Royal Society B. 2007;274:599-604.
166. Wallinga J, Teunis P. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. Am J Epidemiol. 2004;160(6):509-16.
167. White LF, Pagano M. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. Stat Med. 2008;27(16):2999-3016.
168. Howard S, Donnelly CA. Estimation of a time-varying force of infection and basic reproduction number with application to an outbreak of classical swine fever. J Epidemiol Biostat. 2000;5(3):161-8.
169. Cauchemez S, Hoze N, Cousien A, Nikolay B, ten Bosch QA. How Modelling Can Enhance the Analysis of Imperfect Epidemic Data. Trends in Parasitology. 2019;35(5):369-79.
170. Fraser C. Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic. PLoS ONE. 2007;2(8):e758.
171. Mishra S, Berah T, Mellan TA, Unwin HJT, Vollmer MA, Parag KV, et al. On the derivation of the renewal equation from an age-dependent branching process: an epidemic modelling perspective. arXiv. 2020:2006.16487v1.

172. Nishiura H. Correcting the actual reproduction number: a simple method to estimate R(0) from early epidemic growth data. International Journal of Environmental Research and Public Health. 2010;7(1):291-302.
173. Fraser C, Cummings DAT, Klinkenberg D, Burke DS, Ferguson NM. Influenza Transmission in Households During the 1918 Pandemic. Am J Epidemiol. 2011;174(5):505-14.
174. Chowell G, Hincapie-Palacio D, Ospina J, Pell B, Tariq A, Dahal S, et al. Using Phenomenological Models to Characterize Transmissibility and Forecast Patterns and Final Burden of Zika Epidemics. PLoS Currents Outbreaks. 2016.
175. Nouvellet P, Cori A, Garske T, Blake IM, Dorigatti I, Hinsley W, et al. A simple approach to measure transmissibility and forecast incidence. Epidemics. 2018;22:29-35.
176. Anderson R, Donnelly CA, Hollingsworth D, Keeling M, Vegvari C, Baggaley R, et al. Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation. The Royal Society; 2020.
177. Peña-García V, Christofferson R. Correlation of the basic reproduction number (R0) and eco-environmental variables in Colombian municipalities with chikungunya outbreaks during 2014-2016. PLoS Neglected Tropical Diseases. 2019;13(11):e0007878.
178. Rojas D, Dean N, Yang Y, Kenah E, Quintero J, Tomasi S, et al. The epidemiology and transmissibility of Zika virus in Girardot and San Andres island, Colombia, September 2015 to January 2016. Euro Surveill. 2016;21(28):pii=30283.
179. Towers S, Brauer F, Castillo-Chavez C, Falconar AKI, Mubayi A, Romero-Vivas CME. Estimate of the reproduction number of the 2015 Zika virus outbreak in Barranquilla, Colombia, and estimation of the relative role of sexual transmission. Epidemics. 2016;17:50-5.
180. Ospina J, Hincapie-Palacio D, Ochoa H, Molina A, Rúa G, Pájaro D, et al. Stratifying the potential local transmission of Zika in municipalities of Antioquia, Colombia. Tropical Medicine and International Health. 2017;22(10):1249-65.
181. Viboud C, Simonsen L, Chowell G. A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. Epidemics. 2016;15:27-37.
182. Hsieh Y. Temporal patterns and geographic heterogeneity of Zika virus (ZIKV) outbreaks in French Polynesia and Central America. PeerJ. 2017;5:e3015.
183. Chowell G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. Infectious Disease Modelling. 2017;2:379-98.
184. Nishiura H, Mizumoto K, Villamil-Gomez W, Rodríguez-Morales A. Preliminary estimation of the basic reproduction number of Zika virus infection during Colombia epidemic, 2015-2016. Travel Medicine and Infectious Disease. 2016;14:274-6.
185. Rocklöv J, Quam M, Sudre B, German M, Kraemer MUG, Brady O, et al. Assessing Seasonal Risks for the Introduction and Mosquito-borne Spread of Zika Virus in Europe. EBioMedicine. 2016;9:250-6.
186. Majumder M, Santillana M, Mekaru S, McGinnis D, Khan K, Brownstein J. Utilizing Nontraditional Data Sources for Near Real-Time Estimation of Transmission Dynamics During the 2015-2016 Colombian Zika Virus Disease Outbreak. JMIR Public Health and Surveillance. 2016;2(1):e30.

187. Sasmal S, Ghosh I, Huppert A, Chattopadhyay J. Modeling the Spread of Zika Virus in a Stage-Structured Population: Effect of Sexual Transmission. Bulletin of Mathematical Biology. 2018;80:3038-67.
188. Gibbons C, Mangen M, Plass D, Havelaar A, John Brooke R, Kramarz P, et al. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. BMC Public Health. 2014;14:147.
189. Pacheco O, Martínez M, Alarcón A, Bonilla M, Caycedo A, Valbuena T, et al. Estimación del subregistro de casos de enfermedad por el virus del chikungunya en Girardot, Colombia, noviembre de 2014 a mayo de 2015. Biomédica. 2017;37:507-15.
190. Martínez Duran M, Díaz J, Gómez N, López B, Rodríguez A, Montana C, et al. Estimación del subregistro de casos de enfermedad por el virus de chikunguña en el municipio de El Espinal, Tolima, octubre de 2014 a junio de 2015. Bogotá: Ministerio de Salud y Protección Social y Instituto Nacional de Salud; 2016. Contract No.: 22.
191. Nouvellet P, Cucunuba Z, Rodriguez-Barraquer I, Vanhomwegen J, Montoya MC, Correa A, et al. Characterization of population exposure (seroprevalence) to arboviruses after recent outbreaks in Colombia: dengue, chikungunya, and Zika. American Journal of Tropical Medicine and Hygiene. 2018;99(4 Suppl):296. Abstract 941.
192. Moore SM, Oidtman RJ, James Soda K, Siraj AS, Reiner RC, Barker CM, et al. Leveraging multiple data types to estimate the size of the Zika epidemic in the Americas. PLoS Neglected Tropical Diseases. 2020;14(9):e0008640.
193. Martínez Duran M, Alfonso K, Ávila C, Barrios A, Cárdenas J, Causil A, et al. Subregsitro de Zika en Girardot, Cundinamarca, 2015-2016. Bogotá: Instituto Nacional de Salud y 2016. Contract No.: 23.
194. Ciota A, Matacchiero A, Kilpatrick A, Kramer L. The effect of temperature on life history traits of Culex mosquitoes. J Med Entomol. 2014;51(1):55-62.
195. Mordecai E, Cohen J, Evans M, Gudapati P, Johnson L, Lippi CA, et al. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. PLoS Neglected Tropical Diseases. 2017;11(4):e0005568.
196. Paaijmans K, Wandago M, Githeko A, Takken W. Unexpected high losses of Anopheles gambiae larvae due to rainfall. PLoS ONE. 2007;2(11):e1146.
197. Koenraadt C, Harrington L. Flushing effect of rain on container-inhabiting mosquitoes Aedes aegypti and Culex pipiens (Diptera: Culicidae). J Med Entomol. 2008;45(1):28-35.
198. Brady O, Johansson MA, Guerra C, Bhatt S, Golding N, Pigott DM, et al. Modelling adult Aedes aegypti and Aedes albopictus survival at different temperatures in laboratory and field settings. Parasites & Vectors. 2013;6:351.
199. Riou J, Poletto C, Boëlle PY. A comparative analysis of Chikungunya and Zika transmission. Epidemics. 2017;19:43-52.
200. Tesla B, Demakovsky L, Mordecai E, Ryan SJ, Bonds M, Ngonghala C, et al. Temperature drives Zika virus transmission: evidence from empirical and mathematical models. Proc R Soc B. 2018;285:20180795.
201. Kakarla S, Mopuri R, Mutheneni S, Bhimala K, Kumaraswamy S, Kadiri M, et al. Temperature dependent transmission potential model for chikungunya in India. Science of the Total Environment. 2019;647:66-74.

202. Harris M, Caldwell J, Mordecai E. Climate drives spatial variation in Zika epidemics in Latin America. Proc R Soc B. 2019;286:20191578.
203. Chien L, Sy FS, Pérez A. Identifying high risk areas of Zika virus infection by meteorological factors in Colombia. BMC Infectious Diseases. 2019;19:888.
204. Siraj AS, Rodriguez-Barraquer I, Barker CM, Tejedor-Garavito N, Harding D, Lorton C, et al. Data from: Spatiotemporal incidence of Zika and associated environmental drivers for the 2015-2016 epidemic in Colombia. Dryad Digital Repository 2017.
205. National Oceanic and Atmospheric Organization. Climate Data Online Search 2020 [Available from: https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND.
206. Nguyen P, Shearer E, Tran H, Ombadi M, Hayatbini N, Palacios T, et al. The CHRS Data Portal, an easily accessible public repository for PERSIANN global satellite precipitation data. Nature Scientific Data. 2019;6:180296.
207. Siraj AS, Rodriguez-Barraquer I, Barker CM, Tejedor-Garavito N, Harding D, Lorton C. Spatiotemporal Incidence of Zika and Associated Environmental Drivers for the 2015-2016 Epidemic in Colombia. Nature Scientific Data. 2018;5(180073).
208. National Oceanic and Atmospheric Organization. NNDC Climate Data Online 2020 [Available from: https://www7.ncdc.noaa.gov/CDO/cdo.
209. Humanitarian Data Exchange. Colombia - Subnational Administrative Boundaries 2020 [Available from: https://data.humdata.org/dataset/colombia-administrative-boundaries-levels-0-3.
210. Ferguson NM, Cucunubá ZM, Dorigatti I, Nedjati-Gilani GL, Donnelly CA, Basánez M, et al. Countering Zika in Latin America. Science. 2016;353(6297):353-4.
211. Caicedo E, Charniga K, Rueda A, Dorigatti I, Mendez Y, Hamlet A, et al. The epidemiology of Mayaro virus in the Americas: A systematic review and key parameter estimates for outbreak modelling. PLoS Negl Trop Dis. 2021;16(6):e0009418.
212. Singapore Zika Study Group. Outbreak of Zika virus infection in Singapore: an epidemiological, entomological, virological, and clinical analysis. Lancet Infect Dis. 2017;17(8):813-21.
213. Salje H, Lessler J, Kumar Paul K, Azman AS, Waliur Rahman M. How social structures, space, and behaviors shape the spread of infectious diseases using chikungunya as a case study. PNAS. 2016;113(47):13420-5.
214. Cauchemez S, Ledrans M, Polleto C, Quenel P, De Valk H, Colizza V, et al. Local and regional spread of chikungunya fever in the Americas. Euro Surveill. 2014;19(28):1-14.
215. Riou J, Poletto C, Boëlle P. Improving early epidemiological assessment of emerging Aedes-transmitted epidemics using historical data. PLoS Negl Trop Dis. 2018;12(6):e0006526.
216. Majumder M, Cohn E, Fish D, Brownstein J. Estimating a feasible serial interval range for Zika fever. Bulletin of the World Health Organization. 2016:BLT.16.171009.
217. Dorfman R. A Note on the δ-Method for Finding Variance Formulae. The Biometric Bulletin. 1938;1:129-37.
218. Wood S. R Package “mgcv”. Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. 1.8-28 ed: CRAN; 2020.
219. Wood S. Generalized Additive Models: An Introduction with R (2nd edition): Chapman and Hall/CRC Press; 2017.

220. Burnham K, Anderson D. Multimodal inference: Understanding AIC and BIC in model selection. Sociological Methods & Research. 2004;33:261-304.
221. Grubaugh ND, Ladner J, Kraemer MUG, Dudas G, Tan A, Gangavarapu K, et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature. 2017;546:401-4015.
222. Sissoko D, Moendandze A, Malvy D, Giry C, Ezzedine K, Solet J, et al. Seroprevalence and Risk Factors of Chikungunya Virus Infection in Mayotte, Indian Ocean, 2005-2006: A Population-Based Survey. PLoS ONE. 2008;3(8):e3066.
223. Ryan SJ, Lippi CA, Nightingale R, Hamerlinck G, Borbor-Cordova M, Cruz B M, et al. Socio-Ecological Factors Associated with Dengue Risk and Aedes aegypti Presence in the Galápagos Islands, Ecuador. Int J Environ Res Public Health. 2019;16(5):682.
224. Gilks WR, Richardson S, Spiegelhalter DJ. Markov Chain Monte Carlo in Practice: Chapman & Hall/CRC; 1998.
225. Hastings W. Monte Carlo Sampling Methods using Markov Chains and their Applications. Biometrika. 1970;57(1):97-109.
226. Plummer M, Best N, Cowles K, Vines K, Sarkar D, Bates D, et al. R Package "coda". 0.19-4 ed. CRAN2020.
227. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian Measures of Model Complexity and Fit. Journal of the Royal Statistical Society Series B (Statistical Methodology). 2002;64(4):583-639.
228. Villar L, Rojas D, Besada-Lombana S, Sarti E. Epidemiological Trends of Dengue Disease in Colombia (2000-2011): A Systematic Review. PLoS Neglected Tropical Diseases. 2015;9(3):e0003499.
229. Colombia Ministerio de Salud y Protección Social. Situación Actual de Dengue a Semana 12 de 2013 Periodo de Análisis: 2008-2013 2013 [Available from: https://www.minsalud.gov.co/Documentos%20y%20Publicaciones/INFORME%20SITUACION%20DE%20DENGUE.pdf.
230. Lowe R, Barcellos C, Brasil P, Cruz O, Honório NA, Kuper H, et al. The Zika Virus Epidemic in Brazil: From Discovery to Future Implications. Int J Environ Res Public Health. 2018;15(1):96.
231. Watts AG, Miniota J, Joseph HA, Brady OJ, Kraemer MUG, Grills AW, et al. Elevation as a proxy for mosquito-borne Zika virus transmission in the Americas. PLoS ONE. 2017;12(5):e0178211.
232. Ribeiro GS, Hamer G, Diallo M, Kitron U, Ko A, Weaver S. Influence of herd immunity in the cyclical nature of arboviruses. Current Opinion in Virology. 2020;40:1-10.
233. Funk S, Kucharski AJ, Camacho A, Eggo RM, Yakob L, Murray LM, et al. Comparative Analysis of Dengue and Zika Outbreaks Reveals Differences by Setting and Virus. PLoS Neglected Tropical Diseases. 2016;10(12):e0005173.
234. Delamater P, Street E, Leslie T, Yang Y, Jacobsen K. Complexity of the Basic Reproduction Number (R0). Emerging Infectious Diseases. 2019;25(1):1-4.
235. Moore SM, ten Bosch QA, Siraj AS, James Soda K, España G, Campo A, et al. Local and regional dynamics of chikungunya virus transmission in Colombia: the role of mismatched spatial heterogeneity. BMC Medicine. 2018;16:152.
236. Capasso A, Ompad D, Vieira D, Wilder-Smith A, Tozan Y. Incidence of Guillain-Barré Syndrome (GBS) in Latin America and the Caribbean before and during the 2015–2016 Zika virus epidemic: A systematic review and meta-analysis. PLoS Neglected Tropical Diseases. 2019;13(8):e0007622.

237. Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, et al. Stan: A Probabilistic Programming Language. Journal of Statistical Software. 2017;76(1):1-32.
238. Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P. Rank-normalization, folding, and localization: An improved R-hat for assessing convergence of MCMC. Bayesian Analysis. 2021:Advance publication 1-28.
239. Zambrana JV, Carrillo FB, Burger-Calderon R, Collado D, Sanchez N, Ojeda S, et al. Seroprevalence, risk factor, and spatial analyses of Zika virus infection after the 2016 epidemic in Managua, Nicaragua. PNAS. 2018;115(37):9294-9.
240. Lara J, Villegas A, Chavarro D, Silva G, Gómez M, García S, et al. Estudio de disponibilidad y distribución de la oferta de médicos especialistas, en servicios de alta y mediana complejidad en Colombia. Bogotá Cendex; 2013.
241. World Health Organization. Atlas: country resources for neurological disorders – 2nd ed. Geneva 2017.
242. Watson O, Alhaffar M, Mehchy Z, Whittaker C, Akil Z, Brazeau N, et al. Leveraging community mortality indicators to infer COVID-19 mortality and transmission dynamics in Damascus, Syria. Nature Communications. 2021;12:2394.
243. Jorgensen D, Pons-Salort M, Shaw A, Grassly N. The role of genetic sequencing and analysis in the polio eradication programme. Virus Evolution. 2020;6(2):veaa040.
244. Pan American Health Organization. Zika-Epidemiological Report Ecuador. Washington, D.C.; 2017.
245. Charniga K, Cucunubá ZM, Mercado M, Prieto F, Ospina M, Nouvellet P, et al. Spatial and temporal invasion dynamics of the 2014–2017 Zika and chikungunya epidemics in Colombia. PLoS Comput Biol. 2021;17(7):e1009174.
246. Walter SD. Disease Mapping: A Historical Perspective. In Spatial Epidemiology: Methods and Applications. Oxford: Oxford University Press; 2000.
247. Elliot P, Wartenberg D. Spatial Epidemiology: Current Approaches and Future Challenges. Environmental Health Perspectives. 2004;112:998-1006.
248. Chowell G, Rothenberg R. Spatial infectious disease epidemiology: on the cusp. BMC Medicine. 2018;16:1-5.
249. Ferguson NM, Donnelly CA, Anderson RM. The Foot-and-Mouth Epidemic in Great Britain: Pattern of Spread and Impact of Interventions. Science. 2001;292:1155-60.
250. D'Silva JP, Eisenberg MC. Modeling Spatial Invasion of Ebola in West Africa. J Theor Biol. 2017;428:65-75.
251. Vollmer MAC, Mishra S, Unwin HJT, Gandy A, Imperial College COVID-19 Response Team. Report 20: Using mobility to estimate the transmission intensity of COVID-19 in Italy: A subnational analysis with future scenarios. Imperial College London; 2020.
252. Bradley DJ. The Scope of Travel Medicine: An Introduction to the Conference on International Travel Medicine. In: Steffen R., Lobel H., Haworth J., D.J. B, editors. Travel Medicine. Berlin, Heidelberg: Springer; 1989. p. 1-9.
253. Cliff A, Haggett P. Time, travel and infection. British Medical Bulletin. 2004;69(1):87-99.
254. Tatem AJ, Rogers DJ, Hay SI. Global Transport Networks and Infectious Disease Spread. Adv Parasitol. 2006;62:293-343.
255. The World Bank Group. Air transport, passengers carried 2020 [Available from: https://data.worldbank.org/indicator/IS.AIR.PSGR.

256. Vaughan A, Aarons E, Astbury J, Brooks T, Chand M, Flegg P, et al. Human-to-Human Transmission of Monkeypox Virus, United Kingdom, October 2018. Emerging Infectious Diseases. 2020;26:782-5.
257. Findlater A, Bogoch II. Human Mobility and the Global Spread of Infectious Diseases: A Focus on Air Travel. Trends in Parasitology. 2018;34(9):772-83.
258. Rodrigue J, Comtois C, Slack B. The Geography of Transport Systems. Oxon: Routledge; 2006.
259. Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, et al. Human mobility: Models and applications. Physics Reports. 2018;734:1-74.
260. Lessler J, Salje H, Grabowski MK, Cummings DAT. Measuring Spatial Dependence for Infectious Disease Epidemiology. PLoS ONE. 2016;11(5):e0155249.
261. Kuno G. Review of the Factors Modulating Dengue Transmission. Epidemiologic Reviews. 1995;17(2):321-35.
262. Stoddard ST, Forshey BM, Morrison AC, Paz-Soldan VA, Vazquez-Prokopec GM, Astete H, et al. House-to-house human movement drives dengue virus transmission. PNAS. 2012;110(3):994-9.
263. Truscott J, Ferguson NM. Evaluating the Adequacy of Gravity Models as a Description of Human Mobility for Epidemic Modelling. PLoS Computational Biology. 2012;8(10):e1002699.
264. Carey HC. Principles of Social Science. Volume: 1. New York: J.B. Lippincott & Co; 1858.
265. Zipf GK. The P1 P2/D Hypothesis: On the Intercity Movement of Persons. Am Sociol Rev. 1946;11:677-86.
266. Xia Y, Bjørnstad O, Grenfell B. Measles Metapopulation Dynamics: A Gravity Model for Epidemiological Coupling and Dynamics. The American Naturalist. 2004;164(2):267-81.
267. Bertuzzo E, Mari L, Righetto L, Gatto M, Casagrandi R, Blokesch M, et al. Prediction of the spatial evolution and effects of control measures for the unfolding Haiti cholera outbreak. Geophysical Research Letters. 2011;38:L06403.
268. Tuite AR, Tien J, Eisenberg M, Earn DJ, Ma J, Fisman DN. Cholera epidemic in Haiti, 2010: using a transmission model to explain spatial spread of disease and identify optimal control interventions. Ann Intern Med. 2011;154(9):593-601.
269. Bhoomiboonchoo P, Gibbons RV, Huang A, Yoon I, Buddhari D, Nisalak A, et al. The Spatial Dynamics of Dengue Virus in Kamphaeng Phet, Thailand. PLoS Neglected Tropical Diseases. 2014;8(9):e3138.
270. Kramer AM, Pulliam JT, Alexander LW, Park AW, Rohani P, Drake JM. Spatial spread of the West Africa Ebola epidemic. R Soc open sci. 2016;3:160294.
271. Charu V, Zeger S, Bjørnstad ON, Kissler S, Simonsen L, Grenfell BT, et al. Human Mobility and the Spatial Transmission of Influenza in the United States. PLoS Computational Biology. 2017;13(2):e1005382.
272. Eggo RM, Cauchemez S, Ferguson NM. Spatial dynamics of the 1918 influenza pandemic in England, Wales and the United States. Journal of The Royal Society Interface. 2011;8(55):233-43.
273. Gog JR, Ballesteros S, Viboud C, Simonsen L, Bjornstad ON, Shaman J, et al. Spatial Transmission of 2009 Pandemic Influenza in the US. PLoS Computational Biology. 2014;10(6):e1003635.

274. Viboud C, Bjørnstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT. Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza. Science. 2006;312(5772):447-51.
275. Kraemer MUG, Faria NR, Reiner RC, Golding N, Nikolay B, Stasse S, et al. Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: a modelling study. Lancet Infect Dis. 2017;17:330-38.
276. Barrios JM, Verstraeten WW, Maes P, Aerts J, Farifteh J, Coppin P. Using the Gravity Model to Estimate the Spatial Spread of Vector-Borne Diseases. International Journal of Environmental Research and Public Health. 2012;9(12):4346-64.
277. Okubo A. Diffusion and ecological problem: mathematical models. Berlin: Springer; 1980.
278. Huff DL, Jenks GF. A Graphic Interpretation of the Friction of Distance in Gravity Models. Annals of the Association of American Geographers. 1968;58(4):814-24.
279. Chis Ster I, Ferguson NM. Transmission Parameters of the 2001 Foot and Mouth Epidemic in Great Britain. PloS ONE. 2007;2(6):e502.
280. Ferguson NM, Cummings DAT, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. Nature. 2006;442(27):448-52.
281. Fotheringham A. Spatial flows and spatial patterns. Environment and Planning A. 1984;16:529-43.
282. Siraj AS, Sorichetta A, España G, Tatem AJ, Perkins TA. Modeling human migration across spatial scales in Colombia. PLoS ONE. 2020;15(5):e0232702.
283. Simini F, González M, Maritan A, Barabási A. A universal model for mobility and migration patterns. Nature. 2012;484(7392):96-100.
284. Bjørnstad ON, Grenfell BT, Viboud C, King AA. Comparison of alternative models of human movement and the spread of disease. bioRxiv. 2019.
285. Stouffer S. Intervening Opportunities: A Theory Relating Mobility and Distance. American Sociological Review. 1940;5(6):845-67.
286. Kraemer MUG, Golding N, Bisanzio D, Bhatt S, Pigott DM, Ray S, et al. Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. Sci Rep. 2019;9:5151.
287. Roche B, Gaillard B, Léger L, Pélagie-Moutenda R, Sochacki T, Cazelles B, et al. An ecological and digital epidemiology analysis on the role of human behavior on the 2014 Chikungunya outbreak in Martinique. Nature Scientific Reports. 2017;7(5967):1-8.
288. Chadsuthi S, Althouse B, Iamsirithaworn S, Triampo W, Grantz K, Cummings DAT. Travel distance and human movement predict paths of emergence and spatial spread of chikungunya in Thailand. Epidemiology and Infection. 2018;146(13):1654-62.
289. Gardner LM, Bóta A, Gangavarapu K, Kraemer MUG, Grubaugh ND. Inferring the Risk Factors Behind the Geographical Spread and Transmission of Zika in the Americas. PLoS Neglected Tropical Diseases. 2018;12(1):e0006194.
290. Prem K, Lau MSY, Tam CC, Ho MZJ, Ng L, Cook AR. Inferring who-infected-whom-where in the 2016 Zika outbreak in Singapore—a spatio-temporal model. J R Soc Interface. 2019;16:20180604.
291. Rossi G, Karki S, Smith RL, Brown WM, Ruiz MO. The spread of mosquito-borne viruses in modern times: A spatio-temporal analysis of dengue and chikungunya. Spatial and Spatio-temporal Epidemiology. 2018;26:113-25.

292. Nsoesie EO, Ricketts RP, Brown HE, Fish D, Durham DP, Mbah MLN, et al. Spatial and Temporal Clustering of Chikungunya Virus Transmission in Dominica. PLoS Neglected Tropical Diseases. 2015;9:e0003977.
293. Lizarazo E, Vincenti-Gonzalez M, Grillet ME, Bethencourt S, Diaz O, Ojeda N, et al. Spatial Dynamics of Chikungunya Virus, Venezuela, 2014. Emerging Infectious Diseases. 2019;25(4):672-80.
294. Bisanzio D, Dzul-Manzanilla F, Gomez-Dantés H, Pavia-Ruz N, Hladish TJ, Lenhart A, et al. Spatio-temporal Coherence of Dengue, Chikungunya and Zika Outbreaks in Merida, Mexico. PLoS Neglected Tropical Diseases. 2018;12(3).
295. Dalvi APR, Braga JU. Spatial diffusion of the 2015–2016 Zika, dengue and chikungunya epidemics in Rio de Janeiro Municipality, Brazil. Epidemiology and Infection. 2019;147:1-13.
296. Rees EE, Petukhova T, Mascarenhas M, Pelcat Y, Ogden NH. Environmental and social determinants of population vulnerability to Zika virus emergence at the local scale. Parasites & Vectors. 2018;11:290.
297. Perkins TA, Rodriguez-Barraquer I, Manore C, Siraj AS, España G, Barker CM, et al. Heterogeneous local dynamics revealed by classification analysis of spatially disaggregated time series data. Epidemics. 2019;29:100357.
298. Flórez-Lozano K, Navarro-Lechuga E, Llinás-Solano H, Tuesca-Molina R, Sisa-Camargo A, Mercado-Reyes M, et al. Spatial distribution of the relative risk of Zika virus disease in Colombia during the 2015–2016 epidemic from a Bayesian approach. Int J Gynecol Obstet. 2020;148:55-60.
299. Martínez-Bello DA, López-Quílez A, Prieto AT. Spatio-Temporal Modeling of Zika and Dengue Infections within Colombia. Int J Environ Res Public Health. 2018;15:1376.
300. McHale TC, Romero-Vivas CM, Fronterre C, Arango-Padilla P, Waterlow NR, Nix CD, et al. Spatiotemporal Heterogeneity in the Distribution of Chikungunya and Zika Virus Case Incidences during their 2014 to 2016 Epidemics in Barranquilla, Colombia. Int J Environ Res Public Health. 2019;16:1759.
301. Wallace JR. R package Imap. 1.32 ed2010.
302. Weiss DJ, Nelson A, Gibson HS, Temperly WH, Peedell S. A global map of travel time to cities to assess inequalities in accessibility in 2015. Nature. 2018;553:333-6.
303. van Etten J. R package gdistance: Distances and Routes on Geographical Grids. 1.2-2 ed2018.
304. Jarvis A, Reuter HI, Nelson A, Guevara E. Hole-filled SRTM for the globe Version 4: CGIAR-CSI SRTM 90m Database; 2008 [Available from: https://cgiarcsi.community/data/srtm-90m-digital-elevation-database-v4-1/.
305. DANE. Medida de pobreza multidimensional municipal de fuente censal 2018 Bogotá 2020 [Available from: https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/pobreza-y-desigualdad/medida-de-pobreza-multidimensional-de-fuente-censal.
306. Salje H, Lessler J, Berry IM, Melendrez MC, Endy T, Kalayanarooj S, et al. Dengue diversity across spatial and temporal scales: Local structure and the effect of host population size. Science. 2017;355:1302-6.
307. Caminade C, Turner J, Metelmann S, Hesson JC, Blagrove MSC, Solomon T, et al. Global Risk Model for Vector-borne Transmission of Zika Virus Reveals the Role of El Niño 2015. PNAS. 2017;114(1):119-24.

308. Lippi CA, Stewart-Ibarra AM, Franklin Bajaña Loor ME, Dueñas Zambrano JE, Espinoza Lopez NA, Blackburn JK, et al. Geographic shifts in Aedes aegypti habitat suitability in Ecuador using larval surveillance data and ecological niche modeling: Implications of climate change for public health vector control. PLoS Neglected Tropical Diseases. 2019;13:e0007322.
309. Suarez MF, Nelson MJ. Registro de Altitud de Aedes Aegypti en Colombia. Biomedica. 1981;1(4):225.
310. Black A, Moncla LH, Laiton-Donato K, Potter B, Pardo L, Rico A, et al. Genomic epidemiology supports multiple introductions and cryptic transmission of Zika virus in Colombia. BMC Infectious Diseases. 2019;19:963.
311. Oliveira J, Rodrigues M, Skalinski L, Santos A, Costa L, Cardim L, et al. Interdependence between confirmed and discarded cases of dengue, chikungunya and Zika viruses in Brazil: A multivariate time-series analysis. PLoS ONE. 2020;15(2):e0228347.
312. Gordon A, Gresh L, Ojeda S, Katzelnick LC, Sanchez N, Mercado JC, et al. Prior dengue virus infection and risk of Zika: A pediatric cohort in Nicaragua. PLoS Med. 2019;16(1):e1002726.
313. Villero-Wolf Y, Mattar S, Puerta-González A, Arrieta G, Muskus C, Hoyos R, et al. Genomic epidemiology of Chikungunya virus in Colombia reveals genetic variability of strains and multiple geographic introductions in outbreak, 2014. Nature Scientific Reports. 2019;9(9970).
314. Hierlihy C, Waddell L, Young I, Greig J, Corrin T, Mascarenhas M. A systematic review of individual and community mitigation measures for prevention and control of chikungunya virus. PLoS ONE. 2019;14(2):e0212054.
315. Mendoza C, Jaramillo G, Ant T, Power G, Jones R, Quintero J, et al. An investigation into the knowledge, perceptions and role of personal protective technologies in Zika prevention in Colombia. PLoS Neglected Tropical Diseases. 2020;14(1):e0007970.
316. Gestantes con zika deben catalogarse como embarazos de alto riesgo [press release]. Bogotá: Ministerio de Salud y Protección Social, January 6, 2016.
317. Joseph A. Colombia declares end of Zika epidemic, as other experts urge caution 2016 May 5, 2021. Available from: https://www.statnews.com/2016/07/25/colombia-zika-virus-epidemic/.
318. Ferguson NM. Challenges and opportunities in controlling mosquito-borne infections. Nature. 2018;559:490-7.
319. SISPRO Bogotá, Colombia 2021 [Available from: https://www.sispro.gov.co.
320. Farrington C, Andrews N, Beale A, Catchpole M. A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease. Journal of the Royal Statistical Society Series A (Statistics in Society). 1996;159(3):547-63.
321. Angelo J, Fuller T, Leandro B, Praça H, Marques R, Ferreira J, et al. Neurological complications associated with emerging viruses in Brazil. Int J Gynecol Obstet. 2020;148:70-5.
322. Xavier J, Giovanetti M, Fonseca V, Thézé J, Gräf T, Fabri A, et al. Circulation of chikungunya virus East/Central/South African lineage in Rio de Janeiro, Brazil. PLoS ONE. 2019;14(6):e0217871.
323. Grubaugh ND, Faria NR, Andersen K, Pybus O. Genomic Insights into Zika Virus Emergence and Spread. Cell. 2018;172(6):1160-2.

324. U.S. Centers for Disease Control and Prevention. SARS-CoV-2 Variant Classifications and Definitions Atlanta, GA, USA 2021 [Available from: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html#Consequence.
325. Brady O, Wilder-Smith A. What Is the Impact of Lockdowns on Dengue? Current Infectious Disease Reports. 2021;23:2.
326. Lim J, Dickens B, Chew L, Choo E, Koo J, Aik J, et al. Impact of SARS-CoV-2 interventions of dengue transmission. PLoS Neglected Tropical Diseases. 2020;14(10):e0008719.
327. Cavany S, España G, Vazquez-Prokopec GM, Scott TW, Perkins T. The impacts of COVID-19 mitigation on dengue virus transmission: a modelling study. medRxiv. 2020.
328. Pan American Health Organization. Epidemiological Update: Arboviruses in the context of COVID-19. July 2, 2021.
329. Moise I, Ortiz-Whittingham L, Omachonu V, Clark M, Xue R. Fighting mosquito bite during a crisis: capabilities of Florida mosquito control districts during the COVID-19 pandemic. BMC Public Health. 2021;21:687.
330. Mattar S, Alvis-Guzmán N, Garay E, Rivero R, García A, Botero Y, et al. Severe Acute Respiratory Syndrome Coronavirus 2 Seroprevalence Among Adults in a Tropical City of the Caribbean Area, Colombia: Are We Much Closer to Herd Immunity Than Developed Countries? . Open Forum Infectious Diseases. 2020;7(12):ofaa550.
331. Groot H. Estudios sobre virus transmitidos por artrópodos en Colombia. Rev Acad Colomb Cienc Ex Fis Nat. 1964;12:191-217.
332. Prías-Landínez E, Bernal-Cúbides C, de Torres S, Romero-León M. Encuesta serológica de virus transmitidos por artrópodos. Bogotá, Colombia Instituto Nacional de Salud; 1970.
333. 5 reasons why pandemics like COVID-19 are becoming more likely: Gavi; 2020 [Available from: https://www.gavi.org/vaccineswork/5-reasons-why-pandemics-like-covid-19-are-becoming-more-likely.
334. Jones K, Patel N, Levy M, Storeygard A, Balk D, Gittleman J, et al. Global trends in emerging infectious diseases. Nature. 2008;451:990-3.
335. Allen T, Murray K, Zambrana-Torrelio C, Morse S, Rondinini C, Di Marco M, et al. Global hotspots and correlates of emerging zoonotic diseases. Nature Communications. 2017;8:1124.
336. Chretien J, Burkom H, Sedyaningsih E, Larasati R, Lescano A, Mundaca C, et al. Syndromic Surveillance: Adapting Innovations to Developing Settings. PLoS Med. 2008;5(3):e72.
337. Kelly T, Machalaba C, Karesh W, Crook P, Gilardi K, Nziza J, et al. Implementing One Health approaches to confront emerging and re-emerging zoonotic disease threats: lessons from PREDICT. One Health Outlook. 2020;2:1.
338. Kurpiers L, Schulte-Herbrüggen B, Ejotre I, Reeder D. Bushmeat and Emerging Infectious Diseases: Lessons from Africa. Problematic Wildlife. 2015:507-51.
339. Islam M, Sazzad H, Satter S, Sultana S, Hossain M, Hasan M, et al. Nipah Virus Transmission from Bats to Humans Associated with Drinking Traditional Liquor Made from Date Palm Sap, Bangladesh, 2011–2014. Emerging Infectious Diseases. 2016;22(4):664-70.

340. Yozwiak N, Happi C, Grant D, Schieffelin J, Garry R, Sabeti P, et al. Roots, Not Parachutes: Research Collaborations Combat Outbreaks. Cell. 2016;166(1):5-8.
341. PLOS Computational Biology. Authorship 2021 [Available from: https://journals.plos.org/ploscompbiol/s/authorship.
342. Busse C, August E. How to Write and Publish a Research Paper for a Peer-Reviewed Journal. Journal of Cancer Education. 2020.
343. Pre-Publication Support Service. About PREPSS 2021 [Available from: https://sites.google.com/umich.edu/prepss.
344. R Epidemics Consortium. RECON 2021 [Available from: https://www.repidemicsconsortium.org.
345. R Epidemics Consortium. Events 2021 [Available from: https://www.repidemicsconsortium.org/events/.
346. Olivera Mesa D, Hogan A, Watson O, Charles G, Hauck K, Ghani A, et al. Report 43: Quantifying the impact of vaccine hesitancy in prolonging the need for Non-Pharmaceutical Interventions to control the COVID-19 pandemic. Imperial College London. March 24, 2021.
347. World Health Organization. Ten threats to global health in 2019. 2018 [Available from: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019.
348. Rasolt D. Mistrust fuels covid-19 vaccine doubts in Colombia's Indigenous groups. 2021; (3327). Available from: https://institutions-newscientist-com.iclibezp1.cc.ic.ac.uk/article/2272195-mistrust-fuels-covid-19-vaccine-doubts-in-colombias-indigenous-groups/.
349. Campo-Arias A, Pedrozo-Pupo J. COVID-19 vaccine distrust in Colombian university students: Frequency and associated variables. medRxiv. 2021.
350. Anderson RM, May RM. Age-related changes in the rate of disease transmission: implications for the design of vaccination programmes. J Hyg (Lond). 1985;94(3):365-436.
351. Grubaugh ND, Saraf S, Gangavarapu K, Watts A, Tan A, Oidtman RJ, et al. Travel Surveillance and Genomics Uncover a Hidden Zika Outbreak during the Waning Epidemic. Cell. 2019;178:1057-71.
352. Zimmer C. Zika was soaring across Cuba. Few outside the country knew. May 6, 2021. Available from: https://www.nytimes.com/2019/08/22/science/cuba-zika-epidemic.html.
353. Pan American Health Organization. Epidemiological Update: Dengue and other Arboviruses. Washington, D.C. June 10, 2020.
354. Ryan SJ, Carlson C, Mordecai E, Johnson L. Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. PLoS Neglected Tropical Diseases. 2019;13(3):e0007213.

Appendix S1: Additional plots comparing EpiEstim R_ts and model R_ts

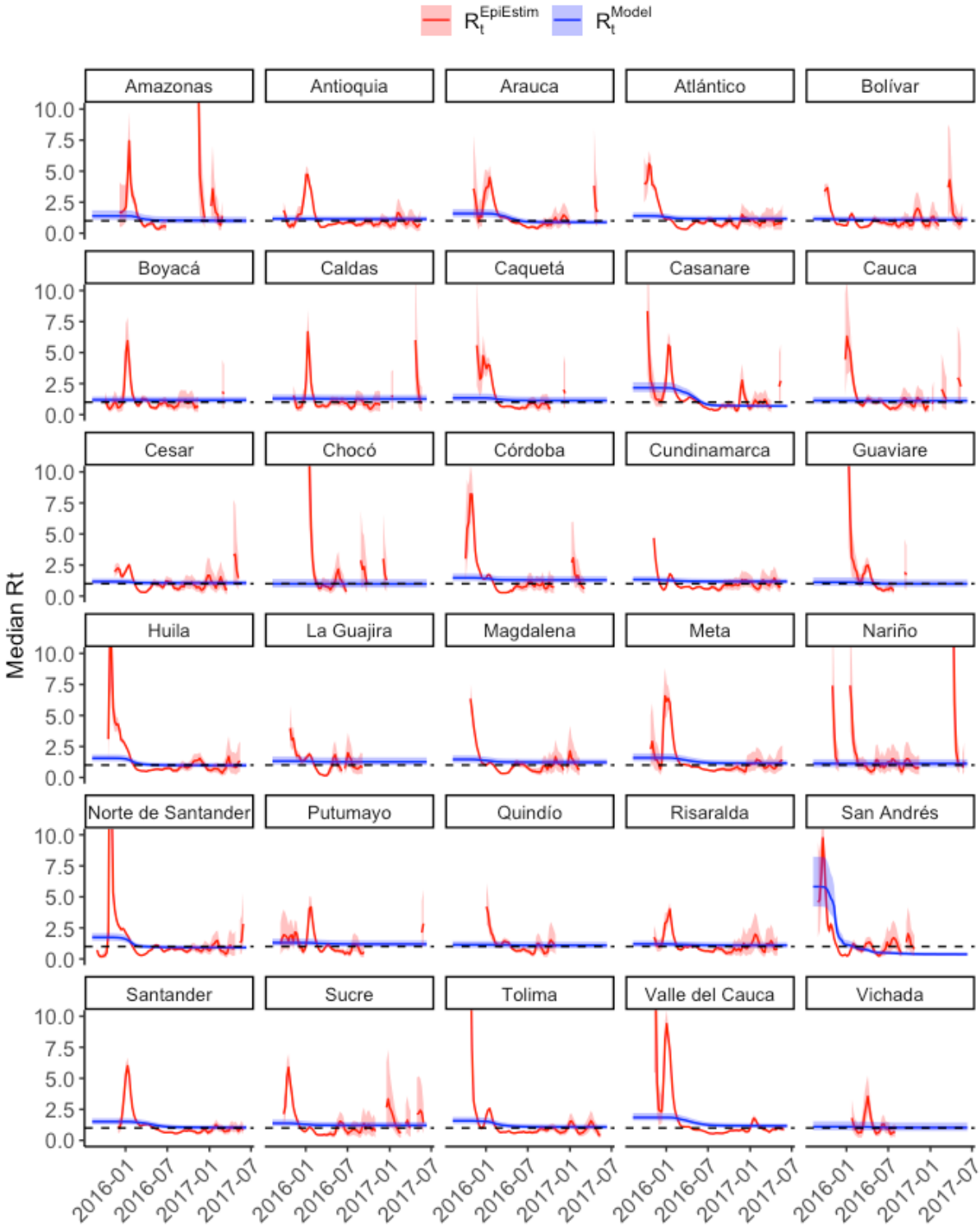


Figure A. Comparing median estimates of R_t from the best-fitting negative binomial model for ZIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines) using a different y-axis. Compared to Figure 3.18 in the main thesis text, the y-axes on these plots have an upper limit of 10 to show the full 95% CrI of the prediction for San Andrés and Providencia. EpiEstim R_ts are plotted in the center of the 5-week window used to compute each estimate. Shaded areas represent 95% CrI. There is a positive statistically significant correlation of 0.31 (Pearson's correlation coefficient, 95% CI: 0.27-0.35, $p < 0.0001$).

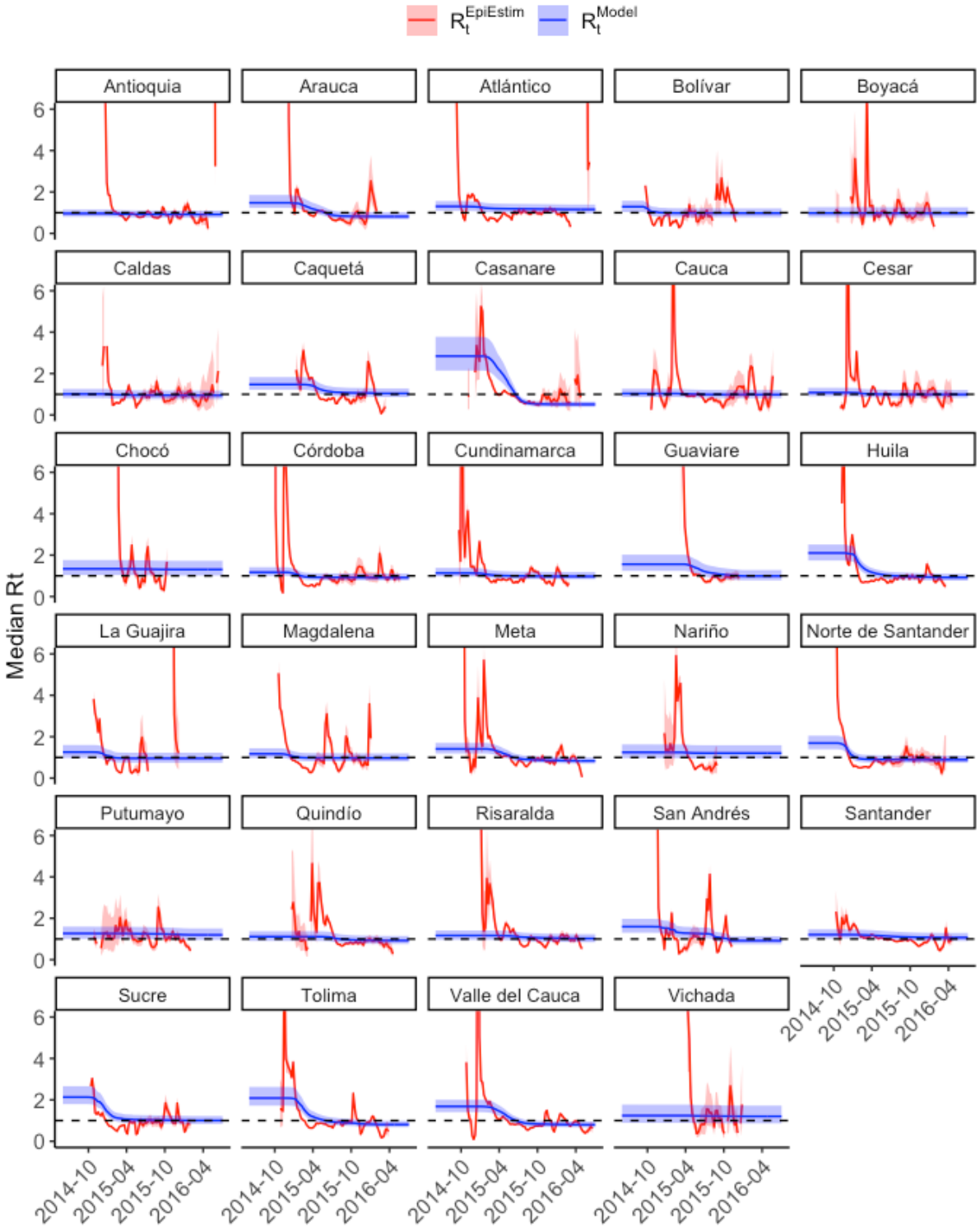


Figure B. Comparing median estimates of R_t from the best-fitting negative binomial model for CHIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines) using a threshold of 15. This figure is nearly indistinguishable from Figure 3.17 in the main thesis text, so the higher threshold of 20 was preferred. EpiEstim R_ts are plotted in the center of the 5-week window used to compute each estimate. Shaded areas represent 95% CrI. There is a positive statistically significant correlation of 0.17 (Pearson's correlation coefficient, 95% CI: 0.12-0.21, $p < 0.0001$).

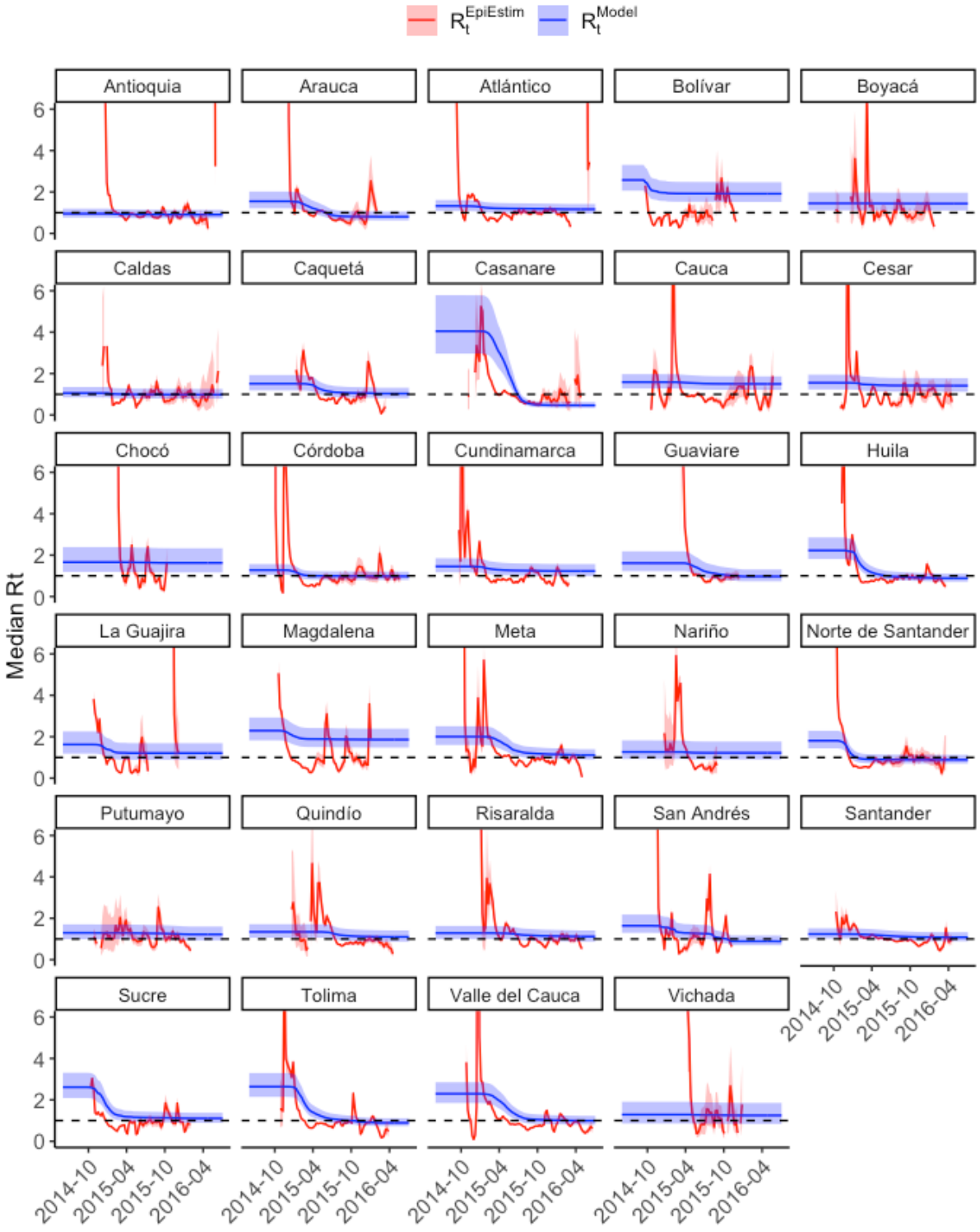


Figure C. Comparing median estimates of R_t from the best-fitting negative binomial model for CHIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines) using a threshold of 55. In contrast to Figure 3.17 in the main thesis text, the predicted values for some departments (especially Bolívar and Magdalena) do not cross 1 at the end of the epidemic. EpiEstim R_ts are plotted in the center of the 5-week window used to compute each estimate. Shaded areas represent 95% CrI. There is a positive statistically significant correlation of 0.12 (Pearson's correlation coefficient, 95% CI: 0.07-0.16, $p < 0.0001$).

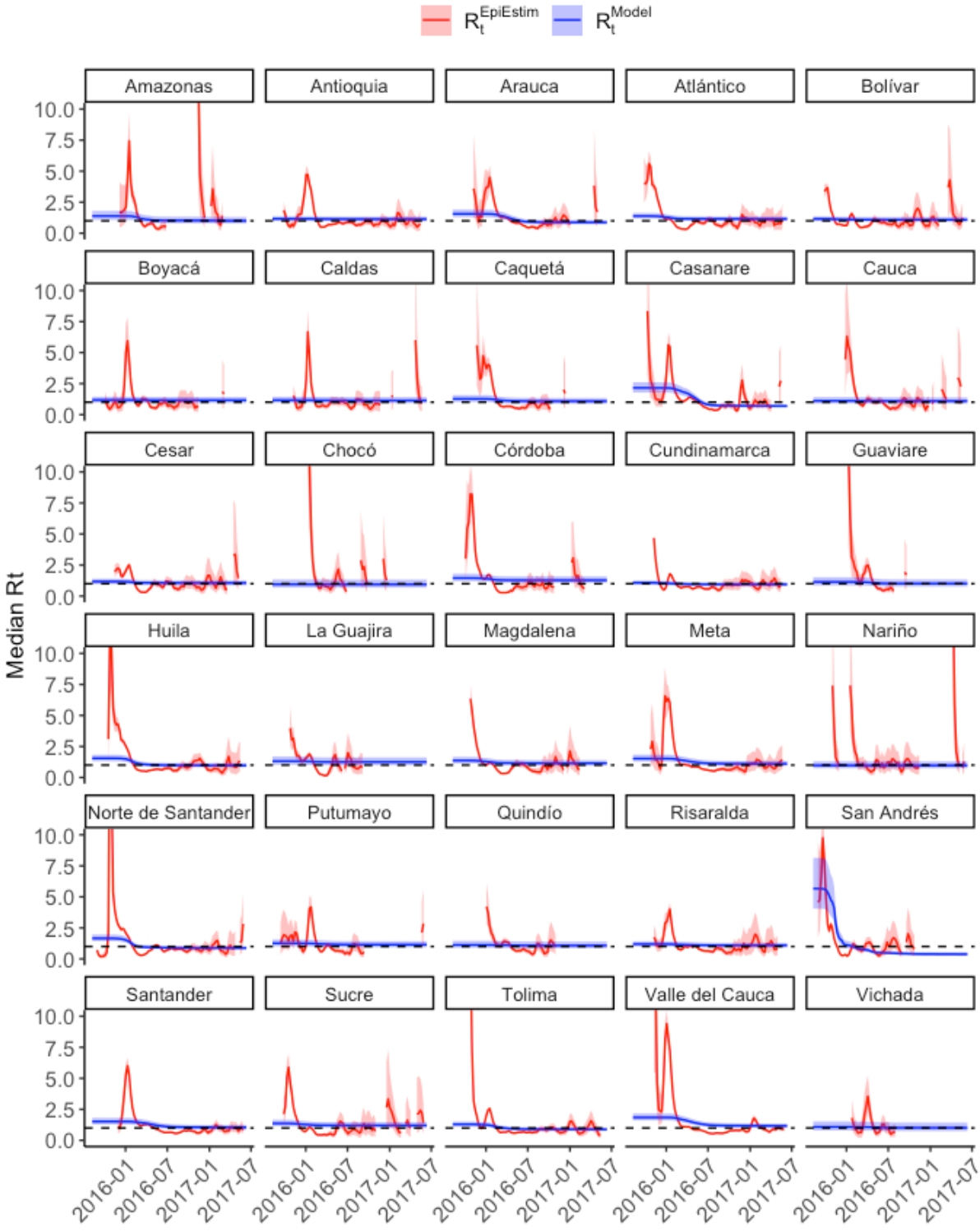


Figure D. Comparing median estimates of R_t from the best-fitting negative binomial model for ZIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines) using a threshold of 15. This figure is nearly indistinguishable from Figure A above, so the more conservative threshold of 20 was preferred. EpiEstim R_ts are plotted in the center of the 5-week window used to compute each estimate. Shaded areas represent 95% CrI. There is a positive statistically significant correlation of 0.31 (Pearson's correlation coefficient, 95% CI: 0.27-0.35, $p < 0.0001$).

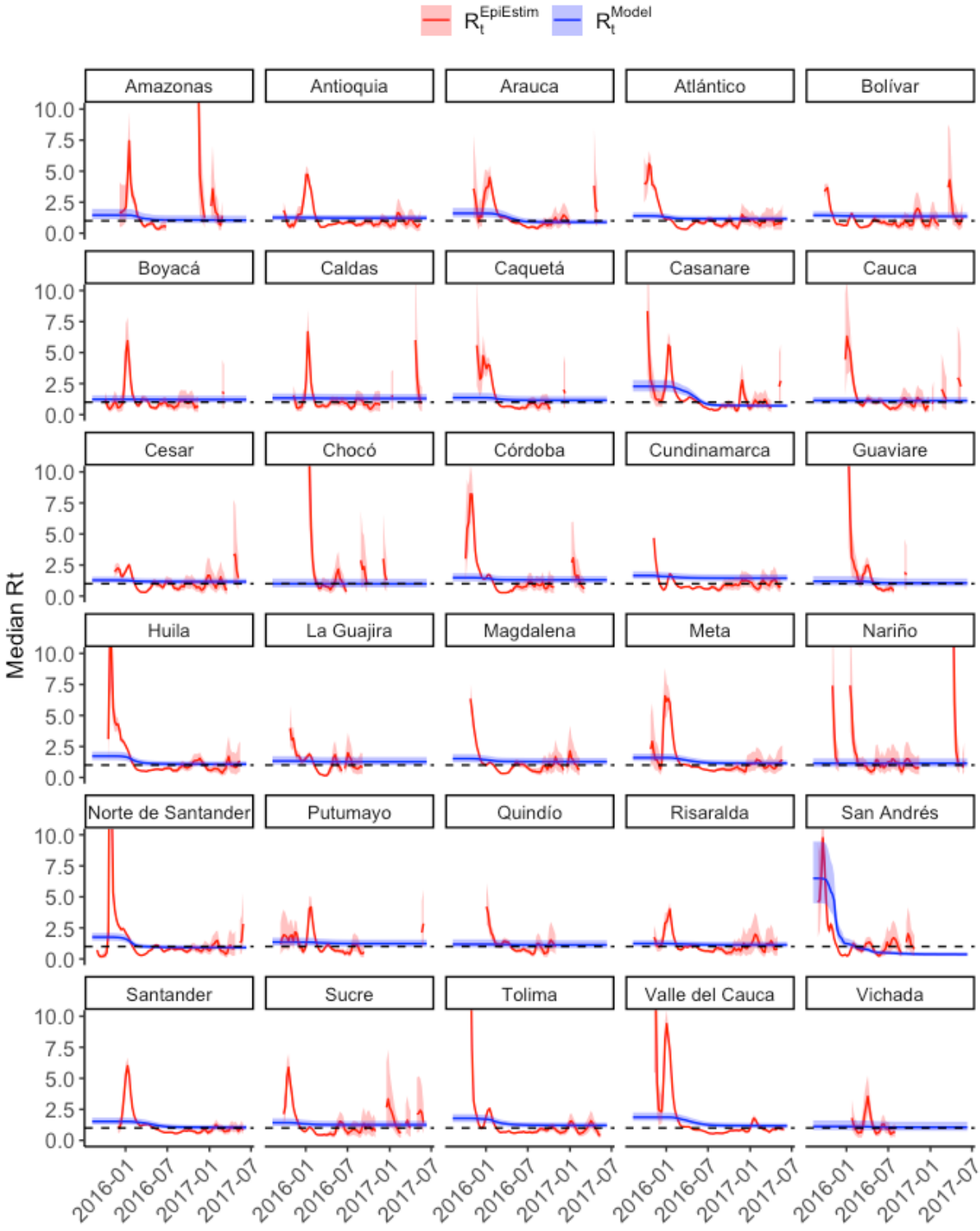


Figure E. Comparing median estimates of R_t from the best-fitting negative binomial model for ZIKV (R_t^{Model}, blue lines) with those obtained from EpiEstim ($R_t^{EpiEstim}$, red lines) using a threshold of 55. Compared to Figure A above, the predictions were slightly worse, with a higher R_t initially estimated for San Andrés and Providencia. Also, Bolívar and Cundinamarca have slightly higher final estimated R_ts, which are above 1. EpiEstim R_ts are plotted in the center of the 5-week window used to compute each estimate. Shaded areas represent 95% CrI. There is a positive statistically significant correlation of 0.29 (Pearson's correlation coefficient, 95% CI: 0.25-0.33, $p < 0.0001$).

Appendix S2: MCMC testing for best-fitting Poisson models

The diagnostics in this section correspond to each virus' best-fitting Poisson model from chapter 3.

Convergence diagnostics

The models were run from three different starting points to ascertain convergence. Table A shows the Gelman-Rubin statistic for each of the Poisson models with multiple R_0s and rainfall after removing the burn-in. All point estimates and 95% CI are at or about 1, suggesting model convergence. Figures A-B show the posterior distributions of one MCMC chain for each parameter after removing the burn-in. All the distributions are close to normal, suggesting that the chains converged.

Table A. Gelman-Rubin statistic for each of the best-fitting Poisson models (after removing the burn-in).

	CHIKV		ZIKV	
Parameter	**Point estimate**	**Upper CI**	**Point estimate**	**Upper CI**
ρ (reporting rate)	1	1	1	1
R_0 (all)	1	1-1.01	1	1
$_X best(rain)$	1.01	1.02	1	1
σ_{rain}	1.01	1.03	1	1

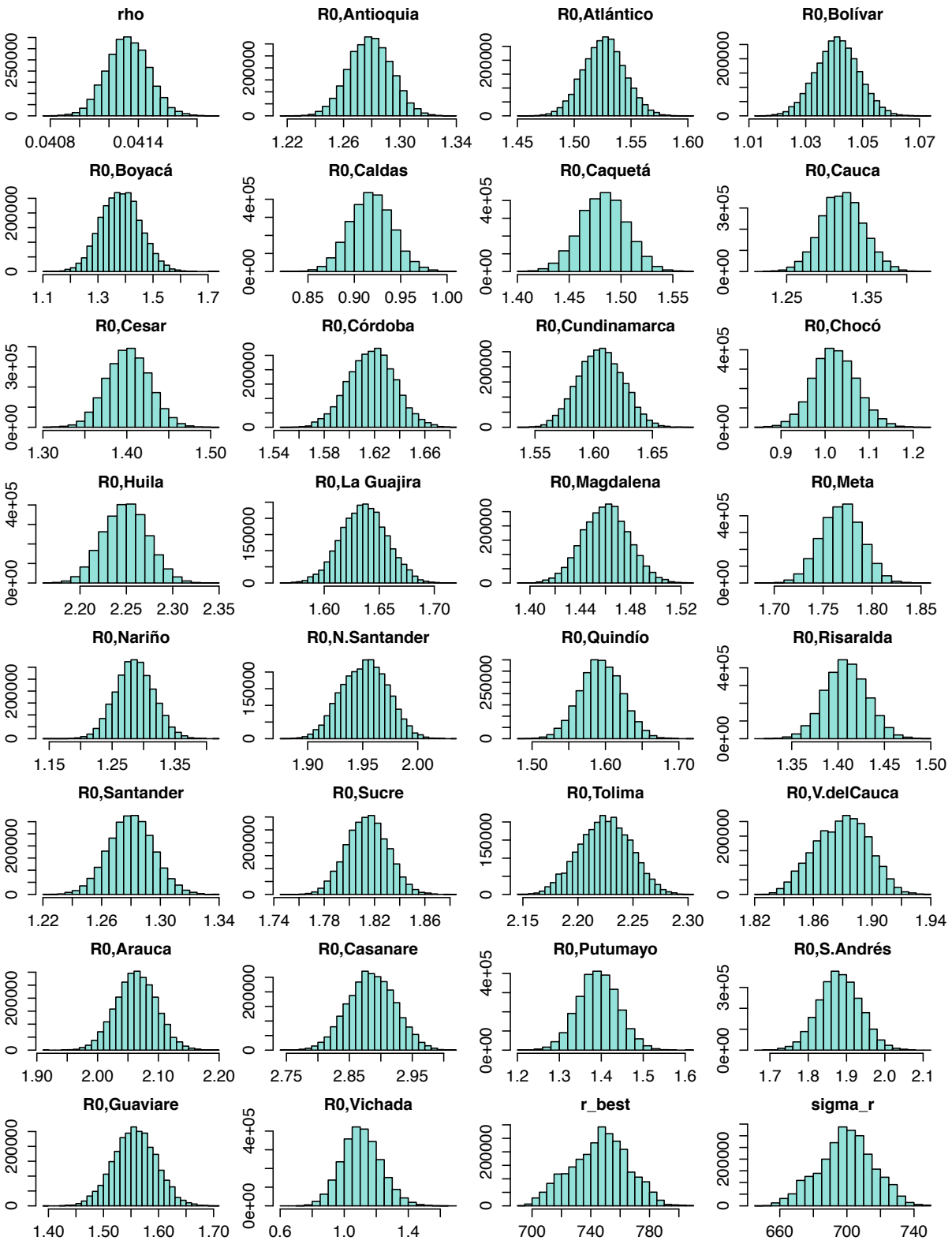


Figure A. Histograms of the posterior distributions of the best-fitting Poisson model for CHIKV.

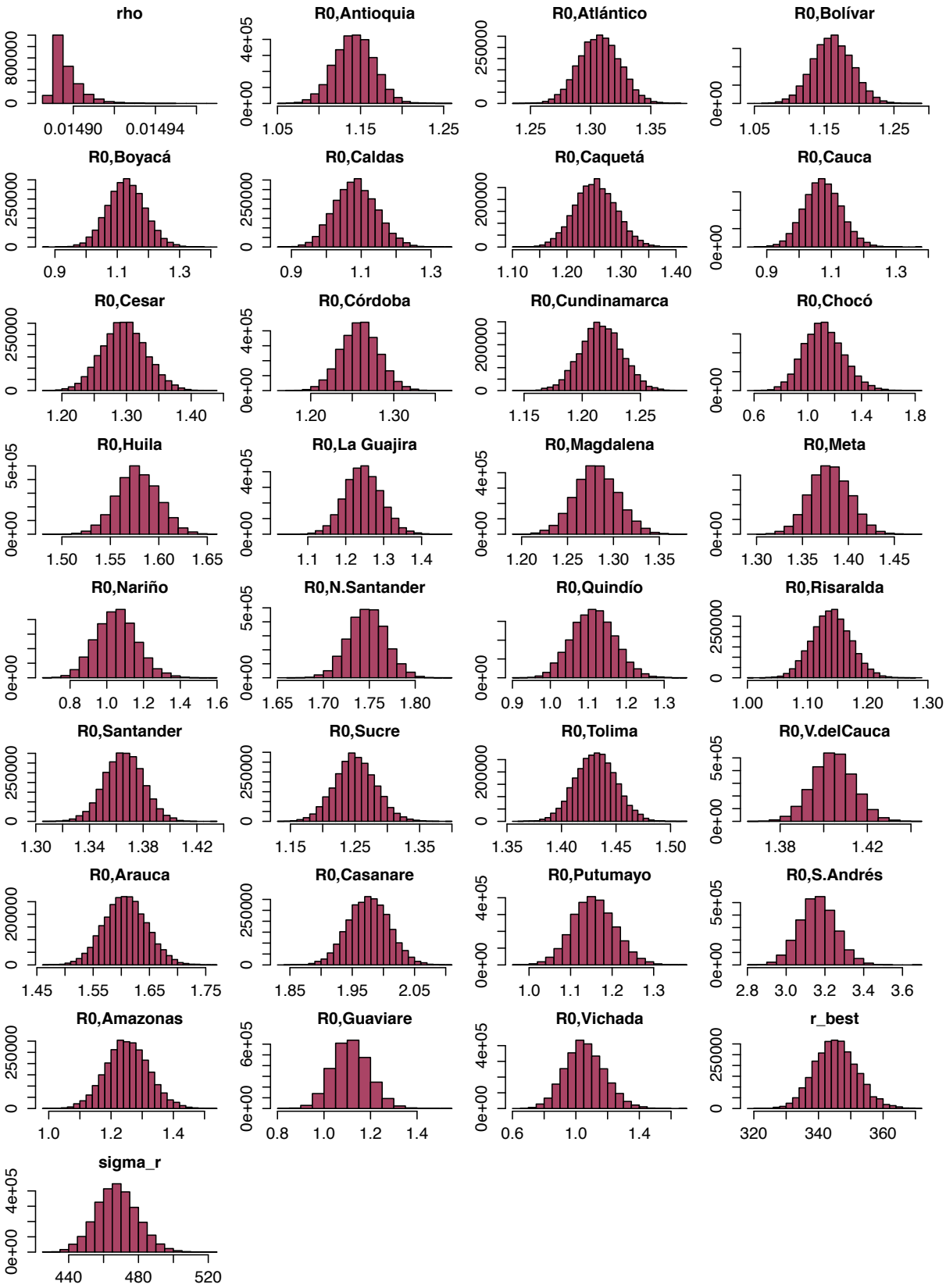


Figure B. Histograms of the posterior distributions of the best-fitting Poisson model for ZIKV.

Traces

Figures C-D show the MCMC traces for three chains of the CHIKV and ZIKV models, respectively. Although mixing is slow for the rainfall parameters in the CHIKV model (r_best and sigma_r), mixing is good for all other parameters based on visual assessment.

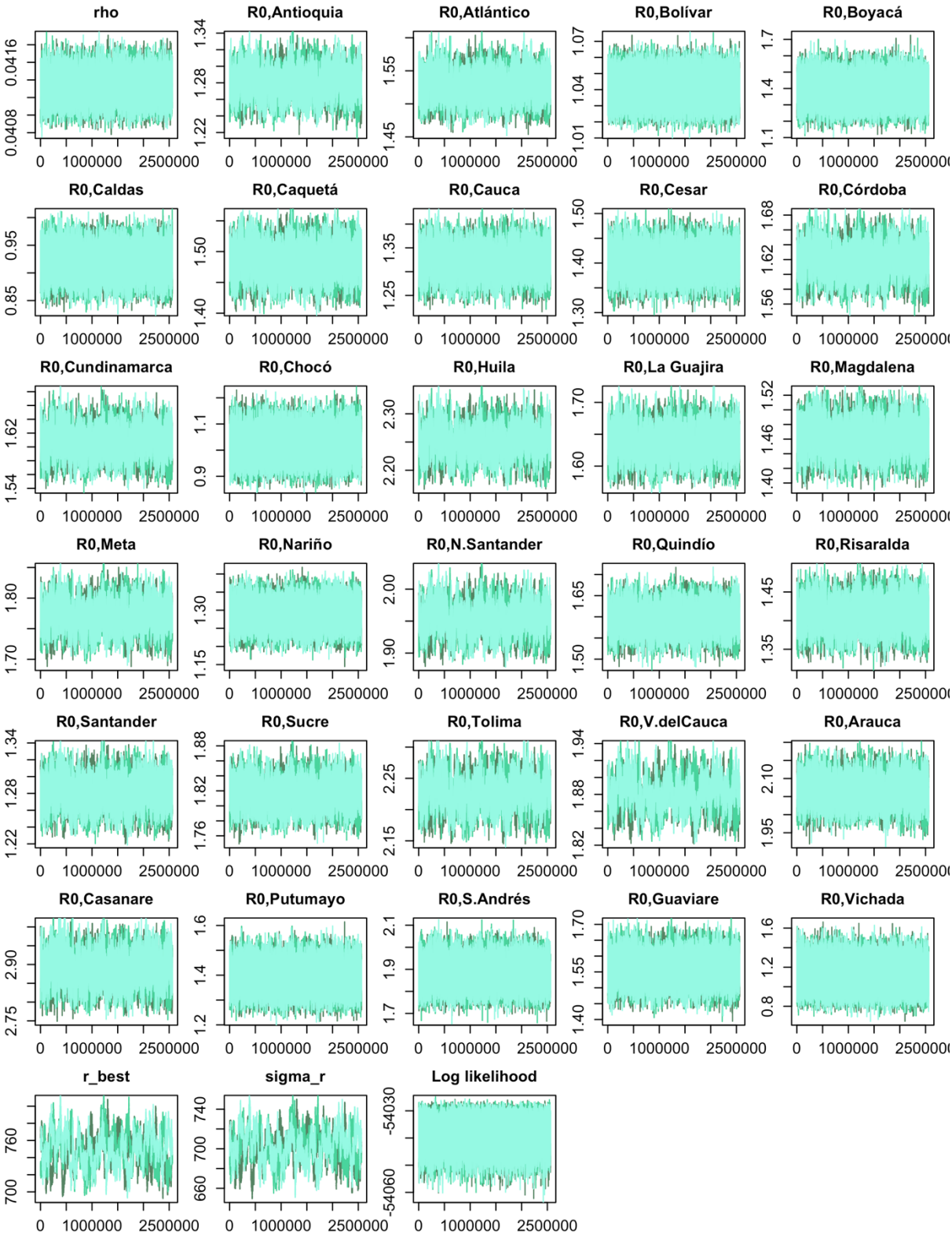


Figure C. MCMC traces for the CHIKV model. Three chains run using different start values are shown.

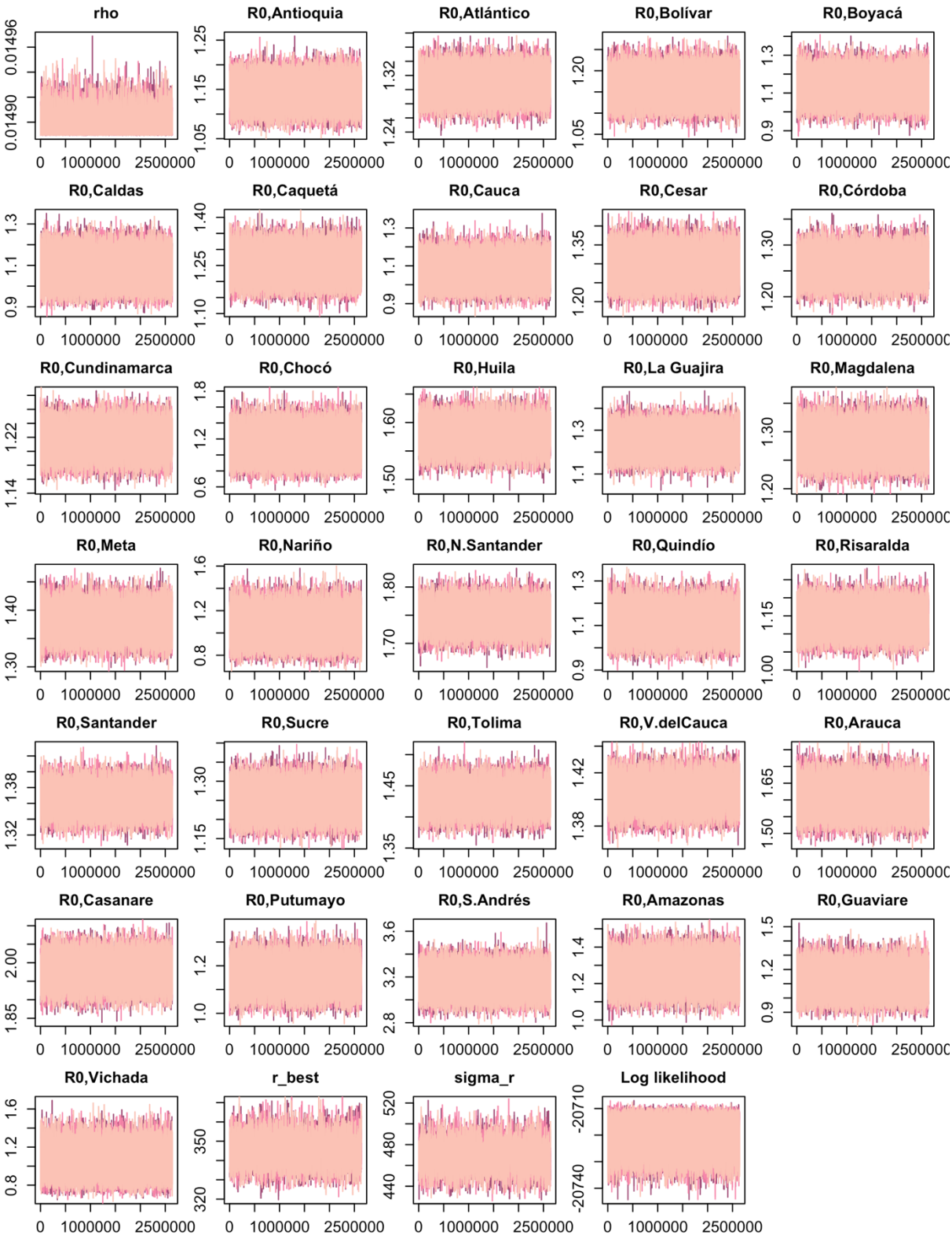


Figure D. MCMC traces for the ZIKV model. Three chains run using different start values are shown.

Acceptance rate and effective sample size

Table B shows the acceptance rate of parameters for the CHIKV and ZIKV models. Both models have good acceptance rates. Table C shows the calculation of the effective sample size for each parameter after removing the burn-in. Except for the rainfall parameters in the CHIKV model, all parameters have good effective sample sizes (several are at least 10% of the total number of iterations).

Table B. Acceptance percentages for parameters of the best-fitting Poisson models for CHIKV and ZIKV (after removing the burn-in).

Parameter	CHIKV	ZIKV
ρ (reporting rate)	20.4	18.5
Department-specific R_0, range	17.8-29.7	16.4-30.0
${}_X{}^{best}(rain)$	24.3	16.9
σ_{rain}	15.7	19.5

Table C. Effective sample sizes from one chain for each of the best-fitting Poisson models (after removing the burn-in).

Parameter	CHIKV	ZIKV
ρ (reporting rate)	5,520	6,873
Department-specific R_0, median (range)	7,038 (779-17,682)	13,600 (6,908-17,257)
${}_X{}^{best}(rain)$	57	1,579
σ_{rain}	85	1,425

Appendix S3: Data for sex and age models, additional sensitivity analyses, and Stan code for chapter 4

Table A. Epidemiological and demographic data for Colombian capital cities by sex.

City	Department	Sex	Population	Reported cases with ZIKV-associated NC*	Reported suspected and laboratory-confirmed cases of ZVD	Estimated post-epidemic seroprevalence and 95% CI
Arauca	Arauca	F M	45,119 44,593	0 1	554 234	
Armenia	Quindío	F M	154,267 143,932	0 0	149 40	
Barranquilla	Atlántico	F M	629,843 593,773	36 44	3,331 1,334	
Bucaramanga	Santander	F M	274,148 254,121	2 6	2,690 1,632	
Cali	Valle del Cauca	F M	1,250,077 1,144,848	11 12	10,072 6,207	
Cartagena	Bolívar	F M	523,353 490,036	2 2	604 417	
Cúcuta	Norte de Santander	F M	338,927 317,453	25 19	4,629 1,856	0.479 (0.440-0.519) 0.479 (0.440-0.519)
Florencia	Caquetá	F M	89,184 86,223	1 2	452 211	
Ibagué	Tolima	F M	287,445 271,360	1 2	2,603 1,473	
Inírida	Guainía	F M	9,721 10,262	0 0	7 5	
Leticia	Amazonas	F M	20,953 20,686	0 0	172 106	
Medellín	Antioquia	F M	1,316,499 1,170,224	3 5	340 209	0.067 (0.048-0.090) 0.067 (0.048-0.090)
Mitú	Vaupés	F M	15,911 15,950	0 0	11 6	
Mocoa	Putumayo	F M	21,864 21,018	0 1	37 20	
Montería	Córdoba	F M	230,424 217,244	1 3	1,318 467	
Neiva	Huila	F M	179,444 164,582	7 6	2,313 1,096	0.578 (0.538-0.618) 0.578 (0.538-0.618)
Pereira	Risaralda	F M	248,342 223,658	0 0	285 178	
Popayán	Cauca	F M	144,266 135,788	0 0	37 14	
Puerto Carreño	Vichada	F M	7,580 8,420	0 0	11 6	
Quibdó	Chocó	F M	57,832 58,075	0 0	9 5	
Riohacha	La Guajira	F M	136,434 132,278	0 0	204 75	

San Andrés	San Andrés & Providencia	F M	36,175 35,771	0 0	685 424	
San José del Guaviare	Guaviare	F M	32,177 33,434	0 0	93 61	
Santa Marta	Magdalena	F M	251,416 240,119	0 2	1,338 575	
Sincelejo	Sucre	F M	141,890 137,141	1 5	596 260	0.659 (0.620-0.696) 0.659 (0.620-0.696)
Valledupar	Cesar	F M	237,186 226,033	2 0	614 174	
Villavicencio	Meta	F M	254,901 240,326	3 2	1,659 718	
Yopal	Casanare	F M	71,401 71,578	1 4	1,367 754	

*Neurological complications

Table B. Epidemiological and demographic data for Colombian capital cities by age group.

City	Department	Age (in years)	Population	Reported cases with ZIKV-associated NC*	Reported suspected and laboratory-confirmed cases of ZVD	Estimated post-epidemic seroprevalence and 95% CI**
Arauca	Arauca	0-39 40+	66,884 22,828	0 1	572 216	
Armenia	Quindío	0-39 40+	180,155 118,044	0 0	123 66	
Barranquilla	Atlántico	0-39 40+	791,190 432,426	31 49	3,526 1,139	
Bucaramanga	Santander	0-39 40+	326,377 201,892	5 3	3,070 1,252	
Cali	Valle del Cauca	0-39 40+	1,513,100 881,825	8 15	11,226 5,053	
Cartagena	Bolívar	0-39 40+	673,499 339,890	2 2	749 272	
Cúcuta	Norte de Santander	0-39 40+	443,112 213,268	21 23	4,928 1,557	0.479 (0.440-0.519) 0.479 (0.440-0.519)
Florencia	Caquetá	0-39 40+	123,947 51,460	2 1	541 122	
Ibagué	Tolima	0-39 40+	354,249 204,556	0 3	2,942 1,134	
Inírida	Guainía	0-39 40+	15,447 4,536	0 0	8 4	
Leticia	Amazonas	0-39 40+	32,316 9,323	0 0	213 65	
Medellín	Antioquia	0-39 40+	1,388,651 1,098,072	3 5	409 140	0.067 (0.048-0.090) 0.067 (0.048-0.090)
Mitú	Vaupés	0-39 40+	24,379 7,482	0 0	16 1	
Mocoa	Putumayo	0-39 40+	30,881 12,001	0 1	43 14	
Montería	Córdoba	0-39 40+	309,496 138,172	2 2	1,415 370	
Neiva	Huila	0-39 40+	224,974 119,052	9 4	2,513 896	0.578 (0.538-0.618) 0.578 (0.538-0.618)
Pereira	Risaralda	0-39 40+	285,659 186,341	0 0	340 123	
Popayán	Cauca	0-39 40+	176,046 104,008	0 0	47 4	
Puerto Carreño	Vichada	0-39 40+	12,968 3,032	0 0	11 6	
Quibdó	Chocó	0-39 40+	91,447 24,460	0 0	9 5	
Riohacha	La Guajira	0-39 40+	206,522 62,190	0 0	217 62	
San Andrés	San Andrés & Providencia	0-39 40+	45,576 26,370	0 0	731 378	
San José del Guaviare	Guaviare	0-39 40+	50,430 15,181	0 0	113 41	

Santa Marta	Magdalena	0-39 40+	343,797 147,738	1 1	1,467 446	
Sincelejo	Sucre	0-39 40+	187,975 91,056	3 3	624 232	0.659 (0.620-0.696) 0.659 (0.620-0.696)
Valledupar	Cesar	0-39 40+	330,123 133,096	2 0	611 177	
Villavicencio	Meta	0-39 40+	338,247 156,980	4 1	1,847 530	
Yopal	Casanare	0-39 40+	103,370 39,609	4 1	1,692 429	

*Neurological complications

**The same post-epidemic seroprevalence was assumed for individuals aged 40 years and older as those under 40 years even though only individuals up to age 45 were sampled in the serological study. This assumption was relaxed in a sensitivity analysis.

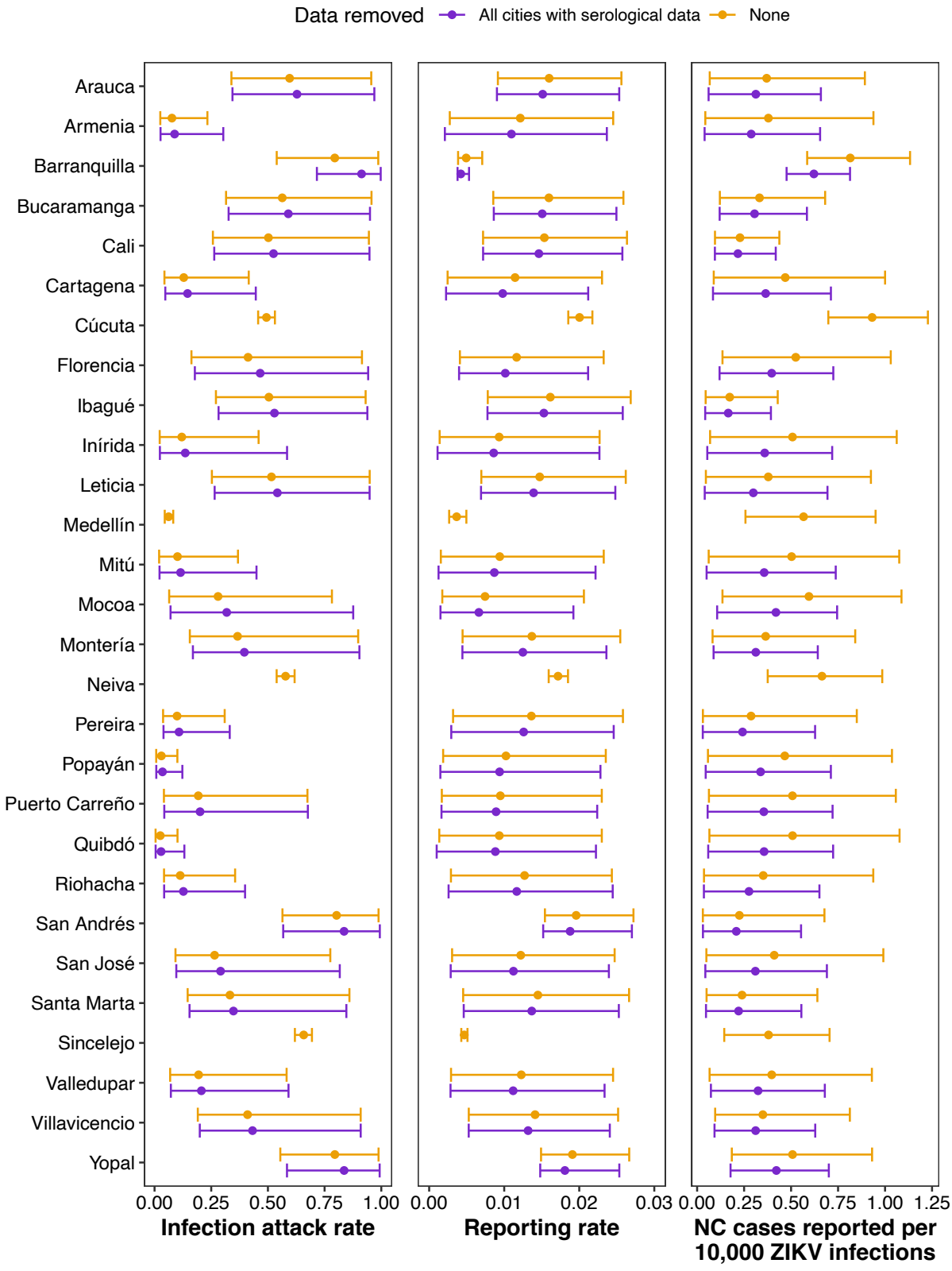


Figure A. Effect of removing all cities with serological data on estimated ZIKV infection attack rates, ZVD reporting rates, and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections. Posterior mean (points) and 95% credible interval (error bars) are shown for each city and overall.

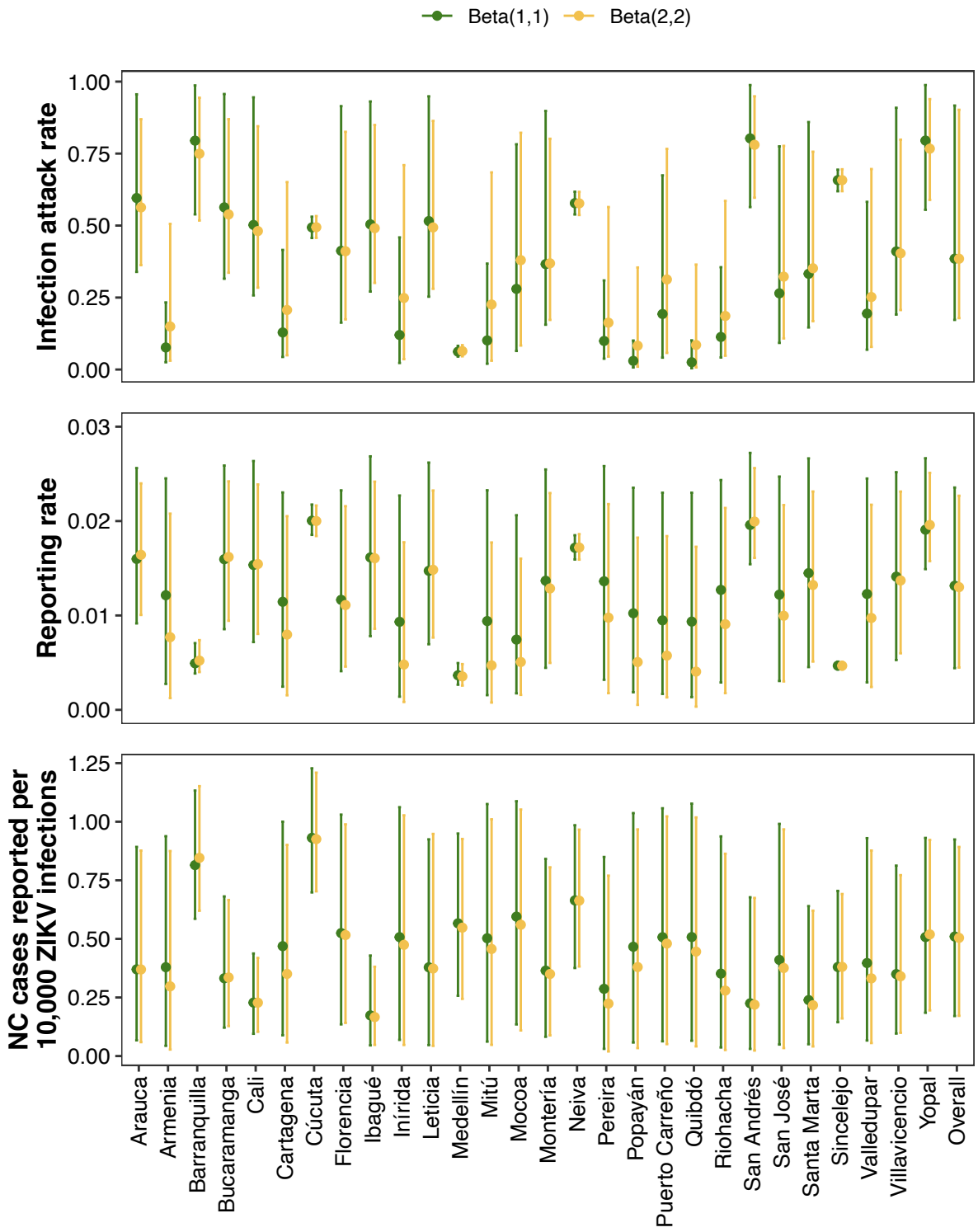


Figure B. Comparison of estimated ZIKV infection attack rates, ZVD reporting rates, and number of ZIKV-associated neurological complications (NC) cases reported per 10,000 ZIKV infections from models that used either Beta(1,1) or Beta(2,2) prior distributions for the ZIKV infection attack rates in cities without seroprevalence data. Posterior mean (points) and 95% credible interval (error bars) are shown for each city and overall.

Code A. Stan code for hierarchical model.

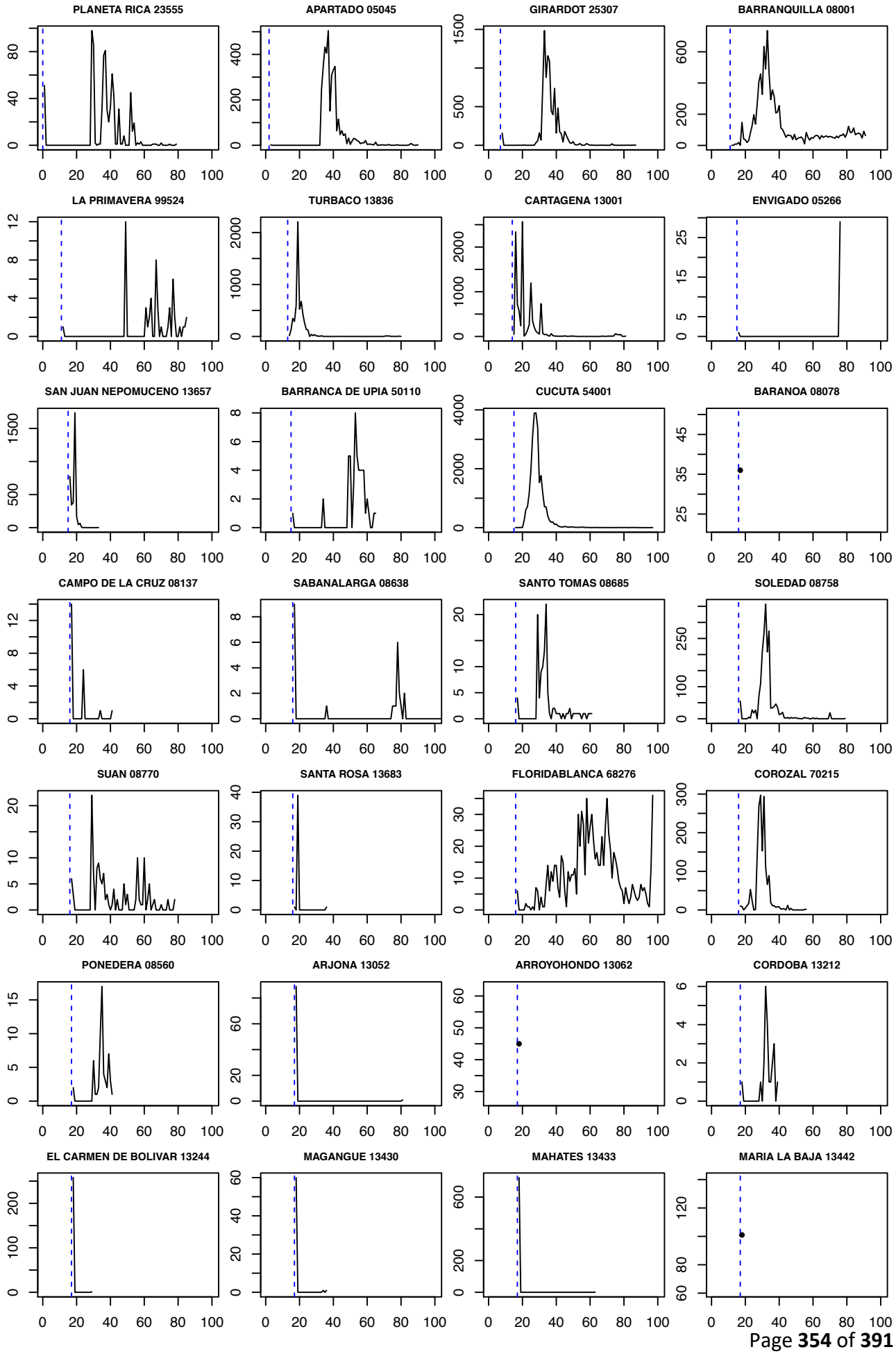
```
data{
  int<lower=0> l; //Number of cities
  vector[l] alphaZ; // From seroprevalence estimates
  vector[l] gammaZ; // From seroprevalence estimates
  int N[l];    //Population size in each city
  int S[l];   //Reported suspect ZVD cases in each city
  int NC[l]; //Reported suspect NC in each city
  int Sall; //Reported suspect ZVD across all cities
  int NCall; //Reported suspect NC across all cities
  int Nall; //total population across all cities
}

parameters{
  real<lower=0,upper=1> pNC_min;
  real<lower=pNC_min,upper=1> pNC_max;
  real<lower=0,upper=1> pS_min;
  real<lower=pS_min,upper=1> pS_max;
  real<lower=0,upper=1> pZ[l]; //Probability of ZIKV infection
  real<lower=pS_min,upper=pS_max> pS[l]; //Probability that a ZIKV
infection is reported as a case to the surveillance system
  real<lower=pNC_min,upper=pNC_max> pNC[l]; //Probability that a
ZIKV infection becomes reported as a case with NC
  real<lower=pS_min,upper=pS_max> pSall; //overall risk that a ZIKV
infection is reported as a case to the surveillance system
  real<lower=pNC_min,upper=pNC_max> pNCall; //overall risk that a
ZIKV infection becomes reported as a case with NC
  real<lower=0,upper=1> pZall;
}

model{

  //Hyperpriors
  pS_min ~ uniform(0,1);
  pS_max ~ uniform(pS_min, 1);
  pNC_min ~ uniform(0,1);
  pNC_max ~ uniform(pNC_min, 1);

  //Priors
  for(i in 1:l){
  pS[i] ~ uniform(pS_min, pS_max); // allows reporting rate of
symptomatic ZIKV infection to vary by city
  pNC[i] ~ uniform(pNC_min, pNC_max); // allows reporting rate of NC
to vary by city
  pZ[i] ~ beta(alphaZ[i], gammaZ[i]); // representing possible
ranges of attack rates from seroprevalence study
  }

  pSall ~ uniform(pS_min, pS_max); //overall risk
  pNCall ~ uniform(pNC_min, pNC_max); //overall risk
  pZall ~ beta(1, 1);

  //Model
  for (i in 1:l){
```

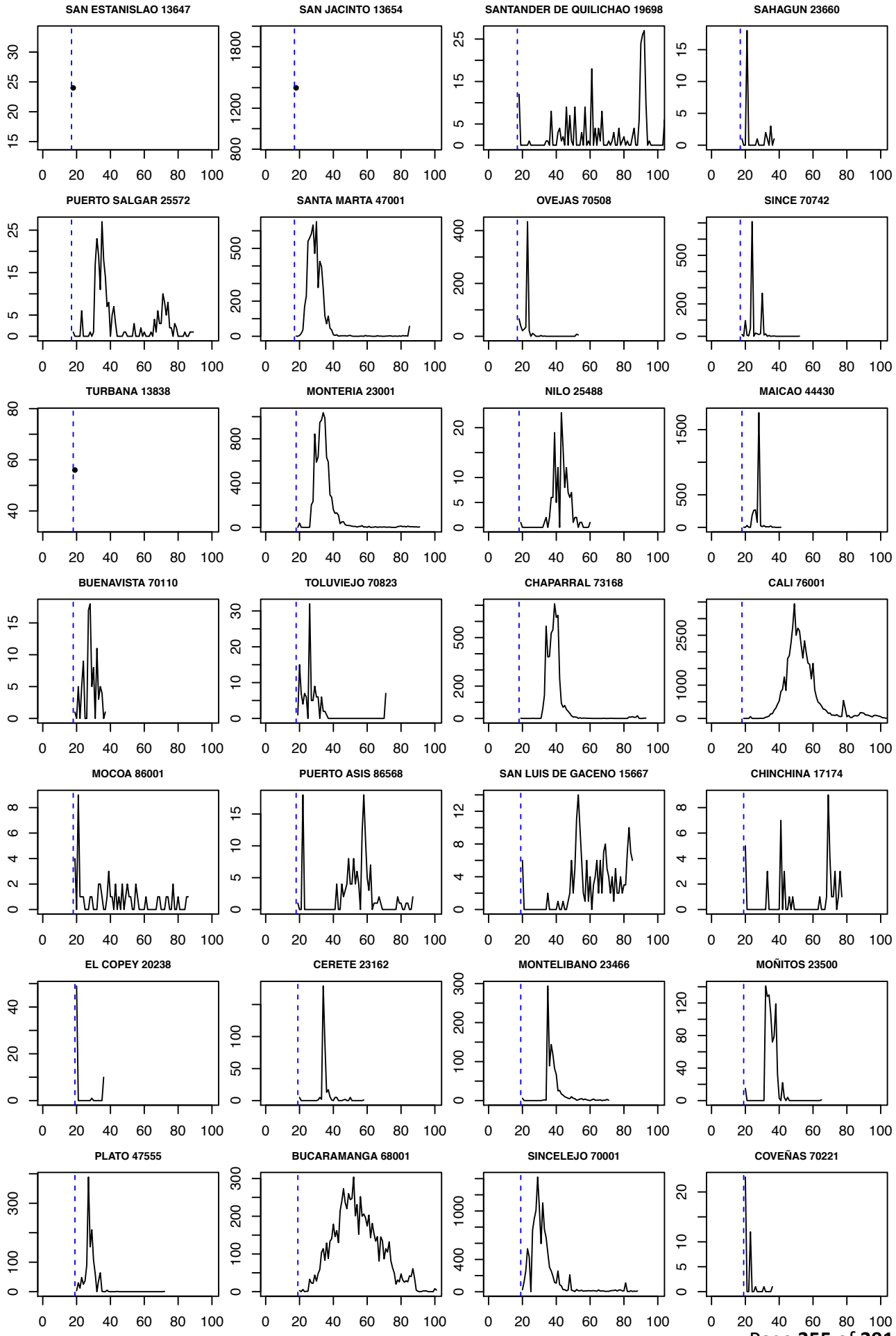
```
    S[i] ~ binomial(N[i], pZ[i]*pS[i]); //ZIKV infections that give
rise to suspected reported cases
    NC[i] ~ binomial(N[i], pNC[i]*pZ[i]); //NC
  }
  Sall ~ binomial(Nall, pSall*pZall);
  NCall ~ binomial(Nall, pNCall*pZall);
}
```

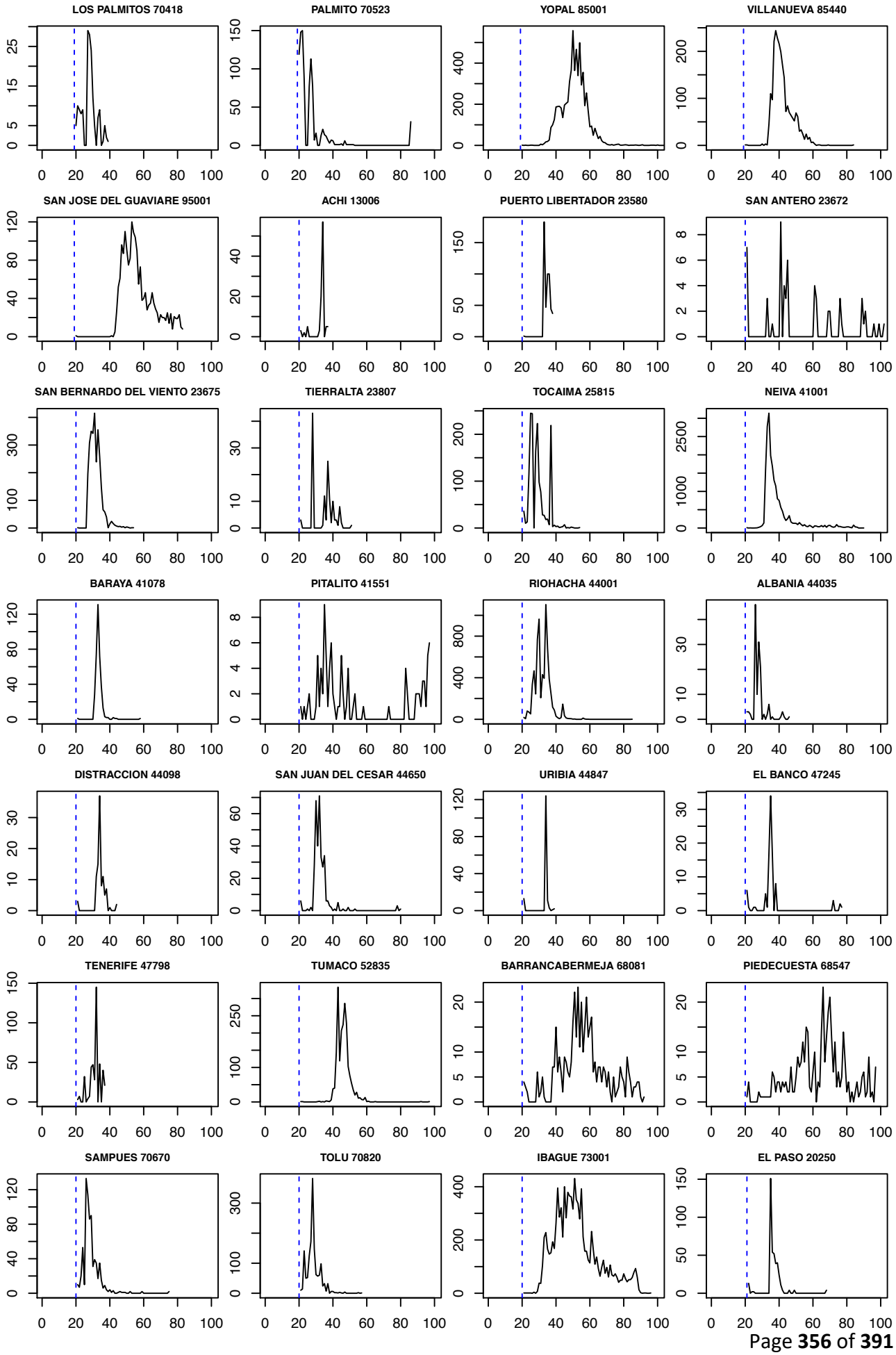
Appendix S4: Week of invasion

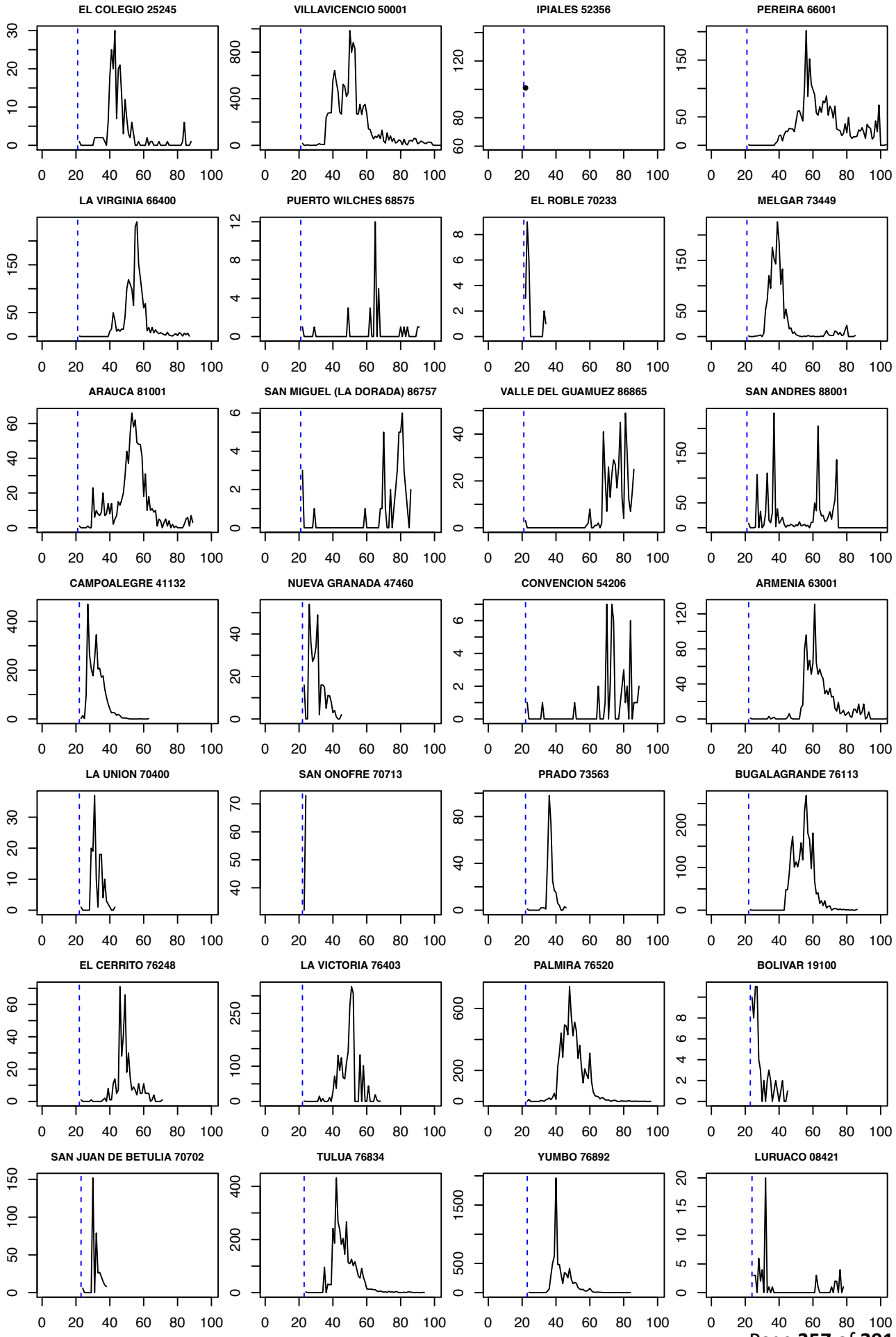
Invasion week, represented by a blue dashed line, is shown for all cities using a method based on the first reported cases for CHIKV and ZIKV, respectively. The y-axis is the number of reported cases, and the x-axis is weeks. Cities were sorted in ascending order by invasion week, and a point instead of a line was plotted for cities that reported all cases in a single week.

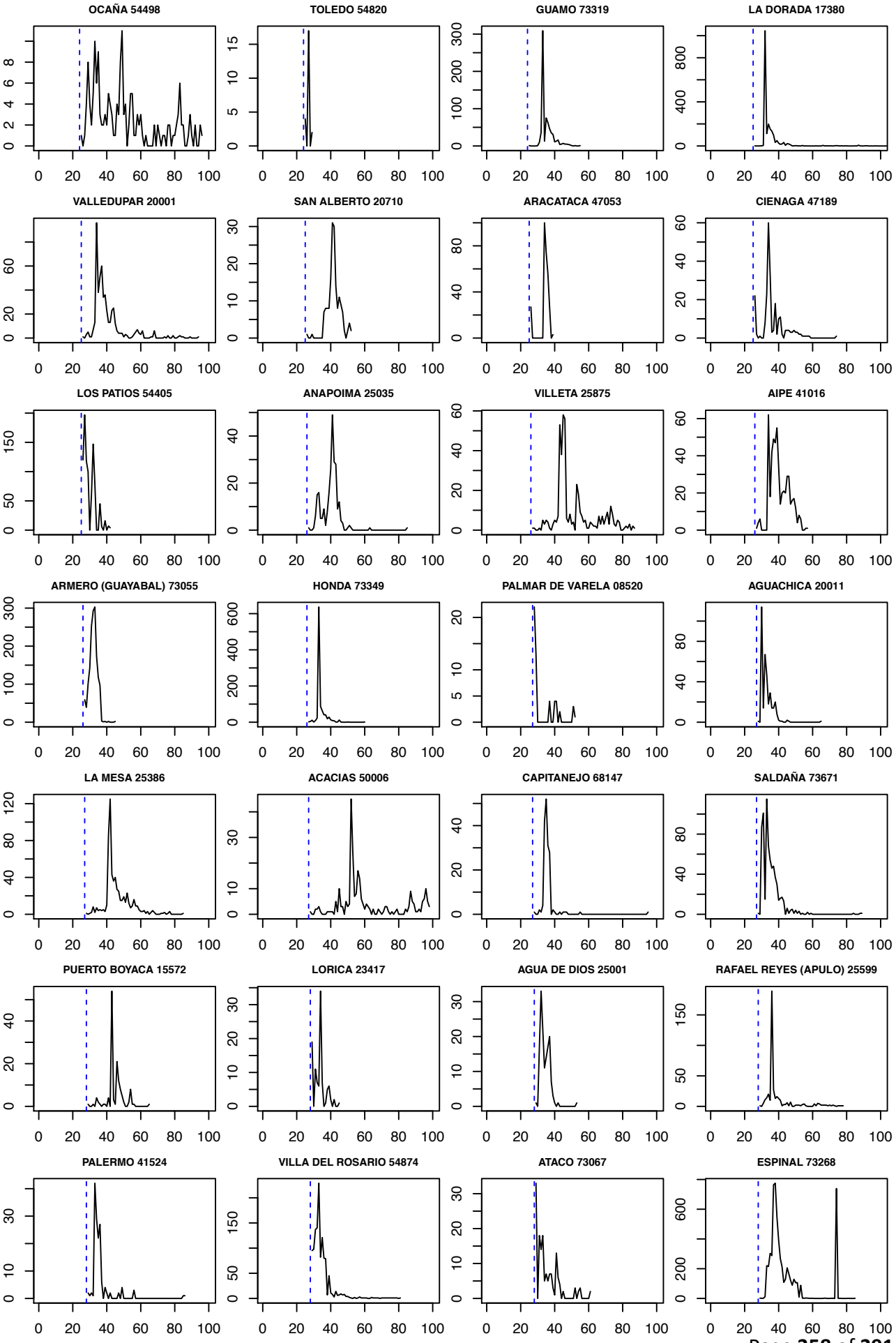
CHIKV

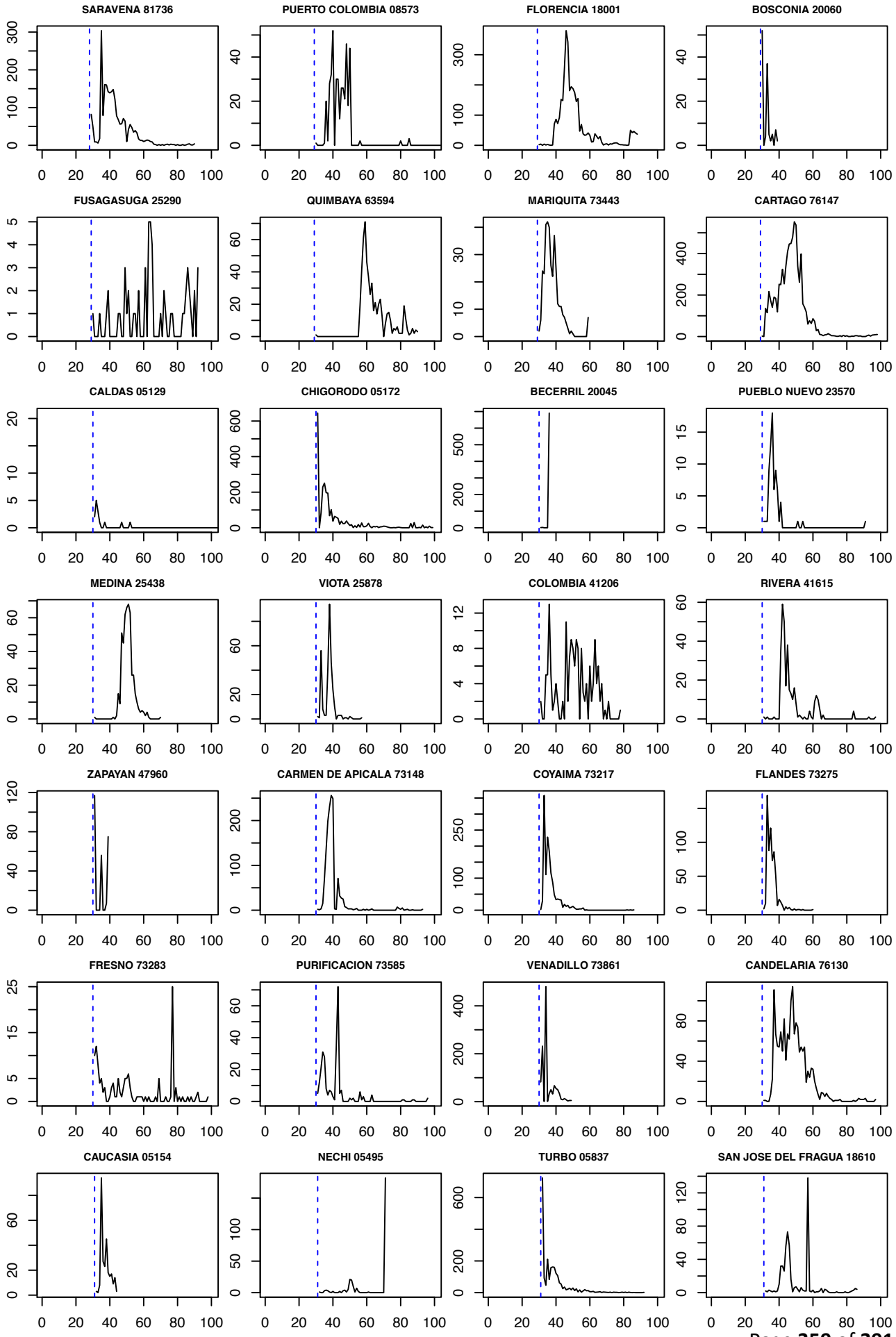

PLANETA RICA 23555
APARTADO 05045
GIRARDOT 25307
BARRANQUILLA 08001
LA PRIMAVERA 99524
TURBACO 13836
CARTAGENA 13001
ENVIGADO 05266
SAN JUAN NEPOMUCENO 13657
BARRANCA DE UPIA 50110
CUCUTA 54001
BARANOA 08078
CAMPO DE LA CRUZ 08137
SABANALARGA 08638
SANTO TOMAS 08685
SOLEDAD 08758
SUAN 08770
SANTA ROSA 13683
FLORIDABLANCA 68276
COROZAL 70215
PONEDERA 08560
ARJONA 13052
ARROYOHONDO 13062
CORDOBA 13212
EL CARMEN DE BOLIVAR 13244
MAGANGUE 13430
MAHATES 13433
MARIA LA BAJA 13442

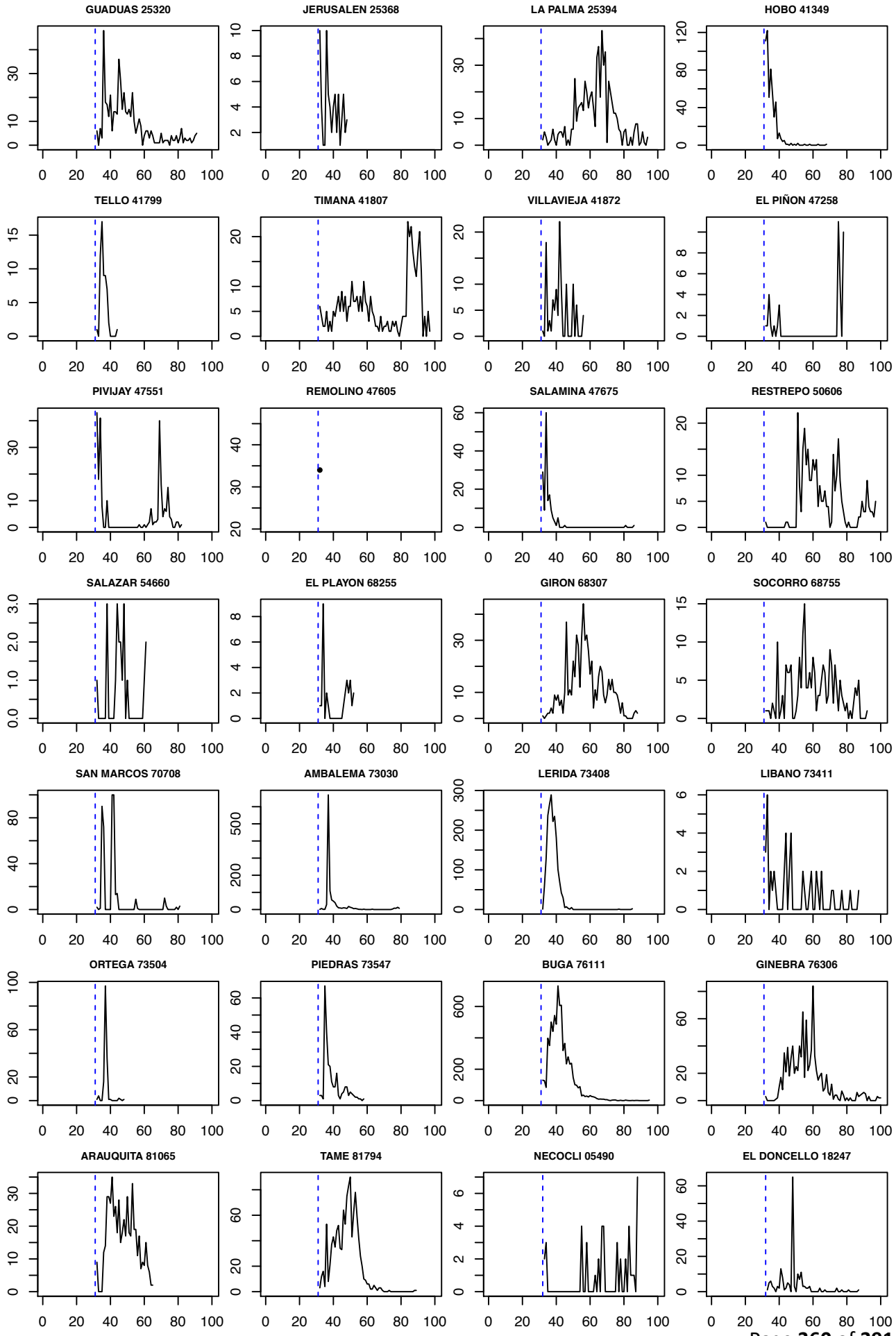

SAN ESTANISLAO 13647
SAN JACINTO 13654
SANTANDER DE QUILICHAO 19698
SAHAGUN 23660
PUERTO SALGAR 25572
SANTA MARTA 47001
OVEJAS 70508
SINCE 70742
TURBANA 13838
MONTERIA 23001
NILO 25488
MAICAO 44430
BUENAVISTA 70110
TOLUVIEJO 70823
CHAPARRAL 73168
CALI 76001
MOCOA 86001
PUERTO ASIS 86568
SAN LUIS DE GACENO 15667
CHINCHINA 17174
EL COPEY 20238
CERETE 23162
MONTELIBANO 23466
MOÑITOS 23500
PLATO 47555
BUCARAMANGA 68001
SINCELEJO 70001
COVEÑAS 70221

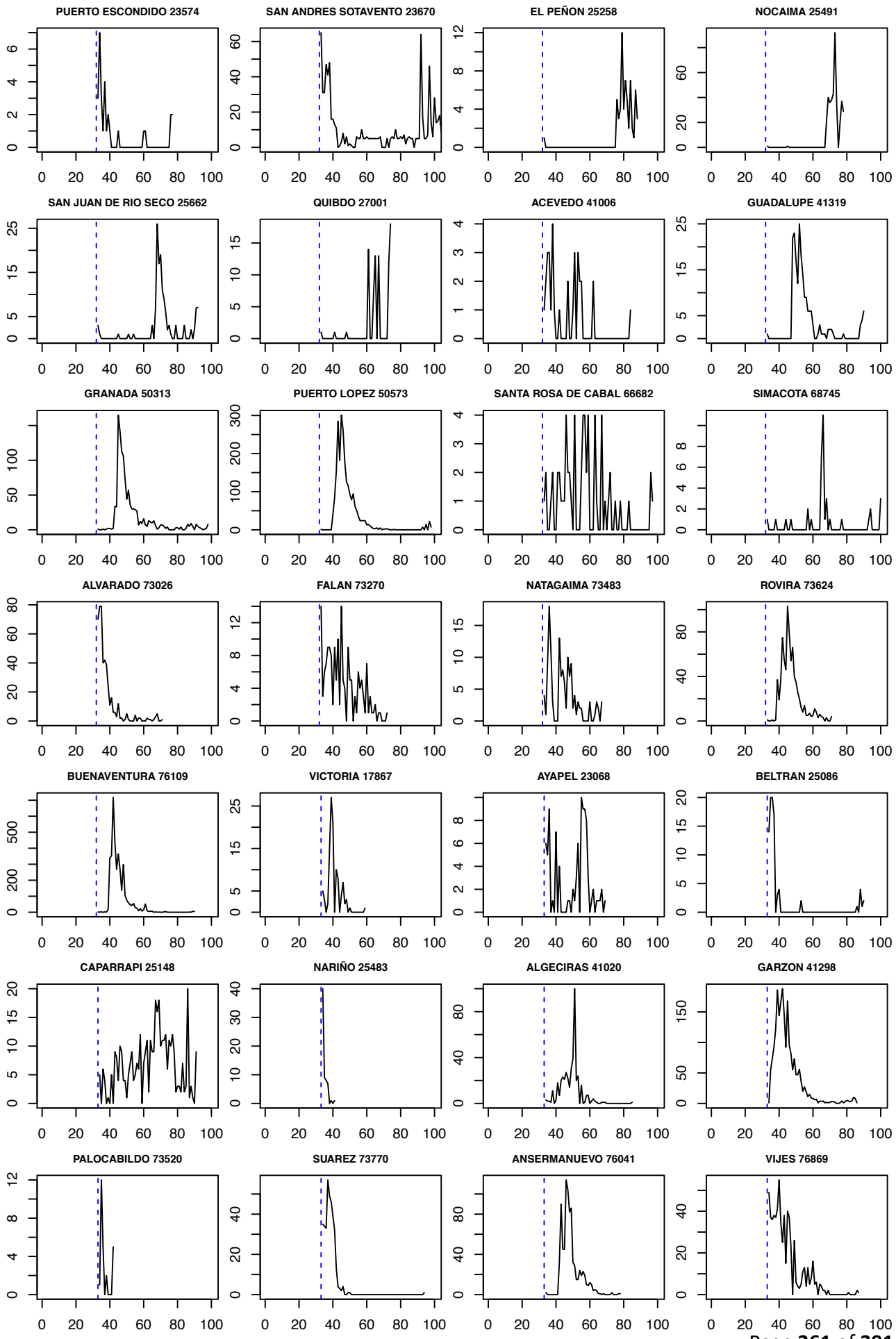

LOS PALMITOS 70418
PALMITO 70523
YOPAL 85001
VILLANUEVA 85440
SAN JOSE DEL GUAVIARE 95001
ACHI 13006
PUERTO LIBERTADOR 23580
SAN ANTERO 23672
SAN BERNARDO DEL VIENTO 23675
TIERRALTA 23807
TOCAIMA 25815
NEIVA 41001
BARAYA 41078
PITALITO 41551
RIOHACHA 44001
ALBANIA 44035
DISTRACCION 44098
SAN JUAN DEL CESAR 44650
URIBIA 44847
EL BANCO 47245
TENERIFE 47798
TUMACO 52835
BARRANCABERMEJA 68081
PIEDECUESTA 68547
SAMPUES 70670
TOLU 70820
IBAGUE 73001
EL PASO 20250

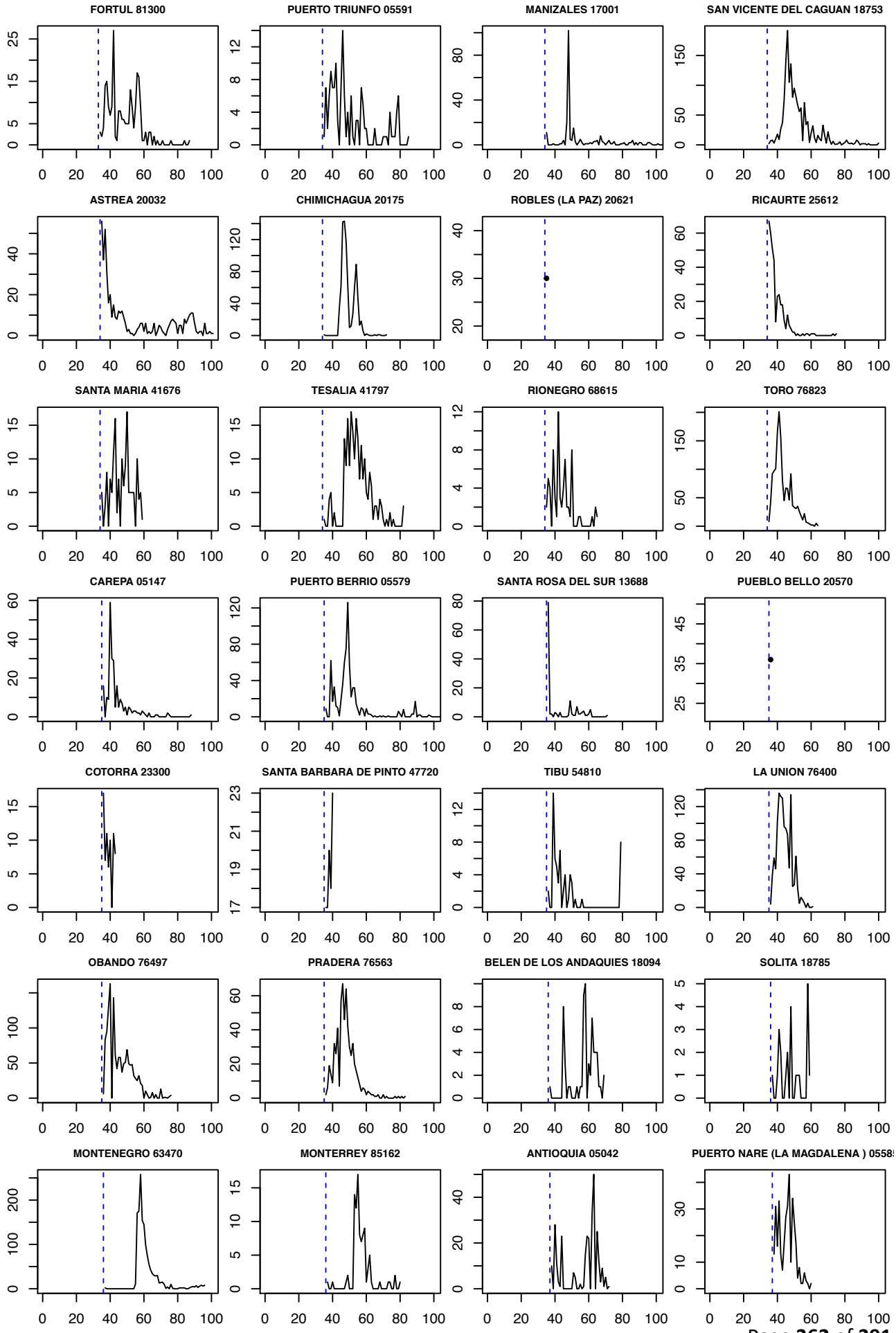

EL COLEGIO 25245
VILLAVICENCIO 50001
IPIALES 52356
PEREIRA 66001
LA VIRGINIA 66400
PUERTO WILCHES 68575
EL ROBLE 70233
MELGAR 73449
ARAUCA 81001
SAN MIGUEL (LA DORADA) 86757
VALLE DEL GUAMUEZ 86865
SAN ANDRES 88001
CAMPOALEGRE 41132
NUEVA GRANADA 47460
CONVENCION 54206
ARMENIA 63001
LA UNION 70400
SAN ONOFRE 70713
PRADO 73563
BUGALAGRANDE 76113
EL CERRITO 76248
LA VICTORIA 76403
PALMIRA 76520
BOLIVAR 19100
SAN JUAN DE BETULIA 70702
TULUA 76834
YUMBO 76892
LURUACO 08421

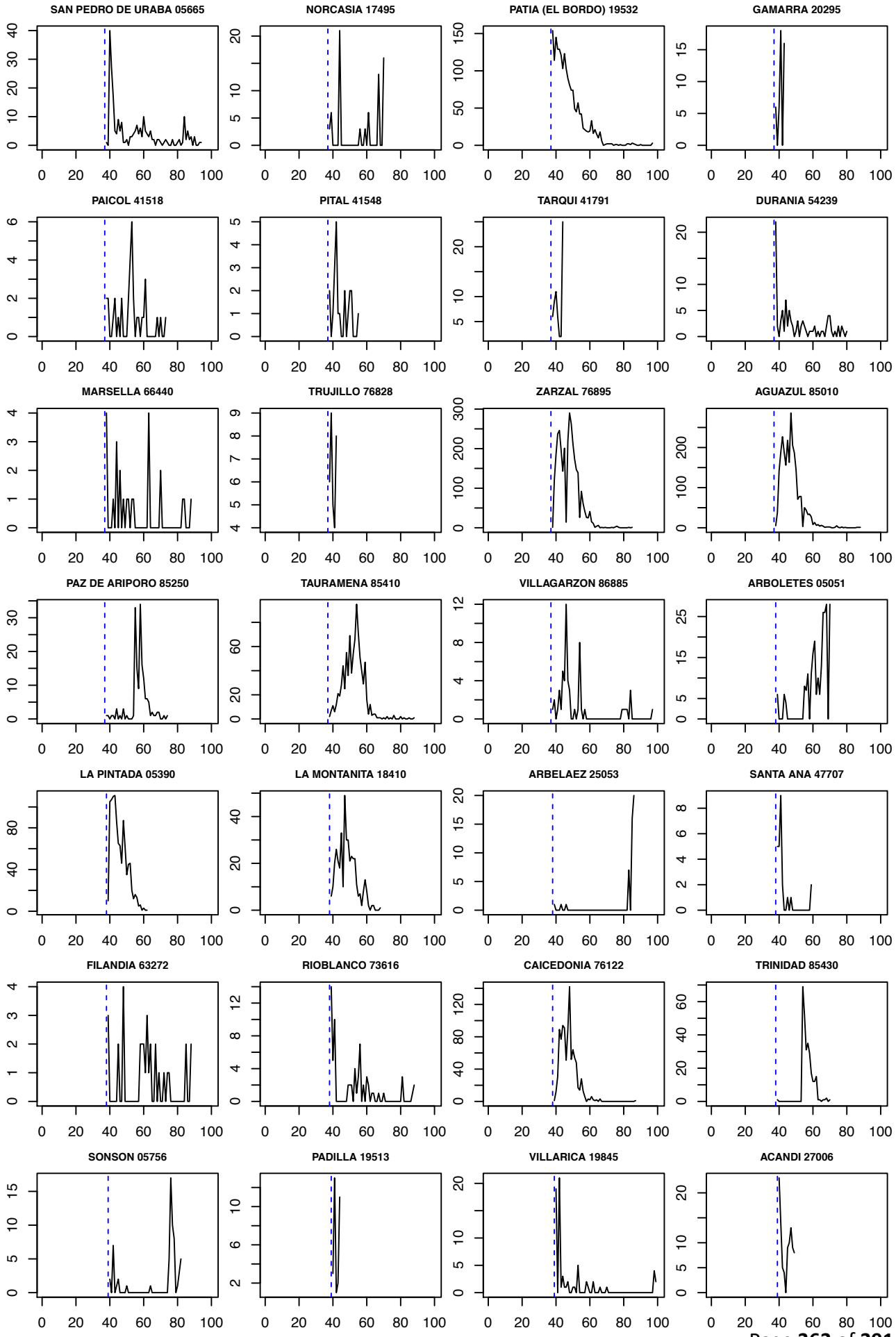

OCAÑA 54498
TOLEDO 54820
GUAMO 73319
LA DORADA 17380
VALLEDUPAR 20001
SAN ALBERTO 20710
ARACATACA 47053
CIENAGA 47189
LOS PATIOS 54405
ANAPOIMA 25035
VILLETA 25875
AIPE 41016
ARMERO (GUAYABAL) 73055
HONDA 73349
PALMAR DE VARELA 08520
AGUACHICA 20011
LA MESA 25386
ACACIAS 50006
CAPITANEJO 68147
SALDAÑA 73671
PUERTO BOYACA 15572
LORICA 23417
AGUA DE DIOS 25001
RAFAEL REYES (APULO) 25599
PALERMO 41524
VILLA DEL ROSARIO 54874
ATACO 73067
ESPINAL 73268

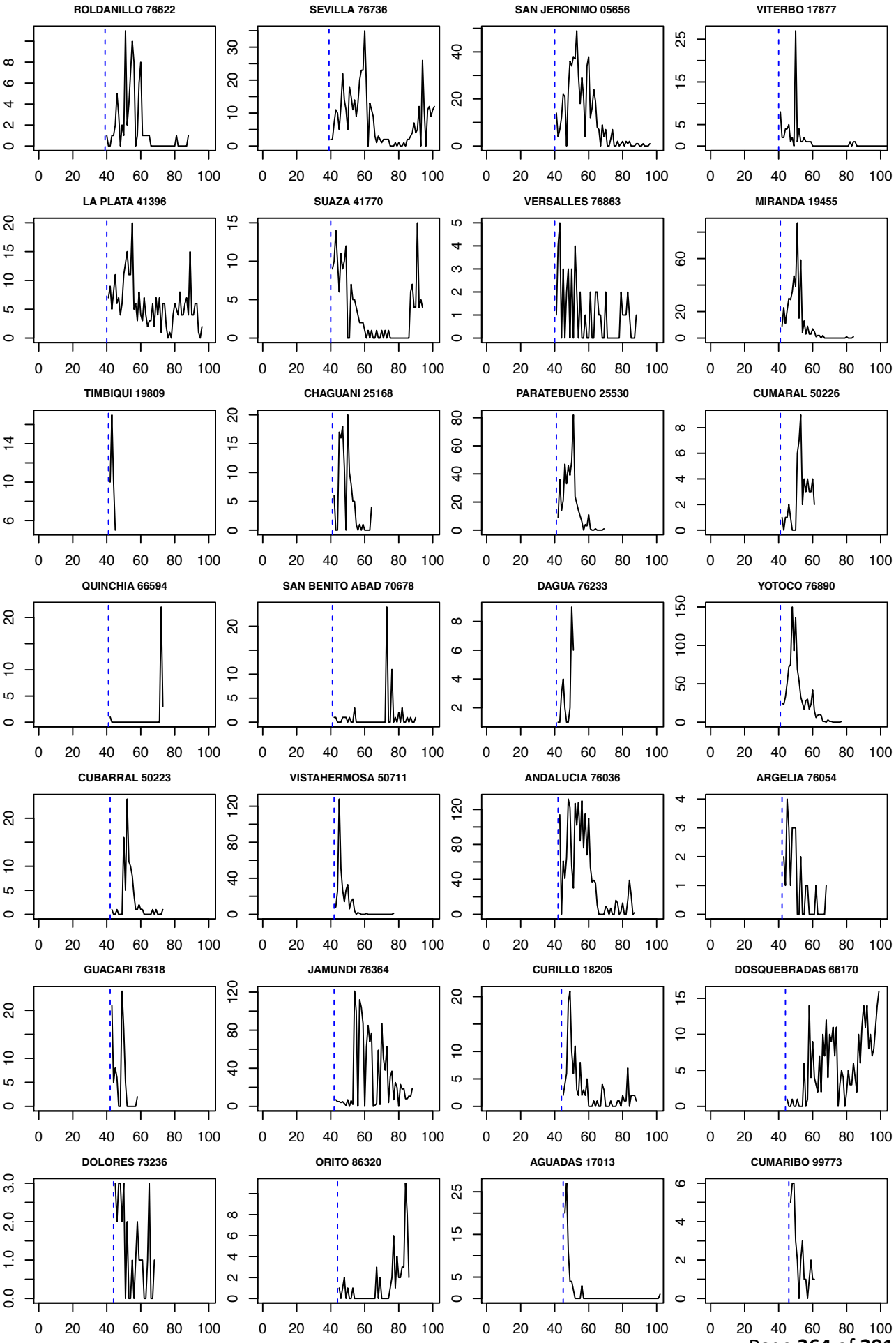

SARAVENA 81736
PUERTO COLOMBIA 08573
FLORENCIA 18001
BOSCONIA 20060
FUSAGASUGA 25290
QUIMBAYA 63594
MARIQUITA 73443
CARTAGO 76147
CALDAS 05129
CHIGORODO 05172
BECERRIL 20045
PUEBLO NUEVO 23570
MEDINA 25438
VIOTA 25878
COLOMBIA 41206
RIVERA 41615
ZAPAYAN 47960
CARMEN DE APICALA 73148
COYAIMA 73217
FLANDES 73275
FRESNO 73283
PURIFICACION 73585
VENADILLO 73861
CANDELARIA 76130
CAUCASIA 05154
NECHI 05495
TURBO 05837
SAN JOSE DEL FRAGUA 18610

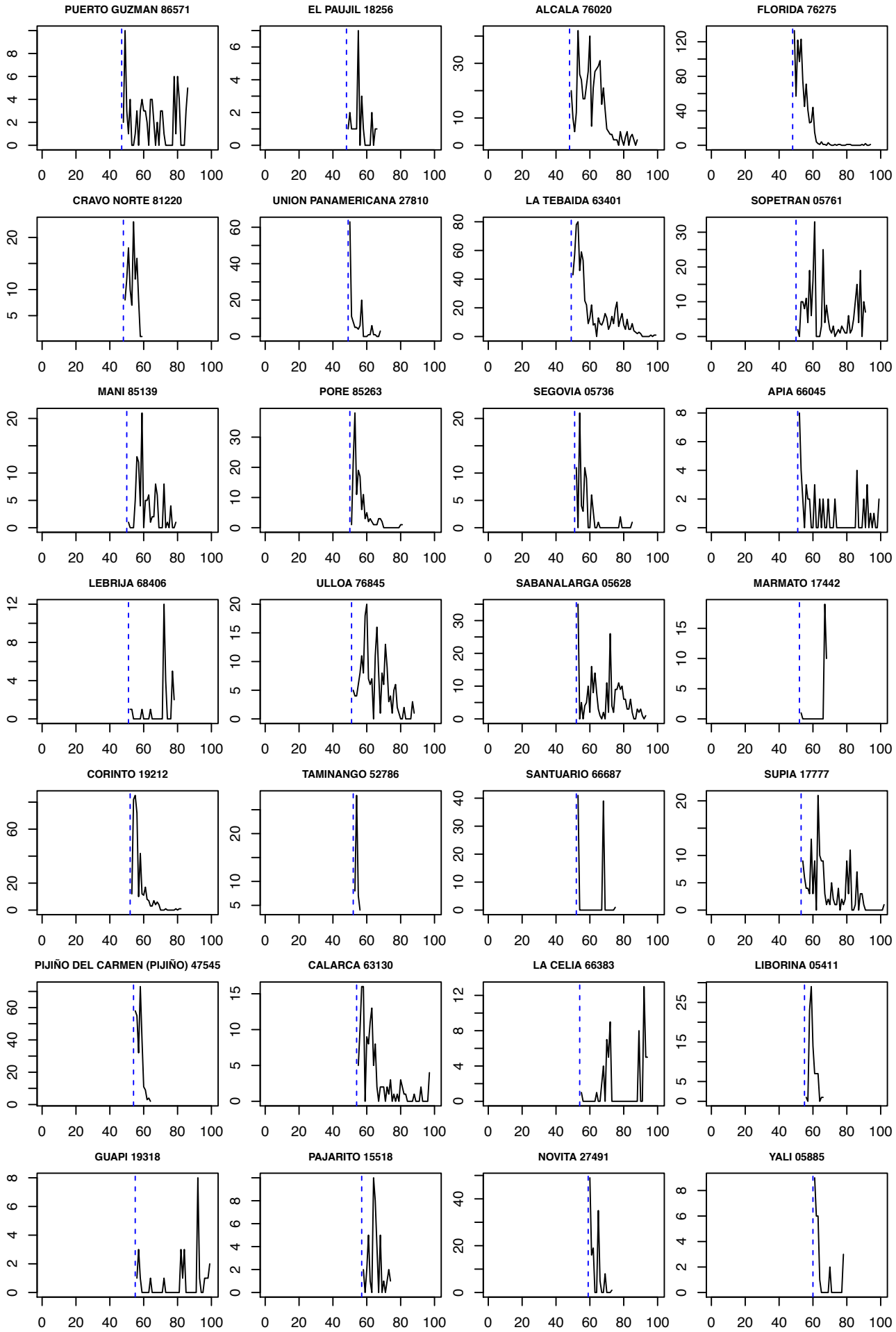

GUADUAS 25320
JERUSALEN 25368
LA PALMA 25394
HOBO 41349
TELLO 41799
TIMANA 41807
VILLAVIEJA 41872
EL PIÑON 47258
PIVIJAY 47551
REMOLINO 47605
SALAMINA 47675
RESTREPO 50606
SALAZAR 54660
EL PLAYON 68255
GIRON 68307
SOCORRO 68755
SAN MARCOS 70708
AMBALEMA 73030
LERIDA 73408
LIBANO 73411
ORTEGA 73504
PIEDRAS 73547
BUGA 76111
GINEBRA 76306
ARAUQUITA 81065
TAME 81794
NECOCLI 05490
EL DONCELLO 18247

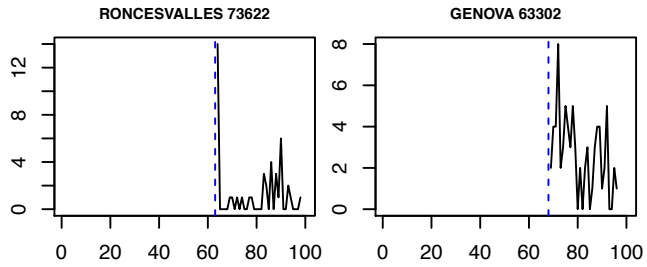

PUERTO ESCONDIDO 23574
SAN ANDRES SOTAVENTO 23670
EL PEÑON 25258
NOCAIMA 25491
SAN JUAN DE RIO SECO 25662
QUIBDO 27001
ACEVEDO 41006
GUADALUPE 41319
GRANADA 50313
PUERTO LOPEZ 50573
SANTA ROSA DE CABAL 66682
SIMACOTA 68745
ALVARADO 73026
FALAN 73270
NATAGAIMA 73483
ROVIRA 73624
BUENAVENTURA 76109
VICTORIA 17867
AYAPEL 23068
BELTRAN 25086
CAPARRAPI 25148
NARIÑO 25483
ALGECIRAS 41020
GARZON 41298
PALOCABILDO 73520
SUAREZ 73770
ANSERMANUEVO 76041
VIJES 76869


FORTUL 81300
PUERTO TRIUNFO 05591
MANIZALES 17001
SAN VICENTE DEL CAGUAN 18753
ASTREA 20032
CHIMICHAGUA 20175
ROBLES (LA PAZ) 20621
RICAURTE 25612
SANTA MARIA 41676
TESALIA 41797
RIONEGRO 68615
TORO 76823
CAREPA 05147
PUERTO BERRIO 05579
SANTA ROSA DEL SUR 13688
PUEBLO BELLO 20570
COTORRA 23300
SANTA BARBARA DE PINTO 47720
TIBU 54810
LA UNION 76400
OBANDO 76497
PRADERA 76563
BELEN DE LOS ANDAQUIES 18094
SOLITA 18785
MONTENEGRO 63470
MONTERREY 85162
ANTIOQUIA 05042
PUERTO NARE (LA MAGDALENA) 0558

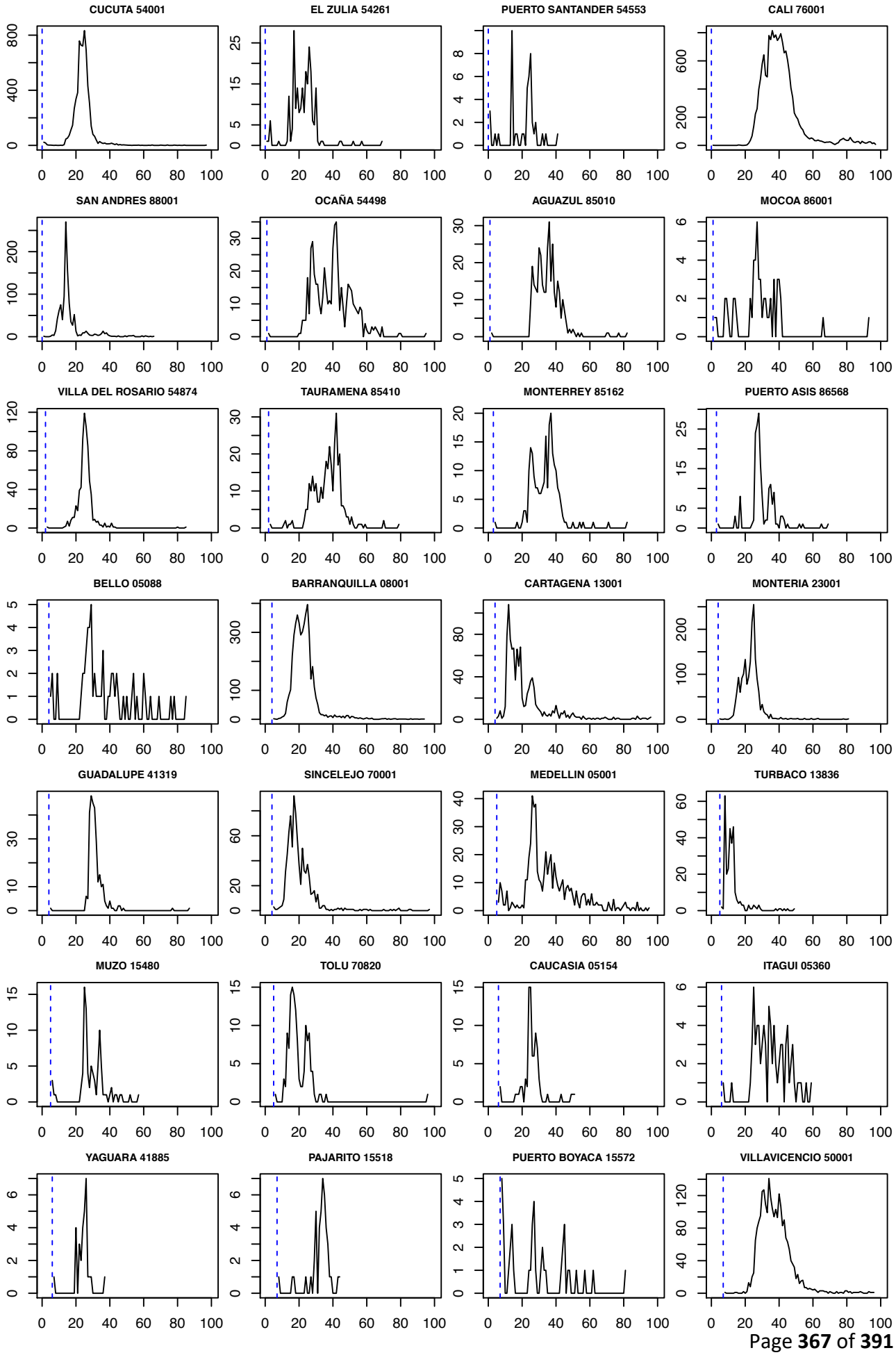

SAN PEDRO DE URABA 05665
NORCASIA 17495
PATIA (EL BORDO) 19532
GAMARRA 20295
PAICOL 41518
PITAL 41548
TARQUI 41791
DURANIA 54239
MARSELLA 66440
TRUJILLO 76828
ZARZAL 76895
AGUAZUL 85010
PAZ DE ARIPORO 85250
TAURAMENA 85410
VILLAGARZON 86885
ARBOLETES 05051
LA PINTADA 05390
LA MONTANITA 18410
ARBELAEZ 25053
SANTA ANA 47707
FILANDIA 63272
RIOBLANCO 73616
CAICEDONIA 76122
TRINIDAD 85430
SONSON 05756
PADILLA 19513
VILLARICA 19845
ACANDI 27006

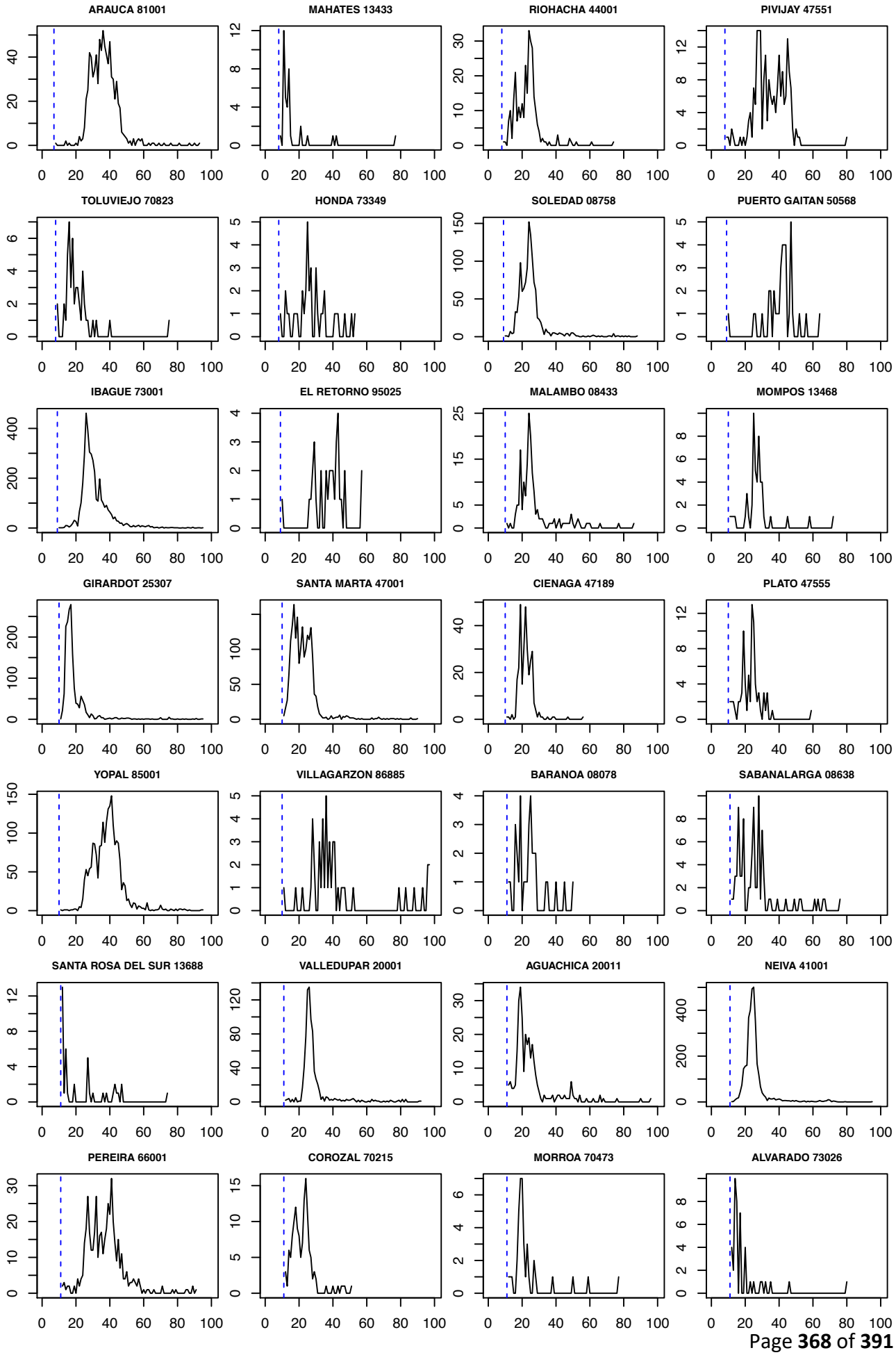

ROLDANILLO 76622
SEVILLA 76736
SAN JERONIMO 05656
VITERBO 17877
LA PLATA 41396
SUAZA 41770
VERSALLES 76863
MIRANDA 19455
TIMBIQUI 19809
CHAGUANI 25168
PARATEBUENO 25530
CUMARAL 50226
QUINCHIA 66594
SAN BENITO ABAD 70678
DAGUA 76233
YOTOCO 76890
CUBARRAL 50223
VISTAHERMOSA 50711
ANDALUCIA 76036
ARGELIA 76054
GUACARI 76318
JAMUNDI 76364
CURILLO 18205
DOSQUEBRADAS 66170
DOLORES 73236
ORITO 86320
AGUADAS 17013
CUMARIBO 99773

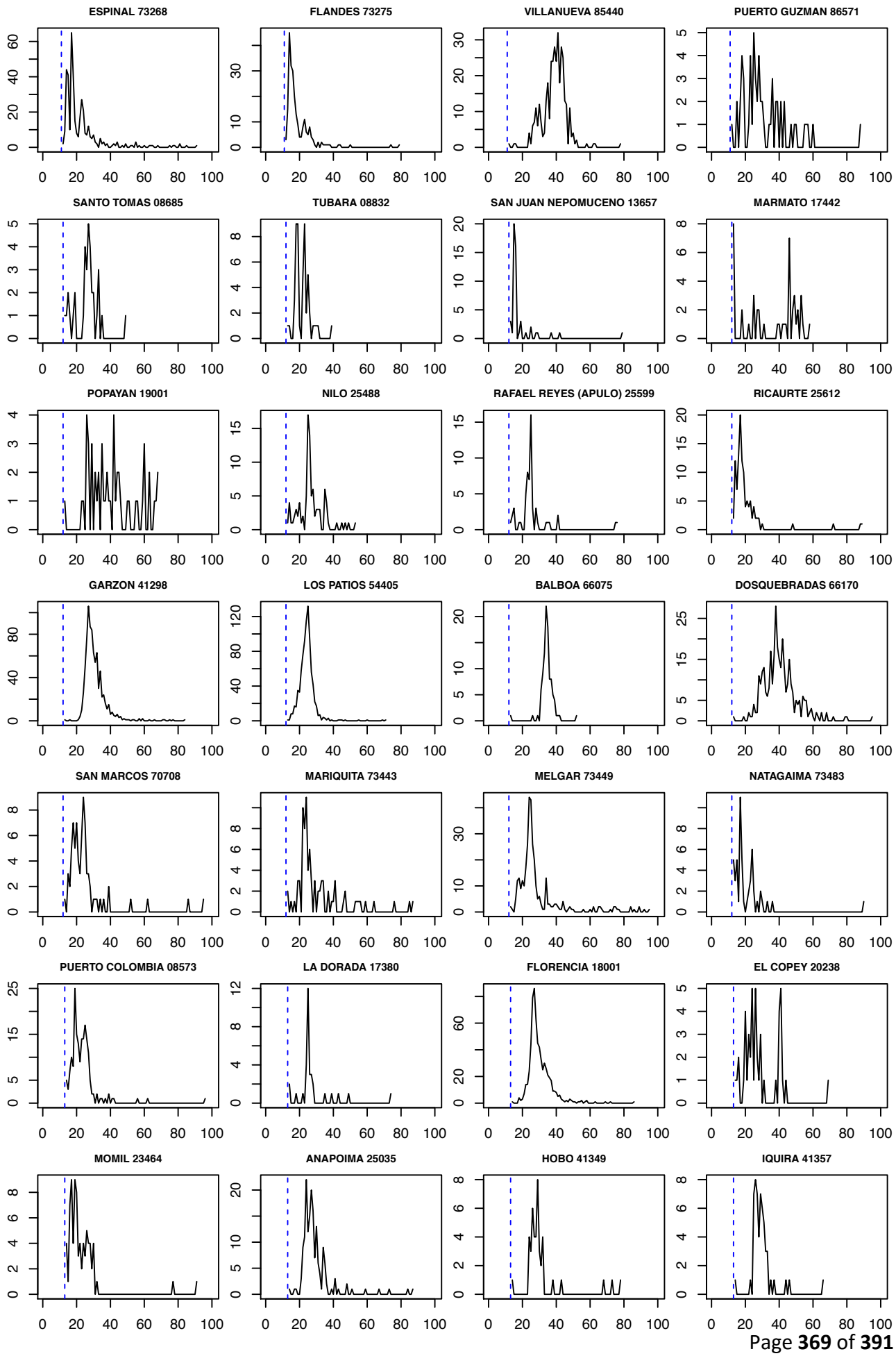

PUERTO GUZMAN 86571
EL PAUJIL 18256
ALCALA 76020
FLORIDA 76275
CRAVO NORTE 81220
UNION PANAMERICANA 27810
LA TEBAIDA 63401
SOPETRAN 05761
MANI 85139
PORE 85263
SEGOVIA 05736
APIA 66045
LEBRIJA 68406
ULLOA 76845
SABANALARGA 05628
MARMATO 17442
CORINTO 19212
TAMINANGO 52786
SANTUARIO 66687
SUPIA 17777
PIJIÑO DEL CARMEN (PIJIÑO) 47545
CALARCA 63130
LA CELIA 66383
LIBORINA 05411
GUAPI 19318
PAJARITO 15518
NOVITA 27491
YALI 05885

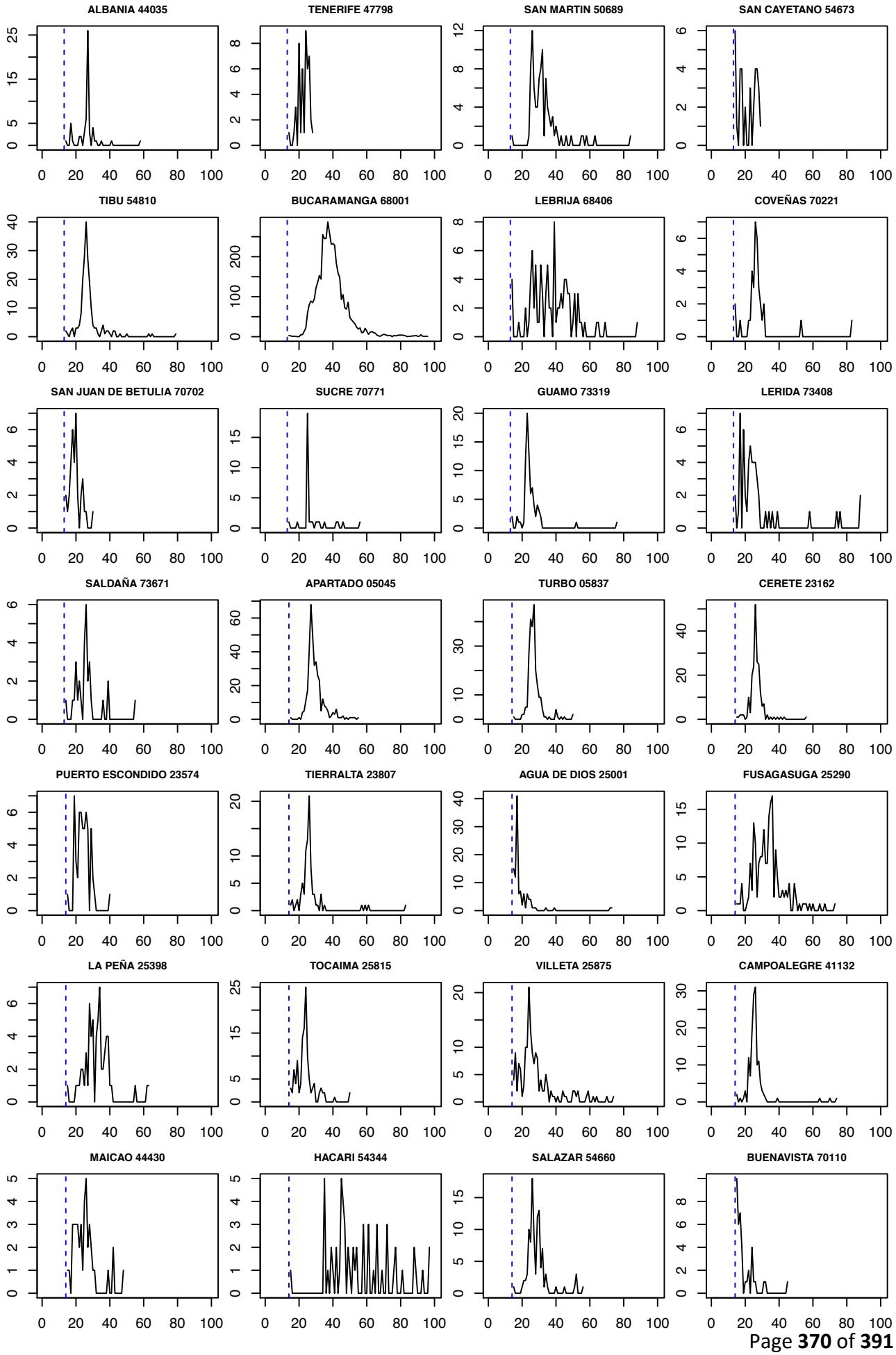

RONCESVALLES 73622
GENOVA 63302
0 4 8 12
0 2 4 6 8
0 20 40 60 80 100
0 20 40 60 80 100

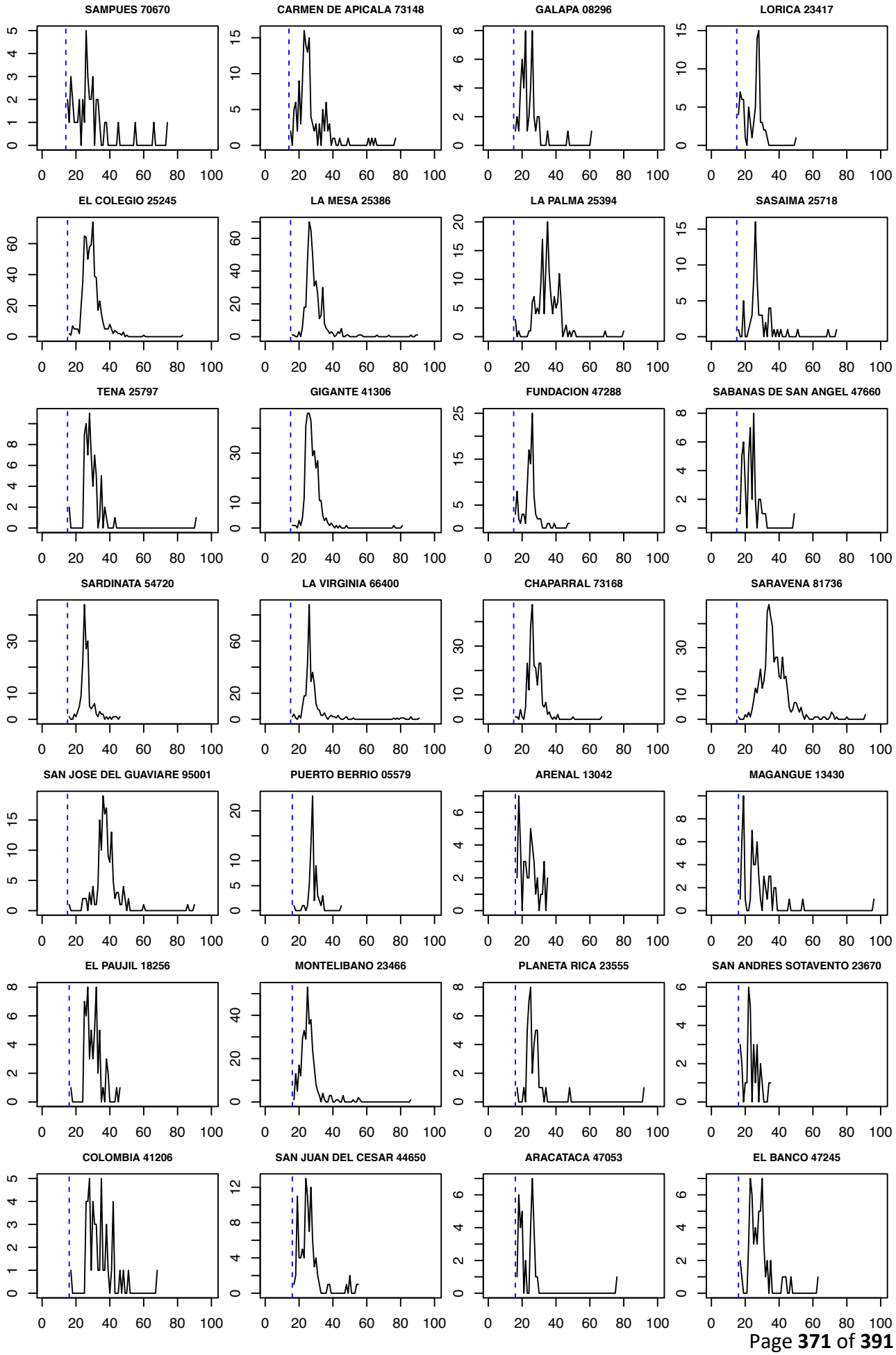
ZIKV

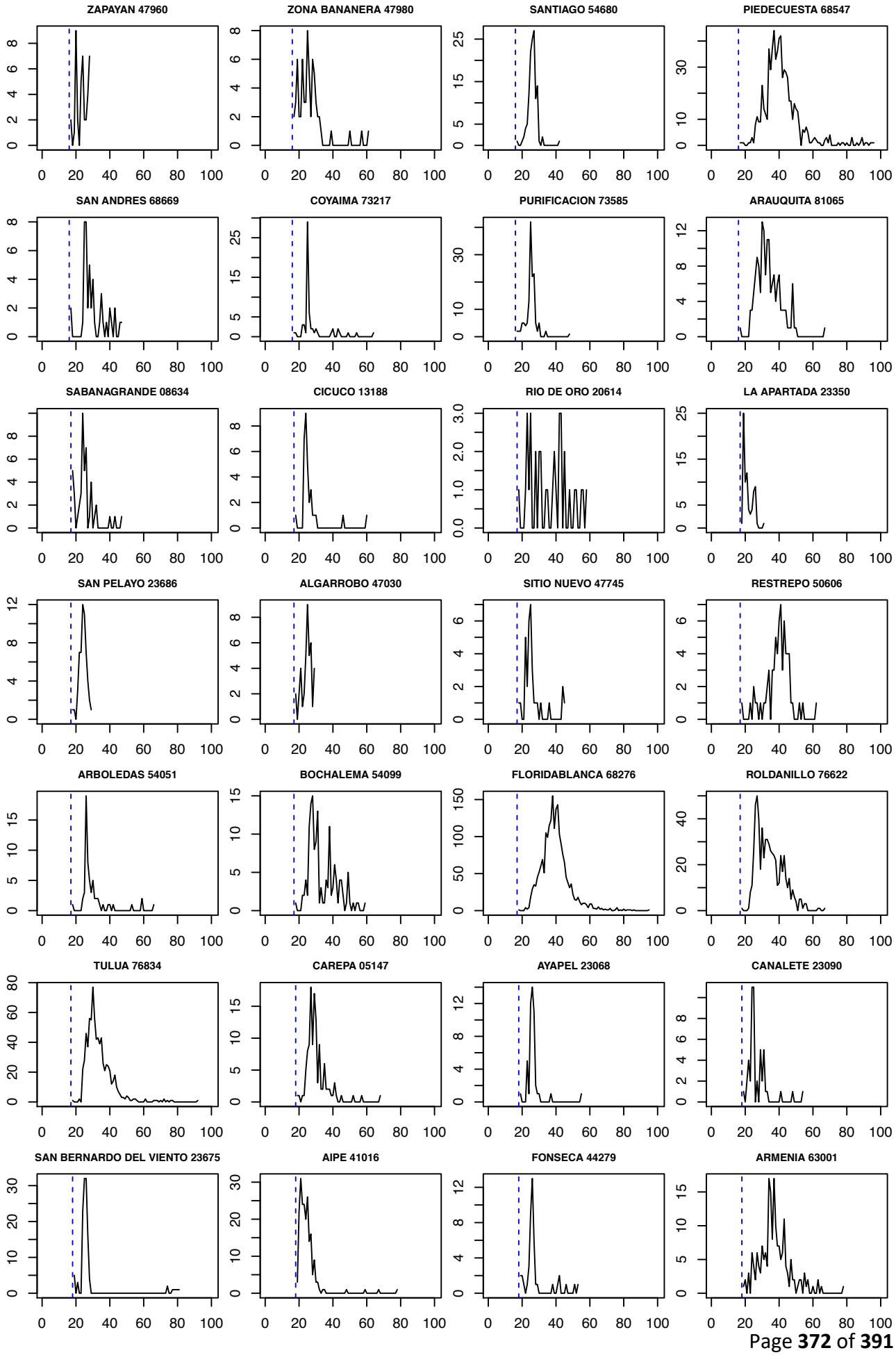

CUCUTA 54001
EL ZULIA 54261
PUERTO SANTANDER 54553
CALI 76001
SAN ANDRES 88001
OCAÑA 54498
AGUAZUL 85010
MOCOA 86001
VILLA DEL ROSARIO 54874
TAURAMENA 85410
MONTERREY 85162
PUERTO ASIS 86568
BELLO 05088
BARRANQUILLA 08001
CARTAGENA 13001
MONTERIA 23001
GUADALUPE 41319
SINCELEJO 70001
MEDELLIN 05001
TURBACO 13836
MUZO 15480
TOLU 70820
CAUCASIA 05154
ITAGUI 05360
YAGUARA 41885
PAJARITO 15518
PUERTO BOYACA 15572
VILLAVICENCIO 50001

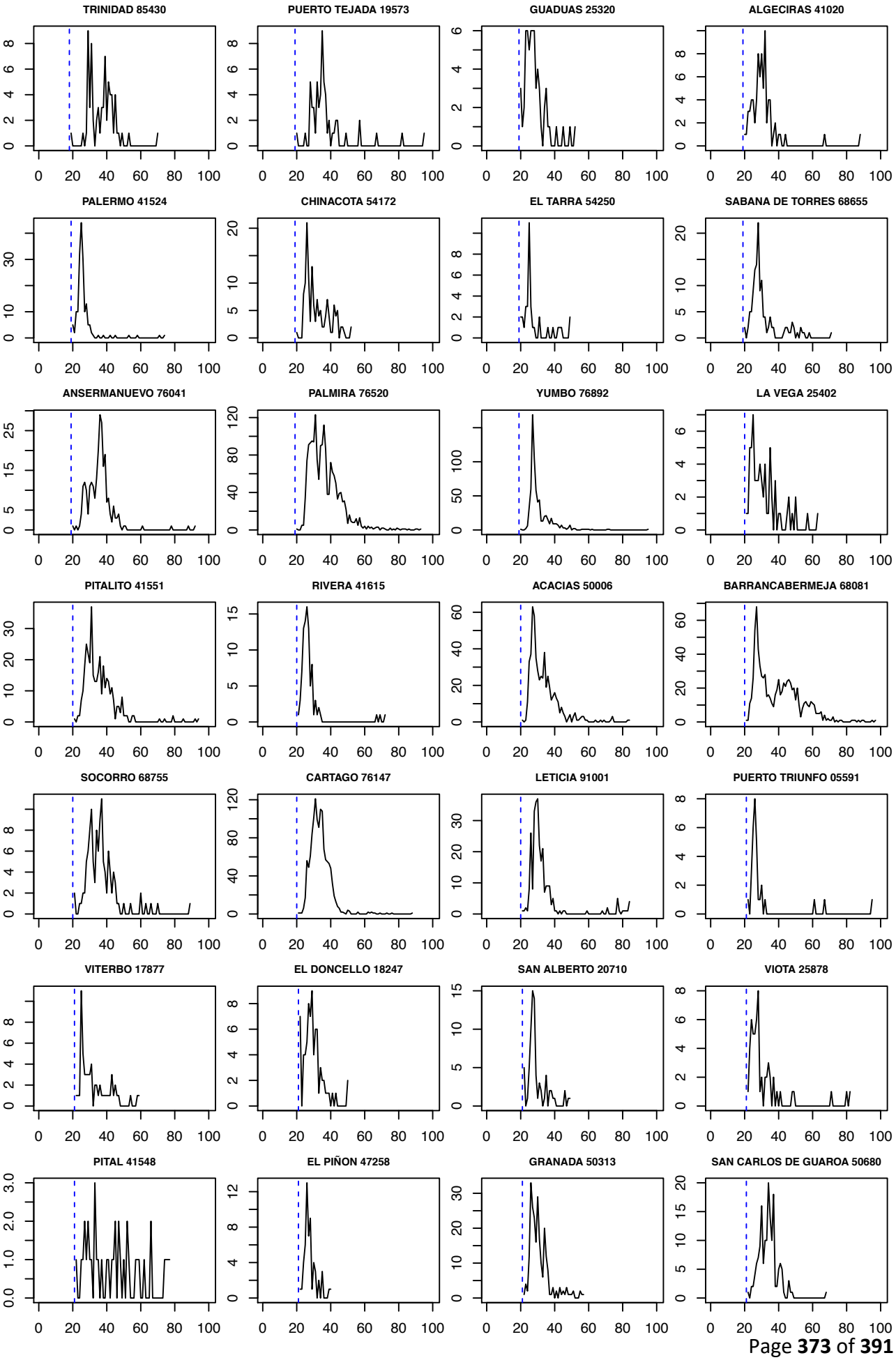

ARAUCA 81001
MAHATES 13433
RIOHACHA 44001
PIVIJAY 47551
TOLUVIEJO 70823
HONDA 73349
SOLEDAD 08758
PUERTO GAITAN 50568
IBAGUE 73001
EL RETORNO 95025
MALAMBO 08433
MOMPOS 13468
GIRARDOT 25307
SANTA MARTA 47001
CIENAGA 47189
PLATO 47555
YOPAL 85001
VILLAGARZON 86885
BARANOA 08078
SABANALARGA 08638
SANTA ROSA DEL SUR 13688
VALLEDUPAR 20001
AGUACHICA 20011
NEIVA 41001
PEREIRA 66001
COROZAL 70215
MORROA 70473
ALVARADO 73026

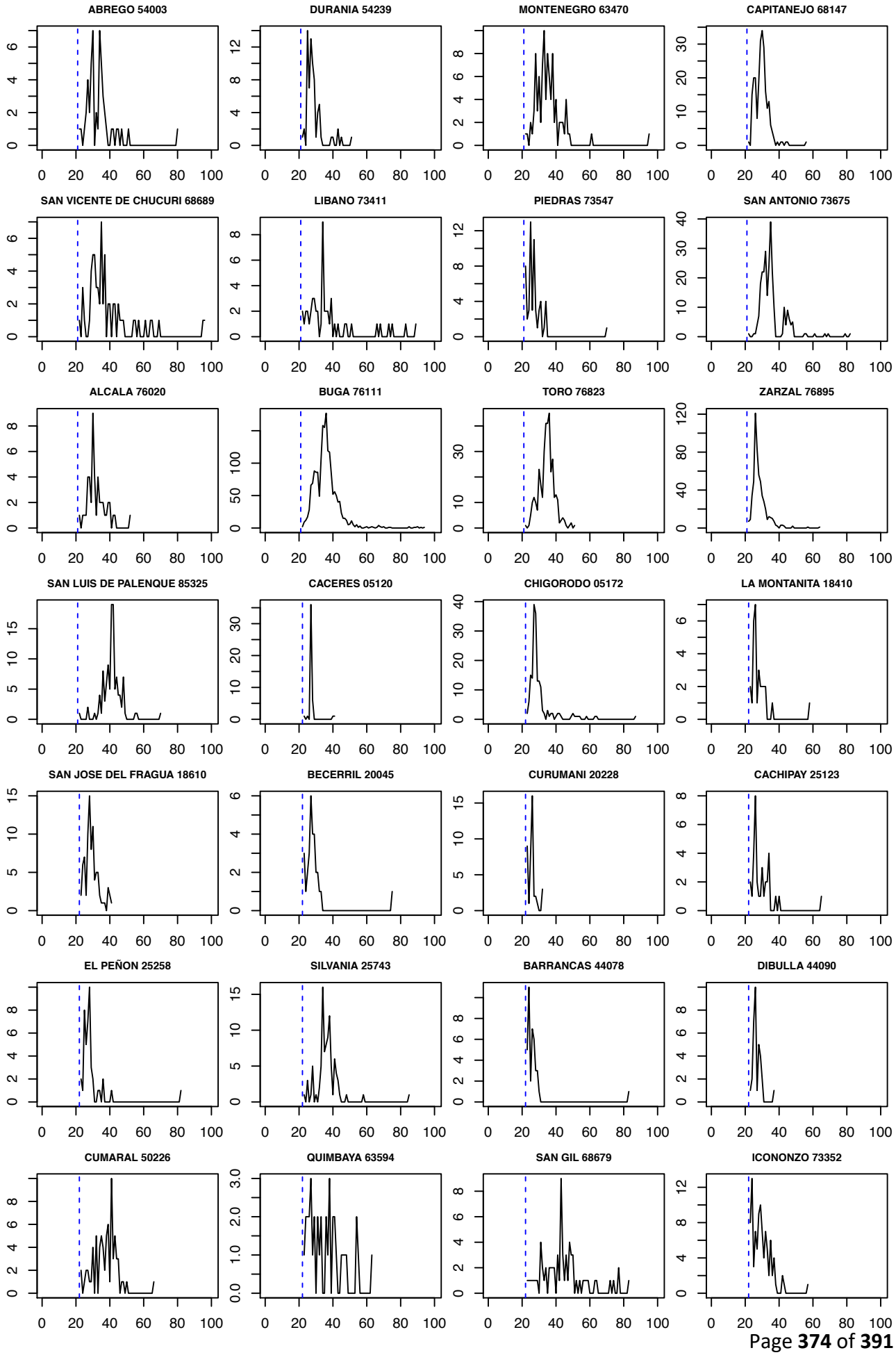

ESPINAL 73268
FLANDES 73275
VILLANUEVA 85440
PUERTO GUZMAN 86571
SANTO TOMAS 08685
TUBARA 08832
SAN JUAN NEPOMUCENO 13657
MARMATO 17442
POPAYAN 19001
NILO 25488
RAFAEL REYES (APULO) 25599
RICAURTE 25612
GARZON 41298
LOS PATIOS 54405
BALBOA 66075
DOSQUEBRADAS 66170
SAN MARCOS 70708
MARIQUITA 73443
MELGAR 73449
NATAGAIMA 73483
PUERTO COLOMBIA 08573
LA DORADA 17380
FLORENCIA 18001
EL COPEY 20238
MOMIL 23464
ANAPOIMA 25035
HOBO 41349
IQUIRA 41357

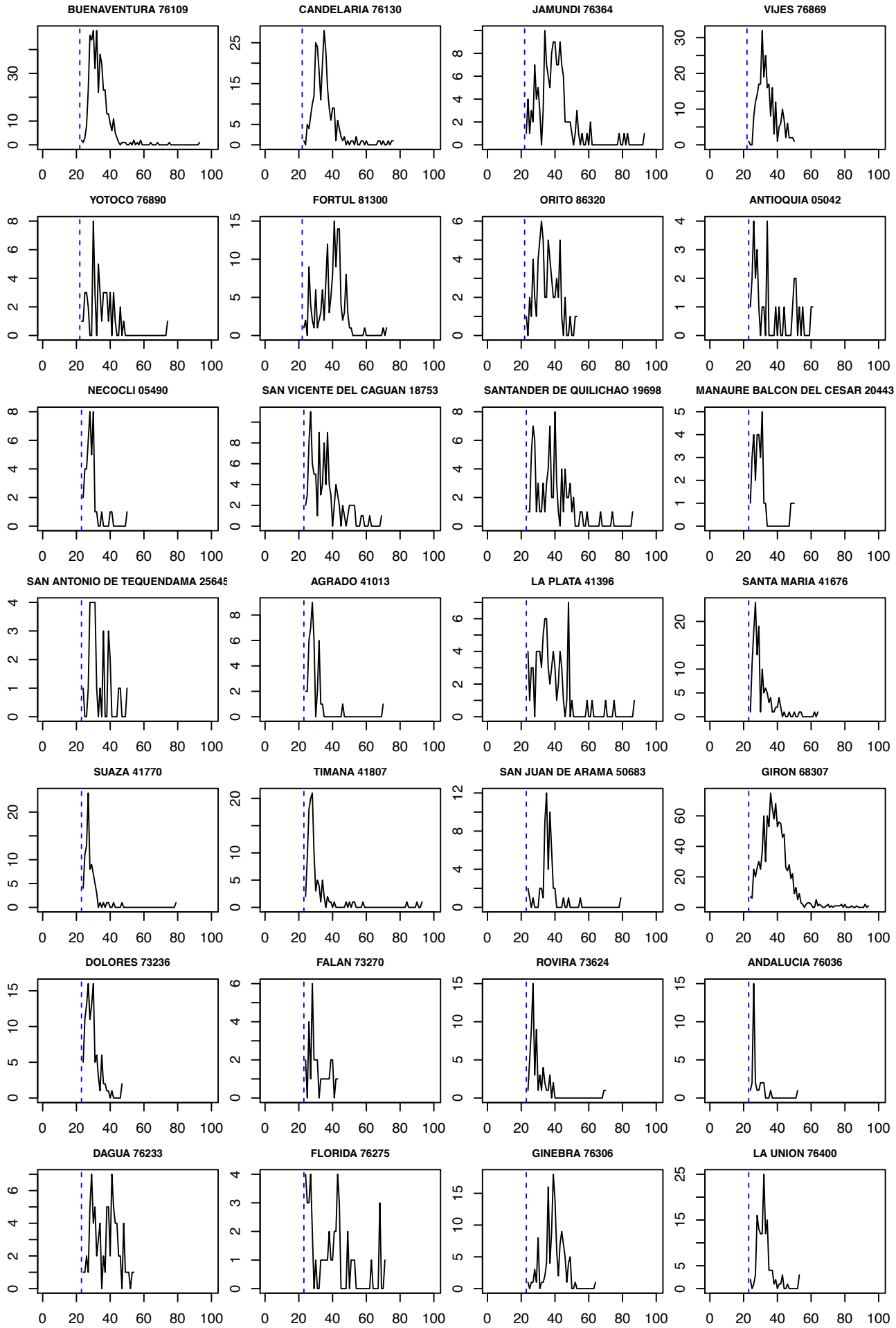

ALBANIA 44035
TENERIFE 47798
SAN MARTIN 50689
SAN CAYETANO 54673
TIBU 54810
BUCARAMANGA 68001
LEBRIJA 68406
COVEÑAS 70221
SAN JUAN DE BETULIA 70702
SUCRE 70771
GUAMO 73319
LERIDA 73408
SALDAÑA 73671
APARTADO 05045
TURBO 05837
CERETE 23162
PUERTO ESCONDIDO 23574
TIERRALTA 23807
AGUA DE DIOS 25001
FUSAGASUGA 25290
LA PEÑA 25398
TOCAIMA 25815
VILLETA 25875
CAMPOALEGRE 41132
MAICAO 44430
HACARI 54344
SALAZAR 54660
BUENAVISTA 70110

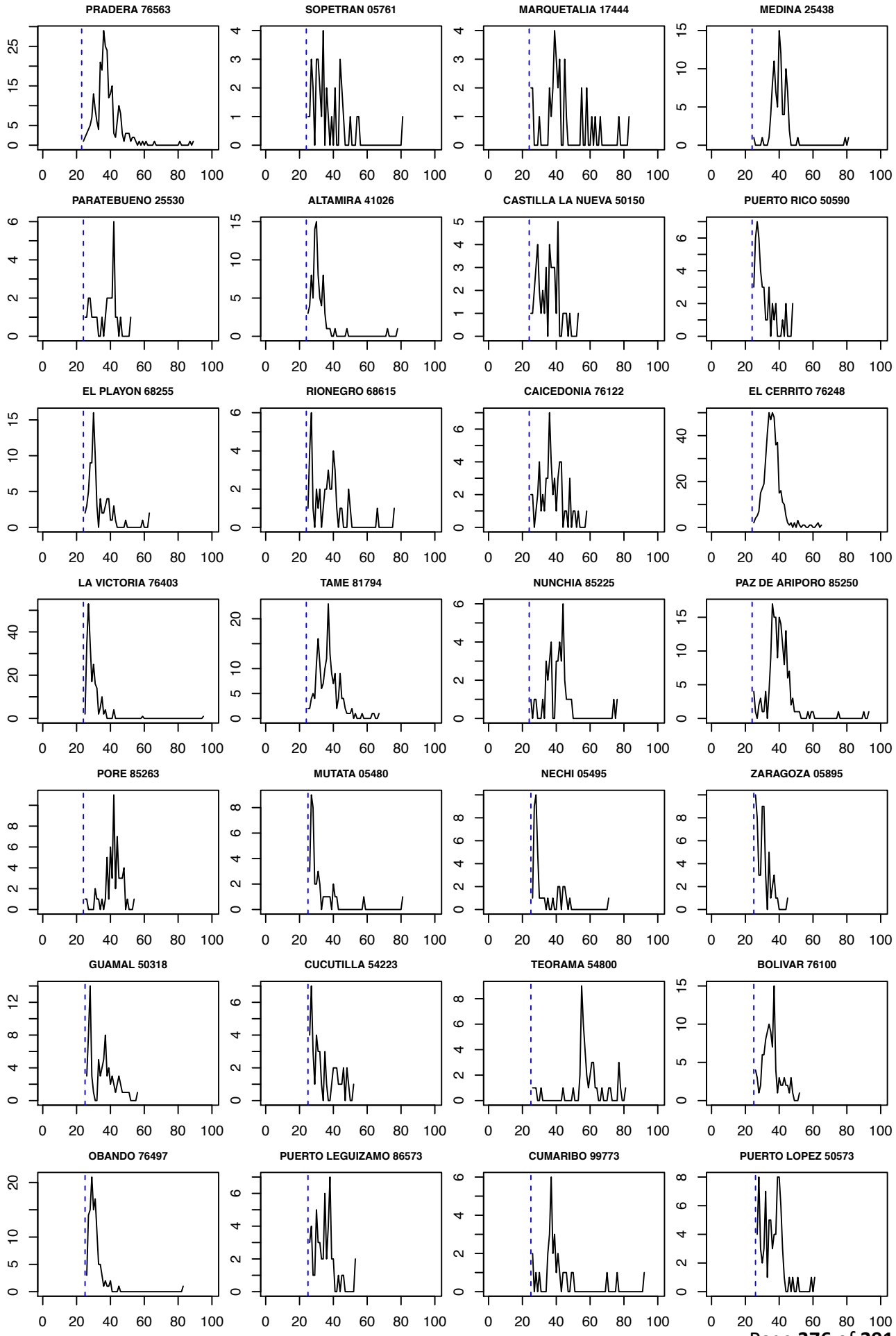

SAMPUES 70670
CARMEN DE APICALA 73148
GALAPA 08296
LORICA 23417
EL COLEGIO 25245
LA MESA 25386
LA PALMA 25394
SASAIMA 25718
TENA 25797
GIGANTE 41306
FUNDACION 47288
SABANAS DE SAN ANGEL 47660
SARDINATA 54720
LA VIRGINIA 66400
CHAPARRAL 73168
SARAVENA 81736
SAN JOSE DEL GUAVIARE 95001
PUERTO BERRIO 05579
ARENAL 13042
MAGANGUE 13430
EL PAUJIL 18256
MONTELIBANO 23466
PLANETA RICA 23555
SAN ANDRES SOTAVENTO 23670
COLOMBIA 41206
SAN JUAN DEL CESAR 44650
ARACATACA 47053
EL BANCO 47245

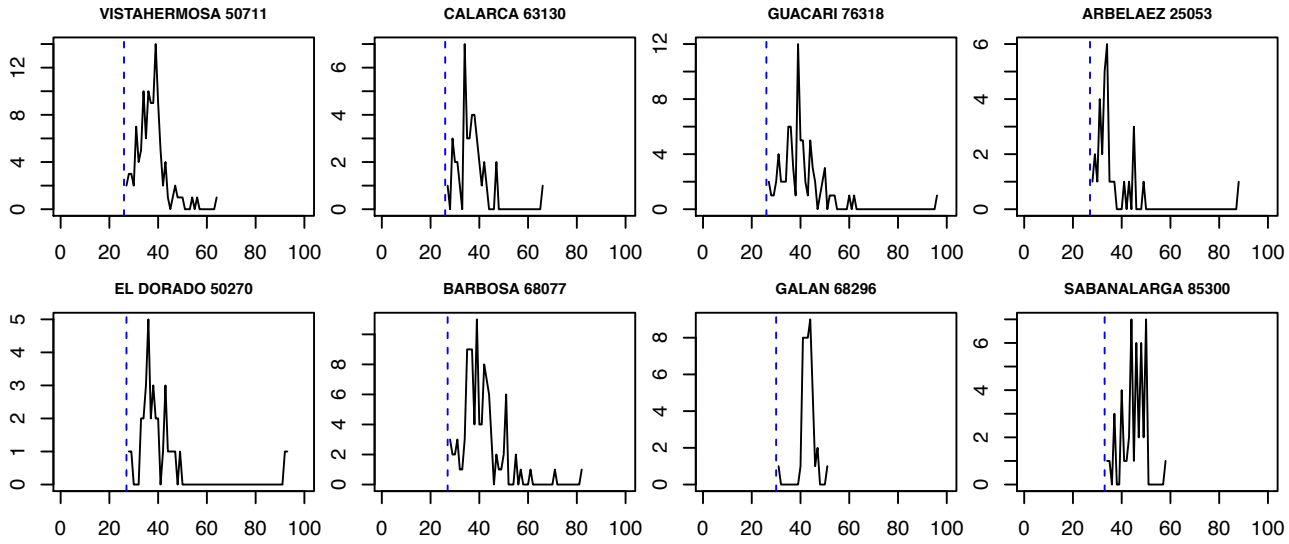

ZAPAYAN 47960
ZONA BANANERA 47980
SANTIAGO 54680
PIEDECUESTA 68547
SAN ANDRES 68669
COYAIMA 73217
PURIFICACION 73585
ARAUQUITA 81065
SABANAGRANDE 08634
CICUCO 13188
RIO DE ORO 20614
LA APARTADA 23350
SAN PELAYO 23686
ALGARROBO 47030
SITIO NUEVO 47745
RESTREPO 50606
ARBOLEDAS 54051
BOCHALEMA 54099
FLORIDABLANCA 68276
ROLDANILLO 76622
TULUA 76834
CAREPA 05147
AYAPEL 23068
CANALETE 23090
SAN BERNARDO DEL VIENTO 23675
AIPE 41016
FONSECA 44279
ARMENIA 63001


TRINIDAD 85430
PUERTO TEJADA 19573
GUADUAS 25320
ALGECIRAS 41020
PALERMO 41524
CHINACOTA 54172
EL TARRA 54250
SABANA DE TORRES 68655
ANSERMANUEVO 76041
PALMIRA 76520
YUMBO 76892
LA VEGA 25402
PITALITO 41551
RIVERA 41615
ACACIAS 50006
BARRANCABERMEJA 68081
SOCORRO 68755
CARTAGO 76147
LETICIA 91001
PUERTO TRIUNFO 05591
VITERBO 17877
EL DONCELLO 18247
SAN ALBERTO 20710
VIOTA 25878
PITAL 41548
EL PIÑON 47258
GRANADA 50313
SAN CARLOS DE GUAROA 50680


ABREGO 54003
DURANIA 54239
MONTENEGRO 63470
CAPITANEJO 68147
SAN VICENTE DE CHUCURI 68689
LIBANO 73411
PIEDRAS 73547
SAN ANTONIO 73675
ALCALA 76020
BUGA 76111
TORO 76823
ZARZAL 76895
SAN LUIS DE PALENQUE 85325
CACERES 05120
CHIGORODO 05172
LA MONTANITA 18410
SAN JOSE DEL FRAGUA 18610
BECERRIL 20045
CURUMANI 20228
CACHIPAY 25123
EL PEÑON 25258
SILVANIA 25743
BARRANCAS 44078
DIBULLA 44090
CUMARAL 50226
QUIMBAYA 63594
SAN GIL 68679
ICONONZO 73352


BUENAVENTURA 76109
CANDELARIA 76130
JAMUNDI 76364
VIJES 76869
YOTOCO 76890
FORTUL 81300
ORITO 86320
ANTIOQUIA 05042
NECOCLI 05490
SAN VICENTE DEL CAGUAN 18753
SANTANDER DE QUILICHAO 19698
MANAURE BALCON DEL CESAR 20443
SAN ANTONIO DE TEQUENDAMA 25645
AGRADO 41013
LA PLATA 41396
SANTA MARIA 41676
SUAZA 41770
TIMANA 41807
SAN JUAN DE ARAMA 50683
GIRON 68307
DOLORES 73236
FALAN 73270
ROVIRA 73624
ANDALUCIA 76036
DAGUA 76233
FLORIDA 76275
GINEBRA 76306
LA UNION 76400

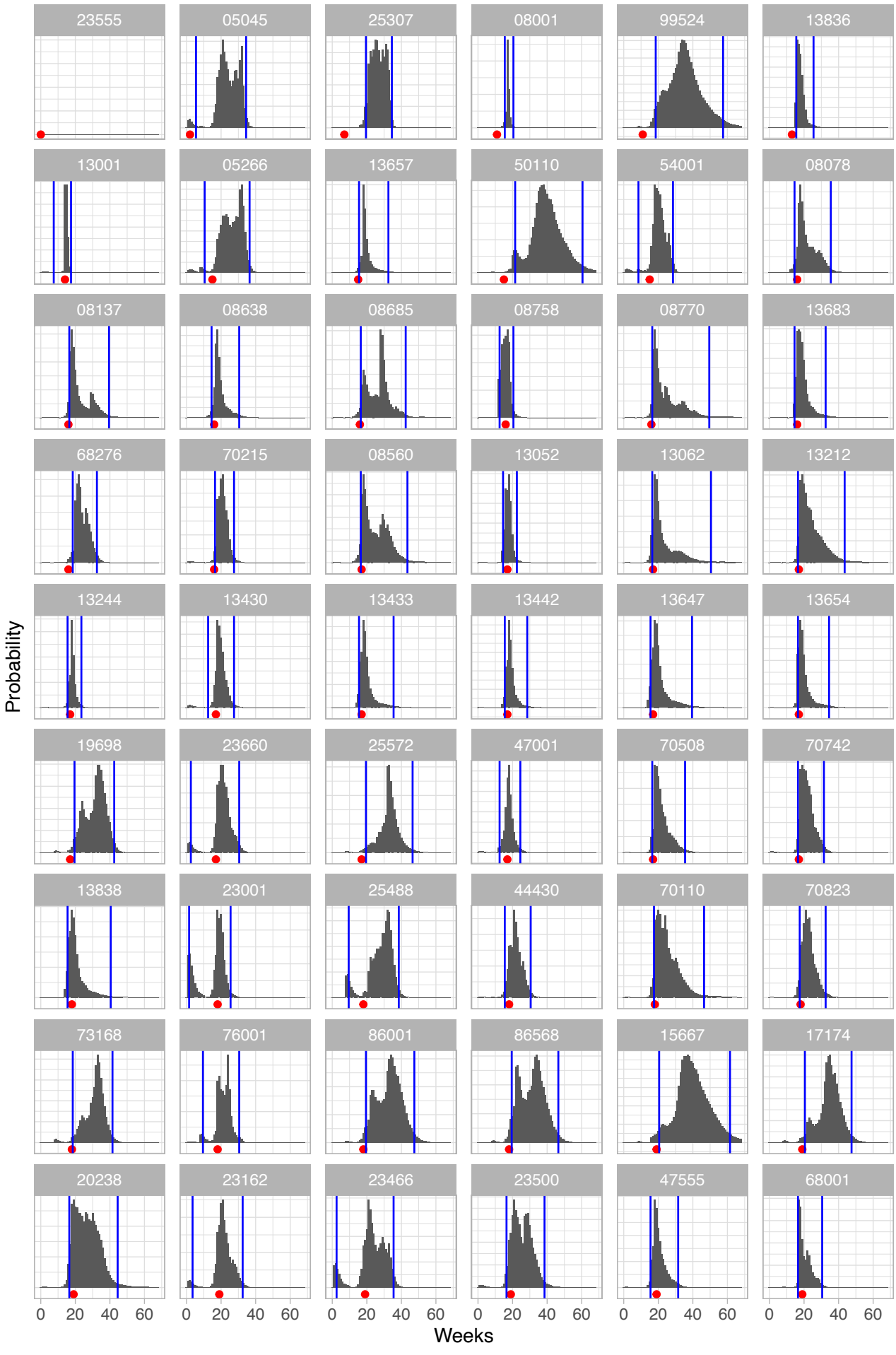

PRADERA 76563
SOPETRAN 05761
MARQUETALIA 17444
MEDINA 25438
PARATEBUENO 25530
ALTAMIRA 41026
CASTILLA LA NUEVA 50150
PUERTO RICO 50590
EL PLAYON 68255
RIONEGRO 68615
CAICEDONIA 76122
EL CERRITO 76248
LA VICTORIA 76403
TAME 81794
NUNCHIA 85225
PAZ DE ARIPORO 85250
PORE 85263
MUTATA 05480
NECHI 05495
ZARAGOZA 05895
GUAMAL 50318
CUCUTILLA 54223
TEORAMA 54800
BOLIVAR 76100
OBANDO 76497
PUERTO LEGUIZAMO 86573
CUMARIBO 99773
PUERTO LOPEZ 50573

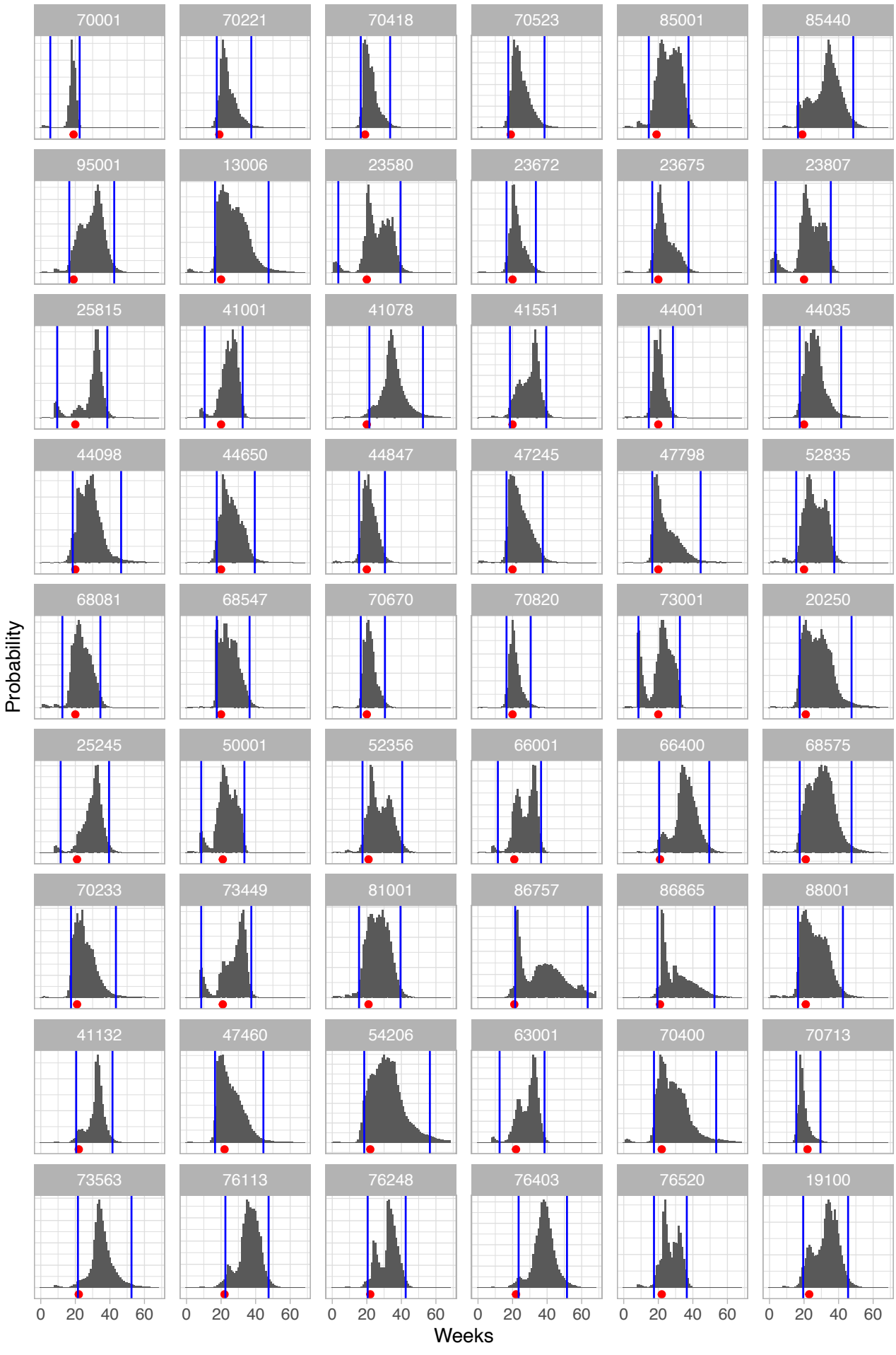

VISTAHERMOSA 50711
CALARCA 63130
GUACARI 76318
ARBELAEZ 25053
EL DORADO 50270
BARBOSA 68077
GALAN 68296
SABANALARGA 85300

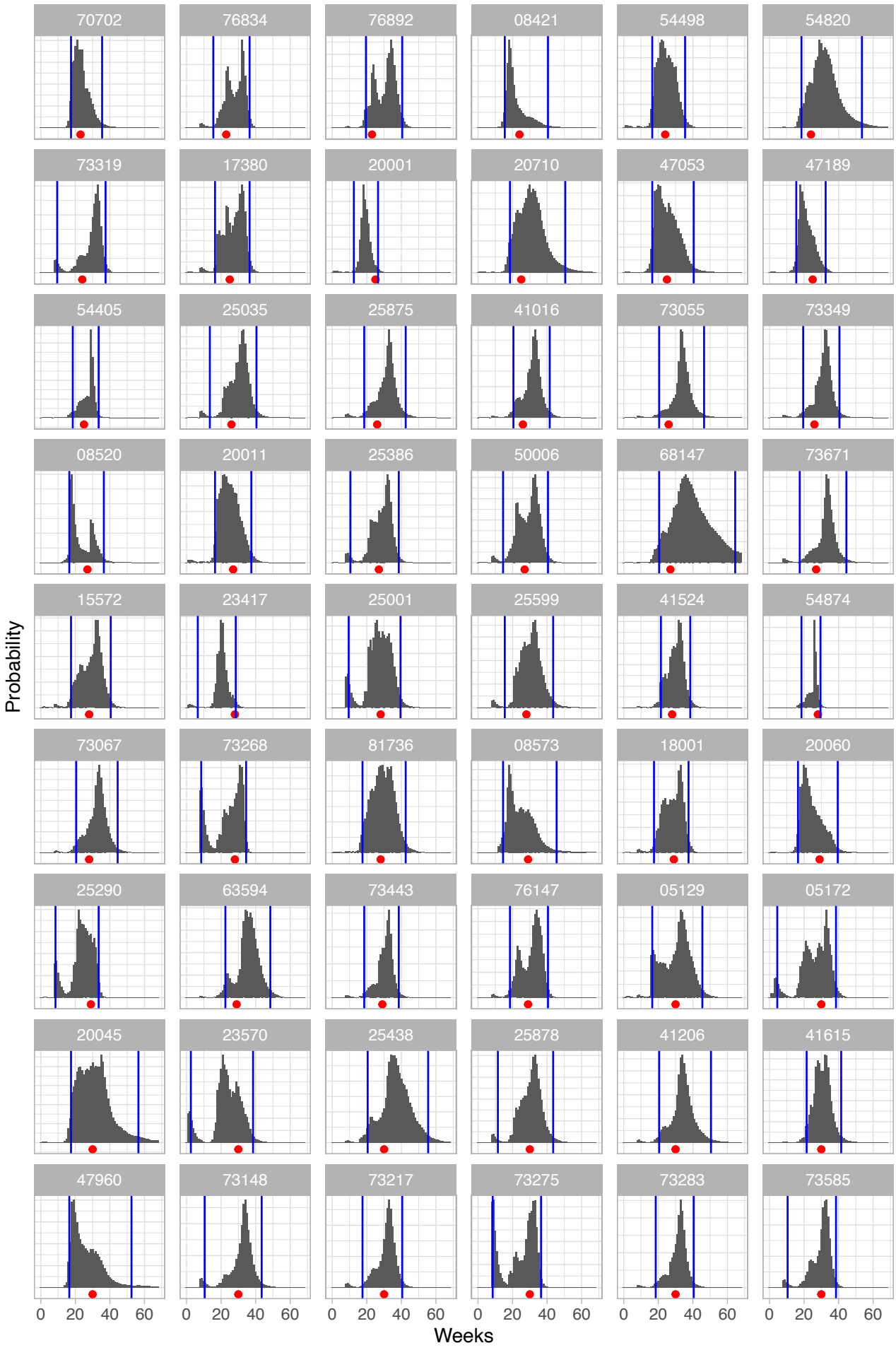
Appendix S5: Probability distribution for week of invasion by city

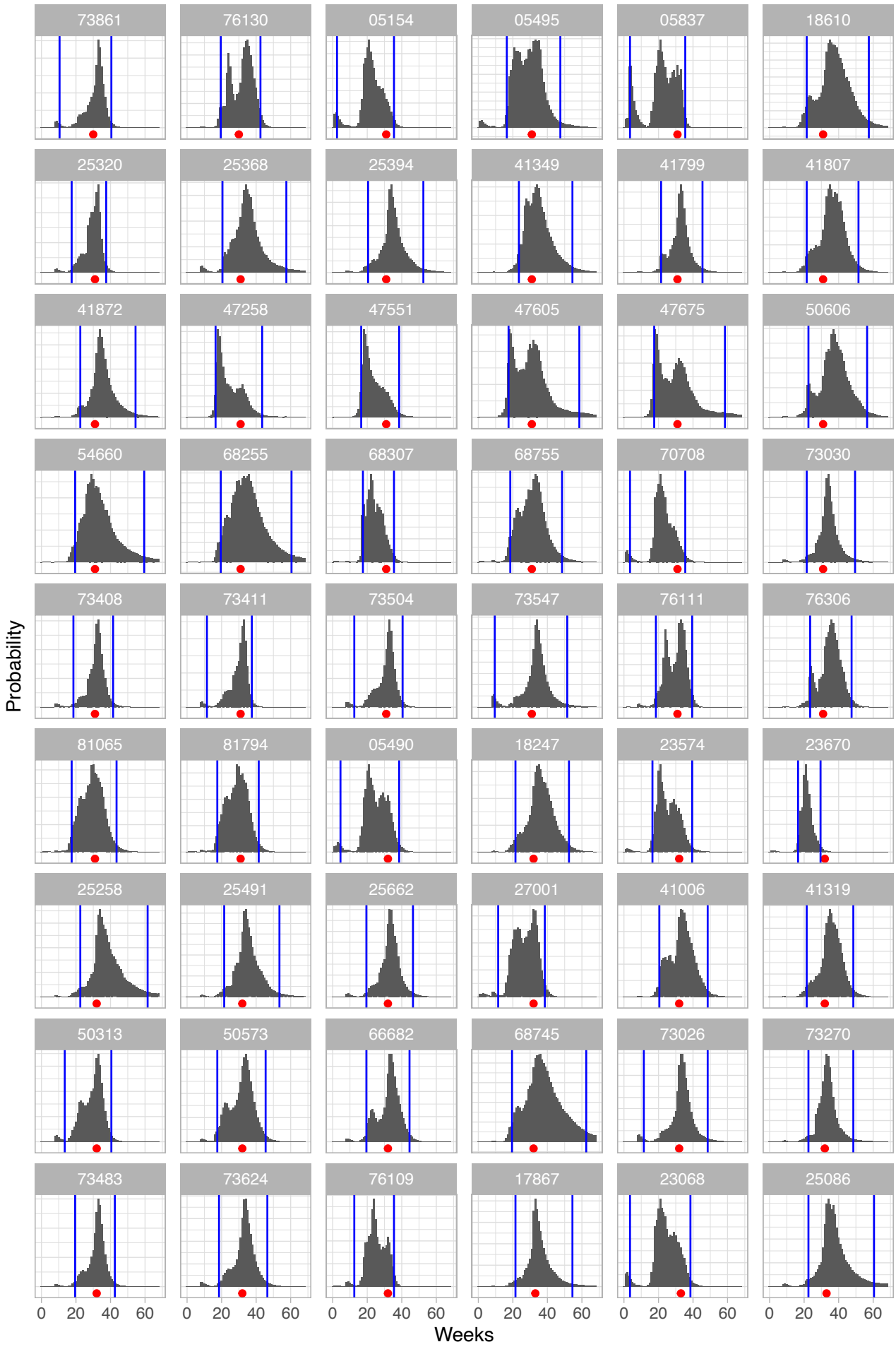
The results in this section are from the best-fitting CHIKV and ZIKV models, respectively. The probability distribution of the expected invasion week for each city is shown in gray. Ninety-five percent of the distribution is contained inside the vertical blue lines, and the red point is the observed week of invasion. Cities without gray bars and blue lines were invaded in week 0. The city code is shown above each plot, and y-axes differ between plots. As in Appendix S4, cities were sorted in ascending order by invasion week.

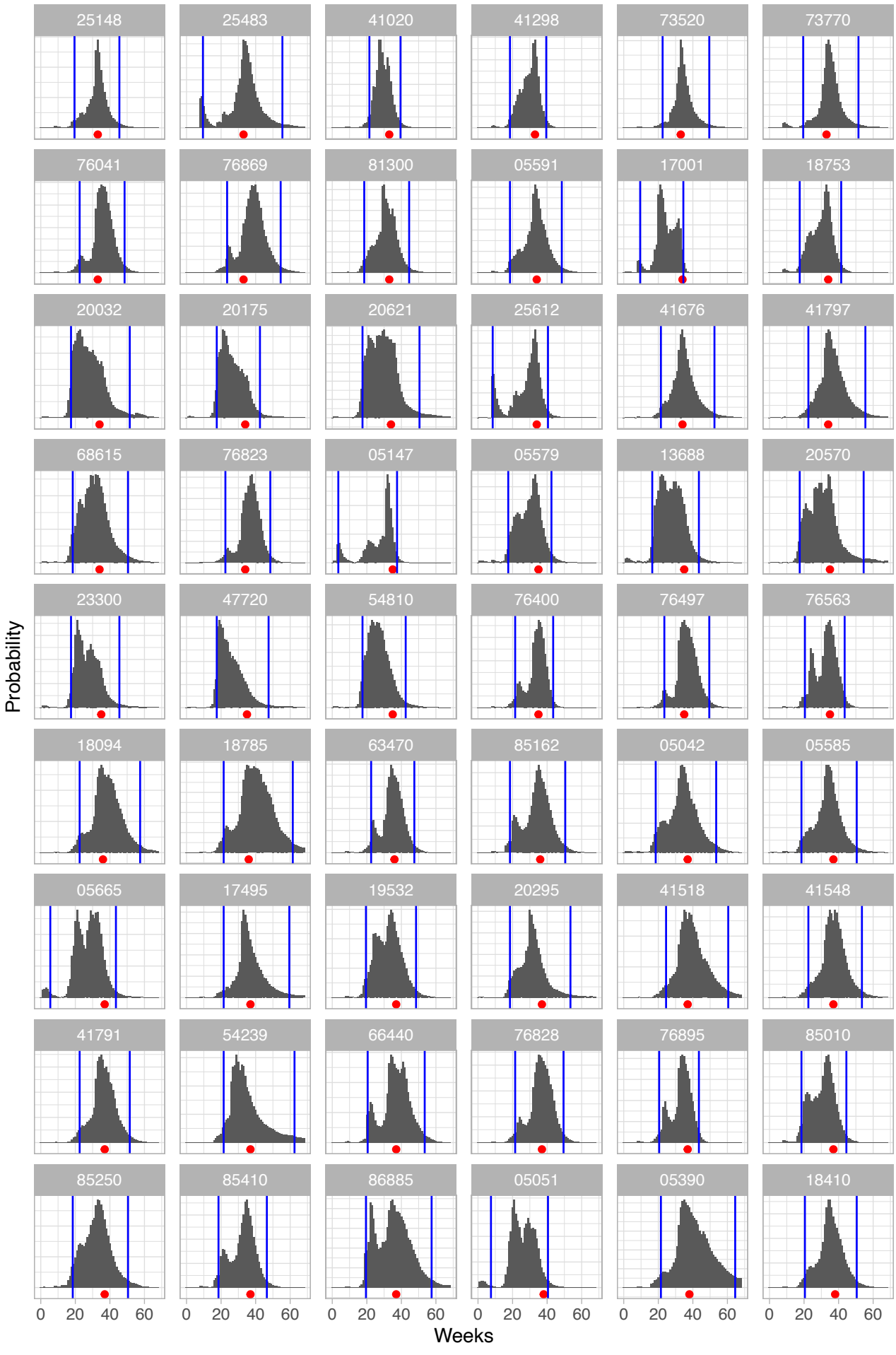
CHIKV


23555
05045
25307
08001
99524
13836
13001
05266
13657
50110
54001
08078
08137
08638
08685
08758
08770
13683
68276
70215
08560
13052
13062
13212
13244
13430
13433
13442
13647
13654
19698
23660
25572
47001
70508
70742
13838
23001
25488
44430
70110
70823
73168
76001
86001
86568
15667
17174
20238
23162
23466
23500
47555
68001
Probability
0 20 40 60
Weeks

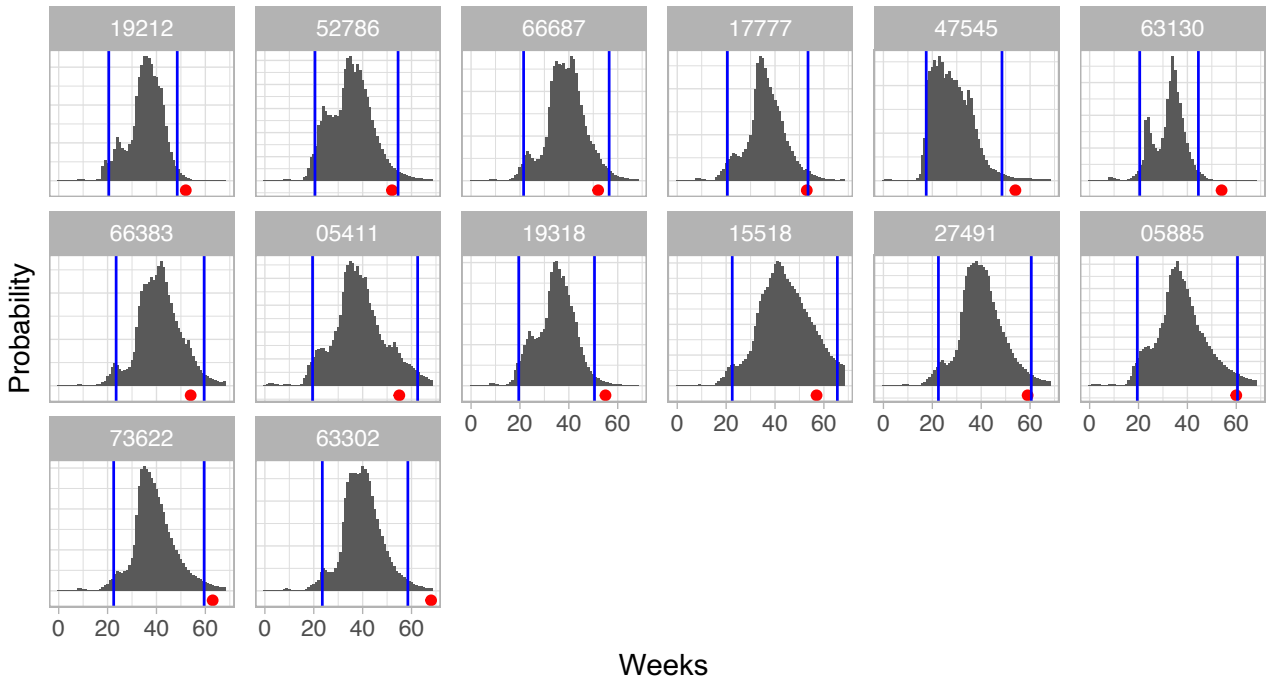

70001
70221
70418
70523
85001
85440
95001
13006
23580
23672
23675
23807
25815
41001
41078
41551
44001
44035
44098
44650
44847
47245
47798
52835
68081
68547
70670
70820
73001
20250
25245
50001
52356
66001
66400
68575
70233
73449
81001
86757
86865
88001
41132
47460
54206
63001
70400
70713
73563
76113
76248
76403
76520
19100
Probability
0
20
40
60
Weeks


70702
76834
76892
08421
54498
54820
73319
17380
20001
20710
47053
47189
54405
25035
25875
41016
73055
73349
08520
20011
25386
50006
68147
73671
15572
23417
25001
25599
41524
54874
73067
73268
81736
08573
18001
20060
25290
63594
73443
76147
05129
05172
20045
23570
25438
25878
41206
41615
47960
73148
73217
73275
73283
73585
Probability
0
20
40
60
Weeks

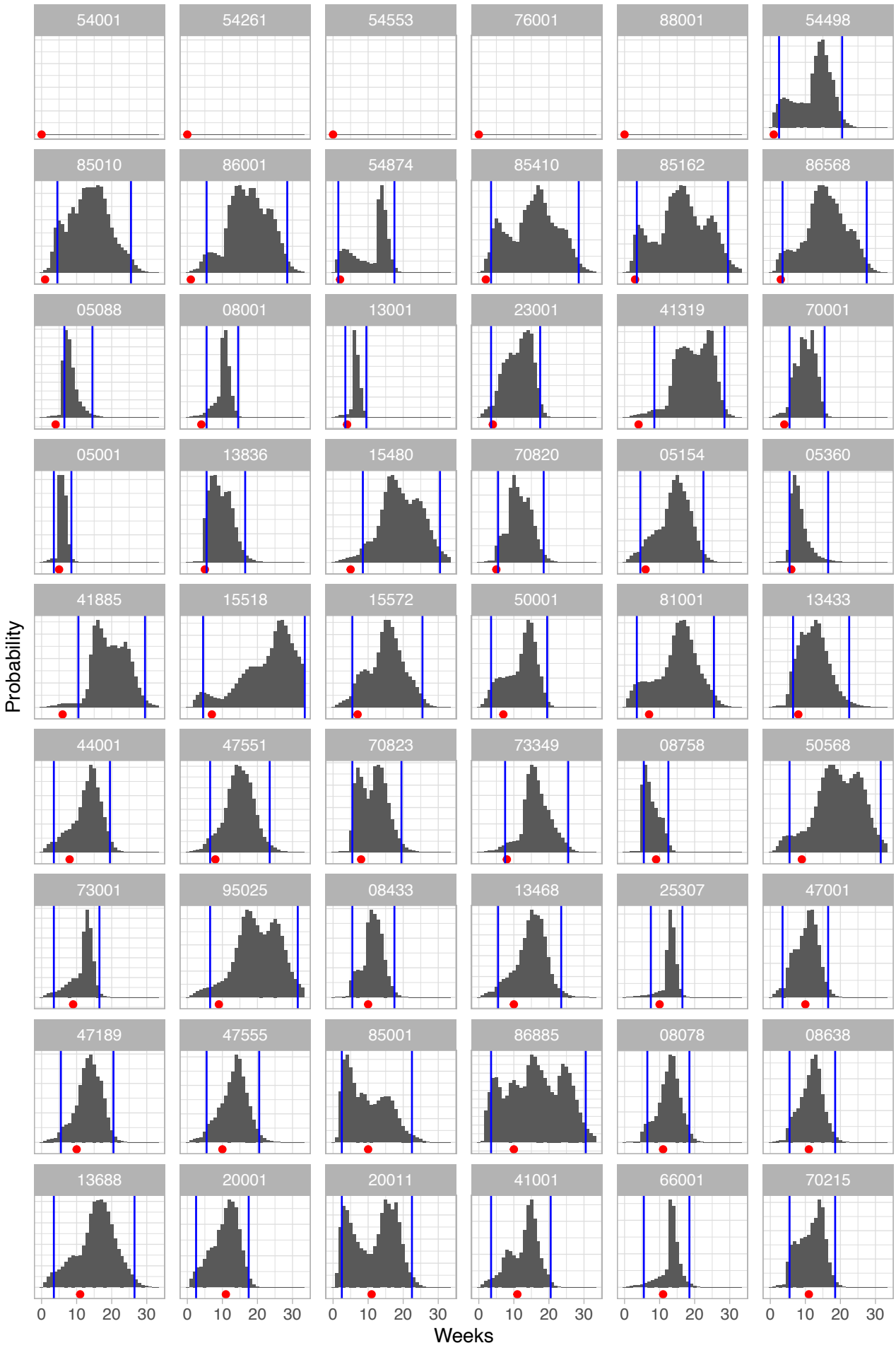

73861
76130
05154
05495
05837
18610
25320
25368
25394
41349
41799
41807
41872
47258
47551
47605
47675
50606
54660
68255
68307
68755
70708
73030
73408
73411
73504
73547
76111
76306
81065
81794
05490
18247
23574
23670
25258
25491
25662
27001
41006
41319
50313
50573
66682
68745
73026
73270
73483
73624
76109
17867
23068
25086
Probability
0 20 40 60
Weeks

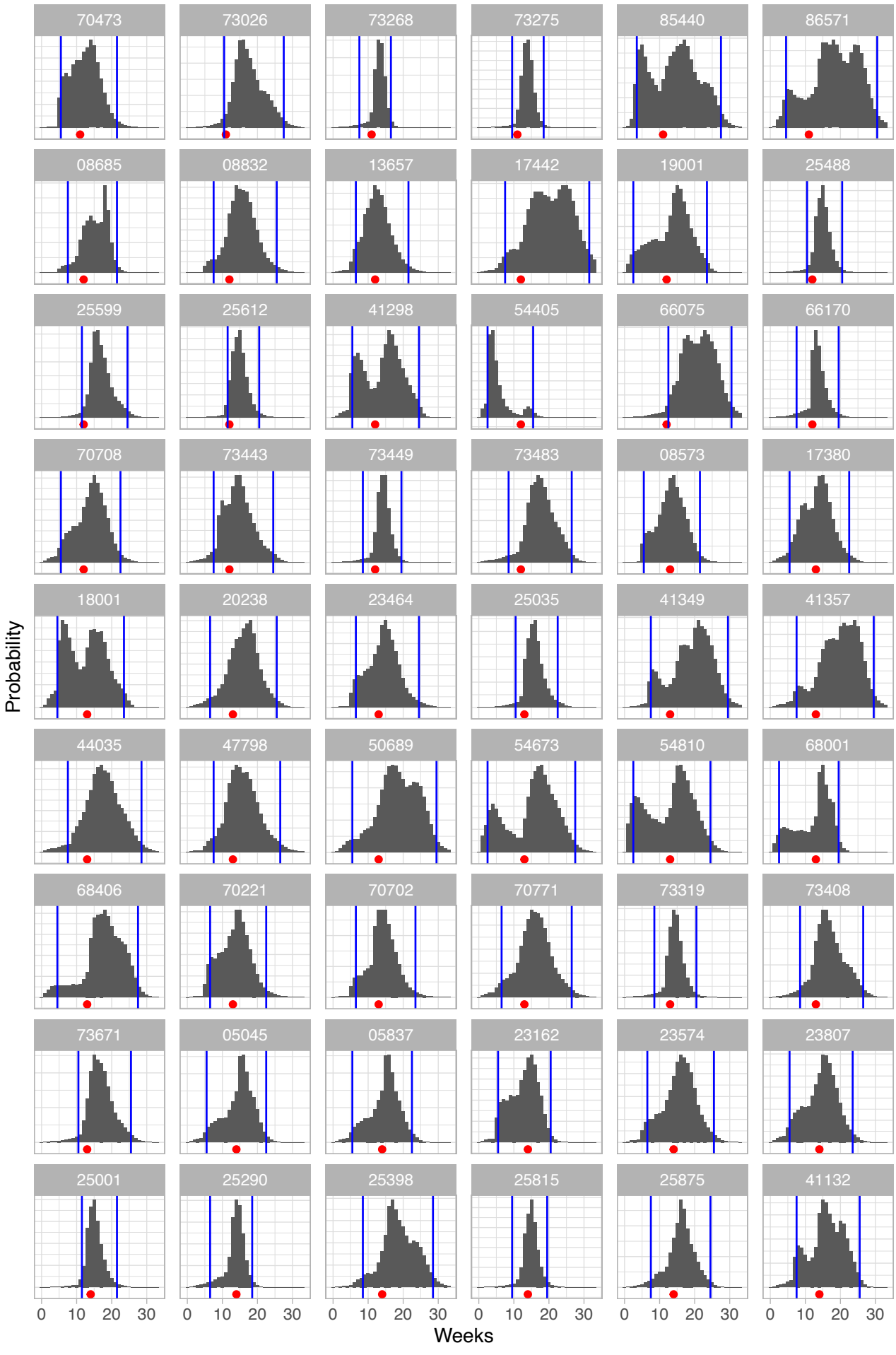

25148
25483
41020
41298
73520
73770
76041
76869
81300
05591
17001
18753
20032
20175
20621
25612
41676
41797
68615
76823
05147
05579
13688
20570
23300
47720
54810
76400
76497
76563
18094
18785
63470
85162
05042
05585
05665
17495
19532
20295
41518
41548
41791
54239
66440
76828
76895
85010
85250
85410
86885
05051
05390
18410
Probability
0
20
40
60
Weeks

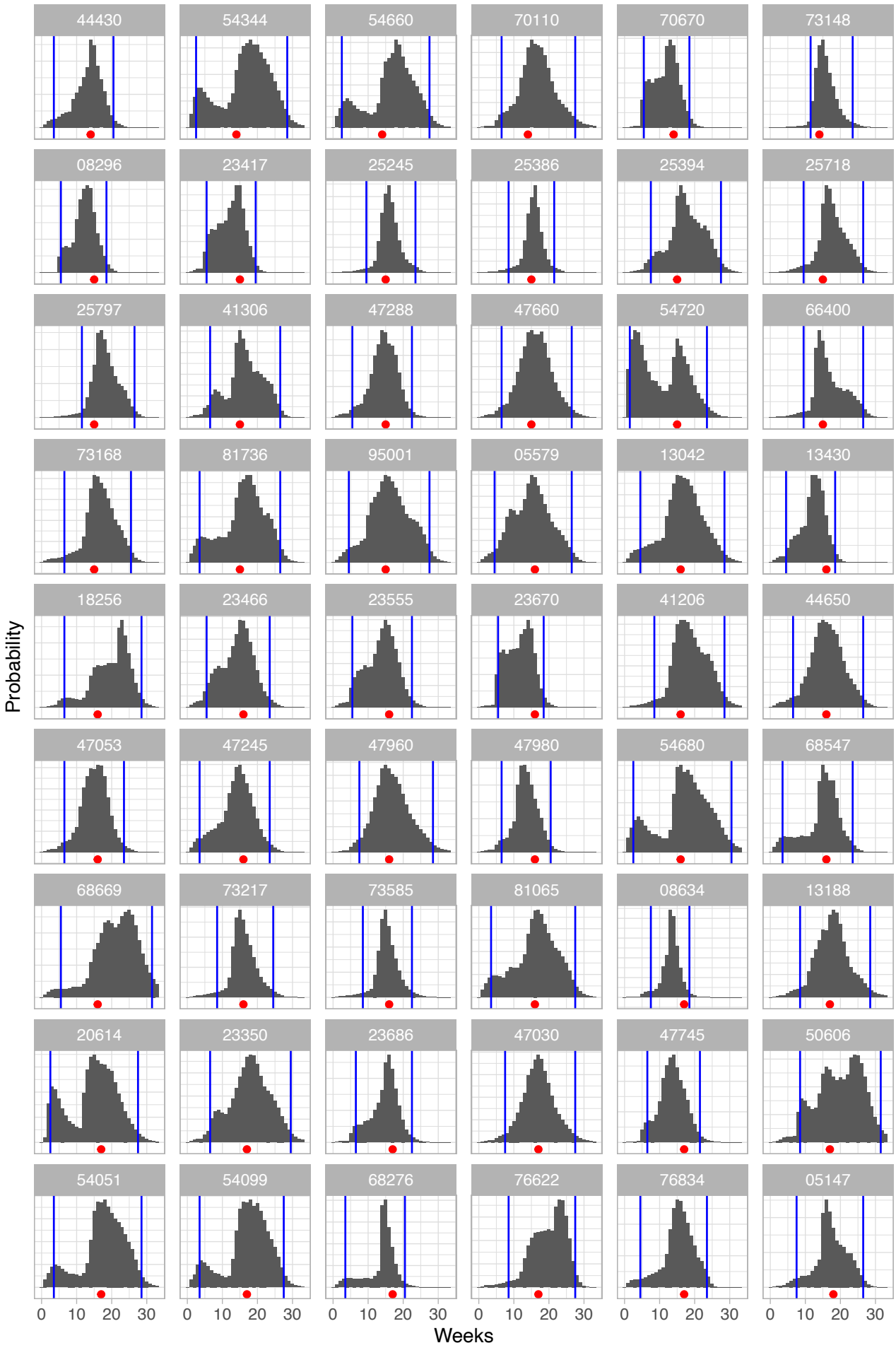

25053
47707
63272
73616
76122
85430
05756
19513
19845
27006
76622
76736
05656
17877
41396
41770
76863
19455
19809
25168
25530
50226
66594
70678
76233
76890
50223
50711
76036
76054
76318
76364
18205
66170
73236
86320
17013
99773
86571
18256
76020
76275
81220
27810
63401
05761
85139
85263
05736
66045
68406
76845
05628
17442
Probability
0
20
40
60
Weeks

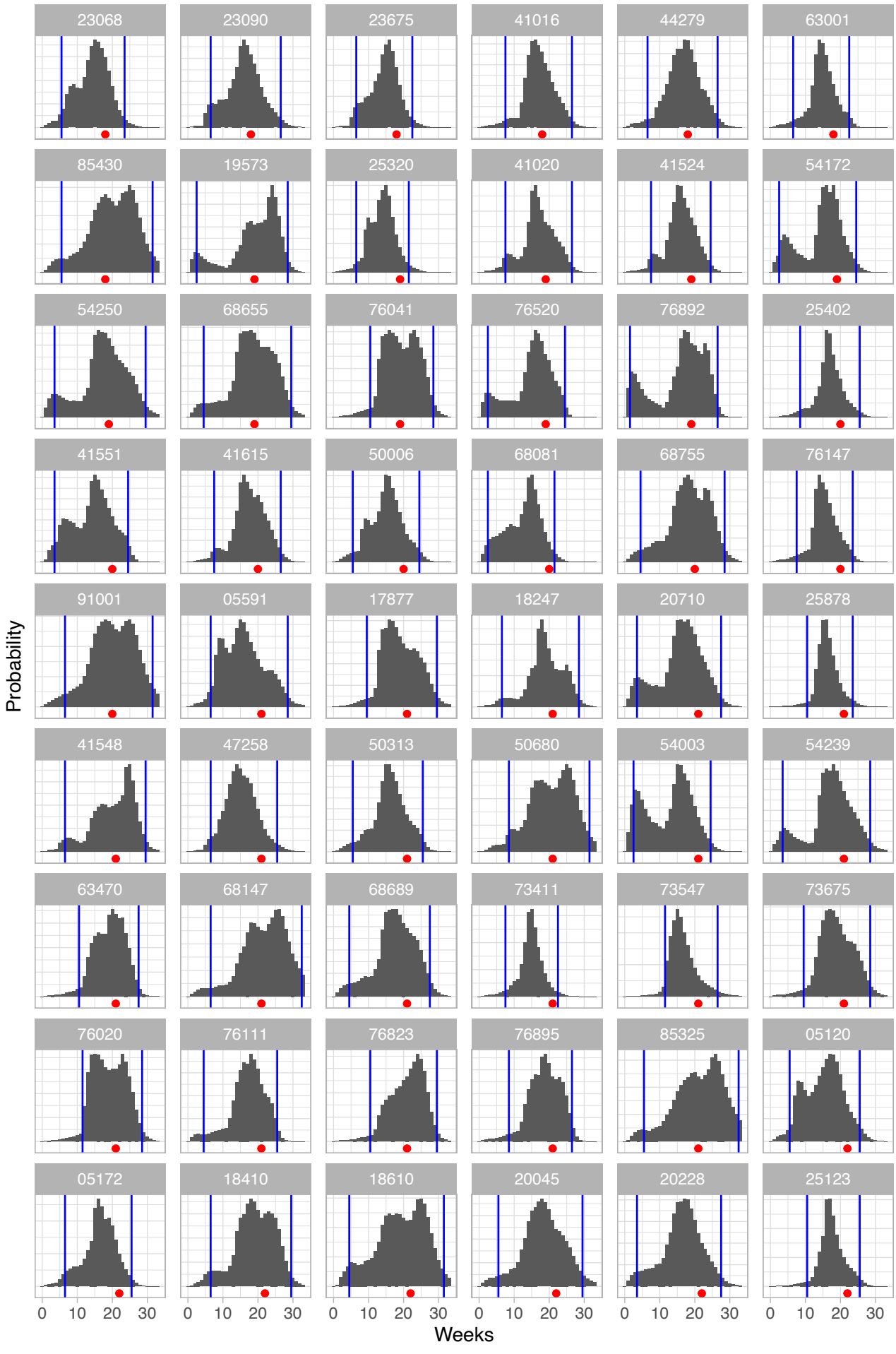

19212
52786
66687
17777
47545
63130
66383
05411
19318
15518
27491
05885
73622
63302
Probability
0 20 40 60
Weeks

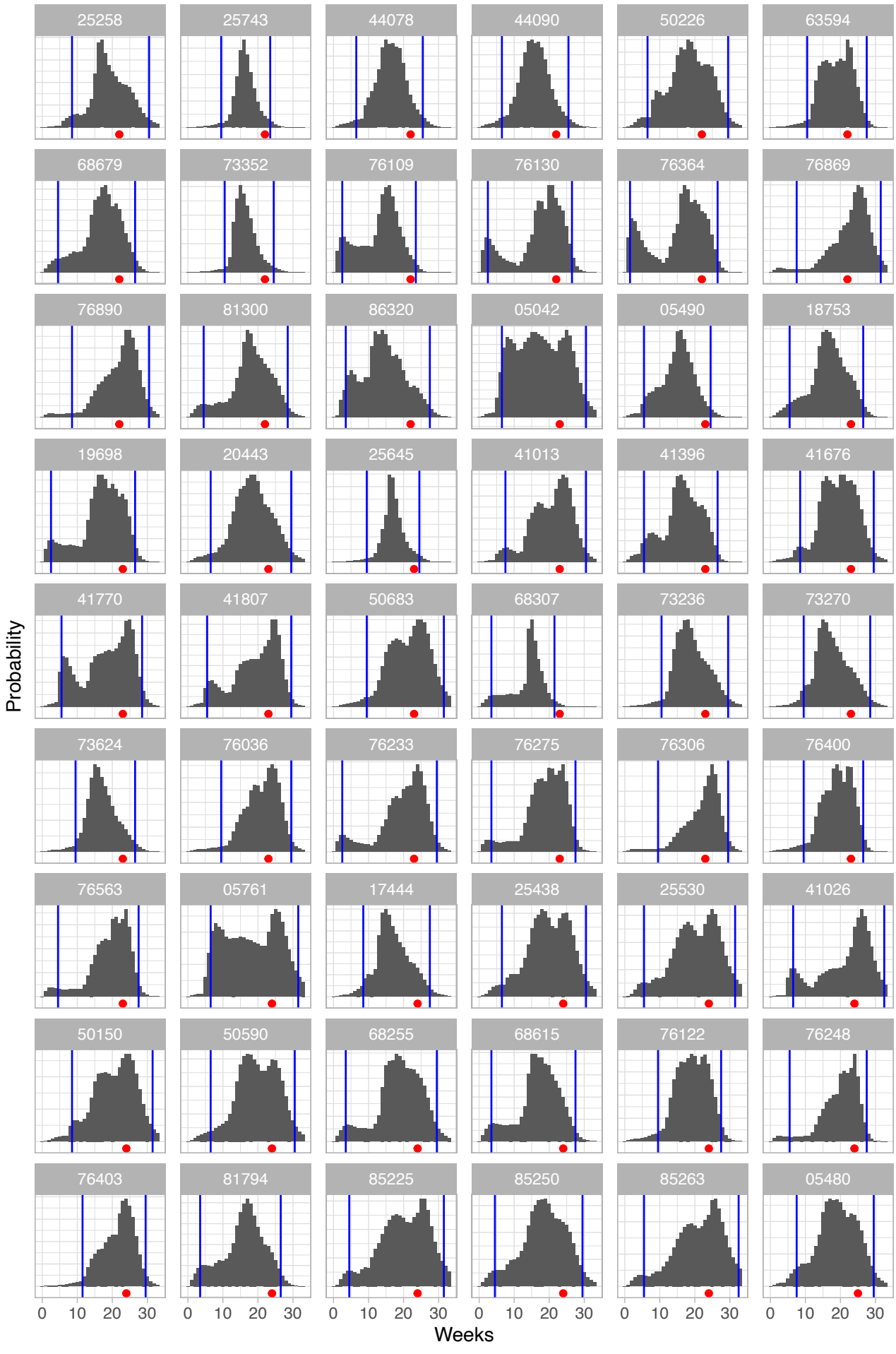
ZIKV

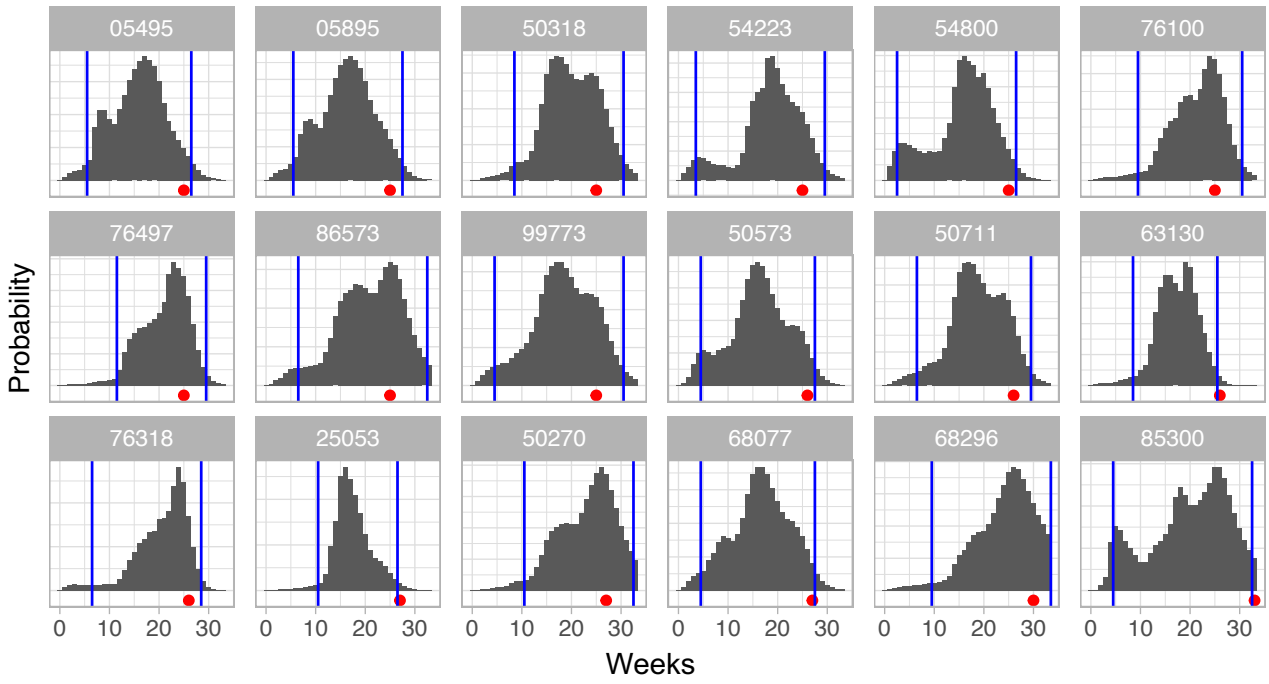

54001
54261
54553
76001
88001
54498
85010
86001
54874
85410
85162
86568
05088
08001
13001
23001
41319
70001
05001
13836
15480
70820
05154
05360
41885
15518
15572
50001
81001
13433
44001
47551
70823
73349
08758
50568
73001
95025
08433
13468
25307
47001
47189
47555
85001
86885
08078
08638
13688
20001
20011
41001
66001
70215
Probability
0 10 20 30
Weeks


70473
73026
73268
73275
85440
86571
08685
08832
13657
17442
19001
25488
25599
25612
41298
54405
66075
66170
70708
73443
73449
73483
08573
17380
18001
20238
23464
25035
41349
41357
44035
47798
50689
54673
54810
68001
68406
70221
70702
70771
73319
73408
73671
05045
05837
23162
23574
23807
25001
25290
25398
25815
25875
41132
Probability
0
10
20
30
Weeks


44430
54344
54660
70110
70670
73148
08296
23417
25245
25386
25394
25718
25797
41306
47288
47660
54720
66400
73168
81736
95001
05579
13042
13430
18256
23466
23555
23670
41206
44650
47053
47245
47960
47980
54680
68547
68669
73217
73585
81065
08634
13188
20614
23350
23686
47030
47745
50606
54051
54099
68276
76622
76834
05147
Probability
0
10
20
30
Weeks


23068
23090
23675
41016
44279
63001
85430
19573
25320
41020
41524
54172
54250
68655
76041
76520
76892
25402
41551
41615
50006
68081
68755
76147
91001
05591
17877
18247
20710
25878
41548
47258
50313
50680
54003
54239
63470
68147
68689
73411
73547
73675
76020
76111
76823
76895
85325
05120
05172
18410
18610
20045
20228
25123
Probability
0
10
20
30
Weeks


25258
25743
44078
44090
50226
63594
68679
73352
76109
76130
76364
76869
76890
81300
86320
05042
05490
18753
19698
20443
25645
41013
41396
41676
41770
41807
50683
68307
73236
73270
73624
76036
76233
76275
76306
76400
76563
05761
17444
25438
25530
41026
50150
50590
68255
68615
76122
76248
76403
81794
85225
85250
85263
05480
Probability
0
10
20
30
Weeks


05495
05895
50318
54223
54800
76100
76497
86573
99773
50573
50711
63130
76318
25053
50270
68077
68296
85300
Probability
0
10
20
30
Weeks