




Estimating transmissibility of Zika virus in Colombia in the presence of surveillance bias

Received: 19 January 2025

Accepted: 28 April 2025

Published online: 08 May 2025

 Check for updatesTim K. Tsang ^{1,2}✉, Diana P. Rojas³, Fei Xu⁴, Yanfang Xu¹, Xiaolin Zhu⁴, M. Elizabeth Halloran ^{5,6}, Ira M. Longini² & Yang Yang^{2,7}✉

The 2015–2016 Zika virus outbreak in the Americas presented significant challenges in understanding the transmission dynamics due to substantial reporting biases, as women of reproductive age (15–39 years) were disproportionately represented in the surveillance data when public awareness of relationship between Zika and microcephaly increased. Using national surveillance data from Colombia during July 27, 2015–November 21, 2016, we developed a Bayesian hierarchical modeling framework to reconstruct the true numbers of symptomatic cases and estimate transmission parameters while accounting for differential reporting across age-sex groups. Our model revealed that the detection rate of symptomatic cases among women of reproductive age was 99% (95% CI: 98.7–100), compared to 85.4% (95% CI: 84.7–86.1) in other demographic groups. After correcting for these biases, our results showed that females aged 15–39 years remained 82.8% (95% CI: 80.2–85.2%) more susceptible to Zika symptomatic infection than males of the same age, independent of differential reporting areas. Departments with medium-high altitude, medium-high population density, low coverage of forest, or high dengue incidence from 2011–2015 exhibited greater Zika risk. This study underscores the importance of accounting for surveillance biases in epidemiological studies to better understand factors influencing Zika transmission and to inform disease control and prevention.

The Zika virus is an emerging infectious disease first identified in 1947 in the Zika Forest of Uganda¹. During 2015–2016, the Zika virus spread across the Americas, including the Caribbean countries², likely due to their geographic location within the range of *Aedes aegypti* mosquitoes, the major vector of the Zika virus³. Environmental and climatic conditions in the Americas such as high population density, warm temperature and abundant precipitation provide suitable ecological niches for mosquitoes^{4–10} and consequently the outbreak in the Americas was extensive, affecting 48 countries and territories as of

December 2017⁴. In Colombia, preliminary monitoring of Zika began following the recognition of the outbreak in Brazil in May 2015. The Colombian Instituto Nacional de Salud (INS) initiated official surveillance for Zika in August 2015. By early October 2015, a Zika outbreak was declared after a cluster of laboratory-confirmed cases was identified among nine patients in northern Colombia.

Although most Zika infections result in a self-limiting, dengue-like illness characterized by exanthema, conjunctivitis, arthralgia, and low-grade fever⁵, severe outcomes such as microcephaly and Guillain-Barré

¹World Health Organization Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China. ²Department of Biostatistics, University of Florida, Gainesville, FL, USA.

³Department of Epidemiology, University of Florida, Gainesville, FL, USA. ⁴Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region, China. ⁵Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center, Seattle, WA, USA.

⁶Department of Biostatistics, University of Washington, Seattle, WA, USA. ⁷Department of Statistics, Franklin College of Arts and Sciences, University of Georgia, Athens, GA, USA. ✉e-mail: timtsang@connect.hku.hk; yang.yang4@uga.edu

syndrome were recognized in Brazil in 2015 and the World Health Organization (WHO) declared a Public Health Emergency of International Concern in February 2016⁶. This recognition led to intensified surveillance and control efforts in many countries². Although large outbreaks of Zika have not been reported since 2016, it continued to emerge and reemerge, e.g., locally infected cases were first reported in southern France in 2019, and India reported hundreds of cases in 2021 during the COVID-19 pandemic^{7,8}. Given the association of Zika infection with Congenital Zika Syndrome and other neurological complications, it is crucial to understand Zika epidemiology and transmission characteristics for effective disease control.

These associations had also driven surveillance bias where case screening and detection efforts targeted pregnant women and women of reproductive age more than other demographic groups^{9–11}, potentially introducing bias into epidemiological analyses of the transmission dynamics. Thus far, previous studies addressing surveillance bias in the Zika epidemic have predominantly employed hierarchical Bayesian models integrating multiple data streams, such as seroprevalence surveys among blood donors, to estimate infection attack rates and other epidemiological parameters^{12–15}. While these analyses tackled challenges like under-ascertainment and misdiagnosis, most models assumed a uniform reporting probability across demographic groups^{12–15}, despite many had noted that certain demographic groups, particularly women of reproductive age, might be overrepresented in surveillance data¹².

Here, we develop a Bayesian model to account for reporting differences among age-sex groups when estimating transmission dynamics. We chose a Bayesian approach because it naturally accommodates multi-level data structures with a substantial amount of latent data, while integrating prior information from the literature and experts, and providing robust uncertainty quantification. These features make it particularly suitable for our complex, heterogeneous surveillance data, e.g., unobserved true numbers of symptomatic cases across space-time and differential reporting rates across demographic groups. We apply this model to the Zika outbreak in Colombia during 2015–2016. Our approach jointly estimates reporting differences across age-sex groups and identifies factors influencing transmission dynamics, including biological effects of age and sex, meteorological variables such as temperature and precipitation, and geographical factors like population density and altitude.

Results

Zika outbreak in Colombia, 2015–2016

The study period for this analysis spanned from July 27, 2015, to November 21, 2016. During this time, a total of 103,875 symptomatic cases were reported in Colombia, and 1283 (1.2%) of them were excluded due to missing onset date or date of birth. Hence, 102,592 symptomatic cases were included in further analyses, of whom 8048 (7.8%) were laboratory-confirmed (Fig. 1A). The number of cases among women of reproductive age (defined as ages 15–39) began to increase significantly around December 2015, coinciding with rising public awareness of the association between Zika and microcephaly (Fig. 1B). Overall, the reported attack rates among women of reproductive age were higher than other age-sex group in all departments (Fig. 2). Before December 1, 2015, women of reproductive age accounted for 25.4% of all reported cases across age and sex groups. This proportion increased to 34.5% over the course of the entire outbreak. Based on surveillance data, the observed attack rate for women of reproductive age was 4.34 per 1000 individuals, higher than other age-sex groups, which ranged from 1.26 to 2.15 per 1000 individuals (Fig. 2, and Supplementary Figs. S1–2). We also plotted the epidemics curve by age-sex group for the 5 regions (Caribbean, Andean, Amazon, Pacific and Orinoquia; Supplementary Fig. S2) in Colombia (Fig. 3). The trend of the number of female cases of reproductive age far exceeding that of male cases in the same age group, this potential surveillance

bias started as early as November 2015 in the Caribbean Region and persisted throughout the outbreaks in each department. These observations suggest that the probability of symptomatic cases being captured by the surveillance system (hereafter referred to as reporting probability) likely varied by age and sex, particularly among women of reproductive age, due to heightened awareness of Zika-related microcephaly.

Cumulative attack rate and reporting probability

To address potential differences in reporting probabilities among age-sex groups when estimating the transmission dynamics of Zika in Colombia, we developed a Bayesian hierarchical model. The model treated the true weekly numbers of symptomatic cases for each age-sex group as latent variables and linked these latent symptomatic case numbers to the observed weekly numbers of symptomatic cases via a sub-model of under-reporting. We assumed that the surveillance capacity increased linearly across all age-sex groups, starting from an initial reporting probability on July 14, which is one tenth of the final reporting probability, and reaching a final reporting probability by November 30, 2015, and staying at the same capacity afterwards. To avoid non-identifiability, we assume all age-sex groups shared the same reporting probabilities except those women of reproductive age had a separate final reporting probability. The final reporting probability was estimated to be 0.990 (95% CI: 0.987–1.00) for women of reproductive age, compared to 0.854 (95% CI: 0.847–0.861) for all other age-sex groups.

Using the posterior samples of the latent true symptomatic case numbers, we estimated the true attack rates for each age-sex group, defined as the number of symptomatic cases divided by the population size for each group (Fig. 4A). The true attack rate for women of reproductive age was estimated at 4.92 (95% CI: 4.91–4.94) per 1000 individuals, about 13.4% higher than the observed attack rate. For all other age-sex groups, the true attack rates ranged from 1.57 to 2.77 per 1000 individuals, about 24.4% to 36.2% higher than the observed attack rates.

Transmission dynamics

After adjusting for differential reporting probabilities, our model revealed significant geographical variation in attack rates (Fig. 4B). Estimated attack rates in the Caribbean, Andean, and Amazon regions ranged from 2.14 to 2.59 per 1000 individuals, while the Pacific and Orinoquia regions exhibited higher rates at 3.71 (95% CI: 3.69–3.73) and 6.64 (95% CI: 6.59–6.71) per 1000 individuals, respectively. The estimated attack rates for each department were shown in Supplementary Fig. S3.

We estimated that children aged 0–14 and adults aged 40 or older were 52.7% (95% CI: 51.9–53.6%) and 38.4% (95% CI: 37.4–39.3%) less susceptible than adults aged 15–39 (Fig. 5). Females were 82.8% (95% CI: 80.2–85.2%) more susceptible than males. Environmental factors also influenced susceptibility. A 1°C increase in temperature was associated with a 13.0% increase in susceptibility to symptomatic infection (odds ratios (OR): 1.130; 95% CI: 1.125–1.135), while a 1 cm/hr increase in precipitation corresponded to a 9.4% reduction in susceptibility to symptomatic infection (OR: 0.906; 95% CI: 0.896–0.916).

Departments with low dengue incidence (0.09–0.57 cases per 100 person-years) or medium dengue incidence (0.58–1.81 cases per 100 person-years) exhibited significantly lower odds of Zika transmission compared to high-incidence departments (1.82–4.33 cases per 100 person-years), with odds ratios of symptomatic infections of 0.280 (95% CI: 0.270–0.290) and 0.228 (95% CI: 0.219–0.237), respectively. Compared to departments with a high forest coverage (16%–45%), departments with a low forest coverage (0%–5%) exhibited substantially increased odds of transmission (OR = 9.55, 95% CI: 9.06–10.08), while those with a medium forest coverage (0.06–0.15) showed only a slight increase (OR: 1.05; 95% CI: 1.02–1.07).

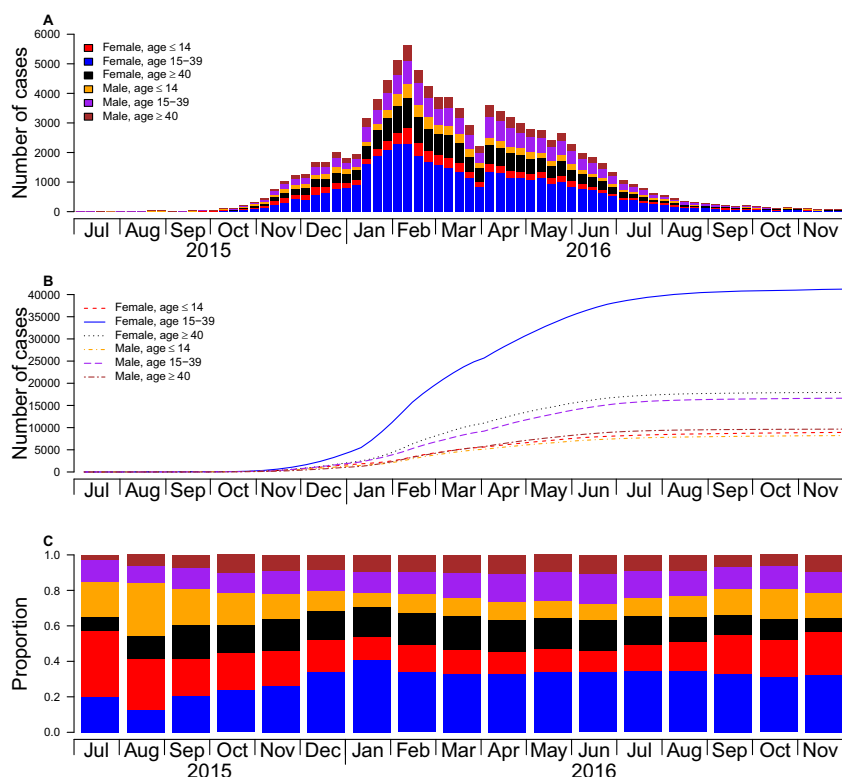


Fig. 1 | Epidemic curve for the Zika outbreak in Colombia from July 27, 2015 to November 21, 2016. A Weekly numbers of cases across six age-sex groups. **B** Cumulative numbers of cases across age-sex groups. **C** Proportions of monthly cases across age-sex groups.

Population density also influenced Zika transmission. Compared to densely populated departments (67–1470 people per km²), those with a low population density (0.6–24 people per km²) had significantly reduced odds of transmission (OR = 0.411, 95% CI: 0.337–0.489), while those with a medium population density (25–66 people per km²) were associated with slightly higher odds (OR = 1.171, 95% CI: 1.135–1.207). Altitude further contributed to variations in transmission risk. In reference to departments with high altitudes (1438–2411 meters), departments at low altitudes (53–215 meters) had 17% lower odds of transmission (OR = 0.83, 95% CI: 0.78–0.88), whereas departments at medium altitudes (216–1437 meters) exhibited 24% higher odds (OR = 1.24, 95% CI: 1.21–1.27).

Model adequacy and validation

We assessed the model adequacy by comparing model-predicted numbers of reported cases with observed surveillance data (Supplementary Information, Section 3.1). The results showed close alignment (Supplementary Fig. S4), suggesting a satisfactory goodness of fit. A simulation study was also conducted to validate the algorithm's ability to recover unknown parameters (Supplementary Information, Section 3.3). The results indicated minimal bias in model estimates of the parameters, and 84–100% of the 95% credible intervals contained the true parameter values across 50 simulated epidemics (Supplementary Information, Table S1).

Comparison analysis

To assess how reporting bias might affect the quantification of predictors' effects, we reran the analyses using a model ignoring the reporting bias (Supplementary Fig. S5). We observed that the directions (positive vs. negative) of the effects of predictors remain consistent between this naïve model and our main model, but there are two notable differences in magnitude. The most substantial differences were observed in the effects of low (vs. high) population density and low (vs. high) proportion of forest coverage, with changes of 50%

and 16% respectively. Other covariates showed smaller relative differences in odds ratios, generally < 10%. This comparison further confirms the importance of adjustment for reporting bias.

Sensitivity analysis

In a sensitivity analysis, when the initial reporting probability was assumed 5% of the final reporting probability, the final reporting probability was estimated at 0.990 (95% CI: 0.989–1.00) for women of reproductive age and 0.836 (95% CI: 0.828–0.846) for other groups, close to the primary estimates based on 10%. The assumption of 20% for the initial reporting capacity produced similar estimates of 0.990 (95% CI: 0.989–1.00) for women and 0.848 (95% CI: 0.842–0.856) for others. All other model parameters also remained similar (see Supplementary Figs. S6–S7 for the 0.05 scenario and Supplementary Figs. S8–S9 for the 0.20 scenario), indicating that our results are robust to the assumption of the initial reporting capacity.

In the primary analysis we assumed the full reporting capacity was reached from 10% to 100% in 20 weeks before Dec. 1, 2015, and at that day there was another instant jump in the reporting probability for women of reproductive age, to differentiate this group from others. In a sensitivity analysis, we allowed that jump to fully scale up in 2 weeks instead of instantly. The estimated reporting probability was 0.990 (95% CI: 0.989–1.00) for women of reproductive age and 0.854 (95% CI: 0.846–0.861) for other groups. Similarly, allowing a 4-week scaling-up produced estimates of 0.990 (95% CI: 0.989–1.00) and 0.855 (95% CI: 0.846–0.863), respectively, all similar to the primary results. Notably, the similarity holds for all model parameters (Supplementary Figs. S10–S13), demonstrating the robustness of our findings to the assumption about the temporal dynamics of reporting changes.

Discussion

In response to emerging public health threats, it is not uncommon for surveillance systems to be biased towards vulnerable subpopulations, which is exemplified by the multinational Zika outbreaks during 2015–

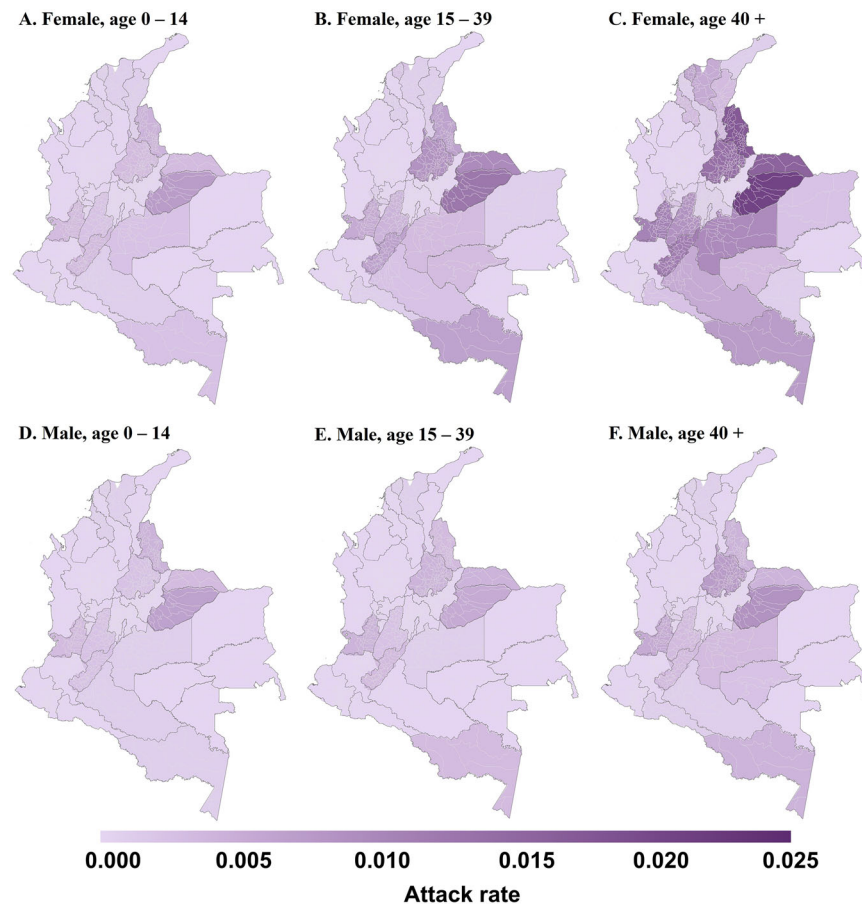


Fig. 2 | Zika Attack Rate Heat Map Across Demographic Groups in Colombia. The maps display the Zika virus attack rates stratified by age group and gender across municipalities in Colombia. **A–C** represent attack rates for females, while **(D–F)** represent males. Each panel is further subdivided into three age categories:

0–14 years (Panels A and D), 15–39 years (**B, E**), and 40+ years (**C, F**). The intensity of the purple shading corresponds to the attack rate, with darker shades indicating higher attack rates.

2016^{16–20}. In this study, we proposed a Bayesian hierarchical modeling framework to account for differential reporting of Zika cases among women of reproductive age in comparison to other age-sex groups, which provided more accurate understanding of the transmission burden and associated risk modifiers of Zika. Our analysis suggested that the detection proportion of symptomatic cases in Colombia increased from 8.5% to 86% in four months since the outbreak began, and detection rates for women of reproductive age reached nearly 100%. This high detection rate among women of reproductive age aligns with the implementation of targeted screening and testing aimed at preventing microcephaly and Congenital Zika Syndrome in Colombia and other affected regions^{9–11}.

After adjusting for surveillance bias towards women of reproductive age using information contained in the data, our findings indicate that females remained more susceptible to Zika symptomatic infection than males. This is consistent with previous studies, which reported higher attack rates among females^{3,10,21–27}, attributed not only to higher reporting rates but also to real biological effects¹⁰, as evidenced by seroprevalence data²⁴. Higher attack rates among women outside reproductive age groups further support this hypothesis²⁵. The potential for sexual transmission of Zika is likely another factor contributing to the high disease burden in women. One study identified an increased infection risk if Zika infection for sexually active women aged 15–65 years, but not for other age groups¹⁰. Another study suggested the possibility that women are more likely to exhibit symptomatic infections than men²⁶.

We found that ambient temperature was positively associated with Zika infection risk, consistent with most previous studies^{28–33}, though one study reported no such association³⁴. Temperature influences the ecological suitability of the vector, *Aedes aegypti*, and consequently, Zika transmission³⁰. Conversely, higher precipitation was associated with lower transmission risks, in line with two prior studies^{29,33}. However, conflicting evidence also exists: one study suggested rainfall might increase Zika risk²⁸, while another reported a non-linear relationship³⁴.

Departments with higher dengue incidence from 2011 to 2015 were associated with higher Zika infection risk. Several factors could have contributed to this association. First, dengue and Zika share the same mosquito vectors, implying that similar ecological and environmental conditions support their co-circulation; consequently, departments with greater vector abundance tend to have higher risks of both dengue and Zika. Another possible reason is the misdiagnosis of dengue case as Zika cases and vice versa as a result of their similarity in symptoms (e.g., fever, rash and myalgia) and antibody responses. The challenges in clinical and laboratory diagnosis may create an epidemiological picture where high dengue areas appear to also have high Zika incidence simply because of misdiagnosis³⁴. While misdiagnosis could be associated with under-reporting or over-reporting, its impact on the parameter estimation of our transmission model is likely limited, as long as the misdiagnosis level did not vary much across space-time or age-sex groups. Misdiagnosis could be relatively high during the early phase of the epidemic due to either technological barrier or lack of awareness; however, we assumed a linear growth in the

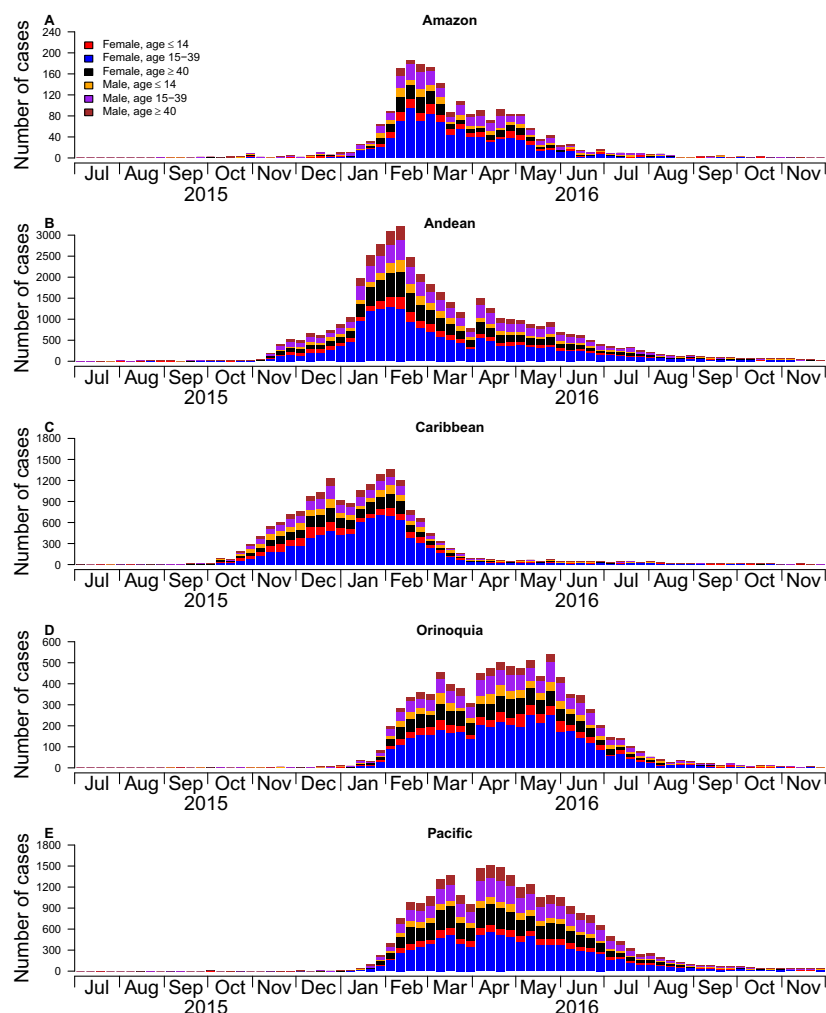


Fig. 3 | Temporal distribution of Zika cases across Colombian regions by gender and age group. This figure illustrates the monthly temporal trends in Zika virus cases stratified by gender and age groups across five regions of Colombia during the epidemic period (2015–2016). Each panel corresponds to a specific region: (A)

Amazon, (B) Andean, (C) Caribbean, (D) Orinoquia, and (E) Pacific. The y-axis represents the number of cases, while the x-axis spans the time from July 2015 to November 2016. Colored lines represent different demographic groups: females aged ≤ 14 years, 15–39 years, and ≥ 40 years, and males in the same age brackets.

reporting rates before November 30, 2015, which partially alleviates the possibility of under-reporting due to misclassification of Zika cases as dengue cases. Finally, similarity in antigenicity between the two viruses could introduce complex interference into population-level transmission dynamics of Zika. Prior results on the dengue-Zika relationship are mixed: one study reported no impact of dengue incidence on Zika⁹, while another suggested that dengue antibodies enhanced Zika infections³⁵. In addition, three studies found that prior dengue infection provided short-term protection against Zika^{36–38}, yet another study suggested minimal impact of pre-existing dengue immunity³⁹. A modeling study highlighted the complexity of the dengue-Zika relationship which might be influenced by their reproductive numbers⁴⁰.

In our study, only 4% of cases were laboratory-confirmed, and the majority were clinically diagnosed. The potential of misclassification of non-Zika cases as Zika cases or vice versa will have similar impacts as the potential misdiagnosis between dengue and Zika cases on our result. If the misclassification was uniform across space-time and across sex-age groups, it should have very limited effects on our estimates. If the misclassification was differential, e.g., if certain demographic groups are more likely to have non-Zika cases misclassified as Zika cases, it could bias the estimated predictor effects as well as reporting probabilities in unpredictable ways.

Geographic factors also played an important role in Zika transmission. Departments with lower forest coverages were associated with

higher infection risk, consistent with findings that cities exhibit greater risk than rural areas³¹. Altitude showed a non-linear relationship with Zika risk: transmission risk was the lowest in low-altitude regions but highest in medium-altitude regions, which was probably related to the role of altitude in ecological niches of the vector. This aligns with studies showing minimal Zika risk at high altitudes⁴¹, though some reported a negative association between altitude and infection risk⁴². Population density similarly exhibited a non-linear relationship. Low population density was associated with reduced infection risk, while medium density correlated with slightly increased risk. Our findings are largely consistent with previous studies reporting a positive relationship between Zika risk and population density^{42,43}, suggesting that densely populated urban areas experienced greater risks of Zika.

The model-corrected attack rates are ~10–20% higher (relative scale) than the reported values across various regions and departments. These findings are broadly consistent with those reported by Moore et al.¹³, who found that the attack rates were generally less than 1%. However, discrepancies in a few departments likely reflect differences in data sources and case definitions; while Moore et al. employed serological data that capture both symptomatic and asymptomatic infections, our analysis is based on suspected symptomatic cases, with only a fraction confirmed by laboratory tests.

Our approach categorized the population into six age-sex groups, enabling the estimation of three parameters related to age-specific

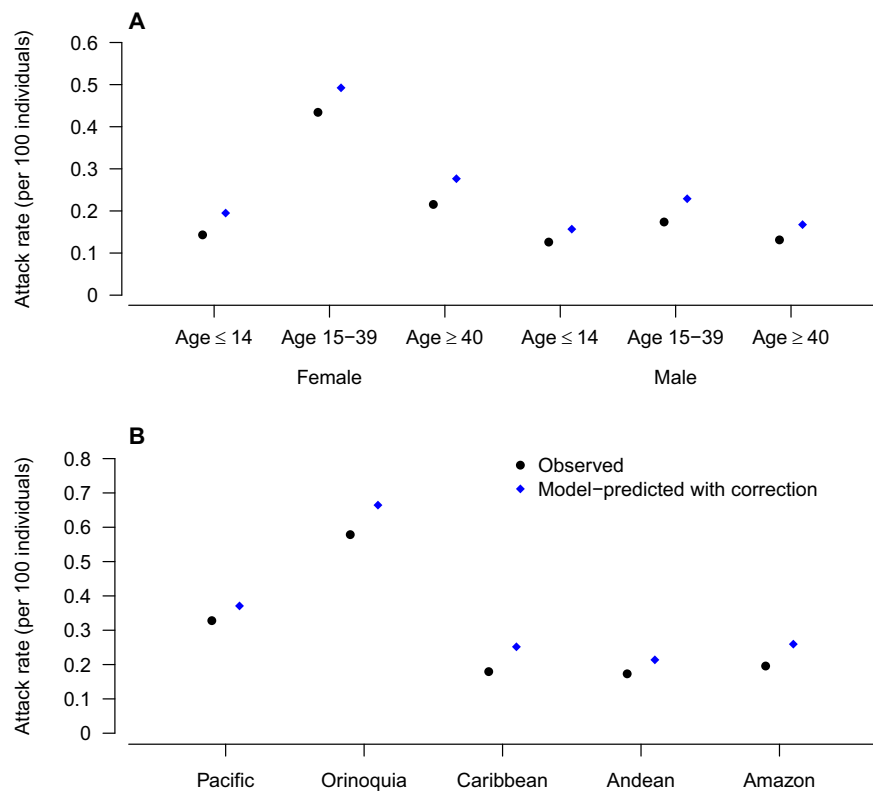


Fig. 4 | Estimated attack rates from the model. A The estimated attack rates for age-sex groups. **B** The estimate attack rates for regions. The black circle and blue diamond are the observed, and model-predicted with correction for reporting differences respectively.

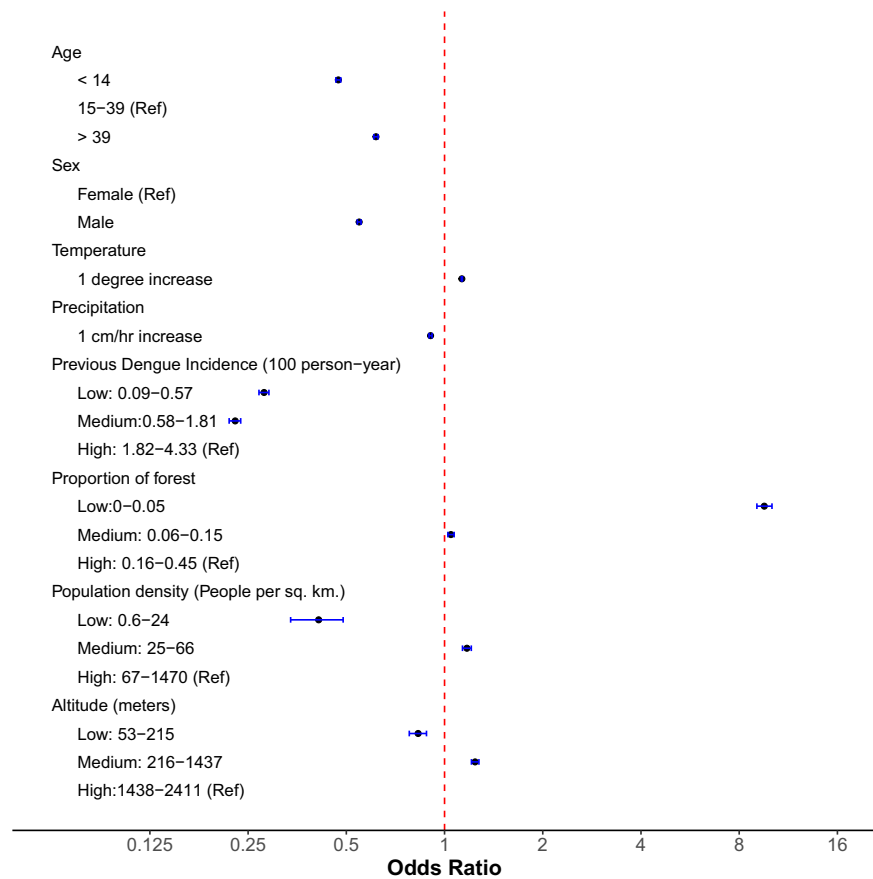


Fig. 5 | Estimates of covariate effects on transmissibility of Zika. The blue points and horizontal bars are the posterior median of odd ratios and 95% credible intervals.

susceptibility to symptomatic infection (corresponding to the three age groups: 0–14, 15–39, and 40+ years) and two parameters related to sex-specific susceptibility to symptomatic infection (male and female). To account for surveillance bias across age-sex groups, we included two additional parameters: one representing the reporting probability for females of reproductive age (15–39 years), and another for all other age-sex groups combined. A finer grouping of reporting probability will result in identifiability issues. Such identifiability issue could be alleviated with additional data, e.g., serosurvey studies in Colombia with age-sex grouping. Via a literature search, we found very limited number of such studies and most of these studies were targeting a special group of people or a special region^{44–46}, insufficient to inform a finer grouping or improve inference for our model.

Our study has limitations. First, we conducted the analysis at the department level, without accounting for finer spatial scales, which could improve result accuracy; yet the computational burden will increase exponentially as we model pairwise transmission among spatial units. Second, we lacked direct abundance data for the Zika vector, relying instead on proxy predictors, such as environmental factors. Finally, as an observational study, the possibility of unmeasured confounders remains.

In conclusion, our study introduced a statistical approach to correct surveillance biases caused by the clinical association of Zika infection with Congenital Zika Syndrome. We estimated that almost all cases among women of reproductive age were detected, compared to 86% for other age-sex groups. Our method can be generalized to other infectious diseases with similar surveillance bias, e.g., the hand, foot and mouth disease that causes neurological complications more often in young children than in older children or adults⁴⁷. These findings underscore the need for statistical and epidemiological methods adjusting for reporting biases to better understand and manage emerging infectious diseases and call for timely collection of relevant data, such as serosurvey data, that can further help correcting surveillance bias.

Methods

Data collection

The details of Colombia's surveillance system have been described previously²³. In brief, the Instituto Nacional de Salud (INS) operates a national public health surveillance system for notifiable conditions, including Zika. Case information is collected by healthcare centers and transmitted to the national health surveillance system, with an approximate reporting delay of 1.5 weeks. We downloaded an anonymised line-list of all suspected and laboratory-confirmed Zika, released by INS as an open-access dataset (<https://portalsivigila.ins.gov.co/>). Cases are primarily identified based on clinical symptoms and are subsequently confirmed by RT-PCR, following established protocols²³. However, only a small fraction (~4%) of reported cases are laboratory-confirmed, highlighting potential gaps in coverage. This low confirmation rate largely reflects two key factors. First is the narrow window during which Zika virus RNA is detectable in serum. If samples are collected too late after symptom onset, the virus may no longer be detectable. Second, resource and logistical constraints limit confirmatory testing to only a subset of suspected cases, with most diagnoses relying on clinical criteria.

For this study, the analysis period was from July 27, 2015, to November 21, 2016. All Zika cases reported to the surveillance system, whether laboratory-confirmed or not, were included in the analysis. Initially, Zika cases were defined as fever accompanied by at least one of the following symptoms: non-purulent conjunctivitis, headache, rash, pruritus, or arthralgia, with no known alternative cause. On December 24, 2015, the case definition was revised to include fever and rash, along with at least one of the following symptoms: non-purulent conjunctivitis, headache, pruritus, arthralgia, myalgia, or malaise. Pacheco

et al.²³ noted this change but did not perform a formal validation analysis to quantify its direct effect on reported case counts.

Several socio-environmental factors that are known to be correlated with Zika transmissions were collected and included in our analysis. Population density captures variations in human-vector contact rates, as denser populations tend to facilitate more interactions between people and mosquitoes^{42,43}. Department-level demographic statistics and population densities were collected from national statistics department of Colombia. Temperature is critical because it affects mosquito development, survival, and viral replication within the vector, thereby influencing transmission potential³⁰. Precipitation is similarly important since it creates and sustains breeding sites for Aedes mosquitoes, driving fluctuations in their abundance^{29,33}. Daily data on temperature and precipitation were extracted from the Environmental Information database in National Oceanic and Atmospheric Administration. Together, these factors provide essential context for understanding spatial and temporal variations in Zika transmission.

The statistical transmission model

Based on individual-level data, we created time series of weekly case numbers stratified by six age-sex groups (ages: 0–14, 15–39, 40+ years; sex: female, male) for 32 departments in Colombia.

Due to increased awareness of the association between Zika infection and microcephaly^{16,48,49}, cases among women of reproductive age (15–39 years) were likely oversampled than other demographic groups in the surveillance data. To address this, we developed a three-level hierarchical model to estimate transmission dynamics while adjusting for differential reporting probabilities among age-sex groups (see Supplementary Information). The model provided estimates of the true case counts and adjusted attack rate for each age-sex group.

The first (or top) level of the model, the reporting sub-model, defined the reporting probability as the likelihood of a symptomatic case being recorded in the surveillance system. Weekly case numbers for each age-sex group were modeled as latent variables representing the “true” case counts after correcting for differential reporting. These true counts were linked to observed case counts using a binomial distribution. The model assumed that surveillance capacity improved gradually, increasing from 10% to 100% over a 20-week period from July 14 to November 30, 2015, with full capacity reached by the latter date. Given the limited empirical information on the initial reporting capacity, we conducted sensitivity analyses assuming the initial reporting capacity was 5% and 20% of the final reporting capacity, in addition to our primary assumption of 10%. Before December 1, 2015, the reporting probability and its linear increase were assumed to be the same across all age-sex groups. From that date and onwards, the reporting probability for women aged 15–39 was allowed to be higher, reflecting increased public awareness of the link between Zika and microcephaly (modeled as an additional parameter). We conducted a sensitivity analysis that allows the reporting capacity for women of reproductive age to gradually scale up for another 2 or 4 weeks after December 1, 2015, instead of an instant jump.

The second level was a Poisson transmission model used to estimate Zika transmission dynamics. The infection probability for a given week in a department and age group was modeled as a function of case numbers in the same department, neighboring departments, and non-neighboring departments. Different transmission probabilities were assumed for within-department, neighbor-department, and non-neighbor-department interactions. Regional variation in transmission probabilities was also modeled across Colombia's five geographic regions: Pacific, Orinoquia, Caribbean, Andean, and Amazon. The model accounted for several covariates, including age group (0–14, 15–39, ≥40 years), sex, temperature, precipitation, cumulative dengue incidence (2010–2015), forest cover proportion, population density, and altitude. The serial interval was derived by combining distributions for the latent and infectious periods in both mosquitoes and humans (see

Table S2 and Supplementary Information for details). We also conducted a comparison analysis that reporting bias was ignored to explore the impact on the estimates of the effects of predictors.

Model Inference

To estimate unobserved “true” case numbers corrected for detection bias, we employed a Bayesian data augmentation framework. Model parameters were updated using the Metropolis-Hastings algorithm, and true weekly case numbers for each age-sex group in each department were jointly estimated. Statistical analyses were conducted using R version 4.0.5 (R Foundation for Statistical Computing, Vienna, Austria).

Model validation and adequacy

To evaluate goodness-of-fit of the model, a simulation study was conducted by comparing observed case numbers in surveillance data with those predicted by the model. For each simulation, a parameter vector was randomly drawn from the posterior distribution, and a dataset was simulated based on the transmission model and the selected parameter vector. Additionally, a separate simulation study was performed to validate the ability of the model to recover unknown parameters. Simulated epidemics were generated using parameter vectors set to the mean of the posterior distribution, and the algorithm was applied to estimate model parameters for each simulated epidemic. The results demonstrated the robustness of the model in estimating key parameters and accurately capturing the dynamics of Zika transmission.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All the data used in the analysis is available at https://github.com/timktsang/zika_surveillance_bias.

Code availability

All codes are available at https://github.com/timktsang/zika_surveillance_bias.

References

- Musso, D. & Gubler, D. J. Zika Virus. *Clin. Microbiol. Rev.* **29**, 487–524 (2016).
- World Health Organization. *Situation Report-Zika Virus, Microcephaly and Guillain-barré Syndrome*. Available from: <https://apps.who.int/iris/handle/10665/204961> (2016).
- Rojas, D. P. et al. The epidemiology and transmissibility of Zika virus in Girardot and San Andres island, Colombia, September 2015 to January 2016. *Euro Surveill.* <https://doi.org/10.2807/1560-7917.ES.2016.21.28.30283> (2016).
- World Health Organization. *Situation Report-zika Virus, Microcephaly, Guillain-barré Syndrome*. Available from: <https://www.who.int/publications/m/item/zika-situation-report> (2017).
- European Centre for Disease Prevention and Control (ECDC). *Rapid Risk Assessment: Zika Virus Infection Outbreak, French Polynesia*. (Stockholm: ECDC, 2014).
- World Health Organization. *Zika virus*. Available from: <https://www.who.int/news-room/fact-sheets/detail/zika-virus> (2022).
- Giron, S. et al. Vector-borne transmission of Zika virus in Europe, southern France, August 2019. *Euro Surveill.* <https://doi.org/10.2807/1560-7917.ES.2019.24.45.1900655> (2019).
- Asokan, D., Bomm, S. P. R., Mall, A. & Pardeshi, G. Geographic distribution and temporal trends of Zika virus outbreaks in India (2016–2023): insights from the integrated disease surveillance programme data. *Cureus* **16**, e67068 (2024).
- Bowman, L. R., Rocklöv, J., Kroeger, A., Oliaro, P. & Skewes, R. A comparison of Zika and dengue outbreaks using national surveillance data in the Dominican Republic. *PLOS Neglected Tropical Dis.* **12**, e0006876 (2018).
- Coelho, F. C. et al. 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.* **51**, 128–132 (2016).
- Ruchusatsawat, K. et al. Long-term circulation of Zika virus in Thailand: an observational study. *Lancet Infect. Dis.* **19**, 439–446 (2019).
- Charniga, K. et al. Estimating Zika virus attack rates and risk of Zika virus-associated neurological complications in Colombian capital cities with a Bayesian model. *R. Soc. Open Sci.* **9**, 220491 (2022).
- Moore, S. M. et al. Leveraging multiple data types to estimate the size of the Zika epidemic in the Americas. *PLOS Neglected Tropical Dis.* **14**, e0008640 (2020).
- Oidtmann, R. J., España, G. & Perkins, T. A. Co-circulation and misdiagnosis led to underestimation of the 2015–2017 Zika epidemic in the Americas. *PLoS Negl. Trop. Dis.* **15**, e0009208 (2021).
- Quandelacy, T. M. et al. Estimating incidence of infection from diverse data sources: Zika virus in Puerto Rico, 2016. *PLOS Computational Biol.* **17**, e1008812 (2021).
- Microcephaly Epidemic Research Group. Microcephaly in infants, Pernambuco State, Brazil, 2015. *Emerg. Infect. Dis.* **22**, 1090–1093 (2016).
- Zhang, Q. et al. Spread of Zika virus in the Americas. *Proc. Natl Acad. Sci. USA* **114**, E4334–E4343 (2017).
- Cauchemez, S. et al. Association between Zika virus and microcephaly in French Polynesia, 2013–15: a retrospective study. *Lancet* **387**, 2125–2132 (2016).
- Brady, O. J. et al. The association between Zika virus infection and microcephaly in Brazil 2015–2017: an observational analysis of over 4 million births. *PLOS Med.* **16**, e1002755 (2019).
- de Oliveira, W. K. et al. Infection-related microcephaly after the 2015 and 2016 Zika virus outbreaks in Brazil: a surveillance-based analysis. *Lancet* **390**, 861–870 (2017).
- Tolosa, N. et al. Zika virus disease in children in Colombia, August 2015 to May 2016. *Paediatr. Perinat. Epidemiol.* **31**, 537–545 (2017).
- Maxian, O., Neufeld, A., Talis, E. J., Childs, L. M. & Blackwood, J. C. Zika virus dynamics: when does sexual transmission matter? *Epidemics* **21**, 48–55 (2017).
- Pacheco, O. et al. Zika virus disease in Colombia - preliminary report. *N. Engl. J. Med.* **383**, e44 (2020).
- Zambrana, J. V. et al. Seroprevalence, risk factor, and spatial analyses of Zika virus infection after the 2016 epidemic in Managua, Nicaragua. *Proc. Natl Acad. Sci.* **115**, 9294–9299 (2018).
- Lozier, M. et al. Incidence of Zika virus disease by age and sex — Puerto Rico, November 1, 2015–October 20, 2016. *Morbidity Mortal. Wkly. Rep.* **65**, 1219–1223 (2016).
- Lozier, M. J. et al. Differences in prevalence of symptomatic Zika virus infection, by age and sex—Puerto Rico, 2016. *J. Infect. Dis.* **217**, 1678–1689 (2018).
- Rosenberg, E. S. et al. Prevalence and incidence of Zika virus infection among household contacts of patients with Zika virus disease, Puerto Rico, 2016–2017. *J. Infect. Dis.* **220**, 932–939 (2019).
- Chien, L.-C., Sy, F. & Pérez, A. Identifying high risk areas of Zika virus infection by meteorological factors in Colombia. *BMC Infect. Dis.* **19**, 888 (2019).
- Sharp, T. M. et al. Epidemiologic and spatiotemporal trends of Zika Virus disease during the 2016 epidemic in Puerto Rico. *PLOS Neglected Tropical Dis.* **14**, e0008532 (2020).
- Mordecai, E. A. et al. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. *PLOS Neglected Tropical Dis.* **11**, e0005568 (2017).
- Tesla, B. et al. Temperature drives Zika virus transmission: evidence from empirical and mathematical models. *Proc. R. Soc. B: Biol. Sci.* **285**, 20180795 (2018).

32. Ngonghala, C. N. et al. Effects of changes in temperature on Zika dynamics and control. *J. R. Soc. Interface* **18**, 20210165 (2021).
 33. Rees, E. E., Petukhova, T., Mascarenhas, M., Pelcat, Y. & Ogden, N. H. Environmental and social determinants of population vulnerability to Zika virus emergence at the local scale. *Parasites Vectors* **11**, 290 (2018).
 34. Riou, J., Poletto, C. & Boëlle, P.-Y. A comparative analysis of Chikungunya and Zika transmission. *Epidemics* **19**, 43–52 (2017).
 35. Castanha, P. M. S. et al. Dengue virus-specific antibodies enhance Brazilian Zika virus infection. *J. Infect. Dis.* **215**, 781–785 (2017).
 36. Collins, M. H. et al. Lack of durable cross-neutralizing antibodies against Zika virus from dengue virus infection. *Emerg. Infect. Dis.* **23**, 773–781 (2017).
 37. Rodriguez-Barraquer, I. et al. Impact of preexisting dengue immunity on Zika virus emergence in a dengue endemic region. *Science* **363**, 607–610 (2019).
 38. Gordon, A. et al. Prior dengue virus infection and risk of Zika: a pediatric cohort in Nicaragua. *PLOS Med.* **16**, e1002726 (2019).
 39. Michlmayr, D. et al. Comprehensive immunoprofiling of pediatric Zika reveals key role for monocytes in the acute phase and no effect of prior dengue virus infection. *Cell Rep.* **31**, 107569 (2020).
 40. Borchering, R. K. et al. Impacts of Zika emergence in Latin America on endemic dengue transmission. *Nat. Commun.* **10**, 5730 (2019).
 41. Cetron, M. Revision to CDC's Zika travel notices: minimal likelihood for mosquito-borne Zika virus transmission at elevations above 2000 meters. *MMWR Morb. Mortal. Wkly Rep.* **65**, 267–268 (2016).
 42. Weinstein, J. S., Leslie, T. F. & von Fricken, M. E. Spatial associations between land use and infectious disease: Zika virus in Colombia. *Int. J. Environ. Res. Public Health* **17**, 1127 (2020).
 43. Costa, S. et al. Spatial analysis of probable cases of dengue fever, chikungunya fever and Zika virus infections in Maranhao State, Brazil. *Rev. Inst. Med Trop. Sao Paulo* **60**, e62 (2018).
 44. Cardona-Ospina, J. A. et al. Susceptibility to endemic Aedes-borne viruses among pregnant women in Risaralda, Colombia. *Int J. Infect. Dis.* **122**, 832–840 (2022).
 45. Marban-Castro, E. et al. High seroprevalence of antibodies against arboviruses among pregnant women in rural Caribbean Colombia in the context of the Zika virus epidemic. *Antibodies (Basel)* <https://doi.org/10.3390/antib9040056> (2020).
 46. Bayona-Pacheco, B. et al. Seroprevalence of Zika virus among blood donors before the epidemic in Barranquilla, Colombia, 2015–2016. *Acad. Bras. Cienc.* **91**, e20180860 (2019).
 47. Tang, X., Yang, Y., Yu, H. J., Liao, Q. H. & Bliznyuk, N. A spatio-temporal modeling framework for surveillance data of multiple infectious pathogens with small laboratory validation sets. *J. Am. Stat. Assoc.* **114**, 1561–1573 (2019).
 48. Jaenisch, T. et al. Risk of microcephaly after Zika virus infection in Brazil, 2015 to 2016. *Bull. World Health Organ* **95**, 191–198 (2017).
 49. Flamand, C. et al. The proportion of asymptomatic infections and spectrum of disease among pregnant women infected by Zika virus: systematic monitoring in French Guiana, 2016. *Euro Surveill.* <https://doi.org/10.2807/1560-7917.ES.2017.22.44.17-00102> (2017).
- enhanced our understanding of the Zika outbreak in Colombia but also contributed significantly to the public health response during this critical period. This research was supported by NIH grant U54-GM11274 (T.K.T., Y.Y., I.M.L., and M.E.H.), Fulbright Colciencias Scholarship (DPR), and the Research Assessment Exercise (RAE) 2026 Development Fund from The University of Hong Kong (T.K.T.).

Author contributions

Y.Y. and T.K.T. designed the research. D.P.R. collected surveillance data. T.K.T. analyzed the data. F.X, Y.X., and X.Z. conducted the geographical analysis. T.K.T. wrote the draft manuscript. Y.Y., T.K.T., I.M.L., M.E.H., D.P.R., F.X, and X.Z. finalized the paper.

Competing interests

All authors report no potential conflicts of interest.

Ethics and inclusion statement

This study used publicly available, de-identified surveillance data from the Instituto Nacional de Salud (INS) and regional health departments in Colombia. Since the data were de-identified and publicly available, and no interaction with human subjects occurred, ethical approval was not required for this analysis. Direct collaboration was not established with local researchers; their critical role in data collection is acknowledged in the Acknowledgements. The study adhered to relevant local ethical standards and incorporated local and regional research in its citations, underscoring our commitment to ethical and equitable global research collaborations.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-59655-9>.

Correspondence and requests for materials should be addressed to Tim K. Tsang or Yang Yang.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

Acknowledgements

We gratefully acknowledge the indispensable contributions of the data collection team at the Instituto Nacional de Salud (INS) and the Secretaries of Health from the various departments, districts, and municipalities in Colombia. Their rigorous surveillance efforts, as exemplified in the comprehensive Zika virus disease data detailed in Pacheco et al. (NEJM, 2016), provided the critical dataset underpinning our analysis. The dedication of these teams to timely and accurate data collection, coupled with the INS arboviral reference laboratory's extensive RT-PCR testing, has been vital for our ability to assess transmission dynamics and adjust for differential reporting biases. Their collective work not only