Epidemics of chikungunya, Zika, and COVID-19 reveal bias in case-based mapping

Fausto Andres Bustos Carrillo[1,2]; Brenda Lopez Mercado[3]; Jairo Carey Monterrey[3]; Damaris Collado[3]; Saira Saborio[3]; Tatiana Miranda[3]; Carlos Barilla[3]; Sergio Ojeda[3]; Nery Sanchez[3]; Miguel Plazaola[3]; Harold Suazo Laguna[3]; Douglas Elizondo[3]; Sonia Arguello[3]; Anna M. Gajewski[3]; Hannah E. Maier[4]; Krista Latta[4]; Bradley Carlson[4]; Josefina Coloma[1]; Leah Katzelnick[1]; Hugh Sturrock[5,6]; Angel Balmaseda[3,7]; Guillermina Kuan[3,8]; Aubree Gordon[4#]*; Eva Harris[1#]*

Affiliations:

[1] Division of Infectious Diseases and Vaccinology, School of Public Health, University of California, Berkeley, Berkeley, California, 94720-3370, USA

[2] Division of Epidemiology and Biostatistics, School of Public Health, University of California, Berkeley, Berkeley, California, 94720-7360, USA

[3] Sustainable Sciences Institute, Managua, 14006, Nicaragua

[4] Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, 48109, USA

[5] Department of Epidemiology and Biostatistics, School of Medicine, University of California, San Francisco, San Francisco, California, 94158, USA

[6] Locational, Poole, Dorset, BH16 6FA, United Kingdom

[7] National Virology Laboratory, Centro Nacional de Diagnóstico y Referencia, Ministry of Health, Managua, 16064 Nicaragua

[8] Centro de Salud Sócrates Flores Vivas, Ministry of Health, Managua, 12005, Nicaragua

# These authors contributed equally

***Co-corresponding authors:**

Eva Harris, Division of Infectious Diseases and Vaccinology, School of Public Health, University of California, Berkeley, 185 Li Ka Shing Center, 1951 Oxford Street, Berkeley, CA 94720-3370; Tel. 1-510-642-4845; eharris@berkeley.edu

Aubree Gordon, Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, 5622 SPH I, 1415 Washington Heights, Ann Arbor, MI 48109-2029; Tel. 1-510-409-5495; gordonal@umich.edu

ABSTRACT

Explosive epidemics of chikungunya, Zika, and COVID-19 have recently occurred worldwide, all of which featured large proportions of subclinical infections. Spatial studies of infectious disease epidemics typically use symptomatic infections (cases) to estimate incidence rates (cases/total population), often misinterpreting them as infection risks (infections/total population) or disease risks (cases/infected population). We examined these three measures in a pediatric cohort (N≈3,000) over two chikungunya epidemics and one Zika epidemic and in a household cohort (N=1,793) over one COVID-19 epidemic in Nicaragua. Across different analyses and all epidemics, case incidence rates considerably underestimated both risk-based measures. Spatial infection risk differed from spatial disease risk, and typical case-only approaches precluded a full understanding of the spatial seroprevalence patterns. For epidemics of pathogens that cause many subclinical infections, relying on case-only datasets and misinterpreting incidence rates, as is common, results in substantial bias, a general finding applicable to many pathogens of high human concern.

INTRODUCTION

In recent years, successive epidemics of chikungunya virus (CHIKV),[1] Zika virus (ZIKV),[1] and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[2] caused millions of symptomatic infections (cases) and subclinical (clinically inapparent) infections across the Americas. Chikungunya manifests with high fever and arthralgia, which can persist for months or years.[1] Zika typically presents non-specifically during childhood and as a mild, dengue-like illness during adolescence and adulthood,[3] although ZIKV infection may infrequently trigger Guillain-Barré Syndrome.[1] ZIKV infection during pregnancy can cause neurodevelopmental complications, including microcephaly, in infants.[1] While coronavirus disease 2019 (COVID-19) mostly presents as a mild illness, severe cases can experience respiratory distress, multiorgan failure, and death.[2]

CHIKV and ZIKV are vectored by *Aedes* mosquitoes,[1] whereas respiratory droplets primarily spread SARS-CoV-2.[2] A large proportion of infections caused by ZIKV, the Asian lineage of CHIKV, and SARS-CoV-2 are subclinical, yet they contribute substantially to transmission.[4–6] Standard spatial studies of these viruses[7–11] and other pathogens nevertheless focus exclusively on cases, neglecting subclinical infections and thereby precluding a full understanding of the spatial contours of immunity. Further, such studies are unable to distinguish between uninfected persons and those with subclinical infections, as only cases are evaluated. These studies often characterize epidemics with the case-only incidence rate (cases/total population), also called the incidence proportion.

Other measures besides the incidence rate need to be examined to assess the full extent of an epidemic with respect to infection and disease. In epidemiological terms, risk is the probability of a susceptible individual experiencing an outcome. For an immunologically naïve population, all persons are at risk for an initial infection, and only infected individuals are at risk for experiencing illness, as uninfected persons cannot experience a given infectious disease. Consequently, measuring infection status is necessary to estimate infection risk (infections/total population) and disease risk (cases/infected population) for pathogens that

cause substantial subclinical infections. For a naïve population that experiences an epidemic, infection risk is equivalent to seroprevalence. The incidence rate is commonly misinterpreted in spatial health research as the risk of infection[7–9] or disease.[10,11] Below, we provide three forms of the same relationship between these three measures:

Epidemiological: Infection risk × Disease risk = Incidence rate (Eq. 1)

Algebraic: $$\frac{\text{Infections}}{\text{Total population}} \times \frac{\text{Cases}}{\text{Infections}} = \frac{\text{Cases}}{\text{Total population}}$$

Statistical: $$P(\text{Infection}) \times P(\text{Disease} \mid \text{Infection}) = P(\text{Disease and Infection})$$

A recent comprehensive review of spatial research on COVID-19 epidemics highlighted the lack of fine-scaled spatial data and preponderance of ecological study designs,[12] which aggregate data at arbitrary spatial districts and thereby induce bias.[13] The under-representation of study populations outside of China, Brazil, and the US was also identified as a critical research gap. Many of these issues also apply to spatial research on epidemics of chikungunya and Zika. Our study site in Managua, Nicaragua (Fig. S1), recently experienced epidemics of chikungunya, Zika, and COVID-19, giving us the unique opportunity to spatially assess infection and disease outcomes among immunologically naïve participants who subsequently lived through four epidemics of three distinct viruses. A joint analysis was undertaken to identify commonalities that generalize across epidemics of different viruses and routes of transmission.

We assessed children 2-14 years old in the Nicaraguan Pediatric Dengue Cohort Study (PDCS) who experienced two chikungunya epidemics (2014, 2015) and one Zika epidemic (2016).[3,14–16] We also evaluated household members 0-87 years old from our Household Influenza Cohort Study (HICS) who experienced the first wave (2020) of the COVID-19 epidemic. We use epidemic-specific datasets from both prospective cohort studies to explore the fine-scale spatial consequences of using standard, incidence-based approaches to characterize these epidemics versus a more comprehensive approach that also measures infection and disease risks.

RESULTS

Participant characteristics

We refer to the first chikungunya epidemic as *ChikE1*, the second as *ChikE2*, the Zika epidemic as *ZikaE,* and the COVID-19 epidemic as *CovidE*. Our study assessed infection and disease outcomes for 4,884 distinct individuals, including 3,693 unique PDCS participants across ChikE1, ChikE2, and ZikaE. Approximately 3,000 PDCS participants were analyzed for each chikungunya and Zika epidemic (Table 1). Of the 1,793 HICS participants, 602 children were also enrolled in the PDCS, and 1,192 mostly adult participants were only enrolled in the HICS. From 2014-2016, the PDCS population experienced ChikE1, ChikE2, and ZikaE during Managua's rainy period of June to November (Fig. S2); these months are generally marked by an abundance of mosquitoes in the study area. Unlike the mosquito-borne epidemics, CovidE peaked from March to August in 2020.

In the PDCS population, the distribution of age and sex was constant across ChikE1, ChikE2, and ZikaE (Table S1). While all PDCS participants are pediatric, the age structure of the HICS is representative of Managua's general population. Approximately 50% of PDCS participants were female in each epidemic. In the HICS, there was an over-enrollment of adult females relative to adult males.

Infection, disease, and case-based incidence measures

We first examined summary statistics of the four epidemics. In the PDCS, ChikE1 exhibited the lowest infection risk (6.4%) while ZikaE had the highest infection risk (47.1%) (Table 1). The first wave of the COVID-19 epidemic in the HICS exhibited the highest infection risk of all four epidemics (57.5%). However, CovidE featured the lowest disease risk (28.9%) as it had the largest proportion of subclinical infections, whereas ChikE2 had the highest disease risk (58.7%). Despite substantial differences in infection and disease risks across ChikE2, ZikaE, and CovidE, all three epidemics had similar case-based incidence rates. Moreover, the incidence rate underestimated both infection and disease risks across epidemics, often considerably, since it has the smallest numerator and the largest denominator (Tables 1, S2).

118 **Table 1.** Summary and descriptive statistics of infection and disease outcomes across four epidemics in the

119 PDCS and HICS.

	First chikungunya epidemic (ChikE1)	Second chikungunya epidemic (ChikE2)	Zika epidemic (ZikaE)	COVID-19 epidemic (CovidE)
Cohort	PDCS	PDCS	PDCS	HICS
Participant age range	2-14	2-14	2-14	0-92
Epidemic period	9/2014-2/2015	7/2015-2/2016	1/2016-1/2017	3/2020-10/2020
Primary transmission pathway	*Aedes* mosquitoes	*Aedes* mosquitoes	*Aedes* mosquitoes	Respiratory droplets
Number at risk of infection	3,124	2,864	3,017	1,793
Number of infections	199	710	1,416	1,039
Number of cases	90	416	494	306
Risk of infection (95% CI)[1]	6.4% (5.5%, 7.4%)	24.8% (23.2%, 26.6%)	47.1% (45.1%, 49.1%)	57.5% (54.1%, 60.9%)
Risk of disease (95% CI)[1]	45.6% (38.6%, 52.6%)	58.7% (54.9%, 62.4%)	35.4% (32.8%, 38.1%)	28.9% (25.5%, 32.5%)
Incidence rate (95% CI)[1]	2.9% (2.3%, 3.6%)	14.5% (13.2%, 16.0%)	16.6% (15.2%, 18.1%)	17.1% (14.8%, 19.6%)
ANOVA-based ICC for intra-household correlation of infection risk (95% CI)[2]	0.22 (0.17, 0.28)	0.21 (0.16, 0.27)	0.22 (0.17, 0.27)	0.30 (0.25, 0.35)
ANOVA-based ICC for intra-household correlation of disease risk (95% CI)[2]	0.14 (0.00, 0.51)	0.26 (0.09, 0.42)	0.28 (0.18, 0.37)	0.21 (0.15, 0.28)

120 [1]GEE model estimates.

121 [2]Table S7 contains additional information.

122 Abbreviations: ANOVA, analysis of variable; CI, confidence interval; GEE, generalized estimating
123 equations; HICS, Household Influenza Cohort Study; ICC, intracluster correlation coefficient; PDCS,
124 Pediatric Dengue Cohort Study

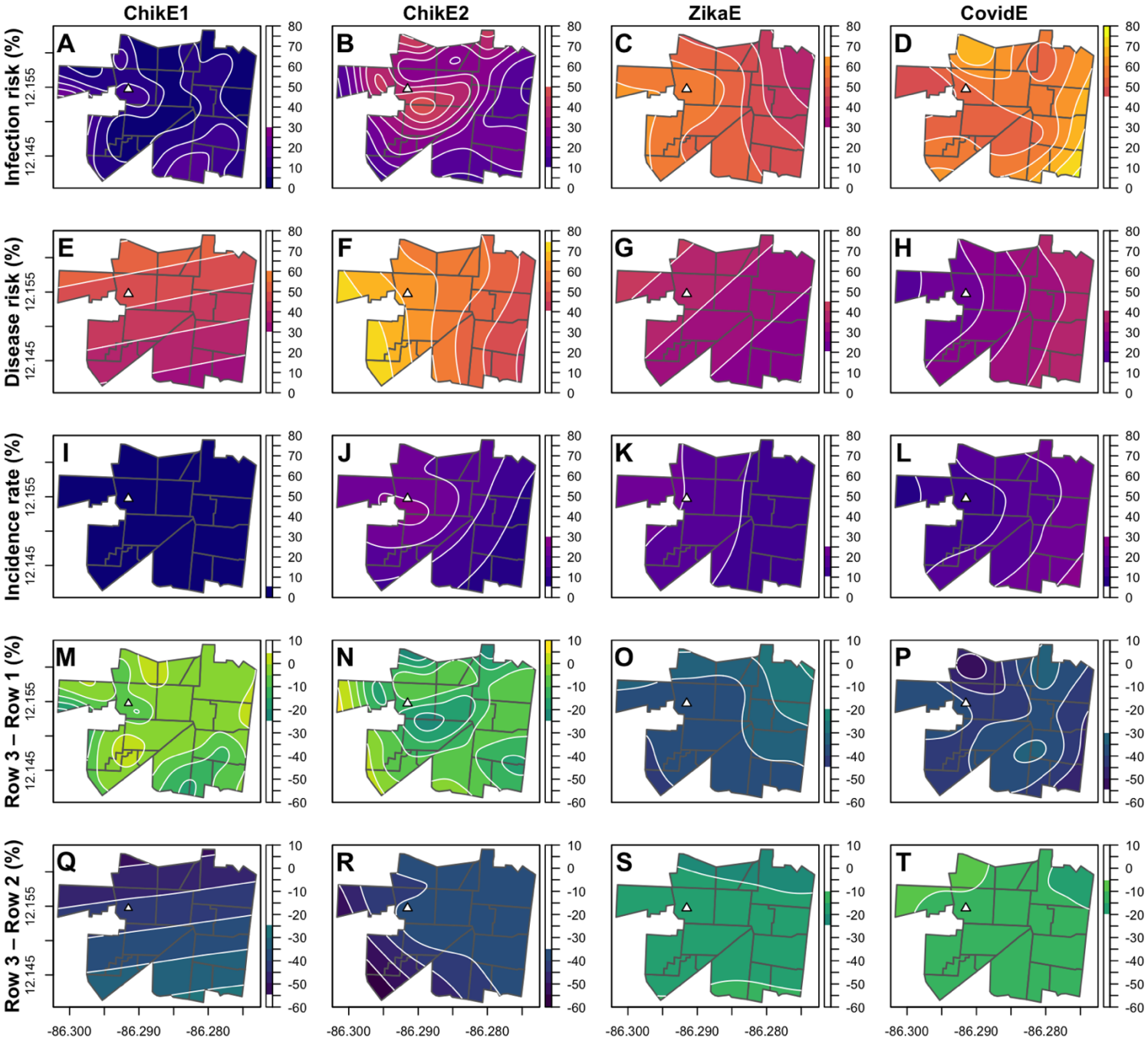
125 We then assessed the effect of sex and age, key demographic variables, on summary statistics. Sex
126 differences for infection and disease risks, even when statistically significant, tended to be small, as when
127 females had an infection risk 6% higher than males during ZikaE (Table S1, Fig. S3-S4). Similarly,
128 accounting for the over-enrollment of adult females in the HICS had little effect (~1%) on overall estimates
129 (Fig. S5-S6). However, there were surprisingly strong age trends for all epidemics. Infection risk increased
130 linearly across pediatric age for all epidemics, especially during ChikE2 and ZikaE (Fig. S7). In contrast,
131 SARS-CoV-2 infection risk was >50% across most ages and stable at ~62% throughout adulthood.

132 Disease risk increased with age during all epidemics except ZikaE (Fig. S8), reflecting the established cross-
133 protective effect of anti-DENV antibodies on Zika occurrence (Fig. S9).[17,18] Unlike the relatively stability
134 of infection risk across age during CovidE, disease risk for COVID-19 increased dramatically throughout
135 childhood and early adulthood. COVID-19 incidence was low across age, particularly during childhood
136 (Fig. S10). That the age trend for COVID-19 incidence reflected age-based changes in disease risk much
137 more than infection risk was an insight not apparent from case data alone.

138 **Spatial mapping of infection, disease, and incidence**

139 Next, we mapped the infection risk, disease risk, and case-only incidence rate across our study area to
140 characterize spatial trends. For all epidemics, infection risk varied at small spatial scales (Fig. 1A-D),
141 suggesting the local environment was an important determinant of infection risk. During ChikE1, ChikE2,
142 and ZikaE, infection risk was elevated in western neighborhoods of the study area adjacent to a large
143 cemetery known to be heavily infested with *Aedes* mosquitoes. Only after adjusting for distance to the
144 cemetery did the spatial patterns of infection risk for these mosquito-vectored epidemics change appreciably
145 (Figs. S11-S16). Conversely, SARS-CoV-2 infection risk was high in eastern neighborhoods that contain
146 large parks and tourist attractions (Fig. 1D, S17-S18), implying that SARS-CoV-2 infections were
147 differently spatially mediated than CHIKV and ZIKV.

Figure 1



Maps of the infection risk, disease risk, case-based incidence rate, and bias. The infection risk (**A-D**), disease risk (**E-H**), and incidence rate (**I-L**) across four epidemics are shown in one color palette. The difference between infection risk and the incidence rate (bias induced by interpreting the incidence rate as the infection risk) (**M-P**) and the corresponding bias for the disease risk (**Q-T**) are shown in another color palette. Maps were generated from generalized additive mixed models. A white triangle indicates the study health center. Neighborhoods are outlined in gray.

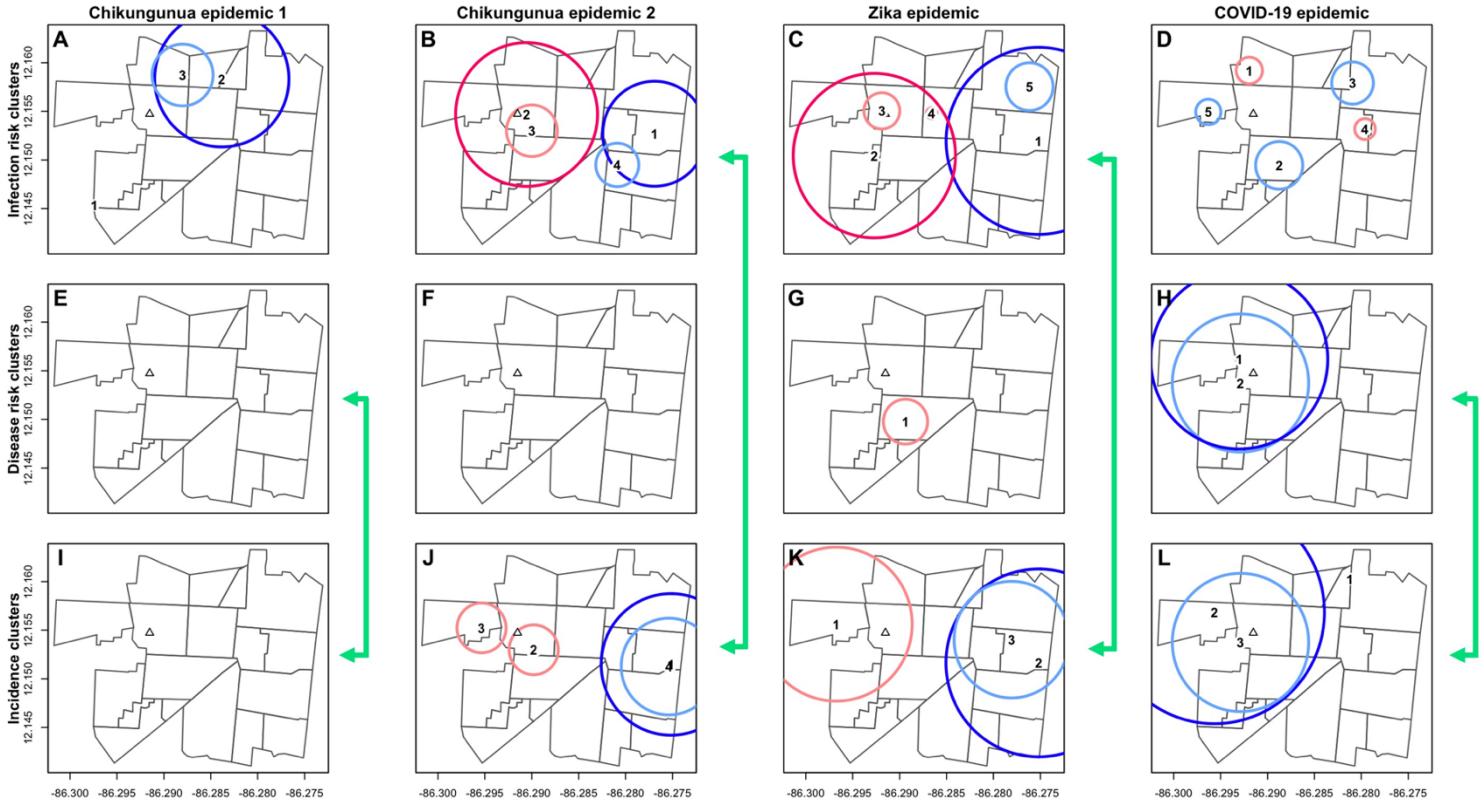
155 Across all epidemics, disease risk also varied at small spatial scales (Fig. 1E-H). After adjusting for
156 variables plausibly related to disease occurrence (Figs S19-26), spatial patterns of disease risk remained
157 non-uniform and distinct from spatial patterns of infection risk, demonstrating that areas of high infection
158 risk did not always have commensurate levels of disease risk.

159 As the case-based incidence rate is the product of two risks (Eq. 1) with different spatial patterns (Fig. 1A-
160 H), incidence rate maps (Fig. 1I-L) – the most common maps of health outcomes found in the literature –
161 underestimated risk-based maps and did not recapitulate spatial patterns of disease or infection risk. Large
162 and spatially non-uniform differences were observed between incidence-based and risk-based maps across
163 all epidemics (Fig. 1M-T), demonstrating that high levels of spatially heterogenous bias often occur when
164 risk is inferred from case-only incidence maps.

165 **Cluster detection**

166 In the context of epidemics, clusters of high disease incidence are often treated as areas with elevated
167 infection risk, and public resources are often directed to interventions in such areas. We used standard
168 hierarchical and the newer Gini cluster detection methods to identify significant clusters of infection risk,
169 disease risk, and case incidence (Fig. 2, Table S3). For each epidemic, we identified clusters of high and
170 low infection risk. Clusters of elevated infection risk for the larger mosquito-borne epidemics, ChikE2 and
171 ZikaE, encompassed the cemetery and western study neighborhoods adjacent to it. Large clusters of
172 diminished infection risk were observed in ChikE1, ChikE2, and ZikaE. These clusters highlighted areas
173 with excess uninfected persons who remain susceptible to future infection and thus areas where infection-
174 reducing interventions will be most needed. Such clusters are only identifiable after ascertaining the
175 infection status of samples representative of the total population, not just the cases. CovidE featured small
176 clusters of both elevated and diminished infection risk. In general, clusters of infection risk were often in
177 different locations and of different sizes than clusters of disease risk, demonstrating that infection and
178 disease risk cluster differently in space. Indeed, we detected no disease risk clusters during the chikungunya
179 epidemics despite finding large clusters of infection risk.

Figure 2



Cluster detection analyses of the infection risk, disease risk, and case-based incidence rate. Clusters of infection risk (**A-D**), disease risk (**E-H**), and the incidence rate (**I-L**) across four epidemics are shown. Panels depict the results of Kulldorf's spatial scan statistic conducted in SaTScan. Hotspots are shown in pink; coldspots are shown in blue. Hierarchical clusters are shown in dark colors; Gini clusters are shown in light colors. Cluster centers are numerically labeled. Table S3 contains additional information for this analysis. Arrows show the kind of risk clusters that incidence clusters resemble. A white triangle indicates the study health center. Neighborhoods are outlined in gray.

Surprisingly, standard incidence clusters – which identify areas of elevated or diminished case counts among the overall population – sometimes missed risk-based hierarchical and Gini clusters, even when risk-based clusters were quite large (Fig. 2). Moreover, incidence clusters failed to display a general pattern across epidemics, variably identifying areas of either infection or disease risk independent of epidemic size, transmission pathway, or causative virus. Thus, the type of risk captured by incidence clusters varied from

195 epidemic to epidemic in an unpredictable manner. This observation would not be apparent without jointly
196 analyzing multiple epidemics.
197
198 **Geostatistical models**
199 We next conducted geostatistical multivariable modeling to identify factors associated with infection and
200 disease outcomes (Tables S4-S6). Surprisingly, analyses that did and did not account for spatial and
201 household clustering yielded very similar results for all epidemics, suggesting that participants' infection
202 and disease outcomes were weakly correlated within homes and over large distances. Directly measuring
203 intra-household clustering of infection and disease outcomes with three statistical approaches likewise
204 identified weak correlations across epidemics (Tables 1, S7). We also observed that infection risk did not
205 scale with household size (Fig. S27). Together, the data demonstrated that our participants' infection
206 outcomes were less clustered by household than is typically assumed. Explicitly plotting the correlation of
207 infection outcomes across space (Fig. S28) showed that they were spatially correlated across short distances
208 (<200m) for all epidemics, reinforcing earlier findings (Fig. 1) that infection risk was moderated by the
209 local spatial environment. As before, these insights would be impossible without infection data.

210 Geostatistical modeling (Tables S5-6) revealed that distance to the cemetery was significantly associated
211 with ZIKV infection such that the odds of ZIKV infection among participants living 1 km from the cemetery
212 were 0.63 (95% CI: 0.55, 0.73) times that of participants living next to the cemetery, conditional on age,
213 sex, and indoor water availability; a similar 1-km odds ratio was observed during ChikE2. In all
214 geostatistical models, age was significantly associated with increasing infection risk, as in bivariate analyses
215 (Fig. S7).

216 Using geostatistical models, we did not identify variables that were consistently related to disease risk
217 across epidemics (Table S6). Rather, model results were epidemic-specific, as when prior DENV infection
218 was negatively associated with disease risk during ZikaE after accounting for age and sex. Age was
219 positively associated with disease risk during ChikE1 and CovidE after accounting for sex. We did not

220 conduct multivariable modeling of case incidence, as such models erroneously assume uninfected persons
221 are at risk for disease, biasing results (appendix page 18).

222 **Spatiotemporal modeling of epidemic dynamics**

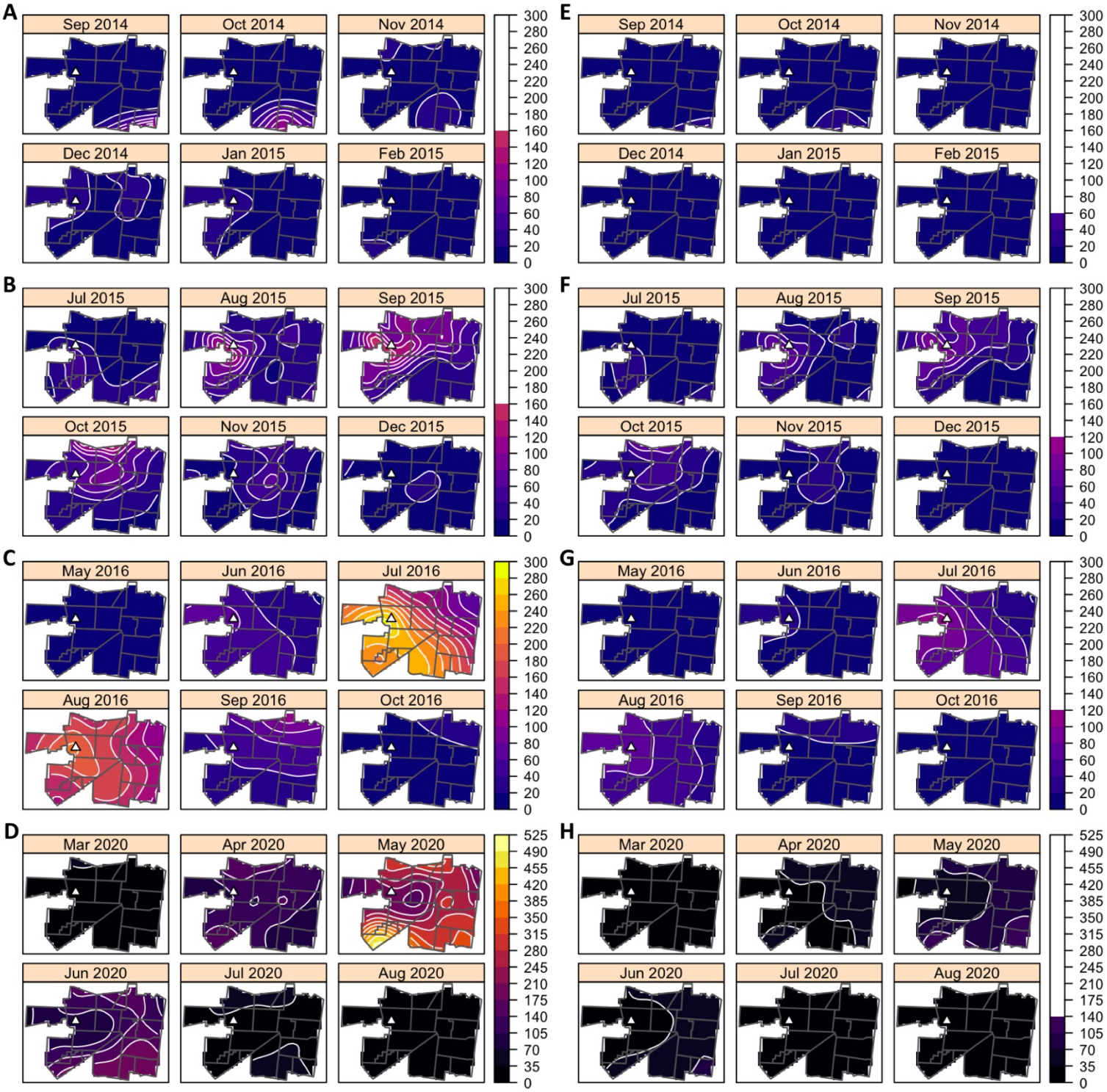
223 Spatiotemporal analyses depict epidemic progression across time and space. By harnessing Eq. 1, we
224 estimated the spatiotemporal dynamics of infection risk (Fig. 3A-D, S29-S32), which were substantially
225 underestimated by the less nuanced but standard spatiotemporal dynamics of case incidence (Fig. 3E-H,
226 S33-S36). ChikE1 was primarily characterized by moderate infection risk in southeastern neighborhoods
227 during September-October 2014, with elevated infection risk in western, cemetery-adjacent neighborhoods
228 from December 2014 to January 2015. ChikE1 subsided shortly afterward. For both ChikE2 and ZikaE,
229 infection risk was high in cemetery-adjacent neighborhoods for the first three epidemic months before the
230 focus of infection risk shifted to northern neighborhoods and concluded thereafter.

231 CovidE exhibited different spatiotemporal dynamics than ChikE1, ChikE2, and ZikaE. Large swaths of the
232 study area had elevated SARS-CoV-2 infection risk during April 2020, particularly in northern and southern
233 neighborhoods. These neighborhoods, along with eastern ones, exhibited very high infection risk during
234 May-June 2020. When CovidE was diminishing in July 2020, elevated infection risk was still observed in
235 northern and southern neighborhoods. In contrast to the dynamics of the mosquito-borne epidemics and as
236 expected, cemetery-adjacent neighborhoods on the western edge of our study site were never the focal point
237 of CovidE. Because CovidE featured the lowest disease risk (Table 1), its infection dynamics differed most
238 from its incidence dynamics, underscoring the substantial and spatially heterogeneous bias arising from
239 common misinterpretations of case-based mapping.

240

241

Figure 3



Spatiotemporal dynamics across four epidemics in our study area. Model predictions of the infection risk (**A-D**) and incidence rate (**E-H**) are reported per-month and per-1,000 population. Figures S28-S35 contain additional information for this analysis. The PDCS epidemics are shown in a different color palette than CovidE given the different age structure, and hence infection risk, across cohort studies. A white triangle indicates the study health center. Neighborhoods are outlined in gray.

248 **DISCUSSION**

249 Across multiple analyses and four epidemics of three different viruses, we showed that traditional case-
250 only spatial approaches, which focus on the incidence rate, considerably underestimated the risks of
251 infection and disease, broadly impacting how epidemics were characterized. In particular, we demonstrated
252 that case-based spatial analyses were insufficient to recover the magnitude and spatial pattern of infection
253 risk, which conveys the extent of pathogen transmission and the spatial landscape of natural immunity. We
254 also showed that without infection data, the case-based incidence rate offered only a limited understanding
255 of all epidemics. For example, ChikE2, ZikaE, and CovidE appear similar based on their comparable
256 incidence rates. However, this resemblance reflects very different infection and disease risks, belying large
257 differences in population-level immunity and obscuring complex, epidemic-specific spatial patterns of
258 infection and disease risk that are critical for the appropriate design of effective interventions. Importantly,
259 we conducted our analyses in a population where it was possible to measure infection and disease outcomes
260 reliably for all participants,[14] showing that even in settings with ideal case ascertainment and serological
261 testing strategies, substantial biases occurred when the case-based incidence rate was the sole summary
262 measure for an epidemic. Such biases would be exacerbated in more typical settings with limited
263 surveillance and laboratory-testing capacity, including instances when only a sample of cases are available
264 for analysis.

265 ZikaE and CovidE exhibited remarkably high infection risks for their epidemic timescales, which only
266 lasted 3-4 months at their peak. In particular, the SARS-CoV-2 infection risk in the HICS (57.5%) was
267 higher than ZIKV infection risk in Managua's general population (45.8%), as previously reported.[16] Our
268 study population also experienced a higher SARS-CoV-2 infection risk than in Léon (33.8%), Nicaragua's
269 second largest city, after its first COVID-19 wave.[19] The more urban, more densely populated setting in
270 Managua may have contributed to more explosive SARS-CoV-2 transmission, and hence more population-
271 level immunity to SARS-CoV-2, than in Léon.

272 The association between socioeconomic status (SES) and infection risk has differed across epidemics and
273 pathogens. While we did not assess whether SES was associated with infection in our study, previous
274 seroprevalence studies of ZIKV in Managua[16] and SARS-CoV-2 in Léon[19] found no such association. The
275 pediatric nature of the PDCS precludes assessments of spatial trends for Managua's general population.
276 However, a previous analysis compared the ZIKV infection risk for children and adults in Managua;[16]
277 although ZIKV infection risk was higher among adults than children, reflecting the strong age-based
278 patterns we identify here, spatial trends were comparable between the two groups. It is therefore likely that
279 analyses of Managua's general population during ChikE1, ChikE2, and ZikaE would reveal similar spatial
280 trends, but with elevated magnitude, as those identified in the PDCS.

281 Contrary to the assumption that disease risk is spatially constant,[20] we observed that disease risk varied
282 spatially. Thus, the incidence rate at a given location was influenced by two spatially varying risks (infection
283 and disease). This may explain why the incidence clusters unpredictably resembled only one type of risk
284 cluster for each epidemic, and why seroprevalence and cluster analyses have seemed to diverge in areas
285 with large COVID-19 epidemics.[21] By identifying clusters of low infection risk, which contain many
286 uninfected persons, cluster detection analyses can reveal areas where policy interventions will be most
287 needed to prevent many infections in subsequent epidemics. The inability of case-based incidence clusters
288 to consistently mirror clusters of infection risks implies that policy decisions based on case-based incidence
289 clusters may misdirect limited public resources available for epidemic response.

290 Across epidemics, we found little evidence that infections were clustered within households in our study
291 area. Although lacking entomological survey data, our epidemiological analyses indirectly suggest that viral
292 transmission rarely occurred within study households. This conclusion is directly supported by a separate,
293 full-length sequencing analysis of ZIKV genomes in our study population (O. Sessions, personal
294 communication, July 2021), which found that many households had Zika cases whose most recently
295 sampled ancestors lived in different households. Together, the evidence suggests that non-household
296 transmission plays an important role in the epidemic dynamics of our study setting. To decrease epidemic

297 potential for mosquito-driven epidemics, better management of mosquito habitats, particularly cemeteries,
298 which are prolific mosquitoes breeding grounds,[16,22] is essential. The spread of SARS-CoV-2 in future
299 epidemic waves may be reduced by limiting access to highly visited locales such as recreational areas and
300 tourist attractions.

301 The case-based incidence rate captures the risk of disease when all individuals are susceptible to an outcome
302 (e.g., cardiovascular disease, death). However, for pathogens that cause subclinical infections, maps of the
303 incidence rate only convey *where* disease occurred, not the *spatial risk* of infection or disease.
304 Consequently, even with perfect case ascertainment, the incidence rate's underestimation of true risk
305 weakens the rationale to enact strong enough policy measures to control epidemics. As the case-based
306 incidence rate is frequently misinterpreted as a measure of infection and/or disease risk for pathogens of
307 high human concern that cause many subclinical infections (including malaria-causing *Plasmodium*;
308 *Mycobacterium tuberculosis*; and many pathogens transmitted by sex, air, vectors, and soil), our findings
309 concerning the limitations of case-based spatial studies likely generalize widely and impact many diseases
310 of global health importance that disproportionately affect neglected populations.

311 Measuring a population's infection status (not just that of suspected cases) has many benefits, the most
312 obvious being to guide interventions to areas of high transmission. As another example, due to antibody-
313 dependent enhancement, prior infection with DENV or ZIKV can lead to severe and life-threatening dengue
314 upon a subsequent DENV infection under certain circumstances.[18,23] Given widespread ZIKV immunity
315 across Latin America and the high prevalence of DENV across the tropics and subtropics, knowing where
316 ZIKV and DENV seroprevalence is high can highlight where episodes of severe dengue are likely to occur
317 and burden medical facilities. For SARS-CoV-2, knowledge of areas with a high proportion of uninfected
318 individuals is critical for public health measures, both to prioritize vaccine rollout, particularly in settings
319 where vaccine availability is severely limited, and to inform the need for other interventions. Others have
320 shown how to combine regional serosurvey data with real-time hospitalization data to estimate infection
321 risk for SARS-CoV-2 in near real-time at larger spatial scales.[24] As epidemic management necessitates

evaluating the risks of infection and disease across space, we argue for the expanded use of serosurveillance to overcome the inherit limitations of case-based spatial approaches.

MATERIALS AND METHODS

Ethical statement

The PDCS was approved by Institutional Review Boards of the University of California, Berkeley, the University of Michigan, Ann Arbor, and the Nicaraguan Ministry of Health. The HICS was approved by the University of Michigan, Ann Arbor and the Nicaraguan Ministry of Health. Participants' parents or legal guardians provided written informed consent. Subjects six years and older provided verbal assent.

Study design and eligibility criteria

The PDCS[14] is an open, population-based, prospective cohort of children initiated in 2004 to study dengue virus (DENV) and later expanded to include CHIKV and ZIKV. We assessed ~3,000 PDCS participants 2-14 years old who experienced two chikungunya epidemics (2014, 2015) and one Zika epidemic (2016).[3,5,15,16] The HICS is an open, population-based, prospective cohort of households that began studying influenza virus and other respiratory pathogens, including coronaviruses, in 2017. We evaluated 1,793 HICS participants 0-87 years old who experienced the first wave (2020) of the COVID-19 epidemic. While all PDCS participants are pediatric, the age structure of the HICS is representative of Managua's general population.

Both cohort studies share the same study site (Fig. S1) in Managua. During the studies' annual sampling in March/April, participants provide blood samples to ascertain infection status during the prior year. A mid-year sampling was instituted in the HICS in October/November 2020 to measure SARS-CoV-2 infections after the first COVID-19 wave but before the second. Both studies provide participants with primary care, and participants agree to visit the study health center at the first indication of any illness.

344 Analysis of each epidemic was restricted to participants who lived within the catchment area of the study
345 health center, were enrolled before the start of the epidemic, and were at risk for incident infection. See
346 appendix (pages 5-7) for additional study design information.

347 **Laboratory methods**

348 Upon collection, annual blood samples were immediately transported to the Nicaraguan National Virology
349 Laboratory of the Ministry of Health for processing and storage at -80°C. Paired annual samples (2014-
350 2015 and 2015-2016) demonstrating seroconversion by CHIKV Inhibition ELISA[25] indicated CHIKV
351 infection. ZIKV infection status was confirmed by the 2017 result of the ZIKV NS1 blockade-of-binding
352 assay[26] on paired 2017-2018 annual samples. SARS-CoV-2 infection status was confirmed by the "Mount
353 Sinai ELISA" protocol,[27] primarily on 2020 midyear samples. Participants with laboratory-confirmed
354 infections who did not seek medical care were categorized as having had inapparent infections. Acute and
355 convalescent samples from participants suspected of chikungunya, Zika, or COVID-19 were tested using
356 molecular, virological, and serological assays.[3,14,16] See appendix (pages 7-8) for additional information on
357 laboratory methods.

358 **Statistical analyses**

359 We characterized the four epidemics non-spatially, spatially, and spatiotemporally by assessing the
360 infection risk, disease risk, and traditional case-based incidence rate. All statistical tests were two-tailed.
361 Overall values for these three measures were estimated using intercept-only logistic models. Predicted
362 values across the study area were estimated with generalized additive models[28] using two-dimensional
363 splines on households' longitude and latitude. The intracluster correlation coefficient was used to measure
364 the intra-household correlation of infection and disease outcomes. We used SaTScan v9.4.4 to identify
365 hierarchical and the newer Gini clusters of infection risk, disease risk, and case incidence.[29,30] Hierarchical
366 clusters identify the most statistically likely clusters and Gini clusters maximize outcome rates.
367 Geostatistical mixed models[31] were used to evaluate the association of risk factors with infection and
368 disease outcomes. Infection dynamics were estimated by treating cases as a spatiotemporal Poisson point

369 process arising from the overall population and then accounting for the spatial distribution of the disease
370 risk, assumed to be time-invariant. Participants were geolocated to their households' location. Initially
371 uninfected participants were considered at risk for infection and infected participants were considered at
372 risk for disease. Analyses used the EPSG:4326 coordinate reference system and were performed in R v3.6.2.
373 See appendix (pages 9-17) for additional information on statistical methods.

374 **Data availability:** Individual-level data may be shared with outside investigators following UC Berkeley
375 and UM Ann Arbor IRB approval. Data has been deposited in a HIPAA-compliant Dropbox account hosted
376 by the University of Michigan and will be available by request, as is required by the IRB-approved protocols
377 for the PDCS and the HICS. Please contact the UC Berkeley Center for the Protection of Human Subjects
378 (ophs@berkeley.edu), Eva Harris (eharris@berkeley.edu), the University of Michigan Health Sciences and
379 Behavioral Sciences IRB (irbhsbs@umich.edu), and Aubree Gordon (gordonal@umich.edu) to arrange for
380 data access. Databases without names and other identifiable information were used for all analyses.
381 Collaborating research groups and institutions will be sent coded data with all personal identifiers unlinked
382 as well as data dictionaries and the accompanying R code.

383

Acknowledgments: We are extremely appreciative of our dedicated study team at the Centro de Salud Sócrates Flores Vivas and the Laboratorio Nacional de Virología at the Centro Nacional de Diagnóstico y Referencia, Nicaraguan Ministry of Health; and the Sustainable Sciences Institute in Nicaragua. We are grateful to Art Reingold for his thoughtful review of the manuscript, and we thank Burke Bundy and Suzanne Default at the University of California, Berkeley, for enabling us to use the computer cluster of the Division of Epidemiology and Biostatistics for our geostatistical modeling. We thank François Rousset for expert consultation regarding spatial generalized linear mixed models and their implementation in the spaMM R package. Most importantly, we thank the PDCS and HICS study participants and their families for engaging with us in the endeavor of science.

Funding: This study was supported by grants R01 AI099631 (AB), P01 AI106695 (EH), R01 AI120997 (AG), and U19 AI118610 (EH) from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health; the National Institutes of Health Centers of Excellence for Influenza Research and Surveillance [contract: HHS 272201400006C (AG)]; and the Open Philanthropy Project Fund for the production of recombinant SARS-CoV-2 spike protein, its receptor binding domain, and antibodies at the University of Michigan Center for Structural Biology. FBC was partially supported by a supplement to grant P01 AI106695.

Author contributions: FBC and EH conceived the study. AB, JC, AG, KG, and EH developed the study design. SO, NS, MP, HSL, AB, AMG, and KG implemented the study design and collected field data. AB developed and supervised the laboratory assays. SS, CB, DC, and TM performed laboratory testing. BLM, JCM, DE, SA, HM, KL, and BC cleaned and verified the data. LCK developed serological algorithms. FBC analyzed the data, performed the statistical analyses, and generated the figures. FBC, HS, AG, and EH drafted and revised the manuscript, and all authors reviewed the manuscript. FBC, AB, AG, and EH acquired funding.

Competing interests: Authors declare that they have no competing interests.

408 **REFERENCES**

409 1. Higgs S, Vanlandingham D, Powers A, editors. Chikungunya and Zika Viruses: Global Emerging
410 Health Threats. 1st ed. Academic Press; 2018.

411 2. Cevik M, Kuppalli K, Kindrachuk J, Peiris M. Virology, transmission, and pathogenesis of SARS-
412 CoV-2. BMJ. 2020 Oct 23;371.

413 3. Burger-Calderon R, Bustos Carrillo F, Gresh L, Ojeda S, Sanchez N, Plazaola M, et al. Age-
414 dependent manifestations and case definitions of paediatric Zika: a prospective cohort study. Lancet
415 Infect Dis. 2020 Dec;20(3):371–80.

416 4. Haby M, Pinart M, Elias V, Reveiz L. Prevalence of asymptomatic Zika virus infection: a systematic
417 review. Bull World Health Organ. 2018 Jun 1;96(6):402-413D.

418 5. Bustos Carrillo F, Collado D, Sanchez N, Ojeda S, Lopez Mercado B, Burger-Calderon R, et al.
419 Epidemiological evidence for lineage-specific differences in the risk of inapparent chikungunya
420 virus infection. J Virol. 2019 Nov 21;93(4):e01622-18.

421 6. Johansson M, Quandelacy T, Kada S, Prasad PV, Steele M, Brooks J, et al. SARS-CoV-2
422 transmission from people without COVID-19 symptoms. JAMA Netw Open. 2021 Jan
423 1;4(1):2035057.

424 7. McHale T, Romero-Vivas C, Fronterre C, Arango-Padilla P, Waterlow N, Nix C, et al.
425 Spatiotemporal heterogeneity in the distribution of chikungunya and Zika virus case incidences
426 during their 2014 to 2016 epidemics in Barranquilla, Colombia. Int J Environ Res Public Health.
427 2019 May 2;16(10).

428 8. Aguiar BS, Lorenz C, Virginio F, Suesdek L, Chiaravalloti-Neto F. Potential risks of Zika and
429 chikungunya outbreaks in Brazil: A modeling study. Int J Infect Dis. 2018 May 1;70:20–9.

430 9. Campos de Lima EE, Gayawan E, Baptista EA, Queiroz BL. Spatial pattern of COVID-19 deaths
431 and infections in small areas of Brazil. PLoS One. 2021 Feb 1;16(2 Febuary).

432 10. Vissoci JRN, Rocha TAH, Silva NC da, de Sousa Queiroz RC, Thomaz EBAF, Amaral PVM, et al.
433 Zika virus infection and microcephaly: Evidence regarding geospatial associations. PLoS Negl Trop
434 Dis. 2018 Apr 25;12(4).

435 11. Bonilla-Aldana DK, Bonilla-Aldana JL, García-Bustos JJ, Lozada CO, Rodríguez-Morales AJ.
436 Geographical trends of chikungunya and Zika in the Colombian Amazonian gateway department,
437 Caqueta, 2015–2018 – Implications for public health and travel medicine. Travel Med Infect Dis.
438 2020 May 1;35.

439 12. Fatima M, O'keefe KJ, Wei W, Arshad S, Gruebner O. Geospatial analysis of covid-19: A scoping
440 review. Vol. 18, Int J Environ Res Public Health. MDPI AG; 2021. p. 1–14.

441 13. Wang Y, Di Q. Modifiable areal unit problem and environmental factors of COVID-19 outbreak.
442 Sci Total Environ. 2020 Oct 20;740:139984.

443 14. Kuan G, Gordon A, Aviles W, Ortega O, Hammond SN, Elizondo D, et al. The Nicaraguan Pediatric
444 Dengue Cohort Study: Study design, methods, use of information technology, and extension to other
445 infectious diseases. Am J Epidemiol. 2009 Jul 1;170(1):120–9.

446 15. Gordon A, Gresh L, Ojeda S, Chowell-Puente G, Gonzalez K, Sanchez N, et al. Differences in
447 transmission and disease severity between two successive waves of chikungunya. Clin Infect Dis.
448 2018 Apr 25;67(11):1760–7.

449 16. Zambrana JV, Bustos Carrillo F, Burger-Calderon R, Collado D, Sanchez N, Ojeda S, et al.
450 Seroprevalence, risk factor, and spatial analyses of Zika virus infection after the 2016 epidemic in
451 Managua, Nicaragua. Proc Natl Acad Sci. 2018 Sep 11;115(37):9294–9.

452 17. Gordon A, Gresh L, Ojeda S, Katzelnick LC, Sanchez N, Mercado JC, et al. Prior dengue virus
453 infection and risk of Zika: A pediatric cohort in Nicaragua. von Seidlein L, editor. PLOS Med. 2019
454 Jan 22;16(1):e1002726.

455 18. Katzelnick LC, Narvaez C, Arguello S, Mercado BL, Collado D, Ampie O, et al. Zika virus infection

456 enhances future risk of severe dengue disease. Science. 2020 Aug 28;369(6507):1123–8.

457 19. González F, Vielot NA, Sciaudone M, Toval-Ruíz C, Premkumar L, Gutierrez L, et al.
458 Seroepidemiology of SARS-CoV-2 infections in an urban Nicaraguan population. medRxiv Prepr
459 Serv Heal Sci. 2021 Mar 1.

460 20. Moore SM, ten Bosch QA, Siraj AS, Soda KJ, España G, Campo A, et al. Local and regional
461 dynamics of chikungunya virus transmission in Colombia: The role of mismatched spatial
462 heterogeneity. BMC Med. 2018 Aug 30;16(1):152.

463 21. Eckerle I, Meyer B. SARS-CoV-2 seroprevalence in COVID-19 hotspots. Lancet. 2020 Aug
464 22;396(10250):514–5.

465 22. Vezzani D. Review: Artificial container-breeding mosquitoes and cemeteries: A perfect match. Trop
466 Med Int Heal. 2007 Feb;12(2):299–313.

467 23. Katzelnick LC, Gresh L, Halloran ME, Mercado JC, Kuan G, Gordon A, et al. Antibody-dependent
468 enhancement of severe dengue disease in humans. Science. 2017 Nov 17;358(6365):929–32.

469 24. Hozé N, Paireau J, Lapidus N, Kiem CT, Salje H, Severi G, et al. Monitoring the proportion of the
470 population infected by SARS-CoV-2 using age-stratified hospitalisation and serological data: a
471 modelling study. Lancet Public Heal. 2021 Jun 1;6(6):e408–15.

472 25. Saborío Galo S, González K, Téllez Y, García N, Pérez L, Gresh L, et al. Development of in-house
473 serological methods for diagnosis and surveillance of chikungunya. Rev Panam Salud Publica. 2017
474 Aug 21;41:e56.

475 26. Balmaseda A, Stettler K, Medialdea-Carrera R, Collado D, Jin X, Zambrana JV, et al. Antibody-
476 based assay discriminates Zika virus infection from other flaviviruses. Proc Natl Acad Sci. 2017
477 Aug 1;114(31):8384–9.

478 27. Amanat F, Stadlbauer D, Strohmeier S, Nguyen THO, Chromikova V, McMahon M, et al. A

479 serological assay to detect SARS-CoV-2 seroconversion in humans. Nat Med. 2020 Jul
480 1;26(7):1033–6.

481 28. Hastie T, Tibshirani R. Generalized Additive Models. New York: Chapman & Hall; 1990.

482 29. Kulldorff M. A spatial scan statistic. Commun Stat - Theory Methods. 1997;26(6):1481–96.

483 30. Han J, Zhu L, Kulldorff M, Hostovich S, Stinchcomb DG, Tatalovich Z, et al. Using Gini coefficient
484 to determining optimal cluster reporting sizes for spatial scan statistics. Int J Health Geogr. 2016
485 Aug 3;15(1):27.

486 31. Rousset F, Ferdy J-B. Testing environmental and genetic effects in the presence of spatial
487 autocorrelation. Ecography (Cop). 2014;37(8):781–90.

488