The duration of travel impacts the spatial dynamics of infectious diseases

John R. Giles[a,1] (ORCID), Elisabeth zu Erbach-Schoenberg[b,c], Andrew J. Tatem[b,c] (ORCID), Lauren Gardner[d], Ottar N. Bjørnstad[e] (ORCID), C. J. E. Metcalf[f,g] (ORCID), and Amy Wesolowski[a] (ORCID)

[a]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205; [b]Department of Geography and the Environment, University of Southampton, Southampton SO17 1BJ, United Kingdom; [c]WorldPop, University of Southampton, Southampton SO17 1BJ, United Kingdom; [d]Department of Civil and Systems Engineering, Johns Hopkins Whiting School of Engineering, Baltimore, MD 21218; [e]Department of Entomology, Pennsylvania State University, University Park, PA 16802; [f]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544; and [g]Princeton School of Public and International Affairs, Princeton University, Princeton, NJ 08544

Humans can impact the spatial transmission dynamics of infectious diseases by introducing pathogens into susceptible environments. The rate at which this occurs depends in part on human-mobility patterns. Increasingly, mobile-phone usage data are used to quantify human mobility and investigate the impact on disease dynamics. Although the number of trips between locations and the duration of those trips could both affect infectious-disease dynamics, there has been limited work to quantify and model the duration of travel in the context of disease transmission. Using mobility data inferred from mobile-phone calling records in Namibia, we calculated both the number of trips between districts and the duration of these trips from 2010 to 2014. We fit hierarchical Bayesian models to these data to describe both the mean trip number and duration. Results indicate that trip duration is positively related to trip distance, but negatively related to the destination population density. The highest volume of trips and shortest trip durations were among high-density districts, whereas trips among low-density districts had lower volume with longer duration. We also analyzed the impact of including trip duration in spatial-transmission models for a range of pathogens and introduction locations. We found that inclusion of trip duration generally delays the rate of introduction, regardless of pathogen, and that the variance and uncertainty around spatial spread increases proportionally with pathogen-generation time. These results enhance our understanding of disease-dispersal dynamics driven by human mobility, which has potential to elucidate optimal spatial and temporal scales for epidemic interventions.

spatial disease dynamics | human mobility | call data records | trip duration | spatial TSIR

Modern human populations are characterized not only by their wide-ranging spatial distribution, but also by the behavioral travel patterns that form the basis of human-population mobility (1). Individual-level travel is often characterized by short movements to few locations, with occasional travel to distant locales, giving human-mobility patterns a long-tailed distribution (2). The sum effect of these individual trajectories comprises connectivity among human communities, which is a fundamental driver of the spatial transmission of infectious diseases (3–7). Infected individuals who travel into new susceptible populations may introduce pathogens that result in disease outbreaks (8–10). Traveling susceptibles may also result in "spatial back spill" if encountering infecteds in a neighboring (or distant) town. Therefore, quantifying human travel on scales relevant for disease transmission is necessary to predict pathogen spatial spread and populations at risk (11, 12).

In the context of spatial disease dynamics, travel encompasses not only the number of individuals who move between locations, but also the duration of these trips. Although the number of trips has been extensively studied by using both individual- and population-level data, less emphasis has been placed on how the duration of these trips impacts spatial transmission (13–17).

The amount of time an infected individual spends in a destination while traveling (trip duration) may impact the probability of onward transmission, where longer trips increase the likelihood of transmission in the visited location (18). Conversely, shorter turn-around times of wayward exposed susceptibles may also enhance spatial contagion. In either case, the trip duration modulates these aspects of spatial contagion based on the proportion of the pathogen's generation time spent in the destination.

Few datasets exist that can comprehensively and systematically quantify both the number of trips between locations and the duration of those trips for populations. Census-based datasets—including migration data (19), which measure a change in the residence location, and journey-to-work surveys (20), which measure the frequency and location of occupation travel—have limited additional information on the time spent in these locations (13, 14). Individual-level mobility data, such as global-positioning system loggers and travel surveys, are able to include information about the duration of travel, but are rarely generalizable across large populations (21, 22). Mobility data derived from mobile-phone records can add an additional source of information on both locational travel and trip duration. However, these data can be difficult to obtain and computationally intensive to process. In previous analyses, often only the number of trips between locations has been analyzed (6, 23–26), despite the fact that other aspects of travel can be mined from these data.

Significance

The spatial dynamics of infectious-disease spread are driven by the biology of the pathogen and the connectivity patterns among human populations. Models of disease spread often use mobile-phone calling records to calculate the number of trips made among locations in the population, which is used as a proxy for population connectivity. However, the amount of time people spend in a destination (trip duration) also impacts the probability of onward disease transmission among locations. Here, we developed models that incorporate trip duration into the mechanism of disease spread, which helps us understand how fast and how far a pathogen might spread in a human population.

Integrating data and models of human travel into disease models has been extensively used to predict or understand the spatial spread of various infectious diseases (7, 9, 27–29). In these models, infected humans in an index location introduce pathogens into areas where they are currently absent—either because of previous control, previous susceptible depletion, or because they are novel pathogens—which leads to new disease outbreaks. The strategic deployment of public health interventions, such as reactive vaccination campaigns (30–32), increased access to preventative and treatment options (33–35), and travel restrictions (36, 37), all require an accurate prediction of where and when introduction events will occur. Data describing aspects of human travel on spatial and temporal time scales relevant to disease transmission are often unavailable; hence, researchers and public health officials have relied on models of mobility. However, adding duration to these models of trip counts may help capture aspects of travel to more realistically describe epidemiologically relevant mobility patterns.

Here, we analyzed mobility data derived from mobile-phone calling data in Namibia to develop a model that directly integrates both the number of trips and the duration of travel. Although these data have been studied previously (24, 26), the importance of including duration in a more general framework has yet to be considered. Using a hierarchical model of mobility, we investigate how the duration of travel varies with the commonly used geographic variables of population size and distance between the origin and destination. We also assess the importance of including trip duration in spatial disease-transmission models for a range of pathogen life histories and locales of introduction. We show that including the duration of travel into transmission models delays the timing of spatial spread by making transmission dynamics less coupled among locations. We further discuss the variability of this effect across pathogens and its implications for the timing of control and outbreak responses.

Results

We analyzed daily call data records (CDRs), which estimate the movement of mobile-phone subscribers between 105 districts in Namibia to build daily origin–destination travel matrices (Fig. 1). In total, over 259 million trips made by ~2.5 million subscribers from October 1, 2010, to April 29, 2014, were analyzed.

We aggregated the data by assigning each subscriber to a daily location (district) based on which district contained their most-used mobile-phone tower. A "trip" was counted if the most-used mobile-phone towers were in different districts on subsequent days; otherwise, a subscriber was classified as staying in the same location. For each trip, the date the trip began and the number of days the subscriber spent in the destination district (trip duration) were recorded. We used daily trip-duration counts to estimate the decay rate in duration for each route and then used these estimates in a modified gravity model fitted to mean monthly trip counts. In total, we analyzed travel between 62 of the 105 districts that contained a minimum of 20 observations to ensure model convergence (*Materials and Methods* and Fig. 1 *A* and *B*). Districts were further classified based on their population density, resulting in 10 high-density (>1,000 people per km^2) and 52 low-density ($\leq$1,000 people per km^2) districts (*Materials and Methods* and *SI Appendix*, Fig. S1).

The Relationship between Trip Counts and Duration with Distance and Population. The majority of trips were made over short geographic distances and for a short duration, though the actual trip counts were likely biased due to the distribution of the mobile-phone towers. In terms of distance, half of the total trips traversed less than 25 km (46%), and a moderate percentage of the total trips were farther than 100 km (21%). Overall, these trips lasted up to a week, with a mean of 6.4 d (95% CI: 1 to 44) and maximum of 1,307 d. Approximately half of the total trips remained in the destination for 1 d (47%)—the temporal scale over which these data were aggregated—or stayed for durations longer than 2 wk (8%; *SI Appendix*, Table S1). Overall, counts of both trip duration and trip distance decayed rapidly, consistent with an exponential distribution in the case of trip duration and a Gamma distribution in the case of trip distance (*SI Appendix*, Fig. S4).

To summarize the distribution of observed trip durations along each route, we estimated the exponential rate of decay in trip count as a function of trip duration using a hierarchical Bayesian model. The model estimated the mean decay rate in trip duration at the population level (λ') and the decay rate of each i to j route (λ_{ij}; *Materials and Methods*). We estimated the overall mean number of trips per route lasting 1 d to be 2,220 and the population-level decay parameter (λ') to be 0.43 among all
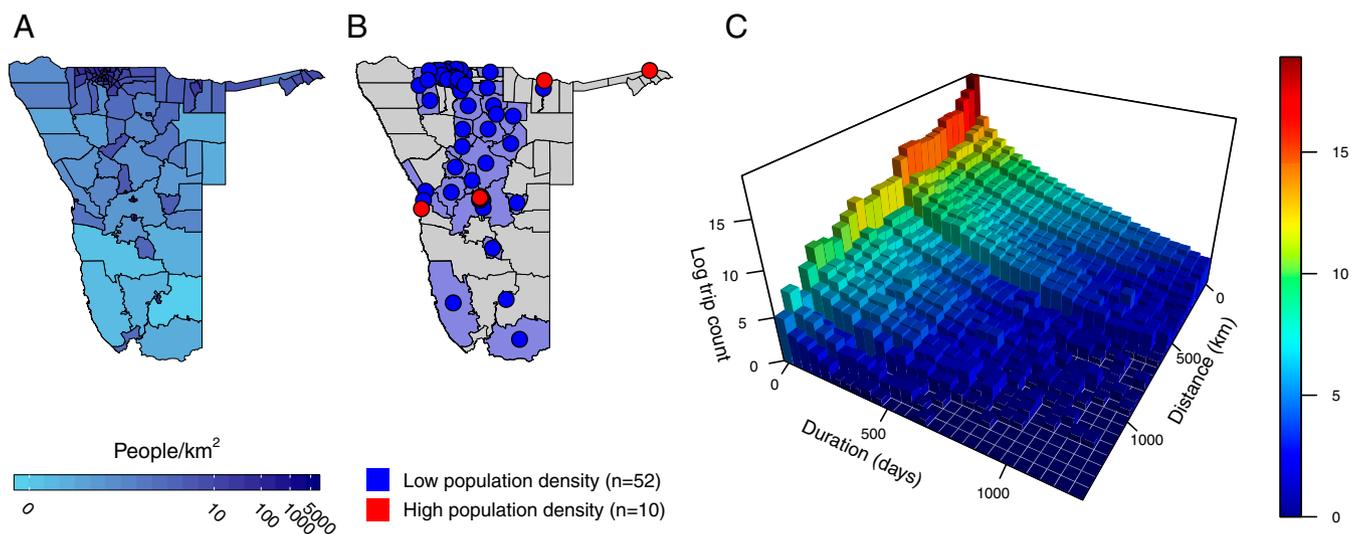
Fig. 1. The distribution of the population in Namibia and mobility data used in analyses. (*A*) Log-transformed population density for all districts in Namibia with districts with lower and higher relative population density in light blue and dark blue, respectively. (*B*) Districts included in the analysis are shaded with colored centroids (n = 62). Districts with low population density (n = 52) are shown in blue, and districts with high population density (n = 10) are shown in red. (*C*) The log count of trips made for a given trip duration (days) and trip length (kilometers).

People/km²

Low population density (n=52)
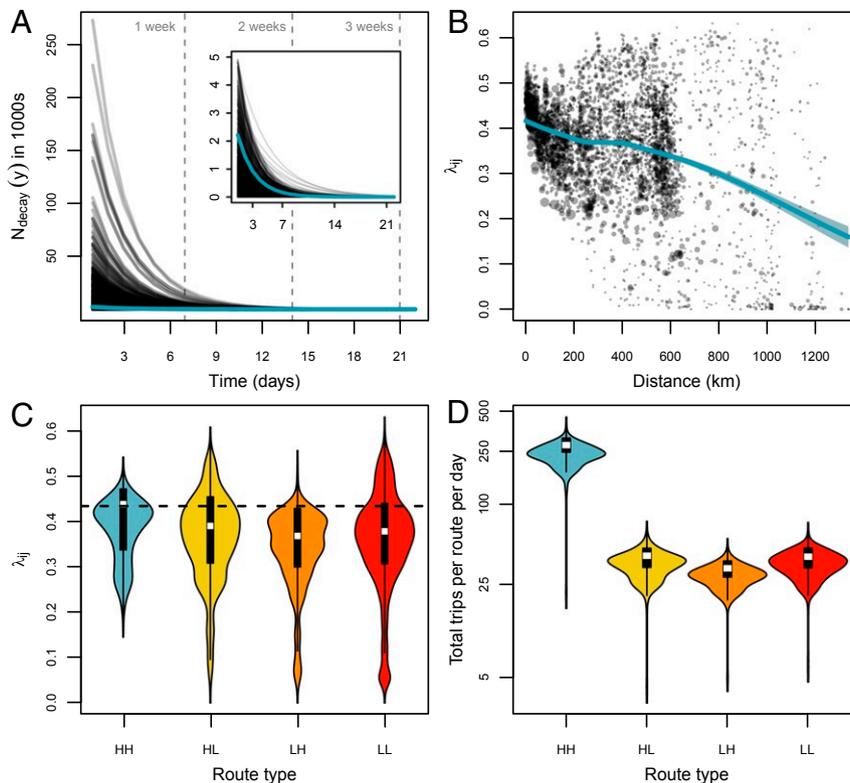High population density (n=10)

Fig. 2. Results from the hierarchical Bayesian model showing that most trips are short and among high-density districts, but duration increases for longer trip distances and low-density destinations. (*A*) Trip-duration decay rate ($\hat{\lambda}_{ij}$) plotted over time (days) for each ij route (transparent black lines) and population mean (teal line). *Inset* shows close-up of the population mean for smaller values of N_{decay} (y). (*B*) The relationship between $\hat{\lambda}_{ij}$ and the distance separating districts i and j. The teal line shows the trend between these two variables represented as a LOESS regression with 95% CIs (shaded region). In *C* and *D*, the values of $\hat{\lambda}_{ij}$ and proportion of total trips are plotted using a cutoff of 1,000 people per km² to define the origin i and destination j, according to "high" or "low" population density. (*C*) Violin plots show the distribution of $\hat{\lambda}_{ij}$ for four route types compared to the population mean (dashed line). (*D*) Violin plots show the log-scaled distribution of the proportion of total trips for the four route types over each day. Abbreviation HH indicates travel from high to high population density, and LL indicates low to low.

districts (0.43 to 0.44 95% highest posterior density [HPD]), which means that, on average, the number of trips decayed at the rate of roughly 0.37% per day added to the trip duration (Fig. 2*A* and *SI Appendix*, Fig. S4).

Estimates of the route-level trip-duration decay rate (λ_{ij}) varied widely compared to the population mean and were largely dependent on both trip distance and the population density of the origin and destination districts. We estimated the correlation between trip-duration decay parameter (λ_{ij}) and observed trip distance to be −0.29 (−0.32, −0.26 95% CI). This significant negative relationship—where trip-duration decay rate decreased as trip distance increases—indicated that mean trip length ($1/\lambda_{ij}$) increased with distance. We fit a locally estimated scatterplot smoothing (LOESS) regression weighted by sample size to show that the strength of this relationship varied over distance, with an accelerated decline from 0 to 200 km and after 600 km (Fig. 2*B*). Estimates of the duration decay rate (λ_{ij}) were higher than the population mean for trips between high-density districts ($\lambda_{ij} = 0.45$, 0.2 to 0.53 95% HPD), which indicated that trips of this route type were consistently shorter than all other route types when considering a threshold of 1,000 people per km² (Fig. 2*C* and *SI Appendix*, Fig. S5 and Table S2). When we used a threshold of 2,500 people per km², this relationship was more pronounced ($\lambda_{ij} = 0.48$, 0.4 − −0.53 95% HPD), with an increase in λ_{ij} for the low- to high-density route type as well. Our analysis thus revealed that density of the destination district was inversely related to trip duration, where higher population density translated to higher decay rates and shorter duration (*SI Appendix*, Fig. S5 and Table S2). For

the total observed trips taken in each route type, we found that routes between high-density districts had seven to eight times more trips per route per day compared to other route types (Fig. 2*D*). Overall, these results suggest that trips typically have a short duration (<1 wk) which increases with distance; when the destination is a high-density district, the duration is short, but when the destination is a low-density district, duration is longer.

Updated Gravity Model Incorporating Both Trip Counts and Duration. Spatial-interaction models of mobility, particularly the gravity model, are commonly used to estimate the number of trips between locations. Based on our mobility analyses, we formulated a gravity model that explicitly used data from trip counts between locations and the duration of those trips (*Materials and Methods*). To compare, we also fit a basic gravity model that used only data from trip counts. Both gravity models estimate connectivity values (π_{ij}), which are the expected proportion of trips leaving an origin district i that travel to destination district j. We found that the duration gravity model, like the basic gravity model, was primarily dependent on distance and population size of the destination, where π_{ij} was higher for routes with a trip distance <250 km (*SI Appendix*, Fig. S8*A*) and routes where the destination population density was ∼100 people per km² (*SI Appendix*, Fig. S8*C*). There were significant differences in π_{ij} according to route type, with the highest values observed for routes among high-density districts, with much lower values estimated for routes where a low-density district was the origin, destination, or both (*SI Appendix*, Figs. S8*D* and S12). The

lowest values of π_{ij} were fitted for routes that went from high to low population density (*SI Appendix*, Fig. S8D). In addition, there was a general pattern where districts of similar population density had higher connectivity values (*SI Appendix*, Fig. S12). These results suggest that, while the broad patterns in connectivity estimated by both the basic and duration gravity models are similar, including information on trip duration can help model fit for less populated routes and provide an explicit model formulation that accounts for the observed dependence of trip distance and trip duration.

We also explored differences in gravity model fit to observed trip-count data and found that the duration gravity model provided marginally better fit to the data compared to the basic gravity model (Fig. 3 *C* and *D*; basic gravity model: $r = 0.6$ [-0.04, 0.98 95% CI]; duration gravity model: $r = 0.62$ [0.1, 0.98 95% CI]). Although the mean change in model fit across all origin districts was incremental, model fit was improved most for

districts that had lower initial fit and origins with low population density (Fig. 3 *C* and *E*). Goodness of fit did not change for districts with high population density where the basic gravity model already performed well (Fig. 3 *D* and *F*). District 24 (Luderitz) was an outlier, where goodness of fit decreased drastically for the duration gravity model (Fig. 3 *E* and *F*). This particular district includes a large desert on the southwestern coast of Namibia with a total population size of $13,500$ and the lowest population density in the country (0.26 people per km^2). Although this district met minimum sample-size criteria, it had the lowest number of unique observations of trip duration. Reduced performance here was likely due to poor estimation of the $\lambda_{24.}$ parameters in the decay model stemming from low sample size and distance from other districts (*SI Appendix*, Fig. S7 *A4* and *B4*). Overall, this indicates that including trip duration in the gravity model can help to improve the fit to observed trip counts along some routes of travel more so than others and that
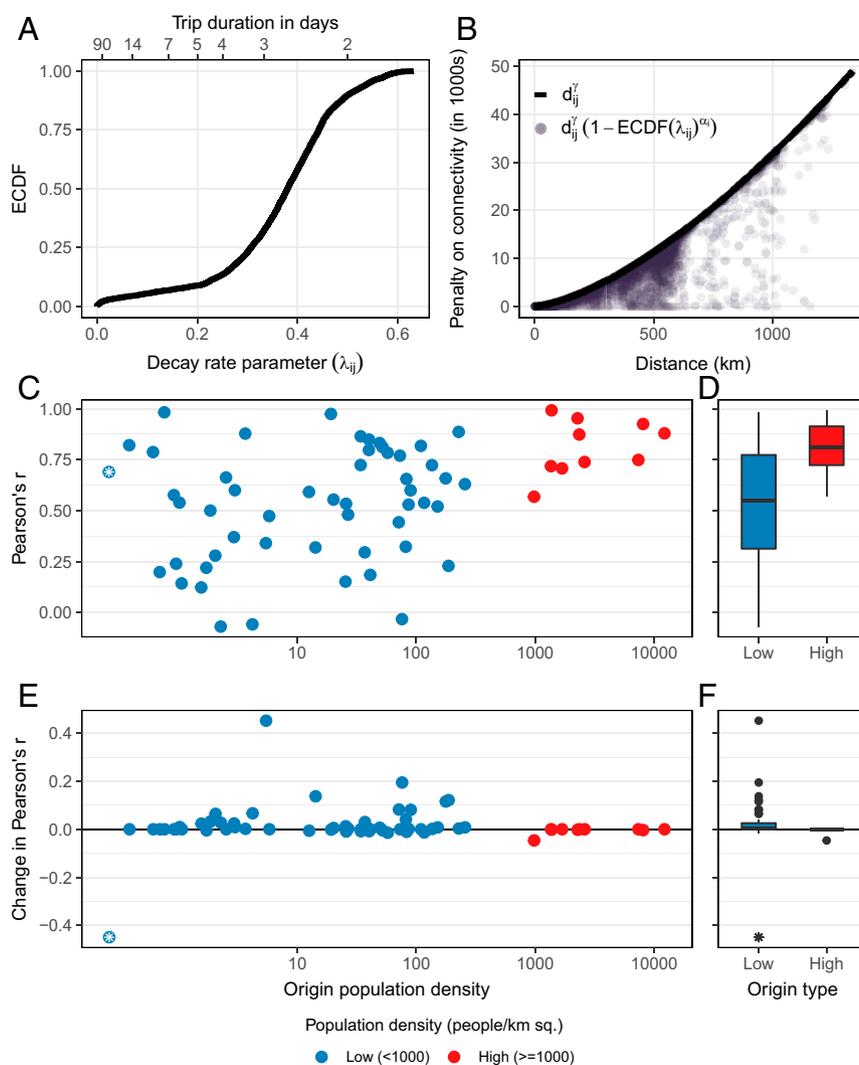

Fig. 3. Inclusion of the conditional dispersal kernel in the duration gravity model improves fit to data when the destination is a low-density district. (*A*) The Empirical Cumulative Distribution Function (ECDF) of the trip-duration decay parameter λ_{ij}. The mean trip duration in days ($1/\lambda_{ij}$) is shown on the top axis. (*B*) The distance-based functions that act as a penalty on fitted connectivity values in both the basic (d_{ij}^{γ}; black line) and duration gravity models ($d_{ij}^{\gamma}(1 - ECDF(\lambda_{ij})^{\alpha_i})$; purple circles). (*C* and *D*) Goodness of fit (Pearson's *r*) of basic gravity model for each of the 105 districts plotted by population density (*C*) and the overall distribution Pearson's *r* for origin districts with low and high population density (*D*). (*E* and *F*) The change in Pearson's *r* for each origin district when duration is included in the gravity model is plotted against origin population density (*E*) and the overall distribution of the change in Pearson's *r* for low- and high-density districts (*F*). Low-density origin districts are shown in blue ($n = 52$), and high-density districts are shown in red ($n = 10$). Sq., squared. District 24 (Luderitz) has unusually low population density and is indicated with an asterisk.

this is contingent upon robust estimation of trip-duration decay (*SI Appendix*, Fig. S11).

Impact of Trip Duration on Spatial Disease Dynamics. We incorporated trip duration decay into a stochastic disease-diffusion model to evaluate the impact of including trip duration on spatial-transmission dynamics. In the model, we used a Time-Series Susceptible Infected Recovered (TSIR) framework and calculated the spatial waiting time hazard of disease introductions to each district (*Materials and Methods*). We compared the resulting spatial dynamics for simulations from a range of pathogens and introduction events that included trip duration (duration TSIR model) to those that did not (basic TSIR model; *Materials and Methods*). Importantly, the duration TSIR model explicitly includes an interaction between the length of a trip and the generation time of the pathogen where trips are weighted based on the generation time.

We simulated scenarios for six different types of pathogens with various R_0 and generation times (*Materials and Methods* and *SI Appendix*, Table S3). Unsurprisingly, introduction events occurred earlier for districts with high population density that neighbor the index district. For simulations using the duration TSIR model, peak time until introduction was generally later, with a wider range of introduction times compared to the basic

TSIR model (Fig. 4). Including trip duration delayed spatial spread by reducing the spatial coupling of transmission among districts (*SI Appendix*, Fig. S14), although the magnitude of this effect varied based on the type of pathogen and initial introduction district (*SI Appendix*, Fig. S15). For a small number of simulations of pathogens with high R_0 values, introduction times were earlier for the duration TSIR model compared to the basic TSIR model, which suggests that in some instances, the effect of a higher R_0 negates any impact of including duration (Fig. 5*A* and *SI Appendix*, Fig. S17). Further, we found that the overall magnitude of the delay caused by the inclusion of trip duration was comparable between high- and low-density introductions (*SI Appendix*, Fig. S16*A*); however, there were measurable differences in the uncertainty and variance of importation times to other districts. In simulations with high-density introductions, variability in peak importation times and uncertainty around each peak was increased compared to simulations with a low-density introduction (Fig. 5*B* and *SI Appendix*, Fig. S17*B*). That is to say, introduction events to other locations are more spread out over time, but the window of time in which they may occur is longer. The magnitude of this increase was proportional to the generation time of the pathogen, which suggests that pathogens with longer generation times exhibit more variable spatial dynamics because the mean trip duration (6.4 d) comprises a smaller proportion of the generation time compared to a shorter generation time, which is more robust to degradation by trip duration.

Discussion

Spatial infectious-disease dynamics are often driven by human travel. Both the number of trips and duration of those trips may be relevant, but the latter has rarely been considered to date. Here, using a dataset of human mobility quantified from mobile-phone calling records in Namibia, we developed a spatial interaction model that incorporates both factors and assess the impact of including duration on disease dynamics. We find that, while the duration of trips is positively related to the distance between origin and destination, the number of those trips is inversely related with distance. Although including trip duration in a gravity model only marginally improves model fit to the data, we see a larger improvement in estimating trips from less-populated areas, which is a well-known limitation of the basic gravity model (38, 39). Overall, including duration into a spatial disease transmission model decreases spatial coupling between locations, resulting in longer waiting times until disease introduction compared to a basic transmission model. These results are fairly robust to the type of pathogen and the location of the first introduction event.

The gravity model that includes the duration of trips developed here is reliant on highly detailed mobility data. One major critique of the gravity model is that it is overparameterized, leading people to propose a parameter-free alternative named the Radiation model (40). In scenarios where data of human mobility is lacking or unavailable, estimating fewer model parameters is more tractable. However, availability of CDRs supports estimation of additional parameters required by the gravity model because there is a greater amount of information regarding human movement that increases the explanatory power of the data. Although mobile-phone data can provide highly detailed information about travel patterns, it remains unclear to what extent these data and models are generalizable to other countries than those analyzed. We found that trip duration was exponentially distributed in all locations and identified relationships between model parameters and commonly available covariates, such as distance and population size (*SI Appendix*, Fig. S7 *B1* and *D1*), which suggests that a general model framework could be defined. However, the relationship between trip duration and available covariates may not hold in all settings due to geographic
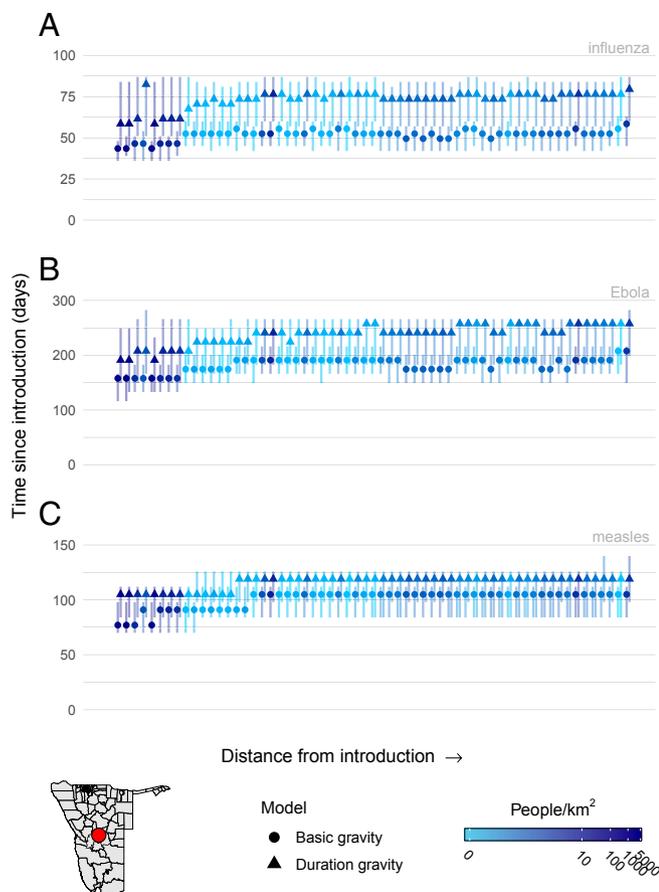

Fig. 4. Spatial TSIR simulations of infectious-disease dispersal for three pathogens—influenza (*A*), Ebola (*B*), and measles (*C*)—introduced into the high-density district Windhoek East (red circle). Caterpillar plots represent aggregated waiting-time distributions for all simulations, where the peak waiting time is indicated with a circle for the basic gravity model or a triangle for the duration gravity model with vertical lines showing the 95% HPD intervals with the color of each caterpillar representing the log population density of that district.
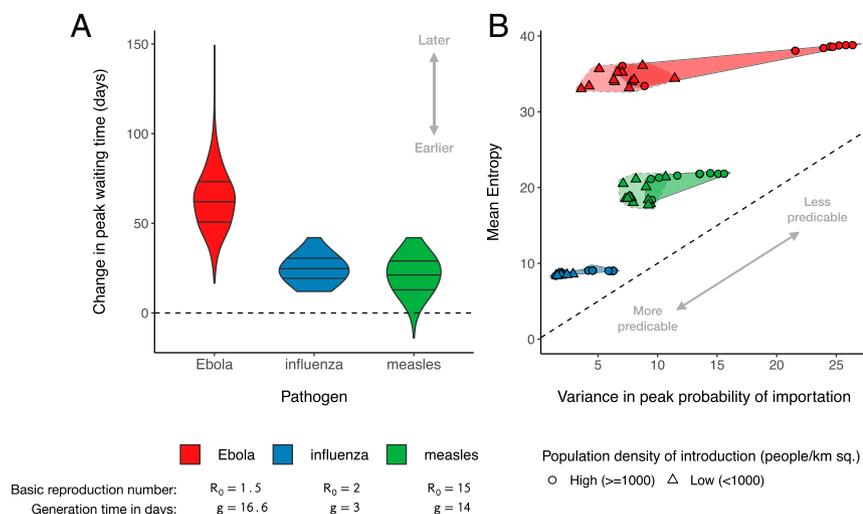
Fig. 5. Overall patterns in simulations of spatial transmission show a general delay in the rate of introduction when trip duration is included in the TSIR model. (*A*) Changes in peak waiting-time distributions for three pathogens (Ebola, influenza, and measles) when trip duration is included. (*B*) Relationship between the variance in peak introduction times and uncertainty around each peak for the 10 districts with the highest (circles) and lowest (triangles) population density. Darker shaded regions indicate high-density introduction events, where transmission to other locations is more spread out over time and the window of time in which each introduction is likely to occur is also longer. For example, simulations with Ebola (red) showed the largest variance in peak introduction times (i.e., introductions are very early or very late), but they also exhibited high levels of entropy in each waiting-time distribution (i.e., exact timing of each introduction event is less predictable). Sq., squared.

and socio-economic differences among countries and bias introduced by mobile phone ownership and coverage (12, 41). Future work must therefore validate the generality of the duration gravity model against other mobility datasets that include both the number of trips and duration of those trips and further explore model-selection studies to determine scenarios where duration information is essential.

We simulated spatial disease dynamics for a range of pathogens, but we only included differences in the generation time, transmission rate, recovery rate, and R_0 to illustrate the impact of duration on disease spread. We did not include differences in susceptibility, which inevitably would vary spatially, and further impact the waiting-time distributions between locations (31, 42). Overall, we find that epidemics that begin in high-density districts exhibit greater variability in peak introduction times and uncertainty around each peak (Fig. 5*B*), which is congruent with previous findings of Colizza et al. (43) that show decreased predictability in the initial stages of epidemics on heterogeneous networks when they begin in a travel hub. Predictability in the quantitative sense, however, has multiple dimensions in the context of epidemic interventions because the increased variance in introduction times allows us to distinguish which locations are at risk early in an epidemic, which would enable interventions to be prioritized and timed accordingly. There is also more uncertainty in the waiting-time distributions, which makes targeted interventions more challenging. In contrast, when the initial introduction event is in a low-density district, waiting-time distributions are more certain for each district. This certainty would help prioritizing and planning interventions in this scenario; however, many locations have the same introduction time that would require deploying more resources simultaneously. By simulating introduction events in both high- and low-density districts, we show that the location of the first introduction event ultimately impacts spatial spread. Therefore, in order to fully predict the waiting-time distributions necessary to plan public health interventions, the exact initial introduction district would be needed.

Although often excluded from spatial models of mobility and disease transmission, the duration of travel can be quantified

by using novel sources of data that can then be incorporated into these models. In particular, we find that including duration can improve model fit in places where traditional models typically perform poorly, such as those with low population sizes. As Balcan et al. (44) note, changes in coupling can impact initial stages of disease spread; similarly, our results show that accounting for trip duration induces differences in the waiting-time distribution until an introduction event occurs and the variance of those distributions, which impacts measures of uncertainty in early spatial spread. Recent work on symbolic entropy may provide further traction on how trip duration affects uncertainty in endemic settings (45) and potential codependencies among locations that drive predictable patterns of spatial spread (46). Since human mobility is a crucial component of predicting the spatial spread of many infectious diseases, models that have a more sophisticated representation of how individuals contribute to the transmission process in places to which they travel can help to elucidate the spatial and temporal dynamics of transmission and identify the best strategy for interventions, such as reactive vaccination campaigns and travel restrictions.

Materials and Methods

Population and Mobility Data. We analyzed CDRs from districts in Namibia from October 1, 2010, to April 29, 2014. For each of the ∼2.5 million subscribers in the dataset, a trip was counted if the most used mobile-phone tower was in a different district compared to the previous day; otherwise, the subscriber was classified as staying in the same location. For each observed trip, the date the trip began and the number of days the subscriber spent in the destination district (trip duration) was recorded. In total, over 259 million trips among 105 districts in Namibia were analyzed. Although we analyzed data spanning 4 y, the majority of variation in trip duration was among routes (spatial) rather than temporal; therefore, we temporally aggregated the data and analyzed the mean of the total monthly trip counts. Trip-duration data for each route were aggregated into 1-d intervals spanning 1,307 unique days in the 4-y period.

Preliminary models of trip duration indicated that at least 20 trips per route were required for adequate model convergence. We constructed a subsample of 62 of the 107 districts which contained a minimum of 20 unique observations of trip duration by sequentially removing the district with the lowest number of observations until routes among all districts contained a minimum of 20 observations. The remaining 62 districts were then

classified based on their population density as high ($>$1,000 people/km^2) or low ($\leq$1,000 people per km^2). Data of district-level population sizes was obtained from the WorldPop Project (https://www.worldpop.org), and population density was calculated as total population size divided by the square-kilometer area of the district.

Estimating Trip Duration with Exponential Decay. In order to incorporate trip duration into a route-level model of travel, we reduced the trip-duration data to a route-level summary statistic. Given the exponential distribution of trip-duration counts (*SI Appendix*, Fig. S3 and Table S1), we estimated the mean trip duration (in days) using an exponential-decay model and then used the decay-rate parameter λ_{ij} as a proxy for the mean trip duration of each route in subsequent models. To model the exponential decay in duration of stay for commuter trips across different routes of travel, we estimated $N_{\text{decay}}(y_{ij})$, which is the expected number of commuters making a trip of duration y when traveling from origin i to destination j. The model fits an exponential-decay function based upon the time spent y in destination j to observed counts of trip duration for each ij route:

$$N_{\text{decay}}(y_{ij}) = N_{ij}^0 e^{-\lambda_{ij} y_{ij}}. \quad [1]$$

We estimated λ_{ij} hierarchically at both the population and route levels to facilitate comparison of decay rates across different route types and compensate for routes that have lower sample sizes. The population-level hyperparameter λ' was given the uninformative prior of Unif(0,25) with route-level λ_{ij} parameters defined as the product of λ' and a scaling factor with the prior Gamma(2,1). The intercept term (N_{ij}^0) is the observed number of trips at $y = 0$ for each ij route (*SI Appendix*, Fig. S2).

Based upon preliminary models of trip duration, convergence of Markov chain Monte Carlo (MCMC) chains was poor for routes with a low number of observations, so we enforced a minimum sample size of 20 observations (or unique durations of travel). For the 62 districts that met the minimum sample size, we fit decay models to observed trip counts for each 1-d interval in the data, and we also explored aggregation of the trip duration data using 3- and 5-d intervals. While this shifted mean estimates of λ_{ij} upward, the relative proportions among route types remained the same, indicating that the model is robust to temporal aggregation (*SI Appendix*, Fig. S5).

Gravity Model Incorporating Both Trip Counts and Trip Duration. The gravity model is commonly used to relate covariates such as the population sizes of locations and the distances among them to the connectivity parameter which represents the proportion of trips or probability of movement from origin i to destination j (39, 47, 48). We fit the basic gravity-model formula to observed trip counts m_{ij} using a normalized connectivity parameter π_{ij} and a Poisson error structure:

$$m_{ij} = \text{Pois}(\pi_{ij} N_i)$$
$$\pi_{ij} \propto \theta \left(\frac{N_i^{\omega_1} N_j^{\omega_2}}{d_{ij}^\gamma} \right), \quad [2]$$

where the exponential parameters ω_1 and ω_2 are weights that scale the contribution of origin and destination population sizes to the numerator, γ controls how quickly the penalty on connectivity increases with distance, and θ is a proportionality constant.

Initial data exploration suggested that the distribution of trip duration may be dependent upon trip distance. Therefore, we developed a formulation of the gravity model that accounts for this interdependence by incorporating the trip-duration decay parameter λ_{ij} into the dispersal kernel so that the probability of movement to destination j also depends on the duration of stay at destination j:

$$\pi_{ij} \propto \theta \left(\frac{N_i^{\omega_1} N_j^{\omega_2}}{f(d_{ij} \mid \lambda_{ij})} \right). \quad [3]$$

The denominator of the gravity model, $f(d_{ij} \mid \lambda_{ij})$, is a conditional dispersal kernel, which we used Bayes' theorem to define as:

$$f(d_{ij} \mid \lambda_{ij}) = d_{ij}^\gamma \left(1 - \text{ECDF}(\lambda_{ij})^{\alpha_i} \right). \quad [4]$$

We refer to this updated formulation as the duration gravity model and compare it to the basic gravity model that is commonly used. When fitting these models to the trip-count data, we first fit the basic gravity model, which did not have any data or parameters associated with trip duration (*SI Appendix*). The basic gravity model used an uninformative prior

of Gamma(1,1) for θ, ω_1, ω_2, and γ parameters. We then fit the duration gravity model, which includes the conditional dispersal kernel, by using the posterior distributions of θ, ω_1, ω_2, and γ estimated in the basic gravity model as priors and a Gamma(1,1) prior for α_i (*SI Appendix*). Both gravity models where fitted to mean monthly trip counts by using a Poisson likelihood function.

TSIR Simulation with Mobility and Length of Stay. Spatial disease transmission was simulated by using a stochastic TSIR model (4, 28, 31, 49, 50). This TSIR framework is a spatial diffusion process that operates over metapopulations comprising the 62 Namibian districts analyzed in the trip-duration decay and gravity models. We used the estimated posterior distributions of the trip-duration decay parameter $\hat{\lambda}_{ij}$ and connectivity parameter $\hat{\pi}_{ij}$ to simulate human mobility and duration of stay. These terms drive local epidemic dynamics through the spatial force of infection, which is the expected number of new infections at location j and time step $t + 1$:

$$\mathbb{E}[I_{j,t+1}] = \frac{\beta S_{jt}(I_{jt} + \iota_{jt} + \kappa_{jt})^\alpha}{N_{jt}}. \quad [5]$$

The epidemic process relies on the movement of infected individuals to track spatial diffusion of the pathogen; therefore, the change in the susceptible population is defined as: $S_{j,t+1} = S_{jt} - I_{j,t+1}$. Parameters governing epidemic dynamics, such as transmission rate β and recovery rate γ, were parameterized for each pathogen, where $R_0 = \beta/\gamma$ and all t time steps were set to the generation time of the pathogen (see *SI Appendix*, Table S3 for parameter values). The exponent α was set to 0.97 to relax the mass-action assumption and allow for discrete-time approximation of the continuous transmission process (51).

In Eq. 5, the ι_{jt} term is a Poisson random variable with a mean equal to m_{jt}, which we define as the number of infected individuals migrating to destination j from all other locations at time step t. The ι_{jt} term is typically used to model the effect of transient infections that arrive in location j and remain for all of the tth epidemic generation. However, data on trip duration allow us to adjust the temporal contribution of infected individuals traveling along each ij route based on how long individuals typically remain in destination j.

$$\iota_{jt} = \text{Pois}(m_{jt})$$
$$m_{jt} = \sum_{\forall i \neq j} \left(\hat{\rho}_{ij} \hat{\pi}_{ij} \hat{\tau}_i I_{it} \right). \quad [6]$$

where m_{jt} is the mean number of infectious individuals immigrating to destination j at time step t scaled by three terms: the probability that an individual leaves district i ($\hat{\tau}_i$), the estimated probability of travel from i to j ($\hat{\pi}_{ij}$), and the probability that an individual remains in destination j for a full epidemic generation when traveling from i ($\hat{\rho}_{ij}$). Both $\hat{\tau}_i$ and $\hat{\rho}_{ij}$ were simulated by using Beta distributions fitted to the CDRs (*SI Appendix*).

In addition to the infectious individuals that visit district j in time t (ι_{jt}), there are also infectious individuals that remain in district j from previous time steps, which we include as κ_{jt} in the spatial force of infection. The κ_{jt} term is the number of infectious individuals that have traveled to district j in a previous time step and remain for a full epidemic generation after δ generations have passed (*SI Appendix*, Fig. S12). The summation over all previous time steps gives the estimated mean number of remnant infectious individuals due to previous immigration events r_{jt}.

$$\kappa_{jt} = \text{Pois}(r_{jt})$$
$$r_{jt} = \bar{\rho}_j \sum_{\delta=1}^{t} \left(\iota_{j,t-\delta} e^{-\delta \bar{\lambda}_j} \right). \quad [7]$$

Using this TSIR framework, we explored the impact of trip duration on the spatial dynamics of disease spread based on different scenarios of connectivity, pathogen life history, and location of disease introduction. Specifically, we compared the TSIR model, which includes both trip counts and trip duration in the gravity model and force of infection, to one that uses only trip counts (*SI Appendix*). For both TSIR model types, we further explored the spatial dynamics for six pathogens with different life histories (influenza, measles, Ebola, severe acute respiratory syndrome (SARS-CoV-1), pertussis, and malaria; *SI Appendix*, Table S3) and introduction of these pathogens at each of the 62 districts in the analysis. We ran each simulation scenario for 100,000 iterations and then—following Bjørnstad and Grenfell (49)—assessed spatial spread using the time-varying spatial-hazard function (*SI Appendix*). We then calculated the waiting-time distributions for each

district over all of the time steps in each simulation and summarized the probability of importation time over all simulations using a simple linear combination of all simulated realizations (52) (*SI Appendix*, Figure S13). We then calculated the peak of the aggregate probability of importation along with its 95% HPD intervals.

The trip-duration decay model and gravity models were fitted to data by using the JAGS (Just Another Gibbs Sampler) Bayesian MCMC algorithm and the "rjags" R package (53). Posterior parameter estimates were then used to simulate population mobility in disease-transmission simulations, which were written in R (54).

Data Availability. Code for data analyses, model fitting, and simulations is available in the "hmob" R package version 0.2.0 (55). The CDRs in this study were published with permission from MTC Mobile. Strict license agreement prohibits direct sharing of these data by the authors; however, other researchers can request these data from MTC Mobile independently. These data were deemed exempt from Institutional Review Board approval because they were deidentified by the mobile phone provider and aggregated within each cell phone tower catchment area prior to our analysis; informed consent was not required.

1. D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
2. M. C. González, C. A. Hidalgo, A.-L. Barabási, Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
3. C. Viboud et al., Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451 (2006).
4. B. T. Grenfell, O. N. Bjørnstad, B. F. Finkenstädt, Dynamics of measles epidemics: Scaling noise, determinism, and predictability with the TSIR model. *Ecol. Monogr.* **72**, 185–202 (2002).
5. A. M. Kramer et al., Spatial spread of the West Africa Ebola epidemic. *R. Soc. Open Sci.* **3**, 160294 (2016).
6. A. Wesolowski et al., Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11114–11119 (2015).
7. A. Wesolowski et al., Quantifying the impact of human mobility on malaria. *Science* **338**, 267–270 (2012).
8. D. Leon, T. House, M. J. Keeling, The role of routine versus random movements on the spread of disease in Great Britain. *Epidemics* **1**, 250–258 (2009).
9. M. S. Y. Lau et al., Spatial and temporal dynamics of superspreading events in the 2014–2015 West Africa Ebola epidemic. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 2337–2342 (2017).
10. A. J. Tatem, D. J. Rogers, S. I. Hay, "Global transport networks and infectious disease spread" in *Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications*, S. I. Hay, A. Graham, D. J. Rogers, Eds. (Advances in Parasitology, Academic Press, Amsterdam, Netherlands, 2006). vol. 62, pp. 293–343.
11. E. C. Lee et al., Mind the scales: Harnessing spatial big data for infectious disease surveillance and inference. *J. Infect. Dis.* **214** (suppl. 4), S409–S413 (2016).
12. A. Wesolowski, C. O. Buckee, K. Eng-Monsen, C. J. E. Metcalf, Connecting mobility to infectious diseases: The promise and limits of mobile phone data. *J. Infect. Dis.* **214** (suppl. 4), S414–S420 (2016).
13. A. Wesolowski et al., The use of census migration data to approximate human movement patterns across temporal scales. *PloS One* **8** (1), e52971 (2013).
14. A. Wesolowski et al., Quantifying travel behavior for infectious disease research: A comparison of data from surveys and mobile phones. *Sci. Rep.* **4**, 5678 (2014).
15. J. M. Marshall et al., Key traveller groups of relevance to spatial malaria transmission: A survey of movement patterns in four sub-Saharan African countries. *Malar. J.* **15**, 200 (2016).
16. S. F. A. Batista, L. Leclercq, N. Geroliminis, Estimation of regional trip length distributions for the calibration of the aggregated network traffic models. *Transp. Res. Part B Methodol.* **122**, 192–217 (2019).
17. J. Pawlak, J. W. Polak, A. Sivakumar, A framework for joint modelling of activity choice, duration, and productivity while travelling. *Transp. Res. Part B Methodol.* **106**, 153–172 (2017).
18. M. J. Keeling, P. Rohani, Estimating spatial coupling in epidemiological systems: A mechanistic approach. *Ecol. Lett.* **5**, 20–29 (2002).
19. US Department of Transportation. National Household Travel Survey (2017). https://nhts.ornl.gov/. Accessed 11 December 2019.
20. US Census Bureau, Commuting (Journey to Work) (2016). https://www.census.gov/topics/employment/commuting.html. Accessed 11 December 2019.
21. M. C. Schneider, V. Belik, T. Couronn, Z. Smoreda, M. C. Gonzlez, Unravelling daily human mobility motifs. *J. R. Soc. Interface* **10**, 20130246 (2013).
22. G. M. Vazquez-Prokopec et al., Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. *PLoS One* **8**, e58802 (2013).
23. H. Barbosa et al., Human mobility: Models and applications. *Phys. Rep.* **734**, 1–74 (2018).
24. N. W. Ruktanonchai et al., Identifying malaria transmission foci for elimination using human mobility data. *PLoS Comput. Biol.* **12**, e1004846 (2016).
25. M. U. G. Kraemer et al., Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. *Sci. Rep.* **9**, 5151 (2019).
26. A. Wesolowski et al., Multinational patterns of seasonal asymmetry in human movement influence infectious disease dynamics. *Nat. Commun.* **8**, 2069 (2017).
27. C. J. E. Metcalf, C. V. Munayco, G. Chowell, B. T. Grenfell, O. N. Bjørnstad, Rubella metapopulation dynamics and importance of spatial coupling to the risk of congenital rubella syndrome in Peru. *J. R. Soc. Interface* **8**, 369–376 (2011).
28. O. N. Bjornstad, B. F. Finkenstadt, B. T. Grenfell. Dynamics of measles epidemics: Estimating scaling of transmission rates using a time series SIR model. *Ecol. Monographs* **72**, 169–184 (2002).
29. L. Bengtsson et al., Using mobile phone data to predict the spatial spread of cholera. *Sci. Rep.* **5**, 8923 (2015).
30. R. F. Grais, X. De Radigues, C. Dubray, F. Fermon, P. J. Guerin, Exploring the time to intervene with a reactive mass vaccination campaign in measles epidemics. *Epidemiol. Infect.* **134**, 845–849 (2006).
31. C. J. E. Metcalf et al., Implications of spatially heterogeneous vaccination coverage for the risk of congenital rubella syndrome in South Africa. *J. R. Soc. Interface* **10**, 20120756 (2013).
32. A. Wesolowski et al., Measles outbreak risk in Pakistan: Exploring the potential of combining vaccination coverage and incidence data with novel data-streams to strengthen control. *Epidemiol. Infect.* **146**, 1575–1583 (2018).
33. I. Casas, E. Delmelle, E. C. Delmelle, Potential versus revealed access to care during a dengue fever outbreak. *J. Transp. Health* **4**, 18–29 (2017).
34. J. R. Mahler, F. K. Kombe. "Context and ethical challenges during the Ebola outbreak in West Africa" in *Socio-Cultural Dimensions of Emerging Infectious Diseases in Africa: An Indigenous Response to Deadly Epidemics*, G. B. Tangwa, A. Abayomi, S. J. Ujewe, N. S. Munung, Eds. (Springer International Publishing, Cham, Switzerland, 2019), pp. 191–202.
35. J. Lessler et al., Mapping the burden of cholera in sub-Saharan Africa and implications for control: An analysis of data across geographical scales. *Lancet* **391**, 1908–1915 (2018).
36. T. D. Hollingsworth, N. M. Ferguson, R. M. Anderson, Will travel restrictions control the international spread of pandemic influenza? *Nat. Med.* **12**, 497–499 (2006).
37. I. I. Bogoch et al., Assessment of the potential for international dissemination of Ebola virus via commercial air travel during the 2014 West African outbreak. *Lancet*, **385**, 29–35 (2015).
38. A. Wesolowski, W. P. O'Meara, E. Nathan, A. J. Tatem, C. O. Buckee, Evaluating spatial interaction models for regional mobility in sub-Saharan Africa. *PLoS Comput. Biol.* **11**, e1004267 (2015).
39. T. James, N. M. Ferguson, Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling. *PLoS Comput. Biol.* **8**, e1002699 (2012).
40. F. Simini, M. C. Gonzlez, A. Maritan, A.-L. Barabsi, A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).
41. A. Wesolowski, E. Nathan, A. M. Noor, R. W. Snow, C. O. Buckee, The impact of biases in mobile phone ownership on estimates of human mobility. *J. R. Soc. Interface* **10**, 20120986 (2013).
42. J. Lessler, W. J. Moss, S. A. Lowther, D. A. T. Cummings. Maintaining high rates of measles immunization in Africa. *Epidemiol. Infect.*, **139**, 1039–1049 (2011).
43. V. Colizza, A. Barrat, M. Barthlemy, A. Vespignani, The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2015–2020 (2006).
44. D. Balcan et al., Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21484–21489 (2009).
45. S. V. Scarpino, G. Petri, On the predictability of infectious disease outbreaks *Nature Comm.* **10**, 898 (2019).
46. S. M. Kissler, C. Viboud, B. T. Grenfell, J. R. Gog, Symbolic transfer entropy reveals the age structure of pandemic influenza transmission from high-volume influenza-like illness data. *J. R. Soc. Interface* **17**, 20190628 (2020).
47. J. M. Marshall et al., Mathematical models of human mobility of relevance to malaria transmission in Africa. *Sci. Rep.* **8**, 7713 (2018).
48. Y. Xia, O. N. Bjrnstad, B. T. Grenfell, Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics. *Am. Nat.* **164**, 267–281 (2004).
49. O. N. Bjørnstad, B. T. Grenfell, Hazards, spatial transmission and timing of outbreaks in epidemic metapopulations. *Environ. Ecol. Stat.* **15**, 265–277 (2008).
50. B. F. Finkenstädt, B. T. Grenfell, Time series modelling of childhood diseases: A dynamical systems approach. *J. Roy. Stat. Soc. C Appl. Stat.* **49**, 187–205 (2000).
51. K. Glass, Y. Xia, B. T. Grenfell, Interpreting time-series analyses for continuous-time biological models—Measles as a case study. *J. Theor. Biol.* **223**, 19–25 (2003).
52. R. T. Clemen, R. L. Winkler, Combining probability distributions from experts in risk analysis. *Risk Anal.* **1**, 187–203 (1999).
53. M. Plummer, *rjags: Bayesian Graphical Models using MCMC* (R Package Version 4-8, The Comprehensive R Archive Network, 2018). https://cran.r-project.org/web/packages/rjags/index.html. Accessed 20 July 2017.
54. R Core Team, *R: A Language and Environment for Statistical Computing* (Version: 3.6.2, R Foundation for Statistical Computing, Vienna, Austria, 2018). https://www.R-project.org/. Accessed 12 December 2019.
55. J. R. Giles, gilesjohnr/hmob: hmob R package (Version v0.2.0). Zenodo. http://doi.org/10.5281/zenodo.3974977. Deposited 6 August 2020.

POPULATION BIOLOGY