Comparing metapopulation dynamics of infectious diseases under different models of human movement

Daniel T. Citron1,2, Carlos A. Guerra3,4, Andrew J. Dolgert3,4, Sean L. Wu3,5, John M. Henry3, Héctor M. Sánchez C.6,7, and David L. Smith8,9

1Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA 98121; 2Medical Care Development International, Silver Spring, MD 20910; and 3Division of Biostatistics & Epidemiology, University of California, Berkeley, CA 94720

Edited by Alan Hastings, University of California, Davis, CA, and approved March 8, 2021 (received for review April 21, 2020)

Newly available datasets present exciting opportunities to investigate how human population movement contributes to the spread of infectious diseases across large geographical distances. It is now possible to construct realistic models of infectious disease dynamics for the purposes of understanding global-scale epidemics. Nevertheless, a remaining unanswered question is how best to leverage the new data to parameterize models of movement, and whether one’s choice of movement model impacts modeled disease outcomes. We adapt three well-studied models of infectious disease dynamics, the susceptible–infected–recovered model, the susceptible–infected–susceptible model, and the Ross–Macdonald model, to incorporate either of two candidate movement models. We describe the effect that the choice of movement model has on each disease model’s results, finding that in all cases, there are parameter regimes where choosing one movement model instead of another has a profound impact on epidemiological outcomes. We further demonstrate the importance of choosing an appropriate movement model using the applied case of malaria transmission and importation on Bioko Island, Equatorial Guinea, finding that one model produces intelligible predictions of $R_0$, whereas the other produces nonsensical results.

Mathematical models are important tools for understanding disease transmission, making it possible to estimate the size, timing, and impact of epidemics and the effectiveness of interventions. There have been many such mathematical models adapted and applied to study epidemics that spread across large geographical regions (1, 2). Many such models are often motivated by recent pandemics, such as severe acute respiratory syndrome (3), Ebola virus disease (4), Zika fever (5), and the 2019 novel coronavirus (6). In each of these cases, long-distance travel of infected human hosts proved to be an important driver of the spread of infectious disease between geographically separated populations. Real-world studies of the spatial dynamics of human infectious diseases must choose an underlying modeling framework for simulating human mobility which is then parameterized with mobility data and coupled with a model of disease dynamics. Here, we have conducted a study that focuses on the consequences which the choice of movement modeling framework can have on various disease dynamical processes.

Parameterizing realistic models of disease transmission occurring across wide geographical ranges is now possible because of the recent availability of large, highly detailed datasets describing human movement patterns (7). Census data attempt to describe population migration that occurs between census years (8). Traditional surveys of commuters (9, 10), patients (11, 12), or residents in disease-affected areas (13, 14) provide another description of recent travel activity. More recently, mobile phone service providers have shared privacy-protected datasets, showing how large numbers of users tend to move between cell phone towers (12, 15–18). Global Positioning System (GPS) trip logs have also recently been used to track study participants’ movement activities (19).

The recent abundance of movement data provides an exciting new opportunity to use models to quantify how human host movement affects epidemiological outcomes. Nevertheless, each method of data collected only provides a partial picture of true movement patterns, and for many studies it is necessary to build a model of host movement to fill gaps in the data or to synthesize data coming from many sources. It remains an unanswered question how best to use the data to build and parameterize models representing host movement.

How should one choose to represent movement rates based on available data? An important consideration is finding a statistical model that fits adequately well to movement data, accurately predicting the frequency of travel or flow volume between two locations. There are already many candidate movement models, such as gravity and radiation models, which have been evaluated against data in a variety of different settings (13, 18, 20–22). A fitted movement model, once it has been evaluated as sufficiently accurate, may then be used to set the movement-related parameters of a mechanistic disease transmission model.

Criteria for selecting a model to parameterize and whether the choice of model would affect the conclusions of a study has received much less attention (9, 23). For the purposes of modeling movement to understand infectious disease dynamics, the

Significance

Newly available large-scale datasets of human population movement represent an opportunity to model how diseases spread between different locations. Combining infectious disease models with mechanistic models of host movement enables studies of how movement drives disease transmission and importation. Here we explore how modeled epidemiological outcomes may be sensitive to the modeler’s choice of movement model structure. We use three different mathematical models of disease transmission to show how a model’s epidemiological predictions can change dramatically depending on the chosen host movement model. We find these different outcomes are robust to using the same data sources to parameterize each candidate model, which we illustrate using an example of real-world malaria transmission and importation in Bioko Island, Equatorial Guinea.

Author contributions: D.T.C. and D.L.S. designed research; D.T.C., C.A.G., A.J.D., S.L.W., J.M.H., H.M.S.C., and D.L.S. performed research; D.T.C. and C.A.G. analyzed data; and D.T.C., A.J.D., and D.L.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

1To whom correspondence may be addressed. Email: dtcitron@uw.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2007488118/-/DCSupplemental.

Published April 29, 2021.

PNAS 2021 Vol. 118 No. 18 e2007488118

https://doi.org/10.1073/pnas.2007488118
structure of the movement model—the rules which govern how hosts’ movement patterns are represented—affect the model’s quantitative behavior and subsequent predicted epidemiological outcomes. The same travel datasets used to calibrate different movement models can have different disease dynamics as a consequence of the movement model choice (9). One’s choice of movement model structure defines the set of parameters which require fitting and remains distinct from the question of how best to fit those parameters.

There is an important structural difference between two of the simplest and most commonly used classes of movement models. Eulerian movement models specify the rates at which hosts from one location travel to any other location. Eulerian models do not track individual behavior, such that after a host moves it has no memory of its previous location. Lagrangian movement models, by way of contrast, follow individuals. The class of Lagrangian movement models includes any model that specifies how frequently hosts travel away from home before returning (23, 24). The authors of ref. 24 formulate models of vector-borne disease with Eulerian and Lagrangian host movement but they do not directly compare the models’ quantitative behavior, inviting further questions of when one host movement model might be more appropriate than another. Depending on the level of detail contained in the available data, one might be able to use the data to select a class of candidate mechanistic movement models, but the challenge here is that it provides some useful information about which model might be more appropriate (23). Modelers whose aim is to understand how movement influences epidemiological outcomes must choose one such mechanistic description to represent movement.

More generally, there has never been a thorough exploration of how the modeler’s choice of movement model may influence modeled epidemiological outcomes. Most early discussions of metapopulation modeling (e.g., ref. 25) primarily focused on formulating the models, without discussing how one might make use of travel data for movement model calibration or how to select a movement model depending on epidemiological context. More recently, the authors of ref. 26 offer the heuristic advice that a Eulerian movement model is suited to describing animal migration while a Lagrangian movement model is more suited to describing human commuting behavior, but they do not provide supporting quantitative analysis. At least one study has used the same travel datasets to calibrate different movement models only to find that the disease dynamics can change as a consequence of the movement model choice (9). In some cases, it may be possible to derive and compare simpler models of disease transmission by carefully rescaling movement parameters to capture the essential features of spatial dynamics in a generalized metapopulation (27). Generally speaking, however, there has not yet been a thorough quantitative investigation into whether the modeler is free to substitute one model for another for a given epidemiological setting without affecting the model’s results and conclusions.

When it comes to incorporating movement data into an epidemiological model, the problem becomes whether choosing one mechanistic movement model over another can impact modeled epidemiological outcomes, and whether there are certain epidemiological settings where one type of movement model may be more appropriate or accurate. In the present analysis we explore this problem using compartmental metapopulation models which integrate together host movement and infectious disease dynamics. We use three infectious disease transmission models—the susceptible–infected–recovered (SIR) model, the susceptible–infected–susceptible (SIS) model, and the Ross–Macdonald model—which represent a suite of tools for modeling the transmission dynamics of a wide variety of pathogens. We adapt each of the disease transmission models to incorporate two mechanistic representations of movement: The first is a Eulerian movement model which we call the Flux model, and the second is a Lagrangian movement model which we call the Simple Trip model. We directly compare the Flux and Simple Trip models by setting parameters such that the total flux of travelers between each subpopulation remains constant, thus emulating a case where one might have a single dataset from which one could calibrate either movement model. We examine how disease model outcomes can change with different movement models for each transmission model. For all three transmission models, we find that the modeled quantities of interest relating to disease dynamics can differ dramatically depending on one’s choice of movement model. We conclude by applying this analysis to a model of malaria transmission and importation on Bioko Island in Equatorial Guinea (EG) and use intuition based on our prior analysis to frame our understanding for why the Flux model surprisingly fails to produce meaningful predictions of transmission intensity in that context.

**Methods**

**Host Movement Models.** We use metapopulation models to represent a network of geographically isolated populations of hosts. Each population occupies a location where local conditions may affect transmission intensity. Each population contains N_i hosts. We assume that the population at each location remains stable over time—in applied contexts each population may be calibrated using census or other population data. Disease transmission is assumed to be completely local, such that hosts from different subpopulations come into contact with one another only if they travel to occupy the same location. We describe and compare two simple models of host movement, each of which represents a different set of rules governing how hosts move from one location to another.

The Flux model is a Eulerian movement model which describes hosts as diffusing from one metapopulation to another (24):

\[
\frac{dN_i}{dt} = -\sum_{j=1}^{K} f_{ij} N_j + \sum_{j=1}^{K} f_{ji} N_i,
\]

where N_i counts the number of hosts currently located at site i. The total number of hosts remains constant over time (N = \sum_{i=1}^{K} N_i). The constant f_{ij} represents the rate at which hosts located at i travel to j, f_{ji} = 0 for all i. The fully specified Flux model requires (K(K - 1)) parameters.

The Simple Trip model is a Lagrangian movement model which assigns home locations to each host and describes how hosts travel temporarily to other locations before returning home (23, 24). Unlike the Flux model, the Simple Trip model differentiates between the residents and the visitors currently located at the same site. Visitors are able to interact with residents at their home sites: A host from location i will be counted as belonging to the N_i population, and the number of hosts whose home is i remains constant over time, even if members visit other sites (N_i = \sum_{j=1}^{K} N_{ij}). The constant \phi_{ij} represents the rate at which hosts whose home is i travel to j, while the constant \tau_{ij} is the rate at which hosts visiting j return home to i. Both \phi_{ij} = 0 and \tau_{ij} = 0 for all i. The fully specified Simple Trip model requires 2(K(K - 1)) parameters.

When the movement equations reach a steady state and the derivatives return home to site i, the notation in Eq. 2 has changed to account for hosts retaining memory of their home site:

\[
\frac{dN_{ij}}{dt} = -\sum_{j=1}^{K} \phi_{ij} N_{ij} + \sum_{j=1}^{K} \tau_{ij} N_{ij},
\]

\[
\frac{dN_{ij}}{dt} = -\tau_{ij} N_{ij} + \phi_{ij} N_{ij},
\]

The notation in Eq. 2 has changed to account for hosts retaining memory of their home sites: A host from i who is currently located at j will be counted as belonging to the N_{ij} population, and the number of hosts whose home is i remains constant over time, even if members visit other sites (N_{ij} = \sum_{k=1}^{K} N_{ik}, N_{ik}).

\[
N_{ij} = \frac{1}{\tau_{ij} + \sum_{j=1}^{K} \phi_{ij}},
\]

\[
N_{ij} = \frac{1}{\tau_{ij} + \sum_{j=1}^{K} \frac{\phi_{ij} N_{ij}}{N_{ij}}},
\]

We can use Eq. 3 to define a “time at risk matrix” ν where \psi_{ij} = N_{ij}/N_i, the average number of individuals from i found at site j in the steady state, or the fraction of time on average that hosts who live in site i spend in site j.
To directly compare the quantitative behavior of both models, we will assume that in the steady state the total number of hosts moving from one site to another is the same across both models. That is to say, the number of people traveling $i \to j$ in the Flux model is the same as the number of $j$ residents traveling to $i$ in the Simple Trip model:

$$f_{ij} = \frac{\phi_{ij} N^i_j/N_j + \phi_{ji} N^j_i/N_i}{1 + \sum_{k=1}^K \frac{\phi_{ik}}{\phi_{jk}}}$$

It can be shown under these circumstances that the steady-state populations in the Flux model $N^i_j$ equal the stable populations in the Simple Trip model. Another property of Eq. 4 is that when we constrain the Flux model parameters to match the fluxes in the Simple Trip model, the Flux model parameters become symmetric $f_{ij} = f_{ji}$. Under this condition the fully specified Flux model only requires $K(K-1)/2$ parameters, meaning that more detailed information is required to parameterize the Simple Trip model than the Flux model.

SIR Models. The next step is to incorporate both movement models into three compartmental models of disease transmission. The first such model is the SIR model, which describes a single outbreak of disease. For a single population, the deterministic ordinary differential equation SIR model is as follows:

$$\frac{dS}{dt} = -\beta SL/N$$
$$\frac{dI}{dt} = \beta LI/N - I\gamma$$
$$\frac{dR}{dt} = I\gamma$$

Combining the basic SIR model with the Flux movement model 1, we obtain an analogous set of 3$K$ equations (23):

$$\frac{dS_i}{dt} = -\beta \frac{S_i}{N} \sum_{j=1}^K f_{ij} S_j - \sum_{j=1}^K f_{ij} S_i$$
$$\frac{dI_i}{dt} = \beta \frac{I_i}{N} - \gamma I_i - \sum_{j=1}^K f_{ij} I_j$$
$$\frac{dR_i}{dt} = \gamma I_i - \sum_{j=1}^K f_{ij} R_j$$

The parameters $\beta$ and $\gamma$ in Eq. 6 represent the transmission rate and recovery rate, respectively.

Combining the basic SIR model with the Simple Trip model 2, we obtain an analogous set of 3$K$ equations describing transmission between hosts while they are at their home location or while they are traveling:

$$\frac{dS_i}{dt} = -\beta \frac{S_i}{N} \sum_{j=1}^K f_{ij} S_j - \sum_{j=1}^K f_{ij} S_i$$
$$\frac{dI_i}{dt} = \beta \frac{I_i}{N} - \gamma I_i - \sum_{j=1}^K f_{ij} I_j$$
$$\frac{dR_i}{dt} = \gamma I_i - \sum_{j=1}^K f_{ij} R_j$$

Note that for both movement models we allow the parameter representing transmission intensity $\beta$ to vary across locations, not across individuals, representing how transmission intensity reflects the contact network structure or other local conditions affecting transmission. $\gamma$, by contrast, is held constant for all individuals regardless of location.

SIS Models. Our second disease transmission model is the SIS model, which can be used to describe persistent endemic disease brought about by cycles of repeated infection. We adapt the SIS model to include either the Flux or Simple Trip models to describe movement between subpopulations. For a single population, the SIS model equation is as follows:

$$\frac{dI}{dt} = \beta I(N-I) - I\gamma - I.$$

Combining the basic SIS model with the Flux model (Eq. 1), we obtain a set of $K$ equations for transmission in the $i$th site:

$$\frac{dI_i}{dt} = \beta \frac{I_i}{N} - \gamma I_i - \sum_{j=1}^K f_{ij} I_j$$

Again, in both movement models the transmission intensity $\beta$ varies across locations, not across individuals, and $\gamma$ is held constant for all individuals in all locations.

Ross–Macdonald Models. Our third disease model is the Ross–Macdonald model, which was developed to describe malaria transmission (28, 29). The Ross–Macdonald model is similar to the SIS model, except it includes a vector population which serves as a mechanism of transmission from infected to susceptible human hosts. The Ross–Macdonald model tracks infection dynamics among human hosts ($X$) and vectors ($Z$) and depends on bionomic parameters ($M$, $a$, $b$, $c$, $g$, $n$; see SI Appendix, section 3 for definitions) which reflect the environment’s support of the vectors which transmit disease.

The basic Ross–Macdonald model is as follows:

$$\frac{dX}{dt} = -baZ(N-X) - rX,$$
$$\frac{dZ}{dt} = acM(e^{-gn} - Z) - gZ.$$

The first of Eqs. 11 describes how human hosts become infected through contact with infectious mosquitoes, while the second describes how mosquitoes become infected through contact with infectious humans. Setting the derivatives in Eq. 11 to zero represents an endemic equilibrium in which the local environment may support endemic malaria under the right conditions. Standard fixed-point analysis of the endemic equilibrium yields an expression for $R_0$ as a function of bionomic parameters. As long as the endemic prevalence $X^* > 0$ it is also possible to reformulate $R_0$ as a function of prevalence (28, 29):

$$R_0 = \frac{rX}{N} \left(1 + \frac{r}{1 + \frac{r}{b}} \right)$$

As before, we adapt the basic Ross–Macdonald model by allowing human hosts to move from one metapopulation to another. At present we will hold the parameters ($r$, $a$, $b$, $c$, $g$, $n$) to be constant across all sites, while the stable mosquito population supported by the local environment ($M$) is allowed to vary across all sites. Incorporating the Flux model into the metapopulation Ross–Macdonald model (24),

$$\frac{dX_i}{dt} = (N_i - X_i)baZ_i/NX_i - rX_i - \sum_{j=1}^K f_{ij} X_i$$

$$\frac{dZ_i}{dt} = acM(e^{-gn} - Z_i) - gZ_i$$

We again perform a standard analysis of the endemic equilibrium of Eq. 13 and obtain an expression for the $R_{0(i)}^{(1)}$ in each site as a function of prevalence at each site $X_i$:

$$R_{0(i)}^{(1)} = \frac{rX_i + \sum_{j=1}^K f_{ij} X_i - \sum_{j=1}^K f_{ij} X_j}{N_i - X_i} \left(1 + \frac{r}{b} \right)$$
We also extend the basic Ross–Macdonald model by allowing human hosts to move from one metapopulation to another as described by the Simple Trip model. Following the analysis of refs. 17, 24, and 30, we simplify the analysis of the equilibrium behavior of the metapopulation Ross–Macdonald model using the time-at-risk matrix \( \psi \). We rewrite the force of infection experienced by human hosts as a weighted average which combines local transmission and the average fraction time spent at risk in different locations:

\[
\frac{dX_i}{dt} = (N_i - X_i) \sum_{j=1}^{k} \psi_{ij} \frac{Z_j}{\sum_{k=1}^{K} \psi_{ik} N_k} - r X_i
\]

\[
\frac{dZ_i}{dt} = a c_i N_i e^{-\theta_0} Z_i - g Z_i
\]

The expression \( \kappa_j \) represents the fraction of infectious humans located at \( j \), accounting for both local residents as well as visitors from other locations. From the endemic equilibrium of Eq. 15, and following the analysis of ref. 17, we obtain an expression for \( R_0^{(i)} \), which reflects how the movement model alters transmission at each metapopulation site:

\[
R_0^{(i)} = \sum_{k=1}^{K} \left( \psi_{ik} \right)^{1/k} \frac{r X_i}{N_k - X_k} \left[ 1 + \frac{\tau_{ij}}{\kappa_j} \right].
\]

Results

SIR Models. We begin by directly comparing the predicted outcomes of the two versions of the SIR model with Flux and Simple Trip movement. We simulate epidemics in two coupled metapopulations of \( N_i = 500 \) human hosts by numerically integrating Eqs. 6 and 7 and measuring the size of the outbreak. The population of susceptibles remaining after the outbreak has ended \( S_\infty \) quantifies the size of the outbreak and allows one to infer \( R_0 \) (26). Furthermore, \( S_\infty \) quantifies the population at risk in subsequent future epidemics, which is relevant for modeling seasonally recurring pathogens such as influenza (31).

We set \( \gamma = 1 \) in all locations, hold \( \beta_j \) constant, and measure the residual population \( S_\infty \) in location 1 as \( \beta_j \) varies. We identify two transmission parameter regimes: \( \beta_j > \gamma \), which is high enough to cause an outbreak in location 2, and \( \beta_j < \gamma \), which in isolation does not lead to an outbreak in location 2. We also define two travel parameter regimes, one where travel occurs frequently and lasts short periods of time (\( \tau_{ij} > \gamma \)) and one where travel occurs infrequently and lasts long periods of time (\( \tau_{ij} < \gamma \)). We specify the parameter values for \( \phi_{ij} \) and \( \tau_{ij} \) in each of these regimes for the Simple Trip model and use Eq. 4 to also parameterize the Flux model, such that we emulate using the same data inputs to parameterize both movement models.

Fig. 1 compares how the two SIR models with metapopulation movement behave in different parameter regimes. In all cases the outcomes are dramatically different, showing that SIR model predictions depend strongly on how host movement is represented in the metapopulation model. The two movement models do agree when transmission parameters are equal in both locations (\( \beta_1 = \beta_2 \)), but there are large quantitative differences whenever there is some heterogeneity in the local transmission conditions across the different locations.

The major difference between the Flux and Simple Trip models is that the Simple Trip model constrains the amount of time that a traveler spends away. For example, the larger residual population for the Simple Trip model, seen in the right-hand column of Fig. 1, follows from a limit on the amount of time that residents of population 1 spend in the high-transmission location 2. Another difference between the Flux and the Simple Trip models is that hosts in the Flux model have no home residence and continue to move freely between populations after the outbreak ends. As a result, as the epidemic dies out the fractions of susceptibles become constant across all locations. In contrast with the Simple Trip model, the Flux model effectively erases any evidence that the transmission intensity differed between the two locations.

**Fig. 1.** Comparing SIR model results. We numerically integrate Eqs. 6 and 7 to find the size of the residual population of susceptibles following the outbreak. We plot the residual population fraction in location 1 (\( S_\infty / N_1 \)) as a function of transmission intensity in location 2 (\( \beta_j / \gamma \)) constant. We explore four different parameter regimes defined by the transmission intensity of location 2 and the duration of travel. In all four parameter regimes there is a dramatic separation between the predicted outbreak sizes that depends on whether one uses the Flux (red) or the Simple Trip (blue) model. As long as residents of location 1 travel from a low- to a high-transmission environment (\( \beta_j < \beta_1 \)) the Simple Trip model’s outbreak will be smaller and the residual population will be larger; the opposite is true when residents of location 1 travel from a high- to a low-transmission environment (\( \beta_j > \beta_1 \)). (A) Short travel duration regime (\( \phi_{ij} = 1, \tau_{ij} = 10 \)); low transmission in location 2 (\( \beta_2 = 0.5 \)). (B) Short travel duration regime (\( \phi_{ij} = 1, \tau_{ij} = 10 \)); high transmission in location 2 (\( \beta_2 = 2.5 \)). (C) Long travel duration regime (\( \phi_{ij} = 0.01, \tau_{ij} = 0.1 \)); low transmission in location 2 (\( \beta_2 = 0.5 \)). (D) Long travel duration regime (\( \phi_{ij} = 0.01, \tau_{ij} = 0.1 \)); high transmission in location 2 (\( \beta_2 = 2.5 \)).
SIS Models. We next compare the respective behavior of the SIS Flux metapopulation model (Eq. 9) and the SIS Simple Trip metapopulation model (Eq. 10) for $K = 2$ metapopulations of $N_i = 500$ coupled together. Again, we set $\gamma = 1$ at all locations. The outcome of interest for the SIS model is now prevalence, which we find by numerically solving for the endemic equilibria in Eqs. 9 and 10. We hold $\beta_2$ constant and solve for the prevalence in location 1 ($I_1/N_i$ for the Flux model and $(I_{1,1} + I_{1,2})/N_1$ for the Simple Trip model) as a function of the transmission intensity in location 1 ($\beta_1$).

We define two transmission parameter regimes: $\beta_2 > \gamma$, which is high enough to sustain endemic transmission locally in location 2, and $\beta_2 < \gamma$, which cannot locally sustain endemic transmission. We again define two travel parameter regimes, one where travel occurs frequently and over short periods of time ($\tau_{i,j} > \gamma$) and one where travel occurs infrequently and over long periods of time ($\tau_{i,j} < \gamma$). In each regime we specify travel rate parameter values for the Simple Trip model and use Eq. 4 to set parameters for the Flux model.

Fig. 2 shows the results from both the Simple Trip and Flux models in each of the four parameter regimes. In the short travel duration regime (Fig. 2 A and B), as long as the transmission intensity differs between the two locations there is a noticeable separation between the two movement models’ predicted relationships between prevalence and local transmission intensity. In Fig. 2 A it is even the case that the apparent critical point, above which location 1 has sustained endemic disease, changes depending on which movement model is used.

The discrepancy in the short travel duration regime arises because the Simple Trip model imposes an extra constraint on the amount of time a traveler spends away from home. The Flux model does not specify the amount of time spent away, and its amount of time a traveler spends away from home. The Flux (red) and Simple Trip (blue) models produce dramatically different relationships between transmission intensity and prevalence. When travel duration is long, however, the two models’ predicted relationships between prevalence and $R_0$ do approach one another. The relationships between prevalence and transmission intensity for the two Ross–Macdonald models in each of the four parameter regimes defined by the transmission intensity of location 2 and the duration of travel. When travel duration is short at risk in the high transmission location 2 is constrained to be smaller. In the long travel duration regime (Fig. 2, Lower), however, if the trip time is longer than the disease infectious period there is enough time for prevalence among travelers to become equal to the prevalence among residents of location 2. In this parameter regime, the Flux model and Simple Trip models converge, because the constraint on time spent away from home makes less of a difference.

Ross–Macdonald Models. We examine the behavior of the Ross–Macdonald models with Flux and Simple Trip movement. We use the model to represent a malaria-causing pathogen with a mean infectious period of $r^{-1} = 200$ d. The outcome of interest in the Ross–Macdonald model is $R_0$, which quantifies the local transmission intensity required to sustain endemic malaria at observed prevalence levels. We compare the two movement models using $K = 2$ connected subpopulations of $N_i = 500$ human hosts by specifying prevalence and using Eqs. 14 and 16 to calculate $R_0$ for the Flux and Simple Trip models, respectively. Again, we define two parameter regimes based on high ($X = 0.25$) versus low ($X = 0.025$) prevalence in location 2 and another pair of parameter regimes based on high-frequency, short-duration trips ($\phi_i = 1/360$, $\tau_{i,j} = 1/10$) versus low-frequency, long-duration trips ($\phi_i = 1/72,000$, $\tau_{i,j} = 1/4,000$). As before, Flux model parameters are defined from the Simple Trip model parameters using Eq. 4.

Fig. 3 illustrates how prevalence in location 1 varies in relation to $R_0$ in location 1. As long as the prevalences in each location are different, there is a dramatic separation between the Flux and Simple Trip model results, particularly in the short travel duration regime (Fig. 3 A and B). If travel duration is long, however, the two models’ predicted relationships between prevalence and $R_0$ do approach one another. The relationships between prevalence and transmission intensity for the two Ross–Macdonald models are illustrated in Fig. 4 A and B. In this parameter regime, the Flux model and Simple Trip models converge, because the constraint on time spent away from home makes less of a difference.

Fig. 2. Comparing SIS model results. We numerically solve Eqs. 9 and 10 for the prevalence at endemic equilibrium. We plot the endemic prevalence in location 1 ($I_1/N_i$) as a function of transmission intensity in location 1 ($\beta_1/\gamma$) while holding transmission intensity in location 2 ($\beta_2/\gamma$) constant. We explore four different parameter regimes defined by the transmission intensity of location 2 and the duration of travel. When travel duration is short (Upper), the Flux (red) and Simple Trip (blue) models produce dramatically different relationships between transmission intensity and prevalence. When travel duration is long (Lower), the two movement models’ results converge.

- (A) Short travel duration regime ($\phi_i = 0.5$, $\tau_{i,j} = 40$); low transmission in location 2 ($\beta_2 = 0.5$).
- (B) Short travel duration regime ($\phi_i = 0.5$, $\tau_{i,j} = 40$); high transmission in location 2 ($\beta_2 = 2.5$).
- (C) Long travel duration regime, $\phi_i = 0.005$, $\tau_{i,j} = 0.4$; low transmission in location 2 ($\beta_2 = 0.5$).
- (D) Long travel duration regime, $\phi_i = 0.005$, $\tau_{i,j} = 0.4$; high transmission in location 2 ($\beta_2 = 2.5$).
Comparing Ross–Macdonald model results. We numerically solve Eqs. 14 and 16 to find the relationship between prevalence in location 1 ($X_1/N_1$) and transmission intensity ($R_0$). We plot the relationship between location 1 prevalence ($X_1/N_1$) as a function of transmission intensity in location 1 while holding prevalence constant ($X_2/N_2$) in location 2. Note that we plot prevalence on the y axis and transmission intensity on the x axis, to emphasize how the Ross–Macdonald model behaves similarly to the SIS model (Fig. 2). We explore four different parameter regimes defined by the prevalence of location 2 and the duration of travel in relation to the average duration of infection ($\tau^{-1} = 200$ d). In the short travel duration regime, there is a large quantitative difference in the relationship between prevalence and $R_0$ for the two different movement models. In the regime with short travel duration and high location 2 prevalence (Upper Right), for a given prevalence of 0.2 using the Flux model would suggest low transmission ($R_0^{(i)} = 1$), whereas using the Simple Trip model would suggest high transmission ($R_0^{(i)} > 1$). In the long travel duration regime, the quantitative disagreement between the two movement models is less significant. (A) Short travel duration regime ($\phi_i = 1/180$, $\tau_{ij} = 1/10$); low prevalence in location 2 ($X_2/N_2 = 0.025$). (B) Short travel duration regime ($\phi_i = 1/180$, $\tau_{ij} = 1/10$); high prevalence in location 2 ($X_2/N_2 = 0.25$). (C) Long travel duration regime ($\phi_i = 1/72,000$, $\tau_{ij} = 1/4000$); low prevalence in location 2 ($X_2/N_2 = 0.025$). (D) Long travel duration regime ($\phi_i = 1/72,000$, $\tau_{ij} = 1/4000$); high prevalence in location 2 ($X_2/N_2 = 0.25$).
Fig. 4. Parameterizing Ross–Macdonald model from data. (A) Prevalence (parasite rate, \(X_i/N_i\)) map, estimated from the MIS collected parasitemia measurements, data from ref. 14. Each pixel shows the mean estimated prevalence within a 1\(\times\)1-km\(^2\) area. (B) Travel “prevalence” map, reproduced with permission from ref. 14, where each pixel shows the probability that a resident leaves home during an 8-wk survey period. (C) Modeled rates of travel \(\{\phi_{ik}\}\) for each of four example pixels, estimated from MIS data reporting recent travel. Arrows pointing off-island represent travel to mainland EG. Arrow weights are proportional to rate of travel from origin pixel to each of the seven destination regions, or the sum total rate of travel to all of the destination pixels within each region.

We frame our understanding of these results in terms of the \(K = 2\) subpopulation example discussed previously for the Ross–Macdonald model. Each subpopulation on Bioko Island exists in a parameter regime where travel duration is short (\(\tau_{i,j} > r\)) and the prevalence in the off-island travel destination is higher than the local prevalence. We recall the results shown in Fig. 3A, where we see how for low prevalence values the Flux model tends to underestimate \(R_0\) compared to the Simple Trip model. The Simple Trip model constrains the amount of time that travelers spend on mainland EG and therefore reduces the number of people who return home from traveling with malaria infections. The Flux model does not include such a constraint, and in this parameter regime we see how in some locations the rate of imported infections is so high that the model breaks down and does not allow meaningful values of \(R_{0,\text{Flux}}\).

Fig. 5. Modeled estimates of \(R_0\). All modeled estimates were generated using the same prevalence and travel data, illustrated in Fig. 4. (A) \(R_{0,\text{Flux}}\) calculated using the Flux model (Eq. 14) parameterized with the input data shown in Fig. 4. Dark gray denotes areas where there is no valid solution for \(R_{0,\text{Flux}}\), meaning that according to the Flux model travelers are subject to too high transmission risk when on the mainland. (B) \(R_{0,\text{ST}}\) calculated using the Simple Trip model (Eq. 16) parameterized with the input data shown in Fig. 4. (C) Histograms comparing the distributions of \(R_{0,\text{Flux}}\) and \(R_{0,\text{ST}}\), further contrasting the two sets of predictions and further illustrating how the Flux model fails to produce meaningful results in some areas.
Discussion

We have shown that models of infectious disease dynamics are sensitive to the choices that modelers make when including movement-mediated interactions between subpopulations. To do so, we have adapted each of three transmission models to include two types of movement between subpopulations. For all three transmission models we have identified parameter regimes where choosing to incorporate one movement model instead of the other will result in dramatically different results. Specifically, when there is a difference in transmission intensity between two interacting locations, and when mean travel duration is short compared to the mean duration of infection.

In all instances we set movement parameters to match the total flux of individuals moving between each pair of subpopulations. In this way, we have emulated a situation where the modeler has a single movement dataset and must choose which movement model to implement. The full impact of the choice of movement model became clear from the example of modeling malaria transmission and importation on Bioko Island: We calibrated each of the two movement models using the same dataset and found that each movement model produced dramatically different results. In some areas, the Flux model and was unable to produce meaningful values of $R_0$. The Simple Trip model, by specifying the rate at which travelers return home, adds a constraint on the amount of time that travelers spend while away from home. As a consequence, the Simple Trip model constrains the amount of risk experienced by travelers. The Flux model has no such constraint: As long as there are variations in the transmission intensity across different locations there will be significant differences between the two movement models’ results.

The way that each of the movement models contributes to the infectious disease model results can lead to real confusion over the true relationship between transmission intensity and epidemiological outcomes. In some applications, as with the Bioko Island example, the purpose of the model is to infer transmission conditions, which may be difficult to measure empirically, based on more readily available prevalence or incidence measures. Returning to Fig. 2B, we can imagine trying to determine the transmission intensity $\beta_0$ based on a measured prevalence of 0.25. From the Flux model we infer $\beta_0 < 1$, which suggests that local transmission risk is low. From the Simple Trip model, however, we infer $\beta_0 > 1$, which suggests the opposite. A similar inconsistency is illustrated for the Ross–Macdonald model in Fig. 3B and appears again in the Bioko Island example in Fig. 5.

In the case of modeling malaria transmission and importation on Bioko Island we have shown how the Flux model is not suitable for producing meaningful results. Despite this drawback, the Flux model remains an attractive option for creating a basic quantitative description of host movement in other epidemiological contexts. The Flux model requires fewer parameters to specify completely. For this reason, parameterization is possible with relatively simple datasets which count the number of individuals moving from one each location to each other location over time. Unfortunately, datasets which record flux volumes do not necessarily lend themselves to parameterizing Simple Trip models, as they do not distinguish between the rates of leaving and returning home.

For example, mobile phone call data records (CDRs) are a source of high-volume datasets for measuring the number of individuals who travel from one location to another. Often, for the sake of preserving subscriber privacy, CDRs are aggregated across individual users, time, and geographical area before becoming public. That is to say, publicly available CDR datasets count the number of mobile phones traveling from one service tower catchment area to another on a daily or weekly basis; these datasets do not record an individual’s rate of leaving home or time spent while traveling away from home. The loss of individual identity from datasets has already been shown to impact the predicted outcome of disease models in other contexts (9).

While the limits and biases inherent to CDR datasets have been well-documented (43), the present study suggests an additional drawback when it comes to using CDR datasets for epidemiological modeling: Because it is not clear that a Flux model is sufficient to produce accurate modeling results in some contexts, it is likewise not clear that flux data are sufficient to quantify travel behavior when building a movement model.

This is not to say that CDR or flux models are useless in every context. Our aim is that the present study serves as a starting point for understanding when the flux model and Simple Trip model are interchangeable. In some settings, the conclusions may not be affected by the choice of a model, but one needs to carefully consider the epidemiological context before choosing a movement model. More generally, the question of which model is appropriate to use in context or how to model the spatial dynamics of pathogens merits greater attention (9, 23, 27).

The present work is restricted to discussing two basic deterministic models of movement. In theory, one may imagine similarly comparing more detailed and complicated movement models, such as incorporating travel behavior with multiple stops, adding stochastic effects, or using agent-based models to incorporate movement behavior heterogeneity among the travelers. It remains to be seen the extent to which such additional model features might further affect the epidemiological outcomes of interest. In practice, however, the data required for calibrating such models may not be readily available, and using a Flux or Simple Trip description of travel behavior may be more practical. Even for such simple movement models there is a risk of misusing the travel data that are available to calibrate a movement model which cannot be reconciled with the data and produces outputs that are difficult or impossible to interpret.

Supporting Datasets

Refer to SI Appendix, section 4 to find code and data supporting the analysis presented in Malaria Modeling Example and Figs. 4 and 5.

Data Availability. Data and code supporting the analysis presented in Malaria Modeling Example have been deposited in Figshare (https://doi.org/10.6084/m9.figshare.12084831.v2). All study data are included in the article and/or SI Appendix.

ACKNOWLEDGMENTS. We acknowledge support from the Bill & Melinda Gates Foundation, Grant OPP1110495. Additionally, we thank the participants of Bioko Island who have taken part in the surveys from which the Bioko data were drawn, the BIMEP field surveyors and supervisors who collected these data, as well as the National Malaria Control Program and the Ministry of Health and Social Welfare of Equatorial Guinea, Marathon Oil, Nobel Energy, and partners for their continued support in producing these datasets. We thank the following collaborators for their helpful insights and discussion: Joel C. Miller, Nick W. Ruktanonchai, M. U. G. Kraemer, Aliu M. Wasolowski, John R. Gilles, Guillermo A. Garcia, Dianna Hergott, Su Yun Kang, Katherine E. Battle, and Daniel J. Weiss.