

Linked within-host and between-host models and data for infectious diseases: a systematic review

Lauren M. Childs¹, Fadoua El Moustaid^{2,3}, Zachary Gajewski^{2,3,4}, Sarah Kadelka¹, Ryan Nikin-Beers^{1,5}, John W. Smith, Jr⁴, Melody Walker¹ and Leah R. Johnson^{2,3,4,6}

¹ Department of Mathematics, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA, USA

² Department of Biological Sciences, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA, USA

³ Global Change Center, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA, USA

⁴ Department of Statistics, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA, USA

⁵ Department of Mathematics, University of Florida, Gainesville, FL, USA

⁶ Computational Modeling and Data Analytics, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA, USA

ABSTRACT

The observed dynamics of infectious diseases are driven by processes across multiple scales. Here we focus on two: within-host, that is, how an infection progresses inside a single individual (for instance viral and immune dynamics), and between-host, that is, how the infection is transmitted between multiple individuals of a host population. The dynamics of each of these may be influenced by the other, particularly across evolutionary time. Thus understanding each of these scales, and the links between them, is necessary for a holistic understanding of the spread of infectious diseases. One approach to combining these scales is through mathematical modeling. We conducted a systematic review of the published literature on multi-scale mathematical models of disease transmission (as defined by combining within-host and between-host scales) to determine the extent to which mathematical models are being used to understand across-scale transmission, and the extent to which these models are being confronted with data. Following the PRISMA guidelines for systematic reviews, we identified 24 of 197 qualifying papers across 30 years that include both linked models at the within and between host scales and that used data to parameterize/calibrate models. We find that the approach that incorporates both modeling with data is under-utilized, if increasing. This highlights the need for better communication and collaboration between modelers and empiricists to build well-calibrated models that both improve understanding and may be used for prediction.

Submitted 7 January 2019

Accepted 28 April 2019

Published 19 June 2019

Corresponding author

Leah R. Johnson, lrjohn@vt.edu

Academic editor

Andrew Tredennick

Additional Information and
Declarations can be found on
page 13

DOI 10.7717/peerj.7057

© Copyright

2019 Childs et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Computational Biology, Mathematical Biology, Epidemiology, Infectious Diseases

Keywords Multit-scale modeling, Linking mechanism, Infectious disease models, SIR models, Data-model integration, Within-host, Between-host, Pathogen transmission

INTRODUCTION

In the study of biological systems, phenomena are often observed at multiple scales—from sub-cellular to entire populations. Here, we focus on the between-host scale and the within-host scale, given their frequent appearance as a dichotomy in the study of infectious disease (Mideo, Alizon & Day, 2008; Handel & Rohani, 2015; Willem et al., 2017). The between-host scale may include how a disease spreads among organisms in a population, while the within-host scale may include intra-cellular or inter-cellular interactions with an invading pathogen. Since biological systems often do not exhibit a clear separation of temporal or spatial scales, there has been increased interest in recent years in how interactions at one scale can affect interactions at the other (Handel & Rohani, 2015; Alizon, Luciani & Regoes, 2011; Mideo et al., 2013; Theys et al., 2018; Dorratoltaj et al., 2017).

Mathematical and computational modeling, which has a rich history of application to the dynamics of ecological systems and infectious diseases, has been used to study phenomena at both within-host and between-host scales (as well as other scales as is reviewed in Garira (2017)). At the between-host scale, classic compartmental models, like the *SIR* model, which represents the interactions between susceptible individuals *S*, infected individuals *I*, and recovered individuals *R*, have been used to predict the spread of infectious diseases between individuals in a population (Kermack & McKendrick, 1991, 1927; Anderson & May, 1992). At the within-host scale, models such as the *TIV* model of viral dynamics, which represents the interactions between target cells *T*, infected cells *I*, and virus *V*, have been used to understand viral load within hosts (Perelson et al., 1996; Nowak & May, 2000).

To understand the outcomes produced by the interactions in and between different scales, a multi-scale model that links the scales may be constructed. For example, an *SIR* model may be used to describe the spread of a viral disease in a population. If the transmission rate between hosts is dependent on the outcome of the viral load from a *TIV* model (since higher viral loads often are associated with higher disease transmission, e.g., Nguyen et al. (2013)), the models at the between-host scale and the within-host scale depend on one another, and are thus considered linked. These models can be diverse in their structure and formulation (Garira, 2017; Garira, Mathebula & Netshikweta, 2014). To be clear, multi-scale models encompass a wide range of possibilities, as reviewed in (Garira, 2017). Here, we focus on the within-host and between-host scales for infectious diseases.

Thinking about the implications across scales is important but is also challenging as the relationships are often complex, nonlinear and, therefore, un-intuitive. Previously, theoretical models of multi-scale phenomena have been reviewed (Mideo, Alizon & Day, 2008; Reiner et al., 2013; Dorratoltaj et al., 2017; Murillo, Murillo & Perelson, 2013; Severins, 2012). Repeated themes of these works and others over the past two decades have included: the need for more data (Alizon & Van Baalen, 2008; Alizon, Luciani & Regoes, 2011; Handel & Rohani, 2015; Lavine, Poss & Grenfell, 2008; Pollitt et al., 2011); the challenge of integrating scales (Frost et al., 2015; Perelson et al., 1996;

Handel & Rohani, 2015; Mideo et al., 2013); and the role of heterogeneity (*Lavine, Poss & Grenfell, 2008; VanderWaal & Ezenwa, 2016*). Furthermore, there was an emphasis on the role of particular quantities such as within-host trade-offs (*Martinez-Bakker & Helm, 2015; Pollitt et al., 2011*) and immune response factors (*Graham et al., 2007; Hawley & Altizer, 2011*).

Of the 22 reviews found by our search, two were themselves systematic reviews (*Dorratoltaj et al., 2017; Willem et al., 2017*). The former had a narrow focus on multi-scale models of HIV infections in humans (and returned nine papers), while the latter examined all individual-based models (IBMs) of infectious disease (698 in total) from 2006 to 2015. Of the remainder of reviews, several focused mainly on biological aspects of multi-scale dynamics (*Dolan, Whitfield & Andino, 2018; Forrester, Gutiérrez & Coffey, 2016; Pimenoff et al., 2018; Wilks et al., 2012*), or were more methodologically based (*Kao, 2010; Willem et al., 2017*). Two recent works (*Dorratoltaj et al., 2017; Theys et al., 2018*) focus specifically on multi-scale models of HIV infections of humans. In particular, *Dorratoltaj et al. (2017)* only returned nine papers, a relatively low number, that is, similar to our own study. However, they actually found three papers that did not appear in our search (*Saenz & Bonhoeffer, 2013; Metzger, Lloyd-Smith & Weinberger, 2011; Yeghiazarian, Cumberland & Yang, 2013*). As we required within- and between-host components as well as data, none of these three papers match our criteria, and thus it is unremarkable that we did not find these papers through our search.

In this review, we aim to illuminate the state of the field joining experimental data with mathematical and computational models that bridge within-host and between-host scales. By doing so in a systematic manner we expect to identify potential gaps in understanding and methodology. Thus, we examine papers that incorporate models that contain linked within-host and between-host model components as well as explicitly utilize data. While we have related an example that involves the linking of two compartmental models in the context of a viral disease, we do not restrict our search to only compartmental models or to models of viral disease. We find 24 papers which contain both (i) the within-host and between-host scales and their connection and (ii) data. In “Survey Methodology,” we describe how we searched for and chose papers. In “Results,” we explain trends of the models in the papers we selected. We then conclude in “Discussion” with some overall thoughts on the current literature using multi-scale models with data.

SURVEY METHODOLOGY

To perform our systematic review we followed the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines (*Moher et al., 2009*). PRISMA is a standard protocol for conducting a systematic review or a meta-analysis. The flowchart showing our procedures are presented in [Fig. 1A](#).

We searched “All Databases” on Web of Science using the search terms (*within-host* OR in-host-model* OR among-host-model**) AND (*between-host* OR nested-model* OR cross-scale-model**) AND (*pathogen* OR parasite**) AND (*transmi**) for papers published up to December 31, 2018. Terms combined within parentheses with “OR” require at least

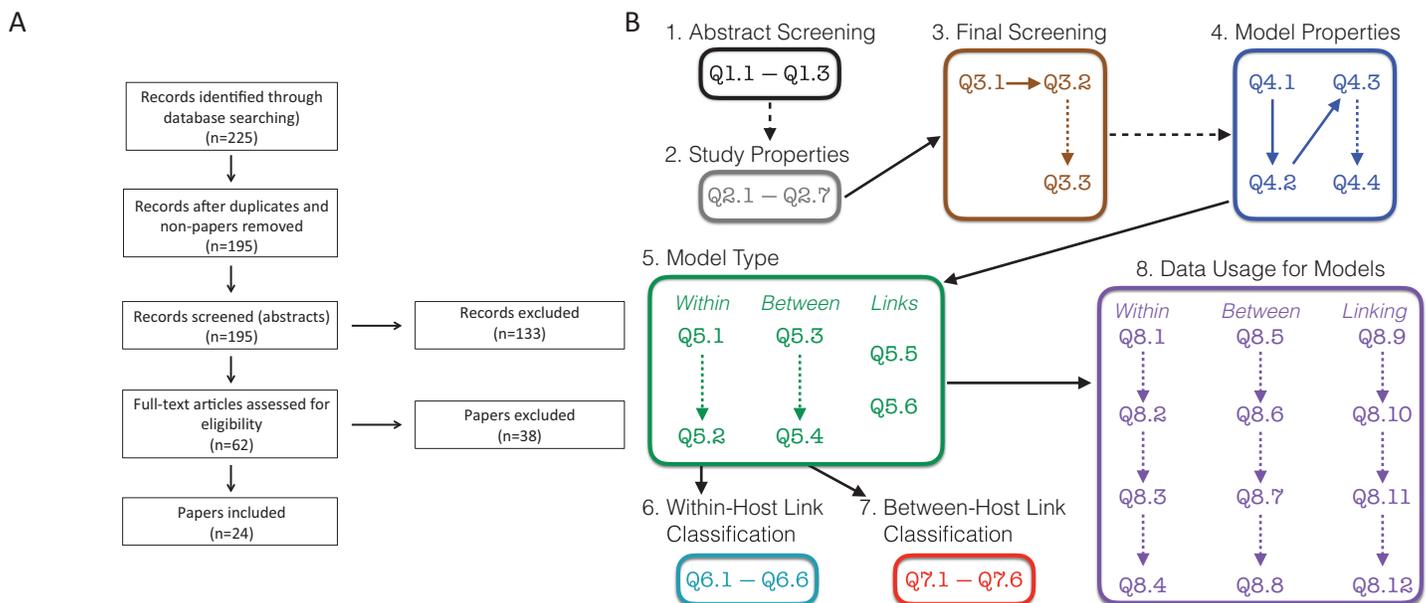


Figure 1 Schematic of survey methodology. (A) PRISMA flowchart showing the inclusion of papers. “Non-papers” refers to database entries that were figures or codes. (B) Schematic of the screening and evaluation questions used. Dashed lines indicate links between questions that were conditional, that is, answering the second question/box depended on the answers to the earlier question. For example, details on the study properties (Q2.1–Q2.7) and questions from the final screening stage (Q3.1–Q3.3) were only collected for the 195 papers that were retained following the abstract screening stage. Questions in boxes 4–8 were completed for all 24 papers that remained following the final screening stage. Questions are found in [Text S1](#); Responses are found in [Tables S1–S8](#); References to all included papers are found in [Text S2](#); References to all excluded papers are found in [Text S3](#); All recorded data can be found in our [Supplemental Data Sets](#). Full-size [DOI: 10.7717/peerj.7057/fig-1](https://doi.org/10.7717/peerj.7057/fig-1)

one of these terms while the “AND” between terms in parentheses requires something from each group of terms. Terms that end in a * indicate that any form of the ending of the word would be acceptable. For example, “transmi*” would return papers with transmit or transmission. We note that this search type accesses the “title,” “abstract,” “author keywords,” and “keywords plus.” If the search terms are not found in these locations, the papers will not be returned from the search.

Based on these search terms, we obtained 225 results ([Fig. 1A](#)). We initially eliminated 29 search results, which included duplicates and other results that were not papers. Further, there was one paper that could not be obtained in English ([Verenini, 1983](#)); only an Italian version was found. This left us with 195 papers, which we initially screened based on the abstract. Two additional papers were suggested during the review process ([Smith & Mideo, 2017](#); [Greenspoon, Banton & Mideo, 2018](#)), for which the abstracts were screened, and the papers were ultimately excluded. These two papers are included in our presentation of data but do not appear in the survey methodology as they were not recovered by the search.

In the initial abstract screening phase, two randomly assigned people (i.e., two of LMC, FEM, ZG, SK, RNB, MW, or LRJ) separately categorized each of the 195 papers into three categories based on whether it appeared to include a linked model with data based purely on the abstract: “Yes,” “Maybe,” and “No.” A linked model was defined as a mathematical model that includes at least two scales, within-host and between-host, as well

as some explicit link between the scales. The abstract was labeled as follows: “Yes” if it appeared to describe both a linked model with data; “Maybe” if it either (i) clearly described a linked model but was unclear on data, or (ii) clearly referred to data, but was unclear if the included model was linked; “No” if it did not meet any of the above criteria, was obviously a review, or obviously outside the scope of our review. A set of study properties (Fig. 1B, Q2.1–Q2.7) were also collected for each of 158 papers at the abstract screening stage including the focal host species, other mentioned species, the type of infection, and the main results of the paper. Study properties were not recorded for the other 37 papers as they were either review papers not out of scope (19) or deemed out of scope (18). The total number of papers for which we recorded the study properties is 160. This includes the two additional papers that were suggested during the review process. If an abstract was labeled with two “Yes” or with one “Yes” and one “Maybe,” we retained the paper for full paper screening; if an abstract was labeled with two “No” we excluded the paper from screening. If an abstract was labeled with one “Yes” and one “No,” we reviewed the abstract collectively to relabel it to two “Yes,” two “No,” or one “Maybe.” If an abstract was labeled with one “Maybe” and one “No,” the person who labeled “Maybe” was assigned to skim the paper to decide if the paper should be kept or eliminated. If an abstract was labeled with two “Maybe,” a third randomly chosen person was assigned to skim the paper to decide whether it should be kept or eliminated. A record was kept if it appeared to have a linked model and/or data, but still was unclear if it had both; the paper was excluded otherwise. Once this process was completed, we kept 62 papers for further screening, and excluded 133 papers based on the abstracts. The reason for exclusion (lacking data, lacking a model, lacking a within-host component, lacking a between-host component, review, or another reason, which needed to be described) was recorded for all 133 papers excluded at this stage (Fig. 1B, Q3.2–Q3.3). Although there may have been multiple reasons to exclude papers, only one reason was recorded.

We then conducted a final screening of the remaining 62 papers by having two individuals (randomly assigned from the full author list) read through the full text of each paper. During this step, a final determination was made for each paper whether to keep it for further analysis or to exclude it. A paper was kept if it contained a linked model with data; a paper was otherwise excluded. The reason for exclusion (lacking data, lacking a model, lacking a within-host component, lacking a between-host component, review, or another reason, which needed to be described) was recorded for all 38 papers excluded at this stage (Fig. 1B, Q3.2–Q3.3). In all, we included 24 papers (Althouse & Hanley, 2015; Chaves, Kaneko & Pascual, 2009; Chen, Sanderson & Lanzas, 2013; Cooper & Heinemann, 2005; Day, Alizon & Mideo, 2011; Dennehy et al., 2006; Dwyer, Levin & Buttel, 1990; Fryer et al., 2012; Giardina et al., 2017; Hall & Mideo, 2018; Handel et al., 2013, Handel et al., 2014; Kennedy & Dwyer, 2018; Leclerc et al., 2014; Lindberg et al., 2018; McKenzie & Bossert, 2005; Mideo et al., 2011; Reperant et al., 2012; Stephenson et al., 2017; Takumi et al., 2010; Tuncer et al., 2016; Van Dorp, Van Boven & De Boer, 2014; Volz, Romero-Severson & Leitner, 2017; Vrancken et al., 2014) in the full analyses (Fig. 1A).

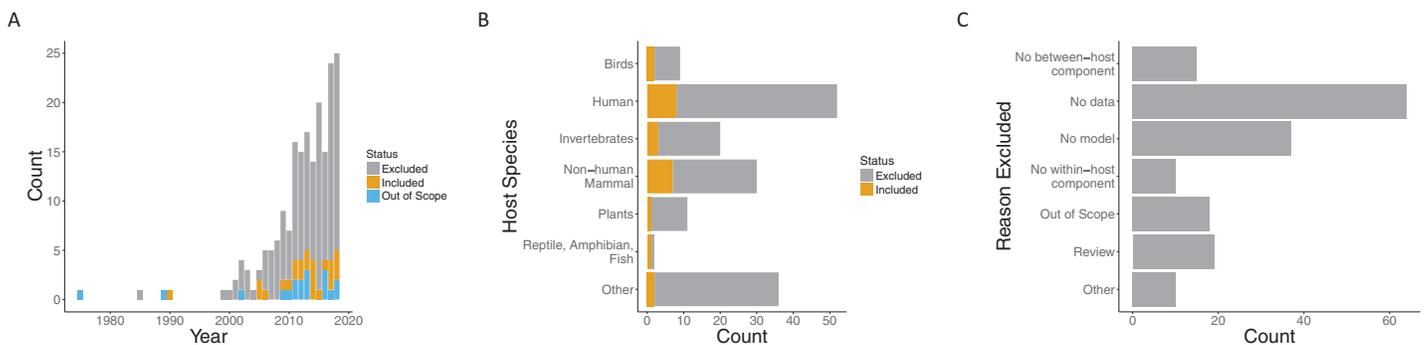


Figure 2 Summary of papers considered. Both included and excluded papers by (A) year of publication, (B) host species, and (C) the reason for exclusion (only for excluded papers). Papers were classified as included (gray), out of scope (orange) or excluded (blue) for (A). “Out of scope” designated papers that literally included the search terms but were not topically related. [Full-size !\[\]\(1679558f37f6db0dd8360a2a7e913e90_img.jpg\) DOI: 10.7717/peerj.7057/fig-2](https://doi.org/10.7717/peerj.7057/fig-2)

For the papers that were included, we answered a detailed set of questions, which described important aspects of the model (such as the transmission route), how the models were linked, how the data was used in the model, etc. (Text S1; Fig. 1B, Q4–Q8). We further characterized the journal in which each paper appeared as a general audience journal, a specialized biological journal, a primarily mathematical/computational journal, or a biology sub-discipline journal (Fig. S1).

RESULTS

Traits of included compared to excluded papers

Our search process (including suggestions in the review process) yielded 197 papers published over the span of more than 30 years. While the earliest included paper was from 1990 (Dwyer, Levin & Buttel, 1990), the next papers that met our requirements were published 15 years later (Cooper & Heinemann, 2005; McKenzie & Bossert, 2005). In the interim, a few more papers were published, but interest in this general area grew quickly starting in 2005. Both the number of papers loosely related to the topic (i.e., those excluded) and papers meeting our criteria to include both models and data (i.e., those included) increased in that time frame (Fig. 2A).

Papers spanned a variety of host species systems (Fig. S2). Infections of humans were, not surprisingly, the most common in both the excluded (44/135) and included categories (8/24), followed by non-human mammals (30 overall) and invertebrates (20 overall). Although human infections were considered in the largest number of included papers overall, the proportion of included papers when broken down by focal host species is largest for non-human mammals (7/30) and approximately the same for invertebrates (3/20) and humans (8/52). The most common reason for exclusion was a lack of data being used with the model (64/135) followed by no model (37/135) (note, that only one reason was recorded for each paper). That is, many papers explore within- to between-host transmission either from a modeling or empirical perspective, but many fewer link the models robustly to data. Recently, there have been a number of review papers on multi-scale models with data, another common reason for exclusion (22/135).

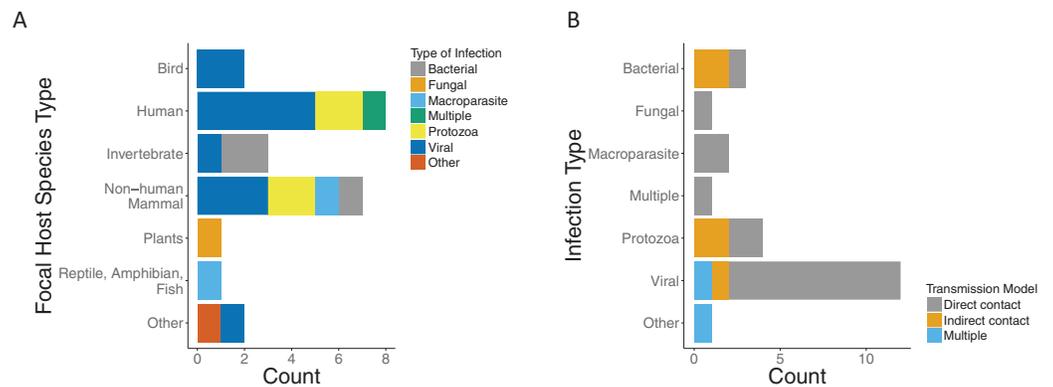


Figure 3 Focal host species and infection types for included papers. (A) Type of infection across host species for included papers as bacterial (gray), fungal (orange), macroparasite (blue), multiple (green), protozoa (yellow), viral (dark blue), or other (red). (B) Modeled transmission route across infection types for included papers as direct contact (gray), indirect contact (orange) or multiple routes (blue).

Full-size [DOI: 10.7717/peerj.7057/fig-3](https://doi.org/10.7717/peerj.7057/fig-3)

Traits of included papers

We considered whether the aim of each paper was primarily strategic (trying to understand underlying dynamics) or primarily tactical (trying to make predictions) (Nisbet & Gurney, 1982). Of the papers examined, most were classified as primarily strategic and very few papers as primarily tactical. Only one paper was classified as both strategic and tactical (Vrancken et al., 2014) (Table S4.2). Included papers were rarely found in highly specialized non-mathematical journals (2/24), but were relatively equally spread between mathematically focused journals, biology focused, and for a general audience (Fig. S1).

Infection, host, and transmission categorization

We found that the majority of the papers and models focused on a single infection. Most infection types were viral (12/24), with protozoa being the second most common (4/24). The host species were predominately mammals (15/24), of which eight were human hosts (Fig. 3A). Most papers modeled transmission as direct contact across infection types. Half of those with protozoa infection type (2/4) and two thirds of those with bacterial infection (2/3) were modeled by indirect contact. Of the papers which modeled viral infections, all but two model direct transmission. In addition, there was one which was indirect (Handel et al., 2014) and one with multiple modes of transmission (Handel et al., 2013) (Fig. 3B).

Model characteristics

The multi-scale models reviewed are composed of three parts: the within-host model, the between-host model, and the linking mechanism. Although we only considered papers with models including all of these three components, as well as data, the papers varied on the focus of their results. Approximately two-thirds (15/24) of papers primarily investigated how the within-host dynamics affect the between-host dynamics; only two papers focused on the impact of the between-host dynamics on the within-host dynamics. The remaining third

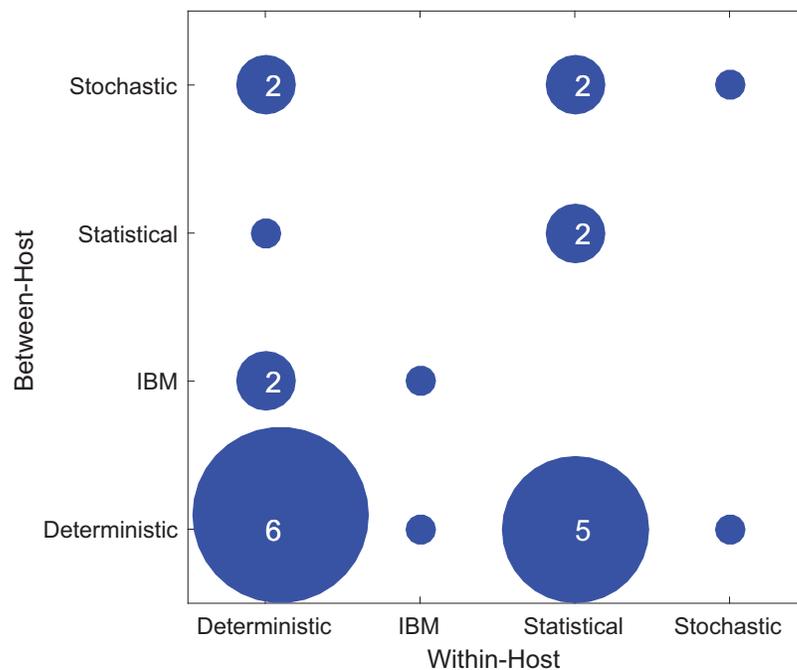


Figure 4 Types of modeling framework used in included papers. The x -axis shows the model types used in the within-host part of the model while the y -axis shows the model types used in the between-host model. The dots' diameter represents how many papers used a particular framework.

Full-size  DOI: [10.7717/peerj.7057/fig-4](https://doi.org/10.7717/peerj.7057/fig-4)

of the papers either examined both of the above directions of impact (i.e., how the within-host and between-host dynamics effect each other) or the influence of within- or between-host dynamics on another factor in their model (Table S4.3).

In the multi-scale models we considered, the within-host component and between-host component were both modeled explicitly. We characterized each of the within-host and between-host models used as either a deterministic model, a stochastic model, an IBM, or a statistical model. In all cases, our categorization is based on the process/mechanism portion of the model being considered. Deterministic models included any mechanistic/process based model that did not include stochasticity in the process. For example, this may include ordinary differential equation (ODE)/compartmental models. These models may have been fit assuming a stochastic observation model overlaid on the dynamics. Stochastic models include any model that includes stochasticity in the description of the mechanism/process. Examples of this would include stochastic SIR models, or stochastic differential equations, but *not* IBMs. Here, IBMs include any models in which individuals or agents were modeled separately from each other and allowed to interact within a simulated environment. This is in contrast to models that include data collected on individuals and where these are used to parameterize models (we would refer to these as “trait based”). IBMs as defined here may contain both deterministic and stochastic components. These models are typically much more difficult to analyze and fit to data than either of the other two flavors of mechanistic models, hence why we considered them separately. Finally, statistical models are those that seek to fit a function to data without the

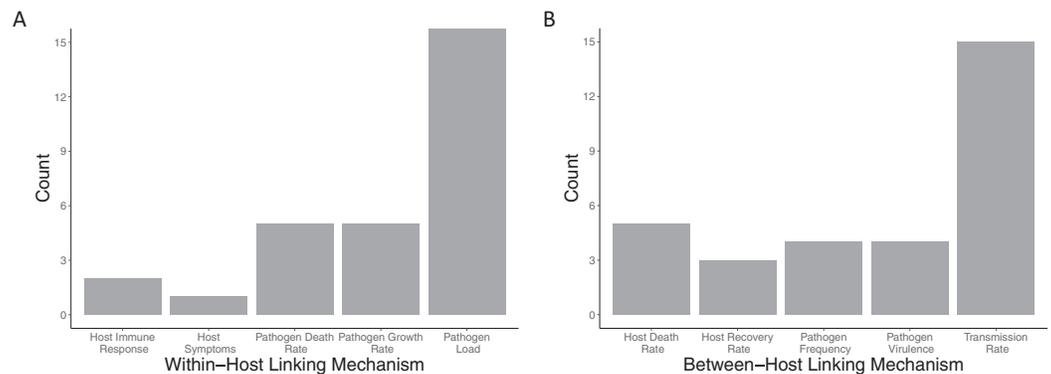


Figure 5 Mechanisms used to link between and within-host models together. The number of included papers that used the each of the (A) within-host linking mechanisms and (B) between-host linking mechanisms to connect the models together. [Full-size !\[\]\(43e4e85f3e07c993849f98a1fa87e486_img.jpg\) DOI: 10.7717/peerj.7057/fig-5](https://doi.org/10.7717/peerj.7057/fig-5)

function having an explicit link to a mechanism for what produces a pattern. This would include all flavors of regression, and many machine-learning approaches.

Figure 4 shows the types of within-host and between-host models used in the included papers. Most studies used the deterministic model type at least once, either for within- or between-host models and sometimes for both. In the included papers, within-host models were most commonly deterministic (11/24), followed by statistical (9/24), individual-based (2/24), and stochastic (2/24). In contrast, for the between-host models, the majority were deterministic (13/24), with a lower and more evenly distributed representation of stochastic (5/24), individual based (3/24), and statistical (3/24). One study used an IBM model type for both the within-host model and the between-host model (*Van Dorp, Van Boven & De Boer, 2014*). In general, studies did not typically use the same modeling approach for both the within- and between-host components. As for host type, there was no evident correlation between model types and the focal host species used in the model (Fig. S3).

Within- and between-host models can be linked in three different ways: within- to between-, between- to within-, or bidirectionally. Regardless of the varied emphasis of the results of the papers, as described above, we examined the formulation of the linking mechanism of the model. Among the included papers, 12 of the studies linked the within-host model to the between-host model while 12 linked bidirectionally—both within-host to between-host and between-host to within-host (Table S5.5).

To link the within-host and between-host models, a linking mechanism was needed, which we categorized either as a state or a trait. Linking via a state meant that an outcome of the model was used; for example, the pathogen load at the within-host scale or the number of infected individuals at the between-host scale. In contrast, a trait was a parameter of the model; for example, the pathogen growth rate at the within-host scale or the transmission rate at the between-host scale. The model framework was categorized in one of three ways: linked only by states, only by traits, or by both traits and states. Furthermore, models could also have multiple linking mechanisms. In the included papers, eleven studies used state variables, four used trait variables, and nine used both (Table S5.6).

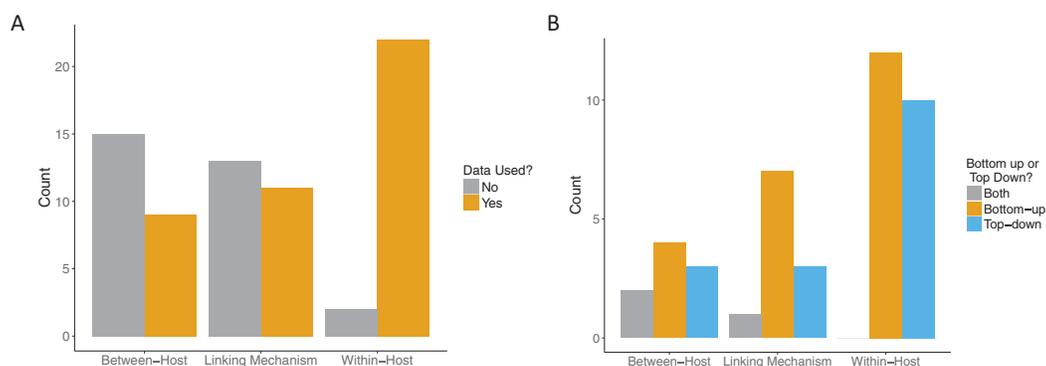


Figure 6 Role of data in multi-scale modeling efforts. (A) Scale (within-host, linking, or between-host) at which data was incorporated (orange) in the multi-scale models. Some models used data at more than one level. (B) How the data was incorporated into the models: bottom-up, that is, fitting traits (orange); top-down, that is, fitting states (blue) or both (gray). [Full-size](#) DOI: 10.7717/peerj.7057/fig-6

Within-host models (Fig. 5A) are linked to the between-host models mostly via the pathogen load, with more than half the papers using this linking mechanism (18/24). Pathogen growth rate was the second most used trait to link the within- to between-host model (5/24 papers). All other within-host linking mechanisms were used in two or fewer papers. Between-host models were also linked into the within-host models (Fig. 5B) based on primarily a single trait, the transmission rate (15/24). All other between-host linking mechanisms were used in five papers or fewer.

Role and method of data incorporation

All papers that passed the screening criteria utilized data in at least one component: the within-host component, the linking mechanism, or the between-host component. Our threshold for data use for model fitting was quite high. We typically required that data be explicitly fitted to some component of the model, for example, fitting the “state” of the system (e.g., the number of observed infectious individuals in an SIR model) to case data. However, one included paper did not explicitly fit data. Instead *Kennedy & Dwyer (2018)* constructed sophisticated linked within and between host models for baculovirus in gypsy moths and then compared the predictions of the calibrated model to newly obtained experimental data in a validation step. Even among the relatively small sample of papers that included data at all in these multi-scale models, most did not use it for more than one scale of their model (Fig. 6A). While most of the included papers (22/24) used data at the within-host scale, only seven papers used data at both the within-host and between-host scales, of which only four also used data for the linking mechanism. Papers that included data for both the linking mechanism and the between-host scale also included data for the within-host scale.

Across all model scales, bottom-up, that is, fitting of traits, was utilized more than top-down, that is, fitting of states, or other methods (Fig. 6B). Only one paper used a mixture of bottom-up and top-down data fitting methods (maximum likelihood and least squares, respectively) at different scales, although one paper did not specify explicitly how the data was incorporated into the model (*Cooper & Heinemann, 2005*). For data fitting that

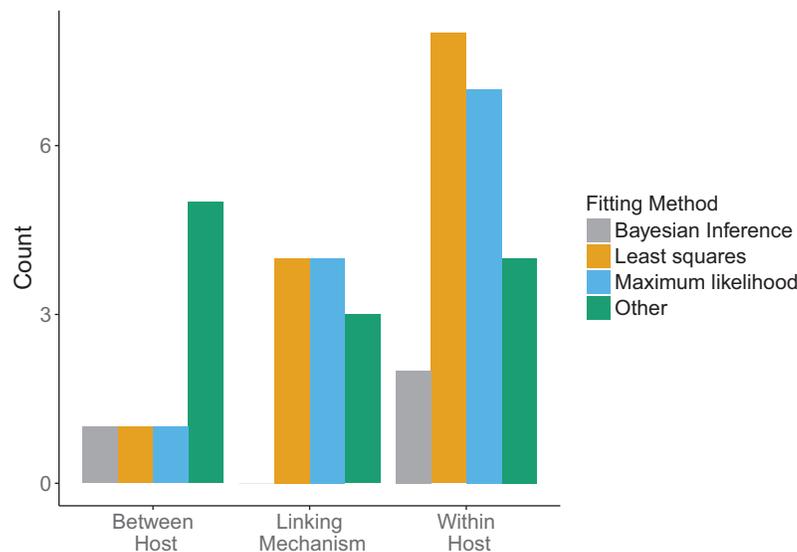


Figure 7 Method used in data fitting at each scale. Three fitting methods were considered: Bayesian inference (gray), least squares (orange), maximum likelihood (blue). All other fitting methods were included under “Other” (green). Different fitting methods could be used in the same papers for different scales. [Full-size](#) DOI: [10.7717/peerj.7057/fig-7](https://doi.org/10.7717/peerj.7057/fig-7)

was bottom-up, the majority of papers (5/12 within-host, 3/7 linking mechanism) used least-squares or maximum likelihood (6/12 within-host, 4/7 linking mechanism, 1/4 between-host). Bayesian inference, although a popular statistical method, was only used three times in the included papers (Fig. 7). Only a single paper (Volz, Romero-Severson & Leitner, 2017) recorded using multiple fitting methods at the same scale, and most papers used the same fitting method across all scales.

There was a diversity of fitting methods used across scales (Fig. 7). Different fitting methods could, and often were, used in the same papers for incorporation of data at different scales. The least squares method was used most when fitting data at the within-host scale, followed by maximum likelihood. Similarly, linking mechanisms primarily used least squares and maximum likelihood to fit data. In contrast, between-host models were less consistent. Across model components the category “Other” was mostly comprised of qualitative fitting methods, or papers for which authors were imprecise about how they fit the data, but it also included methods such as cubic-spline interpolation.

DISCUSSION

Our objective in this review was to determine how multi-scale infectious disease models, focusing on within-host and between-host scales, are used when they directly incorporate data. We focused on which host species are modeled, which pathogens are modeled, which types of models are used, how the within-host and between-host dynamics are linked, and at what scale data has been used. We found that it was most common for these models to describe a human population, to model a viral disease, to use a deterministic model at either of the two scales considered, to link the pathogen load at the within-host scale, to link the transmission rate at the between-host scale, and to use data at the within-host scale. It was

least common for these models to describe a plant, fish, reptile, or amphibian population, to model a bacterial, macroparasite, or fungal infection, to use a stochastic model at either of the two scales considered, to link host symptoms at the within-host scale, to link the host recovery rate at the between-host scale, and to use data at the between-host scale.

We speculate on the reasons for these outcomes. As human disease has tangible consequences directly impacting the wider population, it is unsurprising that the primary host species to examine these multi-scale interactions was found to be humans. However, the importance of other species both economically and ecologically leaves the door open for further study of these interactions. The dominance of viral disease as the focal pathogen likely results from the rich history of mathematical modeling in viral disease as well as their prominence in the human community. In choosing which type of model to use, deterministic models do not include the mathematical and computational complication of stochasticity, making them often easier to simulate and analyze than stochastic or IBMs. Further, explicitly linking the within-host and between-host scales is challenging. Many studies defaulted to the standard assumption that a higher pathogen load often correlates with a higher chance of disease transmission, making pathogen load the simplest way to link the within-host and between-host scales. Other linking mechanisms are often difficult to model because there may not be an obvious relationship in how two elements at different scales affect one another. The incorporation of data was primarily at the within-host scale, perhaps stemming from the fact that some of these relationships can be obtained through laboratory-based research. In contrast, between-host data may often require large-scale resources and monitoring.

We were quite surprised that our search yielded only 24 papers that included both across-scale modeling and substantial use of data. It is possible that our particular search terms may have been overly restrictive. For instance, the search term “pathogen” may be less likely to be used to describe infectious macro-parasites (e.g., worms). Nonetheless, our relatively small included set indicates that there is considerable scope for further work to be done in the area of data-rich multi-scale modeling of infectious diseases. Given the specific results of our review, we propose that future research could productively focus on (i) exploring alternative linking mechanisms and (ii) incorporating more and varied data at all scales.

Most studies we reviewed appeared to use the simplest assumption to link the within-host and between-host scales, namely, linking the pathogen load at the within-host scale to the transmission rate at the between-host scale. While this assumption may be appropriate for some diseases, there are other potential mechanisms that could be used to provide links between scales, separately or perhaps in addition to pathogen load. For example, these could include how host immunity affects the transmission rate or how pathogen load affects the pathogen virulence among the population. The possibility of using immune factors to link multi-scale models has been raised as a potential simplifying method (*Graham et al., 2007; Hawley & Altizer, 2011*). Within-host data on antibodies, when available, could be used as a measure of host immunity. Such simplifications, however, raise the question of which immune components are sufficient. Another major barrier, particularly in wild populations, is the lack of validated assays (*Hawley &*

Altizer, 2011). Accounting for these interactions could produce models that make complementary or potentially divergent predictions of transmission outcomes, and in turn be used to elucidate the effects different treatments have on disease spread.

The lack of data was the major reason that our search only uncovered 24 papers (Fig. 2C). Thus, a major gap in bridging within-host infection dynamics and between-host transmission is the existence and incorporation of data. This appears not to have improved significantly since similar observations in 2008 (*Alizon & Van Baalen, 2008*) and in 2015 (*Handel & Rohani, 2015*). Although there are cases where appropriate data for a model does not currently exist and must be collected in a new experiment, a greater effort should be put forth to work with and incorporate existing data sets. This is especially true at the population level, where data are particularly difficult and expensive to collect. Far more of the papers we examined included data at the within-host scale (Fig. 6A), likely due to the accessibility and scale of data that can be collected in a lab setting. Along with more data overall, the incorporation of more varied data at a variety of scales will enhance the utility of multi-scale disease modeling.

CONCLUSION

Important results about disease spread can be gleaned from modeling the interactions at both the within-host and between-host scales. While current research has mainly focused on simple assumptions, we believe that including additional complexities in future models may help to better explain observations from the field. Multi-scale modeling provides a great opportunity for empiricists and theorists to work together, and to contribute to the understanding of the drivers, treatments, and control of infectious disease.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Lauren M. Childs conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- Fadoua El Moustaid conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Zachary Gajewski conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Sarah Kadelka conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, approved the final draft.

- Ryan Nikin-Beers conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- John W. Smith, Jr performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Melody Walker conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- Leah R. Johnson conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data (files with the assessments of each paper) are available as [Supplemental Files](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.7057#supplemental-information>.

REFERENCES

- Alizon S, Luciani F, Regoes RR. 2011.** Epidemiological and clinical consequences of within-host evolution. *Trends in Microbiology* **19**(1):24–32 DOI [10.1016/j.tim.2010.09.005](https://doi.org/10.1016/j.tim.2010.09.005).
- Alizon S, Van Baalen M. 2008.** Acute or chronic? Within-host models with immune dynamics, infection outcome, and parasite evolution. *American Naturalist* **172**(6):E244–E256 DOI [10.1086/592404](https://doi.org/10.1086/592404).
- Althouse BM, Hanley KA. 2015.** The tortoise or the hare? Impacts of within-host dynamics on transmission success of arthropod-borne viruses. *Proceedings of the Royal Society of London B: Biological Sciences* **370**(1675):20140299.
- Anderson RM, May RM. 1992.** *Infectious diseases of humans: dynamics and control*. New York: Oxford University Press.
- Chaves LF, Kaneko A, Pascual M. 2009.** Random, top-down, or bottom-up coexistence of parasites: malaria population dynamics in multi-parasitic settings. *Ecology* **90**(9):2414–2425.
- Chen S, Sanderson M, Lanzas C. 2013.** Investigating effects of between- and within-host variability on *Escherichia coli* O157 shedding pattern and transmission. *Preventive Veterinary Medicine* **109**(1-2):47–57.
- Cooper T, Heinemann J. 2005.** Selection for plasmid post-segregational killing depends on multiple infection: evidence for the selection of more virulent parasites through parasite-level competition. *Proceedings of the Royal Society B: Biological Sciences* **272**(1561):403–410 DOI [10.1098/rspb.2004.2921](https://doi.org/10.1098/rspb.2004.2921).
- Day T, Alizon S, Mideo N. 2011.** Bridging scales in the evolution of infectious disease life histories: theory. *Evolution* **65**(12):3448–3461.
- Dennehy JJ, Friedenberg NA, Holt RD, Turner PE. 2006.** Viral ecology and the maintenance of novel host use. *The American Naturalist* **167**(3):429–439.
- Dolan PT, Whitfield ZJ, Andino R. 2018.** Mechanisms and concepts in RNA virus population dynamics and evolution. *Annual Review of Virology* **5**(1):69–92 DOI [10.1146/annurev-virology-101416-041718](https://doi.org/10.1146/annurev-virology-101416-041718).

- Dorratoltaj N, Nikin-Beers R, Ciupe SM, Eubank SG, Abbas KM. 2017.** Multi-scale immunoepidemiological modeling of within-host and between-host HIV dynamics: systematic review of mathematical models. *PeerJ* 5:e3877 DOI 10.7717/peerj.3877.
- Dwyer G, Levin SA, Buttel L. 1990.** A simulation model of the population dynamics and evolution of myxomatosis. *Ecological Monographs* 60(4):423–447 DOI 10.2307/1943014.
- Forrester NL, Gutiérrez S, Coffey LL. 2016.** The evolution and transmission of vector-borne viruses. In: Weaver SC, Denison M, Roossinck M, Vignuzzi M, eds. *Virus Evolution: Current Research and Future Directions*. Norfolk, GBR: Caister Academic Press, 61–100.
- Frost SD, Pybus OG, Gog JR, Viboud C, Bonhoeffer S, Bedford T. 2015.** Eight challenges in phylodynamic inference. *Epidemics* 10:88–92 DOI 10.1016/j.epidem.2014.09.001.
- Fryer HR, Frater J, Duda A, Palmer D, Phillips RE, McLean AR. 2012.** Cytotoxic T-lymphocyte escape mutations identified by HLA association favor those which escape and revert rapidly. *Journal of Virology* 86(16):8568–8580.
- Garira W. 2017.** A complete categorization of multiscale models of infectious disease systems. *Journal of Biological Dynamics* 11(1):378–435 DOI 10.1080/17513758.2017.1367849.
- Garira W, Mathebula D, Netshikweta R. 2014.** A mathematical modelling framework for linked within-host and between-host dynamics for infections with free-living pathogens in the environment. *Mathematical Biosciences* 256:58–78 DOI 10.1016/j.mbs.2014.08.004.
- Giardina F, Romero-Severson EO, Albert J, Britton T, Leitner T. 2017.** Inference of transmission network structure from HIV phylogenetic trees. *PLOS Computational Biology* 13(1):e1005316.
- Graham AL, Cattadori IM, Lloyd-Smith JO, Ferrari MJ, Bjørnstad ON. 2007.** Transmission consequences of coinfection: cytokines writ large? *Trends in Parasitology* 23(6):284–291 DOI 10.1016/j.pt.2007.04.005.
- Greenspoon PB, Banton S, Mideo N. 2018.** Immune system handling time may alter the outcome of competition between pathogens and the immune system. *Journal of Theoretical Biology* 447:25–31 DOI 10.1016/j.jtbi.2018.03.010.
- Hall MD, Mideo N. 2018.** Linking sex differences to the evolution of infectious disease life histories. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373(1757):20170431.
- Handel A, Brown J, Stallknecht D, Rohani P. 2013.** A multi-scale analysis of influenza A virus fitness trade-offs due to temperature-dependent virus persistence. *PLOS Computational Biology* 9(3):e1002989 DOI 10.1371/journal.pcbi.1002989.
- Handel A, Lebarbenchon C, Stallknecht D, Rohani P. 2014.** Trade-offs between and within scales: environmental persistence and within-host fitness of avian influenza viruses. *Proceedings of the Royal Society of London B: Biological Sciences* 281(1787):20133051 DOI 10.1098/rspb.2013.3051.
- Handel A, Rohani P. 2015.** Crossing the scale from within-host infection dynamics to between-host transmission fitness: a discussion of current assumptions and knowledge. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 370(1675):20140302 DOI 10.1098/rstb.2014.0302.
- Hawley DM, Altizer SM. 2011.** Disease ecology meets ecological immunology: understanding the links between organismal immunity and infection dynamics in natural populations. *Functional Ecology* 25(1):48–60 DOI 10.1111/j.1365-2435.2010.01753.x.
- Kao RR. 2010.** Networks and models with heterogeneous population structure in epidemiology. In: Ernesto E, Desmond JH, Maria F, Gian-Luca O, eds. *Network Science*. London: Springer, 51–84.

- Kennedy DA, Dwyer G. 2018. Effects of multiple sources of genetic drift on pathogen variation within hosts. *PLoS Biology* **16**(3):e2004444 DOI [10.1371/journal.pbio.2004444](https://doi.org/10.1371/journal.pbio.2004444).
- Kermack WO, McKendrick AG. 1927. Contributions to the mathematical theory of epidemics. *Proceedings of the Royal Society Edinburgh A* **115**(772):700–721 DOI [10.1098/rspa.1927.0118](https://doi.org/10.1098/rspa.1927.0118).
- Kermack WO, McKendrick AG. 1991. Contributions to the mathematical theory of epidemics-I. *Bulletin of Mathematical Biology* **53**(1-2):33–55 DOI [10.1016/S0092-8240\(05\)80040-0](https://doi.org/10.1016/S0092-8240(05)80040-0).
- Lavine JS, Poss M, Grenfell BT. 2008. Directly transmitted viral diseases: modeling the dynamics of transmission. *Trends in Microbiology* **16**(4):165–172 DOI [10.1016/j.tim.2008.01.007](https://doi.org/10.1016/j.tim.2008.01.007).
- Leclerc M, Doré T, Gilligan CA, Lucas P, Filipe JAN. 2014. Estimating the delay between host infection and disease (incubation period) and assessing its significance to the epidemiology of plant diseases. *PLOS ONE* **9**(1):e86568.
- Lindberg HM, McKean KA, Caraco T, Wang N. 2018. Within-host dynamics and random duration of pathogen infection: Implications for between-host transmission. *Journal of Theoretical Biology* **446**:137–148.
- Martinez-Bakker M, Helm B. 2015. The influence of biological rhythms on host-parasite interactions. *Trends in Ecology & Evolution* **30**(6):314–326 DOI [10.1016/j.tree.2015.03.012](https://doi.org/10.1016/j.tree.2015.03.012).
- McKenzie FE, Bossert WH. 2005. An integrated model of *Plasmodium falciparum* dynamics. *Journal of Theoretical Biology* **232**(3):411–426 DOI [10.1016/j.jtbi.2004.08.021](https://doi.org/10.1016/j.jtbi.2004.08.021).
- Metzger VT, Lloyd-Smith JO, Weinberger LS. 2011. Autonomous targeting of infectious superspreaders using engineered transmissible therapies. *PLOS Computational Biology* **7**(3):e1002015 DOI [10.1371/journal.pcbi.1002015](https://doi.org/10.1371/journal.pcbi.1002015).
- Mideo N, Acosta-Serrano A, Aebischer T, Brown MJF, Fenton A, Friman V-P, Restif O, Reece SE, Webster JP, Brown SP. 2013. Life in cells, hosts, and vectors: parasite evolution across scales. *Infection, Genetics and Evolution* **13**:344–347 DOI [10.1016/j.meegid.2012.03.016](https://doi.org/10.1016/j.meegid.2012.03.016).
- Mideo N, Alizon S, Day T. 2008. Linking within-and between-host dynamics in the evolutionary epidemiology of infectious diseases. *Trends in Ecology & Evolution* **23**(9):511–517 DOI [10.1016/j.tree.2008.05.009](https://doi.org/10.1016/j.tree.2008.05.009).
- Mideo N, Nelson WA, Reece SE, Bell AS, Read AF, Day T. 2011. Bridging scales in the evolution of infectious disease life histories: application. *Evolution* **65**(11):3298–3310.
- Moher D, Liberati A, Tetzlaff J, Altman DG. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine* **151**(4):264–269 DOI [10.7326/0003-4819-151-4-200908180-00135](https://doi.org/10.7326/0003-4819-151-4-200908180-00135).
- Murillo LN, Murillo MS, Perelson AS. 2013. Towards multiscale modeling of influenza infection. *Journal of Theoretical Biology* **332**:267–290 DOI [10.1016/j.jtbi.2013.03.024](https://doi.org/10.1016/j.jtbi.2013.03.024).
- Nguyen NM, Kien DTH, Tuan TV, Quyen NTH, Tran CN, Thi LV, Le Thi D, Nguyen HL, Farrar JJ, Holmes EC, Rabaa MA, Bryant JE, Nguyen TT, Nguyen HTC, Nguyen LTH, Pham MP, Nguyen HT, Luong TTH, Wills B, Nguyen CVV, Wolbers M, Simmons CP. 2013. Host and viral features of human dengue cases shape the population of infected and infectious *Aedes aegypti* mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America* **110**(22):9072–9077 DOI [10.1073/pnas.1303395110](https://doi.org/10.1073/pnas.1303395110).
- Nisbet RM, Gurney WSC. 1982. *Modelling fluctuating populations*. Chichester: John Wiley and Sons.
- Nowak M, May RM. 2000. *Virus dynamics: mathematical principles of immunology and virology*. New York: Oxford University Press.
- Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**(5255):1582–1586 DOI [10.1126/science.271.5255.1582](https://doi.org/10.1126/science.271.5255.1582).

- Pimenoff V, Houldcroft C, Rifkin R, Underdown S. 2018.** The role of aDNA in understanding the coevolutionary patterns of human sexually transmitted infections. *Genes* **9**(7):317 DOI [10.3390/genes9070317](https://doi.org/10.3390/genes9070317).
- Pollitt LC, MacGregor P, Matthews K, Reece SE. 2011.** Malaria and trypanosome transmission: different parasites, same rules? *Trends in Parasitology* **27**(5):197–203 DOI [10.1016/j.pt.2011.01.004](https://doi.org/10.1016/j.pt.2011.01.004).
- Reiner RC, Perkins TA, Barker CM, Niu T, Chaves LF, Ellis AM, George DB, Le Menach A, Pulliam JR, Bisanzio D, Buckee C, Chiyaka C, Cummings DAT, Garcia AJ, Gatton ML, Gething PW, Hartley DM, Johnston G, Klein EY, Michael E, Lindsay SW, Lloyd AL, Pigott DM, Reisen WK, Ruktanonchai N, Singh BK, Tatem AJ, Kitron U, Hay SI, Scott TW, Smith DL. 2013.** A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. *Journal of the Royal Society Interface* **10**(81):20120921 DOI [10.1098/rsif.2012.0921](https://doi.org/10.1098/rsif.2012.0921).
- Reperant LA, Kuiken T, Grenfell BT, Osterhaus ADME, Dobson AP. 2012.** Linking influenza virus tissue tropism to population-level reproductive fitness. *PLOS One* **7**(8):e43115.
- Saenz RA, Bonhoeffer S. 2013.** Nested model reveals potential amplification of an HIV epidemic due to drug resistance. *Epidemics* **5**(1):34–43 DOI [10.1016/j.epidem.2012.11.002](https://doi.org/10.1016/j.epidem.2012.11.002).
- Severins M. 2012.** A modeller's perspective on infection dynamics within and between hosts. *Veterinary Quarterly* **32**(3-4):169–178 DOI [10.1080/01652176.2012.738634](https://doi.org/10.1080/01652176.2012.738634).
- Smith DRM, Mideo N. 2017.** Modelling the evolution of HIV-1 virulence in response to imperfect therapy and prophylaxis. *Evolutionary Applications* **10**(3):297–309 DOI [10.1111/eva.12458](https://doi.org/10.1111/eva.12458).
- Stephenson JF, Young KA, Fox J, Jokela J, Cable J, Perkins SE. 2017.** Host heterogeneity affects both parasite transmission to and fitness on subsequent hosts. *Philosophical Transaction of the Royal Society of London B: Biological Sciences* **372**(1719):20160093.
- Takumi K, Franssen F, Fonville M, Grasset A, Vallée I, Boireau P, Teunis P, van der Giessen J. 2010.** Within-host dynamics of *Trichinella spiralis* predict persistent parasite transmission in rat populations. *International Journal for Parasitology* **40**:1317–1324.
- Theys K, Libin P, Pineda-Peña A-C, Nowé A, Vandamme A-M, Abecasis AB. 2018.** The impact of HIV-1 within-host evolution on transmission dynamics. *Current Opinion in Virology* **28**:92–101 DOI [10.1016/j.coviro.2017.12.001](https://doi.org/10.1016/j.coviro.2017.12.001).
- Tuncer N, Gulbudak H, Cannataro VL, Martcheva M. 2016.** Structural and practical identifiability issues of immuno-epidemiological vector–host models with application to rift valley fever. *Bulletin of Mathematical Biology* **78**:1796–1827.
- VanderWaal KL, Ezenwa VO. 2016.** Heterogeneity in pathogen transmission: mechanisms and methodology. *Functional Ecology* **30**(10):1606–1622 DOI [10.1111/1365-2435.12645](https://doi.org/10.1111/1365-2435.12645).
- Van Dorp CH, Van Boven M, De Boer RJ. 2014.** Immuno-epidemiological modeling of HIV-1 predicts high heritability of the set-point virus load, while selection for CTL escape dominates virulence evolution. *PLOS Computational Biology* **10**(12):e1003899 DOI [10.1371/journal.pcbi.1003899](https://doi.org/10.1371/journal.pcbi.1003899).
- Verenini M. 1983.** Parasitoid development within hosts infected by microorganisms. *Bollettino dell'Istituto di Entomologia della Università degli Studi di Bologna* **37**:225–265.
- Volz EM, Romero-Severson E, Leitner T. 2017.** Phylodynamic inference across epidemic scales. *Molecular Biology and Evolution* **34**(5):1276–1288 DOI [10.1093/molbev/msx077](https://doi.org/10.1093/molbev/msx077).
- Vrancken B, Rambaut A, Suchard MA, Drummond A, Baele G, Derdelinckx I, Van Wijngaerden E, Vandamme A-M, Van Laethem K, Lemey P. 2014.** The genealogical population dynamics of HIV-1 in a large transmission chain: bridging within and among host

evolutionary rates. *PLOS Computational Biology* **10(4)**:e1003505

[DOI 10.1371/journal.pcbi.1003505](https://doi.org/10.1371/journal.pcbi.1003505).

Wilks S, De Graaf M, Smith DJ, Burke DF. 2012. A review of influenza haemagglutinin receptor binding as it relates to pandemic properties. *Vaccine* **30(29)**:4369–4376

[DOI 10.1016/j.vaccine.2012.02.076](https://doi.org/10.1016/j.vaccine.2012.02.076).

Willem L, Verelst F, Bilcke J, Hens N, Beutels P. 2017. Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015). *BMC Infectious Diseases* **17(1)**:612 [DOI 10.1186/s12879-017-2699-8](https://doi.org/10.1186/s12879-017-2699-8).

Yeghiazarian L, Cumberland WG, Yang OO. 2013. A stochastic multi-scale model of HIV-1 transmission for decision-making: application to a MSM population. *PLOS ONE* **8(11)**:e70578

[DOI 10.1371/journal.pone.0070578](https://doi.org/10.1371/journal.pone.0070578).