ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Powell-Romero F, Wells K, Clark NJ. 2024 A systematic review and guide for using multi-response statistical models in co-infection research. *R. Soc. Open Sci.* **11**: 231589.
https://doi.org/10.1098/rsos.231589

Received: 20 October 2023
Accepted: 5 August 2024

Subject Category:
Mathematics

Subject Areas:
health and disease and epidemiology, ecology, statistics

Keywords:
co-infection, ecology, epidemiology, multi-response, multivariate, statistical modelling

Author for correspondence:
Francisca Powell-Romero
e-mail: francisca.powell@uq.net.au

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7425681.

THE ROYAL SOCIETY PUBLISHING

A systematic review and guide for using multi-response statistical models in co-infection research

Francisca Powell-Romero[1], Konstans Wells[2] and Nicholas J. Clark[1]

[1]School of Veterinary Science, The University of Queensland, 5391 Warrego Hwy, Gatton, Queensland 4343, Australia
[2]Department of Biosciences, Swansea University, Singleton Park, Swansea SA2 8PP, UK

FP-R, 0000-0001-9800-3100; KW, 0000-0003-0377-2463; NJC, 0000-0001-7131-3301

The simultaneous infection of organisms with two or more co-occurring pathogens, otherwise known as co-infections, concomitant infections or multiple infections, plays a significant role in the dynamics and consequences of infectious diseases in both humans and animals. To understand co-infections, ecologists and epidemiologists rely on models capable of accommodating multiple response variables. However, given the diversity of available approaches, choosing a model that is suitable for drawing meaningful conclusions from observational data is not a straightforward task. To provide clearer guidance for statistical model use in co-infection research, we conducted a systematic review to (i) understand the breadth of study goals and host–pathogen systems being pursued with multi-response models and (ii) determine the degree of crossover of knowledge among disciplines. In total, we identified 69 peer-reviewed primary studies that jointly measured infection patterns with two or more pathogens of humans or animals in natural environments. We found stark divisions in research objectives and methods among different disciplines, suggesting that cross-disciplinary insights into co-infection patterns and processes for different human and animal contexts are currently limited. Citation network analysis also revealed limited knowledge exchange between ecology and epidemiology. These findings collectively highlight the need for greater interdisciplinary collaboration for improving disease management.

1. Introduction

Simultaneous infection with multiple pathogens, or co-infections, holds significant importance in infectious disease research. The possible detrimental effects of co-infection on host health have been reported for many pathogens, including those from secondary bacterial infections during the 1918 influenza and the Coronavirus disease 2019 (COVID-19) pandemics [1,2]. Other studies have highlighted their significance by observing high frequencies of pathogen co-occurrence in both animals [3,4] and humans [5,6], suggesting that co-infections may be much more common than anticipated in epidemiological research and prevention programmes. As recognition of this has increased in recent years, so too has the availability of multi-pathogen occurrence data, offering researchers vast opportunities to characterize co-infections, provide new insights into disease dynamics and better explain pathogen co-occurrence, which are essential to improving surveillance and disease control.

At the host level, co-infection risk is a consequence of a complex interplay between the host response to infection with a given pathogen, and the invasion and persistence strategy of a second pathogen. While some pathogens do simply co-occur as a result of shared environmental affinities and stochasticity (that is, simultaneous but independent ecological fitting for pathogens invading host species), co-infection can also be a product of interspecific pathogen–pathogen and host–pathogen interactions [3]. These interactions may include direct or indirect processes within the host, such as through toxin-mediated competition [7,8] or immune modulation [9,10], which may alter the likelihood of co-infection depending on the order of infection [11–13]. Alternatively, co-infections may also occur through co-exposure or shared vectors [14]. As such, efforts to characterize pairwise associations between pathogens have risen [3,5]. Beyond pathogen–pathogen interactions, research has also expanded to account for within-host pathogen and microbial community dynamics [15,16], adding additional layers of complexity to understanding the drivers of co-infection.

Acknowledging the patterns and processes that drive co-infections throughout the diversity of host–parasite systems and the potential for interspecific pathogen–pathogen and pathogen–microbiota interactions is paramount, since they can have a cascading, and often detrimental, effect on the course of disease in a population. This can occur through various mechanisms such as by altering pathogen shedding patterns and increasing the risk of transmission [17,18], impairing host immunity and increasing host susceptibility to subsequent illnesses [19,20], or intensifying disease severity [21–23]. With the multifaced ways changes in climate drive pathogen spread and emergence, evaluating how environmental drivers may change the frequencies of co-infections across populations and their geographical distribution is also at the forefront of co-infection research [24–28].

Sound inference for understanding the patterns and processes of when and how co-infections occur from cross-sectional surveillance data requires statistical modelling approaches capable of handling multiple response variables. Fortunately, a wide array of methods is already available for these tasks. Some commonly used methods tailored around classifying infection status as multinomial response variables, for example, enable the comparison of groups by infection status in order to gain insights into variation in the relative frequencies of co-infections versus single infections [27,28]. Another important class of more recently proposed co-infection models include multivariate frameworks that account for conditional dependencies to deal with the joint occurrences of free-living species ('joint species distribution models') [29,30]. These have also been adapted to study pathogen community structures [3,31] and microbiome profiles [32,33], as well as to quantify interspecific associations [5,34].

Addressing specific study goals requires careful selection of methods that are well suited to the data and the intended purpose, as the choice will directly impact the ability to draw inferences or make predictions. Unfortunately, determining the suitability of methods is not always straightforward, and this lack of clarity may prevent researchers from making full use of their data's potential. One additional hinderance that may contribute to this is limited knowledge exchange about conceptual and computational advances between different disciplines, particularly ecology and epidemiology [6,35], which have likely developed largely independently from one another due to differences in host systems and study objectives. To provide clearer guidance for model use in co-infection research, this review seeks to (i) understand the breadth of study goals and host–pathogen systems being pursued with multi-response models and (ii) determine whether there is crossover of knowledge among disciplines. In doing so, we identify challenges and opportunities for expanding the use of these models.

2. Material and methods

2.1. Data collection

We conducted a systematic review of papers that used multi-response statistical methods to model co-infection patterns in animal and human populations (see search queries in electronic supplementary material, file 1). Systematic searches were conducted on 3 May and 9 August 2022, across four databases: Embase, PubMed, Web of Science and Scopus (figure 1). Duplicates, as well as papers classified as 'Conference Abstract', 'Conference Papers', 'Erratum', 'Preprint' and 'Review' in 'Type of Work', were removed from the initial search results, leaving 746 out of 1661 papers for screening. Papers were screened using seven criteria, which included studies that (i) were peer-reviewed or primary studies, (ii) utilized observational data, (iii) involved multiple species or variants of infective agents, (iv) involved two or more pathogens, (v) measured infection patterns (i.e. presence–absence or abundance of infections with multiple pathogens of host individuals or populations) as the response variable, (vi) used multi-response methods for analysis and (vii) focused on pathogens of animals or humans. Based on these criteria, 187 papers were retained after screening titles and abstracts only and a final selection of 69 papers after screening the full papers. A full list of these papers can be found in electronic supplementary material, file 3.

2.2. Data extraction and validation

Data relating to four study features were extracted from eligible papers: (i) model type, (ii) study goal(s), (iii) purpose, and (iv) study field indexing. Data on 'model type' were extracted to summarize the modelling approaches used to analyse co-infection data with multiple outcomes. This study feature encompassed six analytical strategies that were not mutually exclusive, and summarized whether multinomial, multivariate, permutation, classification, network and clustering approaches were used. 'Study goal' classified the primary aims of studies for modelling co-infections, grouping studies into 'association/interaction', 'risk factor', 'community structure', and 'spatial distribution'. The study feature describing the purpose of models was denoted by a single binary variable called 'prediction', where a value of 1 was used to indicate whether models were used for prediction in addition to inference, and a value of 0 to indicate inference only. 'Study field' classified studies as either 'Ecology' or 'Epidemiology' based on either journal indexing, study keywords, author affiliations or terminology within the paper. Definitions of variables for model types, study goals and purpose are described in table 1. In addition, a pairwise citation matrix, denoting which pair of studies cited each other, was computed as a proxy for levels of information sharing between included studies. Citation records for each study were manually obtained from the reference list of each study.

To verify the accuracy of the data, both the screening phase and the data extraction phase were externally validated by peers on 40 papers from the search results (approx. 5%), and 17 papers from the eligible subset (approx. 25%) of papers, respectively (see electronic supplementary material, file 2).

2.3. Identifying natural study clusters and assessing correlation with study discipline

A primary goal of our review was to assess whether co-infection studies tend to be clustered with respect to their study goals and methods used for analysis. To identify natural clusters of studies based on the features, principal component analysis (PCA) was conducted. This clustering analysis was performed using nine binary variables related to model type, study goals and purpose (i.e. indexing focal categories into single binary variables).

In addition to PCA, partitioning around medoids (PAM) was used to divide the studies into two groups, using Jaccard distance (d_J) as a dissimilarity metric to assess whether clustering resembles study field groupings. This metric was computed from the nine binary variables to measure the dissimilarity between any two given studies (A and B) such that

$$d_J(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$

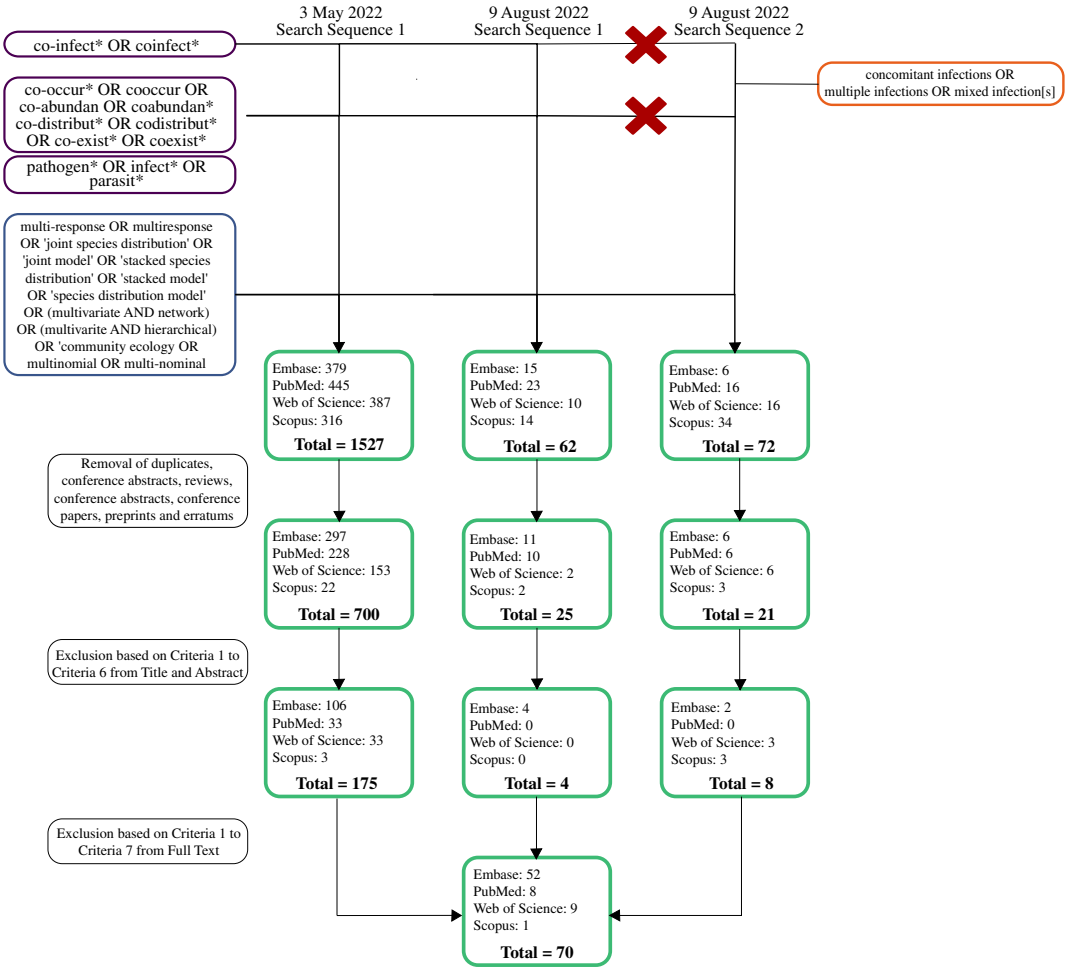


Figure 1. Flowchart of systematic search and filtering process for identifying co-infection studies using multi-response models. Terms to target co-infection studies are presented in purple boxes, and terms to target multi-response models are presented in the blue box. Red crosses indicate terms that were excluded in the second search sequence in order to avoid replicates from the first search when adding the new search terms (in the orange box). To obtain the final paper subset, duplicates were deleted, and papers outside of the scope were filtered first by title and abstracts and then by full texts. A total of 66 papers were kept for data extraction from the first search (3 May 2022), and an additional three papers were added after the second search (9 August 2022).

where the intersection ($\cap$) denotes the number of shared attributes and the union ($\cup$) denotes the number of attributes present in either A or B. The Jaccard distances with a value approaching 0 represent studies with shared attributes and those approaching 1 represent studies with no shared attributes. This dissimilarity metric was used in the PAM analysis to identify two clusters of minimal distances. The two clusters generated through this approach were compared to the manual indexing of papers using cross-tabulation and graphical visualization to assess similarities in groupings. All clustering analyses were conducted using the *vegan* [36], *cluster* [37] and *factoextra* [38] packages in R version 4.3.2.

2.4. Quantifying the association between model type and study goals

To better understand which models are favoured among studies with different goals, we quantified the relationships between study goals and model choice, irrespective of cluster assignment. To guide variable selection, pairwise correlations between study goals and other predictors were examined using the *corrplot* package [39] prior to conducting the analysis. The 'community structure' and 'risk factor' study goals were negatively correlated (−0.68), and therefore the 'risk factor' study goal was excluded from the final multivariable model. A positive correlation between 'prediction' as the study purpose and the 'spatial distribution' study goal was also found (0.63), and so we removed the variable for prediction from the final model. The final adjusted model included three study goal variables (community structure, pathogen

Table 1. Definitions of different categories of study features used to distinguish studies in terms of model type, study goals and purpose.

variable	definition
model type	
multinomial	models where the response variable is categorical and can include two or more categories. One category is used as a baseline reference category. This model assumes that each category is independent of each other and that the outcome is influenced only by the predictors in the model
multivariate	models where multiple response variables are estimated simultaneously, that is, the infection statuses of hosts for different pathogens are considered as separate response variables. These models account for correlations between the outcome variables and seek to understand the effect of predictors on the entire set of outcome variables
permutation	permutation models are methods that use randomization techniques to assess the significance of an observed pattern in the response variable. In the context of co-infections, they are often used to understand whether the observed frequencies of co-occurrences among different groups or across predictor gradients are likely to be due to chance
classification	classification models are models that use a statistical or machine learning algorithm to categorize data into predefined categories. These models use predictors to make decisions about which category an outcome is most likely to belong to
network	network/graphical models are models that are used to characterize the pairwise relationships between objects of interest. In the case of co-infections, the nodes or vertices in an acyclic graph would represent individual pathogen species or variants, and the edges the relationships with each other in terms of co-infection frequencies
clustering	clustering models are statistical models that are used to group data points based on similarities
study goals	
association/interaction	association/interaction refers to the goal of measuring and/or quantifying the relative association strength of different pathogens in terms of co-infection frequencies given in the cross-sectional data. Note that, associations refer to the patterns of co-occurrence that are observed, while interactions refer specifically to the effect that one pathogen has on another and vice versa. Papers with either objective have been grouped for the purpose of this study, but the terms are not interchangeable
risk factor	risk factor studies seek to measure and/or quantify the relationships between infection status (e.g. risk of co-infection versus single infection) and predictor such as host or environmental attributes characterizing all samples from the cross-sectional survey
community structure	studies seeking to measure and/or quantify the relationships between pathogen species or variants within a microbial community or understand the (meta)community composition in which a pathogen is situated, either within a single host or across an entire host population
distribution	distribution studies refer to those seeking to understand the geographical patterns of pathogens and co-infections
purpose	
prediction	binary variable to indicate whether the purpose of the model was for prediction in addition to inference (1) or inference only (0)
study field	
ecology	study field denoted by (i) journal indexing, (ii) paper keywords, (iii) author affiliations, (iv) related text or (v) relevant fields
epidemiology	study field denoted by (i) journal indexing, (ii) paper keywords, (iii) author affiliations, (iv) related text or (v) relevant fields

associations and spatial distribution), in addition to the variable denoting human (as opposed to animal) hosts. The outcomes of the model were classified into one of three categories: 'multinomial', 'multivariate' or 'other models'. One paper that included both multinomial and

multivariate models was excluded from the analysis (n = 68). The key differences between multinomial and multivariate models are described in Box 1.

Box 1. Key differences between multinomial and multivariate modelling frameworks.

	Multinomial	Multivariate
Modelling Framework	Categorization of response variables into discrete disease states	Consider multiple response variables simultaneously
Response Variable Representation	Discrete or nominal categories reflecting infection groups (e.g. no infection, mono-infection, and co-infection)	Multiple response variables, where each outcome represents the presence/absence or abundance of a particular pathogen
Parameterization	Estimates parameters as the probability or odds associated with each response category, compared to a baseline category	Estimates parameters as the variance-covariance or correlation structure between multiple response variables
Data Structure	Requires response variables that can be grouped into mutually exclusive categories (i.e. a host categorised with co-infection cannot be also categorised with mono-infection or no infection)	Can be used to model data with different formats or measurements of the response variables, such as presence/absences or abundance data for pathogens

The adjusted Bayesian multinomial logistic regression model was conducted using the *brms* and *cmdstanr* packages in R [40,41]. These models were built using the 'categorical' likelihood family. Priors with a normal distribution with a mean of 0 and a standard deviation of 1 were specified for the regression coefficients in the models. Odds ratios were calculated for each variable in all three models by taking the exponential of the coefficient and confidence as 95% highest posterior density confidence intervals. Additional information on variable selection and the model equations can be found in electronic supplementary material, file 4.

2.5. Citation analysis

To evaluate the extent of inter-disciplinary collaboration, a directed citation network analysis was conducted using citation record information. The cluster groups determined by PAM as well as the three variables for model type were included as node features in the analysis to explore whether they were linked to network topology. Eigenvector centrality was calculated as an individual-level centrality measure to quantify node importance, and modularity was calculated as a community-level metric to measure the degree of clustering within the network with respect to cluster membership and model membership. This analysis was conducted using the *igraph* [42,43] and *ggplot2* [44] packages in R.

All data and code used in the analysis of this study are openly available in the Zenodo repository (see link in data availability statement).

3. Results

A total of 69 papers published between 2005 and 2022 were eligible for the study [3,5,15,25–28,31,34,45–104]. A considerable proportion of studies were published in 2018 or later (39 papers), showcasing that modelling co-infection is an emerging research frontier. The details for each paper and their identification number as referred to throughout the results can be found in electronic supplementary material, file 3.

Most of these studies used cross-sectional data (approx. 88%), quantified pathogen occurrences using binary data (approx. 97%) and studied co-infections of human hosts (approx. 72%). The subset of eligible papers in this review encompassed a broad variety of study goals, host species and modelling approaches. For example, Mair *et al*. sought to identify associations between pathogens in human populations using only multivariate modelling approaches [75]. Aivelo & Norberg similarly sought to describe associations between pathogens as well as between pathogens and other commensal species in animal hosts [15]. In contrast, multinomial models were often used to identify risk factors of co-infections, such as was done in an animal host by Pigeault *et al*., or to predict the spatial distribution of co-infections [87], as was done

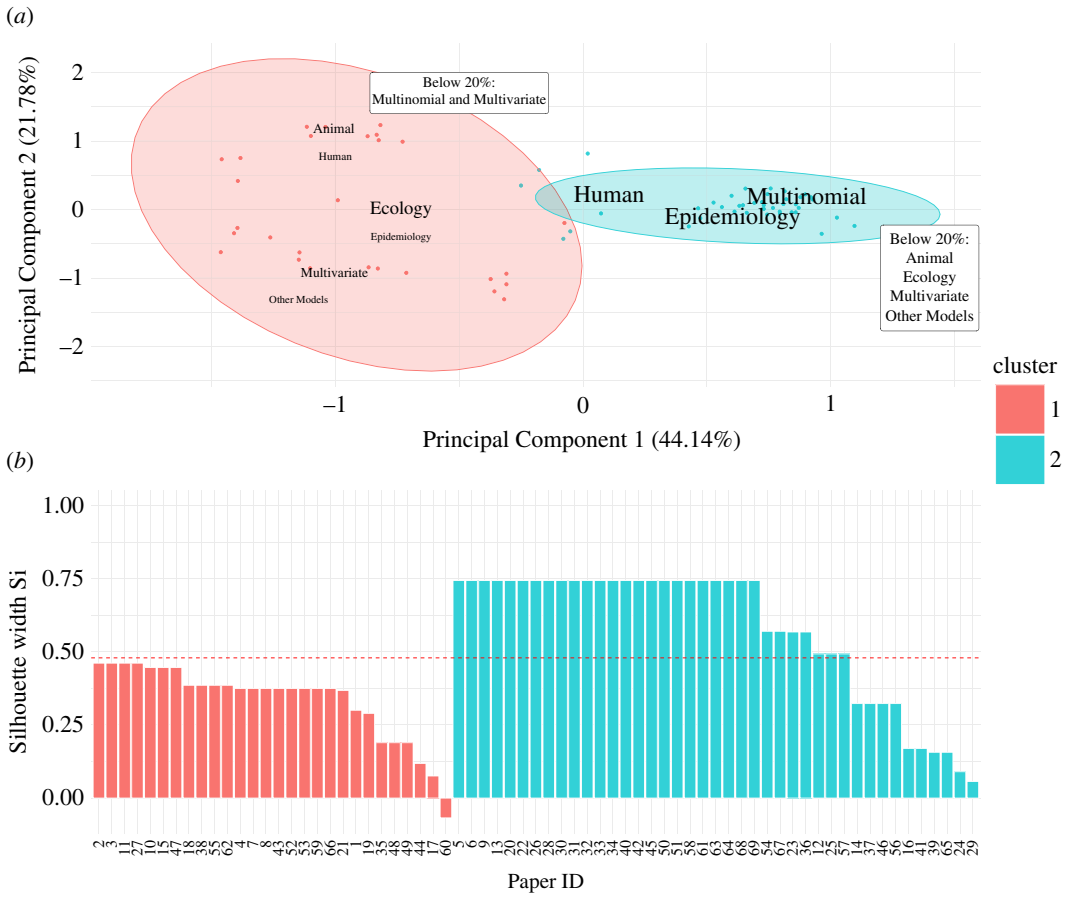


Figure 2. Principal component analysis (PCA) and silhouette plot. This combined plot showcases the cluster groupings obtained through partitioning around medoids (PAM) analysis applied to the dataset of 69 papers. The two clusters are depicted, with cluster 1 in orange and cluster 2 in blue. (*a*) PCA plot. This plot shows the two clusters reduced in a two-dimensional space, whereby each point represents a study according to its PC1 and PC2 values. The cumulative proportion explained by PC1 and PC2 is 65.92% (44.14% by PC1 and 21.78% by PC2). The text within each cluster indicates the host, indexed discipline and model type of studies within each cluster scaled by the proportion. Variables with a proportion below 20% are listed in the textbox beside each cluster. (*b*) Silhouette plot. This plot depicts the silhouette coefficient of papers as obtained through PAM as a measure of the quality of clustering. Each paper is depicted as a vertical bar, with the height determined by the silhouette width, indicating the paper's similarity to its assigned cluster. Positive silhouette width values approaching 1 indicate well-clustered papers, while negative values approaching −1 suggest that these papers may be better assigned to a different cluster. The average silhouette widths of clusters 1 and 2 are 0.33 and 0.58, indicating fair and good cluster matching respectively, while the total average silhouette width is 0.48, depicted by the dotted red line.

in human hosts by Soares Magalhães *et al*. [25]. Additionally, some studies explicitly sought to address multiple goals, such as Dallas *et al*., who sought to identify both associations between pathogens and risk factors using multivariate models in multiple animal host species [59]. Others used a combination of methods to address a single main goal, such as Choi *et al*., who used a combination of network models, permutation methods and clustering to understand the community structure of pathogens in human hosts [57].

Of the 69 co-infection studies, 40 (58.0%) used multinomial models, 17 (24.6%) used multivariate models and 13 (18.8%) used only other models. Of the 19 studies that modelled co-infections in animal hosts, 3 (15.8%) used multinomial models, 10 (52.6%) used multivariate models and 6 (31.6%) used other models. In contrast, in the 50 studies modelling human co-infections, 36 (72.0%) used multinomial models only, 6 (12.0%) used multivariate models only and 7 (14.0%) used other models. One study modelling human co-infections used both multinomial and multivariate models.

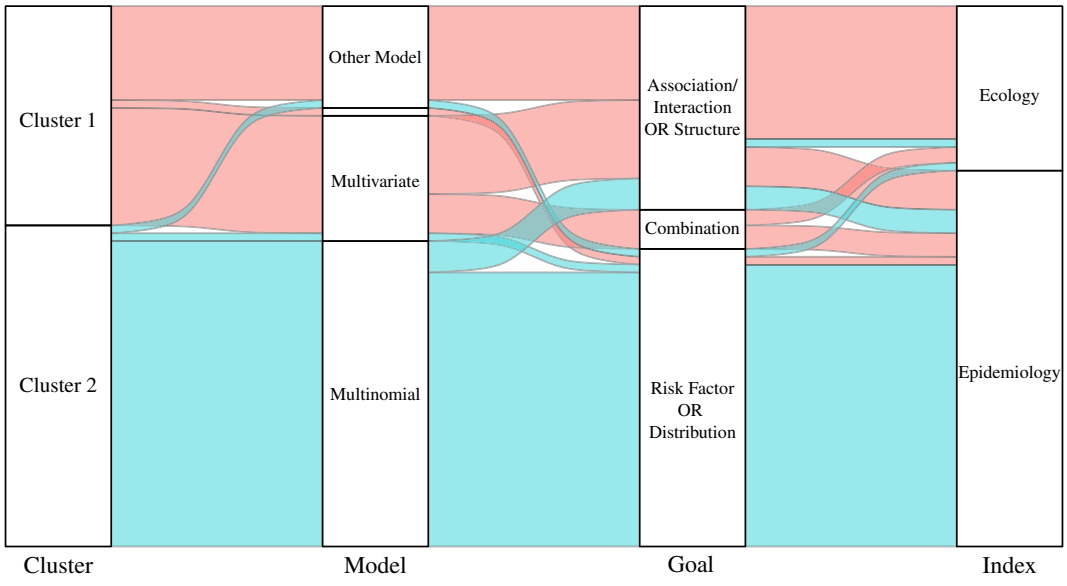


Figure 3. Parallel plot visualizing four study features of the 69 papers included in analysis. Alluvia (bands spanning the plot) are coloured by cluster, with cluster 1 depicted in orange and cluster 2 in blue. The widths of the alluvia correspond to the proportion of papers that belong to the strata on the right side of the band, given their placement on the axis to its left. The left axis represents the cluster assigned through partitioning around medoids (PAM), and contains two strata for cluster 1 and cluster 2. The middle-left axis represents model type, stratified into four categories: 'other models', 'multinomial & multivariate' (not labelled on plot due to size), 'multivariate and 'multinomial'. The middle-right axis represents study goals, grouped and stratified into three categories: (i) those where 'association/interaction' and/or 'community structure' were the only study goals, (ii) those where 'risk factor' and/or 'spatial distribution' were the only study goals, and (iii) those that combined goals from (i) and (ii). The right axis represents the manual indexing of papers into disciplines, with two strata for Ecology and Epidemiology.

3.1. Current co-infection research diverges into ecology and epidemiology clusters

PCA was conducted using variables relating to study goals, prediction and model types to identify natural clusters in the data. Principal components (PCs) 1 and 2 captured the variance in the data relatively well, with a cumulative proportion explained value of 65.92%, showcasing a strong divide in multivariate studies originating mostly from the field of 'Ecology' and multinomial studies mostly originating from the field of 'Epidemiology' (figure 2*a*). The strongest positive loading factors associated with PC1 with values above 0.2 were for the variables relating to 'multinomial' models (0.55) and the 'risk factor' study goal (0.47). The variables with the strongest negative factor loadings were the 'association/interaction' (−0.39) and 'community structure' (−0.39) study goals, along with those associated with the use of 'multivariate' models (−0.29) and 'other models' (−0.26). For PC2, the variables with the strongest positive loading factors above 0.2 were 'other models' (0.47) and the 'community structure' study goal (0.39), while the variables with the strongest negative factors were 'multivariate' models (−0.56) and the 'association/interaction' study goal (−0.46).

The resulting PAM clusters contained 28 (40.6%) and 41 (59.4%) papers, respectively. The average silhouette width between both clusters was 0.48 (0.33 for cluster 1 and 0.58 for cluster 2) (figure 2*b*). Notably, 67.9% of cluster 1 was comprised of 'Ecology' papers and 95.1% of cluster 2 was comprised of 'Epidemiology' papers according to their study field attributes (figure 2*a*). In cluster 1, 57.1% of studies dealt with animal hosts opposed to 7.3% of studies in cluster 2. In terms of model use, 53.6% of studies in cluster 1 used multivariate models, 3.6% used both multivariate and multinomial models, 42.9% used other models. In contrast, 95.1% of studies in cluster 2 used multinomial models, 2.4% used multivariate models and 2.4% used other models.

Table 2. The relative use of multinomial versus multivariate or other models for co-infection studies for different goals. Given as the estimates of odds ratios from Bayesian multinomial logistic regression model using study goals to predict model type with multinomial models as the baseline ($n = 39$). Estimates in bold represent 95% confidence intervals that do not include 0.

	multivariate ($n = 16$)		other models ($n = 13$)	
	OR estimate	95% CI	OR estimate	95% CI
study goals				
human host	**0.26**	**0.07, 0.95**	0.69	0.18, 2.62
associations/interaction	**17.04**	**5.42, 56.87**	1.44	0.40, 5.09
community structure	1.24	0.32, 4.65	**11.74**	**3.52, 41.66**
spatial distribution	1.12	0.27, 4.58	0.5	0.10, 2.33

3.2. Division between study goals and model types used in each discipline

With regard to model type, all papers using multinomial models only were assigned to cluster 2 (i.e. the 'Epidemiology' cluster), while 12 of the 13 papers (92.3%) that used other models were assigned to cluster 1 (i.e. the 'Ecology' cluster) (figure 3). Of the 16 papers that used multivariate models only, 15 (93.8%) were assigned to cluster 1. The one study that used multivariate models that were assigned to cluster 2 looked at risk factors or spatial distribution as a study goal. Of 26 papers that looked at either pathogen associations or community structure only, 22 (84.6%) were grouped into cluster 1. Thirty-seven of the 38 papers (97.4%) that looked either at risk factor identification or spatial distributions were assigned to cluster 2. All five studies combined both pathogen association or community structure goals with risk factor identification or spatial distribution analysis assigned to cluster 1. Only one paper used both multinomial and multivariate models, which was assigned to cluster 1 but was indexed as 'Epidemiology'. Of the remaining 47 papers indexed as 'Epidemiology', 39 (83.0%) were assigned to cluster 2. Of the 21 papers indexed as ecology, 19 papers were assigned to cluster 1 (90.5%). Of the two that were assigned to cluster 2, one used multinomial models to study risk factors or spatial distributions, and the other used multinomial models to understand associations or community structure.

Independent from the discipline clusters, the results from the Bayesian multinomial logistic regression model further suggest that the study goals are associated with model choice between the three model types (table 2). Specifically, studies that sought to detect pathogen associations were more likely to use multivariate models over multinomial models (OR = 17.04 [5.42, 56.87]). Studies that sought to understand pathogen or pathogen–microbial community structure were more likely to use other models compared to multinomial models (11.74 [3.52, 41.66]). No associations were found between host type and the spatial distribution study goal in the final adjusted model. Additionally, studies that involved human hosts were less likely to use multivariate models over multinomial models (OR = 0.26 [0.07, 0.95]).

3.3. Limited knowledge exchange between disciplines

Of the 69 papers in this review, 39 papers (56.5%) cited, or were cited by, at least one other paper in the eligible subset, which are shown in figure 4. Brooker *et al*. [27], Raso *et al*. [93], Clark *et al*. [5] and Soares Magalhães *et al*. [25] were the most cited or citing papers of others within this citation network and were thus most central to the citation network, as indicated by eigenvector centrality values of 1.00, 0.89, 0.87 and 0.82, respectively.

Within the citation network, 18 papers belonged to cluster 1, accounting for 64.3% of papers within the cluster, while the remaining 21 papers accounted for 51.2% of cluster 2. Of those papers that used multinomial methods, 21 papers (52.5%) were connected in the citation network, whereas 15 papers (88.2%) that used multivariate models, and four papers (30.8%) that only used other methods were connected in the citation network.

To evaluate network division based on cluster membership and model use, two modularity values were calculated for the subnetwork represented by the 39 papers. Moderate network clustering was found based on cluster membership (i.e. papers belonging to either cluster 1 or

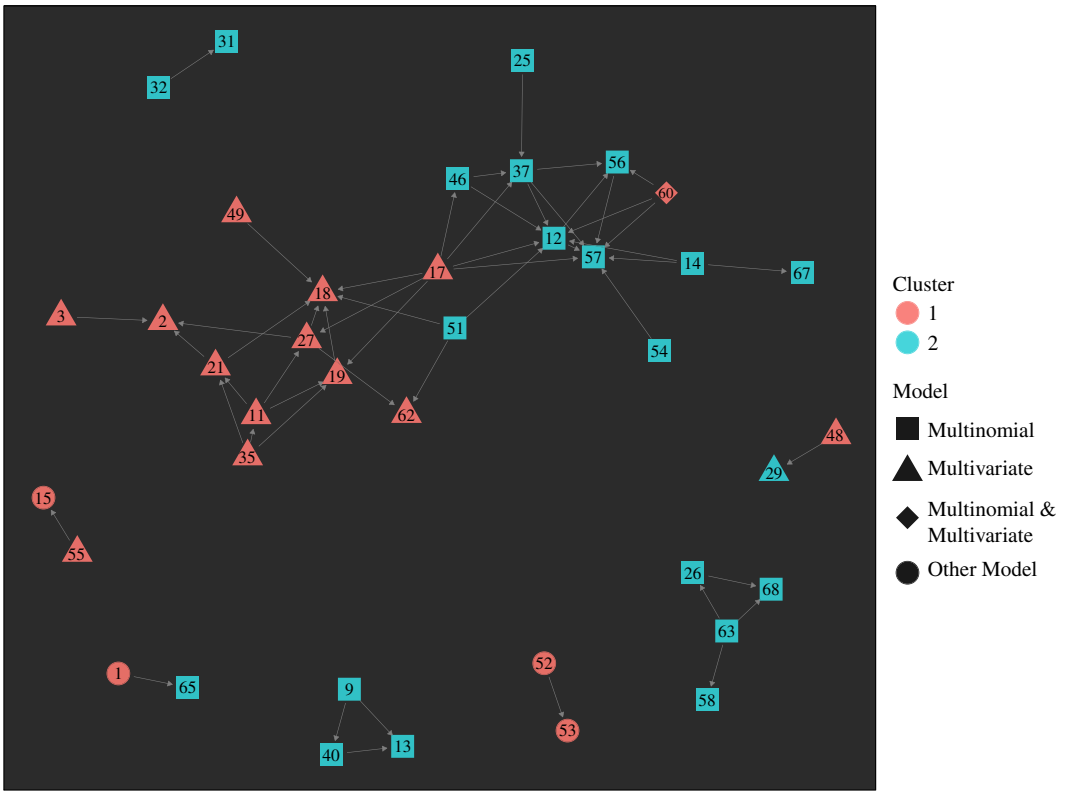


Figure 4. Citation network showing citations between those 39 out of the 69 papers connected by other studies by citations (30 papers excluded from graph due to no citations of other papers within the subset of papers for analysis). Nodes are characterized by model type, where squares represent multinomial models, triangles represent multivariate models, diamonds represent multinomial and multivariate models and circles represent other models. Numbers within nodes correspond to the paper ID numbers listed in electronic supplementary material, file 3. The colour of nodes corresponds to model type where orange nodes represent papers assigned to cluster 1, and blue represents papers assigned cluster 2 in the PAM analysis. The main network is comprised of 22 nodes, 12 from cluster 1 and 10 from cluster 2.

2) and model type use (i.e. multinomial, multivariate, multinomial and multivariate or other models) with modularity values of 0.30 and 0.35, respectively.

4. Discussion

While a multitude of models exists for modelling patterns in multiple response data, their applications in co-infection research for understanding different study goals and approaches used across disciplines have been unclear up to this point. This systematic review does not only suggest a strong link between specific study goals and the choices between the most commonly used model types, but also highlights a clear divergence in research focus among ecologists and epidemiologists, calling for better cross-disciplinary approaches in future research endeavours. Specifically, we found that the majority of studies that sought to understand the role of risk factors on co-infection, and those that sought to model the spatial distribution of co-infections were predominately linked to multinomial models being the most often used in studies from the field of epidemiology, whereas studies that sought to understand individual associations or interactions between pathogens were found to be linked to multivariate model use, and those that sought to understand the overall structure or composition of pathogen communities were linked to the use of other multi-response models, as shown in table 2. These findings reinforce the notion that specific and distinct goals and methods are studied between the two disciplines.

The results from the citation analysis further suggest that cross-disciplinary collaboration and exchange among epidemiologists and ecologists is currently limited, as summarized in Box 2.

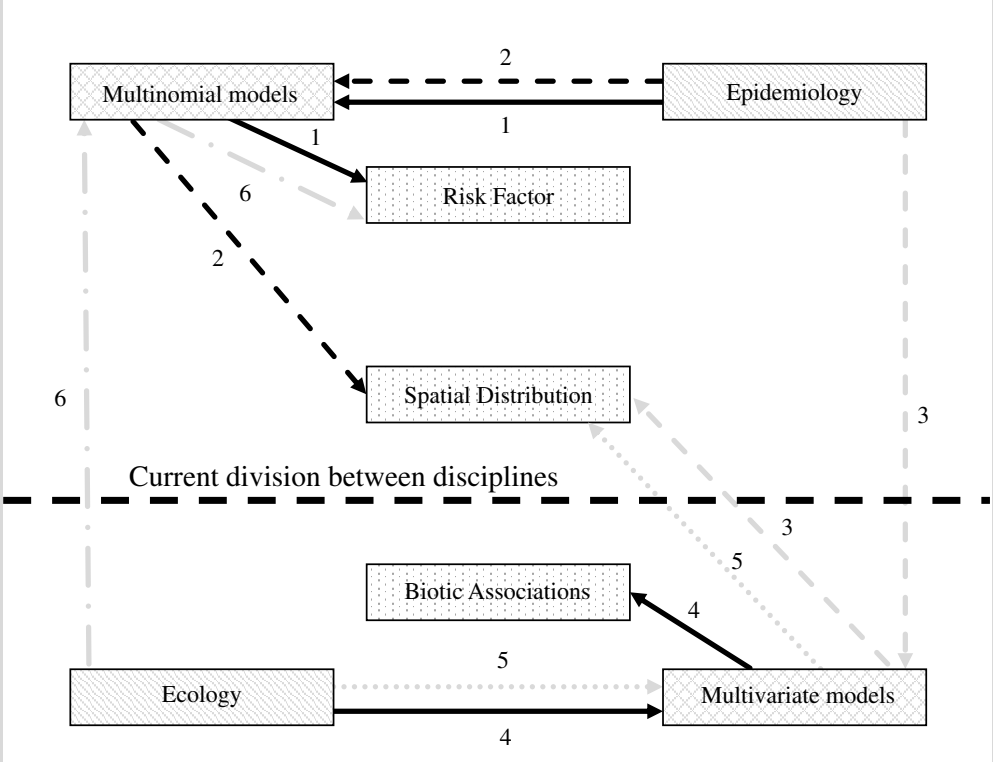
These findings corroborate previous calls for increased collaboration between the two disciplines, which have highlighted the limited overlap between work by epidemiologists and ecologists working on parasite communities in human and animal hosts [6] and multi-species infection dynamics [35]. Similar findings have previously been documented in other fields, with one study conducting a similar citation network noted the limited citating practices of researchers of work outside of the discipline [105], suggesting that the siloing of information within disciplines might be one barrier to cross-disciplinary collaboration. Moreover, the findings from our citation network also highlighted limited citations between studies that used multinomial models and multivariate models. These findings could suggest that a lack of awareness of different model types may drive study objectives, rather than model choice being driven by the study goals, marking a unique opportunity for the sharing of knowledge on these methodologies. This is consistent with the growing recognition of interdisciplinary research and transdisciplinary collaboration for better understanding and control of infectious diseases, with increasing momentum growing for collaborative approaches, such as the One Health framework, which recognizes the interconnected nature between the health of humans, animals and the environment [106]. By increasing the sharing of modelling practices between in ecology and epidemiology, researchers have the potential to instigate more profound discussions on their practical application, embracing a more holistic framework to contribute to a more resilient and integrated approach for disease management.

A clear preference among epidemiologists for multi-response modelling of co-infections, multinomial models are an effective tool for capturing risk factors associated with individual-level risk. As highlighted in this review, this was the most common study goal in studies indexed as epidemiological and is a more suitable choice over multi-response modelling approaches, such as multivariate models, for this task, because they enable easy quantification of the difference in disease risk attributable to specific exposures with commonly used statistical models. This is a particularly useful approach for evaluating behavioural risk factors associated with particular types of infections, such as sexually transmitted infections (STIs), such as, for example, Culbreth *et al*. who explored how engaging in drinking and sexual behaviours is associated with self-reported HIV–STI co-infection among youths [58], and Fotiou *et al*. who evaluated the association between drug injecting behaviours and HCV–HIV co-infection [64]. Alternatively, a similar modelling approach was used to describe the clinical profiles associated with co-infection in de Souza *et al*., who applied these methods for the purpose of associating symptoms for classifying viral and bacterial respiratory co-infections [60]. Because of the interpretability that these models offer, they also allow for comparison between multiple models as well, such as Binka *et al*. who built multinomial models for different ethnic groups to compare the difference in risks associated with co-infection with hepatitis B, hepatitis C and HIV co-infection [51].

Besides risk factor identification, another study goal commonly pursued by epidemiologists included modelling the spatial distributions of co-infections. Among these studies, multinomial models were also the preferred method for understanding the geographical spread of disease. This was especially popular among researchers looking at helminth co-infections in human hosts, making up nine of the 12 studies that sought to model spatial distributions using multinomial models [25–27,55,63,90,92,93,102]. These methods are effective at capturing the co-infection patterns and detecting locations associated with no infections, mono-infections and co-infections, allowing researchers to draw inferences from the data and identify hotspots areas. Of these nine studies, seven also sought to predict areas associated with higher infection risk at the population level using these models [25–27,55,63,92,93]. Only two of the 12 studies applied multivariate models to inform predictions [5,96]. Interestingly, one of these studies used both a multinomial model and two-component models: an independent component model that considered probability of single-infection status, and a shared component model, which considers conditional dependencies between diseases through location-specific shared components [96]. While the study found the multinomial model to outperform the other models across most performance metrics, it was not able to be applied to single-disease survey data. In these cases, the shared component model showed the best performance over the individual models. This, therefore, represents an underexplored opportunity for epidemiologists to leverage existing data from single-disease surveys to formulate better predictions on disease risk by accounting for potential conditional dependencies between multiple pathogen occurrences.

Clark *et al*., who utilized a multivariate model for predicting spatial distributions, also demonstrate the utility of a multivariate network model called conditional random fields, revealing a comparable performance for predictions of human helminth infections at the individual level between the model and single-parasite gradient boosted machine models, and higher predictive accuracy at the school level [5]. These findings highlight the potential of using these models as a way of accounting for conditional dependencies within the models to generate accurate predictions. While inconsistencies

Box 2. Summary of modelling frameworks including their current use in ecology and epidemiology, and opportunities for cross-disciplinary collaboration.



Current & Future* Avenues for Multi -Response Model Use and Cross-Disciplinary Collaboration

1. The most common modelling framework used by epidemiologists for co-infection modelling are multinomial models. They are commonly applied to understand risk factors associated with disease states, and are an appropriate model choice for this purpose.

2. Epidemiologists also use multinomial models to predict spatial distributions of disease states, particularly to help guide infection control strategies.

3. While the second pathway is not necessarily a flawed approach, epidemiologists could benefit from multivariate alternatives, which account for co-occurrence dependencies and may provide more precise spatial predictions. This application would involve crossing the existing disciplinary boundary.

4. Ecologists typically rely on multivariate models, which they use primarily to understand statistical associations between co-occurring pathogens.

5. However, few ecological co-infection studies included spatial data. Including this information can lead to more accurate inference as well as improve the practical utility of findings. This application would involve crossing the existing disciplinary boundary.

6. Discretising co-infections into carefully considered discrete states could be beneficial for ecologists to consider how predictors modify risk of one category versus another, to ask more targeted questions about how and why ecological communities change over space and time.

*Current uses and disciplinary divisions are represented in black, while potential future avenues are represented in grey.

in predictive performances of different multivariate models have been noted [107–116], these models have been shown to often lead to more accurate predictions of the co-occurrence of free-living species and plants, particularly for rare species [110,111,115,116]. These models are therefore also worth considering for modelling the geographical distributions of co-infections, either in place of multinomial models for predicting high-risk areas, or in addition to multinomial models when also seeking to make inferences about risk factors. Moreover, these methods provide a useful avenue for utilizing big data with large numbers of pathogens (and microbiomes) that might be oversimplified if grouped into a few broader categories, as would be feasible for multinomial modelling.

Accounting for conditional dependencies between pathogens using multivariate modelling was included more frequently in studies indexed as ecological; however, these models were primarily used to measure pathogen–pathogen associations as opposed to predicting spatial distributions of co-infections. One popular method for doing this was the hierarchical modelling of species communities (HMSC) framework [15,46,53,59,73], a joint species distribution model (JSDM) that accounts for interspecific associations within a residual variance–covariance matrix [117]. While this method has been previously applied to incorporate spatial data for free-ranging organisms [118,119], our review highlights that these models have primarily been used to quantify interactions between pathogens (i.e. the co-occurrence of pathogens in host individuals, regardless of the spatial context of the infections). Only one study accounted for conditional dependencies by including information about the spatial context of sampled host individuals [46], while no study applied this model to predict the geographical distributions of pathogen co-occurrences. Expanding the study goals and including spatial parameters within these models offer ecologists the opportunity to improve the utility of the models for practical applications and create an overlap with the research interests of epidemiologists, that may subsequently help improve collaboration practices between the two disciplines.

In relation to the fourth study goal identified in this review, understanding pathogen community composition, the findings suggest that other methods, including permutation analyses, network analyses, clustering and classification methods are preferred by researchers over multinomial or multivariate models. Notably, of the 18 studies that sought to understand pathogen community structure, nine studies also included commensal species in addition to pathogenic species [15,46,52,57,78,80,83,88,101]. One such example is by Bouillaguet *et al*., who sought to characterize the microbiota associated with apical periodontitis using a combination of network, permutation and clustering methods [52]. Given that the inclusion of these commensals introduces an additional layer of complexity to understanding co-infections, the choice of methods in these instances may be reflective of the number of species (i.e. response variables) included in the analyses. However, while we have included these studies within the scope of this review to showcase the breadth of co-infection research that is being pursued, further investigation into this subset of studies would be required to provide guidance on model suitability for the task and discipline-specific recommendations. Moreover, while the binary indexing categories for disciplines as 'Ecology' and 'Epidemiology' were utilized here to describe two common fields concerned with infectious diseases for the purpose of this review, allowing generalizability of the findings within these two disciplines, these classifications may be overly simplistic to draw accurate conclusions regarding this subset of studies looking at community structure and may be better represented under another classification such as 'Microbiology'.

While the findings from this study have the potential to serve as a guide for the use of multi-response models for co-infection research and future avenues for cross-disciplinary collaboration, the limitations of this study should also be noted. First, this study focuses on the use of statistical models for modelling observational co-infection data and does not consider mechanistic models within its scope. With regards to the analyses conducted in the review, it should be noted that the odds ratio values from the Bayesian multinomial logistic regression models are not reliable quantitative measures but are deemed fit for purpose here to showcase the divergence in model applications. In terms of the citation network, it should also be noted that there is bias regarding the number of citations pertaining to each study, and in particular studies containing no citations, given that more recent publications have had less time to be cited and therefore the number of citations is likely to be correlated to the year of publication. Thus, cross-collaboration between disciplines within the network cannot be inferred for more recent studies from the citation network alone. Nevertheless, the citation analysis included in this study provides a good indication of trends occurring with earlier work. Lastly, while our systematic approach aimed at providing a comprehensive overview of relevant literature, the degree of variability in terminology for various modelling approaches means our search is unlikely to be exhaustive, particularly where methods did not include multinomial and multivariate models.

5. Conclusion

Selecting statistical models for data analysis can be an arduous task and requires an understanding of how core research questions can be supported by suitable models fit for purpose. The strong divide in both study goals and statistical model choice in studies dealing with co-infections from epidemiological and ecological perspectives highlights that improving communication and sharing of ideas across disciplines is needed if we aim to understand different model systems to an equal extent and aim to synthesize ecological and epidemiological insights together into One Health solutions.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. Data and supporting code to replicate analysis have been made available from the Zenodo repository [120].

Supplementary material is available online [121].

Declaration of AI use. ChatGPT (GPT–3.5) by OpenAI was used minimally as an aid during the text-editing process whereby particular words may have been selected to improve the clarity of specific phrases. It was also used as a search engine to identify suitable code, such as relevant packages, and for troubleshooting some errors during the statistical analysis. All ideas, systematic search, filtering, data collection, data analysis and manuscript drafts were performed and produced by the authors of the study.

Authors' contributions. F.P.-R.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, visualization, writing—original draft, writing—review and editing; K.W.: data curation, methodology, supervision, validation, writing—review and editing; N.J.C.: conceptualization, data curation, funding acquisition, methodology, supervision, validation, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This project was supported by an ARC Discovery Early Career Researcher Award (DE210101439) and The Royal Society (RGS\R2\222152).

Acknowledgements. Thank you to Weronika Konarska, Uttara Kennedy and Thomas Callaghan for their important contributions to the data extraction and validation process, and to Professor Ricardo Soares Magalhães and Dr Nicholas Fountain-Jones for their valuable suggestions during the conceptualization of this study and the visualization of the results.

References

1. Walters KA *et al*. 2016 1918 pandemic influenza virus and *Sreptococcus pneumoniae* co-infection results in activation of coagulation and widespread pulmonary thrombosis in mice and humans. *J. Pathol.* **238**, 85–97. (doi:10.1002/path.4638)
2. Patton MJ *et al*. 2023 COVID-19 bacteremic co-infection is a major risk factor for mortality, ICU admission, and mechanical ventilation. *Crit. Care* **27**, 34. (doi:10.1186/s13054-023-04312-0)
3. Clark NJ, Wells K, Dimitrov D, Clegg SM. 2016 Co-infections and environmental conditions drive the distributions of blood parasites in wild birds. *J. Anim. Ecol.* **85**, 1461–1470. (doi:10.1111/1365-2656.12578)
4. Moutailler S *et al*. 2016 Co-infection of ticks: the rule rather than the exception. *PLoS Negl. Trop. Dis.* **10**, e0004539. (doi:10.1371/journal.pntd.0004539)
5. Clark NJ *et al*. 2020 Parasite associations predict infection risk: incorporating co-infections in predictive models for neglected tropical diseases. *Parasit. Vect.* **13**, 138. (doi:10.1186/s13071-020-04016-2)
6. Petney TN, Andrews RH. 1998 Multiparasite communities in animals and humans: frequency, structure and pathogenic significance. *Int. J. Parasitol.* **28**, 377–393. (doi:10.1016/s0020-7519(97)00189-6)
7. Schoener TW. 1983 Field experiments on interspecific competition. *Am. Nat.* **122**, 240–285. (doi:10.1086/284133)
8. Bashey F. 2015 Within-host competitive interactions as a mechanism for the maintenance of parasite diversity. *Phil. Trans. R. Soc. B* **370**, 20140301. (doi:10.1098/rstb.2014.0301)
9. Smith AM, Adler FR, Ribeiro RM, Gutenkunst RN, McAuley JL, McCullers JA, Perelson AS. 2013 Kinetics of coinfection with influenza a virus and *Streptococcus pneumoniae*. *PLoS Pathog.* **9**, e1003238. (doi:10.1371/journal.ppat.1003238)
10. Ezenwa VO, Etienne RS, Luikart G, Beja-Pereira A, Jolles AE. 2010 Hidden consequences of living in a wormy world: nematode-induced immune suppression facilitates tuberculosis invasion in African buffalo. *Am. Nat.* **176**, 613–624. (doi:10.1086/656496)
11. Clay PA, Dhir K, Rudolf VHW, Duffy MA. 2019 Within-host priority effects systematically alter pathogen coexistence. *Am. Nat.* **193**, 187–199. (doi:10.1086/701126)
12. Clay PA, Cortez MH, Duffy MA, Rudolf VHW. 2019 Priority effects within coinfected hosts can drive unexpected population-scale patterns of parasite prevalence. *Oikos* **128**, 571–583. (doi:10.1111/oik.05937)
13. Hoverman JT, Hoye BJ, Johnson PTJ. 2013 Does timing matter? How priority effects influence the outcome of parasite interactions within hosts. *Oecologia* **173**, 1471–1480. (doi:10.1007/s00442-013-2692-x)

14. Hakimi H, Sarani A, Takeda M, Kaneko O, Asada M. 2019 Epidemiology, risk factors, and co-infection of vector-borne pathogens in goats from Sistan and Baluchestan province, Iran. *PLoS One* **14**, e0218609. (doi:10.1371/journal.pone.0218609)
15. Aivelo T, Norberg A. 2018 Parasite-microbiota interactions potentially affect intestinal communities in wild mammals. *J. Anim. Ecol.* **87**, 438–447. (doi:10.1111/1365-2656.12708)
16. Fountain-Jones NM, Packer C, Jacquot M, Blanchet FG, Terio K, Craft ME. 2018 Chronic infections can shape epidemic exposure: pathogen co-occurrence networks in the Serengeti lions. *bioRxiv*. (doi:10.1101/370841)
17. Peel AJ *et al*. 2019 Synchronous shedding of multiple bat paramyxoviruses coincides with peak periods of hendra virus spillover. *Emerg. Microbes Infect.* **8**, 1314–1323. (doi:10.1080/22221751.2019.1661217)
18. Lass S, Hudson PJ, Thakar J, Saric J, Harvill E, Albert R, Perkins SE. 2013 Generating super-shedders: co-infection increases bacterial load and egg production of a gastrointestinal helminth. *J. R. Soc. Interface* **10**, 20120588. (doi:10.1098/rsif.2012.0588)
19. Su Z, Segura M, Morgan K, Loredo-Osti JC, Stevenson MM. 2005 Impairment of protective immunity to blood-stage malaria by concurrent nematode infection. *Infect. Immun.* **73**, 3531–3539. (doi:10.1128/IAI.73.6.3531-3539.2005)
20. Broughton H, Govender D, Serrano E, Shikwambana P, Jolles A. 2021 Equal contributions of feline immunodeficiency virus and coinfections to morbidity in African lions. *Int. J. Parasitol. Parasites Wildl.* **16**, 83–94. (doi:10.1016/j.ijppaw.2021.07.003)
21. Graham AL, Lamb TJ, Read AF, Allen JE. 2005 Malaria-filaria coinfection in mice makes malarial disease more severe unless filarial infection achieves patency. *J. Infect. Dis.* **191**, 410–421. (doi:10.1086/426871)
22. Risco D *et al*. 2014 Severity of bovine tuberculosis is associated with co-infection with common pathogens in wild boar. *PLoS One* **9**, e110123. (doi:10.1371/journal.pone.0110123)
23. Zhang SX *et al*. 2016 Impact of co-infections with enteric pathogens on children suffering from acute diarrhea in southwest China. *Infect. Dis. Poverty* **5**, 64. (doi:10.1186/s40249-016-0157-2)
24. Ruberanziza E *et al*. 2019 Mapping soil-transmitted helminth parasite infection in Rwanda: estimating endemicity and identifying at-risk populations. *Trop. Med. Infect. Dis.* **4**, 93. (doi:10.3390/tropicalmed4020093)
25. Soares Magalhães RJ, Biritwum NK, Gyapong JO, Brooker S, Zhang Y, Blair L, Fenwick A, Clements ACA. 2011 Mapping helminth co-infection and co-intensity: geostatistical prediction in Ghana. *PLoS Negl. Trop. Dis.* **5**, e1200. (doi:10.1371/journal.pntd.0001200)
26. Owada K, Lau CL, Leonardo L, Clements ACA, Yakob L, Nielsen M, Carabin H, Soares Magalhães RJ. 2018 Spatial distribution and populations at risk of *A. lumbricoides* and *T. trichiura* co-infections and infection intensity classes: an ecological study. *Parasit. Vectors* **11**, 535. (doi:10.1186/s13071-018-3107-y)
27. Brooker S, Clements ACA. 2009 Spatial heterogeneity of parasite co-infection: determinants and geostatistical prediction at regional scales. *Int. J. Parasitol.* **39**, 591–597. (doi:10.1016/j.ijpara.2008.10.014)
28. Bassa FK *et al*. 2022 Prevalence of *Schistosoma* mono- and co-infections with multiple common parasites and associated risk factors and morbidity profile among adults in the Taabo health and demographic surveillance system, south-central Côte d'Ivoire. *Infect. Dis. Poverty* **11**, 3. (doi:10.1186/s40249-021-00925-1)
29. Pollock LJ, Tingley R, Morris WK, Golding N, O'Hara RB, Parris KM, Vesk PA, McCarthy MA. 2014 Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (JSDM). *Methods Ecol. Evol.* **5**, 397–406. (doi:10.1111/2041-210X.12180)
30. Pichler M, Hartig F. 2021 A new joint species distribution model for faster and more accurate inference of species associations from big community data. *Methods Ecol. Evol.* **12**, 2159–2173. (doi:10.1111/2041-210X.13687)
31. Clark NJ, Wells K, Lindberg O. 2018 Unravelling changing interspecific interactions across environmental gradients using Markov random fields. *Ecology* **99**, 1277–1283. (doi:10.1002/ecy.2221)
32. Fountain-Jones NM *et al*. 2020 Microbial associations and spatial proximity predict North American moose (*Alces alces*) gastrointestinal community composition. *J. Anim. Ecol.* **89**, 817–828. (doi:10.1111/1365-2656.13154)
33. Connor N, Barberán A, Clauset A. 2017 Using null models to infer microbial co-occurrence networks. *PLoS One* **12**, e0176751. (doi:10.1371/journal.pone.0176751)
34. Stutz WE, Blaustein AR, Briggs CJ, Hoverman JT, Rohr JR, Johnson PTJ. 2018 Using multi-response models to investigate pathogen coinfections across scales: insights from emerging diseases of amphibians. *Methods Ecol. Evol.* **9**, 1109–1120. (doi:10.1111/2041-210X.12938)
35. Johnson PTJ, de Roode JC, Fenton A. 2015 Why infectious disease research needs community ecology. *Science* **349**, 1259504. (doi:10.1126/science.1259504)
36. Oksanen J *et al*. 2022 Vegan: community ecology package. See https://CRAN.R-project.org/package=vegan.
37. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. 2022 cluster: cluster analysis basics and extensions. See https://CRAN.R-project.org/package=cluster.
38. Kassambara A, Mundt F. 2020 factoextra: extract and visualize the results of multivariate data analyses. See https://CRAN.R-project.org/package=factoextra.
39. Wei T, Simko V. 2021 R package 'corrplot': visualization of a correlation matrix. See https://github.com/taiyun/corrplot.
40. Bürkner PC. 2017 Brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28. (doi:10.18637/jss.v080.i01)
41. Gabry J, Češnovar R, Johnson A. 2022 Cmdstanr: R interface to 'cmdstan'. See https://mc-stan.org/cmdstanr/.
42. Csárdi G, Nepusz T, Traag V, Horvát S, Zanini F, Noom D, Müller K. 2024 Igraph: network analysis and visualization in R. (doi:10.5281/zenodo.7682609). See https://CRAN.R-project.org/package=igraph.
43. Csardi G, Nepusz T. 2006 The igraph software package for complex network research. *InterJornal Complex Syst.* **1695**
, 1–9.

44. Wickham H. 2016 *ggplot2: elegant graphics for data analysis*. New York, NY: Springer-Verlag.
45. Abbate JL, Ezenwa VO, Guégan JF, Choisy M, Nacher M, Roche B. 2018 Disentangling complex parasite interactions: protection against cerebral malaria by one helminth species is jeopardized by co-infection with another. *PLoS Negl. Trop. Dis.* **12**, e0006483. (doi:10.1371/journal.pntd.0006483)
46. Aivelo T, Norberg A, Tschirren B. 2019 Bacterial microbiota composition of *Ixodes ricinus* ticks: the role of environmental variation, tick characteristics and microbial interactions. *PeerJ* **7**, e8217. (doi:10.7717/peerj.8217)
47. Álvarez-Mendizábal P, Villalobos F, Rodríguez-Hernández K, Hernández-Lara C, Rico-Chávez O, Suzán G, Chapa-Vargas L, Santiago-Alarcon D. 2021 Metacommunity structure reveals that temperature affects the landscape compositional patterns of avian malaria and related haemosporidian parasites across elevations. *Acta Oecol.* **113**, 103789. (doi:10.1016/j.actao.2021.103789)
48. Anabire NG, Aryee PA, Abdul-Karim A, Abdulai IB, Quaye O, Awandare GA, Helegbe GK. 2019 Prevalence of malaria and hepatitis B among pregnant women in northern Ghana: comparing RDTs with PCR. *PLoS One* **14**, e0210365. (doi:10.1371/journal.pone.0210365)
49. Beechler BR *et al*. 2019 Bovine tuberculosis disturbs parasite functional trait composition in African buffalo. *Proc. Natl Acad. Sci. USA* **116**, 14645–14650. (doi:10.1073/pnas.1903674116)
50. Bertolino S, Hofmannová L, Girardello M, Modry D. 2010 Richness, origin and structure of an *Eimeria* community in a population of eastern cottontail (*Sylvilagus floridanus*) introduced into Italy. *Parasitology* **137**, 1179–1186. (doi:10.1017/S0031182009992095)
51. Binka M *et al*. 2021 Differences in risk factors for hepatitis B, hepatitis C, and human immunodeficiency virus infection by ethnicity: a large population-based cohort study in British Columbia, Canada. *Int. J. Infect. Dis.* **106**, 246–253. (doi:10.1016/j.ijid.2021.03.061)
52. Bouillaguet S, Manoil D, Girard M, Louis J, Gaïa N, Leo S, Schrenzel J, Lazarevic V. 2018 Root microbiota in primary and secondary apical periodontitis. *Front. Microbiol.* **9**, 2374. (doi:10.3389/fmicb.2018.02374)
53. Brian JI, Aldridge DC. 2021 Abundance data applied to a novel model invertebrate host shed new light on parasite community assembly in nature. *J. Anim. Ecol.* **90**, 1096–1108. (doi:10.1111/1365-2656.13436)
54. Butt ZA *et al*. 2017 A syndemic approach to assess the effect of substance use and social disparities on the evolution of HIV/HCV infections in British Columbia. *PLoS One* **12**, e0183609. (doi:10.1371/journal.pone.0183609)
55. Chammartin F *et al*. 2014 Bayesian risk mapping and model-based estimation of *Schistosoma haematobium-Schistosoma mansoni* co-distribution in Côte d'Ivoire. *PLoS Negl. Trop. Dis.* **8**, e3407. (doi:10.1371/journal.pntd.0003407)
56. Chaturvedi AK, Myers L, Hammons AF, Clark RA, Dunlap K, Kissinger PJ, Hagensee ME. 2005 Prevalence and clustering patterns of human papillomavirus genotypes in multiple infections. *Cancer Epidemiol. Biomarkers Prev.* **14**, 2439–2445. (doi:10.1158/1055-9965.EPI-05-0465)
57. Choi Y *et al*. 2019 Co-occurrence of anaerobes in human chronic wounds. *Microb. Ecol.* **77**, 808–820. (doi:10.1007/s00248-018-1231-z)
58. Culbreth R, Swahn MH, Salazar LF, Ametewee LA, Kasirye R. 2020 Risk factors associated with HIV, sexually transmitted infections (STI), and HIV/STI co-infection among youth living in the slums of Kampala, Uganda. *AIDS Behav.* **24**, 1023–1031. (doi:10.1007/s10461-019-02444-5)
59. Dallas TA, Laine AL, Ovaskainen O. 2019 Detecting parasite associations within multi-species host and parasite communities. *Proc. R. Soc. B* **286**, 20191109. (doi:10.1098/rspb.2019.1109)
60. de Souza PG, Cardoso AM, Sant'Anna CC, March M de FBP. 2018 Acute lower respiratory infection in Guarani indigenous children, Brazil. *Rev. Paul. Pediatr.* **36**, 123–131. (doi:10.1590/1984-0462/;2018;36;2;00017)
61. Duffy FJ, Thompson EG, Scriba TJ, Zak DE. 2019 Multinomial modelling of TB/HIV co-infection yields a robust predictive signature and generates hypotheses about the HIV+TB+ disease state. *PLoS One* **14**, e0219322. (doi:10.1371/journal.pone.0219322)
62. Castro Sanchez AY, Aerts M, Shkedy Z, Vickerman P, Faggiano F, Salamina G, Hens N. 2013 A mathematical model for HIV and hepatitis C co-infection and its assessment from a statistical perspective. *Epidemics* **5**, 56–66. (doi:10.1016/j.epidem.2013.01.002)
63. Forrer A, Khieu V, Schär F, Vounatsou P, Chammartin F, Marti H, Muth S, Odermatt P. 2018 *Strongyloides stercoralis* and hookworm co-infection: spatial distribution and determinants in Preah Vihear province, Cambodia. *Parasit. Vectors* **11**, 33. (doi:10.1186/s13071-017-2604-8)
64. Fotiou A, Kanavou E, Antaraki A, Richardson C, Terzidou M, Kokkevi A. 2016 HCV/HIV coinfection among people who inject drugs and enter opioid substitution treatment in Greece: prevalence and correlates. *Hepatol. Med. Policy* **1**, 9. (doi:10.1186/s41124-016-0017-5)
65. Fountain-Jones NM, Packer C, Jacquot M, Blanchet FG, Terio K, Craft ME. 2019 Endemic infection can shape exposure to novel pathogens: pathogen co-occurrence networks in the Serengeti lions. *Ecol. Lett.* **22**, 904–913. (doi:10.1111/ele.13250)
66. Fujimoto K, Flash CA, Kuhns LM, Kim JY, Schneider JA. 2018 Social networks as drivers of syphilis and HIV infection among young men who have sex with men. *Sex. Transm. Infect.* **94**, 365–371. (doi:10.1136/sextrans-2017-053288)
67. Ghebremichael M. 2015 Joint modeling of correlated binary outcomes: HIV-1 and HSV-2 co-infection. *J. Appl. Stat.* **42**, 2180–2191. (doi:10.1080/02664763.2015.1022138)
68. Hamill MM *et al*. 2022 High burden of untreated syphilis, drug resistant *Neisseria gonorrhoeae*, and other sexually transmitted infections in men with urethral discharge syndrome in Kampala, Uganda. *BMC Infect. Dis.* **22**, 440. (doi:10.1186/s12879-022-07431-1)
69. Heinze K, Kabeto M, Martin ET, Cassone M, Hicks L, Mody L. 2019 Predictors of methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant enterococci co-colonization among nursing facility patients. *Am. J. Infect. Control* **47**, 415–420. (doi:10.1016/j.ajic.2018.09.026)
70. Htun HL, Hon PY, Tan R, Ang B, Chow A. 2023 Synergistic effects of length of stay and prior MDRO carriage on the colonization and co-colonization of methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant *Enterococcus*, and carbapenemase-producing enterobacterales across healthcare settings. *Infect. Control Hosp. Epidemiol.* **44**, 31–39. (doi:10.1017/ice.2022.57)
71. Hu QH *et al*. 2017 Prevalence and determinants of herpes simplex virus type 2 (HSV-2)/syphilis co-infection and HSV-2 mono-infection among human immunodeficiency virus positive men who have sex with men: a cross-sectional study in Northeast China. *Jpn. J. Infect. Dis.* **70**, 284–289. (doi:10.7883/yoken.JJID.2016.177)

72. Kengne-Nde C *et al*. 2020 Highlighting a population-based re-emergence of syphilis infection and assessing associated risk factors among pregnant women in Cameroon: evidence from the 2009, 2012 and 2017 national sentinel surveillance surveys of HIV and syphilis. *PLoS One* **15**, e0241999. (doi:10.1371/journal.pone.0241999)
73. Krasnov BR, Spickett A, Junker K, Bugmyrin SV, Ieshko EP, Bespyatova LA, Stanko M, Khokhlova IS, Matthee S. 2021 Parasite counts or parasite incidences? Testing differences with four analyses of infracommunity modelling for seven parasite–host associations. *Parasitol. Res.* **120**, 2569–2584. (doi:10.1007/s00436-021-07217-5)
74. Loum MA, Poursat MA, Sow A, Sall AA, Loucoubar C, Gassiat E. 2019 Multinomial logistic model for coinfection diagnosis between arbovirus and malaria in Kedougou. *Int. J. Biostat.* **15**. (doi:10.1515/ijb-2017-0015)
75. Mair C, Nickbakhsh S, Reeve R, McMenamin J, Reynolds A, Gunson RN, Murcia PR, Matthews L. 2019 Estimation of temporal covariances in pathogen dynamics using Bayesian multivariate autoregressive models. *PLoS Comput. Biol.* **15**, e1007492. (doi:10.1371/journal.pcbi.1007492)
76. Mazigo HD, Kidenya BR, Ambrose EE, Zinga M, Waihenya R. 2010 Association of intestinal helminths and *P. falciparum* infections in co-infected school children in northwest Tanzania. *Tanzan. J. Health Res.* **12**. (doi:10.4314/thrb.v12i4.56152)
77. McKee G *et al*. 2018 Syndemic characterization of HCV, HBV, and HIV co-infections in a large population based cohort study. *eClinicalMedicine* **4–5**, 99–108. (doi:10.1016/j.eclinm.2018.10.006)
78. Mehta SD, Donovan B, Weber KM, Cohen M, Ravel J, Gajer P, Gilbert D, Burgad D, Spear GT. 2015 The vaginal microbiota over an 8- to 10-year period in a cohort of HIV-infected and HIV-uninfected women. *PLoS One* **10**, e0116894. (doi:10.1371/journal.pone.0116894)
79. Mkhize BT, Mabaso M, Mamba T, Napier CE, Mkhize-Kwitshana ZL. 2017 The interaction between HIV and intestinal helminth parasites coinfection with nutrition among adults in Kwazulu-Natal, South Africa. *Biomed. Res. Int.* **2017**, 1–12. (doi:10.1155/2017/9059523)
80. Muhlebach MS *et al*. 2018 Anaerobic bacteria cultured from cystic fibrosis airways correlate to milder disease: a multisite study. *Eur. Respir. J.* **52**, 1800242. (doi:10.1183/13993003.00242-2018)
81. Okango E, Mwambi H, Ngesa O, Achia T. 2015 Semi-parametric spatial joint modeling of HIV and HSV-2 among women in Kenya. *PLoS One* **10**, e0135212. (doi:10.1371/journal.pone.0135212)
82. Ong JJ, Aung E, Read TRH, Fairley CK, Garland SM, Murray G, Chen MY, Chow EPF, Bradshaw CS. 2018 Clinical characteristics of anorectal mycoplasma genitalium infection and microbial cure in men who have sex with men. *Sex. Transm. Dis.* **45**, 522–526. (doi:10.1097/OLQ.0000000000000793)
83. Pacha-Herrera D, Erazo-Garcia MP, Cueva DF, Orellana M, Borja-Serrano P, Arboleda C, Tejera E, Machado A. 2022 Clustering analysis of the multi-microbial consortium by *Lactobacillus* species against vaginal dysbiosis among Ecuadorian women. *Front. Cell. Infect. Microbiol.* **12**, 863208. (doi:10.3389/fcimb.2022.863208)
84. Pal A, Das A. 2021 Joint modeling of HIV and tuberculosis through copula-based bivariate binary model. *Stat. Appl.* **19**, 139–146.
85. Pascall DJ, Tinsley MC, Clark BL, Obbard DJ, Wilfert L. 2021 Virus prevalence and genetic diversity across a wild bumblebee community. *Front. Microbiol.* **12**, 650747. (doi:10.3389/fmicb.2021.650747)
86. Peci A, Winter A, Gubbay JB, Skowronski DM, Balogun EI, De Lima C, Crowcroft NS, Rebbapragada A. 2013 Community-acquired respiratory viruses and co-infection among patients of Ontario sentinel practices, April 2009 to February 2010. *Influenza Other Respir. Viruses* **7**, 559–566. (doi:10.1111/j.1750-2659.2012.00418.x)
87. Pigeault R *et al*. 2022 Determinants of haemosporidian single- and co-infection risks in western palearctic birds. *Int. J. Parasitol.* **52**, 617–627. (doi:10.1016/j.ijpara.2022.05.002)
88. Pitarch A, Diéguez-Uribeondo J, Martín-Torrijos L, Sergio F, Blanco G. 2022 Fungal signatures of oral disease reflect environmental degradation in a facultative avian scavenger. *Sci. Total Environ.* **837**, 155397. (doi:10.1016/j.scitotenv.2022.155397)
89. Pitarch A, Gil C, Blanco G. 2020 Vultures from different trophic guilds show distinct oral pathogenic yeast signatures and co-occurrence networks. *Sci. Total Environ.* **723**, 138166. (doi:10.1016/j.scitotenv.2020.138166)
90. Pullan RL, Bethony JM, Geiger SM, Cundill B, Correa-Oliveira R, Quinnell RJ, Brooker S. 2008 Human helminth co-infection: analysis of spatial patterns and risk factors in a Brazilian community. *PLoS Negl. Trop. Dis.* **2**, e352. (doi:10.1371/journal.pntd.0000352)
91. Ranjeva SL, Mihaljevic JR, Joseph MB, Giuliano AR, Dwyer G. 2019 Untangling the dynamics of persistence and colonization in microbial communities. *ISME J.* **13**, 2998–3010. (doi:10.1038/s41396-019-0488-7)
92. Raso G, Vounatsou P, McManus DP, Utzinger J. 2007 Bayesian risk maps for *Schistosoma mansoni* and hookworm mono-infections in a setting where both parasites co-exist. *Geospat. Health* **2**, 85–96. (doi:10.4081/gh.2007.257)
93. Raso G, Vounatsou P, Singer BH, N'Goran EK, Tanner M, Utzinger J. 2006 An integrated approach for risk profiling and spatial prediction of *Schistosoma mansoni*-hookworm coinfection. *Proc. Natl Acad. Sci. USA* **103**, 6934–6939. (doi:10.1073/pnas.0601559103)
94. Ray Saraswati L, Sarna A, Sebastian MP, Sharma V, Madan I, Thior I, Pulerwitz J, Tun W. 2015 HIV, hepatitis B and C among people who inject drugs: high prevalence of HIV and hepatitis C RNA positive infections observed in Delhi, India. *BMC Public Health* **15**, 726. (doi:10.1186/s12889-015-2003-z)
95. Rellstab C, Louhi KR, Karvonen A, Jokela J. 2011 Analysis of trematode parasite communities in fish eye lenses by pyrosequencing of naturally pooled DNA. *Infect. Genet. Evol.* **11**, 1276–1286. (doi:10.1016/j.meegid.2011.04.018)
96. Schur N, Gosoniu L, Raso G, Utzinger J, Vounatsou P. 2011 Modelling the geographical distribution of co-infection risk from single-disease surveys. *Stat. Med.* **30**, 1761–1776. (doi:10.1002/sim.4243)
97. Stokely JN, Niendorf S, Taube S, Hoehne M, Young VB, Rogers MA, Wobus CE. 2016 Prevalence of human norovirus and *Clostridium difficile* coinfections in adult hospitalized patients. *Clin. Epidemiol.* **8**, 253–260. (doi:10.2147/CLEP.S106495)

98. Toro-Tobón D, Berbesi-Fernández D. 2020 Prevalence of HIV/hepatitis C virus co-infection and injection risk correlations in people who inject drugs in Colombia: a cross-sectional study using respondent driven sampling. *Subst. Use Misuse* **55**, 414–423. (doi:10.1080/10826084.2019.1683198)
99. Valdes-Donoso P, Mardones FO, Jarpa M, Ulloa M, Carpenter TE, Perez AM. 2007 Co-infection patterns of infectious salmon anaemia and sea lice in farmed Atlantic salmon, *Salmo salar* L., in southern Chile (2007-2009). *J. Fish Dis.* **36**, 353–360. (doi:10.1111/jfd.12070)
100. Vaumourin E *et al*. 2014 To be or not to be associated: power study of four statistical modeling approaches to identify parasite associations in cross-sectional studies. *Front. Cell. Infect. Microbiol.* **4**, 62. (doi:10.3389/fcimb.2014.00062)
101. Wen A *et al*. 2014 Selected vaginal bacteria and risk of preterm birth: an ecological perspective. *J. Infect. Dis.* **209**, 1087–1094. (doi:10.1093/infdis/jit632)
102. Yapi RB *et al*. 2014 Infection and co-infection with helminths and *Plasmodium* among school children in Côte d'Ivoire: results from a national cross-sectional survey. *PLoS Negl. Trop. Dis.* **8**, e2913. (doi:10.1371/journal.pntd.0002913)
103. Yen YF, Yen MY, Su LW, Li LH, Chuang P, Jiang XR, Deng CY. 2012 Prevalences and associated risk factors of HCV/HIV co-infection and HCV mono-infection among injecting drug users in a methadone maintenance treatment program in Taipei, Taiwan. *BMC Public Health* **12**, 1066. (doi:10.1186/1471-2458-12-1066)
104. Zhang L *et al*. 2015 Prevalence and correlates of HCV monoinfection and HIV and HCV coinfection among persons who inject drugs in Vietnam. *Eur. J. Gastroenterol. Hepatol.* **27**, 550–556. (doi:10.1097/MEG.0000000000000321)
105. McLevey J, Graham AV, McIlroy-Young R, Browne P, Plaisance KS. 2018 Interdisciplinarity and insularity in the diffusion of knowledge: an analysis of disciplinary boundaries between philosophy of science and the sciences. *Scientometrics* **117**, 331–349. (doi:10.1007/s11192-018-2866-8)
106. Conrad PA, Meek LA, Dumit J. 2013 Operationalizing a One Health approach to global health challenges. *Comp. Immunol. Microbiol. Infect. Dis.* **36**, 211–216. (doi:10.1016/j.cimid.2013.03.006)
107. Baselga A, Araújo MB. 2010 Do community-level models describe community variation effectively? *J. Biogeogr.* **37**, 1842–1850. (doi:10.1111/j.1365-2699.2010.02341.x)
108. D'Amen M, Pradervand JN, Guisan A. 2015 predicting richness and composition in mountain insect communities at high resolution: a new test of the SESAM framework. *Glob. Ecol. Biogeogr.* **24**, 1443–1453. (doi:10.1111/geb.12357)
109. Harris DJ, Taylor SD, White EP. 2018 Forecasting biodiversity in breeding birds using best practices. *PeerJ* **6**, e4278. (doi:10.7717/peerj.4278)
110. Leathwick JR, Elith J, Hastie T. 2006 Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol. Modell.* **199**, 188–196. (doi:10.1016/j.ecolmodel.2006.05.022)
111. Maguire KC, Nieto-Lugilde D, Blois JL, Fitzpatrick MC, Williams JW, Ferrier S, Lorenz DJ. 2016 Controlled comparison of species- and community-level models across novel climates and communities. *Proc. R. Soc. B* **283**, 20152817. (doi:10.1098/rspb.2015.2817)
112. Moisen GG, Frescino TS. 2002 Comparing five modelling techniques for predicting forest characteristics. *Ecol. Modell.* **157**, 209–225. (doi:10.1016/S0304-3800(02)00197-7)
113. Norberg A *et al*. 2019 A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecol. Monogr.* **89**, e01370. (doi:10.1002/ecm.1370)
114. Zhang C, Chen Y, Xu B, Xue Y, Ren Y. 2018 Comparing the prediction of joint species distribution models with respect to characteristics of sampling data. *Ecography* **41**, 1876–1887. (doi:10.1111/ecog.03571)
115. Powell-Romero F, Fountain-Jones NM, Norberg A, Clark NJ. 2023 Improving the predictability and interpretability of co-occurrence modelling through feature-based joint species distribution ensembles. *Methods Ecol. Evol.* **14**, 146–161. (doi:10.1111/2041-210X.13915)
116. Bonthoux S, Baselga A, Balent G. 2013 Assessing community-level and single-species models predictions of species distributions and assemblage composition after 25 years of land cover change. *PLoS One* **8**, e54179. (doi:10.1371/journal.pone.0054179)
117. Ovaskainen O, Tikhonov G, Norberg A, Guillaume Blanchet F, Duan L, Dunson D, Roslin T, Abrego N. 2017 How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol. Lett.* **20**, 561–576. (doi:10.1111/ele.12757)
118. Tikhonov G, Duan L, Abrego N, Newell G, White M, Dunson D, Ovaskainen O. 2020 Computationally efficient joint species distribution modeling of big spatial data. *Ecology* **101**, e02929. (doi:10.1002/ecy.2929)
119. Tikhonov G, Opedal Ø, Abrego N, Lehikoinen A, Ovaskainen O. 2019 Joint species distribution modelling with HMSC-R. *bioRxiv*. (doi:10.1101/603217)
120. Powell-Romero F, Wells K, Clark NJ. 2024 Data from: A systematic review and guide for using multi-response statistical models in co-infection research. Zenodo. (doi:10.5281/zenodo.13382166)
121. Powell-Romero F, Wells K, Clark NJ. 2024 Data from: A systematic review and guide for using multi-response statistical models in co-infection research. Figshare. (doi:10.6084/m9.figshare.c.7425681)