Co-infections by non-interacting pathogens are not independent & require new tests of interaction

Frédéric M. Hamelin [1], Linda J.S. Allen [2], Vrushali A. Bokil [3], Louis J. Gross [4], Frank M. Hilker [5], Michael J. Jeger [6], Carrie A. Manore [7], Alison G. Power [8], Megan A. Rúa [9], Nik J. Cunniffe [10,*]

[1] IGEPP, Agrocampus Ouest, INRA, Université de Rennes 1, Université Bretagne-Loire, 35000 Rennes, France; [2] Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409, USA; [3] Department of Mathematics, Oregon State University, Corvallis, OR 97331, USA; [4] National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville, TN 37996, USA; [5] Institute of Environmental Systems Research, School of Mathematics/Computer Science, Osnabrück University, 49069 Osnabrück, Germany; [6] Centre for Environmental Policy, Imperial College London, Ascot, SL5 7PY, UK; [7] Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM 87545, USA; [8] Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA; [9] Department of Biological Sciences, Wright State University, Dayton, OH 45435, USA; [10] Department of Plant Sciences, University of Cambridge, Cambridge, CB2 3EA, UK.;
* Corresponding author: `njc1001@cam.ac.uk`

Abstract If pathogen species, strains or clones do not interact, intuition suggests the proportion of co-infected hosts should be the product of the individual prevalences. Independence consequently underpins the wide range of methods for detecting pathogen interactions from cross-sectional survey data. Surprisingly, however, the results of the very simplest of epidemiological models challenge the underlying assumption of statistical independence. Even if pathogens do not interact, death and/or clearance of co-infected hosts causes net prevalences of individual pathogens to decrease simultaneously. The induced positive correlation between prevalences means the proportion of co-infected hosts is expected to be higher than multiplication would suggest. By modeling the dynamics of multiple non-interacting pathogens, we develop a pair of novel tests of interaction that properly account for this hitherto overlooked coupling. Which test is appropriate for any application depends on the form of the available data, as well as the extent to which the co-infecting pathogens are epidemiologically similar. Our tests allow us to reinterpret data from a number of previous studies, including pathogens of plants and animals, as well as papillomavirus and malaria in humans. We find certain reports of interactions have been overstated, and in some cases are not supported by the available data. Our work demonstrates how methods to identify interactions between pathogens can be updated to account for epidemiological dynamics.

Keywords: coinfection, multiple pathogens, SIS epidemic, statistical independence, pathogen association

Author contributions: Designed research: All authors. Performed mathematical analyses: F.M.Ha., L.J.A., N.J.C. Explored and analyzed data: F.M.Ha., F.M.Hi., M.A.R., N.J.C. Wrote the paper: F.M.Ha., L.J.A., N.J.C.

1 Introduction

It is increasingly recognized that infections often involve multiple pathogen species or strains/clones of the same species (Balmer and Tanner, 2011; Vaumourin et al., 2015). Infection by one pathogen can affect susceptibility to subsequent infection by others (Griffiths et al., 2011; Petney and Andrews, 1998). Co-infection can also affect the severity and/or duration of infection, as well as the extent of symptoms and the level of infectiousness (Graham et al., 2005). Antagonistic, neutral and facilitative interactions are possible (Karvonen et al., 2018; Rigaud et al., 2010). Co-infection therefore potentially has significant epidemiological, clinical and evolutionary implications (Susi et al., 2015; Hilker et al., 2017; Alizon et al., 2013).

However, detecting and quantifying biological interactions between pathogens is notoriously challenging (Johnson and Buller, 2011; Hellard et al., 2015).In pathogens of some host taxa, most notably plant pathogens, biological interactions can be quantified by direct experimentation (Mascia and Gallitelli, 2016). However, often ethical considerations mean this is impossible, and so any signal of interaction must be extracted from population scale data. Analysis of longitudinal data remains the gold standard (Fenton et al., 2014), although the associated methods are not infallible (Telfer et al., 2010). However, collecting longitudinal data requires a dedicated and intensive sampling campaign, meaning in practice cross-sectional data are often all that are available. Methods for cross-sectional data typically concentrate on identifying deviation from statistical independence, using standard methods such as χ^2 tests or log-linear modelling to test whether the observed probability of co-infection differs from the product of the prevalences of the individual pathogens (Booth and Bundy, 1995; Howard et al., 2001; Bogaert et al., 2004; Raso et al., 2004; Regev-Yochay et al., 2004; Nielsen et al., 2006; Chaturvedi et al., 2011; Rositch et al., 2012; Degarege et al., 2012; Malagón et al., 2016; Teweldemedhin et al., 2018). Detecting such a non-random statistical association between pathogens is then taken to signal a biological interaction. The underlying mechanism can range, for example, from individual-scale direct effects on within-host pathogen dynamics (Tollenaere et al., 2016; Mascia and Gallitelli, 2016), to indirect within-host immune-mediated interactions (de Roode et al., 2005), to indirect population-scale “ecological interference” caused by competition for susceptible hosts (Rohani et al., 1998, 2003).

A well-known difficulty is that factors other than biological interactions between pathogens can drive statistical associations. For instance, host heterogeneity – that some hosts are simply more likely than others to become infected – can generate positive statistical associations, since co-infection is more common in the most vulnerable hosts. Heterogeneity in host age can also generate statistical associations, as infections accumulate in older individuals (Lord et al., 1999). Methods aimed at disentangling such confounding factors have been developed, but show mixed results in detecting biological interactions (Hellard et al., 2012; Vaumourin et al., 2014). Methods using dynamic epidemiological models to track co-infections are also emerging, although more often than not require longitudinal data (Shrestha et al., 2011, 2013; Reich et al., 2013; Alizon et al., 2018; Man et al., 2018).

More fundamentally, however, the underpinning and long standing assumption that non-interaction implies statistical independence (Forbes, 1907; Cohen, 1973) has not been challenged. Here we confront the intuition that biological interactions can be detected via statistical associations, demonstrating how simple epidemiological models can change the way we think about biological interactions. In particular, we show that non-interacting pathogens should not be expected to have prevalences that are statistically independent. Co-infection by non-interacting pathogens is more probable than multiplication would suggest, invalidating any test invoking statistical independence.

The paper is organized as follows. First, we use a simple epidemiological model to show that the probability that a host is co-infected by both of a pair of non-interacting pathogens is greater than the product of the net prevalences of the individual pathogens. Second, we extend this result to an arbitrary number of non-interacting pathogens. This allows us to construct a novel test for biological interaction, based on testing the extent to which co-infection data can be explained by our epidemiological models in which pathogens do not interact. Different versions of this test, conditioned on the form of available data and whether or not co-infections are caused by different pathogen species, allow us to reinterpret a number of previous reports (Chaturvedi et al., 2011; López-Villavicencio et al., 2007; Andersson et al., 2013; Koepfli et al., 2011; Nickbakhsh et al., 2016; Seabloom et al., 2009; Moutailler et al., 2016; Howard et al., 2001; Molineaux et al., 1980). Our examples

include plant, animal and human pathogens, and the methodology can potentially be applied to any cross-sectional survey data tracking co-infection.

2 Results

2.1 Two non-interacting pathogens

2.1.1 Dynamics of the individual pathogens

We consider two distinct pathogen species, strains or clones (henceforth pathogens), which we assume do not interact, i.e. the interaction between the host and one of the pathogens is entirely unaffected by its infection status with respect to the other. Epidemiological properties that are therefore unaffected by the presence or absence of the other pathogen include initial susceptibility, within-host dynamics including rates of accumulation and/or movement within tissues, host responses to infection, as well as onward transmission. Assuming a fixed-size host population and Susceptible-Infected-Susceptible (S-I-S) dynamics (Keeling and Rohani, 2007), the proportion of the host population infected by pathogen $i \in \{1, 2\}$ follows

$$\dot{I}_i = \beta_i I_i (1 - I_i) - \mu I_i, \tag{1}$$

in which the dot denotes differentiation with respect to time, β_i is a pathogen-specific infection rate, and μ is the host's renewal rate.

Specifically, the renewal rate μ is the sum of the host's death rate and the pathogens' clearance rate (Box 1). For simplicity of presentation in the main text we assume clearance is unspecific, meaning it simultaneously clears all pathogens from an infected host (Nowak et al., 1990; Hellriegel, 1992; Antia et al., 1996; Lion, 2013; Murall et al., 2014). Infected individuals are therefore assumed to die or to clear pathogens at net rate μ, independent of the pathogens hosted by each individual. However our model can be extended to handle pathogen-specific rates of clearance (Box 1; Supplementary Information: Sections S1.4, S1.5 and S2.3).

Unspecific clearance (i.e. simultaneous purging of all pathogens from a single host) is possible even when pathogens do not interact, so long as clearance occurs at the same rate in singly- and co-infected hosts. For clearance caused by host

immune responses, this corresponds to assuming that no additional host defences are induced by co-infection. For clearance following treatment, this corresponds to application to all hosts regardless of infection status. Such preventive treatments are common in agriculture, in which context systematic and repeated treatments with antibiotics or fungicides are routine in both animal and plant populations. In these cases, unspecific clearance is entirely orthogonal to any pathogen interactions, but can occur at a significant rate.

The death rate of the host in our model may include disease-induced death and natural death. Box 1 shows how disease-induced mortality (virulence) is also compatible with non-interacting pathogen dynamics in certain cases. Yet, pathogens do not necessarily increase the death rate of their host. Similarly, unspecific clearance may not occur. Therefore, the parameter μ may reduce to the natural death rate of the host. While natural mortality may be negligible for acute infections, it cannot be neglected for chronic (i.e. long-lasting) infections, which are responsible for a large fraction of co-infections in humans and animals (Griffiths et al., 2011; Gorsich et al., 2018). Likewise, plants remain infected over their entire lifetime following infection by most pathogens, including almost all plant viruses, as well as the anther smut fungus, which drives one of our examples here (López-Villavicencio et al., 2007).

We assume that host population size is a constant. This makes death and unspecific pathogen clearance epidemiologically equivalent, since both replace an infected host with an uninfected one. The simplification facilitates later extension to an arbitrary number of co-infecting pathogens. Focusing on a generic model also allows us to reasonably use it for pathogens of any host species ranging from humans to animals to plants.

2.1.2 Tracking co-infection

Making identical assumptions, but instead distinguishing hosts infected by different combinations of pathogens, leads to an alternate representation of the dynamics. We denote the proportion of hosts infected by only one of the two pathogens by J_i, with $J_{1,2}$ representing the proportion co-infected. Pathogen-specific net forces of infection are

$$F_i = \beta_i I_i = \beta_i (J_i + J_{1,2}) \,, \tag{2}$$

and so

$$
\begin{aligned}
\dot{J}_1 &= F_1 J_\varnothing - (F_2 + \mu) J_1 \,, \\
\dot{J}_2 &= F_2 J_\varnothing - (F_1 + \mu) J_2 \,, \\
\dot{J}_{1,2} &= F_2 J_1 + F_1 J_2 - \mu J_{1,2} \,.
\end{aligned}
\tag{3}
$$

in which $J_\varnothing = 1 - J_1 - J_2 - J_{1,2}$ is the proportion of hosts uninfected by either pathogen (Fig. 1).

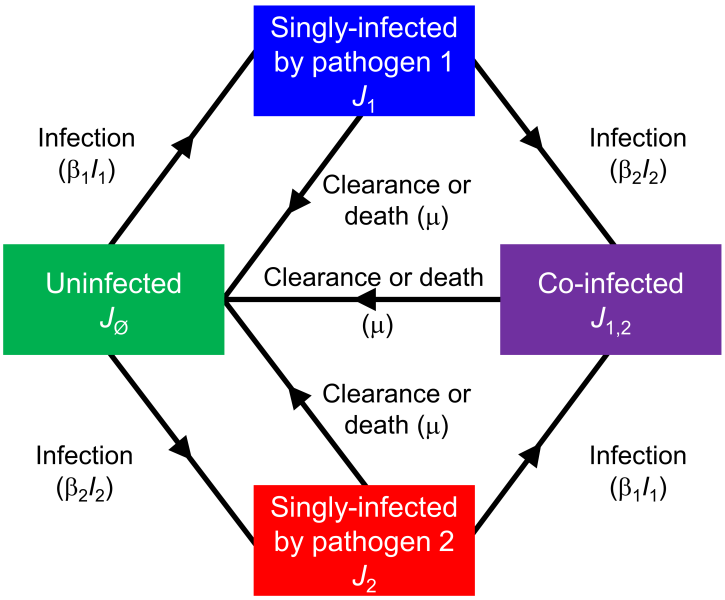

Figure 1: **Schematic of the model tracking a pair of non-interacting pathogens**. The model is defined in Eqs. (1-3): $J_\varnothing$ denotes uninfected hosts, J_1 and J_2 are hosts singly infected by pathogens 1 and 2, respectively, $J_{1,2}$ are co-infected hosts, $I_1 = J_1 + J_{1,2}$ and $I_2 = J_2 + J_{1,2}$ are net densities of hosts infected by pathogens 1 and 2, respectively. The model assumes clearance is unspecific, meaning that all pathogens are simultaneously purged from any infected host. This makes clearance epidemiologically-equivalent to host death. However, this assumption can be relaxed (Box 1; Supplementary Information: Sections S1.4, S1.5 and S2.3).

2.1.3 Prevalence of co-infected hosts

We assume the basic reproduction number, $R_{0,i} = \beta_i/\mu > 1$ for both pathogens. Solving Eq. (3) numerically for arbitrary but representative parameters (Fig. 2A) shows the proportion of co-infected hosts ($J_{1,2}$) to be larger than the product of the individual prevalences ($P = I_1 I_2$ from Eq. (1)). That $J_{1,2}(t) \geq P(t)$ for large t (for all parameters) can be proved analytically (Supplementary Information: Section S1.1).

Simulations of a stochastic analogue of the model (Fig. 2B) reveal the key driver of this behavior. The net prevalences of the pathogens considered in isolation, I_1 and I_2, are positively correlated (Fig. 2C; Eq. (27) in Methods: Section 4.1.4, "Stochastic models"), due to simultaneous reductions whenever co-infected hosts die or clear pathogens.

Box 1. Is a non-interacting model with non-zero virulence an oxymoron?

Here we show that it is only possible to build a model in which two virulent pathogens do not interact at the population scale (i.e., the net prevalence of each pathogen is unaffected by the presence or absence of the other) if: i) both pathogens have the same disease-induced death rate (virulence) α_i; ii) co-infected hosts do not suffer any additional disease-induced mortality. The model is restricted to two pathogen species for simplicity ($i = 1, 2$). The specific clearance rate of each pathogen is γ_i. The disease-induced death rate of co-infected hosts is $\alpha_{1,2}$. The natural death rate of the host is ν. The host population size is assumed to be a constant, with no vertical transmission. Therefore, each death implies the birth of an uninfected individual. Unspecific clearance (purging all pathogens at once) occurs at rate ξ. As in the main text, $I_i = J_i + J_{1,2}$ denotes the prevalence of hosts singly infected or doubly infected by pathogen i ($i = 1, 2$). The full model is:

$$\begin{aligned}
\dot{J}_1 &= F_1 J_\varnothing - (F_2 + \gamma_1 + \alpha_1 + \nu + \xi) J_1 + \gamma_2 J_{1,2}\,, \\
\dot{J}_2 &= F_2 J_\varnothing - (F_1 + \gamma_2 + \alpha_2 + \nu + \xi) J_2 + \gamma_1 J_{1,2}\,, \\
\dot{J}_{1,2} &= F_2 J_1 + F_1 J_2 - (\gamma_1 + \gamma_2 + \alpha_{1,2} + \nu + \xi) J_{1,2}\,,
\end{aligned}$$

with $F_i = \beta_i I_i = \beta_i (J_i + J_{1,2})$ and $J_\varnothing = 1 - J_1 - J_2 - J_{1,2}$. The full model implies:

$$\begin{aligned}
\dot{I}_1 &= F_1(1 - I_1) - (\gamma_1 + \alpha_1 + \nu + \xi) I_1 + (\alpha_1 - \alpha_{1,2}) J_{1,2}\,, \\
\dot{I}_2 &= F_2(1 - I_2) - (\gamma_2 + \alpha_2 + \nu + \xi) I_2 + (\alpha_2 - \alpha_{1,2}) J_{1,2}\,.
\end{aligned}$$

Non-interacting pathogens necessarily have independent dynamics, i.e.,

$$\begin{aligned}
\dot{I}_1 &= F_1(1 - I_1) - (\gamma_1 + \alpha_1 + \nu + \xi) I_1\,, \\
\dot{I}_2 &= F_2(1 - I_2) - (\gamma_2 + \alpha_2 + \nu + \xi) I_2\,.
\end{aligned}$$

This requires $\alpha_1 = \alpha_{1,2} = \alpha_2$. Therefore, non-interacting pathogens must have the same virulence ($\alpha_1 = \alpha_2$) and being infected by one or both pathogens must leave the disease-induced death rate unchanged $\alpha_{1,2} = \alpha$. These assumptions may fit similar pathogens such as different genotypes of the same species. However, the absence of additional disease-induced mortality may imply (indirect) interactions at the within-host scale. This does not compromise non-interaction at the host population scale though.

Otherwise ($\alpha_1 \neq \alpha_2$ or $\alpha_1 = \alpha_2 \neq \alpha_{1,2}$), it is simply not possible to build a non-interacting model with nonzero virulence, since pathogens interact at the host population scale. This is a form of "ecological interference" (Rohani et al., 2003). Consequently, there is no reason to expect such virulent pathogens to be statistically independent.

In conclusion, the renewal rate μ in the main text can be interpreted as $\mu = \nu + \xi$ for pathogens which do not increase their host's death rate ($\alpha_i = 0$ for $i = 1, 2$) or (more cautiously) as $\mu = \nu + \xi + \alpha$ for similar pathogens (restricting "non-interaction" to its population-scale definition: the prevalence of each pathogen is unaffected by the other). The answer to the question posed in the title of this box is therefore "No."

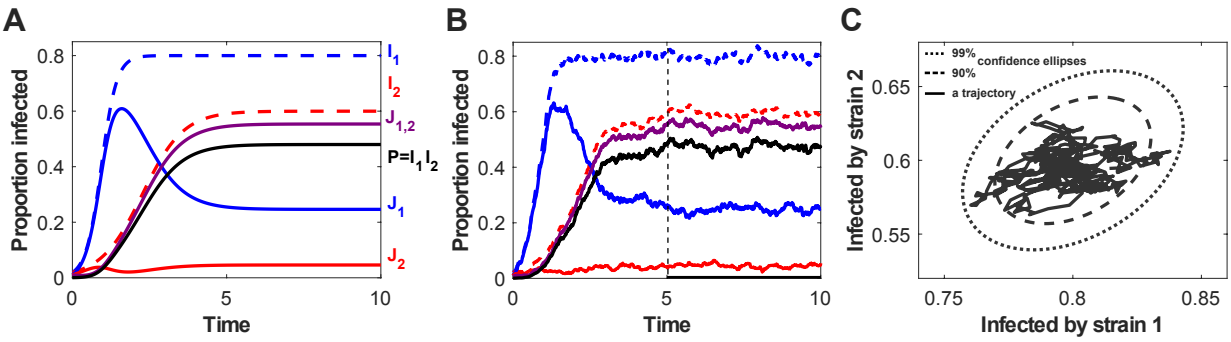

Figure 2: **Simulations of the two-pathogen model show that net densities of the two pathogens are positively correlated**. J_1 and J_2 are hosts singly infected by pathogens 1 and 2, respectively, $J_{1,2}$ are co-infected hosts, $I_1 = J_1 + J_{1,2}$ and $I_2 = J_2 + J_{1,2}$ are net densities of hosts infected by pathogens 1 and 2, respectively. (A) Dynamics of the deterministic model (1-3), with $\beta_1 = 5$, $\beta_2 = 2.5$, and $\mu = 1$ (parameters have units of inverse time). (B) Dynamics of a stochastic version of the model, in a population of size $N = 1000$ (see also Methods: Section 4.1.4, "Stochastic models"). (C) A single trajectory from the stochastic simulation (black line) in panel B (restricted to the time interval starting from the dashed line at $t = 5$) in the phase plane (I_1, I_2), and the 90% and 99% confidence ellipses (dashed and dotted curves, respectively) generated from an analytical approximation to the stochastic model.

2.1.4 Deviation from statistical independence

For $R_{0,i} > 1$ the relative deviation of the equilibrium prevalence of co-infection ($\bar{J}_{1,2}$) from that required by statistical independence ($\bar{P} = \bar{I}_1\bar{I}_2$) is

$$\frac{\bar{J}_{1,2} - \bar{P}}{\bar{P}} = \frac{\mu}{\beta_1 + \beta_2 - \mu} = \frac{1}{R_{0,1} + R_{0,2} - 1} \geq 0 \tag{4}$$

(Eq. (9) in Methods: Section 4.1.1, "Equilibria of the two-pathogen model"). The deviation is therefore zero if and only if the renewal rate $\mu = 0$.

The observed outcome would conform with statistical independence only for non-interacting pathogens where there is no unspecific clearance or host death (at the timescale of an epidemic). This reiterates the role of host renewal in causing deviation from a statistical association pattern. Note that our result is not related to aging (infections accumulating in older individuals) as the deviation increases with the renewal rate μ.

2.2 Testing for interactions between pathogens

Eq. (3) can be straightforwardly extended to track n pathogens which do not interact in any way (including pairwise and three-way interactions). Equilibria of this model are prevalences of different classes of infected or co-infected hosts carrying different combinations of non-interacting pathogens. These can be used to derive a test for interaction between pathogens which properly accounts for the lack of statistical independence revealed by our analysis of the simple two-pathogen model.

2.2.1 Modelling co-infection by n non-interacting pathogens

We denote the proportion of hosts simultaneously co-infected by the (non-empty) set of pathogens Γ to be J_Γ, and use $\Omega_i = \Gamma \setminus \{i\}$ (for $i \in \Gamma$) to represent combinations with one fewer pathogen.

The dynamics of the $2^n - 1$ distinct values of J_Γ follow

$$\dot{J}_\Gamma = \sum_{i \in \Gamma} F_i J_{\Omega_i} - \left(\sum_{i \notin \Gamma} F_i + \mu \right) J_\Gamma, \tag{5}$$

in which the net force of infection of pathogen i is

$$F_i = \beta_i I_i = \beta_i \sum_{\Gamma \in \nabla_i} J_\Gamma, \tag{6}$$

and ∇_i is the set of all subsets of $\{1, \ldots, n\}$ containing i as an element. Eq. (5) can be interpreted by noting

- the first term tracks inflow due to hosts carrying one fewer pathogen becoming infected;
- the second term tracks the outflows due to hosts becoming infected by an additional pathogen, or death/clearance.

Predicted prevalences. If $R_{0,i} = \beta_i/\mu > 1$ for all $i = 1, \ldots, n$, the equilibrium prevalence of hosts infected by any given combination of pathogens, $\bar{J}_\Gamma$, can be obtained by (recursively) solving a system of 2^n linear equations (Eq. (16) in Methods: Section 4.1.2, “Equilibria of the n-pathogen model”).

These equilibrium prevalences are the prediction of our "Non-interacting Distinct Pathogens" (NiDP) model, which in dimensionless form has n parameters (the $R_{0,i}$'s, $i = 1, \ldots, n$; Methods: Section 4.2.2, "Fitting the models").

Epidemiologically interchangeable pathogens. If we simplify the model by assuming that all pathogen infection rates are equal (i.e. $\beta_i = \beta$ for all i), then if $R_0 = \beta/\mu > 1$, the proportion of hosts infected by k distinct pathogens can be obtained by (recursively) solving $n + 1$ linear equations (Eq. (22) in Methods: Section 4.1.3, "Deriving the NiSP model from the NiDP model"). This constitutes the prediction of our "Non-interacting Similar Pathogens" (NiSP) model, a simplified form of the NiDP model requiring only a single parameter (R_0).

2.2.2 Using the models to test for interactions

If either the NiSP or NiDP model adequately explain co-infection data, those data are consistent with the underpinning assumption that pathogens do not interact. Which model is fitted depends on the form of the available data.

Numbers of distinct pathogens. Studies often quantify only the number of distinct pathogens carried by individual hosts, without necessarily specifying the combinations involved (López-Villavicencio et al., 2007; Seabloom et al., 2009; Chaturvedi et al., 2011; Koepfli et al., 2011; Andersson et al., 2013; Moutailler et al., 2016; Nickbakhsh et al., 2016). There are insufficient degrees of freedom in such data to fit the NiDP model, and so we fall back upon the NiSP model. In using the NiSP model, we additionally assume all pathogens within a given study are epidemiologically interchangeable.

We identified four suitable studies reporting data concerning strains/clones of a single pathogen, and tested whether these data are consistent with no interaction. For all four studies (Fig. 3), the best-fitting NiSP model is a better fit to the data than the corresponding binomial model assuming statistical independence (Eq. (28) in Methods: Section 4.2.1, "Models corresponding to assuming statistical independence"). Three additional examples for data-sets considering distinct pathogens, which deviate more markedly from the epidemiological equivalence assumption, are

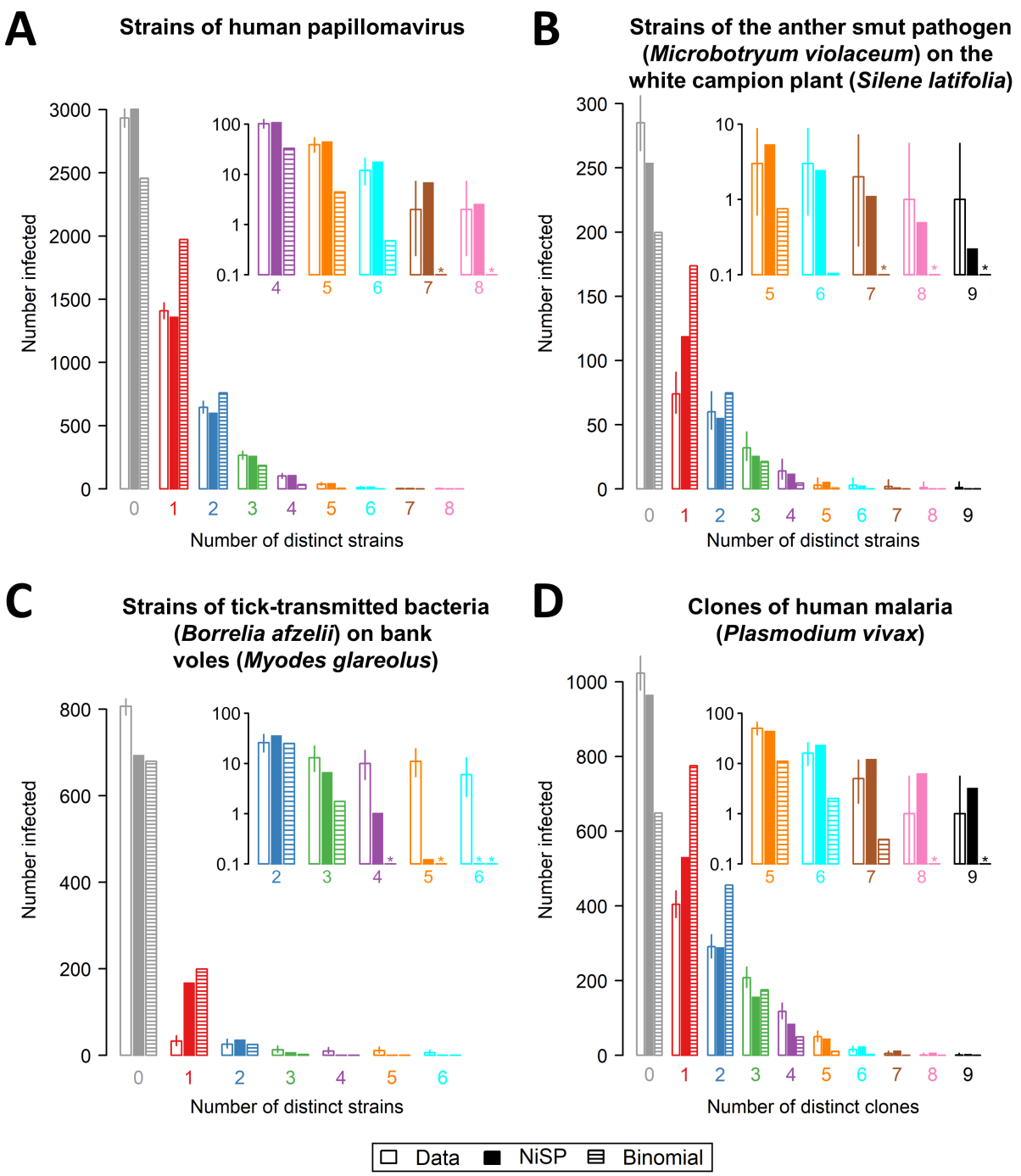

Figure 3: **Comparing predictions of the NiSP model with binomial models assuming statistical independence**. In using the NiSP model, pathogens are assumed to be epidemiologically interchangeable: we have therefore restricted attention to data sets concerning strains/clones of a single pathogen species. (A) strains of human papillomavirus (Chaturvedi et al., 2011); (B) strains of the anther smut pathogen (*M. violaceum*) on the white campion (*S. latifolia*) (López-Villavicencio et al., 2007); (C) strains of tick-transmitted bacteria (*B. afzelii*) on bank voles (*M. glareolus*) (Andersson et al., 2013); and (D) clones of malaria (*P. vivax*) (Koepfli et al., 2011). Insets to each panel show a "zoomed-in" section of the graph corresponding to high multiplicities of clone/strain co-infection. Asterisks indicate predicted counts smaller than 0.1. In all four cases, the NiSP model is a better fit to the data than the binomial model (ΔAIC $= 572.8, 158.6, 293.8$ and 596.3, respectively). For the data shown in panel (A), there is no evidence that the NiSP model does not fit the data (lack of goodness-of-fit $p = 0.08$), and so our test indicates the human papillomavirus strains do not interact. For the data shown in panels (B)-(D), there is evidence of lack of goodness-of-fit (all have lack of goodness-of-fit $p < 0.01$). Our test therefore indicates these strains/clones interact (or are epidemiologically different).

in the Supplementary Information (Section S2.1).

In one case – co-infection by different strains of human papillomavirus (Chaturvedi et al., 2011) (Fig. 3A) – we find no evidence that the reported data cannot be explained by the NiSP model. These data therefore support the hypothesis of no interaction – and indeed no epidemiological differences – between the pathogen strains in question.

In the three other cases we considered – strains of anther smut (*Microbotryum violaceum*) on the white campion (*Silene latifolia*) (López-Villavicencio et al., 2007) (Fig. 3B); strains of the tick-transmitted bacterium *Borrelia afzelii* on bank voles (*Myodes glareolus*) (Andersson et al., 2013) (Fig. 3C); and clones of a single malaria parasite (*Plasmodium vivax*) infecting children (Koepfli et al., 2011) (Fig. 3D) – despite outperforming the model corresponding to statistical independence, the best-fitting NiSP model does not adequately explain the data. We therefore reject the hypotheses of no interaction in all three cases, noting that our use of the NiSP model means it might be epidemiological differences between pathogen strains/clones that have in fact been revealed.

Combinations of pathogens. Other studies report the proportion of hosts infected by particular combinations (rather than counts) of pathogens, although many of those concentrate on helminth macroparasites for which our underlying S-I-S model is well-known to be inappropriate (Anderson and May, 1991).

However, a methodological article by Howard et al. (2001) introduces the use of log-linear modeling to test for statistical associations. Conveniently, that article reports the results of that methodology as applied to a large number of studies focusing on *Plasmodium* spp. causing malaria.

By interrogating the original data sources (Methods: Section 4.3.2, “Combinations of pathogens (NiDP model)”), we found a total of 41 studies of malaria reporting the disease status of at least $N = 100$ individuals, and in which three of *P. falciparum*, *P. malariae*, *P. ovale* and *P. vivax* were considered. Data therefore consist of counts of the number of individuals infected with different combinations of three of these four pathogens, a total of eight classes. There were sufficient degrees of freedom to fit the NiDP model, which here has three parameters, each corresponding to the

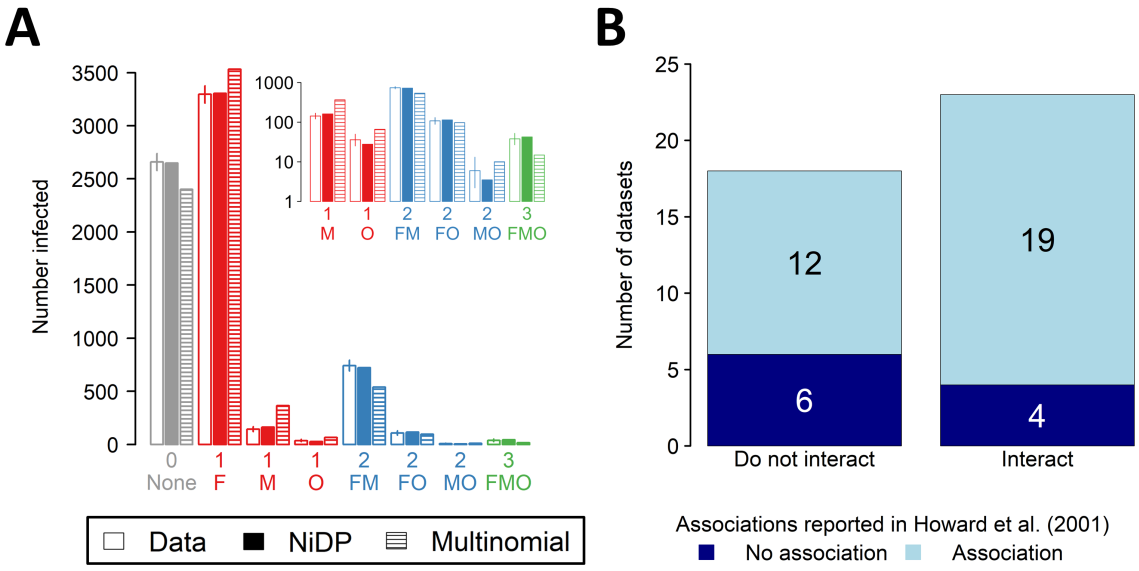

Figure 4: **Using the NiDP model to re-analyse malaria data sets considered by Howard et al. (2001)**. In using the NiDP model there is no need to assume malaria-causing *Plasmodium* spp. are epidemiologically interchangeable. (A) Comparing the predictions of the NiDP model with a multinomial model of infection (i.e. statistical independence) for the data set on *P. falciparum* (F), *P. malariae* (M) and *P. ovale* (O) co-infection in Nigeria reported by Molineaux et al. (1980). The NiDP model is a better fit to the data than the multinomial model (ΔAIC = 326.2); additionally, there is no evidence of lack of goodness-of-fit (p = 0.40). This data set is therefore consistent with no interaction between the three *Plasmodium* species. (B) Comparing the results of fitting the NiDP model and the methodology of Howard et al. (2001) based on log-linear regression and so statistical-independence. For 16 (i.e. 12 + 4) out of the 41 data sets we considered, the conclusions of the two methods differ.

infection rate of a single *Plasmodium* spp. Fig. 4A shows the example of fitting the NiDP model to data from a study of malaria in Nigeria (Molineaux et al., 1980).

Fitting the NiDP model allows us to test for interactions between *Plasmodium* spp., without assuming they are epidemiologically interchangeable. In 18 of the 41 cases we considered, our methods suggest the data are consistent with no interaction (Fig. 4B). We note that in 12 of these 18 cases the methodology based on statistical independence of Howard et al. (2001) instead suggests the *Plasmodium* spp. interact.

3 Discussion

We have shown that pathogens which do not interact and so have uncoupled prevalence dynamics (Eq. (1)) are not statistically independent. For two pathogens, the prevalence of co-infection is always greater than the product of the prevalences (Eq. (4)), unless host renewal (death or unspecific clearance) does not occur. Pathogens share a single host in co-infections, and so when a co-infected host is renewed, net prevalences of both pathogens decrease simultaneously. The prevalences of individual pathogens, regarded as random variables, therefore co-vary positively. As a side result, we note our analysis indicates a high-profile model of May and Nowak (1995) is based on a faulty assumption of probabilistic independence (Supplementary Information: Section S1.3). More importantly, our analysis shows that statistically independent pathogens may well be interacting (Supplementary Information: Section S1.5) which confirms that statistical independence is far from equivalent to the absence of biological interaction between pathogens.

We extended our model to an arbitrary number of pathogens to develop a novel test for interaction that properly accounts for statistical non-independence. Many data sets summarize co-infections in terms of multiplicity of infection, regardless of which pathogens are involved. Since there would be as many epidemiological parameters as pathogens in our default NiDP model, and so as many parameters as data-points, the full model would be over-parameterised. We therefore introduced the additional assumption that all pathogens are epidemiologically interchangeable. This formed the basis of the parsimonious NiSP model, which is most appropriate for

testing for interactions between strains or clones of a single pathogen species.

Despite the strong and perhaps even unrealistic assumption that strains/clones are interchangeable, the NiSP model outperformed the binomial model assuming statistical independence for all four data sets we considered. In particular, the NiSP model successfully captured the fat tails characteristic of observed multiplicity of infection distributions. All four data sets therefore support the idea that co-infection is far more frequent than statistical independence would imply.

For the data set concerning co-infection by different strains of human papillomavirus (Chaturvedi et al., 2011), the NiSP model also passed the goodness-of-fit test, allowing us to conclude strains of this pathogen do not interact. Goodness-of-fit for such a simple model is a particularly conservative test, especially for the NiSP model, when we assume pathogens clones/strains are epidemiologically interchangeable.

We illustrated our methods via case studies for which suitable data are readily available, and our purpose was not to come to definitive conclusions concerning any particular system. That would require dedicated studies. However, by fitting even a highly-simplified version of our model to data, we have demonstrated how results of simple epidemiological models challenge previous methods based on statistical independence.

To explore further the implications of our findings, we analyzed available data sets tracking combinations of pathogens involved in each occurrence of co-infection. For methodological comparison purposes, we restricted ourselves to data referenced in Howard et al. (2001), concerning interactions between *Plasmodium* spp. causing malaria. Relaxing the assumption of epidemiological interchangeability (i.e. using the NiDP model), we found that 43.9% (i.e. 18/41) of data sets considered in Howard et al. (2001) are consistent with no interaction.

Again we do not intend to conclusively demonstrate interactions – or lack of interactions – for malaria. Instead what is important is that our results very often diverge from those originally reported in Howard et al. (2001) using a method based on statistical associations, namely log-linear regression. Log-linear regression suffers from well-acknowledged difficulties in cases in which there are zero counts (i.e. certain combinations of pathogens are not observed) (Fienberg and Rinaldo, 2012).

Such cases often arise in epidemiology. Methods based on epidemiological models therefore offer a twofold advantage: biological interactions are not confounded with statistical associations, and parameter estimation is well-posed, irrespective of zero counts.

We focused here on the simple S-I-S model, since it is sufficiently generic to be applicable to a number of systems. In future work focusing on particular pathosystems, models including additional system-specific detail would of course be appropriate. We leave further analysis of more complex underpinning epidemiological models to such future research.

Lastly, we speculate our results may have implications beyond epidemiology. After all, pathogens are species which form meta-populations occupying discrete patches (hosts) (Seabloom et al., 2015). Meta-community ecology has long been concerned with whether interactions between species can be detected from co-occurrence data (Forbes, 1907; Caswell, 1976; Connor and Simberloff, 1979) and most existing methods are based on detecting statistical associations (Gotelli, 2000; Gotelli and Ulrich, 2012), but see Hastings (1987). Our dynamical modeling approach may also provide a new perspective in this area.

4 Methods

4.1 Mathematical analyses

4.1.1 Equilibria of the two-pathogen model

The 2-pathogen model is given by Eq. (1-2-3). Since the population size is constant, $J_\varnothing = 1 - J_1 - J_2 - J_{1,2}$, and so it follows that

$$\dot{J}_\varnothing = \mu(J_1 + J_2 + J_{1,2}) - (F_1 + F_2)J_\varnothing = \mu(1 - J_\varnothing) - (F_1 + F_2)J_\varnothing \,. \tag{7}$$

It is well-known (Keeling and Rohani, 2007) that if $R_{0,i} = \beta_i/\mu > 1$ and $I_i(0) > 0$, the prevalence of pathogen i will tend to an equilibrium $\bar{I}_i = 1 - 1/R_{0,i}$.

Since $F_i = \beta_i I_i$ and $J_i = I_i - J_{1,2}$, the rate of change of co-infected hosts in Eq. (3) can be recast as

$$\dot{J}_{1,2} = \beta_2 I_2 (I_1 - J_{1,2}) + \beta_1 I_1 (I_2 - J_{1,2}) - \mu J_{1,2} \,. \tag{8}$$

The equilibrium prevalence of co-infected hosts ($\bar{J}_{1,2}$) can therefore be written in terms of the individual net prevalences at equilibrium ($\bar{I}_1$ and $\bar{I}_2$),

$$\bar{J}_{1,2} = \left(\frac{\beta_1 + \beta_2}{\beta_1 + \beta_2 - \mu} \right) \bar{I}_1 \bar{I}_2 . \tag{9}$$

This immediately leads to the result concerning the deviation of $\bar{J}_{1,2}$ from $\bar{P} = \bar{I}_1 \bar{I}_2$ (i.e., the expected prevalence of co-infected hosts given statistical independence) quoted in Eq. (4).

4.1.2 Equilibria of the *n*-pathogen model

The *n*-pathogen model is given by Eq. (1-5-6). Since the host population size is constant, $J_\varnothing = 1 - \sum_{\Gamma \in \nabla} J_\Gamma$, where ∇ is the set of all $2^n - 1$ sets with infected or co-infected hosts. It is also true that

$$\dot{J}_\varnothing = \mu(1 - J_\varnothing) - \left(\sum_{i=1}^{n} F_i \right) J_\varnothing . \tag{10}$$

At equilibrium, Eq. (5) becomes

$$0 = \sum_{i \in \Gamma} \bar{F}_i \bar{J}_{\Omega_i} - \left(\sum_{i \notin \Gamma} \bar{F}_i + \mu \right) \bar{J}_\Gamma , \tag{11}$$

in which $\bar{J}_{\Omega_i}$ and $\bar{J}_\Gamma$ are equilibrium prevalences, and $\bar{F}_i$ is the force of infection of pathogen i at equilibrium, i.e.

$$\bar{F}_i = \beta_i \bar{I}_i = \beta_i \left(1 - \frac{\mu}{\beta_i} \right) = \beta_i - \mu . \tag{12}$$

Since these forces of infection are constant and do not depend on the equilibrium prevalences, the set of $2^n - 1$ equations partially characterizing the equilibrium is linear, with

$$0 = \sum_{i \in \Gamma} (\beta_i - \mu) \bar{J}_{\Omega_i} - \left(\sum_{i \notin \Gamma} (\beta_i - \mu) + \mu \right) \bar{J}_\Gamma . \tag{13}$$

Similarly, Eq. (10) is linear

$$0 = \mu(1 - \bar{J}_\varnothing) - \left(\sum_{i=1}^{n} (\beta_i - \mu) \right) \bar{J}_\varnothing . \tag{14}$$

The equilibrium prevalences can be written very conveniently in a recursive form (i.e. using the first equation to fix $\bar{J}_\varnothing$, using $\bar{J}_\varnothing$ to independently calculate all values of $\bar{J}_\Gamma$ for $|\Gamma| = 1$, then using the set of values of $\bar{J}_\Gamma$ when $|\Gamma| = 1$ to independently calculate all values of $\bar{J}_\Gamma$ for $|\Gamma| = 2$, and so on). A recurrence relation to find all equilibrium prevalences can therefore be initiated with the following expression for the density of uninfected hosts:

$$\bar{J}_\varnothing = \frac{\mu}{\mu + \sum_{i=1}^{n}(\beta_i - \mu)} = \frac{1}{1 + \sum_{i=1}^{n}(R_{0,i} - 1)}\,. \tag{15}$$

Then, one may recursively use the following equation, equivalent to Eq. (13):

$$\bar{J}_\Gamma = \frac{\sum_{i\in\Gamma}(\beta_i - \mu)\bar{J}_{\Omega_i}}{\mu + \sum_{i\notin\Gamma}(\beta_i - \mu)} = \frac{\sum_{i\in\Gamma}(R_{0,i} - 1)\bar{J}_{\Omega_i}}{1 + \sum_{i\notin\Gamma}(R_{0,i} - 1)}\,. \tag{16}$$

Since the densities in Eq. (16) are entirely in terms of the equilibrium densities of hosts carrying one fewer pathogen ($\bar{J}_{\Omega_i}$), this allows us to recursively find the densities of all pathogens given pathogen-by-pathogen values of $R_{0,i}$.

4.1.3 Deriving the NiSP model from the NiDP model

If all pathogens are interchangeable, and so have identical values of $R_{0,i} = R_0\ \forall i$, then for any pair of combinations of infecting pathogens, Γ_1 and Γ_2, it must be the case that $\bar{J}_{\Gamma_1} = \bar{J}_{\Gamma_2}$ whenever $|\Gamma_1| = |\Gamma_2|$. This means the equilibrium prevalences of hosts infected by the same number of distinct pathogens must all be equal, irrespective of the particular combination of pathogens that is carried. In this case solving the system is much simpler. First, Eq. (11) can be rewritten as

$$0 = |\Gamma|\bar{F}\bar{J}_{\Omega_i} - \big((n - |\Gamma|)\bar{F} + \mu\big)\bar{J}_\Gamma\,, \tag{17}$$

in which $\bar{F} = \beta - \mu$. The net prevalence of hosts infected by k distinct pathogens is

$$\bar{M}_k = \sum_{\Gamma\in\nabla(k)} \bar{J}_\Gamma, \tag{18}$$

in which $\nabla(k)$ is the set of combinations of $\{1, \ldots, n\}$ with k elements. Since the form of Eq. (17) depends only on $|\Gamma|$, all individual prevalences involved in $\bar{M}_k$ are

identical, and so

$$\bar{M}_k = C_k^n \bar{J}_{\Gamma,k}, \tag{19}$$

in which C_k^n is a combinatorial coefficient, and $\bar{J}_{\Gamma,k}$ is any of the individual prevalences for which $|\Gamma| = k$. The ratio between successive values of $\bar{M}_k$ is given by

$$\frac{\bar{M}_k}{\bar{M}_{k-1}} = \frac{C_k^n}{C_{k-1}^n} \frac{\bar{J}_{\Gamma,k}}{\bar{J}_{\Gamma,k-1}} = \frac{n-k+1}{k} \frac{\bar{J}_{\Gamma,k}}{\bar{J}_{\Gamma,k-1}}. \tag{20}$$

From Eq. (15), it follows that

$$\bar{M}_0 = \frac{\mu}{\mu + n\bar{F}} = \frac{1}{1 + n(R_0 - 1)}, \tag{21}$$

in which $R_0 = \beta/\mu$. For $1 \leq k \leq n$, Eq. (17) and (20) together imply

$$\bar{M}_k = \frac{(n-k+1)\bar{F}}{(n-k)\bar{F} + \mu} \bar{M}_{k-1} = \frac{(n-k+1)(R_0-1)}{(n-k)(R_0-1)+1} \bar{M}_{k-1}, \tag{22}$$

a form which admits a simple recursive solution.

4.1.4 Stochastic models

Figs. 2B and 2C were generated by simulating the stochastic differential equation corresponding to Eq. (3); simulating a continuous time Markov chain model using Gillespie's algorithm gave consistent results. Confidence ellipses were obtained from an approximate expression for the covariance matrix at equilibrium (see below).

Continuous-time Markov chain. The continuous-time Markov chain model corresponding to the unscaled version of Eq. (3-7) tracks a vector of integer-valued random variables $X(t) = (J_\varnothing(t), J_1(t), J_2(t), J_{1,2}(t))$. Defining $\Delta X = X(t + \Delta t) - X(t) = (\Delta J_\varnothing, \Delta J_1, \Delta J_2, \Delta J_{1,2})$, changes of ± 1 to each element of $X(t)$ occur in small periods of time Δt at the rates given in Table 1. Stochastic trajectories from this model can conveniently be simulated via the Gillespie algorithm (Gillespie, 1977). Note that the numeric values of the infection rates and the host birth rate must be altered to account for the scaling by population size.

Event number	Event	Rate	Change(s) to state variable(s) (ΔX)
1	Infection of uninfected host by pathogen 1	$F_1 J_\varnothing \Delta t + o(\Delta t)$	$J_\varnothing \to J_\varnothing - 1$ $J_1 \to J_1 + 1$
2	Infection of uninfected host by pathogen 2	$F_2 J_\varnothing \Delta t + o(\Delta t)$	$J_\varnothing \to J_\varnothing - 1$ $J_2 \to J_2 + 1$
3	Infection by pathogen 1 of host singly-infected by pathogen 2	$F_1 J_2 \Delta t + o(\Delta t)$	$J_2 \to J_2 - 1$ $J_{1,2} \to J_{1,2} + 1$
4	Infection by pathogen 2 of host singly-infected by pathogen 1	$F_2 J_1 \Delta t + o(\Delta t)$	$J_1 \to J_1 - 1$ $J_{1,2} \to J_{1,2} + 1$
5	Removal of host singly-infected by pathogen 1 and replacement with an uninfected host	$\mu J_1 \Delta t + o(\Delta t)$	$J_1 \to J_1 - 1$ $J_\varnothing \to J_\varnothing + 1$
6	Removal of host singly-infected by pathogen 2 and replacement with an uninfected host	$\mu J_2 \Delta t + o(\Delta t)$	$J_2 \to J_2 - 1$ $J_\varnothing \to J_\varnothing + 1$
7	Removal of co-infected host and replacement with an uninfected host	$\mu J_{1,2} \Delta t + o(\Delta t)$	$J_{1,2} \to J_{1,2} - 1$ $J_\varnothing \to J_\varnothing + 1$

Table 1: Transitions in the two-pathogen stochastic models. The prevalence of uninfected host is $J_\varnothing$, the prevalence of each class of singly-infected hosts is J_i (for $i \in [1,2]$), and the prevalence of co-infected host is $J_{1,2}$. The net force of infection of pathogen i is $F_i = \beta_i I_i / N = \beta_i (J_i + J_{1,2})/N$ (note the scaling by the population size N relative to the forces of infection as used in the deterministic version of the model). To ensure a constant host population size, we have made the simplifying assumption that removal and replacement occur simultaneously; this has no effect on our qualitative results.

Stochastic differential equations. The model can also be written as a system of stochastic differential equations (SDEs), an approximation to the continuous-time Markov chain that is valid for sufficiently large N (Kurtz, 1970) and which is particularly well-suited for simulation of the stochastic model when the population size is large. This form of the model again tracks the seven events in Table 1, although in the SDE formulation the random variables in $X(t)$ are continuous-valued. A heuristic derivation is based on a normal approximation described below. Alternately, the forward Kolmogorov differential equations in the continuous-time Markov chain model are closely related to the Fokker Planck equation for the probability density function of the SDE model (Allen et al., 2008).

The expected change $\mathbb{E}(\Delta X)$ and covariance of the changes $\mathbb{V}(\Delta X)$ can be computed from Table 1 to order Δt via

$$\mathbb{E}(\Delta X) \approx \tilde{f} \Delta t \quad \text{and} \quad \mathbb{V}(\Delta X) \approx \mathbb{E}(\Delta X [\Delta X]^T) = \Sigma \Delta t, \tag{23}$$

where $\mathrm{d}J = \tilde{f}\mathrm{d}t$ is the unscaled version of the deterministic model as specified in Eq. (3-7) with $N = J_\varnothing + J_1 + J_2 + J_{1,2}$ (a constant) and $F_i = \beta_i (J_i + J_{1,2})/N$. In addition, the

matrix Σ is given by

$$\begin{bmatrix} \mu(N-J_\varnothing)+(F_1+F_2)J_\varnothing & -F_1J_\varnothing-\mu J_1 & -F_2J_\varnothing-\mu J_2 & -\mu J_{1,2} \\ -F_1J_\varnothing-\mu J_1 & F_1J_\varnothing+(F_2+\mu)J_1 & 0 & -F_2J_1 \\ -F_2J_\varnothing-\mu J_2 & 0 & F_2J_\varnothing+(F_1+\mu)J_2 & -F_1J_2 \\ -\mu J_{1,2} & -F_2J_1 & -F_1J_2 & F_2J_1+F_1J_2+\mu J_{1,2} \end{bmatrix}. \tag{24}$$

The changes in a small time interval Δt are approximated by a normal distribution via the Central Limit Theorem: $\Delta X(t)-\mathbb{E}(\Delta X(t)) \approx \text{Normal}(0,\Sigma\Delta t)$, where 0 = zero vector. The covariance matrix Σ can be written as $\Sigma = GG^T$. Letting $\Delta t \to 0$, the SDE model can therefore be expressed as

$$\mathrm{d}X = \tilde{f}\mathrm{d}t + G\mathrm{d}W. \tag{25}$$

The matrix G is not unique but a simple form with dimension 4×7 accounts for each event in Table 1 (Allen et al., 2008). Each entry in matrix G involves a square root and W is a vector of seven independent standard Wiener processes, where $\mathrm{d}W_i \approx \Delta W_i(t) = W_i(t+\Delta t) - W_i(t) \sim \text{Normal}(0,\Delta t)$. An explicit form for the SDE model in Eq. (25) is

$$\begin{aligned} \mathrm{d}J_\varnothing &= \tilde{f}_0\mathrm{d}t - \sqrt{F_1J_\varnothing}\,\mathrm{d}W_1 - \sqrt{F_2J_\varnothing}\,\mathrm{d}W_2 + \sqrt{\mu J_1}\,\mathrm{d}W_5 + \sqrt{\mu J_2}\,\mathrm{d}W_6 + \sqrt{\mu J_{1,2}}\,\mathrm{d}W_7\,, \\ \mathrm{d}J_1 &= \tilde{f}_1\mathrm{d}t + \sqrt{F_1J_\varnothing}\,\mathrm{d}W_1 - \sqrt{F_2J_1}\,\mathrm{d}W_4 - \sqrt{\mu J_1}\,\mathrm{d}W_5\,, \\ \mathrm{d}J_2 &= \tilde{f}_2\mathrm{d}t + \sqrt{F_2J_\varnothing}\,\mathrm{d}W_2 - \sqrt{F_1J_2}\,\mathrm{d}W_3 - \sqrt{\mu J_2}\,\mathrm{d}W_6\,, \\ \mathrm{d}J_{1,2} &= \tilde{f}_{1,2}\mathrm{d}t + \sqrt{F_1J_2}\,\mathrm{d}W_3 + \sqrt{F_2J_1}\,\mathrm{d}W_4 - \sqrt{\mu J_{1,2}}\,\mathrm{d}W_7\,. \end{aligned} \tag{26}$$

Covariance matrix at the endemic equilibrium. In Supplementary Information (Section S1.2) we show that the covariance between the prevalences of pathogen 1 and pathogen 2 as they fluctuate in the vicinity of their equilibrium values is approximately

$$\mathrm{cov}\left(\frac{I_1}{N},\frac{I_2}{N}\right) = \frac{\mu\bar{J}_{1,2}}{N^2[(\beta_1-\mu)+(\beta_2-\mu)]} = \frac{(\beta_1+\beta_2)(\beta_1-\mu)(\beta_2-\mu)\mu}{N\beta_1\beta_2(\beta_1+\beta_2-\mu)(\beta_1-\mu+\beta_2-\mu)} \geq 0\,, \tag{27}$$

with equality if and only if $\mu = 0$ (assuming $\beta_i > \mu$, $i = 1, 2$). Only in the specific case $\mu = 0$, is the deviation from statistical independence equal to zero (Eq. (4)).

4.2 Statistical methods

4.2.1 Models corresponding to assuming statistical independence

If data are observations of numbers of individuals infected with k distinct pathogens, O_k, for $k \in [0, n]$, statistical independence corresponds to assuming the infection load of a single individual follows the one-parameter, binomial model $Bin(n, p)$, in which p is the pathogen prevalence (assumed identical for each pathogen, and fitted appropriately to the data), and n is the maximum number of infections that is possible (i.e. the total number of distinct pathogens under consideration). Model predictions are then simply N samples from this binomial distribution, where $N = \sum_k O_k$ is the total number of individuals observed in the data. One interpretation is as a multinomial model in which

$$O_k \sim N q_k \quad \text{where} \quad q_k = C_k^n p^k (1-p)^{n-k}. \tag{28}$$

For the data for malaria corresponding to numbers of individuals, O_Γ, infected by different sets of pathogens, Γ, statistical independence corresponds to an n-parameter multinomial model, parameterised by the prevalences of the individual pathogens p_i (again fitted to the data), i.e.

$$O_\Gamma \sim N \prod_{i \in \Gamma} p_i \prod_{i \notin \Gamma} (1 - p_i). \tag{29}$$

4.2.2 Fitting the models

The host renewal rate, μ, can be scaled out of the equilibrium prevalences by rescaling time. Fitting the models therefore corresponds to finding value(s) for scaled infection rate(s) β_i, i.e. $R_{0,i} = \beta_i / \mu$ (all are equal for the NiSP model).

The method used to fit the model does not depend on whether the data are numbers of hosts infected by a particular combination of pathogens, or numbers of hosts carrying particular numbers of distinct pathogens, since both can be viewed as N samples drawn from a multinomial distribution, with q_j observations of the jth

class. If the corresponding probabilities generated by the model being fitted are p_j, then the log-likelihood is

$$L = \sum_j q_j \log(p_j). \tag{30}$$

The models were fitted by maximizing L via `optim()` in R (R Core Team, 2016). Convergence to a plausible global maximum was checked by repeatedly refitting the model from randomly chosen starting sets of parameters. All models were fitted in a transformed form to allow only biologically-meaningful values of parameters; that is, the basic reproduction numbers were estimated after transformation with $\log(R_{0,i} - 1)$ to ensure $R_{0,i} > 1$.

4.2.3 Model comparison

To compare the best-fitting NiSP or NiDP model and an appropriate model assuming statistical independence (binomial or multinomial), we use the Akaike Information Criterion $AIC = 2k - 2\hat{L}$, in which $\hat{L}$ is the log-likelihood of the best-fitting version of each model and k is the number of model parameters. This is necessary since these comparisons involve pairs of models that are not nested.

4.2.4 Goodness-of-fit

We use a Monte-Carlo technique to estimate p-values for model goodness-of-fit, generating $1,000,000$ independent sets of samples of total size N from the multinomial distribution corresponding to the best-fitting model, calculating the likelihood (Eq. (30)) of each of these synthetic data sets, and recording the proportion with a smaller value of L than the value calculated for the data (Sokal and Rohlf, 2012). This was done using the function `xmonte()` in the R package `XNomial` (Engels, 2015).

4.3 Sources of data and results of model fitting

4.3.1 Numbers of distinct pathogens (NiSP model)

Results of fitting the NiSP model to data from four publications for strains of a single pathogen are presented in Figure 3. Error bars are 95% confidence intervals using exact methods for binomial proportions via `binconf()` in the R package `Hmisc` (Har-

rell Jr et al., 2016). Results for three further data sets concerning different pathogens of a single host (Andersson et al., 2013; Moutailler et al., 2016; Nickbakhsh et al., 2016) are provided as Supplementary Information (Section S2.1).

For convenience the raw data as extracted for use in model fitting are re-tabulated in Table 2. Results of model fitting are summarized in Table 3. We used the value $n = 102$ for the number of distinct strains in López-Villavicencio et al. (2007) following personal communication with the authors; there might be undetected genetic differences due to missing data – which would require a larger value of n in our model fitting procedure – but we confirmed that our inferences are unaffected by taking any value of $n \in [100, 200]$.

Pathogens with n distinct		Observed counts, O_k										Total
types, strains or clones	n	0	1	2	3	4	5	6	7	8	9	N
Human papillomavirus	25	2933	1409	646	267	102	39	12	2	2	-	5412
Anther smut (*M. violaceum*)	102	285	74	60	32	14	3	3	2	1	1	475
Borrelia afzelii on bank voles	7	807	33	26	13	10	11	6	-	-	-	906
Malaria (*Plasmodium vivax*)	57	1023	404	291	208	118	50	16	5	1	1	2117

Table 2: Sources of data for fitting the NiSP model in which pathogen types, clones or strains are assumed to be epidemiologically-interchangeable. The data sets include human papillomavirus (Chaturvedi et al., 2011), anther smut (*M. violaceum*) (López-Villavicencio et al., 2007), *Borrelia afzelii* on bank voles (Andersson et al., 2013), and malaria (*Plasmodium vivax*) (Koepfli et al., 2011).

	NiSP		Binomial			GoF
	R_0	L	p	L	ΔAIC=2ΔL	p
Human papillomavirus	1.032	-6580.9	0.031	-6868.8	575.8	**0.077**
Anther smut (*M. violaceum*)	1.008	-614.0	0.008	-693.3	158.6	0.001
Borrelia afzelii on bank voles	1.044	-652.1	0.040	-799.0	293.8	0.000
Malaria (*Plasmodium vivax*)	1.021	-3169.2	0.021	-3467.3	596.3	0.000

Table 3: Fitting the NiSP model. The NiSP model was highly supported over the binomial model ($\Delta AIC \gg 10$) in all cases tested. The final column of the table – GoF – corresponds to the goodness-of-fit test of the NiSP model; values $p > 0.05$ correspond to lack of evidence for failure to fit the data, and so the NiSP model is adequate for the data concerning human papillomavirus (Chaturvedi et al., 2011).

4.3.2 Combinations of pathogens (NiDP model)

Howard et al. (2001) report results of analyzing 73 data sets concerning multiple *Plasmodium* spp. causing malaria (rows 68–140 of Table 1 in that paper). We re-analyzed the subset of these studies satisfying certain additional constraints as detailed in the main text (see Supplementary Information: Section S2.2 for a full description of how the studies were filtered). This left a final total of 41 data sets taken

from 35 distinct papers: 24 data sets considering the three-way interaction between *P. falciparum*, *P. malariae* and *P. vivax* and 17 data sets considering the three-way interaction between *P. falciparum*, *P. malariae* and *P. ovale*.

We used our method based on the NiDP model to test whether any of these data sets were consistent with no interaction between the *Plasmodium* spp. considered (Table 4). We found 15 data sets for which the NiDP model was: i) a better fit than the multinomial model as indicated by $\Delta\text{AIC} \geq 2$; ii) sufficient to explain the data as revealed by our goodness-of-fit test. In these 15 cases our methods therefore support the hypothesis of no interaction. For 11 of these 15 data sets (76, 109, 118, 130, 132, 68, 69, 70, 79, 95, 97, 98, 99, 100, 102) the results as reported in (Howard et al., 2001) instead suggest the strains interact.

5 Data availability

The cross-sectional survey data extracted from previous publications which we have used to test our methodology are tabulated in Table 2 in the main text, and Tables S2 and S4 in the Supplementary Information. These data are available in electronic format as `.csv` files from the corresponding author upon reasonable request.

6 Code availability

Code illustrating all statistical methods is freely available at: `https://github.com/nikcunniffe/Coinfection`.

7 Acknowledgments

This work was initiated during the Multiscale Vectored Plant Viruses Working Group at the National Institute for Mathematical and Biological Synthesis, supported by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville. This material is based upon research supported by the Thomas Jefferson Fund of the Embassy of France in the United States and the FACE Foundation. We thank S. Alizon, T. Berrett, E. Bussell, V. Calcagno, C. Donnelly, R. Donnelly, T. Giraud, M. López-Villavicencio, T. Obadia,

	NiDP					Multinomial						Association(s) in
	$R_{0,1}$	$R_{0,2}$	$R_{0,3}$	L	p(GoF)	p_1	p_2	p_3	L	p(GoF)	ΔAIC	Howard et al. (2001)
74	1.764	1.256	1.004	-340.7	0.000	0.468	0.220	0.004	-311.0	0.000	-59.3	✓
75	1.694	1.248	1.022	-194.8	0.000	0.445	0.215	0.022	-177.4	0.000	-34.8	✓
76	1.235	1.019	1.005	-492.5	**0.251**	0.190	0.019	0.005	-493.7	0.098	**2.4**	✓
82	1.776	1.165	1.108	-996.0	0.000	0.463	0.147	0.101	-936.3	0.000	-119.2	✓
84	1.212	1.017	1.207	-684.2	0.000	0.180	0.017	0.177	-660.4	0.000	-47.6	✓
88	1.296	1.120	1.260	-314.7	0.000	0.242	0.111	0.217	-295.0	0.000	-39.3	✓
106	1.818	1.146	1.055	-4105.2	0.000	0.442	0.125	0.052	-4296.6	0.000	382.9	✓
108	1.241	1.024	1.096	-1147.5	0.000	0.197	0.023	0.089	-1132.1	0.721	-30.9	✗
109	1.023	1.013	1.045	-359.3	**0.866**	0.023	0.013	0.043	-361.1	0.343	**3.5**	✓
111	1.198	1.005	1.786	-1929.2	0.000	0.175	0.005	0.467	-1798.8	0.000	-260.7	✓
112	1.307	1.086	1.056	-119.6	**0.115**	0.241	0.080	0.054	-116.6	0.552	-6.0	✗
113	1.213	1.007	1.119	-1324.1	0.000	0.179	0.007	0.108	-1290.8	0.000	-66.6	✓
114	1.615	1.084	1.038	-1224.4	0.000	0.392	0.080	0.037	-1182.6	0.000	-83.6	✓
116	1.780	1.124	1.100	-1035.1	0.000	0.471	0.116	0.094	-953.5	0.000	-163.2	✓
117	1.072	1.000	1.268	-31530.5	0.000	0.068	0.000	0.214	-30958.7	0.000	-1143.5	✓
118	1.085	1.039	1.171	-225.3	**0.990**	0.078	0.037	0.146	-227.5	0.515	**4.5**	✓
119	1.433	1.164	1.375	-265.7	0.000	0.325	0.146	0.291	-249.0	0.146	-33.6	✓
123	1.016	1.055	1.098	-6684.7	0.000	0.016	0.052	0.090	-6623.5	0.000	-122.4	✓
124	1.254	1.100	1.082	-3600.6	0.000	0.206	0.092	0.076	-3541.3	0.017	-118.7	✓
127	1.341	1.005	1.266	-1087.4	0.000	0.265	0.005	0.219	-1039.0	0.000	-96.8	✓
130	1.013	1.002	1.350	-352.7	**0.978**	0.013	0.002	0.259	-353.7	0.636	**2.0**	✗
132	1.397	1.027	1.074	-591.8	**0.347**	0.285	0.026	0.068	-594.3	0.067	**4.9**	✓
133	1.571	1.022	1.332	-687.9	0.000	0.375	0.022	0.257	-676.2	0.001	-23.4	✓
137	1.196	1.005	1.130	-2356.8	0.000	0.166	0.005	0.117	-2309.6	0.000	-94.3	✓
68	1.910	1.091	1.021	-152.0	**0.200**	0.469	0.082	0.020	-157.8	0.002	**11.7**	✓
69	4.827	1.443	1.036	-177.2	**0.822**	0.796	0.310	0.035	-181.4	0.121	**8.5**	✓
70	4.612	1.203	1.089	-239.2	**0.953**	0.781	0.168	0.082	-247.4	0.012	**16.4**	✓
71	6.070	1.370	1.181	-310.1	0.001	0.822	0.261	0.148	-336.2	0.000	52.1	✗
77	14.275	1.383	1.142	-155.3	0.032	0.944	0.286	0.127	-150.4	0.931	-9.9	✗
78	4.171	1.178	1.006	-166.2	**0.264**	0.773	0.153	0.006	-163.2	0.997	-5.8	✗
79	1.855	1.033	1.005	-1260.1	**0.969**	0.461	0.032	0.005	-1263.4	0.224	**6.6**	✓
80	1.546	1.062	1.021	-715.1	**0.735**	0.355	0.059	0.020	-715.2	0.675	0.2	✓
95	1.855	1.033	1.005	-1260.1	**0.970**	0.461	0.032	0.005	-1263.4	0.224	**6.6**	✗
96	1.910	1.071	1.017	-240.4	0.019	0.469	0.065	0.016	-248.9	0.000	17.1	✗
97	1.952	1.077	1.004	-242.6	**0.568**	0.486	0.071	0.004	-246.7	0.031	**8.3**	✓
98	1.662	1.014	1.018	-183.7	**0.373**	0.396	0.013	0.018	-187.0	0.030	**6.6**	✗
99	1.627	1.019	1.019	-133.7	**0.823**	0.384	0.019	0.019	-135.6	0.332	**3.8**	✗
100	1.037	1.003	1.000	-432.1	**0.254**	0.035	0.003	0.000	-433.2	0.083	**2.3**	✓
101	3.590	1.269	1.063	-11014.1	0.000	0.720	0.211	0.060	-11392.7	0.000	757.3	✓
102	2.473	1.153	1.027	-8188.9	**0.403**	0.595	0.132	0.027	-8352.0	0.000	**326.2**	✓
103	1.798	1.180	1.015	-7425.7	0.000	0.437	0.150	0.015	-7736.6	0.000	621.8	✓

Table 4: Fitting the NiDP model. Data sets which are consistent with no interaction between the *Plasmodium* spp. considered are highlighted in grey. Such data sets have both p-values for the goodness-of-fit test of the NiDP model $p(\text{GoF}) > 0.05$, and $\Delta AIC \geq 2$, meaning the NiDP model is adequate. The multinomial model corresponds to the statistical independence hypothesis. Parameters $R_{0,1}$ and $R_{0,2}$ are associated with *P. falciparum* and *P. malariae*, respectively. Parameter $R_{0,3}$ corresponds either to *P. vivax* (upper part of the table, data sets 74–137) or to *P. ovale* (lower part of the table, data sets 68–103). The final column contains a tick whenever at least one association between a pair of pathogens was assessed to be significant in Howard et al. (2001). Red ticks correspond to possible statistical associations that are consistent with our no-interaction model (NiDP), i.e. cases in which our methods lead to results diverging from those reported in Howard et al. (2001).

M. Parry, M. Plantegenest, O. Restif, E. Seabloom, J. Shykoff, R. Thompson and C. Trotter for helpful discussions or provision of data.

References

Alizon, S., de Roode, J., and Michalakis, Y. (2013). Multiple infections and the evolution of virulence. *Ecology Letters*, 16:556–567.

Alizon, S., Murall, C. L., Saulnier, E., and Sofonea, M. T. (2018). Detecting within-host interactions using genotype combination prevalence data. *bioRxiv*, (256586). ver. 3 peer-reviewed by Peer Community in Ecology.

Allen, E., Allen, L. J. S., Arciniega, A., and Greenwood, C. (2008). Construction of equivalent stochastic differential equation models. *Stochastic Analysis and Applications*, 26:274–297.

Anderson, R. and May, R. (1991). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford.

Andersson, M., Scherman, K., and Råberg, L. (2013). Multiple-strain infections of *Borrelia afzelii*: a role for within-host interactions in the maintenance of antigenic diversity? *The American Naturalist*, 181:545–554.

Antia, R., Nowak, M. A., and Anderson, R. M. (1996). Antigenic variation and the within-host dynamics of parasites. *Proceedings of the National Academy of Sciences*, 93:985–989.

Balmer, O. and Tanner, M. (2011). Prevalence and implications of multiple-strain infections. *The Lancet Infectious Diseases*, 11:868–878.

Bogaert, D., van Belkum, A., Sluijter, M., Luijendijk, A., de Groot, R., Rümke, H., Verbrugh, H., and Hermans, P. (2004). Colonisation by *Streptococcus pneumoniae* and *Staphylococcus aureus* in healthy children. *The Lancet*, 363:1871–1872.

Booth, M. and Bundy, D. (1995). Estimating the number of multiple-species geohelminth infections in human communities. *Parasitology*, 111:645–653.

Caswell, H. (1976). Community structure: a neutral model analysis. *Ecological Monographs*, 46:327–354.

Chaturvedi, A. K., Katki, H. A., Hildesheim, A., Rodríguez, A. C., Quint, W., Schiffman, M., Van Doorn, L.-J., Porras, C., Wacholder, S., Gonzalez, P., et al. (2011). Human papillomavirus infection with multiple types: pattern of coinfection and risk of cervical disease. *The Journal of Infectious Diseases*, 203:910–920.

Cohen, J. E. (1973). Heterologous immunity in human malaria. *The Quarterly Review of Biology*, 48:467–489.

Connor, E. F. and Simberloff, D. (1979). The assembly of species communities: chance or competition? *Ecology*, 60:1132–1140.

de Roode, J. C., Helinski, M. E., Anwar, M. A., and Read, A. F. (2005). Dynamics of multiple infection and within-host competition in genetically diverse malaria infections. *The American Naturalist*, 166:531–542.

Degarege, A., Legesse, M., Medhin, G., Animut, A., and Erko, B. (2012). Malaria and related outcomes in patients with intestinal helminths: a cross-sectional study. *BMC Infectious Diseases*, 12:291.

Engels, B. (2015). *XNomial: Exact Goodness-of-Fit Test for Multinomial Data with Fixed Probabilities*. R package version 1.0.4.

Fenton, A., Knowles, S. C., Petchey, O. L., and Pedersen, A. B. (2014). The reliability of observational approaches for detecting interspecific parasite interactions: comparison with experimental results. *International Journal for Parasitology*, 44:437–445.

Fienberg, S. E. and Rinaldo, A. (2012). Maximum likelihood estimation in log-linear models. *The Annals of Statistics*, 40:996–1023.

Forbes, S. A. (1907). On the local distribution of certain Illinois fishes: an essay in statistical ecology. *Bulletin of the Illinois State Laboratory of Natural History*, 7:273–303.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81:2340–2361.

Gorsich, E. E., Etienne, R. S., Medlock, J., Beechler, B. R., Spaan, J. M., Spaan, R. S., Ezenwa, V. O., and Jolles, A. E. (2018). Opposite outcomes of coinfection at individual and population scales. *Proceedings of the National Academy of Sciences*, 115:7545–7550.

Gotelli, N. J. (2000). Null model analysis of species co-occurrence patterns. *Ecology*, 81:2606–2621.

Gotelli, N. J. and Ulrich, W. (2012). Statistical challenges in null model analysis. *Oikos*, 121:171–180.

Graham, A., Lamb, T., Read, A., and Allen, J. (2005). Malaria-filaria coinfection in mice makes malarial disease more severe unless filarial infection achieves patency. *Journal of Infectious Disease*, 191:410–421.

Griffiths, E. C., Pedersen, A. B., Fenton, A., and Petchey, O. L. (2011). The nature and consequences of coinfection in humans. *Journal of Infection*, 63:200–206.

Harrell Jr, F. E., with contributions from C Dupont, and many others. (2016). *Hmisc*. R package version 4.1.1.

Hastings, A. (1987). Can competition be detected using species co-occurrence data? *Ecology*, 68:117–123.

Hellard, E., Fouchet, D., Vavre, F., and Pontier, D. (2015). Parasite–parasite interactions in the wild: How to detect them? *Trends in Parasitology*, 31:640–652.

Hellard, E., Pontier, D., Sauvage, F., Poulet, H., and Fouchet, D. (2012). True versus false parasite interactions: a robust method to take risk factors into account and its application to feline viruses. *PLoS ONE*, 7:e29618.

Hellriegel, B. (1992). Modelling the immune response to malaria with ecological concepts: short-term behaviour against long-term equilibrium. *Proceedings of the Royal Society, B.*, 250:249–256.

Hilker, F. M., Allen, L. J., Bokil, V. A., Briggs, C. J., Feng, Z., Garrett, K. A., Gross, L. J., Hamelin, F. M., Jeger, M. J., Manore, C A, P. A. G., Redinbaugh, M. G., Rúa, M. A., and Cunniffe, N. J. (2017). Modeling virus coinfection to inform management of maize lethal necrosis in Kenya. *Phytopathology*, 107:1095–1108.

Howard, S., Donnelly, C., and Chan, M.-S. (2001). Methods for estimation of associations between multiple species parasite infections. *Parasitology*, 122:233–251.

Johnson, P. T. and Buller, I. D. (2011). Parasite competition hidden by correlated coinfection: using surveys and experiments to understand parasite interactions. *Ecology*, 92:535–541.

Karvonen, A., Jokela, J., and Laine, A.-L. (2018). Importance of sequence and timing in parasite coinfections. *Trends in Parasitology*, 35:109–118.

Keeling, M. J. and Rohani, P. (2007). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton, N.J.

Koepfli, C., Ross, A., Kiniboro, B., Smith, T. A., Zimmerman, P. A., Siba, P., Mueller, I., and Felger, I. (2011). Multiplicity and diversity of *Plasmodium vivax* infections in a highly endemic region in Papua New Guinea. *PLoS Neglected Tropical Diseases*, 12:e1424.

Kurtz, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump markov processes. *Journal of Applied Probability*, 7:49–58.

Lion, S. (2013). Multiple infections, kin selection and the evolutionary epidemiology of parasite traits. *Journal of Evolutionary Biology*, 26:2107–2122.

López-Villavicencio, M., Jonot, O., Coantic, A., Hood, M. E., Enjalbert, J., and Giraud, T. (2007). Multiple infections by the anther smut pathogen are frequent and involve related strains. *PLoS Pathogens*, 3:e176.

Lord, C., Barnard, B., Day, K., Hargrove, J., McNamara, J., Paul, R., Trenholme, K., and Woolhouse, M. (1999). Aggregation and distribution of strains in microparasites. *Philosophical Transactions of the Royal Society, B.*, 354:799–807.

Malagón, T., Lemieux-Mellouki, P., Laprise, J.-F., and Brisson, M. (2016). Bias due to correlation between Times-at-risk for infection in epidemiologic studies measuring biological interactions between sexually transmitted infections: A case study using human papillomavirus type interactions. *American Journal of Epidemiology*, 184:873–883.

Man, I., Wallinga, J., and Bogaards, J. A. (2018). Inferring pathogen type interactions using cross-sectional prevalence data: opportunities and pitfalls for predicting type replacement. *Epidemiology*, 29:666–674.

Mascia, T. and Gallitelli, D. (2016). Synergies and antagonisms in virus interactions. *Plant Science*, 252:176–192.

May, R. M. and Nowak, M. A. (1995). Coinfection and the evolution of parasite virulence. *Proceedings of the Royal Society, B.*, 261:209–215.

Molineaux, L., Storey, J., Cohen, J., and Thomas, A. (1980). A longitudinal study of natural malaria in the West African Savanna in the absence of control measures: relationships between different *Plasmodium* species, in particular *P. falciparum* and *P. malariae*. *American Journal of Tropical Medicine and Hygiene*, 29:725–737.

Moutailler, S., Valiente Moro, C., Vaumourin, E., Michelet, L., Tran, F. H., Devillers, E., Cosson, J.-F., Gasqui, P., Van, V. T., Mavingui, P., Vourc'h, G., and Vayssier-Taussat, M. (2016). Co-infection of ticks: the rule rather than the exception. *PLoS Neglected Tropical Diseases*, 10:1–17.

Murall, C. L., McCann, K. S., and Bauch, C. T. (2014). Revising ecological assumptions about human papillomavirus interactions and type replacement. *Journal of Theoretical Biology*, 350:98–109.

Nickbakhsh, S., Thorburn, F., Von Wissmann, B., McMenamin, J., Gunson, R., and Murcia, P. (2016). Extensive multiplex PCR diagnostics reveal new insights into the epidemiology of viral respiratory infections. *Epidemiology & Infection*, 144:2064–2076.

Nielsen, N., Simonsen, P. E., Magnussen, P., Magesa, S., and Friis, H. (2006). Cross-sectional relationship between HIV, lymphatic filariasis and other parasitic infections in adults in coastal northeastern Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100:543–550.

Nowak, M. A., May, R. M., and Anderson, R. M. (1990). The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease. *AIDS*, 4:1095–1103.

Petney, T. N. and Andrews, R. H. (1998). Multiparasite communities in animals and humans: frequency, structure and pathogenic significance. *International Journal for Parasitology*, 28:377–393.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raso, G., Luginbühl, A., Adjoua, C. A., Tian-Bi, N. T., Silué, K. D., Matthys, B., Vounatsou, P., Wang, Y., Dumas, M.-E., Holmes, E., et al. (2004). Multiple parasite infections and their relationship to self-reported morbidity in a community of rural Côte d'Ivoire. *International Journal of Epidemiology*, 33:1092–1102.

Regev-Yochay, G., Dagan, R., Raz, M., Carmeli, Y., Shainberg, B., Derazne, E., Rahav, G., and Rubinstein, E. (2004). Association between carriage of *Streptococcus pneumoniae* and *Staphylococcus aureus* in children. *JAMA*, 292:716–720.

Reich, N. G., Shrestha, S., King, A. A., Rohani, P., Lessler, J., Kalayanarooj, S., Yoon, I.-K., Gibbons, R. V., Burke, D. S., and Cummings, D. A. (2013). Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. *Journal of The Royal Society Interface*, 10:20130414.

Rigaud, T., Perrot-Minnot, M., and Brown, M. (2010). Parasite and host assemblages: embracing the reality will improve our knowledge of parasite transmission and virulence. *Proceedings of the Royal Society, B.*, 277:3693–3702.

Rohani, P., Earn, D. J., Finkenstädt, B., and Grenfell, B. T. (1998). Population dynamic interference among childhood diseases. *Proceedings of the Royal Society, B.*, 265:2033–2041.

Rohani, P., Green, C., Mantilla-Beniers, N., and Grenfell, B. (2003). Ecological interference between fatal diseases. *Nature*, 422:885–888.

Rositch, A. F., Poole, C., Hudgens, M. G., Agot, K., Nyagaya, E., Moses, S., Snijders, P. J., Meijer, C. J., Bailey, R. C., and Smith, J. S. (2012). Multiple human papillomavirus infections and type competition in men. *Journal of Infectious Diseases*, 205:72–81.

Seabloom, E. W., Borer, E. T., Gross, K., Kendig, A. E., Lacroix, C., Mitchell, C. E., Mordecai, E. A., and Power, A. G. (2015). The community ecology of pathogens: coinfection, coexistence and community composition. *Ecology Letters*, 18:401–415.

Seabloom, E. W., Hosseini, P. R., Power, A. G., and Borer, E. T. (2009). Diversity and composition of viral communities: coinfection of barley and cereal yellow dwarf viruses in California grasslands. *The American Naturalist*, 173:E79–E98.

Shrestha, S., Foxman, B., Weinberger, D. M., Steiner, C., Viboud, C., and Rohani, P. (2013). Identifying the interaction between influenza and pneumococcal pneumonia using incidence data. *Science Translational Medicine*, 5:191ra84.

Shrestha, S., King, A. A., and Rohani, P. (2011). Statistical inference for multipathogen systems. *PLoS Computational Biology*, 7:e1002135.

Sokal, R. and Rohlf, F. (2012). *Biometry: the Principles and Practice of Statistics in Biological Research*. W.H. Freeman, New York.

Susi, H., Barrès, B., Vale, P., and Laine, A. (2015). Co-infection alters population dynamics of infectious disease. *Nature Communications*, 6:5975.

Telfer, S., Lambin, X., Birtles, R., Beldomenico, P., Burthe, S., Paterson, S., and Begon, M. (2010). Species interactions in a parasite community drive infection risk in a wildlife population. *Science*, 330:243–246.

Teweldemedhin, M., Asres, N., Gebreyesus, H., and Asgedom, S. W. (2018). Tuberculosis-human immunodeficiency virus (HIV) co-infection in Ethiopia: a systematic review and meta-analysis. *BMC Infectious Diseases*, 18:676.

Tollenaere, C., Susi, H., and Laine, A.-L. (2016). Evolutionary and epidemiological implications of multiple infection in plants. *Trends in Plant Science*, 21:80–90.

Vaumourin, E., Vourc'h, G., Telfer, S., Lambin, X., Salih, D., Seitzer, U., Morand, S., Charbonnel, N., Vayssier-Taussat, M., and Gasqui, P. (2014). To be or not to be associated: power study of four statistical modeling approaches to identify parasite associations in cross-sectional studies. *Frontiers in Cellular and Infection Microbiology*, 4:62.

Vaumourin, E., Vourc'h, G., Gasqui, P., and Vayssier-Taussat, M. (2015). The importance of multiparasitism: examining the consequences of co-infections for human and animal health. *Parasites & Vectors*, 8:545.