Reconstructing heterogeneous pathogen interactions from co-occurrence data via statistical network inference

Irene Man^{1,2*#a}, Elisa Benincà¹, Mirjam E. Kretzschmar², Johannes A. Bogaards^{1,3}

1 Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

2 Julius Centre, UMC Utrecht, Utrecht University, Utrecht, The Netherlands

3 Department of Epidemiology & Data Science, Amsterdam UMC, Amsterdam, The Netherlands

#a Current Address: EPR Branch, IARC, 150 Cours Albert Thomas, 69008 Lyon, France

* irene.y.s.man@gmail.com

Short title: Reconstructing heterogeneous pathogen interactions

Abstract

Infectious diseases often involve multiple pathogen species or multiple strains of the same pathogen. As such, knowledge of how different pathogen species or pathogen strains interact is key to understand and predict the outcome of interventions that target only a single pathogen or subset of strains involved in disease. While population-level data have been used to infer pathogen strain interactions, most previously used inference methods only consider uniform interactions between all strains, or focus on marginal interactions between pairs of strains (without correction for indirect interactions through other strains). Here, we evaluate whether statistical network inference could be useful for reconstructing heterogeneous interaction networks from cross-sectional surveys tracking co-occurrence of multi-strain pathogens. To this end, we applied a suite of network models to data simulating endemic infection states of pathogen strains. Satisfactory performance was demonstrated by unbiased estimation of interaction parameters for large sample size. Accurate reconstruction of networks may require regularization or penalizing for sample size. Of note, performance deteriorated in the presence of host heterogeneity, but this could be overcome by correcting for individual-level risk factors. Our work demonstrates how statistical network inference could prove useful for detecting pathogen interactions and may have implications beyond epidemiology.

Introduction

The rapid expansion of sequencing technologies over the last decades has drastically increased our ability to detect within-host pathogen diversity. As a result, it is recognised that infectious diseases often involve multiple pathogen species. For example, diseases may arise from bacterial and viral co-infections [1,2], or co-infections of multiple strains of the same pathogen species, e.g. serotypes of *Streptococcus pneumoniae* [3], genotypes of the human papillomavirus [4], or clones of *Plasmodium falciparum* [5]. It is also increasingly understood that networks of microbial interactions may determine the outcome of preventive or therapeutic interventions (e.g. vaccination, antibiotics, probiotics) in unexpected ways [6–8]. For infectious agents, these interactions between pathogens or pathogen strains can take various forms: infection by one strain can modify susceptibility to subsequent infection by others, and the simultaneous presence of multiple strains can affect the duration, infectiousness, as well as severity of infection [9,10]. Consequently, pathogen interactions may have far-reaching clinical, epidemiological, and eco-evolutionary implications [11–13].

The acknowledgement that pathogen strain interaction is a crucial factor shaping the ecology for a wide number of infectious agents has spurred the development of epidemiological models where hosts can be infected by different pathogens or different strains of the same pathogen [14]. However, inference of interaction parameters in these models is notoriously challenging, especially when pathogen strain interaction is to be extracted from population-level data. The availability of large-scale cross-sectional surveys, detecting the presence of multiple infectious strains simultaneously, has given rise to a wide range of methods for detecting interactions from co-occurrence data [15–20]. These methods rely on identifying deviation from statistical independence, either in the number of pathogen strains occurring in individual hosts, pairwise co-occurrence between any two strains, or the exact combination of co-occurrence.

Clearly, the validity of such statistical methods for detecting pathogen interactions hinges on the assumption that non-interaction implies statistical independence, but statistical associations can be driven by factors other than biological interactions [21]. This point has been made repeatedly by community ecologists, concerned with the possibility to detect interdependence among species from co-occurrence patterns across

habitats [22], and has been reiterated by epidemiologists [23]. A common criticism of using statistical associations is that hosts or habitats are not identical, and this may favour the presence of one or another (or both) species irrespective of biological interactions. Opportunely, epidemiological studies on human pathogens typically offer rich meta-data, making it possible (in principle) to correct for host heterogeneity. Moreover, when studying different strains of the same pathogen, one might assume that the same hosts are more likely than others to become infected with all strains, justifying correction for host heterogeneity by host-specific covariates in statistical models [23].

So far, many existing methods for detecting multi-strain pathogen interactions are restricted to uniform (or homogeneous) interactions across strains (i.e. all strains under consideration interact with one another in the same manner), or focused on marginal interactions between pairs of pathogen strains. Marginal estimation of pairwise interactions does not necessarily control for indirect interactions through other strains and may therefore be prone to bias, especially when strains interact in a heterogeneous manner. Network inference methods relying on conditional dependency between the co-occurrence of multiple strains might be more suited for detecting heterogeneous pathogen strain interactions (e.g. interactions that are different between each pair of strains), but are not widely considered. Moreover, it is not yet well studied how the performance of these methods might be affected by the presence of host heterogeneity.

The purpose of this paper is two-fold. First, we explore the possibility of detecting heterogeneous pathogen interactions from cross-sectional survey data using statistical network inference methods. For this, we evaluate the performance of a suite of network models in reconstructing simulated networks of pathogen strain interactions for a range of epidemiological and network settings. Second, we investigate the robustness of these methods to settings with host heterogeneity that is relevant for pathogen spread. In this work, we focus on reconstruction of interaction network of pathogen strains, but the methodology is applicable to cross-sectional surveys tracking co-occurrence of potentially interacting pathogen species and may have implications beyond epidemiology.

Results

In order to evaluate the performance of various network inference methods, we simulated random interaction networks with up to 10 pathogen strains, which describe whether and to what extent the strains interact in an epidemiological model. Within each network, the presence of interaction between each pair of strains i and j was established with connection probability σ . Given that interaction between a pair was established, the strength of interaction, indicated by the value of x_{ij} , was drawn uniformly from the range $[-\theta, \theta]$, where θ is a positive number. Values of x_{ij} in the range $[-\theta, 0)$ and $(0, \theta]$ correspond to competitive and mutualistic interactions, respectively, and the greater the value diverges from zero, the stronger the interaction becomes.

Each network generated as such was then used to parametrize a Susceptible-Infected-Susceptible (SIS) epidemiological model [8]. In the model, $\exp(x_{ij})$ was used to as the hazard ratio that determines how the presence of strain i in a host alters the host's hazards to acquire or clear strain j , and vice versa. See Material and methods for the exact description of how parameter x_{ij} is utilized in the parametrization of the epidemiological model. After parametrization, the steady state prevalence of the epidemiological model was obtained and used to sample random cross-sectional datasets with the number of observations being 100, 1,000, 10,000, or 100,000. These datasets were then used to reconstruct the interaction networks. See S1 Fig for a schematic representation of the entire data generation process.

The set of network inference methods that we considered were the Ising model, graphical modelling approaches, and generalized estimating equations (GEE) [24–26]. For the Ising model and GEE, model selection was done based lasso (least absolute shrinkage and selection operator) and Wald tests with false discovery rate control, respectively. Graphical model selection was performed either in a forward or a backward fashion (i.e. by iteratively growing or shrinking networks, respectively), using either AIC or BIC as selection criterion.

In the base-case setting, we considered 500 random networks with 5 strains, each generated with connection probability $\sigma = 0.25$ and a range of interaction parameters of $[-\theta, \theta] = [-\log 3, \log 3]$. In the corresponding epidemiological model, we considered interaction in acquisition only (i.e. no interaction in clearance), a population without

Fig 1. Examples of networks with $n = 10$ strains in the simulation study.

Two random networks generated with connection probability $\sigma = 0.25$ and interaction strengths x_{ij} drawn from $[-\theta, \theta] = [-\log 3, \log 3]$ (A and D); Estimated networks from co-occurrence in 100,000 observations according to the regularized Ising model (B and E); and according to a graphical model using backward model selection by BIC (C and F). Strength and type of interactions (mutualistic versus competitive) are indicated by the thickness and colour (green versus red) of edges, respectively.

host heterogeneity, and type-specific basic reproduction numbers $R_{0,i}$ in the range

[1.5, 2].

In the base-case setting, visual inspection showed satisfactory concordance between the estimated networks and true networks. This is illustrated in Fig 1, which also shows some differences between network inference methods. In order to compare the different methods systematically, we assessed various performance measures, including sensitivity and specificity for indicating the presence (or absence) of pairwise interactions, i.e. the proportions of truly present (or truly absent) interactions that are correctly identified in the reconstructed networks. In addition, we assessed the F1-score, which summarizes sensitivity and positive predictive value (PPV, i.e. the proportion of identified interactions that are truly present) in the form of their harmonic mean.

Trade-off between sensitivity and specificity

Network reconstruction improved with the number of observations according to almost all assessed performance measures for all considered methods (Fig 2A-D, Table 1).

Interactions were recovered with almost 100% sensitivity at 100,000 observations (Fig 2A). Optimal specificity was attained already at small sample size in graphical modelling approaches with BIC-based, whereas specificity remained at a more or less constant suboptimal level in graphical modelling approaches with AIC-based selection (Fig 2D). GEE was good at small sample size, but was the only method considered with slightly deteriorating performance with increasing sample size notwithstanding correction for multiple testing (Fig 2B-D). Overall, methods with less optimal specificity showed better sensitivity, showcasing trade-off between sensitivity and specificity. Finally, the Ising model and the graphical modelling approaches with BIC-based selection were able to achieve relatively high values of sensitivity and specificity at high sample size, which resulted in near-optimal F1-scores. (Fig 2B,C).

Table 1. Performance measures of statistical network inference in the base-case analysis. In the base-case analysis, interactions between strains were generated with connection probability $\sigma = 0.25$, interaction strengths x_{ij} indicating interaction in acquisition drawn from $[-\log 3, \log 3]$, and basic reproduction numbers randomly drawn from $[1.5, 2]$. All methods were evaluated at sample sizes of 100, 1,000, 10,000 and 100,000 observations. Abbreviations: GEE, generalized estimating equations; PPV, positive predictive value; RMSE, root-mean-square error.

Method	Sample size §	Specificity (%) ±	Sensitivity (%) ±	PPV (%) ±	F1-score †±	Pearson's r *±	RMSE ±
<i>Ising model</i>	100,000	99.9 (99.8; 100.)	93.7 (92.3; 95.1)	99.6 (99.2; 99.9)	.966 (.958; .974)	.999 (.998; .999)	.038 (.036; .040)
	10,000	99.6 (99.4; 99.8)	85.1 (82.8; 87.0)	98.6 (97.8; 99.3)	.913 (.899; .925)	.996 (.995; .996)	.079 (.075; .083)
	1,000	99.5 (99.3; 99.7)	51.5 (48.5; 54.3)	97.1 (95.6; 98.4)	.673 (.646; .697)	.969 (.965; .973)	.218 (.208; .228)
	100	99.3 (99.1; 99.6)	7.9 (6.3; 9.7)	78.0 (70.3; 85.4)	.144 (.115; .172)	.847 (.805; .887)	.606 (.537; .675)
<i>GEE</i>	100,000	90.4 (89.1; 91.7)	96.3 (95.3; 97.4)	75.6 (73.8; 77.7)	.847 (.834; .861)	.995 (.994; .996)	.055 (.050; .060)
	10,000	95.8 (95.1; 96.6)	89.9 (88.1; 91.8)	86.9 (85.0; 88.8)	.884 (.871; .895)	.991 (.989; .992)	.085 (.077; .093)
	1,000	98.6 (98.2; 99.0)	60.9 (58.2; 64.1)	92.9 (91.1; 94.6)	.736 (.714; .757)	.969 (.966; .973)	.189 (.177; .201)
	100	99.1 (98.8; 99.4)	7.2 (5.6; 8.1)	70.8 (62.0; 79.0)	.131 (.103; .163)	.876 (.837; .919)	.946 (.851; 1.041)
<i>Graphical model with backward AIC selection</i>	100,000	76.2 (74.4; 78.1)	97.6 (96.7; 98.4)	55.9 (53.7; 58.4)	.711 (.688; .730)	.998 (.997; .998)	.033 (.029; .037)
	10,000	71.9 (70.1; 73.7)	94.7 (93.2; 96.0)	50.9 (48.5; 53.2)	.662 (.643; .681)	.987 (.983; .989)	.075 (.067; .083)
	1,000	71.1 (68.8; 73.1)	79.7 (77.4; 81.9)	46.0 (43.7; 48.4)	.584 (.562; .602)	.918 (.908; .927)	.200 (.190; .210)
	100	67.1 (65.1; 69.1)	50.6 (47.6; 53.7)	32.0 (30.0; 34.3)	.392 (.372; .413)	.619 (.592; .646)	.656 (.633; .679)
<i>Graphical model with backward BIC selection</i>	100,000	99.8 (99.6; 99.9)	94.4 (93.1; 95.7)	99.2 (98.7; 99.7)	.967 (.960; .974)	.999 (.998; .999)	.033 (.027; .040)
	10,000	99.8 (99.6; 99.9)	85.9 (84.0; 87.8)	99.2 (98.5; 99.7)	.921 (.910; .932)	.995 (.994; .996)	.066 (.057; .075)
	1,000	99.2 (99.0; 99.5)	58.7 (56.0; 61.4)	95.9 (94.5; 97.3)	.729 (.704; .751)	.974 (.971; .977)	.178 (.168; .188)
	100	96.0 (95.2; 96.7)	18.4 (16.1; 20.8)	58.5 (53.2; 63.3)	.280 (.252; .306)	.769 (.734; .800)	.918 (.864; .972)
<i>Graphical model with forward AIC selection</i>	100,000	86.2 (85.1; 87.2)	97.2 (96.3; 98.1)	68.4 (66.4; 70.6)	.803 (.788; .819)	.998 (.997; .998)	.034 (.030; .038)
	10,000	85.2 (84.1; 86.3)	93.5 (92.0; 94.9)	66.1 (64.0; 68.3)	.775 (.756; .790)	.988 (.985; .991)	.081 (.071; .091)
	1,000	84.6 (83.3; 85.8)	76.4 (73.8; 78.7)	60.4 (58.0; 62.7)	.675 (.653; .694)	.925 (.914; .933)	.224 (.212; .236)
	100	83.3 (82.2; 84.5)	39.6 (36.9; 42.2)	42.2 (39.0; 45.0)	.408 (.382; .432)	.660 (.634; .685)	.806 (.781; .831)
<i>Graphical model with forward BIC selection</i>	100,000	99.9 (99.8; 100.)	94.1 (92.8; 95.3)	99.6 (99.1; 99.9)	.967 (.960; .974)	.999 (.998; .999)	.033 (.027; .040)
	10,000	99.8 (99.7; 100.)	85.8 (83.8; 87.6)	99.4 (98.7; 99.9)	.921 (.910; .932)	.995 (.994; .996)	.066 (.057; .075)
	1,000	99.2 (98.9; 99.5)	58.7 (56.0; 61.4)	95.8 (94.3; 97.2)	.728 (.704; .751)	.974 (.970; .977)	.179 (.169; .189)
	100	96.5 (96.0; 97.1)	17.5 (15.5; 19.8)	60.8 (55.7; 65.3)	.271 (.241; .298)	.763 (.727; .793)	.987 (.938; 1.036)

§ Number of sampled individuals

± 95% confidence intervals (bootstrap percentile estimates) in brackets

† F1-score denotes the harmonic mean between sensitivity (also termed recall) and PPV (also termed precision)

* Linear correlation coefficient between generated and estimated interaction strengths (non-zero estimates only)

Fig 2. Performance measures of statistical network inference in the base-case analysis. Several measures have been calculated to assess the performance of the statistical network inference as a function of the sample size for the base-case analysis. A) Sensitivity; B) Positive predictive value (PPV); C) F1 score; D) Specificity; E) Pearson correlation; F) Root-mean-square deviation. In the base-case analysis, interactions between strains were generated with connection probability $\sigma = 0.25$, interaction strengths x_{ij} indicating interaction in acquisition drawn from $[-\log 3, \log 3]$, and basic reproduction numbers randomly drawn from $[1.5, 2]$. All methods were evaluated at sample sizes of 100, 1,000, 10,000 and 100,000 observations, but x-axis coordinates are slightly jittered to improve visualization. Abbreviations: bw.aic (turquoise): graphical model with backward AIC selection; bw.bic (red): graphical model with backward BIC selection; fw.aic (purple): graphical model with forward AIC selection; fw.bic (pink): graphical model with forward BIC selection; gee (green): generalized estimating equations; Ising (yellow): Ising model.

Unbiased estimation of the strength of interaction

In addition, we assessed the quantitative agreement between the estimated and the true interaction parameter x_{ij} , which indicates the strength of interaction, based on Pearson correlation and root-mean-square deviation. Average Pearson correlations between the true and estimated interactions were high for all methods with 1,000 observations or more (Fig 2E). Under the smallest sample size involving only 100 observations, fewer interactions were recovered and the estimated interaction strength deviated somewhat from the true parameters, but even in this case the average correlation between true and estimated interaction strength remained substantial, especially under GEE or regularized Ising estimation (Table 1). Moreover, while correlations between true and estimated interactions reached a plateau above 1,000 observations, accuracy continued to improve with increasing sample size, as verified by steadily decreasing root-mean-square deviations for all methods considered, showcasing asymptotically unbiased estimation (Fig 2F).

Performance under alternative epidemiological settings and network configurations

In a series of sensitivity analyses, we modified the base-case setting to alternative epidemiological settings or network configurations: inclusion of interaction in clearance, lower basic reproduction numbers $R_{0,i} \in [1, 1.5]$, stronger interactions $x_{ij} \in [-\log 10, \log 10]$, higher connection probabilities $\sigma = 0.5$, larger networks by increasing the number of strains to 10, or larger networks by connecting two

Fig 3. F1-scores of statistical network inference methods in the sensitivity analyses. The performance of different network inference methods expressed by F1-score as a function of the sample size. A) Ising model; B) generalized estimating equations; and graphical models with C) forward BIC selection; D) forward AIC selection; E) backward BIC selection; F) backward AIC selection. The alternative settings considered in the sensitivity analyses were: acquisition and clearance (pink): interaction strengths x_{ij} including both interactions in acquisition and clearance; reproduction number (turquoise): lower basic reproduction numbers from the range $[1, 1.5]$ instead of $[1.5, 2]$; strength (purple): strong interaction strengths x_{ij} drawn from $[-\log 10, \log 10]$ instead of $[-\log 3, \log 3]$; network size (yellow): larger networks with 10 strains; sub-networks (green): larger networks with 10 strains created by collating sub-networks of 5 strains; connection probability (red): higher connection probability σ being 0.5 instead of 0.25. Performance of the base-case analysis is given by black horizontal lines.

sub-networks of 5 strains.

Network inference was not affected by including interactions in clearance relative to the base-case setting. In this setting, F1-scores remained the same for all methods considered (compare pink dots to black lines in Fig 3). This implies that the combined effect of interactions in acquisition and clearance was well captured by the estimated interaction parameters. Specifically, the ratio of the interaction parameters in acquisition and clearance was well estimated.

Looking across all alternative settings, regularized Ising, GEE estimation and BIC-graphical modelling approaches maintained the high performance of the base-case analysis when the sample size was 10,000 or larger (Fig 3A,B,C,E). For these methods, when the sample size was 1,000 or smaller, performance stayed more or less the same with higher connection probability or lower reproduction numbers (compare red and turquoise dots to black lines), improved with stronger interactions (compare purple dots to black lines) but slightly deteriorated when considering larger networks of 10 strains (compare green and yellow dots to black lines).

As for the AIC-based graphical modelling approaches, the performance diverged more from that of the base-case setting (Fig 3D,F). The performance improved substantially under higher connection probabilities (compare red dots to black lines). However, it suffered also more when networks were larger (compare green and yellow dots to black lines). These patterns are linked to the high sensitivity and poor specificity of the AIC-based approaches (as shown in the base-case analysis).

Fig 4. F1-scores of statistical network inference under host heterogeneity. The performance of the different inference methods evaluated under the setting with host heterogeneity as a function of the sample size. A) Ising model; B) generalized estimating equations; and graphical models with C) forward BIC selection; D) forward AIC selection; E) backward BIC selection; F) backward AIC selection. Similar epidemiological models are used as in Fig 2, but with two sub-populations of hosts. Average contact rate is the same as in the base-case analysis, but 80% of hosts is assumed to have below-average contacts and 20% above-average contacts (coefficient of variation: 80%). Mixing between sub-populations occurred pseudo-assortatively (assortivity fraction: 50%). Performance is investigated in following ways: uncorrected: based on representatively sampled individuals from the total population without correction for contact rate; corrected: idem but with correction for contact rate; low-risk: (stratified) analysis on individuals sampled from sub-population with low contact rate only; high-risk: (stratified) analysis on individuals sampled from sub-population with high contact rate only.

Robustness to host heterogeneity

Finally, we tested the robustness of the network inference methods to host heterogeneity in contact rate relevant to pathogen spread. The host population was divided into two interconnected sub-populations: one with high and one with low contact rate. Network inference strongly deteriorated in the presence of host heterogeneity as compared to the base-case analysis when heterogeneity was not corrected for (compare turquoise dots to black lines in Fig 4, compare D to A in Fig 5). For all methods considered, F1-scores did not improve with increasing sample size. Similarly, root-mean-square deviation remained high at large sample size (S1 Table). These findings show a consistent bias towards positive associations in uncorrected analyses (S2 Fig).

Network inference regained good performance if host heterogeneity was corrected for by performing stratified analyses based on the risk variable indicating membership to either sub-population (compare purple and pink dots to black lines in Fig 4; compare B and C to A in Fig 5). At moderate sample size (1.000 and 10.000), performance was somewhat better in subgroup analyses on high-risk individuals, likely due to increased statistical power at higher prevalence of infection. The alternative correction approach by adding the risk variable as an extra node to the network, representing elevated infection risk, also performed well (compare red to turquoise in Fig 4). Graphically, the risk variable can be represented as a central node equally connected to all other nodes representing carriage of strains (Fig 5E). This dependency illustrates that elevated infection risk is associated with positivity for each strain. After filtering out the risk

Fig 5. An example of network reconstruction under host heterogeneity. The true random network was generated with connection probability $\sigma = 0.25$, interaction strengths drawn from $[-\log 3, \log 3]$, and basic reproduction numbers randomly drawn from $[1.5, 2]$ (A). Estimated networks were obtained using the regularized Ising model from a dataset with 10,000 individuals in a stratified analysis among low-risk individuals only (B), high-risk individuals only (C), among a representative sample of the total population without correction (D), or with correction by augmenting the network with an extra node R, indicating membership to either sub-population (E). The filtered network (F) is obtained from (E) by omitting node R and the corresponding edges. Strength and type of interactions (mutualistic versus competitive) are indicated by the thickness and colour (green versus red) of edges, respectively.

variable, the strain-specific interaction network is retained (Fig 5F)

179

Discussion

In this paper, we evaluated whether heterogeneous pathogen strain interactions can be recovered from cross-sectional surveys tracking co-occurrence of multi-strain pathogens via statistical network inference. Using simulated data, we demonstrated unbiased estimation of pairwise interactions between multiple infectious strains, appropriately correcting for indirect interactions through other strains. Performance of network inference was shown to be strongly influenced by host heterogeneity, but we also demonstrated how this can be overcome by correcting for individual risk factors common to all pathogen strains.

Conceivably, the nature and strength of multi-strain interactions may indeed be heterogeneous, as these are modulated by within-host niche differentiation. Such differentiation may arise from slight differences in host-cell tropism, as e.g. described for oncogenic HPV infection, where genotypes of alpha-7 and alpha-9 species have distinct preference for glandular and squamous epithelium, respectively [4,27]. Moreover, if multi-strain interactions are mediated by within-host immune responses, as hypothesized for different pneumococcal serotypes, strength of interaction may depend on antigenic overlap between strains [3]. To uncover heterogeneous interactions, it is essential to analyse datasets where the exact combination of strains in each host is recorded. Previous endeavours to infer competitive interactions from such genotype combination data found that they only performed satisfactorily in datasets where competitive interactions are particularly strong, and that accounting for host behavioural heterogeneity is more important than adding additional information via the genotype combinations [20,21]. The importance of accounting for host heterogeneity is reiterated in our analysis, but we also demonstrate that network inference is well suited to detect small interactions at large enough sample size, provided that the mechanisms of interaction can be inferred from conditional independence structures.

Our results were obtained under a particular epidemiological model and under a steady state assumption, with between-strain interactions modelled in acquisition and clearance. We have previously shown that the odds ratio of co-occurrence in a two-strain version of this model is an exact estimator of the composite of interaction parameters defined accordingly [28]. The results of the present study should be viewed

as an extension of that result to systems with an arbitrary number of strains. Network models are well suited to capture conditional independence between multiple strains, providing a clear advantage over marginal estimation of pairwise interactions as these may be confounded through indirect interactions with other strains. While we showed excellent performance of the network inference methods in this respect, we cannot ensure good performance under epidemiological models other than the one we considered. For instance, if strains interact through natural cross-immunity that is long-lasting, infections with different strains may be positively associated (i.e. as given by odds ratios greater than one) and co-occurrence patterns of current infection are no longer informative of their interaction structure [28]. Likewise, strains may interact through modification of viral or bacterial load during co-colonization, rather than through modification of acquisition or clearance. Although such interactions are better captured by records of viral or bacterial load, they are likely correlated with interactions in acquisition and clearance. Hence, network inference based on co-occurrence may retain good performance even in such situations. More extensive simulations are needed to demonstrate the robustness of the network inference methods to variations in model specification. In general, more complex mechanisms of interactions may be better identifiable from individual- or population-level longitudinal studies than cross-sectional surveys [29,30].

Likewise, performance of network inference may depend on the way that interactions between multiple strains combine during co-infection. Previously, we demonstrated that predictors of strain-replacement in the wake of vaccination against a subset of pathogen strains, perform best when interactions in acquisition and clearance combine multiplicatively [8]. We envisage that reconstruction of heterogeneous strain interaction networks from co-occurrence data becomes more complicated if interactions combine otherwise than multiplicatively, e.g. if strength of interaction is not affected by multiplicity of infection. It should be noted that estimates obtained under GEE or Ising models are also marginal in this sense, as they only consider pairwise interactions and leave possible higher-order interactions unspecified. Conversely, graphical log-linear models implicitly incorporate a three-way interaction whenever three strains are graphically connected, allowing pairwise interactions to be modulated by a third strain [25]. Suitability of either approach is determined by the likelihood of higher-order

interactions being present in a particular system, and by the quality and resolution of the available data.

While we did not directly compare the methods applied here to previous approaches, we envisage that the comparatively good performance of statistical network inference derives from their rigour and computational efficiency. These methods bypass specification of explicit models for interacting strain dynamics, and instead capitalize on the implicit relation between the interaction parameters and the expected proportions of the host population that carry different combinations of infectious strains. Consistency of methods depends on whether the underlying interaction mechanisms are appropriately captured, but the same applies to explicit epidemiological modelling of co-infections. Both approaches may suffer from model misspecification, but if approximations hold, statistical network inference provides a particularly powerful tool, as shown here.

All network modelling approaches considered in this paper show unbiased estimation of interaction parameters, evidenced by diminishing RMSD for increasing sample size. However, we found that different methods excel in different aspects of estimation, with a general trade-off between non-discovery and false discovery. Network structure was only captured appropriately when applying some form of regularization (as in the Ising model) or penalizing for sample size (as in BIC-based log-linear model selection). In comparison, GEE or AIC-based model selection suffered from poor specificity, yielding comparatively low PPV at large sample size. Even so, these methods exhibited negligible RMSD, demonstrating that false discovery at large sample size predominantly pertains to supposedly small interaction strengths. Penalization ensures that very weak connections are set to zero; hence, the appropriateness of using AIC- or BIC-based model selection depends on the sparsity of the true interaction network being analysed. In our simulations, sparsity of the networks used to generate co-occurrence data was controlled by fixing the rate of true zeroes in the network. The extent to which true independence holds in reality should determine which class of network inference methods is to be preferred [31]. From a practical point of view, however, the important interactions were almost always correctly identified, also at limited sample size.

Finally, we show that a regression framework familiar to most epidemiologists, i.e. GEE, performs satisfactorily in estimating pairwise interactions. The GEE regression

framework has the additional benefit that it offers a flexible way of separating
individual risk factors common to all strains, from interactions between any two strains
[23]. This framework has been employed before to study clustering patterns of HPV
genotypes across risk populations, with correction for known risk factors [18]. While
correction for individual-level characteristics is also possible in other methods, as
demonstrated in the present study, most other graphical modelling approaches rely on
the inclusion of risk factors as binary nodes to the network. This clearly limits the
practical ability to correct for various sources of host heterogeneity, that are more
naturally accommodated in a regression framework [26, 32]. Compared to other
graphical modelling approaches, the GEE framework is not scalable to situations where
the number of strains exceeds the number of observations. However, for most
epidemiological studies this does not pose a real problem.

To summarize, we have demonstrated how interactions between multiple pathogen
strains might be estimated from cross-sectional surveys, detecting the presence of
multiple infectious strains at once. We illustrated this by applying statistical network
models, that properly account for conditional independence between strains and are
able to efficiently and consistently estimate heterogeneous interactions. Our work
demonstrates how widely available multivariate methods may be used to identify
between-strain interactions from readily available data.

Materials and methods

Multi-strain epidemiological model

The epidemiological model we used for simulation was previously introduced to assess the outcome of vaccination against a subset of pathogen strains present in a host population [8]. In short, this model allows for an arbitrary number of infectious strains to circulate, assuming Susceptible-Infected-Susceptible (SIS) dynamics with regard to each of the individual strains. It applies to pathogens for which naturally acquired immunity is limited so that reinfection is possible [33], as for instance could be the case for infection with human papillomavirus or with *Streptococcus pneumoniae*. For a pathogen with n strains, the model state space $\mathcal{S}$ consists of $2n$ infection states, each denoted by the set of strains the host is infected with. See S3 Fig for a schematic representation of the possible infection states with $n = 3$. A system of ordinary differential equations describes how the proportions of individuals in each of the model states evolve over time. The equation for the change in prevalence N_X of any state $X \in \mathcal{S}$ is generally given by

$$\begin{aligned} \frac{dN_X}{dt} = & \sum_{i \in X} N_{X \setminus \{i\}} q_{X \setminus \{i\} \rightarrow X} - \sum_{i \in X} N_X q_{X \rightarrow X \setminus \{i\}} \\ & - \sum_{i \notin X} N_X q_{X \rightarrow X \cup \{i\}} + \sum_{i \notin X} N_{X \cup \{i\}} q_{X \cup \{i\} \rightarrow X}. \end{aligned} \quad (1)$$

The four groups of terms on the right-hand side, moving from left to right, correspond to flows of individuals into or out of state X due to 1) acquisition of a strain $i \in X$, 2) clearance of a strain $i \in X$, 3) acquisition of a strain $i \notin X$, and 4) clearance of a strain $i \notin X$, respectively. Each term is a product of the proportion of the population in a particular state and the corresponding acquisition or clearance hazards (per capita rates) denoted by q . In an example with three strains and $X = \{1\}$, the corresponding terms are 1) $N_{\emptyset} q_{\emptyset \rightarrow \{1\}}$, 2) $N_{\{1\}} q_{\{1\} \rightarrow \emptyset}$, 3) $N_{\{1\}} q_{\{1\} \rightarrow \{1,2\}} + N_{\{1\}} q_{\{1\} \rightarrow \{1,3\}}$, and 4) $N_{\{1,2\}} q_{\{1,2\} \rightarrow \{1\}} + N_{\{1,3\}} q_{\{1,3\} \rightarrow \{1\}}$.

The baseline hazard of clearance for an individual only infected with strain i is denoted by $q_{\{i\} \rightarrow \emptyset}$ and the baseline hazard for a completely susceptible individual to

acquire strain i is given by

$$q_{\emptyset \rightarrow \{i\}} = c\beta_i \sum_{X \in \mathcal{S}_i} I_X. \quad (2)$$

Here, c is the per capita rate at which hosts make contacts that are relevant for pathogen spread, β_i is the probability of successfully acquiring strain i given contact with an individual infected with strain i , and $\mathcal{S}_i \subset \mathcal{S}$ is the subset of states containing strain i .

Potential interactions between strains are modelled through modification of the baseline acquisition or clearance hazards. The model allows for different structures of interactions, but we only consider the pairwise-symmetric multiplicative structure (see [8] for the alternative structures). To be exact, we assume each strain that is carried to contribute multiplicatively to the hazard of acquiring (or clearing) the incoming (or outgoing) strain. Hence, the hazards of acquisition and clearance in the presence of other strains are

$$\begin{aligned} q_{X \rightarrow X \cup \{i\}} &= \prod_{j \in X} k_{ij} q_{\emptyset \rightarrow i}, \\ q_{X \setminus \{i\} \rightarrow X} &= \prod_{j \in X} h_{ij} q_{i \rightarrow \emptyset}, \end{aligned} \quad (3)$$

where the involved interaction parameters are pairwise-symmetric, i.e. $k_{ji} = k_{ij}$ and $h_{ji} = h_{ij}$ for all i and j . Defined as such, k_{ij} and h_{ij} are essentially hazard ratios that act on the baseline hazards. In particular, values equal to one imply no interaction and the deviation from one indicates the strength of interaction. For the mode of acquisition, $k_{ij} < 1$ and $k_{ij} > 1$ indicate competitive and mutualistic interactions, respectively. Conversely, for the mode of clearance, $h_{ij} < 1$ and $h_{ij} > 1$ indicate mutualistic and competitive interactions, respectively. Previously, we have shown that when interaction is present in both modes, the overall interaction could be summarized by the ratio of the two interaction parameters k_{ij}/h_{ij} , with $k_{ij}/h_{ij} < 1$ and $k_{ij}/h_{ij} > 1$ indicating competitive and mutualistic interactions, respectively.

Finally, the basic reproduction number of each strain i is defined by $R_{0,i} \equiv c\beta_i/q_{\{i\} \rightarrow \emptyset}$. In the absence of interactions, only strains having $R_{0,i} > 1$ would be expected to survive, with marginal prevalence given by $\sum_{X \in \mathcal{S}_i} N_X = 1 - 1/R_{0,i}$. Note, however, that $R_{0,i} > 1$ is neither sufficient nor required for survival in the presence of

strain interactions, and that rank marginal prevalence need not reflect rank
reproduction number. For simplicity, we fixed $c = 3$, set $q_{\{i\} \rightarrow \emptyset} = 1$ for all strains, and
drew β_i randomly to obtain random $R_{0,i}$.

Network construction

In the interaction networks we consider in this paper, nodes represent pathogen strains
and edges the presence of a pairwise interaction between two strains. By connecting
each pair of strains with fixed connection probability σ , independent from other edges,
the strain interaction networks resemble Erdős–Rényi random graphs that become more
saturated with higher connection probability [34].

Binary networks only indicating presence or absence of interactions were converted
to weighted networks by attributing strengths of interaction x_{ij} , drawn uniformly
between $[-\theta, \theta]$, with $\theta > 0$ denoting the maximum strength of interaction. In settings
where we considered interaction in acquisition only, $\exp(x_{ij})$ was used as k_{ij} in the
epidemiological model. When we also considered interaction in clearance, aside from x_{ij} ,
which we drew uniformly from $[-\theta, \theta]$, we also drew an additional random number x'_{ij}
uniformly from $[-\theta/2, \theta/2]$. $\exp(x'_{ij})$ was used as k_{ij} and $\exp(x'_{ij} - x_{ij})$ as h_{ij} .

Sampling cross-sectional datasets

We sampled random cross-sectional datasets $[Y_1, Y_2, \dots, Y_m]^T$ from a multivariate
binomial distribution that coincides with the steady state of the described
epidemiological model, where m is the number of i.i.d. observations, i.e. individuals.
For each individual l , $Y_l = [Y_{1,l}, Y_{2,l}, \dots, Y_{n,l}]$ denotes the sequence of Bernoulli random
variables [35], with $Y_{i,l}$ denoting the presence of strain i in individual l . It follows that
 $P(Y_{i,l} = 1 \ \forall i \in X, Y_{j,l} = 0 \ \forall j \notin X) = N_X$, where N_X is the steady state solution of
model (1).

Statistical network inference

In previous work, we and others have shown that in a two-strain version of model (1),
the prevalence of being simultaneously infected with two strains that do not interact,
corresponds to the product of marginal prevalence of both strains, provided that host

death in age groups susceptible to infection is negligible [21], and there are no unobserved common risk factors [28]. Moreover, the odds ratio of co-occurrence in the two-strain model corresponds to the ratio of interaction parameters in acquisition and clearance k_{12}/h_{12} [28]. When there are more than two strains, however, the correspondence between pairwise association measures and corresponding pairwise composite interaction parameters k_{ij}/h_{ij} might no longer hold, and deviation from independence between two non-interacting strains could be induced by a third strain that interacts with both. The objective here is to capture these (possibly heterogeneous) pairwise interaction parameters from cross-sectional prevalence surveys for more than two strains. For this purpose, we make use of various network inference methods (S1 Appendix).

The first statistical network inference method we applied are graphical modelling approaches based on log-linear analysis [36]. This technique is generally used to examine the relationship between more than two categorical variables, here denoting presence-absence of each pathogen strain in a host population. To reconstruct networks of strain-specific interactions from co-occurrence data with up to 10 circulating strains, we searched through the subset of decomposable log-linear models, i.e. models whose dependence graph is triangulated [25]. Selection was applied both in a forward (focussed on adding edges) and backward (focussed on removing edges) fashion, using AIC or BIC as selection criterion. Strengths of interaction were quantified a posteriori by calculating odds ratios using conditional maximum likelihood estimation from the contingency tables implied by the selected model.

Alternatively, we used the Ising model to reconstruct the simulated networks [24,37]. Here, the probability of a certain pathogen strain being present is modelled as a function of other strains being present at the same time. The Ising model can be shown to be equivalent to certain kinds of log-linear models, but interactions are at most pairwise and the dependence graph does not need to be triangulated. To estimate presence and strength of interactions, we made use of regularized logistic regressions: iteratively, one variable is regressed onto all others, with a penalty imposed on the regression coefficients to obtain a sparse network representation [38], and with model selection based on the extended BIC [24].

Lastly, we used generalized estimating equations (GEE) to estimate pairwise odds

ratios between all strains under consideration. In modelling the associations between multiple strains concomitantly, we used the alternating logistic regression algorithm of the GENMOD procedure in SAS statistical software [26, 32]. The regression framework facilitates assessment of strain-specific interactions on the basis of Wald tests. To correct for multiple hypothesis testing, we made use of the Benjamini–Hochberg procedure [39].

Host heterogeneity

In the sensitivity analysis with host heterogeneity we considered a version of the described multi-strain epidemiological model with two sub-populations that differ in their contact rates relevant for pathogen spread. The proportions of hosts in the high- and low-contact sub-populations are denoted by $p_1 = 20\%$ and $p_2 = 80\%$, and the corresponding contact rates by c_1 and c_2 , respectively. The values of c_1 and c_2 were chosen to obtain a coefficient of variation of 80%. Contacts between the two sub-populations follow a classical pattern based on assortativity fraction ϕ within host types, that takes the value 0 when mixing between hosts is random and 1 when fully assortative [33]. Here, we fixed $\phi = 50\%$. The contact rate c_{sr} between a ‘transmitting’ individual from sub-population s and a ‘receiving’ individual from sub-population r is given by (see S2 Appendix for derivation)

$$c_{sr} = \phi c_s \delta_{sr} + (1 - \phi) \frac{p_r c_r}{p_1 c_1 + p_2 c_2}, \text{ for } s, r \in \{1, 2\}. \quad (4)$$

Here, δ_{sr} is the Kronecker delta (with $\delta_{sr} = 1$ if $s = r$ and zero otherwise). Defined accordingly, the baseline hazard for a susceptible individual in sub-population r to acquire strain i is given by

$$q_{r, \emptyset \rightarrow \{i\}} = c_{1r} \beta_i \sum_{X \in \mathcal{S}_i} \frac{N_{1,X}}{p_1} + c_{2r} \beta_i \sum_{X \in \mathcal{S}_i} \frac{N_{2,X}}{p_2}, \text{ for } r \in \{1, 2\}. \quad (5)$$

Here, $N_{r,X}$ is the proportion of individuals in sub-population $r \in 1, 2$ and infection state X , with $\sum_{X \in \mathcal{S}} N_{r,X} = p_r$. Differential equations for $N_{r,X}$ were defined analogously as in model (1) and the hazard for acquisition in presence of infection with other strains as in equation (3), e.g.

$$q_{r, X \rightarrow X \cup \{i\}} = \prod_{j \in X} k_{ij} q_{r, \emptyset \rightarrow \{i\}}. \quad (6)$$

For network reconstruction in the stratified analysis, we performed analyses separately within either sub-population. For the alternative analysis with correction, we included an additional variable R indicating to which sub-population each individual belongs. In the GEE regression framework, we included R as an explanatory variable (irrespective strain). In the other network approaches, we added R to the sequence of variables denoting strain-specific presence, i.e. $Y'_l = [Y_l, R]$. In effect, the conditional independence structure inferred from $[Y'_1, Y'_2, \dots, Y'_m]^T$ is augmented by one node representing elevated infection risk, next to nodes representing strains.

Supporting information

S1 Fig. Schematic representation of the data generation process. For simplicity, the workflow is illustrated for a system with $n = 3$ strains but the same principle was applied in base-case and sensitivity analyses for systems of $n = 5$ or $n = 10$ strains.

S2 Fig. Scatter plots of the estimated against the true interaction parameters. Four scatter plots of the estimated against the true interaction parameters under GEE or Ising models, with or without correction for contact rate.

S3 Fig. Schematic representation of the epidemiological model with $n = 3$ strains. The eight (2^n) infection states and transitions pertaining to acquisitions. For convenience, the reverse transitions (i.e. clearances) are not shown, and infectious states are not shown in set notation with brackets, unlike the formulas in the main text.

S1 Table. Performance measures of statistical network inference in the sensitivity analysis with host heterogeneity. In the base-case analysis, interactions between strains were generated with connection probability $\sigma = 0.25$, interaction strengths indicating interaction in acquisition drawn from $[-\log 3, \log 3]$, and basic reproduction numbers randomly drawn from $[1.5, 2]$. In this sensitivity analysis, the average contact rate relevant for pathogen spread is the same as in the base-case simulations, but 80% of hosts is assumed to have below-average contacts and 20% above-average contact. Mixing between sub-populations occurred pseudo-assortatively

with an assortment fraction of 50%. Abbreviations: GEE, generalized estimating equations; PPV, positive predictive value; RMSE, root-mean-square error.

S1 Appendix. Statistical network inference.

S2 Appendix. Host heterogeneity.

References

1. Budden KF, Shukla SD, Rehman SF, Bowerman KL, Keely S, Hugenholtz P, et al. Functional effects of the microbiota in chronic respiratory disease. *Lancet Respir Med.* 2019;7(10):907–920.
2. Rowe HM, Meliopoulos VA, Iverson A, Bomme P, Schultz-Cherry S, Rosch JW. Direct interactions with influenza promote bacterial adherence during respiratory infections. *Nat Microbiol.* 2019;4(8):1328–1336.
3. Cobey S, Lipsitch M. Niche and neutral effects of acquired immunity permit coexistence of pneumococcal serotypes. *Science.* 2012;335(6074):1376–1380.
4. Egawa N, Egawa K, Griffin H, Doorbar J. Human papillomaviruses; epithelial tropisms, and the development of neoplasia. *Viruses.* 2015;7(7):3863–3890.
5. Lerch A, Koepfli C, Hofmann NE, Kattenberg JH, Rosanas-Urgell A, Betuela I, et al. Longitudinal tracking and quantification of individual *Plasmodium falciparum* clones in complex infections. *Sci Rep.* 2019;9(1):3333.
6. Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: networks, competition, and stability. *Science.* 2015;350(6261):663–666.
7. Bucci V, Tzen B, Li N, Simmons M, Tanoue T, Bogart E, et al. MDSINE: Microbial Dynamical Systems Inference Engine for microbiome time-series analyses. *Genome Biol.* 2016;17(1):121.
8. Man I, Auranen K, Wallinga J, Bogaards JA. Capturing multiple-type interactions into practical predictors of type replacement following human

- papillomavirus vaccination. *Philos Trans R Soc Lond B Biol Sci.* 2019;374(1773):20180298.
9. Balmer O, Tanner M. Prevalence and implications of multiple-strain infections. *Lancet Infect Dis.* 2011;11(11):868–878.
10. Griffiths EC, Pedersen AB, Fenton A, Petchey OL. The nature and consequences of coinfection in humans. *J Infect.* 2011;63(3):200–206.
11. Alizon S, de Roode JC, Michalakakis Y. Multiple infections and the evolution of virulence. *Ecol Lett.* 2013;16(4):556–567.
12. van Dorp CH, van Boven M, De Boer RJ. Immuno-epidemiological modeling of HIV-1 predicts high heritability of the set-point virus load, while selection for CTL escape dominates virulence evolution. *PLoS Comput Biol.* 2014;10(12):e1003899.
13. Susi H, Barrès B, Vale PF, Laine AL. Co-infection alters population dynamics of infectious disease. *Nat Commun.* 2015;6:5975.
14. Wikramaratna PS, Kucharski A, Gupta S, Andreasen V, McLean AR, Gog JR. Five challenges in modelling interacting strain dynamics. *Epidemics.* 2015;10:31–34.
15. Howard S, Donnelly C, Chan MS. Methods for estimation of associations between multiple species parasite infections. *Parasitology.* 2001;122(2):233–251.
16. Regev-Yochay G, Dagan R, Raz M, Carmeli Y, Shainberg B, Derazne E, et al. Association between carriage of *Streptococcus pneumoniae* and *Staphylococcus aureus* in children. *JAMA.* 2004;292(6):716–720.
17. Chaturvedi AK, Katki HA, Hildesheim A, Rodríguez AC, Quint W, Schiffman M, et al. Human papillomavirus infection with multiple types: pattern of coinfection and risk of cervical disease. *J Infect Dis.* 2011;203(7):910–920.
18. Mollers M, Vriend HJ, van der Sande MA, van Bergen JE, King AJ, Lenselink CH, et al. Population- and type-specific clustering of multiple HPV types across

- diverse risk populations in the Netherlands. *Am J Epidemiol*. 2014;179(10):1236–1246.
19. Wilber MQ, Johnson PT, Briggs CJ. When can we infer mechanism from parasite aggregation? A constraint-based approach to disease ecology. *Ecology*. 2017;98(3):688–702.
 20. Alizon S, Murall CL, Saulnier E, Sofonea MT. Detecting within-host interactions from genotype combination prevalence data. *Epidemics*. 2019;29:100349.
 21. Hamelin FM, Allen LJ, Bokil VA, Gross LJ, Hilker FM, Jeger MJ, et al. Coinfections by noninteracting pathogens are not independent and require new tests of interaction. *PLoS Biol*. 2019;17(12):e3000551.
 22. Hastings A. Can competition be detected using species co-occurrence data? *Ecology*. 1987;68(1):117–123.
 23. Plummer M, Vaccarella S, Franceschi S. Multiple human papillomavirus infections: the exception or the rule? (editorial). *J Infect Dis*. 2011;203(7):891–893.
 24. Van Borkulo CD, Borsboom D, Epskamp S, Blanken TF, Boschloo L, Schoevers RA, et al. A new method for constructing networks from binary data. *Sci Rep*. 2014;4(1):1–10.
 25. Højsgaard S, Edwards D, Lauritzen S. Graphical models with R. Springer Science & Business Media; 2012.
 26. Carey V, Zeger SL, Diggle P. Modelling multivariate binary data with alternating logistic regressions. *Biometrika*. 1993;80(3):517–526.
 27. de Villiers EM. Cross-roads in the classification of papillomaviruses. *Virology*. 2013;445(1-2):2–10.
 28. Man I, Wallinga J, Bogaards JA. Inferring pathogen type interactions using cross-sectional prevalence data: opportunities and pitfalls for predicting type replacement. *Epidemiology*. 2018;29(5):666–674.

29. Telfer S, Birtles R, Bennett M, Lambin X, Paterson S, Begon M. Parasite interactions in natural populations: insights from longitudinal data. *Parasitology*. 2008;135(7):767–781.
30. Shrestha S, King AA, Rohani P. Statistical inference for multi-pathogen systems. *PLoS Comput Biol*. 2011;7(8):e1002135.
31. Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag; 2004.
32. Lipsitz SR, Fitzmaurice GM. Estimating equations for measures of association between repeated binary responses. *Biometrics*. 1996;52(3):903–912.
33. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton University Press; 2008.
34. Newman M. Networks. Oxford University Press; 2018.
35. Teugels JL. Some representations of the multivariate Bernoulli and binomial distributions. *J Multivar Anal*. 1990;32(2):256–268.
36. Agresti A. Categorical data analysis. vol. 482. John Wiley & Sons; 2003.
37. Ising E. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*. 1925;31:253–258.
38. Ravikumar P, Wainwright MJ, Lafferty JD, et al. High-dimensional Ising model selection using l1-regularized logistic regression. *Ann Stat*. 2010;38(3):1287–1319.
39. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57(1):289–300.

Figure 5

Figure 4

Figure 3

Figure 2

A

True network 1

B

Ising model 1

C

Graphical model 1

D

True network 2

E

Ising model 2

F

Graphical model 2

Figure 1