

# EPINEST, an agent-based model to simulate epidemic dynamics in large-scale poultry production and distribution networks

Francesco Pinotti<sup>1\*</sup>, José Lourenço<sup>2</sup>, Sunetra Gupta<sup>1</sup>, Suman Das Gupta<sup>3,4</sup>, Joerg Henning<sup>3</sup>, Damer Blake<sup>5</sup>, Fiona Tomley<sup>5</sup>, Tony Barnett<sup>5,6</sup>, Dirk Pfeiffer<sup>5,7</sup>, Md. Ahasanul Hoque<sup>8</sup>, Guillaume Fournie<sup>5,9,10</sup>

**1** University of Oxford, Oxford, UK

**2** BioISI (Biosystems and Integrative Sciences Institute), Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal

**3** School of Veterinary Science, The University of Queensland, Queensland, Australia

**4** Gulbali Institute, Charles Sturt University, Wagga Wagga, NSW, Australia

**5** Royal Veterinary College, London, UK

**6** The Firoz Lalji Centre for Africa, London School of Economics and Political Science, London, UK

**7** City University of Hong Kong, Hong Kong SAR, Hong Kong

**8** Chattogram Veterinary and Animal Sciences University, Chittagong, Bangladesh

**9** INRAE, VetAgro Sup, UMR EPIA, Université de Lyon, Marcy l'Etoile, 69280, France

**10** INRAE, VetAgro Sup, UMR EPIA, Université Clermont Auvergne, Saint Genes Champanelle, 63122, France

\* francesco.pinotti@biology.ox.ac.uk

## Abstract

The rapid intensification of poultry production raises important concerns about the associated risks of zoonotic infections. Here, we introduce EPINEST (EPI- demic NEtwork Simulation in poultry Transportation systems): an agent-based modelling designed to simulate pathogen transmission within realistic poultry production and distribution networks. The modular structure of the model allows for easy parameterization to suit specific countries and system configurations. Moreover, the framework enables the replication of a wide range of eco-epidemiological scenarios by incorporating diverse pathogen life-history traits, modes of transmission and interactions between multiple strains and/or pathogens. EPINEST was developed in the context of an interdisciplinary multi-centre study conducted in Bangladesh, India, Vietnam and Sri Lanka, and will facilitate the investigation of the spreading patterns of various health hazards

such as avian influenza, *Campylobacter*, *Salmonella* and antimicrobial resistance in these countries. Furthermore, this modelling framework holds potential for broader application in veterinary epidemiology and One Health research, extending its relevance beyond poultry to encompass other livestock species and disease systems.

## Introduction

Animal populations act as reservoirs for a wide range of zoonotic pathogens, such as Ebola virus, MERS-CoV, SARS-CoV-2, avian influenza viruses (AIVs), *Campylobacter* and *Salmonella* [1–6]. Within this context, livestock production is known to promote the risk of zoonotic infections [7]. In the case of emerging pathogens of wildlife, livestock may become intermediate or amplifier hosts, increasing odds of spillover into the human population [8]. The ongoing global intensification of livestock production raises critical questions about the role of husbandry and animal trading practices in shaping the risk of zoonotic epidemics or spillover events. [9, 10]. Unfortunately, however, a comprehensive understanding of how such risk is modulated and amplified along production and distribution networks (PDNs) is lacking.

Poultry production has become the fastest growing livestock sector in the last three decades, with rapid intensification occurring in low- and middle-income countries (LMICs) and particularly in South and Southeast Asia [11]. In many of these countries, intensive production did not replace local farming and trading practices completely, resulting in multiple modes of production and distribution articulated in ways that are poorly understood and which vary according to market and other conditions. While such transformative changes have proven instrumental towards improving food security, nutrition and economic and societal development e.g. in China, India, Bangladesh among others, they also require careful monitoring and investigation. Indeed, the growth of poultry production and distribution networks has brought novel challenges in terms of disease management: intensive farming, limited surveillance infrastructure and veterinary services and in many examples poor biosecurity conditions [12,13] can lead to an environment replete with health hazards. For example, widespread sub-optimal use of antimicrobial drugs by poultry farmers represents a leading driver of the emergence of antimicrobial resistance [14–16].

In many LMICs, people prefer to obtain their poultry from live bird markets (LBMs), which are a longstanding feature of poultry trade and of urban and rural life. Within poultry PDNs, LBMs may be considered as hubs, sites wherein large numbers of people, and critically birds, meet and mix [17,18]. Thus, they are major hotspots of AIV amplification and evolution [19], and have been implicated in sustaining viral transmission in domestic poultry [20]. The diverse ecology of AIV strains circulating within LBMs in Asia has been documented extensively [21–24]. Low pathogenic strains such as H9N2 AIV are commonly found among LBMs in Bangladesh, often at higher rates than in surrounding farms [25–27]. Since its first identification in 1996, highly-pathogenic H5N1

influenza has been detected in LBMs in many Asian countries [28–32].

While the biological risks within poultry production systems are widely acknowledged, they remain poorly characterised. This is partly due to the inherent complexity of PDNs, which makes it difficult to understand how such risks are modulated and increased along poultry value chains. Previous modelling efforts have focused on disease transmission within specific PDN settings, e.g. single farms or LBMs [33, 34], or some PDN segment, such as networks of farms or LBMs [18, 35–37]. Attempts to account for poultry or livestock PDN structure in infectious disease modelling are rare and mostly theoretical, often leaving out many epidemiologically relevant details of poultry production and distribution [38, 39]. Recent PDN mapping efforts have provided a clearer picture of PDNs in several Asian countries [40, 41]. A central observation is that PDNs are highly heterogeneous across countries, poultry types, and even within the same country. Therefore, a better understanding can be achieved by extending and developing modelling to increase our understanding structural heterogeneities within and across PDNs.

To address this gap, we introduce EPINEST, a novel agent-based model (ABM) that allows simulation of pathogen transmission on top of realistic, empirically derived assumptions about poultry movements. EPINEST generates synthetic PDNs consisting of the key nodes, e.g. farms, traders, LBMs, that are responsible for the production and transportation of chickens through the PDN until they are sold to end-point consumers. Extensive data about farming and trading practices, collected mainly from field surveys, is used to inform PDN generation and simulation [18, 27]. Farm-specific data, for example, include farm locations, capacity and statistics of distinct stages of production cycles. Trader-level data encompass details of purchases and sales involving individual actors, origins of purchased poultry, and trader movements. EPINEST allows for substantial flexibility for users in terms of specifying PDN structure and functioning, making it a suitable framework to carry both data-driven and more open-ended analyses. In fact, the ABM permits customisation of many PDN properties, thus allowing users to explore a wide range of hypothetical PDN configurations.

This ABM provides a unified and flexible modelling framework to simulate epidemic dynamics in poultry PDNs and is the outcome of a wider interdisciplinary research initiative [42]. Within this context, EPINEST will enable investigating the amplification and dissemination of a wide range of health hazards, including AIV, *Campylobacter* and anti-microbial resistance genes in poultry systems in Bangladesh, India, Vietnam and Sri Lanka. More broadly, our framework may also be tailored to distinct poultry and livestock production realities to tackle a wider range of epidemiological questions.

In this paper, we provide a detailed description of our ABM and illustrate how to use it to explore a range of PDN structures and to better understand aspects of pathogen transmission in PDNs. The examples presented here are based on a broiler (chickens reared for meat) PDN in Bangladesh, which has been characterised extensively [18, 40], while epidemic simulations focus on the paradigmatic case of AIV transmission. The latter also illustrate an important

feature of our framework, namely the ability to simulate multiple co-circulating pathogens and their interactions. 85  
86

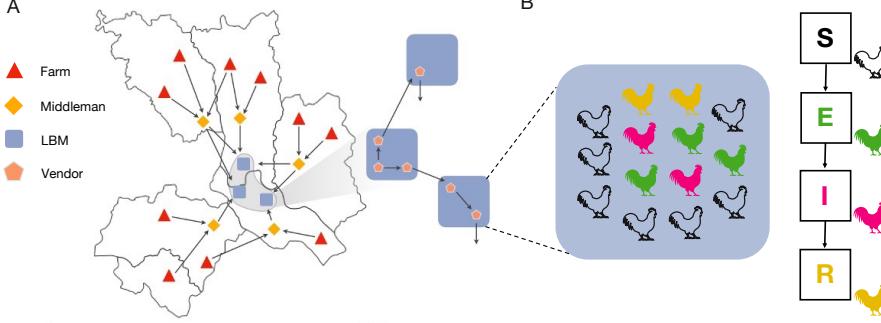
## Results 87

### Synthetic poultry networks 88

To address questions about the eco-epidemiological dynamics of AIVs and other poultry-related pathogens, we implemented an agent-based model to simulate pathogen transmission on top of synthetic PDNs. Within our framework, generated PDNs consists of four main types of nodes: farms, middlemen, vendors and LBMs (Fig. 1A). The system works as a supply chain where chickens are reared in farms starting from day-old chicks and are later transported to LBMs by middlemen (more details can be found in the Materials and Methods section and in Text S1). Once they arrive at the LBM stage, chickens are handled by vendors. These vendors may then sell chickens to other vendors operating in the same or different LBMs, and/or to endpoint consumers, in which case chickens are removed from the PDN. At any stage where chickens are exchanged, other than to the endpoint customer, an opportunity arises for pathogen exchange and mixing. 89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126

To illustrate the ability of the model to synthesize realistic poultry movements, we simulate a small PDN consisting of 1200 farms scattered across the 50 upazilas (sub-districts) that supply the largest amount of broiler chickens to LBMs located in Dhaka (Fig. 2A). The simulated PDN includes 20 distinct LBMs, 163 middlemen and 444 vendors, and allows the trade of chickens between LBMs. Numbers of middlemen and vendors can not be specified a priori; instead, they are determined dynamically by initially calculating the average number of chickens that are sold by farms to each LBM daily. These calculations depend on the spatial arrangement of farms, their sizes and frequency of selling, i.e. parameters that can be specified a priori. The capacity of each trader (middleman or vendor), i.e. the maximum amount of chickens that he/she can purchase daily is also fixed over the course of a simulation. 102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126

Farms sell all their chickens at the end of a production cycle. The trading phase may require multiple days to complete and the flock may be split into multiple transactions involving different middlemen. Fig. 2B,C show that both the distributions of farm trading times and numbers of transactions per production cycle obtained through simulations are consistent with field observations. Upstream transportation and distribution of poultry operated by middlemen represent an important driver of poultry mixing in LBMs [18]. In simulations, middlemen direct previously purchased chickens to LBMs depending on where these have been sourced from. In practice, a chicken bought in upazila  $a$  is sold in market  $l$  with probability  $f_{a,l}$ , as estimated from field questionnaires. Fig. 2D shows that the ABM generates poultry fluxes between individual upazilas and LBMs that are in excellent agreement with the corresponding expected values (i.e.  $f_{a,l}$ ). Moreover, the allocation algorithm ensures that individual 114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126



**Fig 1: Model schematics.** (A) Synthetic PDN and poultry movements. Chickens are produced in farms (red) across the study area, and transported to LBMs (blue) by middlemen (yellow). These are mobile traders that may collect chickens from multiple farms located in one or more upazilas/sub-districts (an administrative area below that of a district in Bangladesh). Within LBMs, chickens are handled by vendors (orange) and may be moved between LBMs as a result of vendors' trading practices. (B) Individual settings associated with farms, middlemen, LBMs (when open) and vendors (overnight, when LBMs are closed) provide the context for pathogen transmission, under the assumption that chickens mix homogeneously within the same setting. The panel zooms in on a single LBM, where chickens are colour-coded according to disease status: susceptible (S), exposed or latent (E), infectious (I) and recovered or immune (R).

middlemen deliver chickens to a desired number of LBMs, as specified by some statistical distribution. The agreement between empirical and simulated frequencies of unique LBMs visited daily is shown in Fig. 2E. At the market level, wholesaling activities and vendor movements between LBMs further contribute to poultry mixing. Once a chicken enters an LBM, it may be sold multiple times to secondary vendors before reaching end-point consumers [18, 43]. In order to better capture the inner organization of LBMs, the model structures vendors in tiers according to their position along transaction chains (Fig. 2F). Finally, we show the realised distribution of poultry marketing times alongside another estimate obtained using a different approach [18] (Fig. 2G). Further statistics about individual actors and poultry transactions can be found in Fig. S1, S2 and S3.

Selected aspects of generated PDNs can be easily manipulated within our framework, allowing flexibility in exploring PDN configurations. In Fig. 3, for example, we examine different distributions of LBMs serviced ( $Pr(k_m)$ ) by individual middlemen on a daily basis (Fig. 3A). As we increase the number of LBMs serviced per middleman,  $\langle k_m \rangle$  on average, middlemen trade with more vendors (Fig. 3B); consequently, individual transactions involve fewer birds since the total cargo is the same (Fig. 3C). Fig. 3B also suggests that the small dis-

127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145

crepancy observed in Fig. 3A at larger  $\langle k_m \rangle$  is due to the limited amount of vendors (inset). 146

We also present the impact of vendors' trading practices on poultry marketing time. In particular, we alter the probability  $p_{empty}$  that a vendor sells its entire cargo in a single day, the fraction  $\rho_{unsold}$  of unsold birds in presence of some surplus (occurring with probability  $1 - p_{empty}$ ). In addition, we consider high and low tendency to prioritise selling older (i.e. previously unsold) chickens over newly purchased ones. Varying parameters  $\rho_{unsold}$  and  $p_{empty}$  affects the average marketing time (Fig. 4A), as well as the proportion of chickens being offered for sale on multiple days (Fig. 4B). Full distributions of marketing times can be found in Fig. S4. Prioritizing the sale of older chickens had a negligible effect on these statistics. Indeed, prioritizing older chickens is compensated by a delay in selling newly purchased chickens (Fig. 4C). 147

As a final example, we examine the role of vendor movements between LBMs in promoting the mixing of chickens from different upazilas/sub-districts. Networks of LBMs defined by trader movements can vary considerably across poultry types, countries, and even cities within the same country [40]. In Chatogram, for example, vendors trading broiler chickens operate almost exclusively in a single market (Fig. 5A). In Dhaka however this is not the case, resulting in frequent vendor movements that are articulated in a top-down structure where central and peripheral markets can be identified (Fig. 5B). In fact, removing a single edge in the network shown in Fig. 5B is sufficient to make it acyclic, suggesting a hierarchical organisation. 148

Within our framework, we encode inter-market mobility in a graph  $G$ , whose entries  $G_{i,j}$  represent the probability that a vendor purchasing in market  $i$  moves to market  $j$  (or remains in  $i$ ) to sell. As outlined above, vendors are further arranged in tiers, so that vendors in tier  $L$  ( $V_L$ ) can only buy poultry from wholesalers located in tier  $L - 1$  or, in the case of  $L = 0$  vendors ( $V_0$ ) from middlemen trading in LBM  $i$ . For each vendor, purchase and sell locations remain fixed throughout a simulation. 149

To explore inter-market mobility, we use a generative network model to create mobility networks  $G$  akin to that of Fig. 5B. In practice, we generate directed acyclic graphs (DAGs) of varying density and amount of hierarchy (see Materials and Methods section), according to parameters  $\rho$  and  $p_{random}$ .  $\rho$  represents the density of connections, while  $p_{random}$  is the probability of an individual connection emanating from a source LBM that is selected randomly, rather than proportionally to their actual number of connections. To quantify the degree of hierarchy in a DAG  $G$ , we measure its global reaching centrality (GRC, see Fig. 5C,E) and tree depth (TD, see Fig. 5D,F). GRC measures how well every node can reach other nodes in the network with respect to the most influential node; it takes value 1 in the case of a star graph and approaches 0 when all nodes have the similar influence (no hierarchy). In contrast, TD represents the longest directed path in  $G$ . Hierarchical DAGs, e.g. stars, tend to be more compact and hence shallower than random structures. Setting  $p_{random} = 1$  yields DAGs with little hierarchy, as edges are allocated randomly. In contrast,  $p_{random} \rightarrow 0$  introduces additional structure. Fig. 5C,D show GRC and TD, re- 150

spectively, for Dhaka's network and for DAGs generated with  $p_{random} = 1$  (red) and  $p_{random} = 0.1$  (cyan) while keeping the density of edges constant. Clearly, Dhaka's network is significantly more hierarchical and compact than random DAGs; in contrast, DAGs generated with  $p_{random} = 0.1$  provide a much closer fit in terms of both GRC and TD.

We also summarize framework output by quantifying poultry mixing across 20 LBMs for different combinations of  $\rho$  and  $p_{random}$  (GRC and TD are shown in Fig. 5E,F, respectively). Here mixing refers to the extent to which chickens from distinct regions are brought together within LBM. Upstream distribution, managed by middleman, and vendor movements between LBM are the factors driving chicken mixing within this model. To quantify the amount of mixing, we record the geographic origins of chickens offered for sale in each LBM and use Pianka's index [44] to make pairwise comparisons of poultry populations marketed in distinct LBM. Mean Pianka's index values are shown in Fig. 5G as a function of parameters  $\rho$  and  $p_{random}$ . Values close to 0 imply low overlap, while a value of 1 corresponds to identical distributions of geographic sources of poultry. Fig. 5H shows another, complementary quantification of poultry mixing in terms of the mean number of LBM where it is possible to find chickens from two randomly chosen upazilas/sub-districts. In general, we find that chicken mixing increases with network density, while hierarchy has the opposite effect: random vendor movements are more effective at mixing chickens within this simplified network model. It should be noted that high levels of mixing can be observed even in the absence of vendor movements due to upstream distribution (overlap between the catchment areas of LBM, of which middlemen are responsible; see caption of Fig. 5 for further details).

## Epidemic dynamics

In this section, we illustrate how our framework can be used to simulate and characterise pathogen transmission across PDNs. We first consider a single, AIV-like pathogen whose dynamics is described by a Susceptible-Exposed-Infectious-Recovered (SEIR) model, as depicted in Fig. 1B: upon infection, susceptible (S) chickens enter an intermediate exposed stage (E) and become infectious (I) after a short latent period  $T_E = 6$  hours. Infectious chickens recover (R) after an infectious period  $T_I = 48$  hours and become immune to further infection. Importantly, we assume that chickens do not die due to the disease. We assume that the pathogen (repeatedly) emerges at rate  $\alpha$  in farms due to external factors (e.g. contacts with wild birds) and spreads through the PDN through a combination of poultry movements, intra- and inter-farm transmission (see Material and Methods section).

Model output comes at different levels of aggregation. Fig. 6A shows for example daily incidence within LBM during the first stages of an outbreak. At the most granular level, individual transmission events and their metadata can be tracked as well. Using this information, we can reconstruct transmission chains originating from individual introduction events and characterise their spatio-temporal evolution (Fig. 6B). Fig. 6C further characterises farm outbreaks by

summarising attack rates by production cycle.

In Fig. 6D-F, we investigate the role of spatial transmission in an endemic context. We do so in a scenario where most transmission events occur within farms (Fig. 6D), while viral amplification in LBM is limited (Fig. 6E). Here, spatial transmission is a crucial factor in determining global levels of infection. Increasing the strength of inter-farm transmission  $\beta_{FF}$  facilitates spatial invasions, thus leading to more outbreaks on farms and infections (Fig. 6D). This results in an increasing number of infected chickens pouring into LBM from farms (Fig. 6F), explaining also the increase in within-market prevalence observed in Fig. 6E.

Another important epidemiological question is whether AIV is transmitted and maintained in LBM despite short marketing times. We address this question by considering an alternative endemic scenario where transmission is contributed mostly by LBM (Fig. 6G). We find that a major limiting factor to viral amplification in LBM is represented by the latent period  $T_E$  (Fig. 6H): delaying the onset of infectiousness corresponds to a shorter window of opportunity for transmission under short marketing times. In order to further demonstrate this point, we quantify persistence of transmission chains within LBM (Fig. 6I). As  $T_E$  increases, opportunities for transmission are diminished and chains of infection stutter, leading to reduced persistence. In this case, the presence of AIV in LBM can only be maintained through repeated introductions of infected poultry.

## Simulating multi-strain pathogens

Genomic surveillance in LBM routinely identifies AIV lineages with distinct genetic signatures [45]. In some instances, the presence of multiple AIV subtypes, including the highly pathogenic H5N1 AIV, is also reported. Understanding this diversity requires, however, accounting for multiple, potentially interacting strains/pathogens that co-circulate in the same PDN. In this section, we use our framework to perform multi-strain simulations in a variety of PDN structures.

We illustrate this in Fig. 7, which shows SEIR simulations with 50 co-circulating strains. For simplicity, we assume that these share the same epidemiological parameters, namely  $T_E$ ,  $T_I$  and  $\beta$ , and generate partial cross-immunity after a single infection.

Our aim is to measure the extent to which PDNs mix viral lineages from distinct geographical regions. To this end, we modify the external seeding protocol so that strain  $s_i$ ,  $i = 1, \dots, 50$  can emerge only from upazila  $i$ . First, we investigate the role of viral amplification during the transport segment, which is operated by middlemen (Fig. 7A-C). To better disentangle the role of these actors, we consider low within-farm transmission and prevent inter-farm transmission fully by setting  $\beta_{FF} = 0$ . Consequently, viral mixing can not occur until chickens from different upazilas/sub-districts are collected by a middleman. As shown in Fig. 7A, increasing transmission during transport by varying  $w_{FF}$  leads to more infected chickens being introduced in LBM, i.e. it results in viral amplification. Note, however, that below a certain value of  $w_{MM}$ , mid-

dlemen may introduce fewer infections in LBMs than those they picked up at farms. Increasing  $w_{MM}$  has a modest positive effect on the average number of strains circulating (i.e. strain richness) in individual LBM, and on the overlap between LBM in terms of circulating strains (light to dark bars in Fig. 7B and C, respectively). We also explored, for fixed  $w_{MM}$ , the role of inter-market mobility on these metrics. In this context, the density  $\rho$  of vendor movements had a positive effect on both strain richness and overlap between LBM as it is promoting the dissemination of multiple strains across LBM. In contrast, a larger degree of hierarchy in movements (striped bars) had the opposite effect, in agreement with findings from Fig. 5.

Finally, we consider a further scenario in which transmission within and between farms plays a central role in shaping epidemic dynamics, while transmission occurring during transport is assumed to be negligible ( $w_{MM} = 0.001$ ). We find that increasing between-farm transmission  $\beta_{FF}$  leads to a wider spatial dissemination of strains even outside their upazila of origin (Fig. 7D). Consequently, a more diverse set of strains is supplied to LBM, as evidenced by the number of strains observed at these locations (Fig. 7E). Also, because larger values of  $\beta_{FF}$  promote strain dispersal across the entire area, LBM are now more similar to each other in terms of their strain populations (Fig. 7F). It should be noted, however, that increased within-farm transmission is responsible, at least in part, for the larger strain numbers and overlap between LBM observed in Fig. 7E,F with respect to panels B,C. Finally, we note that the effects of density and hierarchy of vendor movements on ecological metrics are analogous to those observed in the previous scenario. These results are robust to increasing cross-immunity between strains (Fig. S6).

## Discussion

In this paper we have introduced EPINEST, an agent-based model to simulate the transmission of generic health hazards in the context of realistic poultry or livestock movements within a defined PDN. To the best of our knowledge, this work represents the first attempt to account for the structural complexities of poultry PDNs in the context of epidemic transmission modelling. Our model allows to generate synthetic PDNs consisting of key actors and settings involved in poultry production and distribution, namely farms, middlemen, LBM and market vendors. Using Bangladesh as a case study, we illustrated the ability of our framework to reproduce empirical features of a broiler PDN. We used extensive data from field surveys to inform most aspects of the model, including farming and trading practices of key actors [18, 27, 46]. At the same time, our model offers the possibility to easily manipulate most properties of the network, allowing exploration of alternative PDN configurations. Importantly, we emphasize that our model may be applied to other contexts, e.g. different poultry types and countries for which sufficient data is available.

One of the main purposes of EPINEST was to assess the impact of PDN structure and stakeholders' trading practices on pathogen transmission. For this

reason, we prioritised including PDN components with the highest relevance to transmission dynamics. These include, for example, the time spent by chickens at different locations. All-in/all-out production, which is commonly implemented in commercial broiler farms, results in relatively homogeneous rearing times across farmed chickens, although these vary considerably among different chicken types. In contrast, LBMs are characterised by a much faster turnover, with most chickens being sold within a few hours and unsold chickens remaining for up to a few days. Longer marketing times are a well-established risk factor for AIV infection in LBM, and have been linked to AIV persistence in these settings [18,47]. To account for heterogeneity in marketing times, we explicitly account for a fraction of chickens being offered for sale on consecutive days.

Further basic ingredients of the model are the spatial distribution of poultry farms and their sizes. Both elements are highly relevant to disease transmission. Heterogeneities in farm locations can affect systemic vulnerability to epidemics and pathogen dispersal patterns [48–50], while higher livestock densities are associated with increased intra-farm transmission and may favor the emergence of virulent pathogens [9,51]. In the absence of accurate data about farm locations, we generated random farm distributions complying with reported volumes of poultry production at the upazila level, and used field surveys to assign farm sizes [18,27]. Nonetheless, we stress that our model can accommodate any distribution of farms. These may represent not only higher-resolution data, but also outcomes from more accurate generative models [52–54].

Our model also allows to control the degree of mixing of chickens along distribution and trading channels. The ability of PDNs to mix large numbers of chickens, particularly within LBM, is well-established. The inter-mingling of different types of birds from potentially distant locations is concerning when associated to co-circulation of genetically distinct viruses. A recent phylogenetics study found substantial genetic structuring of H9N2 AIV by city in Bangladesh [55], compatibly with low overlap between the corresponding supplying production areas [18]. In contrast, viral lineages appeared to be highly mixed across LBM within the same city, possibly indicating frequent connections between these markets. Live poultry trade has also been shown to be an important driver of regional AIV dissemination in China [56]. Here, poultry mixing within and between LBM is dictated by two factors: first, upstream distribution via middlemen connects LBM with farmed populations from a wide geographic area. Within our framework, geographic fluxes between regions (upazilas/sub-districts) and LBM are expressed as a matrix that can be informed using field surveys or traceability systems. Second, wholesaling activities and vendor movements further contribute to stirring marketed poultry across LBM. In this manuscript, we used a generative model to sample inter-market mobility networks, and quantified their impact on poultry mixing. We emphasize that more complex mobility patterns, informed either from data or through simulations, can be easily embedded within our framework.

A major feature of our model is that it allows simulating pathogen transmission while accounting for the complexity of poultry movements and PDN structures. Importantly, the epidemic layer is fully uncoupled from PDN generation.

Thus, while current code supports simulations of SIR and SEIR dynamics only, implementing additional epidemic models is a relatively straightforward task. We illustrated how our ABM can be used to model AIV dynamics in both epidemic and endemic settings. In the former case, the ABM makes it possible to map the early dissemination of, e.g., an emerging AIV strain across farms and the rest of the PDN. The second scenario would be more suitable to describe endemic circulation of AIVs. In this context, relevant scientific questions that could be addressed using our framework include understanding how and where an endemic AIV is maintained and amplified along the PDN.

A novel aspect of our ABM is that it enables simulations of multiple co-circulating pathogens/strains and their interactions. This paves the way for a number of eco-epidemiological applications. As an example, we assessed the potential of PDNs to mix viral lineages originating from distinct geographical areas. Additional applications may consider the joint dynamics of endemic and emerging AIVs and simulate the early transmission dynamics of, say, highly-pathogenic H5N1 AIV against a background of (cross-)immunity generated by endemic circulation of H9N2 AIV [57].

As any modelling framework, there are limitations to our ABM. Despite our efforts to account for the structural complexity of PDNs, our focus on epidemiological investigations meant that several aspects of real PDNs could not be included in the model. For example, actors' behaviours are treated as fixed parameters external to, rather than emerging from, the dynamic system being modelled. In reality, the decisions made by individual traders to sell or purchase birds is influenced by social, economic and epidemiological factors. These may include uncertainty about market conditions and fear spurred by disease outbreaks [58, 59]. In addition, unequal power dynamics often constrains trading ties [40, 60]. In this context, we plan on expanding our ABM's capabilities to include simple reactive behaviours, e.g. farmers selling chickens pre-emptively following a surge in bird mortality [59]. Other extensions could include mixing of different poultry species, different farming systems and trading practices, such as second-line middlemen purchasing chickens from other traders, and different biosecurity measures implemented at different LBMs to limit pathogen spread. Finally, although we wrote our model in C++ to improve simulation speed, computational constraints make it difficult to scale up simulations to more than a few millions of farmed chickens. This is a common challenge in agent-based models, where the increased amount of detail is traded off by computational costs.

In conclusion, we implemented a novel agent-based model to jointly simulate realistic poultry movements and epidemic trajectories. Realised structures encompass a wide-range of PDN configurations as encountered in many countries in South and Southeast Asia, and potentially even other livestock production systems with similar structure to the one discussed here. Compared to existing ABMs devoted to veterinary epidemiology applications [61–64], ours offers the ability to run both single- and multi-strain simulations. In addition, the simulator can be programmed to yield a wide range of outputs, including individual transactions and chains of infections, hence providing a full characterisation of

the underlying system. This model is a unique tool in the One Health context as it allows investigation of a range of epidemiological scenarios and helps us to understand better the role of different structural aspects on disease transmission. Immediate applications of this model will allow exploration of the transmission and amplification of AIVs and anti-microbial resistance genes within poultry PDNs. 415  
416  
417  
418  
419  
420

## Materials and Methods 421

### Generating synthetic PDNs 422

In general, a PDN denotes the ensemble of actors that are involved in the production and/or distribution of a product such as poultry and their interactions. At any point in time, a chicken is physically located within one and only one setting, such as a farm, a middleman's truck, an LBM or a vendor-owned shed during the night. 423  
424  
425  
426  
427

Our generative algorithm instantiates a population of actors based on external specifications. First, a spatial distribution of farms must be provided alongside the corresponding geographic setup. The latter consists of a partitioning of the study area into a set of non-overlapping regions. In this study, we take upazilas/sub-districts as regional units. Second, the user specifies a number of LBMs and their catchment areas. In practice, this is achieved by specifying a matrix  $f_{a,l}$  representing the relative fluxes of chickens reaching market  $l = 1, \dots, N_M$  from area  $a = 1, \dots, N_A$ . A full description of LBMs requires a set of weights  $w_{l,l'}$  encoding the probability that a vendor purchases chickens in LBM  $l$  and trades in LBM  $l'$  (with possibly  $l = l'$ ). Finally, a number of parameters influencing farming, distribution and trading practices should be specified as well (these are described in Text S1). With these details, the algorithm computes the expected poultry fluxes between farms and LBMs and allocates enough vendors and middlemen to satisfy such demand. At the LBM stage, vendors are allocated in a tier-wise fashion depending on the volume of chickens supplied by middlemen, inter-market movements, and wholesaling practices. Eventually, it is possible to generate more middlemen and vendors than strictly required based on heuristic calculations by inflating the expected supply of chickens handled by middlemen and vendors through multiplicative factors  $\epsilon_{MM}$  and  $\epsilon_V$ . 428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447

### Modelling inter-market movements 448

As detailed in Text S1, a vendor purchasing chickens in LBM  $i$  is assigned to trade in LBM  $j$  with probability  $G_{i,j}$ . 449  
450

We sample the weights  $G_{i,j}$  from a generative network model defined by a growth mechanism: we add LBMs  $j = 1, \dots, N_M$  one at a time and establish links  $i \rightarrow j$ ,  $i < j$  as follows: first, we draw the number of incoming edges (in-degree)  $z_j \sim \text{Binomial}(\rho, j - 1)$ . Second, we sample  $z_j$  LBMs (with  $i < j$ ) 451  
452  
453  
454

without replacement either at random with probability  $p_{random}$ , or proportionally to their out-degree. This yields a binary matrix  $\mathcal{I}_{i \rightarrow j}$  denoting existing connections. Weights are then calculated as:

$$G_{i,j} = \begin{cases} \delta_{i,j}, & \text{if } k_i = 0. \\ G_{self}\delta_{i,j} + (1 - G_{self})\mathcal{I}_{i \rightarrow j}/k_i, & \text{otherwise,} \end{cases} \quad (1)$$

where  $G_{self}$  is the probability of a vendor operating in a single market  $i$ , conditional on the out-degree  $k_i = \sum_j \mathcal{I}_{i \rightarrow j}$  being positive. Here we set  $G_{self} = 0.8$ .

## Global reaching centrality

GRC is defined based on the notion of local reaching centrality  $C_R(i)$ , which quantifies the proportion of nodes reachable from node  $i$  through directed edges. Based on this definition, we calculate GRC by subtracting  $C_R(i)$  from the maximum observed value  $C_R^{max} = \max_i C_R(i)$ , and averaging over all nodes:

$$GRC = \frac{\sum_i^N C_R^{max} - C_R(i)}{N - 1}. \quad (2)$$

## PDN dynamics

Within simulations, actors follow a daily routine. Let  $t = 0, \dots, 23$  indicate the time of the day (each time step is 1 hour long). Unless otherwise stated, default parameters indicated in Text S1 are considered. LBMs open between  $T_{open}$  and  $T_{close}$ ; at  $t = T_{open}$ , vendors move to LBMs, followed by middlemen. Middlemen then proceed to sell their cargo to frontline vendors, i.e. those in the first LBM tier ( $L = 0$ ). In the next time step ( $t = T_{open} + 1$ ), some of these move to another LBM and trade with second-tier vendors, who in turn sell chickens to vendors in the tier after that, repeating the process until the last tier is reached. Vendor movements and wholesaling are therefore resolved sequentially, in a tier-wise fashion, at time  $t = T_{open} + 1$ . In contrast, retailing activities roll out between  $T_{open} + 1$  and  $T_{close}$ . At  $T_{close}$ , both wholesalers and retailers leave LBMs alongside any unsold chickens. Overnight, these chickens are stored in some other place, e.g. in a shed. Importantly, all chickens from the same vendor are stored in the same place.

At some time  $t = T_{farm}$  we update farms: empty farms may recruit a new batch of chickens, while active farms may offer birds for sale depending on batch age. After that, always at  $t = T_{farm}$ , middlemen are updated: first, they decide whether to cover a different set of upazilas/sub-districts. Then, they contact farms within covered upazilas/sub-districts in order to purchase chickens. At this stage, middlemen only determine how many chickens to collect from each farm; the collection may happen anytime between  $T_{farm}$  and  $T_{open}$  on the following day.

## Epidemic dynamics

In this work, we consider a general transmission model involving a generic number of strains. Each strain, indexed by  $s$ , spreads according to SEIR dynamics. Infected chickens become infectious only after a random latent period  $\hat{\tau}_E$ , sampled from a distribution  $P(\hat{\tau}_E)$  with mean  $T_E$ . Analogously, infectious chickens recover after a random time  $\hat{\tau}_I$ , sampled from a distribution  $P(\hat{\tau}_I)$  with mean  $T_I$ . All epidemiological parameters are listed in Text S1.

Here, transmission is assumed to occur through infectious contacts among chickens from the same setting. Other transmission mechanisms, including external introductions and inter-farm transmission, are described in Text S1. During a time step, an infectious chicken  $i$  contacts a single chicken  $j$ , chosen at random within the same setting, and transmits strain  $s$  with probability:

$$p_{\text{infect}} = 1 - \exp(-\beta(s, i) \cdot S(s, j) \cdot w_X), \quad (3)$$

where  $\beta(s, i)$  transmissibility of strain  $s$  and  $S(s, j)$  is susceptibility of chicken  $j$  to  $s$ . The factor  $w_X$  is a multiplier that depends only on the underlying setting type (F,MM,M,V).

In general, the transmission rate  $\beta(s, i)$  and susceptibility  $S(s, j)$  may depend on the immune state of infector and infectee, respectively. Importantly, different functional forms of  $\beta(s, i)$  and  $S(s, j)$  embody different assumptions about immune cross-reactions induced by previous exposure to other pathogens/strains. In this work we consider uniform transmission  $\beta(s, i) = \beta$ , irrespective of immune state, and susceptibility  $S(s, j) = 0, 1, \sigma$  depending on whether  $j$  has already been infected with  $s$ , is fully naive or was infected with some other strain  $s' \neq s$ , respectively. The parameter  $\sigma \in [0, 1]$  represents reduced susceptibility due to cross-immunity, and interpolates between sterilising cross-immunity ( $\sigma = 0$ ) and no cross-immunity ( $\sigma = 1$ ).

## Supporting information

### Data availability statement

Data and code to generate and analyse simulation output are available at <https://github.com/francescopinotti92/EPINEST>.

### Financial Disclosure Statement

This study was funded by the UKRI GCRF One Health Poultry Hub (Grant No. B/S011269/1), one of twelve interdisciplinary research hubs funded under the UK government's Global Challenge Research Fund Interdisciplinary Research Hub initiative. G.F. is supported by the French National Research Agency and the French Ministry of Higher Education and Research

## Competing interests

523

The authors have no competing interests to declare.

524

## References

1. Barua H, Biswas PK, Olsen KEP, Christensen JP. Prevalence and Characterization of Motile *Salmonella* in Commercial Layer Poultry Farms in Bangladesh. *PLOS ONE*. 2012;7(4):e35914. doi:10.1371/journal.pone.0035914.
2. Kaplan BS, Webby RJ. The avian and mammalian host range of highly pathogenic avian H5N1 influenza. *Virus research*. 2013;178(1):3–11. doi:10.1016/j.virusres.2013.09.004.
3. Mohd HA, Al-Tawfiq JA, Memish ZA. Middle East Respiratory Syndrome Coronavirus (MERS-CoV) origin and animal reservoir. *Virology Journal*. 2016;13(1):87. doi:10.1186/s12985-016-0544-0.
4. Koch LK, Cunze S, Kochmann J, Klimpel S. Bats as putative Zaire ebolavirus reservoir hosts and their habitat suitability in Africa. *Scientific Reports*. 2020;10(1):14268. doi:10.1038/s41598-020-71226-0.
5. Sharun K, Dhama K, Pawde AM, Gortázar C, Tiwari R, Bonilla-Aldana DK, et al. SARS-CoV-2 in animals: potential for unknown reservoir hosts and public health implications. *The Veterinary Quarterly*. 2021;41(1):181–201. doi:10.1080/01652176.2021.1921311.
6. Hoque N, Islam SS, Uddin MN, Arif M, Haque AKMZ, Neogi SB, et al. Prevalence, Risk Factors, and Molecular Detection of *Campylobacter* in Farmed Cattle of Selected Districts in Bangladesh. *Pathogens*. 2021;10(3):313. doi:10.3390/pathogens10030313.
7. Hayek MN. The infectious disease trap of animal agriculture. *Science Advances*. 2022;8(44):eadd6681. doi:10.1126/sciadv.add6681.
8. Jones BA, Grace D, Kock R, Alonso S, Rushton J, Said MY, et al. Zoonosis emergence linked to agricultural intensification and environmental change. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(21):8399–8404. doi:10.1073/pnas.1208059110.
9. Gilbert M, Xiao X, Robinson TP. Intensifying poultry production systems and the emergence of avian influenza in China: a 'One Health/Ecohealth' epitome. *Archives of Public Health = Archives Belges De Sante Publique*. 2017;75:48. doi:10.1186/s13690-017-0218-4.

10. Gilbert W, Thomas LF, Coyne L, Rushton J. Review: Mitigating the risks posed by intensification in livestock production: the examples of antimicrobial resistance and zoonoses. *animal*. 2021;15(2):100123. doi:10.1016/j.animal.2020.100123.
11. Structured Poultry Industry Growing in Size; 2020. Available from: <https://www.lightcastlebd.com/insights/2020/01/structured-poultry-industry-growing-in-size/>.
12. Wong JT, de Bruyn J, Bagnol B, Grieve H, Li M, Pym R, et al. Small-scale poultry and food security in resource-poor settings: A review. *Global Food Security*. 2017;15:43–52. doi:10.1016/j.gfs.2017.04.003.
13. Høg E, Fournié G, Hoque MA, Mahmud R, Pfeiffer DU, Barnett T. Competing biosecurity and risk rationalities in the Chittagong poultry commodity chain, Bangladesh. *BioSocieties*. 2019;14(3):368–392. doi:10.1057/s41292-018-0131-2.
14. Imam T, Gibson JS, Foysal M, Das SB, Gupta SD, Fournié G, et al. A Cross-Sectional Study of Antimicrobial Usage on Commercial Broiler and Layer Chicken Farms in Bangladesh. *Frontiers in Veterinary Science*. 2020;7:576113. doi:10.3389/fvets.2020.576113.
15. Chowdhury S, Ghosh S, Aleem MA, Parveen S, Islam MA, Rashid MM, et al. Antibiotic Usage and Resistance in Food Animal Production: What Have We Learned from Bangladesh? *Antibiotics*. 2021;10(9):1032. doi:10.3390/antibiotics10091032.
16. Tasmim ST, Hasan MM, Talukder S, Mandal AK, Parvin MS, Ali MY, et al. Socio-demographic determinants of use and misuse of antibiotics in commercial poultry farms in Bangladesh. *IJID Regions*. 2023;doi:10.1016/j.ijregi.2023.01.001.
17. Fournié G, Guitian J, Desvaux S, Mangtani P, Ly S, Cong VC, et al. Identifying Live Bird Markets with the Potential to Act as Reservoirs of Avian Influenza A (H5N1) Virus: A Survey in Northern Viet Nam and Cambodia. *PLoS ONE*. 2012;7(6):e37986. doi:10.1371/journal.pone.0037986.
18. Moyen N, Hoque MA, Mahmud R, Hasan M, Sarkar S, Biswas PK, et al. Avian influenza transmission risk along live poultry trading networks in Bangladesh. *Scientific Reports*. 2021;11(1):19962. doi:10.1038/s41598-021-98989-4.
19. Senne DA, Pearson JE, Panigrahy B. Live Poultry Markets: A Missing Link in the Epidemiology of Avian Influenza. *Avian Diseases*. 2003;47:50–58.
20. Kung NY, Morris RS, Perkins NR, Sims LD, Ellis TM, Bissett L, et al. Risk for infection with highly pathogenic influenza A virus

(H5N1) in chickens, Hong Kong, 2002. *Emerging Infectious Diseases*. 2007;13(3):412–418. doi:10.3201/eid1303.060365.

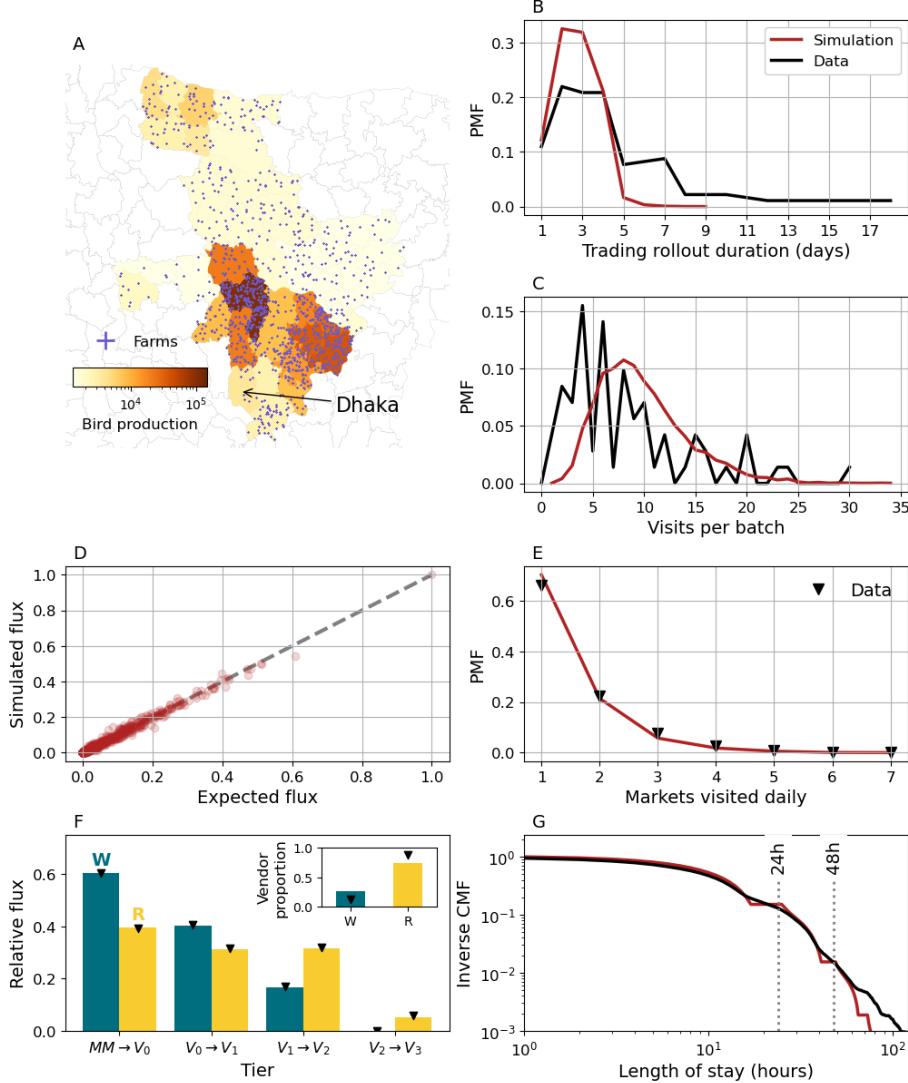
21. Shortridge KF, Butterfield WK, Webster RG, Campbell CH. Isolation and characterization of influenza A viruses from avian species in Hong Kong. *Bulletin of the World Health Organization*. 1977;55(1):15–20.
22. Gerloff NA, Khan SU, Zanders N, Balish A, Haider N, Islam A, et al. Genetically Diverse Low Pathogenicity Avian Influenza A Virus Subtypes Co-Circulate among Poultry in Bangladesh. *PLOS ONE*. 2016;11(3):e0152131. doi:10.1371/journal.pone.0152131.
23. Chen LJ, Lin XD, Guo WP, Tian JH, Wang W, Ying XH, et al. Diversity and evolution of avian influenza viruses in live poultry markets, free-range poultry and wild wetland birds in China. *Journal of General Virology*. 2016;97(4):844–854. doi:10.1099/jgv.0.000399.
24. Khan SU, Gurley ES, Gerloff N, Rahman MZ, Simpson N, Rahman M, et al. Avian influenza surveillance in domestic waterfowl and environment of live bird markets in Bangladesh, 2007–2012. *Scientific Reports*. 2018;8(1):9396. doi:10.1038/s41598-018-27515-w.
25. Turner JCM, Feeroz MM, Hasan MK, Akhtar S, Walker D, Seiler P, et al. Insight into live bird markets of Bangladesh: an overview of the dynamics of transmission of H5N1 and H9N2 avian influenza viruses. *Emerging Microbes & Infections*. 2017;6(3):e12. doi:10.1038/emi.2016.142.
26. Kim Y, Biswas PK, Giasuddin M, Hasan M, Mahmud R, Chang YM, et al. Prevalence of Avian Influenza A(H5) and A(H9) Viruses in Live Bird Markets, Bangladesh. *Emerging Infectious Diseases*. 2018;24(12):2309–2316. doi:10.3201/eid2412.180879.
27. Gupta SD, Hoque MA, Fournié G, Henning J. Patterns of Avian Influenza A (H5) and A (H9) virus infection in backyard, commercial broiler and layer chicken farms in Bangladesh. *Transboundary and Emerging Diseases*. 2021;68(1):137–151. doi:10.1111/tbed.13657.
28. Nguyen DC, Uyeki TM, Jadhao S, Maines T, Shaw M, Matsuoka Y, et al. Isolation and characterization of avian influenza viruses, including highly pathogenic H5N1, from poultry in live bird markets in Hanoi, Vietnam, in 2001. *Journal of Virology*. 2005;79(7):4201–4212. doi:10.1128/JVI.79.7.4201-4212.2005.
29. Hulse-Post DJ, Sturm-Ramirez KM, Humberd J, Seiler P, Govorkova EA, Krauss S, et al. Role of domestic ducks in the propagation and biological evolution of highly pathogenic H5N1 influenza viruses in Asia. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(30):10682–10687. doi:10.1073/pnas.0504662102.

30. Choi YK, Heui Seo S, Kim JA, Webby RJ, Webster RG. Avian influenza viruses in Korean live poultry markets and their pathogenic potential. *Virology*. 2005;332(2):529–537. doi:10.1016/j.virol.2004.12.002.
31. Amonsin A, Choatrakol C, Lapkuntod J, Tantilertcharoen R, Thana-wongnuwech R, Suradhat S, et al. Influenza Virus (H5N1) in Live Bird Markets and Food Markets, Thailand. *Emerging Infectious Diseases*. 2008;14(11):1739–1742. doi:10.3201/eid1411.080683.
32. Pepin KM, Wang J, Webb CT, Smith GJD, Poss M, Hudson PJ, et al. Multiannual patterns of influenza A transmission in Chinese live bird market systems. *Influenza and Other Respiratory Viruses*. 2013;7(1):97–107. doi:10.1111/j.1750-2659.2012.00354.x.
33. Pinsent A, Pepin KM, Zhu H, Guan Y, White MT, Riley S. The persistence of multiple strains of avian influenza in live bird markets. *Proceedings of the Royal Society B: Biological Sciences*. 2017;284(1868):20170715. doi:10.1098/rspb.2017.0715.
34. Guinat C, Tago D, Corre T, Selinger C, Djidjou-Demasse R, Paul M, et al. Optimizing the early detection of low pathogenic avian influenza H7N9 virus in live bird markets. *Journal of The Royal Society Interface*. 2021;18(178):rsif.2021.0074, 20210074. doi:10.1098/rsif.2021.0074.
35. Fournié G, Guitian J, Desvaux S, Cuong VC, Dung DH, Pfeiffer DU, et al. Interventions for avian influenza A (H5N1) risk management in live bird market networks. *Proceedings of the National Academy of Sciences*. 2013;110(22):9177–9182.
36. Hill EM, House T, Dhingra MS, Kalpravidh W, Morzaria S, Osmani MG, et al. Modelling H5N1 in Bangladesh across spatial scales: Model complexity and zoonotic transmission risk. *Epidemics*. 2017;20:37–55. doi:10.1016/j.epidem.2017.02.007.
37. Hill EM, House T, Dhingra MS, Kalpravidh W, Morzaria S, Osmani MG, et al. The impact of surveillance and control on highly pathogenic avian influenza outbreaks in poultry in Dhaka division, Bangladesh. *PLOS Computational Biology*. 2018;14(9):e1006439. doi:10.1371/journal.pcbi.1006439.
38. Fournié G, Guitian FJ, Mangtani P, Ghani AC. Impact of the implementation of rest days in live bird markets on the dynamics of H5N1 highly pathogenic avian influenza. *Journal of The Royal Society Interface*. 2011;8(61):1079–1089. doi:10.1098/rsif.2010.0510.
39. Poolkhet C, Makita K, Thongratsakul S, Leelehapongsathon K. Exponential random graph models to evaluate the movement of backyard chickens after the avian influenza crisis in 2004–2005, Thailand. *Preventive Veterinary Medicine*. 2018;158:71–77. doi:10.1016/j.prevetmed.2018.07.015.

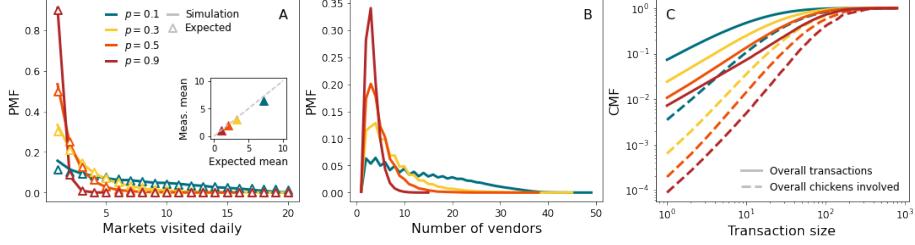
40. Hennessey M, Fournié G, Hoque MA, Biswas PK, Alarcon P, Ebata A, et al. Intensification of fragility: Poultry production and distribution in Bangladesh and its implications for disease risk. *Preventive Veterinary Medicine*. 2021;191:105367. doi:10.1016/j.prevetmed.2021.105367.
41. Thi Dien N, Thi Minh Khue N, Etaba A, Fournié G, Huyen LTT, Van Dai N, et al. Mapping chicken production and distribution networks in Vietnam: An analysis of socio-economic factors and their epidemiological significances. *Preventive Veterinary Medicine*. 2023;214:105906. doi:10.1016/j.prevetmed.2023.105906.
42. One Health Poultry Hub; Available from: <https://www.onehealthpoultry.org/>.
43. Høg E, Fournié G, Hoque MA, Mahmud R, Pfeiffer DU, Barnett T. Avian Influenza Risk Environment: Live Bird Commodity Chains in Chatogram, Bangladesh. *Frontiers in Veterinary Science*. 2021;8.
44. Pianka ER. Niche Overlap and Diffuse Competition. *Proceedings of the National Academy of Sciences*. 1974;71(5):2141–2145. doi:10.1073/pnas.71.5.2141.
45. Liu T, Peng Y, Wu J, Lu S, He Y, Li X, et al. Surveillance of avian influenza viruses in live bird markets of Shandong province from 2013 to 2019. *Frontiers in Microbiology*. 2022;13.
46. Moyen N, Ahmed G, Gupta S, Tenzin T, Khan R, Khan T, et al. A large-scale study of a poultry trading network in Bangladesh: implications for control and surveillance of avian influenza viruses. *BMC Veterinary Research*. 2018;14(1):12. doi:10.1186/s12917-018-1331-5.
47. Pepin KM, Lloyd-Smith JO, Webb CT, Holcomb K, Zhu H, Guan Y, et al. Minimizing the threat of pandemic emergence from avian influenza in poultry systems. *BMC Infectious Diseases*. 2013;13(1):592. doi:10.1186/1471-2334-13-592.
48. Brown DH, Bolker BM. The effects of disease dispersal and host clustering on the epidemic threshold in plants. *Bulletin of Mathematical Biology*. 2004;66(2):341–371. doi:10.1016/j.bulm.2003.08.006.
49. Tildesley MJ, House TA, Bruhn MC, Curry RJ, O’Neil M, Allpress JLE, et al. Impact of spatial clustering on disease transmission and optimal control. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107(3):1041–1046. doi:10.1073/pnas.0909047107.
50. Benincà E, Hagenaars T, Boender GJ, van de Kassteele J, van Boven M. Trade-off between local transmission and long-range dispersal drives infectious disease outbreak size in spatially structured populations. *PLoS computational biology*. 2020;16(7):e1008009. doi:10.1371/journal.pcbi.1008009.

51. Alizon S, Hurford A, Mideo N, Van Baalen M. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *Journal of Evolutionary Biology*. 2009;22(2):245–259. doi:10.1111/j.1420-9101.2008.01658.x.
52. Simpson D, Illian JB, Lindgren F, Sørbye SH, Rue H. Going off grid: computationally efficient inference for log-Gaussian Cox processes. *Biometrika*. 2016;103(1):49–70. doi:10.1093/biomet/asv064.
53. Gilbert M, Nicolas G, Cinardi G, Van Boeckel TP, Vanwambeke SO, Wint GRW, et al. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Scientific Data*. 2018;5(1):180227. doi:10.1038/sdata.2018.227.
54. Chaiban C, Da Re D, Robinson T, Gilbert M, Vanwambeke S. Poultry farm distribution models developed along a gradient of intensification. *Preventive Veterinary Medicine*. 2020;186:105206. doi:10.1016/j.prevetmed.2020.105206.
55. Carnegie L, Hasan M, Mahmud R, Hoque MA, Debnath N, Uddin MH, et al. H9N2 avian influenza virus dispersal along Bangladeshi poultry trading networks. *Virus Evolution*. 2023;9(1):vead014. doi:10.1093/ve/vead014.
56. Yang Q, Zhao X, Lemey P, Suchard MA, Bi Y, Shi W, et al. Assessing the role of live poultry trade in community-structured transmission of avian influenza in China. *Proceedings of the National Academy of Sciences*. 2020;117(11):5949–5954. doi:10.1073/pnas.1906954117.
57. Seo SH, Webster RG. Cross-Reactive, Cell-Mediated Immunity and Protection of Chickens from Lethal H5N1 Influenza Virus Infection in Hong Kong Poultry Markets. *Journal of Virology*. 2001;75(6):2516–2525. doi:10.1128/JVI.75.6.2516-2525.2001.
58. Gregory NG, Biswas PK, Chowdhury SH. Recent Concerns about the Environment in Bangladesh. *Outlook on Agriculture*. 2010;39(2):115–120. doi:10.5367/000000010791745367.
59. Delabougline A, Thanh NTL, Xuyen HTA, Nguyen-Van-Yen B, Tuyet PN, Lam HM, et al. Poultry farmer response to disease outbreaks in smallholder farming systems in southern Vietnam. *eLife*. 2020;9:e59212. doi:10.7554/eLife.59212.
60. Masud AA, Rousham EK, Islam MA, Alam MU, Rahman M, Mamun AA, et al. Drivers of Antibiotic Use in Poultry Production in Bangladesh: Dependencies and Dynamics of a Patron-Client Relationship. *Frontiers in Veterinary Science*. 2020;7.

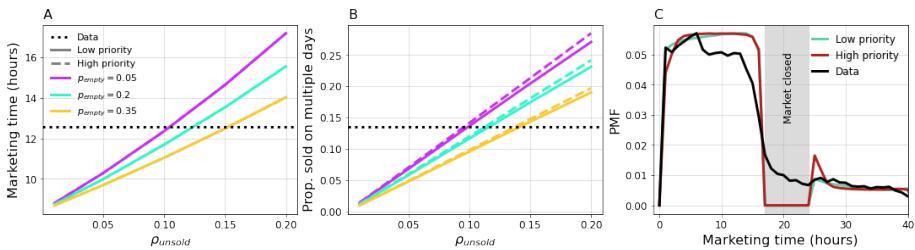
61. Wiltshire SW. Using an agent-based model to evaluate the effect of producer specialization on the epidemiological resilience of livestock production networks. *PLoS ONE*. 2018;13(3):e0194013. doi:10.1371/journal.pone.0194013.
62. Pomeroy LW, Moritz M, Garabed R. Network analyses of transhumance movements and simulations of foot-and-mouth disease virus transmission among mobile livestock in Cameroon. *Epidemics*. 2019;28:100334. doi:10.1016/j.epidem.2019.02.005.
63. Ansari S, Heitzig J, Brzoska L, Lentz HHK, Mihatsch J, Fritzemeier J, et al. A Temporal Network Model for Livestock Trade Systems. *Frontiers in Veterinary Science*. 2021;8.
64. Knight MA, White PCL, Hutchings MR, Davidson RS, Marion G. Generative models of network dynamics provide insight into the effects of trade on endemic livestock disease. *Royal Society Open Science*. 2021;8(3):201715. doi:10.1098/rsos.201715.



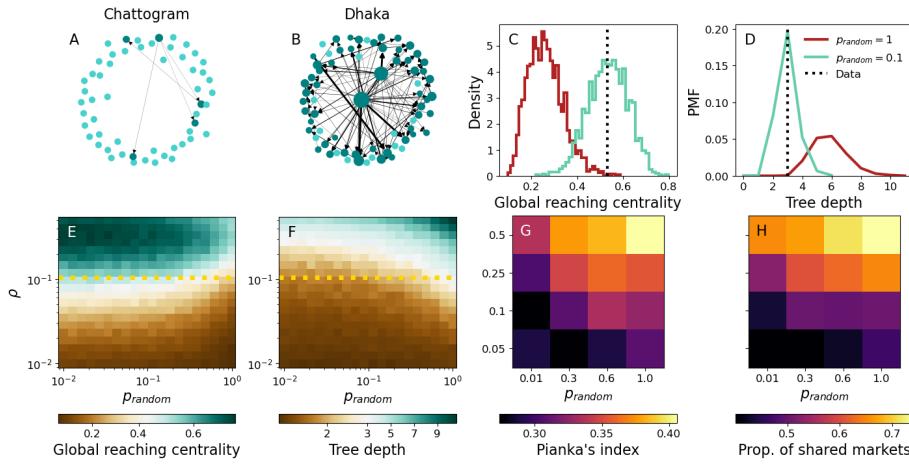
**Fig 2: Simulating poultry movements.** (A) Spatial population of 1200 farms supplying Dhaka. Farm locations are generated as described in Text S1 and assigned preferentially to upazilas with a larger observed outgoing chicken flux (colour scale). (B) Expected (black) and measured (red) distribution of times required to sell an entire batch. (C) Expected and measured distributions of transactions a single batch is split into. (D) Measured vs expected relative flux between individual pairs (dots) of upazilas and LBMs. (E) Distribution of LBMs serviced daily by individual middlemen. (F) Proportion of chickens sold to wholesalers (W, teal) and retailers (R, yellow) by LBM tier in simulations (bars) and data (markers).  $MM \rightarrow V_0$  refers to transactions involving middlemen and first tier vendors, while  $V_L \rightarrow V_{L+1}$  represents inter-tier transactions. For each tier, bars do not add up to 1 since wholesalers can sell to end-point consumers as well. Inset shows proportions of wholesalers and retailers. (G) Marketing time distribution. Results are obtained from a single simulation with default settings. Farm data are obtained from [27]. Data about middlemen and vendor trading practices and marketing times are obtained from [18].



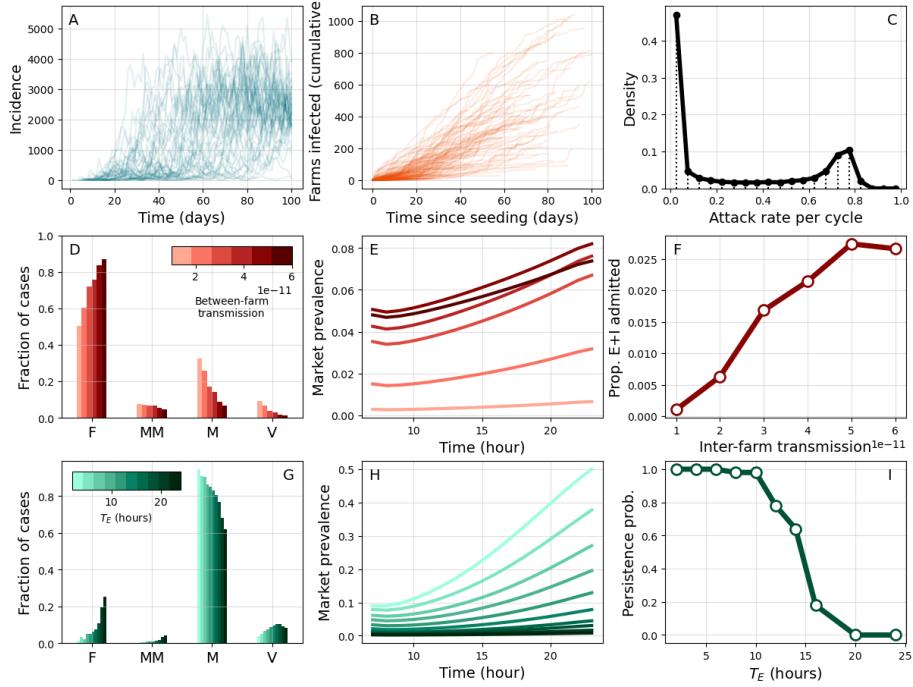
**Fig 3: Markets serviced daily.** (A) Expected (scatters) and measured (lines) distributions of markets serviced daily. Expected distributions are of the form  $Pr(k_m = n) \propto (1 - p_{k_m})^{n-1} p_{k_m}$  where  $n = 1, \dots, k_m$ . Inset compares expected and measured average numbers of markets serviced. (B) Distribution of vendors a single middleman trades daily with. (C) Cumulative distribution of sizes of transactions involving middlemen and vendors (solid lines). Dashed lines represent cumulative proportion of chickens sold in transactions up to a given size. Results are averaged over 50 simulations from 10 different PDN realisations.



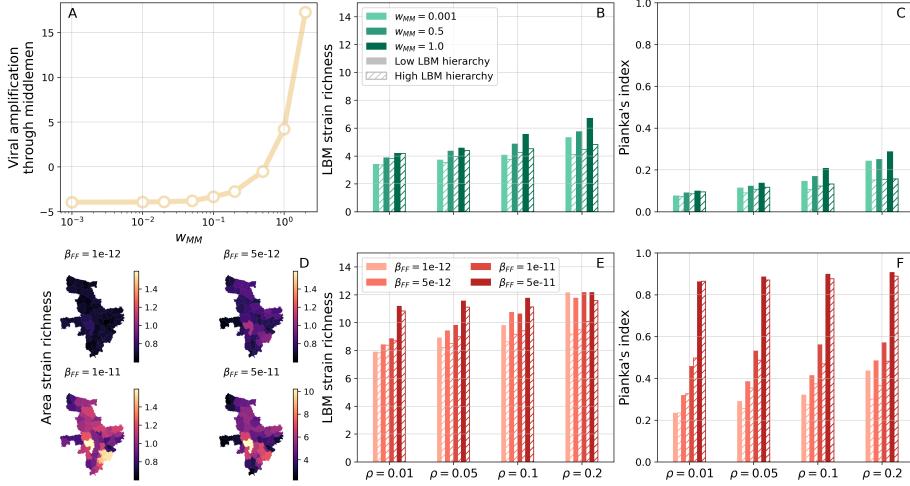
**Fig 4: Vendor trading practices.** (A) Average marketing time as a function of  $\rho_{unsold}$  for different values of  $\rho_{empty}$ . Solid and dashed lines correspond respectively to low (10%) and high (90%) frequency of vendors prioritizing trading older chickens. (B) Proportion of marketed chickens offered for sale on multiple days. (C) Marketing time distributions for low and high frequency of vendors prioritizing older chickens. Here,  $\rho_{unsold} = 0.1$  and  $\rho_{empty} = 0.2$ . Results are averaged over 50 simulations from 10 different PDN realisations.



**Fig 5: LBM networks and poultry mixing.** (A,B) Broiler LBM networks for Chattogram and Dhaka, respectively. An arrow pointing from market  $l$  to  $l'$  indicates at least one movement in that direction, while arrow thickness is proportional to the number of vendors moving on that edge. Node size is proportional to the outgoing weight, i.e. the total number of vendors leaving it. Isolated and connected nodes are shown in cyan and teal, respectively. (C,D) GRC and TD, respectively, for Dhaka's network (line) and ensembles of 2000 synthetic LBM networks with the same density as Dhaka's network and  $p_{random} = 1$  (red) and  $p_{random} = 0.1$  (cyan). (E,F) Average GRC and TD, respectively, across 100 networks with 20 nodes and as a function of  $\rho$  and  $p_{random}$ . Dotted line denotes Dhaka's density. (G,H) Pianka's index of overlap and proportion of markets where it is possible to find chickens from different upazilas/sub-districts, respectively, as a function of network parameters. Performing the same measurement before any vendor movement occurs, yields an overlap (Pianka's) of 0.261, and 25.7% shared markets, on average. This represents the baseline overlap due to middlemen sourcing chickens from farms and selling them to vendors. Results are averaged over 50 simulations from 10 different PDN realisations. All other PDN parameters are set to default values.



**Fig 6: Epidemic dynamics.** (A) Daily incidence in LBMs in multiple simulations. (B) Cumulative number of new farms infected over time from multiple clusters. Each cluster is initiated by a different infectious seed. (C) Distribution of attack rates for individual production cycles, conditional on at least one infection. (D-F) High farm transmission scenario ( $w_F = 0.2, w_M = 0.7$ ). Colour scale corresponds to varying levels of inter-farm transmission  $\beta_{FF}$ . (D) Proportion of incident cases in different setting types (F: farms, MM: middlemen, M: markets, V: vendors). (E) Average hourly prevalence in LBMs at stationarity. (F) Proportion of latent and infectious chickens entering markets daily as a function of  $\beta_{FF}$ . (G-I) High LBM transmission scenario ( $w_F = 0.1, w_M = 2.4$ ). Colour scale corresponds to varying latent period  $T_E$ . (G,H) mirror (D,E). (I) Persistence is measured as the proportion of simulations where at least one transmission chain persisting in markets and vendors for longer than 50 days was observed. Results are qualitatively the same under different different criteria about the duration of transmission chains (Fig. S5). Other parameters are set to default values. Results are based on 50 simulations from 10 different synthetic PDNs.



**Fig 7: Multi-strain dynamics and viral mixing in LBMs.** (A-C) Simulations with no inter-farm transmission ( $\beta_{FF} = 0$ ). (A) Viral amplification happening through transportation from farm to LBM gates as a function of middlemen-specific transmission weight  $w_{MM}$ . This is quantified through the difference between total numbers of exposed and infected chickens sold to vendors and purchased daily by middlemen. (B) Average strain richness (i.e. number of strains) in single LBMs as a function of density  $\rho$  of vendor movements (on the x-axis),  $w_{MM}$  (from light to dark). Solid and striped bars correspond to low and high hierarchy in vendor movements, respectively. (C) Average Pianka's index of overlap between pairs of LBMs in terms of their catchment areas. (D-G) Simulations with inter-farm transmission. (D) Average richness per upazila for increasing  $\beta_{FF}$ . Note that the bottom-right map uses a different colour scale. (E,F) Same as (B,C) but for varying  $\beta_{FF}$  and with  $w_{MM} = 0.001$ . We set  $w_F = 0.1$  in (A-C) and  $w_F = 0.2$  in (D-F), while  $w_M = 2.4$  and  $w_V = 1$  in all panels. Cross-immunity reduces susceptibility to secondary infections to  $\sigma = 0.3$ . Results are averaged over 50 simulations from 10 different synthetic PDNs. In each simulation, statistics are collected for 100 days after an initial transient of 2000 days.

Supplementary Text for the manuscript  
"EPINEST, an agent-based model to simulate  
epidemic dynamics in large-scale poultry  
production and distribution networks"

## Contents

<b>1</b>	<b>Data analysis</b>	<b>1</b>
1.1	Farmers	1
1.2	Middlemen	2
1.3	Vendors	4
<b>2</b>	<b>Actor dynamics</b>	<b>7</b>
2.1	Farms	7
2.2	Middlemen	8
2.3	Markets	10
2.4	Vendors	11
<b>3</b>	<b>PDN setup</b>	<b>12</b>
3.1	Geography setup	13
3.2	Farm generation	13
3.3	Market setup	14
3.4	Vendor setup	15
3.5	Middlemen setup	16
<b>4</b>	<b>Simulating epidemic spread</b>	<b>17</b>
4.1	Within-flock transmission	18
4.2	Between-farm transmission	19
4.3	External introductions	20

## 1 Data analysis

### 1.1 Farmers

#### Questionnaire data

Data were generated through a cross-sectional study that collected information about farming practices from 100 distinct farms in Chattogram division,

Bangladesh [1]. Available data includes farm sizes, the number of production cycles completed in one year, as well as the numbers of traders and transactions involved in clearing individual batches (Fig. A1). Here we focus on the 47 farms that reported trading broilers and raising a single batch at a time in a single shed. It should be noted, however, that while the questionnaire aimed to uncover farmers' practices over one year, it is not possible to know whether farmers raised at most one batch at a time for the entire period as their answers reflect their situation during the interview.

**Farm/batch size.** We modelled batch sizes  $S_F$  by assuming a truncated negative binomial distribution. More in detail, we assumed:

$$S_F \sim \mathcal{NB}(p_{S_F}, n_{S_F}) \text{ if } S_F \in [S_F^{\min}, S_F^{\max}], \quad (1)$$

where  $S_F^{\min}, S_F^{\max}$  are the minimum and maximum observed batch sizes. For simplicity, we estimated parameters  $p_{S_F}, n_{S_F}$  by maximum likelihood without accounting for truncation and verified a posteriori that the resulting bias was small. Note that for this type of farms, batch size is the same as farm size, since they do not rear multiple batches at the same time.

**Replenishment time.** We modelled replenishment time  $\tau_{replenish}$  by assuming a shifted negative binomial distribution, i.e.:

$$\tau_{replenish} - 1 \sim \mathcal{NB}(p_{\tau_{replenish}}, n_{\tau_{replenish}}), \quad (2)$$

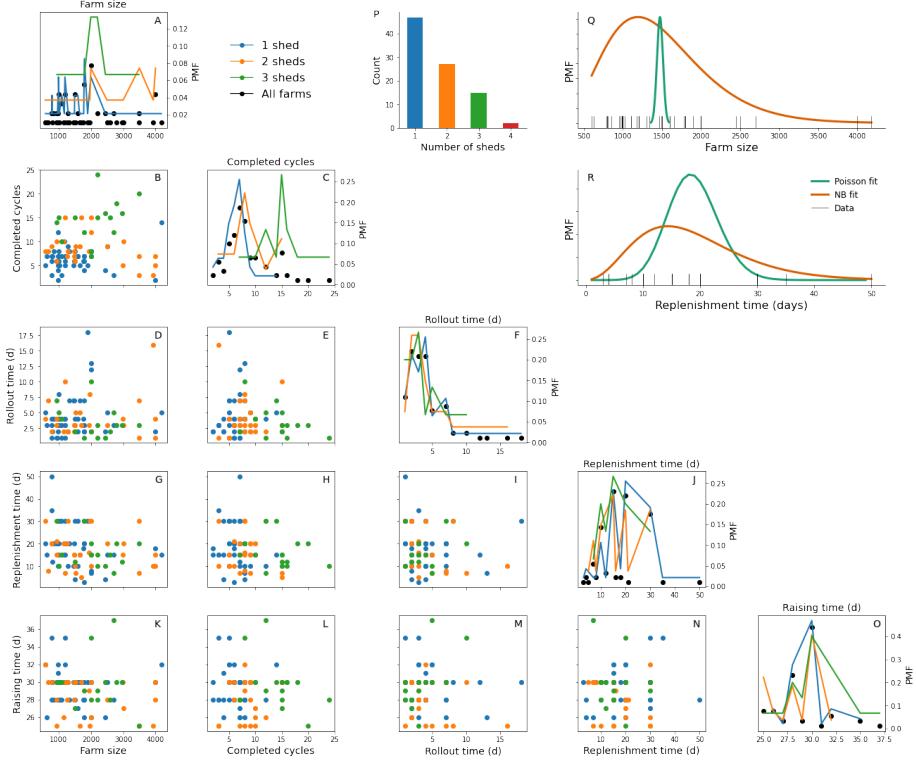
## 1.2 Middlemen

### Individual middlemen properties

We used data from questionnaires to inform middlemen's trading patterns. Consistently with the main manuscript, we restrict our analysis to broiler data only [2].

**Number of chickens bought daily.** We divided the total number of chickens bought during the entire study period by each middlemen with the number of days in which he/she purchased any poultry. We used these raw values to construct a discrete distribution from which to sample middlemen batch sizes  $S_{MM}$  in the ABM.

**Number of markets visited daily.** As a first step, we evaluated the number of working days for each middlemen by combining sale frequencies of different breeds. This was straightforward for middlemen that sold chickens every day during the study or sold a single chicken breed. One middleman declared selling chickens 4 times in a week; in this case, we set the number of working days to 4. We then obtained the mean number of daily market visits  $k^{(daily)}$  by dividing the total number of market visits made by each middleman during the study



**Figure A 1: Farm statistics from survey data and statistical fits.** (A-O) Univariate distributions (diagonal) and pairwise scatter plots (off-diagonal) for several farm statistics, coloured by the number of available sheds. Shown quantities include farm size (A), completed production cycles per year (C), trading rollout duration (F), time to replenish the farm (J) and raising time (O). We did not show farms with 4 sheds, as there were only two (see panel P showing frequencies of sheds per farm). (Q) Poisson (green) and truncated negative binomial (orange) fits to batch size data (black ticks). (R) Poisson (green) and shifted negative binomial (orange) fits to replenishment time data (black ticks). In both cases, a negative binomial distribution provides a better fit than Poisson to underlying data. Analyses in (Q,R) were restricted to farms with a single shed.

period with the corresponding number of working days. Note that some values  $k_i^{daily}$  are not integer numbers due to the previous calculation. Finally, we used maximum likelihood to fit a geometric distribution with parameter  $p_{k_m}$  to data  $\{k_i^{daily}\}$ , with  $i$  ranging from 1 to the total number of middlemen interviewed. More precisely, the maximum likelihood estimate for  $p_{k_m}$  is given by the inverse of the sample mean of data  $\{k_i^{daily}\}$ . This formula works also when some values  $k_i^{daily}$  are not integer numbers (Fig. A2).

40  
41  
42  
43  
44  
45  
46

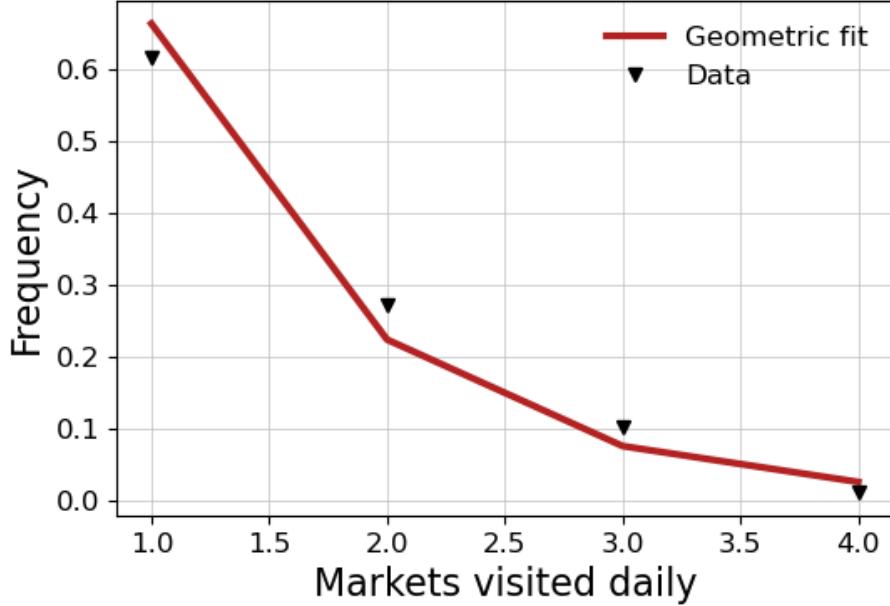


Figure A 2: **Distribution of daily market visits.** Fitted (line) and observed (markers) frequencies of markets visited daily. The fitted distribution is geometric with parameter  $p$  estimated using maximum likelihood. Note that some data points were not integers and were therefore rounded to the nearest integer in this plot. The fitted distribution corresponds to black markers in Fig. 2E, but is here shown up to value 4, i.e. the maximum observed count.

### 1.3 Vendors

#### Individual vendor properties

We used data from questionnaires to inform vendors' trading patterns. Unless otherwise stated, we repeat every analysis for retailers and wholesalers.

**Number of chickens bought daily.** We first reconstructed the number of chickens bought by a single vendor over the survey period by adding counts of sold and unsold chickens. Then, we divided this number by the number of days during which the same vendor bought any chickens, yielding an estimate of the number of chickens bought daily. We used these raw values to construct a discrete distribution from which to sample vendor batch sizes  $S_V$  in the ABM (Fig. A3A).

**Surplus chickens** First, we estimated the probability  $p_{empty}$  to sell the entire batch of chickens before a vendor buys another batch. This was computed as

the ratio between the number of days with any unsold chickens and the number of days with a purchase.

Second, we modelled the counts  $u_{batch}$  of unsold chickens per batch, conditional on batch size  $S_V$  and on the event that the batch is not fully emptied (which happens with probability  $1 - p_{empty}$ ). We consider a negative binomial distribution with mean  $\mu = \rho_{unsold} S_V$  and variance  $\mu \cdot (1 + n_{unsold}^{-1} \cdot \mu)$ , where  $\rho_{unsold}$  is the proportion of chickens that remain unsold before purchasing the next batch and  $n_{unsold}$  affects overdispersion compared to a Poisson distribution. Because our data are aggregated over a week period, we model the total count of unsold chickens  $u_{total}$  rather than  $u_{batch}$ . Because the sum of negative binomial random variables still follows a negative binomial distribution,  $u_{total}$  has the same distribution as  $u_{batch}$ , with  $S_V$  multiplied by the number of days with any unsold chickens (as we are conditioning on having a surplus). Aggregating data from wholesalers and retailers yields maximum likelihood estimates  $\rho_{unsold}^{MLE} = 0.13$  and  $n_{unsold}^{MLE} = 3.4$  (Fig. A3B-C), which are used throughout the main manuscript. For completeness, we also estimate the same parameters for wholesalers and retailers separately, who amount to 55 and 376 data points, respectively. We find that wholesalers generate, on average and conditional on at least one unsold chicken, less surplus chickens than retailers ( $\rho_{unsold}^{MLE,W} = 0.06$  vs  $\rho_{unsold}^{MLE,R} = 0.13$ ), while the amount of overdispersion is roughly the same ( $n_{unsold}^{MLE,W} = 3.4$  and  $n_{unsold}^{MLE,R} = 4.2$ ). Note that we assume  $p_{empty}$  to differ between retailers and wholesalers. It should also be added that in the context of questionnaire data, a surplus of chickens is defined in relation to consecutive purchases, which could be multiple days apart, by the same vendor. In the ABM, however, we conflate these parameters with surplus between consecutive days, as vendors tend to buy new chickens every day. The resulting discrepancy should not be large since 76% of interviewed vendors purchased (broiler) chickens every day, and 93% during at least 6 days in a week. Finally, note that during simulations, daily surplus is also subject to the constrain that it can not be larger than batch size.

**Prioritising unsold chickens.** We estimate the probability of a vendor prioritising the sale of previously unsold chickens as the proportion of vendors declaring to do so in our data.

### Transaction networks and inter-tier fluxes

Here we detail a procedure to estimate  $p_W^{(L)}$  and  $p_R^{(L)}$  for each market tier  $L = 0, 1, \dots, L_{max}$ . Parameters  $p_{W,R}^{(L)}$  represent the proportions of chickens sold respectively to wholesalers and retailers in tier  $L$ , while  $1 - p_W^{(L)} - p_R^{(L)}$  is the proportion of chickens that wholesalers in tier  $L-1$  ( $L > 1$ ) sell to end-point consumers.

The tier  $L = 0$  corresponds to vendors that buy directly from middlemen. Note that  $p_W^{(0)} + p_R^{(0)} = 1$  since middlemen are not allowed to sell chickens to end-point consumers. Wholesalers in tier  $L$  sell to wholesalers and retailers in

tier  $L + 1$  according to probabilities  $p_W^{(L+1)}$  and  $p_R^{(L+1)}$ , respectively. Note that the last tier  $L_{max}$  must consist of retailers only and hence  $p_W^{(L_{max})} = 0$ .

Here we outline a simplified estimation procedure that yields parameters that are common to all markets. We start with reconstructed transaction networks. These networks are DAGs consisting of multiple transactions, weighted by the number  $g$  of chickens involved, between different node types. Nodes can be either middlemen (MM) or vendors, the latter being further classified into either wholesalers (W) or retailers (R). In addition, MM and W nodes are further classified according to their 'depth' within a transaction chain. This latter characterization is somewhat reminiscent of our tiered structure, but a direct mapping is not possible at this stage yet. In particular, in order to infer relevant parameters from reconstructed transaction chains, we must take care of transactions involving middlemen that purchase chickens directly from wholesalers and sell them to other vendors. As our ABM does not allow middlemen to buy from vendors, we replace these 'forbidden' transactions found in reconstructed networks, with sets of 'allowed' transactions. Let  $V_{MM}^*$  denote the set of such middlemen. As a first step, we remove all transactions involving poultry farms, except those involving any middlemen  $k \in V_{MM}^*$ . For each remaining farm  $i$ , we replace middlemen  $k \in V_{MM}^*$  with a new label  $k^*$  so that each transaction  $i \rightarrow k$ ,  $k \in V_{MM}^*$  becomes  $i \rightarrow k^*$ . The label  $k^*$  is just a placeholder representing a fictitious middleman. Second, We then remove transactions of the type  $i \rightarrow k$  and  $k \rightarrow j$ , where  $i, j$  are not farms and  $k \in V_{MM}^*$ , and replace them with viable transactions of the type  $i \rightarrow j$ . Thus, if actor  $i$  sold chickens to  $k \in V_{MM}^*$ , which in turn sold chickens to vendor  $j$ , we remove edges  $(i, k)$  and  $(k, j)$ , and create a new edge  $(i, j)$  with a weight  $g_{i,j} = g_{i,k}g_{k,j}/g_k$ , where  $g_k = \sum_l g_{k,l}$  is the total out-weight of  $k$ . Finally, we remove transactions involving only middlemen and merge any duplicated edges together.

We can now analyse the resulting graph  $G$ . Let  $V_{MM}$  represent the set of middlemen nodes, i.e. the roots of the DAG. Note that  $V_{MM}$  excludes nodes from  $V_{MM}^*$ . We associate middleman  $i$  with a weight  $g_i^{(0)}$  representing the number of chickens it 'injects' into markets. For each middleman node  $i \in V_{MM}$ , we then perform the following calculation:

- Enumerate all chains departing from node  $i$  and ending with a leaf node. The latter might be a W or a R node. We stress that these are observed transactions, not the output of our ABM.
- For each such chain, count the number of mark-ups, distinguishing between R and W nodes. Each chain may involve any number of W nodes and an optional terminal R node. For example, a chain of the type  $W \rightarrow W \rightarrow R$  consists of two wholesalers mark-ups before landing into a retailer. A chain of the type  $W \rightarrow W$  consists of two wholesalers mark-ups, with the terminal wholesaler selling directly to end-consumers. Finally, assign a weight to the current chain by multiplying  $g_i^{(0)}$  by the proportion of chickens that reach end-consumers through this chain. The latter is easily computed by multiplying together the proportions of chickens flowing

through each edge in the chain. It is easy to see that the weights of chains emanating from node  $i$  must add up to  $g_i^{(0)}$ .

By enumerating all chains constructed in this way, we can easily determine the maximum number of tiers as the length of the longest chain in  $G$ . For example, if the longest chain is of the type  $W \rightarrow W \rightarrow R$ , then we need 3 tiers and  $L = 0, 1, L_{max} = 2$ .

Once we have computed the weights for every chain in  $G$ , we can finally obtain  $p_W^{(L)}$  and  $p_R^{(L)}$  by collapsing all chains according to their length and the type of terminal node. The idea is to compute a running sum  $g_G(L, T)$  that counts how many chickens end up in the hands of actors of type T, where T is either W, R or C (end-point consumer) in tier L. Please note that we are deliberately over-counting chickens according to how many mark-ups they make. Let us illustrate this procedure by considering a generic chain with weight  $g_c$ ,  $n_W \geq 0$  wholesaler and  $n_R = 0, 1$  retailer mark-ups. The total length of the chain is  $n_c = n_W + n_R$ . For each  $L = 0, \dots, n_W - 1$  we increase the running sum  $g_G(L, W)$  by  $g_c$ . If the last node is of type W, we increase the running sum  $g_G(n_W, C)$  by  $g_c$ ; else, the last node is of type R and we increase the running sum  $g_G(n_W, R)$  by  $g_c$ .

Finally,  $p_W^{(L)} = g_G(L, W)/[g_G(L, W) + g_G(L, R) + g_G(L, C)]$  and  $p_R^{(L)} = g_G(L, R)/[g_G(L, W) + g_G(L, R) + g_G(L, C)]$ . Note that  $g_G(0, C) = 0$  if we ignore transactions involving MM selling to end-point consumers.

Note that we can include more reconstructed transaction networks in the same analysis: we compute  $g_G(L, T) \forall G$  and then create a consensus quantity  $g_{consensus}(L, T) = \sum_G g_G(L, T)$ , which we can finally use to compute parameters of interest.

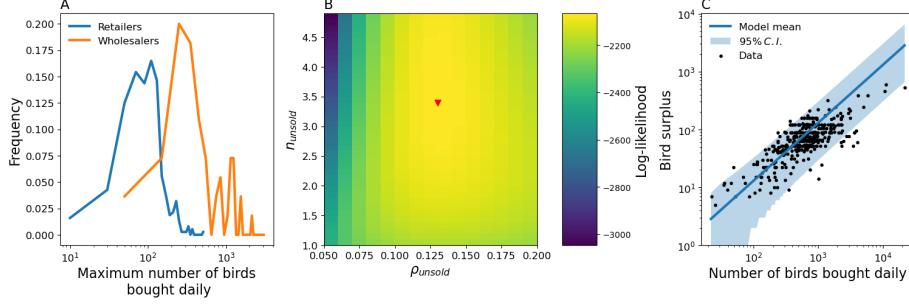
## 2 Actor dynamics

This section describes in detail different action and tasks for each actor. Further details on actor instantiation can be found in the section dedicated to PDN setup.

### 2.1 Farms

At any time, a farm is either empty or raising chickens. We consider farms raising a single batch of chickens at a time, meaning that individual production cycles do not overlap. Chickens from the same batch are introduced at the same time as day-old chicks and are offered for sale as soon as they reach an appropriate age  $\tau_{raise}$ . A production cycle ends when all chickens in a farm are sold.

**Farm replenishment.** After completing a cycle, a farm remains empty for a random time  $\tau_{replenish}$ , sampled from a (shifted) negative binomial distribution. After  $\tau_{replenish}$  time steps, a farm recruits a new batch of day-old chicks.



**Figure A 3: Vendors' parameters analysis.** (A) Raw distributions of number of birds bought daily for R (blue) and W (orange). (B) Log-likelihood function for negative binomial fit to surplus counts. Fitted parameters include the average proportion on unsold chickens  $\rho_{unsold}$  and the shape parameter  $n_{unsold}$ . The maximum likelihood solution ( $\rho_{unsold}^{MLE}, n_{unsold}^{MLE}$ ) is indicated with a red marker. (C) Expected surplus  $\rho_{unsold}^{MLE} \cdot S$  (line) as a function of batch size  $S$ , together with 95% C.I. (shaded area) calculated assuming a negative binomial distribution  $\mathcal{NB}(\rho_{unsold}^{MLE}S, n_{unsold}^{MLE})$  for surplus. Broiler data are shown as well (black dots).

**Offering birds for sale.** After raising a batch for  $\tau_{raise}$  days, a farm is ready to sell chickens. These are offered for sale progressively over a minimum of  $\tau_{rollout}$  days; more precisely, a farm stages new chickens corresponding to a fraction  $1/\tau_{rollout}$  of its batch over the first  $\tau_{rollout}$  days. By day  $\tau_{rollout}$ , all chickens at the farm can be sold to middlemen. Note that a farm could sell chickens for a longer period of time if not enough middlemen are available to trade with. In the case where any chickens remain unsold after  $\tau_{rollout}^{(max)}$  days, the corresponding farm is emptied automatically.

## 2.2 Middlemen

Middlemen daily routine consists of buying and collecting chickens from farms in order to sell them to vendors operating in markets. A single middleman is able to source chickens from multiple farms, potentially from distinct regions, in order to fill up his/her cargo.

**Updating scouted regions.** Each middleman purchases chickens from farms located in a subset  $\mathcal{A}$  of neighboring areas, whose number we denote with  $n_{scout} = |\mathcal{A}|$ . With daily frequency, each middleman may move and change his/her catchment area with probability  $P_{move}$ . In that case, a middleman updates  $\mathcal{A}$  by choosing  $n_{scout}$  new regions as follows: first, a focal region  $a$  is sampled at random with propensity proportional to the number of chickens being currently offered for sale in that region. In the case no farm is currently trading,  $a$  is chosen fully at random. Second,  $n_{scout} - 1$  new regions are chosen

185  
186  
187  
188  
189  
190  
191  
192

193  
194  
195  
196  
197

198  
199  
200  
201  
202  
203  
204  
205

at random among those neighboring  $a$ . More precisely, we say that a region  $a'$  neighbors  $a$  if their centroids are less than 80 km far apart. If  $a$  has less than  $n_{scout} - 1$  neighbors, new regions are chosen among the areas neighboring with  $a$ 's direct neighbors. The choice of a 80 km radius is arbitrary, but it guarantees that middlemen do not cover too long distances in a single day. Note also that according to our algorithm, middlemen are drawn preferentially to areas with the largest offer of chickens, which reduces the odds of farms not being able to sell their chickens.

**Buying birds from farms.** The main purpose of mobile traders is to collect chickens from farms and deliver them to LBMs. With daily frequency, our algorithm allocates amounts of chickens to be moved from any trading farm to any market via middlemen. No chicken is collected at this stage yet. Our simulator resolves allocations one middleman at a time in random order and under the following constraints:

1. Middleman  $j$  can hold at most  $S_{MM}$  chickens.
2. Middleman  $j$  can source chickens from farms located in any area  $a \in \mathcal{A}$ .
3. Middleman  $j$  sells chickens to a number of vendors from exactly  $k_m$  markets.
4. Destination markets must be chosen in a way the preserves the overall proportion of chickens leaving region  $a$  and entering market  $l$ ,  $f_{a,l}$ .

As a consequence of the last two requirements, each middleman keeps track of which market each chicken is scheduled to be moved to.

The first step consists in sampling markets with probabilities  $\{q_l\}_{l=1,\dots,N_M}$ :

$$q_l = \frac{\sum_{a \in \mathcal{A}} \tilde{O}_a f_{a,l}}{\sum_{a \in \mathcal{A}} \tilde{O}_a}, \quad (3)$$

where  $\tilde{O}_a$  is the number of birds currently offered for sale by farms in region  $a$ . We sample markets with replacement until we get a set  $\mathcal{M}$  comprising  $k_m$  distinct markets, or after 1000 draws, in which case  $|\mathcal{M}| < k_m$ . Let  $t_l$  be the number of times market  $l \in \mathcal{M}$  was selected during sampling, and let  $t = \sum_l t_l$  be the total number of draws. Focusing on market  $l$  with  $t_l > 0$ , we construct the distribution  $f_{a|l}$ :

$$f_{a|l} = \begin{cases} \frac{\tilde{O}_a f_{a,l}}{\sum_{a \in \mathcal{A}} \tilde{O}_a f_{a,l}} & \text{if } a \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases}.$$

We then sample tokens  $t_{l,a} \sim \text{Multinomial}(t_l, \{f_{a|l}\})$  such that  $\sum_{a \in \mathcal{A}} t_{l,a} = t_l$ . Given the tokens  $t_{l,a}$ , middleman  $j$  contacts farms in each region  $a \in \mathcal{A}$ ,

securing up to  $\lfloor \tilde{S}_{MM} t_{l,a} / t \rfloor$  birds destined to market  $l$ . Here,  $\tilde{S}_{MM} \leq S_{MM}$  denotes the maximum number of birds that can be purchased during the current day.  $\tilde{S}_{MM}$  is smaller than  $S_{MM}$  if middleman  $j$  is already carrying birds for any reason, e.g. if unsold from previous days (this is usually rare).

Middlemen contact farms in an area  $a$  sequentially, starting with those trading for the longest amount of time. Transactions follow a greedy heuristics: if middleman  $j$  has to collect  $s_a$  birds from region  $a$ , he/she will attempt to buy as many birds as possible from individual farms to meet such quota. In case middleman  $j$  fails to secure the required number of birds from region  $a$ , we allow  $j$  to contact any previously visited farms to meet the original quota. Additional chickens obtained in this way are directed to market  $l$  with probability:

$$\begin{cases} \frac{f_{a,l}}{\sum_{l' \in \mathcal{M}} f_{a,l'}} & \text{if } l \in \mathcal{M} \\ 0 & \text{otherwise} \end{cases} . \quad (5)$$

Finally, it should be noted that  $\tilde{O}_a$  decreases as chickens are allocated progressively to middlemen; in other words, availability of chickens decreases as we iterate over middlemen.

**Bird collection.** Middlemen start collecting birds as soon as the allocation step described in the previous section is terminated. At this point, middleman  $j$  is already aware of the farms to be visited and the amounts of birds to pick up from each of them. In simulations, middleman  $j$  can visit any of these farms in random order at any point in time before the next market opening. This is to ensure that  $j$  collects all of their chickens before bringing them to the market.

**Selling birds to vendors.** Each day, middleman  $j$  visits markets  $l \in \mathcal{M}$  sequentially. There,  $j$  sells chickens to available vendors until all  $s_l$  carried chickens scheduled for delivery to market  $l$  have been sold. These chickens can be sold only to vendors in the first market tier ( $L = 0$ ). A number  $s_l(W) \sim \text{Binomial}(s_l, p_W^{(0)})$  are allocated for sale to wholesalers in this tier. The remaining  $s_l - s_l(W)$  chickens, plus any of the  $s_l(W)$  chickens that could not be sold to wholesalers (e.g. because of limited buying capacity) are directed to retailers. Any unsold chickens remain in middleman's stock and are offered for sale in the same market on the following day. Please note that this event is expected to occur rarely in simulations.

## 2.3 Markets

Markets can be either open or closed: all markets open at  $T_{open}$  and close at  $T_{close}$  every day. Markets contain chickens only during opening hours.

## 2.4 Vendors

Vendors trade chickens within and between markets. There are two types of vendors: retailers, who sell chickens to end-point consumers only, and wholesalers, who can also sell chickens to other vendors. Vendors are organized in tiers: vendors in tier  $L > 0$  buy chickens from wholesalers in tier  $L - 1$ , while vendors in tier  $L = 0$  buy chickens directly from middlemen. Vendors first purchase chickens in market  $l$  and then move to another market  $l'$ , or stay in  $l$ , where they can sell chickens to other vendors and/or end-point consumers.

**Buying birds.** All vendors seek to buy as many chickens as possible during any transaction with middlemen and/or other vendors, compatibly with their own capacity and daily quota  $S_V$ . Note that a vendor may hold more than  $S_V$  chickens at a time due to the presence of surplus chickens. Nonetheless, the same vendor can not hold more chickens than the maximum carrying capacity  $C_V > S_V$ .

**Inter-market movements** At  $T_{open}$ , vendor  $k$  moves to market  $l$  to purchase chickens. Eventually,  $k$  stays in the same market or moves to a second location  $l'$  to trade. Chickens move alongside their owner. Note that  $k$ 's purchase and trade markets  $l, l'$  are invariant. In other words,  $k$  will always purchase chickens in  $l$  and commute to  $l'$  (if  $l \neq l'$ ) or stay in  $l$  (if  $l = l'$ ) in a given PDN realisation.

All vendor movements are resolved at  $T_{open}$  and are therefore instantaneous: vendors move to market  $l$ , purchase chickens and eventually change market during the same time step. Note that vendors in tier  $L$  must move after those in tier  $L - 1$ .

**Wholesaling.** Let us consider a wholesaler  $k$  operating in tier  $L$ . After purchasing chickens,  $k$  carries  $n_{tot}$  chickens, including also older chickens that remained unsold from previous days. In the current day,  $k$  sells  $n_{sale} = n_{tot} - n_{unsold}$  chickens, where  $n_{unsold}$  denotes the number of unsold chickens.  $n_{unsold}$  is a random variable whose sampling procedure is described in sections below. Proportions  $p_W^{(L+1)}$  and  $p_R^{(L+1)}$  of these  $n_{sale}$  are then directed to wholesalers and retailers in tier  $L + 1$ , respectively. The remaining chickens are sold to end-point consumers (retailing).

It is possible that  $k$  sells less chickens to  $L + 1$  tier wholesalers than planned. In this case, any surplus chickens are redirected to retailers. Similarly, any chickens that would remain unsold after the wholesaling phase are redirected to end-point consumers.

Please note that wholesaling is instantaneous, occurring toe-to-toe with inter-market movements as vendor switch from purchasing to trading chickens.

**Retailing.** Both wholesalers and retailers can sell chickens to end-point customers. While wholesaling is instantaneous, retailing rolls out over market opening hours, i.e. between  $T_{open}$  and  $T_{close}$ . More in detail, vendors can sell

chickens to end-point consumers between  $T_{open} + 1$  and  $T_{close} - 1$ ; this means  
310  
that chickens spend at least one time step at the market.  
311

The number of chickens sold in any time step over this period is stochastic,  
312  
but uniformly distributed. More in detail, we assume that the number of  
313  
chickens sold by vendor  $k$  during time step  $t$  is given by:  
314

$$\min\{n_{avail}(t), \text{round}(u + X(t))\}, \quad (6)$$

where  $X$  is a Poisson random variate with mean  $n_{avail}(t)/\delta t$ , with  $n_{avail}(t)$   
315  
and  $\delta t$  denoting the number of chickens destined to retail still in stock and  
316  
the time left until market closes, respectively.  $u$  is a random draw from an  
317  
uniform distribution on the unit interval whose role is to fix rounding issues,  
318  
while  $\text{round}(\cdot)$  guarantees a meaningful result by rounding  $u + X$  to the nearest  
319  
integer. Finally, we require that  $k$  sells all of the remaining chickens that had  
320  
been scheduled for retail, right before market closure, i.e. at  $t = T_{close} - 1$ .  
321

**Surplus chickens.** As explained above, vendor  $k$  may retain a random number  
322  
 $n_{unsold}$  of chickens each day, which are then offered for sale the day after.  
323  
The surplus  $n_{unsold}$  is computed as follows: first, vendor  $k$  sells the entire stock  
324  
with probability  $p_{empty}$ , in which case  $n_{unsold} = 0$ . Alternatively,  $n_{unsold}$  is  
325  
sampled from a negative binomial distribution with mean  $\rho_{unsold} \cdot n_{tot}$ , with  
326  
the constraint  $n_{unsold} \leq n_{tot}$ . More details can be found in the data analysis  
327  
section.  
328

**Prioritising unsold birds.** Vendors keep track of unsold birds. We assume  
329  
that a proportion  $P_{priority}$  of vendors in our simulations prioritise selling these  
330  
birds before recently purchased birds. The remaining vendors, instead, do not  
331  
prioritise repurposed chickens over those bought in the current day.  
332

### 3 PDN setup

This section illustrates the generative algorithm responsible for instantiating the  
333  
PDN. The following subsections reflect the sequential steps of the algorithm.  
335  
The final goal is to generate a kind of supply chain with multiple intermediate  
336  
nodes under certain structural constraints.  
337

The algorithm first instantiates 'source' and 'sink' nodes, i.e. farms and  
338  
markets, where chickens are first introduced and sold last, respectively. Then,  
339  
it populates markets with vendors, compatibly with incoming bird flux as deter-  
340  
mined by farm production and region-to-market fluxes. Finally, our algorithm  
341  
instantiates middlemen, i.e. the intermediate actors.  
342

We assume here for simplicity that all actors handle and trade a single  
343  
chicken breed.  
344

### 3.1 Geography setup

Using external specifications about the study area, the algorithm creates a list of  $N_A$  regions, each being characterized by a given location and a weight proportional to the share  $O_a^{(0)}, a = 1, \dots, N_A$  of chickens produced there. In the case of Bangladesh we identify individual regions with upazilas. In addition, we also specify a matrix  $f_{a,l}$  representing the proportion of chickens that end up in market  $l = 1, \dots, N_M$  from area  $a$ . Note that  $\sum_l f_{a,l} = 1$ .

### 3.2 Farm generation

**Farm properties.** In this work we consider only farms raising a single batch of chickens at a time. For this type of farm then, farm size  $S_F$  is equivalent to batch size; the latter is assumed to vary across farms and is sampled from a truncated negative binomial distribution. In addition, batch size is invariant, in the sense that a particular farm will always raise batches with the same size. The raising time  $\tau_{raise}$  is assumed to be a constant and shared by all farms. Analogously to batch size, a random value of minimum rollout duration  $\tau_{rollout}$  is assigned to each farm according to a probabilistic distributions but does not change over the course of a simulation. Finally, replenishment time  $\tau_{replenish}$  is assumed to be sampled from a shifted negative binomial distribution, namely  $\tau_{replenish} \sim 1 + \mathcal{NB}(p_{\tau_{replenish}}, n_{\tau_{replenish}})$ , whenever a farm completes a rollout. Farm properties and model parameters describing farm generation are listed in Table A1.

**Farm locations.** For each farm we draw a random location according to the following algorithm: we first draw a region  $a$ , either uniformly at random with probability  $P_{random}$ , or proportionally to the outgoing flux  $O_a^{(0)}$  with complementary probability. Then, we sample a random point within the selected region.

Alternatively, empirical or simulated data on farm spatial distributions could be used to fix their locations.

**Farm output.** Once all farms have been generated, we compute a range of quantities that are fundamental to PDN generation and/or dynamics.

First, we compute the expected daily bird output  $O_i$  from farm  $i = 1, \dots, N_F$ . We compute  $O_i$  by assuming that farm  $i$  completes every rollout in exactly  $\tau_{rollout,i}$  days yielding:

$$O_i = \frac{S_{f,i}}{\langle \tau_{replenish} \rangle + \tau_{rollout,i} + \tau_{raise} - 1}, \quad (7)$$

it should be noted, however, that this overestimates the true expected daily output  $O_i$  since  $\tau_{rollout,i}$  represents only the minimum rollout duration, which might take longer in absence of middlemen to collect birds. Nonetheless, we will stick to the heuristic calculation in eq. S(7) to evaluate other quantities. These include expected daily bird output from region  $a$ :

$$O_a = \sum_{i \in a} O_i, \quad (8)$$

where the sum runs over all farms located in region  $a$ . Given the regional outputs, it is possible to compute the following quantities:

•  $\tilde{O}_a = O_a / \sum_{a'} O_{a'}$ , a normalized version of  $O_a$ . 383

•  $q_{a,l} = O_a \cdot f_{a,l}$ , the expected daily bird flux from region  $a$  to market  $l$ . 384

•  $M_l = \sum_a q_{a,l}$ , the expected daily bird flux impinging on market  $l$ . 385

Table A 1: **Farm-specific parameters.** A fraction  $f_{random}$  of farms are allocated in random upazilas, while remaining farms are assigned to upazilas proportionally to their volume of traded chickens. Farm locations are completely random within upazilas. Farm size and replenishment time distributions are  $\mathcal{NB}(p, n)$ . Farm size  $S_F$  is further constrained in the range  $[S_F^{min}, S_F^{max}]$ . Farm size is sampled once per farm per realisation: a farm always recruits the same amount of birds. Refill and rollout times are sampled during each production cycle.

Parameter	Meaning	Distribution/Value	Source
$N_F$	Number of farms	1200	-
$f_{random}$	Prob. random allocation	0.3	Assumed
$T'$	Farm update time	11 a.m.	-
$S_F$	Farm size	$\mathcal{NB}(p_{S_F}, n_{S_F})$	Estimated
$p_{S_F}$	Farm size $\mathcal{NB}$ parameter	0.003544	Estimated
$n_{S_F}$	Farm size $\mathcal{NB}$ parameter	5.245	Estimated
$S_F^{min/max}$	Min/max farm size	600, 4180	Estimated
$\tau_{raise}$	Raising time	32 days	[1]
$p_{\tau_{replenish}}$	$\tau_{replenish}$ $\mathcal{NB}$ parameter	0.2013	Estimated
$n_{\tau_{replenish}}$	$\tau_{replenish}$ $\mathcal{NB}$ parameter	4.488	Estimated
$P(\tau_{rollout})$	Min. rollout time distribution	(0.2, 0.3, 0.3, 0.2) for $\tau_{rollout} = 1 - 4$ days	Assumed

### 3.3 Market setup

Details about markets are provided externally. At this stage, the algorithm instantiates  $N_M$  markets, each structured in  $L_{max} + 1$  empty tiers. Market properties and model parameters describing market generation are listed in Table A2.

Table A 2: Market parameters.

Parameter	Meaning	Distribution/Value	Source
$N_M$	Number of LBMs	20	-
$T_{open}$	LBM opening time	6 a.m.	-
$T_{close}$	LBM closing time	11 p.m.	-
$w_{l,l'}$	Prob. vendor purchasing in LBMs $l$ sells in $l'$	Explored	Modelled
$f_{a,l}$	Prop. of chickens leaving upazila $a$ for LBM $l$	Empirical	[2]

### 3.4 Vendor setup

**Vendor properties.** At setup, each vendor is assigned a batch size  $S_V$ , drawn from a probabilistic distribution, the latter being different for W and R. Here,  $S_V$  denotes the maximum number of chickens he/she can purchase in a single day. Parameters describing properties of vendors and their generation are listed in Table A3, while parameters relating to market tiers are listed in Table A4.

**Tier-by-tier vendor allocation.** As explained in the main text, each vendor operates in up to two markets, and always in the same tier. Our algorithm assigns vendors to markets and tiers within them in a way that is compatible with the expected flux of chickens entering the market, as well as with between-tier fluxes. Importantly, neither the number of W and R can be specified a priori as they are determined by our algorithm, which we define as follows.

Starting from tier  $L = 0$ , we assign as many R and W to that tier so that their combined capacities  $S_V$  match the expected R and W incoming fluxes  $M_l \cdot p_R^{(0)}$  and  $M_l \cdot p_W^{(0)}$  from middlemen. Then, we assign a random destination market  $l'$  with probability  $w_{l,l'}$  to each vendor generated in the current tier; If  $l' = l$ , a vendor will operate in a single market.

We then proceed by evaluating the expected flux of chickens  $M_l^{(1)}$  impinging on tier  $L = 1$  due to vendors in tier  $L = 0$ ; this amounts to compute the combined expected output from all  $L = 0$  wholesalers buying chickens in any market and selling in  $l$ . The expected wholesaler output is taken to be equal to  $S_V$ . Given  $M_l^{(1)}$ , we allocate as many R and W to match fluxes  $M_l^{(1)} \cdot p_R^{(1)}$  and  $M_l^{(1)} \cdot p_W^{(0)}$ . Finally, we repeat the same scheme for the deeper tiers as well.

As a final note, we allow for a multiplicative scaling factor, to be applied to  $M_l^{(0)}$ , whose effect is to allocate more vendors than what would be implied by our heuristic calculation of farm outputs.

Note that a vendor buying  $S_V$  chickens daily also sells the same amount of chickens on average. To see that, let us consider the following discrete-time process: during time step  $t$ , a vendor accumulates  $S_V$  chickens and sells all of his

stock with probability  $p_{empty}$ , or a proportion  $1 - \rho_{unsold}$  with complementary probability. The average surplus at time step  $t + 1$  is given by: 422  
423

$$u_{t+1} = (1 - p_{empty}) \cdot \rho_{unsold} \cdot (u_t + S_V).$$

At stationarity we must have  $u_{t+1} = u_t \equiv u^*$ , which implies that the amounts of chickens sold and purchased must balance each other, hence our claim. Solving for  $u^*$  yields: 424  
425  
426

$$u^* = \frac{(1 - p_{empty}) \cdot \rho_{unsold} \cdot S_V}{1 - (1 - p_{empty}) \cdot \rho_{unsold}},$$

and an average daily stock, after acquiring  $S_V$  chickens: 427

$$S_V + u^* = \frac{S_V}{1 - (1 - p_{empty}) \cdot \rho_{unsold}}.$$

Finally, we take an individual's vendor maximum carrying capacity  $C_V$  as being 428  
1.5 times the average daily stock: 429

$$C_V = \frac{1.5 \cdot S_V}{1 - (1 - p_{empty}) \cdot \rho_{unsold}}.$$

**Table A 3: Vendor-specific parameters.** W,R denote wholesalers and retailers, respectively. When some surplus is generated (with probability  $p_{empty}$ ), it is sampled from a negative binomial distribution with mean  $m = \rho_{unsold} \cdot n$ , where  $n$  denotes total chickens offered for sale, and overdispersion parameter  $\alpha_{unsold}$  (the variance is  $m + \alpha_{unsold} * m^2$ ).

Parameter	Meaning	Distribution/Value	Source
$S_V$	Max chickens purchased	Empirical	[2]
$p_{empty}$	Prob. no surplus	0.32 (W), 0.15 (R)	Estimated
$\rho_{unsold}$	Prop. surplus	0.13 (W & R)	Estimated
$\alpha_{unsold}$	Overdispersion surplus	0.29 (W & R)	Estimated
$P_{priority}$	Prop. vendors selling older chickens first	0.84 (W), 0.49 (R)	Estimated

### 3.5 Middlemen setup

**Middlemen properties.** Middlemen are assumed to trade a single chicken breed. Each middleman is assigned a capacity  $S_{MM}$  from a discrete distribution, denoting the maximum number of chickens it can buy in a single day. Analogously to vendors, the number of middlemen can not be specified a priori, as it is determined dynamically: we allocate as many vendors so that their combined capacity matches the expected daily output from farms, i.e.  $\sum_i O_i$ . 430  
431  
432  
433  
434  
435  
436

Table A 4: **Market tiers’ parameters.** Parameters  $p_{W,R}^{(L)}$  represent the proportions of chickens sold respectively to wholesalers and retailers in tier  $L$  from either middlemen (if  $L = 0$ ) or wholesalers from the previous tier (if  $L > 0$ ).  $1 - p_W^{(L)} - p_R^{(L)}$  represents instead the proportion of chickens that wholesalers in tier  $L - 1$  ( $L > 1$ ) sell to end-point consumers ( $p_W^{(0)} - p_R^{(0)} = 1$  since middlemen do not sell to end-point consumers). Values are the same for all markets. Parameters are estimated from reconstructed transaction networks. The last tier contains only zeros as vendors in the previous tier sell only to end-point consumers.

$L$	$p_W^{(L)}$	$p_R^{(L)}$
0	0.606	0.394
1	0.407	0.315
2	0.169	0.318
3	0	0.058
4	0	0.

In addition, each middleman is assigned an integer  $k_m$ , sampled from a probability distribution, representing the number of markets visited daily. The middleman will then commit to sell chickens to  $k_m$  different markets during each day ( $k_m$  does not change in time). Properties of middlemen and model parameters describing middlemen generation are listed in Table A5.

**Middlemen initial positions.** As explained in the main text, at any point in time, each middleman tracks a set of  $n_{scout}$  distinct regions. During PDN generation, we assign middlemen to regions as follows: we first select an initial area  $a$  uniformly at random. Then,  $n_{scout} - 1$  new areas are chosen among those neighboring  $a$ , i.e. within a distance  $d < d_{MM}$  (based on their centroids). In case not enough areas are selected, we choose among neighbors’ neighbors.

## 4 Simulating epidemic spread

Our simulator allows to simulate transmission of multiple pathogens/strains in the same poultry population. The current version of the simulator allows to simulate transmission across multiple scales, including at the level of the same flock and at the level of farms. All parameters describing pathogen transmission are listed in Table A6.

Table A 5: **Middleman-specific parameters.** Number of markets serviced daily is geometric with probability  $p_{k_m}$  (truncated above by the number of markets).  $k_m$  is sampled only once per middleman per realisation.

Parameter	Meaning	Distribution/Value	Source
$S_{MM}$	Cargo size	Empirical	[2]
$P_{move}$	Daily movement prob.	0.1	Assumed
$n_{scout}$	Number of scouted areas	4	Assumed
$d_{MM}$	distance between neighboring areas	80 km	Assumed
$p_{k_m}$	Markets serviced parameter	0.66	[2]

## 4.1 Within-flock transmission

Let us consider a population of chickens within a single setting, e.g. a farm, a middleman's truck, a market, or a vendor's shed. Note that at any time, any chicken belongs to one and only one setting.

We simulate transmission using Sellke's construction [3]. Briefly, we assign a hazard value  $h$  to each chicken, sampled from an exponential distribution with unit rate; then, any contact with an infectious chicken reduces the target chicken's hazard by an amount  $\delta h$ . Whenever  $h$  hits 0 due to an infectious contact, the target chicken is infected. Later,  $h$  is updated with another draw from an exponential distribution with unit rate.

We make the assumption that chickens within the same setting mix homogeneously at random. Therefore, infectious contacts are directed at random chickens. During simulations, each infectious chicken makes exactly one contact per time step with a randomly chosen chicken from the same setting.

Infection triggers a chain of events depending on the specified compartmental model. In SIR-like dynamics, an infected chicken becomes infectious immediately, but recovers after an infectious period  $\hat{t}_I$  sampled from a geometric distribution with PMF:

$$P(\hat{t}_I = k) = (1 - p_I)^k \cdot p_I, k = 0, 1, \dots, \quad (9)$$

where  $p_I = 1 - \exp(-(T_I + 1)^{-1})$  and  $T_I$  is the average infectious period. It is easy to check that the distribution in Eq. (9) has mean  $T_I$ . In simulations,  $\hat{t}_I$  is sampled immediately after an infection happens, say time  $t$ , and recovery is deferred to time  $t + \hat{t}_I$ . In models with latency, i.e. SEIR-like models, chickens become infectious only after an incubation time  $\hat{t}_E$  sampled from a geometric distribution with mean  $T_E$ .  $\hat{t}_E$  is sampled immediately after infection at time  $t$ , and the chicken becomes infectious only at time  $t + \hat{t}_E$ , at which point the corresponding infectious period is also sampled.

Let us now consider an infectious chicken  $i$  trying to infect chicken  $j$  with pathogen  $x$ . The overall hazard reduction  $\delta h$  can then be written as:

$$\delta h = \beta(x, i) \cdot S(x, j) \cdot w_{setting}, \quad (10)$$

where  $\beta(x, i)$  is the transmissibility of  $x$  and  $S(x, j)$  is susceptibility of chicken  $j$  to infection with  $x$ . The factor  $w_{setting}$  is a multiplier that depends only on the current setting, and accounts for differences in transmission across settings. The factors  $\beta(x, i)$  and  $S(x, j)$  may depend on the state of the infector and infectee, respectively. Susceptibility may account for example for previous exposure to the same pathogen, or cross-reactions induced by exposure to other pathogens/strains. In the single-strain SIR model, for example,  $S(x, j) = 0$  if  $j$  is infectious or recovered.

## 4.2 Between-farm transmission

We allow pathogens to spread between distinct farms. If a farm  $f$  contains infectious chickens, the probability of infecting another farm  $f'$  (provided  $f'$  is not empty), irrespective of whether  $f'$  is already infected or not, is computed as:

$$p_{f,f'} = \min\{1, n_f \cdot n_{f'} \cdot K(d_{f,f'})\}, \quad (11)$$

where  $n_f$ ,  $n_{f'}$  are the average numbers of chickens in farms  $f$  and  $f'$ , respectively,  $d_{f,f'}$  is the distance between  $f$  and  $f'$  and  $K(x)$  is a spatial kernel. If transmission occurs, a random chicken in  $f'$  is set as infected, conditional on not being already immune, and a random infector chicken is selected from farm  $f$ . If the infector is co-infected with multiple strains, a single carried strain  $s$  is chosen at random and transmitted to the infectee. Note that this process bypasses the hazard rate calculation and that all strains are equivalent in the context of between-farm transmission.

We simulate transmission between farms using the Conditional Entry algorithm [4]. The algorithm requires farms to be assigned to cells in a grid in order to exploit the fact that transmission is more likely to occur within cells than between them. We use an adaptive algorithm described in the same paper to construct a grid over the farm population. The algorithm relies on a hyper-parameter  $\lambda$ , here set to 15, that only affects the sizes and number of individual cells. Finally, because performing between-farm transmission is computationally expensive, we run the conditional entry algorithm only once per day rather than every time step.

In this work we consider a power-law transmission kernel:

$$K(d) = \begin{cases} \beta_{FF}, & \text{for } d < d_K \\ \beta_{FF} \left(\frac{d_K}{d}\right)^{\gamma_K}, & \text{for } d \geq d_K \end{cases} \quad (12)$$

where  $\beta_{FF}$  denotes the overall strength of spatial transmission.

### 4.3 External introductions

External transmission events are responsible for seeding and re-seeding pathogens in farms. Once a day we iterate over all farms and reduce the hazard of a randomly selected chicken  $i$  due to pathogen  $s$  by an amount:

$$\delta h_{ext,i} = \beta_{ext}(s)S(x, i). \quad (13)$$

In the main manuscript, we also consider an alternative seeding protocol that introduces different strains in distinct upazilas.

Table A 6: **Epidemic parameters.** Distributions of latent and infectious period are geometric with expected values typical of AIV infections. The transmission kernel's shape and parameters  $\gamma_K$  and  $d_K$  are instead inspired to a study of H5N1 epidemics in Dhaka region.

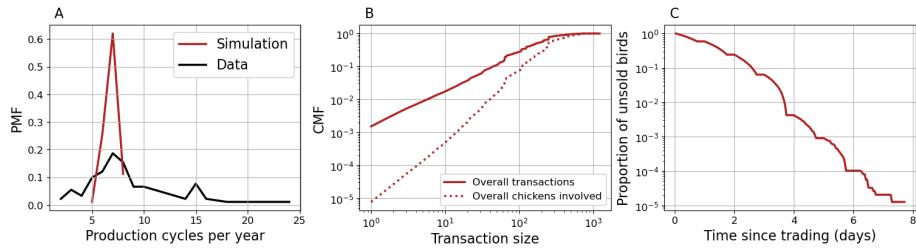
Parameter	Meaning	Distribution/Value	Source
$T_E$	Mean latent period	6 hours	[5]
$T_I$	Mean infectious period	48 hours	[5]
$\beta$	Base within-flock transmissibility	$0.2 \text{ hours}^{-1}$	Assumed
$\beta_{ext}$	Intro. rate	$0.0005 \text{ days}^{-1}$	Assumed
$\beta_{FF}$	Inter-flock transmissibility	$5 \cdot 10^{-11} \text{ days}^{-1}$	Assumed
$\gamma_K$	Exponent transmission kernel	0.94	[6]
$d_K$	Scale transmission kernel	0.1 km	[6]
$w_X$	Setting-specific transmission multiplier	0.2 (F), 1 (MM) 1 (M), 1 (V)	Assumed

## References

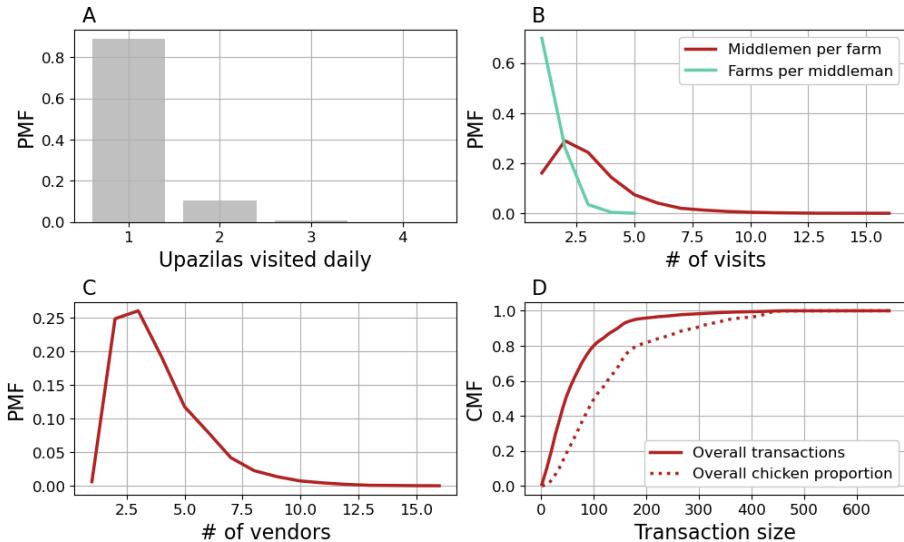
- [1] Suman Das Gupta, Md. Ahasanul Hoque, Guillaume Fournié, and Joerg Henning. Patterns of Avian Influenza A (H5) and A (H9) virus infection in backyard, commercial broiler and layer chicken farms in Bangladesh. *Transboundary and Emerging Diseases*, 68(1):137–151, 2021. eprint: <https://onlinelibrary.wiley.com/doi/10.1111/tbed.13657>.
- [2] Natalie Moyen, Md Ahasanul Hoque, Rashed Mahmud, Mahmudul Hasan, Sudipta Sarkar, Paritosh Kumar Biswas, Hossain Mehedi, Joerg Henning, Punam Mangtani, Meerjady Sabrina Flora, Mahmudur Rahman, Nitish C. Debnath, Mohammad Giasuddin, Tony Barnett, Dirk U. Pfeiffer, and Guillaume Fournié. Avian influenza transmission risk along live poultry trading networks in Bangladesh. *Scientific Reports*, 11(1):19962, October 2021. Number: 1 Publisher: Nature Publishing Group.

- [3] Thomas Sellke. On the Asymptotic Distribution of the Size of a Stochastic Epidemic. *Journal of Applied Probability*, 20(2):390–394, 1983. Publisher: Applied Probability Trust.
- [4] Stefan Sellman, Kimberly Tsao, Michael J. Tildesley, Peter Brommesson, Colleen T. Webb, Uno Wennergren, Matt J. Keeling, and Tom Lindström. Need for speed: An optimized gridding approach for spatially explicit disease simulations. *PLOS Computational Biology*, 14(4):e1006086, April 2018. Publisher: Public Library of Science.
- [5] Annemarie Bouma, Ivo Claassen, Ketut Natih, Don Klinkenberg, Christl A. Donnelly, Guus Koch, and Michiel van Boven. Estimation of Transmission Parameters of H5N1 Avian Influenza Virus in Chickens. *PLOS Pathogens*, 5(1):e1000281, 2009. Publisher: Public Library of Science.
- [6] Edward M. Hill, Thomas House, Madhur S. Dhingra, Wantanee Kalpravidh, Subhash Morzaria, Muzaffar G. Osmani, Mat Yamage, Xiangming Xiao, Marius Gilbert, and Michael J. Tildesley. Modelling H5N1 in Bangladesh across spatial scales: Model complexity and zoonotic transmission risk. *Epidemics*, 20:37–55, September 2017.

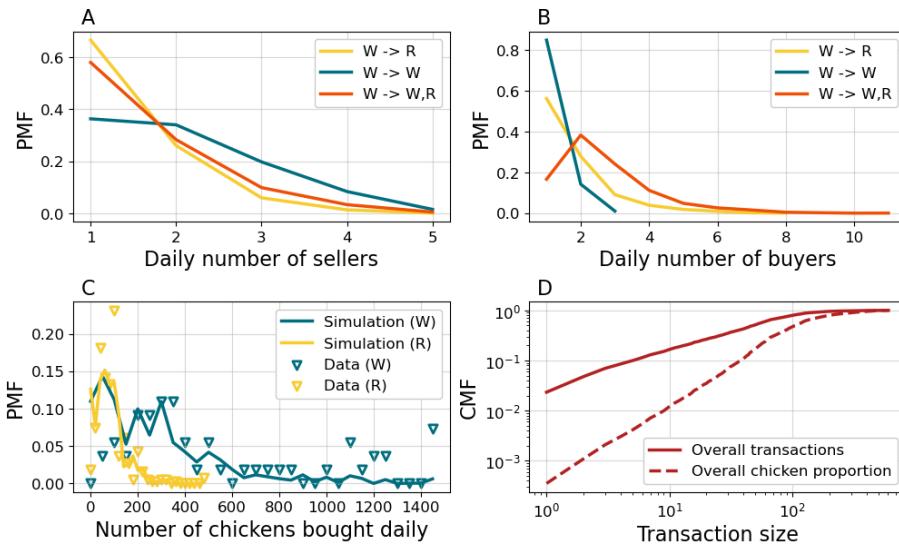
## Supplementary figures



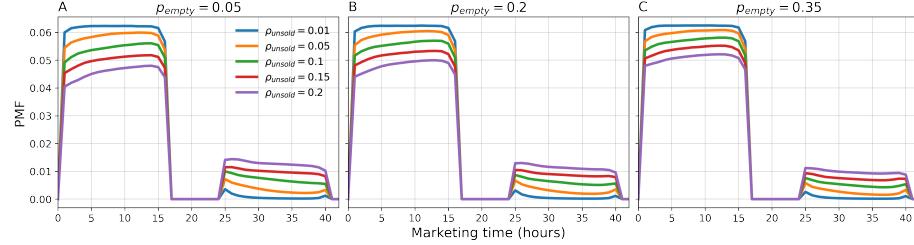
Supplementary Figure 1: **Additional farm statistics from simulations.** (A) Distribution of numbers of production cycles completed per year. The simulated distribution (red) appears narrower compared to empirical data (black) [1]. However, it should be added that several interviewed farmers raised multiple batches simultaneously, and those that declared raising a single batch during the interview may well have been managing 2 or more simultaneously during the previous year. (B) Cumulative distribution of sizes of transactions involving farms and middlemen (solid line). The dotted line represents the cumulative proportion of chickens sold in transactions up to a given size, given size. The corresponding PMFs, denoted with  $p_s$  and  $p'_s$  respectively, are related since  $p'_s = s \cdot p_s \sum_s s \cdot p_s$ . In other words,  $p'_s$  is the size-biased version of  $p_s$ . (C) Proportion of chickens remaining unsold after a given time since being offered for sale for the first time by a farmer. Note that it is highly unlikely for a chicken to remain unsold for more than 5 days. Results are obtained from a single simulation with default settings as in Fig. 2 in the main manuscript.



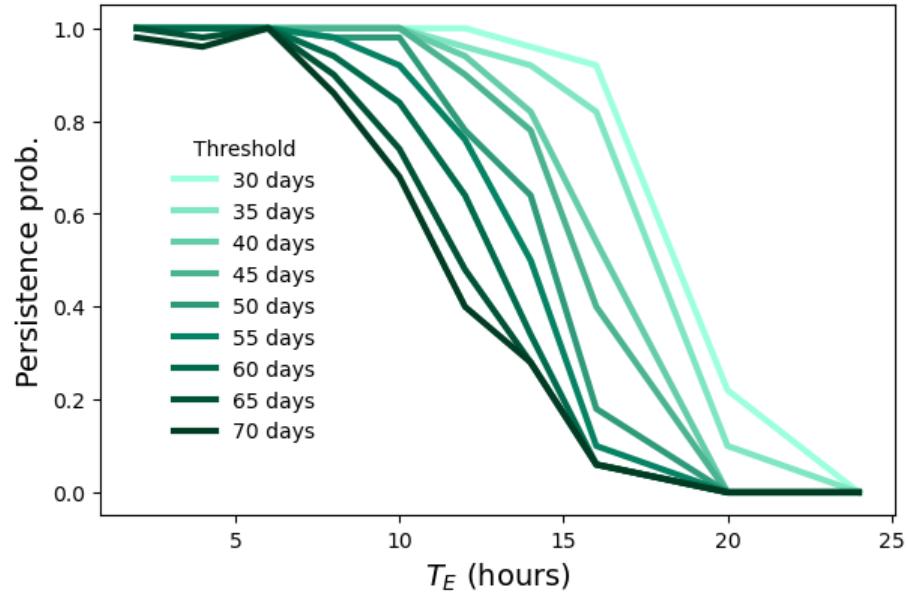
Supplementary Figure 2: **Additional middlemen statistics from simulations.** (A) Proportion of upazilas visited daily by one middleman during a single simulation. Note that a middleman may visit up to 4 upazilas per day, but visiting one or two is usually sufficient to complete a cargo. (B) Distributions of daily numbers of farms visited by one middlemen (green) and middlemen visiting one farm (red). (C) Distribution of numbers of vendors trading daily with a middleman. (D) Cumulative distribution of sizes of transactions involving middlemen and vendors (solid line). The dotted line represents the cumulative proportion of chickens sold in transactions up to a given size. Note that these transactions are typically smaller than those between farms and middlemen since vendors deal with smaller amounts of chickens than other PDN actors. Results are obtained from a single simulation with default settings as in Fig. 2 in the main manuscript.



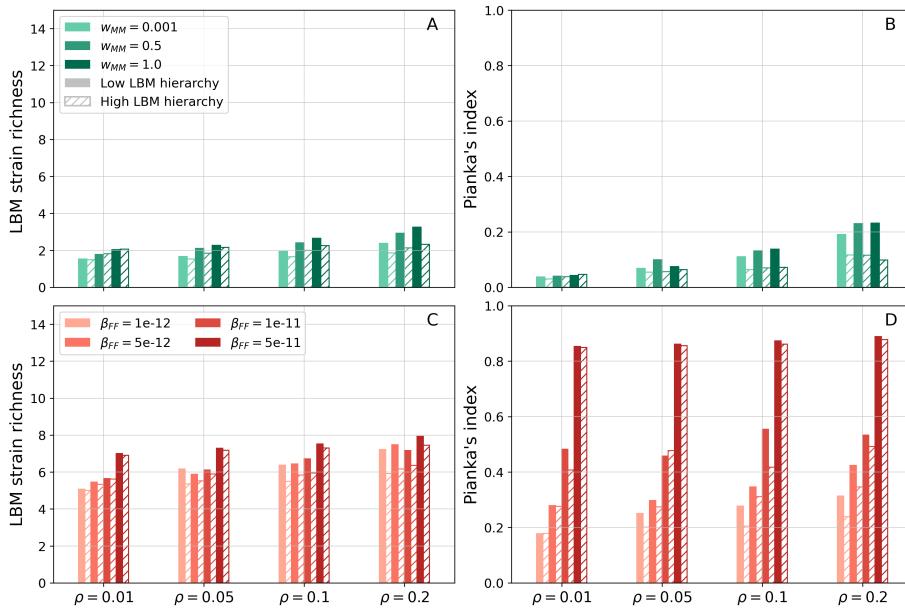
Supplementary Figure 3: **Additional vendor statistics from simulations.** (A) Distribution of numbers of wholesalers supplying a retailer (yellow), another wholesaler (blue) or any vendor (red) on a daily basis. (B) Distribution of numbers of retailers (yellow), wholesalers (blue) or vendors (red), regardless of type, purchasing from a single wholesaler on a daily basis. Note that (A) excludes vendors buying chickens from middlemen, i.e. vendors operating in the first LBM tier. (C) Distributions of daily amounts of chickens bought from retailers (yellow) and wholesalers (blue) in simulations (lines) and data (markers) [2]. (D) Cumulative distribution of sizes of transactions involving middlemen and vendors (solid line). The dotted line represents the cumulative proportion of chickens sold in transactions up to a given size. Results are obtained from a single simulation with default settings as in Fig. 2 in the main manuscript.



Supplementary Figure 4: **Distribution of marketing times.** Each panel shows distributions of marketing times for different average proportions of unsold chickens  $\rho_{unsold}$  and for increasing probability  $p_{empty}$  of a vendor selling all chickens in a single day (from left to right). The marketing time is defined as the time interval elapsed since a chicken enters any LBM for the first time and is sold to an end-point customer. Simulation settings are the same as in Fig. 4 with only 10% of vendors prioritizing the sale of unsold chickens.



Supplementary Figure 5: **Sensitivity of persistence probability to duration of transmission chains.** Lines show how the probability of pathogen persistence varies with both  $T_E$  and the minimum duration to determine whether a transmission chain is persistent or not. The estimation of the probability of persistence as well as simulation settings are the same as in Fig. 6I.



Supplementary Figure 6: **Viral mixing under complete cross-immunity.** Results mirror panels B,C,E,F from Fig. 7 in the main manuscript, under the assumption of complete cross-immunity ( $\sigma = 0$ ). Increasing cross-immunity lowers strain richness in any setting as individual strains face increased competition. Nonetheless, increasing cross-immunity does not significantly affect overlap between LBMs.