1. The intra-herd model: equations

The model for dynamics of *Coxiella burnetii* in a dairy cattle herd used in this study is a variant of the model introduced by Courcoul et al. in 2011 [10].

Here we present the Equations 1-6, which describe the updating (between time steps (*t-1*) and *t*) of variables corresponding to health states of cows in a herd *i*:

(1)

(2)

(3)

(4)

(5)

(6)

Based on Equations 1-6, it is possible to define two main outputs of the model at the herd level as:

(7)

(8)

Equations 9-16 here below define ingoing and outgoing flows in Equations 1-6, with parameters defined in Table 2 (main text):

*, where* (9)

(10)

(11)

(12)

(13)

(14)

(15)

(16)

where*,* and

is the probability of selling a cow in the lactation year *l* and with specific health status to herd *j.* The lactation year (*l)*, number of purchases made by herd *I* from *j (*, and the number of cows sold to herd *j* by herd *i* () were based on the data.

The dynamics of bacterial load in the environment is given by the following equation (17, identical to Equation 3 in the main text):

(17)

The overall dynamics of the environmental bacterial load is governed by animals shedding through different routes at different stages of their reproductive cycle and at different levels of shedding. According to [10], for herd *i* this hence can be summarized in the following equation:

, (18)

where, *origin* are the different health states of cows which can shed the bacteria, is the state of reproductive cycle of the cow, , is the route by which the quantities *Qtyl* of bacteria are shed, is the level of bacterial shedding; with the number of animals in corresponding health state at time *t* and the probability distributions governing shedding levels. The remaining parameters are defined in Tables 2 and 3 of the main text.

1. **Airborne dispersion of infectious particles in herd neighbourhood and incorporation of meteorological data**

The wind speed and direction data from the European Centre for Medium Range Weather Forecasts database were procured for the entire Finistère department (Western France). Data consisted of daily values of northward wind component and eastward wind component. Using this data, the wind speed was estimated as , and direction of the wind flow was estimated through its angle ɸ with the original x-axis, where . Weekly averages of wind direction and speed (a unique value for the whole area considered, here the Finistère department) were used in the Gaussian dispersion model. Adjustment of the frame of the receiving *j* and source *i* herd coordinates according to the direction of wind flow was done based on the distance between the two herds (distanceij), the angle between the line linking the two herds and the x-axis (angleij) and angle ɸ, as described in Stockie [27]: and .

The source for generation of the plume leaving the source herd *i* was the bacteria being lost from the environment compartment of this herd (rate of bacterial removal from the environment due to various mechanisms estimated for the intra-herd dynamics by Courcoul et al. [23]). Here, we assume that a part of bacteria shed by cows through different routes form dust particles, which remain infectious and become source for generation of plume. This plume then transported by the wind to another herd. Indeed, the smaller droplets generated by sneezing, coughing, splashing and other activities remain suspended in the air and dry fast enough to produce smaller particles called droplet nuclei, which can remain suspended in the air for long duration and can be transported along with the wind to distant places, unlike larger particles. Hence, the inherent capacity of airborne transmission of any infectious agent depends on production of appropriate range of droplet particle sizes with viable pathogens [40]. Multiple studies have suggested a higher risk of airborne transmission of Q fever within the radius of 5 km from the source in moderate environmental conditions [41, 35]. Hence, we restricted our dispersion model to a radius of 5 km from the source herd.

1. **Estimation of the optimum cut-off value for probability of infection (PI) for incident herds**

The cut-off PI was used to classify the herds into two categories: herds with a PI larger than the cut-off were considered positive (infected), the others were negative (uninfected). This categorisation concerns the simulated herd status at one year after the onset of pathogen spread into the metapopulation. The optimum cut-off is chosen based on comparison of simulation to data concerning herd status at time zero (of the follow-up) and at one year. The optimum cut-off value is chosen based on the epidemiological situation of the case concerned, such as prevalence in the population and consequences of false positive and false negative results. In the literature, prevalence dependent (Sensitivity (*Se*), Specificity (*Sp*), Youden index (*J*), odds ratio etc) and independent criteria (Efficiency, kappa) are both used to come up to a decision [29]. Here, three criteria were used to determine the optimal cut-off value: , maximum Youden index (Jmax, where ) and maximum accuracy (Accmax, where Acc is equal to the proportion of true negative and true positive into the population). Values of these criteria along with PI cut-off values are provided in Additional file 2 for both herd level and neighbourhood (3 km) based on simulations (description of the simulation settings in the main text).

Predicted incident herds for different values of cut-off at one year after the onset of pathogen spread into the metapopulation, specifying the contamination route, are provided in Additional file 3.