Original article

A deterministic and stochastic simulation model for intra-herd paratuberculosis transmission

Régis POUILLOT[a], Barbara DUFOUR[a]*, Benoît DURAND[b]

[a] Agence Française de Sécurité Sanitaire des Aliments, DERNS, 27–31 avenue du Général Leclerc, BP 19, 94701 Maisons-Alfort Cedex, France
[b] Agence Française de Sécurité Sanitaire des Aliments, LERPAZ, 22 rue Pierre Curie, BP 67, 94703 Maisons-Alfort Cedex, France

Abstract – In France, the considerable economic losses due to paratuberculosis introduction in cattle herds may justify the development of a herd-level certification procedure. For a cost/benefit analysis purpose, a simple deterministic and stochastic simulation model for intra-herd paratuberculosis transmission has been developed to evaluate the economic consequences of the purchase of a single infected heifer in a French average herd. The values of the epidemiological parameters were provided by a panel of French paratuberculosis experts. The results were in adequacy with field observations. A sensitivity analysis was conducted. The model was however difficult to validate rigorously, since few data on the intra-herd paratuberculosis true prevalence level are currently available.

paratuberculosis / stochastic model / deterministic model / sensitivity analysis

1. INTRODUCTION

Johne's Disease, or paratuberculosis, is a chronic wasting disease that causes considerable production losses in adult cattle, sheep and goats. This infection is of great economic importance for producers. Averaged across all herds, the cost of Johne's disease in the US dairy industry was estimated at US$ 200 to US$ 250 million per year [15].

Paratuberculosis is due to *Mycobacterium avium* subsp. *paratuberculosis* infection. Clinical symptoms are dominated by a chronic inflammation of the intestines, resulting in lower milk production, emaci-ation, diarrhea and death. For unknown reasons, young animals are more susceptible to *M. paratuberculosis* than adults, while experimental challenge studies have shown a possible infection of adult cattle. Following infection, a prolonged incubation phase is observed before clinical signs appear. Excretion of *M. paratuberculosis* in the faeces from infected cattle may begin long before the development of the first clinical signs, but the dissemination increases with the progression of the infection. Dissemination of the bacteria to susceptible animals is horizontal through faeces, but also vertical, mainly through the milk route [20]. In the absence of clinical signs,

---

* Corresponding author: bdufour@vet-alfort.fr

paratuberculosis diagnosis is difficult due to the poor positive predictive value of indirect diagnostic tests [7, 13], while the confirmation of a clinical diagnosis is much more reliable.

Paratuberculosis has received increased attention in France in the last years, and a recent study has revealed a cattle herd seroprevalence rate ranging from 0.02% to 4.57% per departement (French administrative subdivision) [17]. In this context, the French association for animal health certification (ACERSA: "Association pour la certification en santé animale") has been working on a certification scheme that would allow French breeders to buy animals from herds certified as being paratuberculosis-free. A working group of French paratuberculosis experts has been created (see members list in the appendix). The objectives of this group were (i) to propose paratuberculosis certification schemes, and (ii) to evaluate the economic interest of these schemes, using a cost/benefit approach.

In order to evaluate the costs and the benefits induced by a certification scheme, it was necessary to evaluate: (i) the cost of the certification, (ii) the reduction of the risk to introduce an infected animal in a herd when the certification scheme is applied, and (iii) the cost of such an introduction, if the certification scheme fails or if it is not applied.

To evaluate the latter cost, a global approach of the dynamics of infection in a herd was necessary through the use of modelling. The objective of this paper was to present a simple model of the paratuberculosis intra-herd dynamics following the purchase of a single infected heifer. This model was developed in collaboration with the ACERSA in order to compute the epidemiological figures that are necessary to evaluate the paratuberculosis cost in a herd.

2. MATERIALS AND METHODS

2.1. Material

The model is based upon a state-transition model developed from a Markov-Chain where the unit of concern is an average composite herd. Time is considered as discrete, using a time-step of one year. We first describe the overall skeleton of the model and the general assumptions that have been made. Then we show how this skeleton is completed to produce on the one hand a deterministic, continuous animal model; and on the other hand a stochastic, discrete animal model.

2.1.1. Assumptions

The objective of this work was to develop a simple model. The individual-level natural history of the disease was thus simplified in order to catch the main influential events on paratuberculosis development according to the panel of experts. The following assumptions were made.

(i) The susceptibility of ≥ 1 year animals is negligible, i.e. animals may only be contaminated during their first year of life.

(ii) Infected heifers are first incubating and begin to shed bacteria in the environment after their first calving. Therefore, the only source of contamination is the infected cows (i.e. females after the first calving). Amongst them, two categories were distinguished: (a) some cows become sick: these animals shed large amounts of bacteria in the environment; (b) the other infected cows remain sub-clinical: these animals are moderately infectious.

(iii) It is well documented that paratuberculosis transmission is highly connected to the confinement of the animals. As a consequence, the transmission will be much more important if the disease onset occurs during the calving months than if it occurs later, when the animals are kept on pastures. The former period (during which animals are kept inside) will be denoted below the "post-partum period", whereas the latter one (during which animals are outside, on pastures) will be denoted the "milking period".

Some additional assumptions were made about the herd management.

(iv) Every year, the breeder keeps the same number of female calves for renewal (the other calves being sold within a 1 year time step). This number is computed to compensate for, in the absence of paratuberculosis, the outflow of culled cows (for other reasons than paratuberculosis) by the herd born and bred heifers that calve for the first time.

(v) Animals culled because of paratuberculosis are replaced by healthy purchased heifers, introduced in the herd when they are 1 year-old (i.e. when they can not be contaminated anymore, according to the assumption (i)).

(vi) The paratuberculosis clinical signs are recognised: the corresponding animals and their calves are systematically culled within the year of the disease onset.

(vii) The cows are also culled for other reasons than paratuberculosis, according to an age-independent culling rate (cows reaching a maximal age being, however, systematically culled). This culling rate is however time-dependant: it is adjusted every year by the breeder, in order to keep the number of cows equal to a constant value (and therefore to maintain a constant production level). The animals are culled at the end of each time step.

2.1.2. General structure

A corollary of the assumption (ii) is that the infected calves and heifers do not play any role in the contagion process, since animals do not become infectious before their first calving. Therefore, among calves and heifers, only those that are kept until their first calving had to be taken into account in the model. The modelled population was then the set of the cows, and the set of calves and heifers that will become cows later. The terms "renewal calves/heifers" will be used below to refer to the latter set of animals.

The animals were represented by pairs of states: their age (from 1 to A, the maximal age of a cow), and their health state:

– Non infected (S): these animals are not infected. Note that the symbol S was chosen for "susceptible" to be in accordance with classical transmission models [2], but that only < 1 year-old animals in state S are really susceptible (according to the assumption (i)).

– Sub-clinically infected (I): these animals are infected, but do not show clinical signs (they are thus culled with the same culling rate as the non infected animals) and are not infectious until the first calving (assumption (ii)). Afterwards they are moderately infectious (assumption (ii.b))

– Clinically infected (D): these animals are infected, show clinical signs of paratuberculosis and are highly infectious (assumption (ii.a)). These animals and their calves are systematically culled within a one year time step (assumption (vi)).

A herd was then modelled by a matrix H of state variables that has A rows and 3 columns; and $H_t(a, h)$ denotes the number of a year-old animals which are in the health state h ($h = S$, I or D) at the time step t. The intra-herd paratuberculosis dynamic is described by the transitions between pairs of states, through difference equations.

Assuming that only < 1 year-old animals may be contaminated led to describe the disease general dynamics by two pairs of equations, the first one for the ≥ 1 year-old animals, and the second one for the < 1 year-old animals.

Since ≥ 1 year-old animals are not susceptible anymore, the dynamics (Eqs. (1a) and (1b)) simply resulted from culling, disease onset and ageing:

$$H_{t+1}(a+1, S) = H_t(a, S) - C_t(a, S) \quad (1a)$$

$$H_{t+1}(a+1, I) = H_t(a, I) - H_t(a, D)$$
$$- C_t(a, I) \quad (1b)$$

where:

– $C_t(a, S)$ denotes the number of a year-old animals which were culled during the time

step t, while they were in the non infected health state (S state);

– $C_t(a, I)$ denotes the number of a year-old animals which were culled (for other reasons than paratuberculosis) during the time step t, while they were in the sub-clinically infected health state (I state).

– $H_t(a, D)$ denotes the number of a year-old sub-clinically infected animals (I state) which became clinically infected (D state) during the time step t (and were culled because of paratuberculosis during this time step).

These quantities ($C_t(a,\{S, I\})$ and $H_t(a, D)$) do not have the same definitions in the deterministic and in the stochastic models: they are given in the corresponding sections.

For the < 1 year-old animals (Eqs. (2a) and (2b)) the dynamics represents both the contamination of the calves born in the herd, and the cow renewal process. Assuming that every year, the breeder keeps the same number F of renewal calves, and that the animals culled because of paratuberculosis are replaced by healthy, unsusceptible purchased heifers (assumption (v)), the number of < 1 year-old non infected (S) calves was F plus the bought heifers (i.e. the sum of clinically infected animals which are culled because of paratuberculosis during the time step) and minus the number of contaminated < 1 year-old renewal calves (Eq. (2a)).

$$H_{t+1}(1, S) = F - I_t + \sum_{1 \le a \le A} H_t(a, D) \quad (2a)$$

$$H_{t+1}(1, I) = I_t. \quad (2b)$$

The number of contaminated renewal calves (I_t in Eqs. (2a) and (2b)) does not have the same definition in the deterministic and in the stochastic models and is given in the corresponding sections. However, in both cases, I_t was modelled as being proportional to a contamination probability at the time step t: β_t. Since the clinically infected cows and their calves were assumed to be culled within the time step of the disease onset, the vertical transmission route from clinically infected cows was not taken into account. Using a Reed-Frost approach, the contamination probability (Eq. (3)) was then defined as being equal to one minus the joint probability to escape the three contamination sources: (i) the animals that become clinically infected during the post-partum period (horizontal transmission), (ii) the animals that become clinically infected during the milking period (horizontal transmission), and (iii) the sub-clinically infected animals that do not become clinically infected during the time step (horizontal and vertical transmission). Each of these contamination sources was associated with a different transmission parameter (respectively τ_X, τ_Y and τ_Z), the parameterisation of which should satisfy the following relations: $\tau_X \ge \tau_Y \ge \tau_Z$.

$$\beta_t = 1 - (1 - \tau_X)^{X_t} (1 - \tau_Y)^{Y_t} (1 - \tau_Z)^{Z_t} \quad (3)$$

where, at each time step t:

– X_t represents the number of clinically infected cows (D state) for which the clinical signs appeared during the post-partum period; its definition varies in the deterministic and in the stochastic models and is given in the corresponding sections;

– Y_t represents the number of clinically infected cows (D state) for which the clinical signs appeared during the milking period, i.e. the total number of clinically infected cows (of V years or older, where V is the age at the first calving) minus X_t:

$$Y_t = \left[\sum_{a = V}^{A} H_t(a, D) \right] - X_t;$$

– Z_t represents the number of sub-clinically infected cows (I state) that did not become clinically infected (D state) during the time step:

$$Z_t = \sum_{a = V}^{A} [H_t(a, I) - H_t(a, D)].$$

2.1.3. The deterministic model

Four definitions are missing to complete the general model structure, namely, for each time step t: (i) the number of contaminated calves (I_t), (ii) the number of cows culled for other reasons than paratuberculosis (C_t), (iii) the number of sub-clinically infected cows that become clinically infected and are culled because of paratuberculosis ($H_t(a, D)$), and (iv) the number of the latter cows for which the clinical signs appear during the post-partum period (X_t). A first way to complete this structure led to a deterministic, continuous animal model.

In such a mean-field model, the number of contaminated calves during the time step t is simply the product of the contamination probability by the number of renewal calves:

$$I_t = \beta_t F. \qquad (4)$$

Because only the heifers that are kept until their first calving are taken into account in the model, a fictitious zero culling rate was applied before the first calving. For cows, the number of culled animals (for other reasons than paratuberculosis) was equal to an age-independent and time-dependent culling rate (ρ_t), times the number of the animals (except for the cows reaching the maximal age A, which were systematically culled).

$$C_t(a, x) = 0 \qquad\quad a < V, x \in \{ S, I \}$$
$$C_t(a, x) = \rho_t H_t(a, x) \; V \le a < A, x \in \{S, I\}$$
$$C_t(a, x) = H_t(a, x) \qquad a = A, x \in \{S, I\}.$$
$$(5)$$

The cows culling rate is, by definition, the number of culled cows per year over the number of cows younger than the maximal age A. To keep the number of cows equal to a constant value of N, the number of cows that have to be culled was equal to the number of supernumerary cows during the

next time step if only A year-old cows were culled, i.e. the number of cows less than A year-old, plus the number of heifers which will calve for the first time during the next time step (the animals of V–1 year-old), minus the desired number of cows (N).

$$\rho_t = \left[\sum_{V-1 \le a < A, h \in \{ S, I \}} H_t(a, h) - N \right]$$
$$\times \frac{1}{\displaystyle\sum_{V \le a < A, h \in \{ S, I \}} H_t(a, h)}.$$

The number of clinically infected cows (Eq. (6)) was modelled by the product of the number of sub-clinically infected animals by an age-dependent probability of disease onset for these animals (α_a).

$$H_t(a, D) = \alpha_a H_t(a, I). \qquad (6)$$

Finally, the number of cows which become clinically infected during the post-partum period (Eq. (7)) was defined as being a fixed proportion of the number of clinically infected animals.

$$X_t = \sum_{V \le a \le A} \delta H_t(a, D). \qquad (7)$$

2.1.4. The stochastic model

A second way to complete the general model structure led to a stochastic, discrete animal model. In this case, the four missing quantities are considered to be random variables. Defining these quantities consisted of specifying the probability distribution that these random variables follow.

The number of infected calves followed a binomial distribution B(n; p), n being the number of renewal calves, and p the contamination probability at the time step (Eq. (8))

$$H_t(a, D) \sim \mathrm{B}(F; \beta_t). \qquad (8)$$

As for the deterministic model, a fictitious zero culling rate was applied before the first calving, and cows reaching the maximal age A were systematically culled. For younger cows, the number of culled animals for other reasons than paratuberculosis

Table I. Herd management parameter values for a mean French beef herd and for a mean French dairy herd, estimated from agricultural statistics and from expert opinions.

Parameter	Description	Value	
		Dairy	Beef
V	Mean age at the first calving (years)	2	3
N	Mean number of cows (females of V years or older)	40	50
E	Mean number of calves per cow over the lifetime	3	5
A	Maximal age of cows (at which they are systematically culled)	15	15

(Eq. (9)), followed a generalised hypergeometric distribution GH($M; x$), which has the following semantic: choose x units among M without replacement, M and the result of the random process being matrices of the same dimensions. In this case, x is the number of animals which have to be culled to keep the number of cows constant, as defined in the deterministic model; and M is the set of animals which may be culled.

$$C_t \sim \text{GH}(\{ H_t(a, h), V \le a < A, e \in \{ S, I \} \};$$

$$\sum_{v-1 \le a < A, h \in \{ S, I \}} H_t(a, h) - N\Big). \quad (9)$$

The number of clinically infected animals (Eq. (10)) followed a binomial distribution B($n; p$), n being the number of subclinically infected animals, and p the age-dependent probability of disease onset (α_a).

$$H_t(a, D) \sim \text{B}(H_t(a, I); \alpha_a. \quad (10)$$

Finally, the number of animals which become clinically infected during the postpartum period (Eq. (11)) was a sum of independent random variables, one for each age class. Each of these variables followed a binomial distribution, the parameters of which were the number of clinically infected animals and the probability, for the disease onset, to occur during the postpartum period (δ).

$$X_t = \sum_{V \le a \le A} X_t(a)$$

with

$$X_t(a) \sim \text{B}(H_t(a, D); \delta). \quad (11)$$

2.2. Methods

2.2.1. Parameterisation

Due to their differences in herd management and size, two kinds of herds were considered: beef herds and dairy herds. For both herd types, the maximal age of cows (A) was set to an arbitrary value of 15 years. The values of the other herd management parameters for a mean beef herd and for a mean dairy herd were estimated using French agricultural statistics data (Tab. I).

In paratuberculosis-free herds, the cows culling rate was assumed to be age-independent (except for A year-old cows which were systematically culled). Taking into account only the heifers kept for renewal that will become cows later in their life, the number of these "renewal heifers" kept each year by the breeder (F) can be derived from the number of cows in a mean herd (N estimated at 50 for beef herds and at 40 for dairy herds), and from the mean number of calves per cow over total lifetime (E), estimated at 5 for beef herds and at 3 for dairy herds. This yielded a value of 10.7 heifers for a mean dairy herd and of 13.4 heifers for a mean beef herd (note that these values would correspond to a mean yearly cows culling rate of 0.20 in disease-free beef herds and of 0.33 in disease-free dairy herds).

The values of the epidemiological parameters (Tab. II) were estimated using expert opinions using a classical expert elicitation method [22]. For measurable parameters (the age-dependent disease onset probability α_a and the probability that the disease onset occurs during the post-partum

Table II. Epidemiological parameter values, estimated using expert opinions.

Parameter	Description	Value
α_a	Age-dependent probability of disease onset for sub-clinically infected animals:	
	$a < V$	0.00
	$a = V$	0.40
	$a = V + 1$	0.40
	$a > V + 1$	0.00
δ	Proportion of the clinically infected animals for which the disease onset occurs during the post-partum period	0.90
τ_X	Disease transmission probability during the post-partum period, from a clinically infected cow to a susceptible calf	0.25
τ_Y	Disease transmission probability during the milking period, from a clinically infected cow to a susceptible calf	0.05
τ_Z	Yearly disease transmission probability, from a sub-clinically infected cow to a susceptible calf	0.05

period δ), experts were asked to propose initial point estimates, which were then compared and discussed until a consensus was obtained. For non-measurable parameters (the three transmission parameters τ_X, τ_Y and τ_Z), the elicitation process started in the same way, with experts being asked to propose initial estimates. However, instead of simply comparing and discussing these values, experts were then asked to validate (or reject) the disease dynamics obtained from these initial estimates: the deterministic and the stochastic models were run using these values and the resulting paratuberculosis dynamics in both herd types were presented to the experts panel. This allowed them to refine their initial estimates. The process continued until a consensus was obtained, with the predicted dynamics being considered as satisfactory by the experts. Note that the estimated parameters appeared to be identical for beef herds and for dairy herds.

2.2.2. Model exploitation

The model was exploited through simulations, each simulation lasting 30 simulated years, and starting with the introduction in the herd of a single sub-clinically infected 1 year-old heifer. Two simulation programs were developed using R [9]: one for the deterministic model and the other for the stochastic model (source code available upon request). The stochastic model was exploited by Monte-Carlo simulation, 10 000 simulations being run for each set of parameter values.

The predicted dynamics of the number of clinically infected and sub-clinically infected cows were first studied using the deterministic model, and these results were compared with the mean values obtained using the stochastic model. The latter model was also used to obtain the predicted cumulative disease extinction probability over time, and the evolution of the incidence distribution through the predicted time variations of the quartiles of this distribution.

The effect of the herd size and of the culling rate were then studied, using the deterministic model for a dairy herd.

Finally, a sensitivity analysis was conducted [22], running the deterministic model from 10 000 independent sets of randomly chosen parameter values. Again, the number of heifers kept every year by the breeder (F) was not directly randomly chosen, but was computed for each simulation run from the randomly chosen number of cows (N), and from the randomly chosen number of calves per cow (E). The random values of N and of E (and of the other

Table III. Sensitivity analysis: parameterisation and results.

Parameter	Description	Distribution	Spearman rank correlation
V	Mean age at the first calving (years)	$P(V = 2) = 2/3$[a] $P(V = 3) = 1/3$	–0.29
N	Mean number of cows (females of V years or older)	$N_{RT}(40; 20; 1)$[b]	0.55
E	Mean number of calves per cow over the lifetime	$P(E = 2) = 1/3$ $P(E = 3) = 1/3$ $P(E = 4) = 1/3$	–0.37
α_V	Probability of disease onset at the first calving	$U(0.3; 0.5)$[c]	0.16
α_{V+1}	Probability of disease onset at the second calving	$U(0.3; 0.5)$[b]	0.10
δ	Proportion of the clinically infected animals for which the disease onset occurs during the post-partum period	$U(0.8; 1)$[c]	0.09
τ_X	Disease transmission probability during the post-partum period, from a clinically infected cow to a susceptible calf	$U(0.1; 0.4)$[c]	0.53
τ_Y	Disease transmission probability during the milking period, from a clinically infected cow to a susceptible calf	$U(0; 0.1)$[b]	0.02
τ_Z	Yearly disease transmission probability, from a sub-clinically infected cow to a susceptible calf	$U(0; 0.1)$[c]	0.20

[a] $P(X = x)$: probability that parameter X equals x. [b] N_{RT} (mean; standard deviation; minimum): rounded normal truncated distribution. [c] U(minimum; maximum): uniform distribution.

parameters) were chosen according to likely distributions (Tab. III) reflecting the variability of the parameters and/or the uncertainty about them: many of these distributions were uniform distributions around the point estimates provided by the expert group. The Spearman rank correlation coefficients were estimated independently between each parameter and a single output parameter: the incidence rate (i.e. the proportion of clinically infected cows) observed 10 years after the introduction of the infected heifer.

3. RESULTS

3.1. Disease dynamics

Using the deterministic model, the yearly disease incidence (i.e. the yearly number of clinically infected cows, culled because of paratuberculosis) was predicted to rise slowly until it reaches a stable value of 8.4 cases in a mean dairy herd, and 7.1 cases in a mean beef herd (Fig. 1A). These figures correspond to a yearly incidence rate (i.e. the number of clinically infected cows over the number of cows) of 21.1% in a mean dairy herd, and 14.2% in a mean beef herd. The delay to reach the final value was about 15 years for the dairy herd; it was approximately 10 years longer for the beef herd.

The predicted dynamics of the yearly number of sub-clinically infected animals (Fig. 1B) had similar shapes. The stable value was 12.6 cases in the mean dairy herd (31.5% of the cows) and 13.0 cases in the mean beef herd (25.9% of the cows).

Using the same set of parameter values, each run of the stochastic model gave a different dynamics. Therefore, this model can produce a wide range of behaviour, as shown in Figure 2. The disease may progress
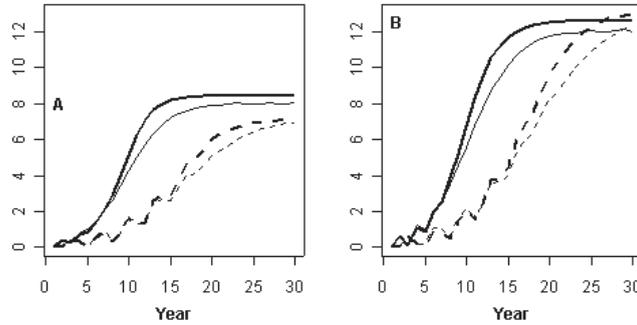
Figure 1. Predicted dynamics of the number of clinical (A) and sub-clinical (B) paratuberculosis cases in a mean French dairy herd (solid lines) and in a mean French beef herd (dashed lines), obtained from the deterministic model (bold lines) and from the stochastic model (thin lines, 10 000 runs mean values).
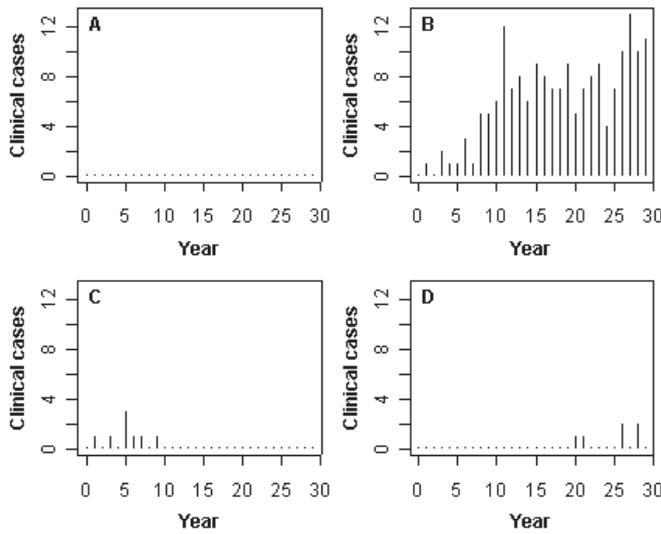
Figure 2. Predicted number of clinical cases in four examples of simulation runs of the stochastic model in a mean French dairy herd: extinction without clinical cases (A), rapid progression (B), late extinction (C), prolonged sub-clinical propagation (D).

rapidly inside the herd (Fig. 2B), the first clinical cases being seen soon after the introduction of the infected heifer. Conversely, the introduced infected heifer may remain a sub-clinical case and be culled before having contaminated any other animal (Fig. 2A): in this case, the disease disappears (i.e.

there are no more infected animals) without having caused any clinical cases. Between these two extreme behaviours, some intermediate situations may be encountered (but at a low rate). On the one hand, the disease may cause several clinical cases before it disappears: Figure 2C shows a simulation
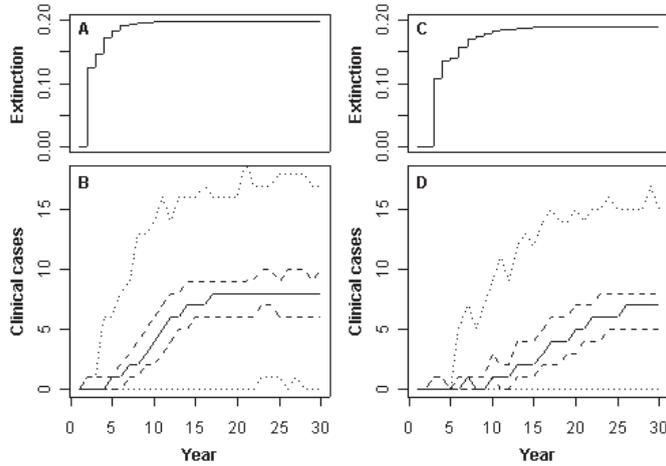
Figure 3. Predicted evolution of the cumulated yearly disease extinction probability (A, C) and of the extreme values (dotted lines), 25th and 75th percentiles (dashed lines), and median (solid lines) of the yearly incidence, in a mean French dairy herd (A, B) and in a mean French beef herd (C, D) (results obtained from 10 000 runs of the stochastic model).

run during which 8 clinical cases are observed before the disease disappears, 12 years after the introduction of the infected heifer. On the other hand, the disease may evolve only sub-clinically during a prolonged period, before the appearance of the first clinical cases: Figure 2D shows an extreme case of this kind of dynamic, with a simulation run during which the disease evolves silently for 21 years before the first clinical case is observed.

Besides these examples, some yearly statistics can be computed to catch the distribution of the possible behaviours, considering every simulated year the runs from herds where the disease has not disappeared (i.e. there are still infected animals). Figure 1 shows the mean yearly number of clinically infected and sub-clinically infected cows, obtained from 10 000 runs: the curves obtained in this way were always slightly under those obtained with the deterministic model. The final yearly mean of the number of clinical cases was 8.1 cases for the dairy herd (20.2% of the cows) and 6.9 cases for the beef herd (13.8% of the

cows). Concerning the yearly mean of the number of sub-clinically infected cows, the final figures were 12.0 cases in the dairy herd (30.1% of the cows) and 12.3 cases in the beef herd (24.6% of the cows).

Figure 3 provides an insight of the shapes of the yearly distributions of the number of clinically infected cows through the time evolution of their extreme values (minimum and maximum), of their 25th and 75th percentiles, and of their medians. The yearly minimal values were found to be zero most of the time: for almost every year and in both herd types, there were simulation runs in which no clinical case occurred, despite the fact that there were still infected animals. Furthermore, the 25th percentile remained equal to zero during the 6 first years for the mean dairy herd (the first 9 years for the mean beef herd): there were thus many cases in which the disease remained silent during a prolonged period after the introduction of the infected heifer. If the final value of the median (8 cases for the dairy herd and 7 cases for the beef herd) was similar to the deterministic result (shown in Fig. 1), the median
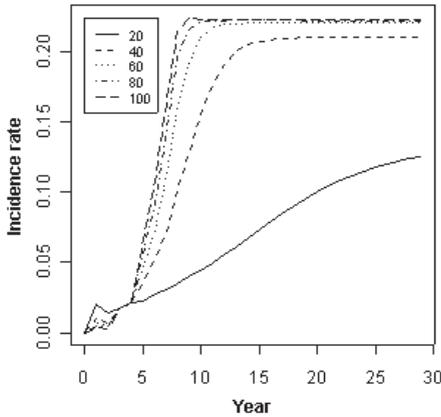
Figure 4. Predicted evolution of the incidence rate in a mean dairy herd according to the number of cows, obtained from the deterministic model.
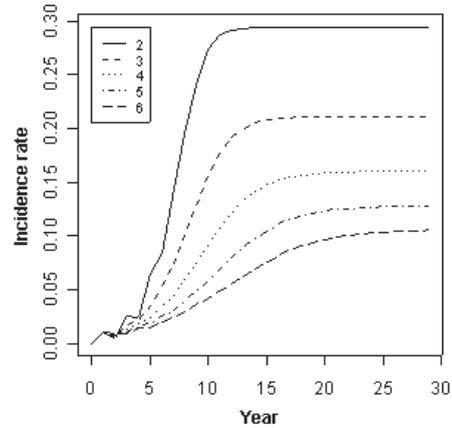
Figure 5. Predicted evolution of the incidence rate in a mean dairy herd according to the number of calves per cow, obtained from the deterministic model.

curves rose more slowly: before this stable situation was reached, the yearly distribution of the number of clinical cases was left-skewed, low incidence values being more frequent than high values.

Finally, the cumulated yearly extinction probability (i.e. for each year, the proportion of runs in which the disease disappears, with neither sub-clinically infected nor clinically infected animals anymore) is also shown in Figure 3. In both herd types, the curve rose rapidly and reached a stable value of approximately 20%. Thus, in one case out of five, the disease disappeared from the herd. The last extinction event (cull of the last infected animals) occurred earlier in the mean dairy herd (16th year) than in the mean beef herd (20th year).

3.2. Effect of the herd size and of the culling rate

The deterministic model was used to generate the predicted yearly incidence rate in a dairy herd, with herd sizes varying from 20 to 100 cows (Fig. 4). In more than 40 cow herds, the incidence rate rose rapidly and became stable whatever the herd size around 0.22, but the higher the herd size, the faster the incidence rate reached this stable value. In a 20 cow herd, the incidence rate rose very slowly, and the stable value of the incidence rate reached a lower value of 0.13.

The effect of the culling rate was also studied, using the deterministic model to generate the predicted yearly incidence rate in a dairy herd, with a mean number of calves per cow (over the lifetime) varying from 2 calves per cow (i.e. a 50% culling rate) to 6 calves per cow (i.e. a 16.7% culling rate) (Fig. 5). The simulation results showed that the higher the culling rate (i.e. the lower the mean number of calves per cow), the higher the equilibrium incidence rate was reached, and the faster it was reached.

It was then interesting to study the joint effect of an increase of the herd size and a decrease of the culling rate. The analysis of this joint effect on the incidence rate at the 30th time step (Fig. 6) showed that the effect of the number of calves per cow dominated the effect of the herd size since the slope of the surface was steeper along the y-axis than along the x-axis.
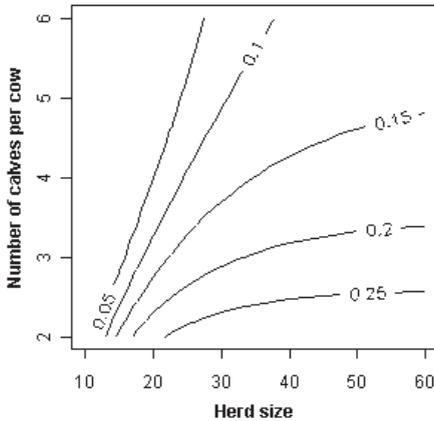
Figure 6. Predicted variations of the incidence rate at the 30th year according to the herd size and to the number of calves per cow, obtained from the deterministic model.

3.3. Sensitivity analysis

The sensitivity analysis using as an output variable the incidence rate at the 10th time step allowed to distinguish two kinds of parameters according to the sign of the Spearman rank correlation coefficient (Tab. III): those that are positively linked with the output variable, and those that are negatively linked with it.

Among the former, the higher correlation coefficients were obtained with the number of cows (0.55: this confirms the above herd size effect) and with the disease transmission probability during the post-partum period, from a clinically infected cow to a susceptible calf (0.53). The association was found to be weaker for the yearly disease transmission probability, from a sub-clinically infected cow to a susceptible calf (0.20), whereas the association was very low for the third disease transmission parameter, the disease transmission probability during the milking period, from a clinically infected cow to a susceptible calf (0.02).

Two parameters were found to be negatively linked with the output variable: the mean number of calves per cow (–0.37: this confirmed the above culling rate effect) and the age at the first calving (–0.29).

4. DISCUSSION

The model is a probabilistic model that uses both chain binomial and Markov-Chain principles. It partially uses the Reed-Frost model and assumptions to calculate the number of animals in each state at each time period. A major depart from the classical Reed-Frost model is the possibility for an animal to remain infectious during more than one time interval. While this kind of model is relatively easy to simulate, its analytic behaviour is difficult to explore particularly when the number of states is high: simulation is then the only way to explore the model.

The individual health states used in the model (non infected, sub-clinically infected, clinically infected) and the distinction between the sub-clinically infected animals (which shed moderate amounts of bacteria in the environment) and the clinically infected animals (heavy shedders), are in accordance with classical descriptions [12]. Within herd paratuberculosis transmission is complex. The transmission route from a cow to a susceptible calf are (i) foetal infections (vertical); (ii) infections around birth (vertical); (iii) infections due to colostrum contamination (vertical); (iv) infections due to milk contamination (vertical); and (v) infections due to faecal contamination (vertical and horizontal). Since we made the assumption that the clinical signs were recognised by the breeder and that clinically infected cows and their calves were systematically culled, we only had to consider the horizontal transmissions from all infected cows, and the vertical transmissions from sub-clinically infected cows. The model assumed a random mixing of infectious cows and susceptible calves, a general underlying assumption of Reed-Frost models. This assumption is not correct for the vertical transmissions from sub-clinically infected

cows. However, for simplicity purposes, this vertical transmission has been aggregated with the horizontal transmission from these animals, using a single transmission parameter. This simplification appeared to be globally acceptable by the ACERSA expert group.

According to experts and the literature, however, the infectious attack rate should be considered as different if the cow excretes *M. paratuberculosis* during the post partum period, while all animals are confined in the stable, and if the cow excretes *M. paratuberculosis* later, during the milking period, while the animals are less confined. Moreover, clinical symptoms of paratuberculosis and antibody levels are often exacerbated around the time of calving [5, 10]. These facts were taken into account in the model through the use of two different transmission parameters according to the delay between the calving and the appearance of clinical signs.

When modelling such a complex disease, the main difficulty is the parameterisation of the model. Defining parameters from expert opinion is a rather complicated objective. Classical expert elicitation methods have been used [22]. The confrontation of experts with stochastic model results using an initial set of parameter estimation was of great help to refine the estimates. It has to be noted that, in North America, it is commonly considered that the disease transmission is lower in beef herds than in dairy herds, since herd management is less intensive in the former case [21]. This was not taken into account by French experts, since beef herds are more intensively managed in France.

Collins and Morgan [6] developed a deterministic model for within herd paratuberculosis, using assumptions different from ours. The link between the animals confinement level and the disease transmission is not specifically taken into account in their model. The general contagion model is not the same as the one we used: briefly, the authors consider a transmission scheme based upon contact between a cow and a calf, with a fixed number of cow-calf contacts per time period. In this case, for a given number of infectious cows, the contamination probability of the calves decreases when the number of calves increases. Such a scheme is better adapted to model the vertical transmission of paratuberculosis than its horizontal transmission, in which the environment (and its contamination level) plays an important role: in Collins' and Morgan's work, the systematic culling of clinically infected animals is not taken into account, the vertical transmission is thus predominant. Conversely, in our model a consequence of this culling was that the horizontal transmission became the main transmission route. Therefore, and according to the expert opinions, we considered that the contamination probability was not influenced by the number of calves. Finally, Collins and Morgan simulate the possibility of purchasing other infected heifers than the first one. Since our final goal was to compute the economic losses due to the purchase of a single infected heifers, we did not take into account this opportunity which could be easily implemented.

A stochastic simulation model has also been implemented by the Farm Management Group of the Wageningen Agricultural University. Its application for evaluating several paratuberculosis control strategies is presented in [8]. This model is much more complete and complex than ours. It incorporates a large number of parameters which determine herd dynamics, production, disease spread, but also disease control strategies. This kind of model allows to control a wide variety of factors and to simulate a wide range of situations. A counterpart is that, as pointed out by Baldock and Cameron [3], the multiplicity of parameters makes this model complex to use and parameterise. For this reason, we chose to develop a simpler mathematical model, better adapted to our objectives.

Model validation using field data was difficult because *M. paratuberculosis* infected herds have been poorly monitored in

France. Moreover, the model was used to simulate the paratuberculosis dynamics without any control, while control interventions in affected herds are generally performed. When evaluating Johne's disease within-herd prevalence, problems arise because of the poor sensitivity of the available diagnostic tests, the low within herd prevalence of infection, and the clustering of false positives within a herd [11]. To date, few field surveys are available. These surveys can hardly be compared, due to the large discrepancies in the studied population (herd management, age), in the diagnostic test used and the epidemiological indexes analysed (herd, individual, withinherd prevalence, seroprevalence, or corrected prevalence). Little data exist on the "true" within-herd prevalence (i.e. corrected for test diagnostic performance) of paratuberculosis [4]. Moreover, corrections of the estimated prevalence from literature data are almost impossible since paratuberculosis test sensitivities are described as highly different according to the infection stage [18, 19]. Collins et al. [7] estimate this within-herd prevalence in Wisconsin (USA) to 20.3% on a sample survey using ELISA (sensitivity estimated to 0.509, specificity estimated to 0.949). In New-York State (USA), Obasanjo et al. [14] estimate this within-herd prevalence to 5.2% using faecal culture on whole herds. In Belgium, the median within-herd prevalence estimated by Boelaert et al. [4] is 7% (quartiles 4%–12%), based on an ELISA serological survey (sensitivity estimated to 0.45, specificity estimated to 0.99). Keeping in mind the described biases in withinherd prevalence estimation [4], the results of our models are in agreement with these relatively low estimates.

Despite what could be generally considered [6], deterministic models do not describe the average result obtained from an infinite number of stochastic simulations (see, for instance, [16]), and stochastic models result in lower epidemics than deterministic models. This observation, confirmed here (Fig. 1), is partly due to the possibility of rapid extinction of the epidemic, the basic example being when the introduced infected heifer does not infect any calf. Obviously, the development of a stochastic model is very helpful in comparing model behaviour and field observations since the stochastic model results reflect the epidemiological distribution of cases in the population. As an example, it is well recognised that, without any animal purchase, paratuberculosis may seem to disappear in a herd and recur some years later, a kind of behaviour often produced by the stochastic model.

As pointed out above, it is particularly difficult to do a detailed comparison of the predicted paratuberculosis prevalence change over time with current data from naturally or artificially infected herds. However, field observations seem to be qualitatively consistent with the behaviour produced by the stochastic model: paratuberculosis may be observed only through one or two clinical observations, vanishing without any control measure. In other cases, paratuberculosis may become endemic, with silent periods of two or three years occurring from time to time.

The sensitivity analysis of a model is of great interest. It allows to point out which parameters are the most influent, and thus, those for which estimate refinements should be performed or risk management should be focused. However, the sensitivity analysis conducted here evaluated the parameters influence independently and did not deal with probable interactions. For example, the disease transmission probability during the milking period, from a clinically infected cow to a susceptible calf (τ_Y) would probably be a more influent parameter if the proportion of clinically infected animals for which the disease onset occurs during the milking period $(1-\delta)$ was higher. Nevertheless, according to the literature, this method is the first step to explore global associations, excluding the interactions [22].

Two groups may be distinguished among the model parameters: the herd management

parameters and the epidemiological parameters. The distributions used in the sensitivity analysis for the parameters of the first group reflected herds variability. The results showed opposite effects of the herd size (positively linked with the incidence rate) and of the number of calves per cow (negatively linked with it). The herd size effect may be explained considering that, when the number of cows increases, the number of susceptible 1 year heifers increases as well, and more animals are thus infected. Because of the non-linearity of the contamination probability (Eq. (3)), such an increase accelerated the disease transmission. With more than 40 cows, the system was saturated at equilibrium, i.e. the contamination probability was very high; therefore the stable values of the incidence rate were close, whatever the herd size. Conversely, in a small herd, the system was not saturated at equilibrium (i.e. the contamination probability remained in medium values), and the corresponding incidence rate was lower. For the number of calves per cow (i.e. the culling rate), two opposite effects may be expected. On the one hand, a high number of calves per cow (i.e. a low culling rate) is necessarily associated with a low number of heifers born in the herd and kept every year. For the reasons detailed above, this should decrease the disease transmission. On the other hand, a high number of calves per cow means that the life expectancy of infected cows is high, which should increase the disease transmission. With the default set of parameter values (and probably because subclinically infected animals become clinically infected and shed heavily bacteria in the environment at their first or second calving only), the first effect dominated the second one.

Epidemiological studies suggest that the odds of seropositivity increase with herd size [10, 15]. Nevertheless, it is not possible according to the published data to investigate whether this observation is linked to a greater risk to introduce an infected animal when the herd is large, or if it is linked to a different within herd disease behaviour, or

both. A herd size effect is also found by Collins and Morgan [6], but at a lower level, because of their assumption of a constant effective cow-calf contact per year. It is well documented that dairy herds show higher prevalence rates than beef herds [21]. While the higher number of purchased (and possibly infected) animals in dairy herds may be a cause, our model explains this observation only on the basis of the differences between the herd management, with a higher size, but a lower replacement rate in beef herds.

The distributions used in the sensitivity analysis for the herd management parameters reflected herd-to-herd variability. In the same way, the epidemiological parameter distributions we used reflected a variability, but also an uncertainty about their point estimates [1]. The results suggested that the probability of transmission from a cow developing paratuberculosis during the post-partum phase has a major influence, compared to the other parameters. Experimental transmission could be conducted to estimate this probability more accurately in order to refine the estimates.

The simple model we presented here was built in order to allow economists to assess the costs and the benefits of paratuberculosis certification schemes in France. The overall structure of the model was in agreement with classical knowledge on the subject, and the results were in agreement with the available field knowledge. It could be adapted to other countries as long as the assumptions we made remain valid. This is particularly important for the hypothesis about the culling of the clinically infected cows and of their calves: this assumption makes the horizontal transmission predominant, and the formulation of the contamination probability is based upon such a predominance. It would be different in a context of vertical transmission predominance. Finally, using the model in other countries would imply to carefully adapt the parameter values to the local context: we are definitely convinced that this parameterisation is the

most sensitive and difficult aspect of paratuberculosis modelling.

APPENDIX

Members of the ACERSA panel of experts: Dr D. Repiquet (ACERSA), Dr G. Jubert and Dr T. Delcroix (FNGDSB), Dr P. Holleville, H. Petit, D. Guerin (GDS), Dr G. Gounot, Dr Y. Frigère, Dr F. Lars (field vets), Dr V. Carpinschi, Dr G. Manet, Dr C. Rose, Dr M. Lavoix, Dr F. Chaduc (veterinary laboratory), Dr F. Dion, J.N. Bonnet (association of sheep breeders), Dr Gregory (DGAl National veterinary officer), M.F. Thorel, Dr B. Garin-Bastuji, K. Laroucau (Afssa), Dr A. Drouart, Dr J. Vialard (veterinary schools).

REFERENCES

[1] Anderson E.L., Hattis D., When and how can you specify a probability distribution when you don't know much. Foundations: A. Uncertainty and Variability, Risk Anal. 19 (1999) 47–49.

[2] Bailey N.T.J., The mathematical theory of epidemics, Charles Griffin & Co Ltd, London, 1957.

[3] Baldock C., Cameron A., Evaluation of simulation models for Johne's disease, Animal Health Australia, [on line] (2002) http://www.aahc.com.au/bjd/simulationmodels.pdf [consulted 20 August 2002].

[4] Boelaert F., Walravens K., Biront P., Vermeersch J.P., Berkvens D., Godfroid J., Prevalence of paratuberculosis (Johne's disease) in the Belgian cattle population, Vet. Microbiol. 77 (2000) 269–281.

[5] Chiodini R.J., van Kruiningen H.J., Merkal K.L., Ruminant paratuberculosis (Johne's disease): the current status and future prospects, Cornell Vet. 74 (1984) 218–262.

[6] Collins M.T., Morgan I.R., Epidemiological model of paratuberculosis in dairy cattle, Prev. Vet. Med. 11 (1991) 131–146.

[7] Collins M.T., Sockett D.C., Goodger W.J., Conrad T.A., Thomas C.B., Carr D.J., Herd prevalence and geographic distribution of, and risk factors for, bovine paratuberculosis in Wisconsin, Am. J. Vet. Med. Assoc. 204 (1994) 636–641.

[8] Groenendaal H., Nielen M., Jalvingh A., Horst H., Galligan D.T., Hesselink J.W., A simulation of Johne's disease control, Prev. Vet. Med. 54 (2002) 225–245.

[9] Ihaka R., Gentleman R., R: a language for data analysis and graphics, J. Comput. Graph. Stat. 5 (1996) 299–314.

[10] Jakobsen M.B., Alban L., Nielsen S.S., A cross-sectional study of paratuberculosis in 1155 Danish dairy cows, Prev. Vet. Med. 46 (2000) 15–27.

[11] Jordan D., Aggregate testing for the evaluation of Johne's disease herd status, Aust. Vet. J. 73 (1996) 16–19.

[12] Larsen A.B., Johne's disease – immunization and diagnosis, Am. J. Vet. Med. Assoc. 163 (1973) 902–904.

[13] McNab W.B., Meek A.H., Duncan J.R., Brooks B.W., Van Dreumel A.A., Martin S.W., Nielsen K.H., Sugden E.A., Turcotte C., An evaluation of selected screening tests for bovine paratuberculosis, Can. J. Vet. Res. 55 (1991) 252–259.

[14] Obasanjo I.O., Grohn Y.T., Mohammed H.O., Farm factors associated with the presence of Mycobacterium paratuberculosis infection in dairy herds on the New York State Paratuberculosis Control Program, Prev. Vet. Med. 32 (1997) 243–251.

[15] Ott S.L., Wells S.J., Wagner B.A., Herd-level economic losses associated with Johne's disease on US dairy operations, Prev. Vet. Med. 40 (1999) 179–192.

[16] Renshaw E., Modelling biological populations in space and time, Cambridge University Press, Cambridge, 1991, p. 20.

[17] Repiquet D., Prévalence de la paratuberculose bovine en France, Société Française de Buiatrie, Paris, 2001.

[18] Ridge S.E., Morgan I.R., Sockett D.C., Collins M.T., Condron R.J., Skilbeck N.W., Webber J.J., Comparison of the Johne's absorbed EIA and the complement-fixation test for the diagnosis of Johne's disease in cattle, Aust. Vet. J. 68 (1991) 253–257.

[19] Sockett D.C., Conrad T.A., Thomas C.B., Collins M.T., Evaluation of four serological tests for bovine paratuberculosis, J. Clin. Microbiol. 30 (1992) 1134–1139.

[20] Sweeney R.W., Transmission of paratuberculosis, Vet. Clin. North Am. Food Anim. Pract. 12 (1996) 305–312.

[21] Thoen C.O., Baum K.H., Current knowledge on paratuberculosis, Am. J. Vet. Med. Assoc. 192 (1988) 1609–1611.

[22] Vose D., Risk Analysis, A quantitative guide, Wiley and Sons, Chichester, 2000.