A Data-driven Individual-based Model of Infectious Disease in Livestock Operation: A Validation Study for Paratuberculosis

Mohammad. A. Al-Mamun[1*], Rebecca. L. Smith[2], Annette. Nigsch[3], Ynte. H. Schukken[3,4] & Yrjo. T. Gröhn[5]

[1]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, Connecticut, United States of America.

[2]Department of Pathobiology, University of Illinois, College of Veterinary Medicine, Urbana, Illinois, 61802, United States of America.

[3]Department of Animal Sciences, Wageningen University, Wageningen, 6700 AH, The Netherlands.

[4]GD Animal Health, Arnsbergstraat 7, 7411 EZ, Wageningen, The Netherlands.

[5]Department of Population Medicine and Diagnostic Sciences, Cornell University, College of Veterinary Medicine, Tower Road, Ithaca, New York, 14853, United States of America.

*Email: mohammad.al-mamun@yale.edu

Abstract

Chronic livestock diseases cause large financial loss and affect the animal health and welfare. Controlling these diseases mostly requires precise information on both individual animal and population dynamics to inform farmer's decision. Mathematical models provide opportunities to test different control and elimination options rather implementing them in real herds, but these models require valid parameter estimation and validation. Fitting these models to data is a difficult task due to heterogeneities in livestock processes. In this paper, we develop an infectious disease modeling framework for a livestock disease (paratuberculosis) that is caused by *Mycobacterium avium* subsp. *paratuberculosis* (MAP). Infection with MAP leads to reduced milk production, pregnancy rates, and slaughter value and increased culling rates in cattle and causes significant economic losses to the dairy industry in the US. These economic effects are particularly important motivations in the control and elimination of MAP. In this framework, an individual-based model (IBM) of a dairy herd was built and a MAP infection was integrated on top of it. Once the model produced realistic dynamics of MAP infection, we implemented an evaluation method by fitting it to data from three dairy herds from the Northeast region of the US. The model fitting exercises used least-squares and parameter space searching methods to obtain the best-fitted values of selected parameters. The best set of parameters were used to model the effect of interventions. The results show that the presented model can complement real herd statistics where the intervention strategies suggested a reduction in MAP but no elimination was observed. Overall, this research not only provides a complete model for MAP infection dynamics in a cattle herd, but also offers a method for estimating parameter by fitting IBM models.

Introduction

Chronic livestock diseases like paratuberculosis (PTB) and bovine tuberculosis (bTB) are commonly reported worldwide (1,2). Bovine TB is caused by the pathogen Mycobacterium bovis (*M. bovis*) while PTB is caused by *Mycobacterium avium* subsp. *paratuberculosis* (MAP). In the UK, bTB has been spreading over the last two decades, putatively due to the presence of a wildlife reservoir in badgers(3). In United States (US), 68% of dairy herds have apparently at least one cow that is infected with MAP (4). Both diseases pose a potential threat not only to animal health and production, but also to public health. Historically, bTB has been a contributor to human TB cases worldwide and PTB infections in humans have been associated with an increased risk of Crohn's disease in humans(5). Recently, it has been reported that these diseases may induce additional collateral risks for public health due to dispensed antibiotics as a treatment in some cases can potentially contribute to the spread of antibiotic resistance(6).

In the US cattle industry, the cost of PTB was estimated at $250 million every year (7). Infection by MAP usually occurs in the first year of life(8) and transmission can occur vertically (9) and/or horizontally via ingestion of fecal material contaminated by MAP (10). As PTB is a slowly progressive disease, progression of individual animals through different MAP infection states is a complex continuous process alternating excreting and non-excreting stages with a late onset of clinical signs (11,12). It has a large economic impact for producers due to decreased milk production (13–15), premature culling (16,17), reduced slaughter value (18), low fertility (19,20), and an increased animal replacement rate (21). However, tests routinely used on individuals have low sensitivities, especially in the early stages of the disease (22).

In last two decades, different mathematical models have been developed on a within-herd scale to understand MAP transmission dynamics (23,24) and effectiveness of recommended control

strategies (25–28). These models were simulated to assess the impact of contact structure on the MAP transmission (23) , efficacy of test-and-cull policy (24,25,29,30), impact of low diagnostic test sensitivity in decision making (8,31), stopping some transmission pathways using hygiene improvement (32), improved calf management (33), impact of super-shedders in transmission(34,35) , and economic efficacy of recommended programs (29). Most of these studies suggest that culling a test positive animal is an effective solution to reduce the prevalence. However, none of the previous models considered the pervasiveness of MAP in the farm environment and the value of information of individual animal along with real dairy herd data. Moreover, controlling MAP requires significant management of testing and culling strategies to reduce the prevalence, which are normally unregulated and reliant on farmers' decisions(36). The decision of culling an animal is not straightforward and poses a multiscale problem where an individual animal, farm dynamics, infectious status and disease symptoms, and management profit are related (37). Substantial costs are also related to the implementation of control measures and prevention (21,34,38). Though previous compartmental MAP models have shown many potential interventions programs, most considered population-level decision making rather than individual-level animal information. Recently, individual-based models (IBMs) have been proposed to show the value of the information about the infection, daily life events and management policy for each individual animal within the farm (32,37,39–42)

Mathematical models of infectious diseases often direct us to understand both infection biology and efficacy of intervention policies taken in human and veterinary medicine(43). However, translating modeling results into practice require appropriately real-world assumptions to be built into the model. We hypothesize that in case of MAP, the use of model results will more realistic when the model has been built on up-to-date infection biology and epidemiology, parametrized from adequate real herd data and fitted back to that real-world scenario to test the recommended

intervention strategies. In this paper, our aim is to build an IBM framework of MAP infection that is fitted to and validated by in-depth longitudinal data from three northeastern dairy farms. The objective of this study was four-fold: first, we extended an existing IBM of a dairy herd to resemble the population level parameters (i.e. milk yield, herd size) with three real herds to create three *in silico* herds; second, we fitted the milk-yield measurement of individual animal to those herds; third, we fitted the model-predicted apparent MAP prevalence to the observed data to obtain herd-specific important infection parameters; and fourth, we integrated a risk-based control strategies on those three *in silico* herds to evaluate the efficacy. Finally, we discuss the value of observational data to feed information to simulation models, thereby making simulations more reflective and predictive of real-world circumstances.

Materials and Method

The Individual-based model

We used the dairy herd model named a multiscale agent-based simulation of a dairy herd (MABSdairy), an improved version of dairy herd published in Al-Mamun et al. (32,40). The MABSdairy is a multiscale stochastic IBM that simulates individual cows in a standard US cattle herd with a daily time step. In brief, each cow resides in one of three different management operations: adult/milking (aged >720 days), calf (aged 1-60 days) and heifer rearing housing (aged 61-719 days). Adult cows must calve to produce milk and the lactation cycle refers to the period between one calving and the next. The lactation cycle included the processes of a voluntary waiting period (interval during the postpartum period), insemination, and the dry off period (a non-lactating period prior to an impending parturition to optimize milk production in the subsequent lactation). For the fitting purpose, we modified the milk production Wood lactation curve by adding a herd-specific term and a herd-specific random component(44). The function

is defined as

$$Y_t = ad^b e^{ct} + f_i * f_r \quad (1)$$

$$where\ \ i = Parity\ 1\ and\ 2\ by\ farm\ A, B\ and\ C$$

where Y_t is the yield on day t after calving, d is days in milk (DIM), *a* is a scaling factor for initial yield, *b* is a rate factor for the increase in yield to peak, *c* is a rate factor for the decline after the peak, f_i farm specific factor and f_r is a random number. We used base milk yield parameters from Dematawewa et al. for parities 1 and ≥2 in the basic model (45).

MAP infection dynamics

The infection compartments in the milking herd were divided into four categories: susceptible (X_A), latent (H), low shedding (Y_1), and high shedding (Y_2). In calf rearing housing, there were two infection categories: susceptible (X_C) and infected (Y_C). In heifer rearing housing, there were also two infection categories: susceptible (X_H) and infected (Y_H). We included six different transmission routes: adult-to-adult, adult-to-calf (vertical transmission), adult-to-calf (horizontal transmission), environmental contamination, calf-to-calf, and heifer-to-heifer. The detailed infection structure is shown in Fig1.

Fig1. A flow diagram of animal movement among infection categories for the adult, calves, and heifers within the herd. Each horizontal gray box classifies the animals according to their initial age group. The green and red boxes define the susceptible and infected states, respectively, for each animal in the three age categories. The probabilities of exit at each time point from susceptible to latent, latent to low shedding and low shedding to high shedding animals are s_1, h_1, and y_1, respectively. Vertical transmission probabilities from latent, low shedding and high

144 shedding animals are V_h ,V_{y1}, and V_{y2}, respectively. Horizontal transmission probabilities to calves
145 from low shedding and high shedding animals are H_{y1} and H_{y2}, respectively. The probability an
146 animal gets infected by the environment is $\beta_{environment}$. Calf-to-calf and heifer-to-heifer
147 transmission probabilities are C_{inf} and Y_{inf}, respectively. Stochastic death/sale probabilities for
148 adult, calves, and heifers are μ_a, μ_c, and μ_h, respectively. μ is the replacement animals coming
149 from heifer compartment upon completion of two years.
150
151 In the milking herd group, adult animals could be infected by low and high shedding adults. The
152 probability of fecal-oral transmission for adult animals can be given by:

$$Inf_{adult-adult} = \beta_A \left(\frac{(\beta_{direct} + \beta_{environment})}{N} \right) \quad (2)$$

$$\beta_{direct} = \beta_{Y_1} Y_1 + \beta_{Y_2} Y_1$$

$$\beta_{environment} = U\,(0,1)$$

153 Susceptible adult animals in the milking herd compartment were susceptible to MAP infection
154 by contact with low shedding (Y_1) and high shedding (Y_2) animals with transmission rates of β_{Y_1}
155 and β_{Y_2}, respectively. β_A is the adult-to-adult transmission coefficient, $\beta_{environment}$ is the MAP
156 contamination risk from the environment and N is the total number of animals in the milking herd,
157 $N = X_A + H + Y_1 + Y_2$. The horizontal infection probability to calves can be determined by

$$Inf_{adult-calf} = \beta_a \left(\frac{\beta_{direct} + \beta_{environment}}{N_c} \right) \quad (3)$$

158 β_a is the horizontal transmission coefficient for an adult to newborn calves and N_c is the total
159 number of calves at every day, $N_c = X_c + Y_c$. A calf can also become infected vertically (i.e., in
160 utero infection) by an adult and which are modelled using the certain proportions(25) .

161 A calf stays in calf rearing housing for the first 60 days after birth. The probability of direct
162 transmission was calculated as

$$Inf_{calf-calf} = \alpha + \beta_c \left(\frac{Y_c + \beta_{environment}}{N_c} \right) \quad (4)$$

163 β_c is the horizontal calf-to-calf transmission coefficient, $\mathrm{N_c}$ is the total number of calves at each
164 day, X_c is susceptible calves, Y_c is infected calves. During the first day after birth, a calf may also
165 be infected horizontally by infected adults present in the maternity pen or vertically by an infected
166 dam.
167 Susceptible calves became susceptible heifers and infected calves became infected heifers.
168 Infected heifers could infect susceptible heifers by the heifer-to-heifer transmission path

$$Inf_{heifer-heifer} = \beta_h \left(\frac{Y_h + \beta_{environment}}{N_{X_H}} \right) \quad (5)$$

169 β_h is the horizontal heifer-to-heifer transmission coefficient, and the total number of heifers is
170 $\mathrm{N_{X_H}} = \mathrm{X_H} + \mathrm{Y_H}$. After one year, the infected heifers became latent heifers and eventually entered
171 the milking herd as latent adults. For simplifying the model, we assumed that heifer remains in
172 the heifer rereading housing are transiently shedding while they ended up in the adult herd as
173 latent animals.
174
175 **Observed herd data**
176 The longitudinal dataset used here was obtained from a longitudinal study of three commercial
177 dairy farms in the northeastern US: farm A in New York State, farm B in Pennsylvania, and farm
178 C in Vermont (46,47). All three farms participated in the Regional Dairy Quality Management
179 Alliance (RDQMA) project, which was a multistate research program conducted under a
180 cooperative research agreement between the USDA Agricultural Research Service (ARS) and

four Universities: Cornell University, Pennsylvania State University, University of Pennsylvania, and University of Vermont. The project consisted of longitudinal data collection for endemic infectious diseases of public and animal health concern in dairy herds. For a more complete description, including information on farms, samplings, and microbial analyses, see Pradhan et al.(46) Briefly, the milking herds consisted of approximately 330, 100, and 140 cows on farms A, B, and C, respectively. Sampling commenced in February, March, and November 2004 on farms A, B, and C, respectively, and continued for approximately 7 years, until 2010. The project design included a biannual collection of individual fecal samples and a quarterly collection of individual serum samples from all milking and non-lactating cows. Additionally, culled cows were tracked as much as possible from the farm to the slaughterhouse, where four gastrointestinal (GI) tissues and a fecal sample were collected with the cooperation of USDA Food Safety and Inspection Service (FSIS) personnel. The harvested tissues included two lymph nodes located at the ileocecal junction and two pieces of ileum, one taken from 20 cm proximal to the ileocecal valve and the other taken from very near the ileocecal valve. In addition to the sampling of animals, the farm environment was sampled in approximately 20 locations on a biannual basis. All fecal and environmental samples were tested by 4-tube culture for presence of viable MAP organisms, reported as colony-forming units per tube. All serum samples were tested using the ParaCheck ELISA (Prionics USA Inc., La Vista, NE) for antibody reactions to MAP antigens. On each of the farms, demographic data, production data and herd management information was collected. Precise demographic data included birth date, birth location, calving dates, fertility data, animal location data (pen status at any point in time), dry-off dates and culling information and cull dates. These demographic data were collected for each animal present on the farms. All infection data, strain typing data, herd management, demographic and production data was maintained in a relational database.

206 **Model parameters**

207 The parameterization of the base dairy herd model is described in Al-Mamun et al. (32,48). Initial
208 infection parameter values were updated according to Mitchell et al. 2015(43). Table 1 provides
209 the base parameters for the initial MAP transmission before fitting the model to the RDQMA
210 herds.

211 Table 1. Base parameter values of *Mycobacterium avium* subsp. *paratuberculosis* (MAP)
212 infection within a dairy herd.

Symbols	Description	Initial value	References
V_h	The proportion of calves from latent animals infected at birth	0.15	(25)
V_{y1}	The proportion of calves from low-shedding animals infected at birth	0.15	(25)
V_{y2}	The proportion of calves from high-shedding animals infected at birth	0.17	(25)
β_A	Adult-to-adult transmission coefficient	0.05	(32)
β_a	Adult-to-calf transmission coefficient	0.383	(32)
β_c	Calf-to-calf transmission coefficient	0.0025	(32)
β_h	Heifer-to-heifer transmission coefficient	0.001	Calibrated in the model
β_{y_1}	Transmission rate between low shedders (Y_1) and susceptible (X_A)	2/year	Calibrated in the model
β_{y_2}	Transmission rate between high shedders (Y_2) and susceptible (X_A)	20/year	Calibrated in the model

213

214

215 **Model fitting method**

216 The goal of the model-fitting exercise was to estimate key parameters in order to produce results
217 consistent with the epidemiologic data from three farms. Our fitting exercise was two-fold: first,
218 we fitted our base dairy herd models with farm-specific parameters (total population and milk
219 yield), then we fitted the model predicted apparent prevalence results based on antemortem
220 ELISA and fecal testing and postmortem tissue and fecal testing results for the farm. To assess
221 the goodness-of-fit we sampled from the defined parameter ranges in multiple rounds and the
222 model was run for three different scenarios of each of the three farms. The model fitting was
223 done using a non-linear fitting method named Nelder-Mead Simplex Method (49), which is used
224 for unconstrained optimization. While fitting the milk yield and apparent prevalence, the best-fit
225 parameters were extracted.

226 To determine the specific range for each parameter, we used multidimensional parameter space
227 searching method. The point estimate of each parameter was taken as a mean value and, using
228 Latin Hypercube Sampling, 100,000 parameter combinations were generated spanning the
229 specified range ±75% of the mean values. The searching was done in two stages. In the first
230 stage, we set a broad range to identify the particular regions of the parameter range and chose
231 the best 10000 (1%) parameter sets. In the next stage, we ran the simulation with 10% parameter
232 sets to compare with the best fit curve by minimizing the sumsquare error. The parameter ranges
233 presented in the results section were calculated from the top 1% simulations.

234

235 **Intervention strategies**

236 Once the three *in silico* herds were stable using fitted values, we tested a proposed intervention
237 strategy. We chose risk-based testing and culling strategies suggested by Al-Mamun et al. (32).

238 In brief, all cows that tested negative throughout testing were marked as low risk or green cows.
239 The cows that tested positive were divided into two groups: yellow and red. Red animals had at
240 least 2 positive tests out of the last 4 tests and yellow cows had one positive test. We proposed
241 two controls: control I, culling red animals straightway (aggressive culling); and control II, culling
242 only red animal with a delay of 305 DIM (delayed culling). The simulations results were then
243 compared against the observed pre-fitted data from the three herds. To evaluate the efficacy of
244 the intervention, we divided our seven years of observation into two parts: the first 4 years were
245 used for pre-intervention fit, while the last 3 years were used for validation against the model
246 results, in which the intervention was introduced and run for 3 years. Moreover, we extended the
247 intervention for more two years to see the long-term efficacy.
248
249 **Simulation background**
250 First, the base dairy herd model was initiated with a certain proportion of adult animals for farms
251 A (330), B (100) and C (140). Second, after a 2-year burn-in period the model was run for 7 more
252 years to resemble the observations of the real herds. During the 2 years burn-in period, each
253 farm was assumed to be self-sufficient in producing their own replacement, so that no animal
254 purchase from outside was needed. The model was initiated with a pre-determined distribution
255 of animals with different parities. Every day, the algorithm first determined the group of animals.
256 If it found adult animals, it checked reproductive status (voluntary waiting period (VWP), waiting
257 to be inseminated, and pregnant) and milk yield status. Any cow on the 280th day of pregnancy
258 was assumed to calve. For a newborn calf, the stillbirth probability was checked; if the calf was
259 not stillborn, it was flagged as a calf. Only female calves were kept in the herd, and male calves
260 were removed/sold immediately after birth. Once an adult animal calved, it transitioned to VWP
261 status and continued in the milking herd loop until it was removed due to culling or death.
262 Mortality was allowed in the calf rearing loop; otherwise, calves were transferred into the heifer

263 loop at the 61st day of age. In the heifer loop, heifers were inseminated at the 400th day of age
264 in order to become pregnant, so that they would calve at the 680th day of age. When heifers
265 were ready to calve for the first time, they transitioned to the milking herd in the model. The
266 model was fitted for the 7 years data for each farm. Third, for testing intervention strategies, each
267 model was fitted to the first 4 years of data- that is called pre-intervention fit, and then the
268 intervention was tested in 2 phases. In the first phase, 3 years and then extended more 2 years
269 to see how the suggested strategies result in long term. The base model was developed as
270 custom codes in MATLAB and other data analysis were done using R.
271
272 **Results**
273 The purpose of the fitting exercises was to obtain a better fit to the estimates of three herds prior
274 fitting to the apparent prevalence. The model predicted total number of animals (adult, calves,
275 and heifers) closely resembles the data from the three real farms (shown in table 2).
276
277 Table 2. The comparison of observed and predicted values from three *in silico* farms in terms of
278 a total number of animals, and average daily milk yield (in kg) for 305 days, presented as Mean
279 (95% Confidence Interval).

Herd A	Total number of animals	Milk yield: parity 1	Milk yield: parity ≥ 2
Observed	720 (708-754)	36.07 (29.61-40.73)	39.48 (27.11-49.86)
Predicted	714 (693-737)	36.15 (30.44-40.73)	39.49 (27.56- 50.18)
Herd B			
Observed	194 (102-230)	33.38 (24.34-40.35)	34.52 (17.42-47.40)
Predicted	200 (182–219)	32.97 (26.51-38.31)	34.97 (21.86-48.29)
Herd C			

Observed	262 (116-339)	27.49 (19.03-34.68)	27.49 (19.03-34.68)
Predicted	221 (184-257)	27.16 (20.23-32.98)	27.90 (17.12-38.06)

Fig2 shows the concordance between predicted and observed milk yield data from three herds. It is evident that the models predicted milk yield estimations matched with the observed milk yield from three northeastern herds. The best fit model predictions to the observed milk yield curve for parity 1 and parity ≥2 are shown in supplementary FigS1 . The best fitting lines also describe that the model was able to capture inherent randomness from the data into the model. The estimation of the critical parameters *a, b, c,* and f_i of the modified lactation curve are presented in table 3.

Fig2. The comparison of observed and model predicted milk yield distribution for 1% simulation using best fit parameters for the milk yield. In the box plot, the bottom and top end of the bars are minimum and maximum values respectively, the top of the box is the 75th percentile, the bottom of the box is the 25th percentile, and the horizontal line within the box is median; outliers are presented as a solid black circle and the density of the milk yield is presented by the width of the violin.

Table 3. The estimated parameters from the fitting exercise for the modified milk yield function for three farms A, B, and C.

	a	*b*	*c*	Herd specific parameter (f_i)
Farm A-Parity 1	17.87	0.207	0.00199	3.59
Farm A-Parity≥2	25.23	0.199	0.00329	5.19
Farm B-Parity 1	16.09	0.198	0.00196	6.60

Farm B-Parity≥2	23.38	0.200	0.00392	8.33
Farm C-Parity 1	15.25	0.193	0.00269	6.55
Farm C-Parity≥2	16.50	0.215	0.00399	9.00

Model fitting exercises

Table 4 represents the observed apparent prevalence and apparent incidence and the tracking of the animals in the next biannual testing for three farms for seven years, 2004-2010. The observed prevalence shows zero infected animals in the last half of 2010, for the sake of persistence scenario we replace that with the previous quarter value. During our simulation, we normalized the prevalence with the previous half of the year so that it remains consistent for our simulation. We simulated the three *in silico* farms to fit with the observed apparent prevalence data from herd A, B and C.

Fig3 shows the model predicted prevalence with a 95% confidence interval while fitting against the observed prevalence. It should be noted that our model confidence interval overpredicts the prevalence of herd B, but for other two herds it forecasts the better fitting. Through this model fitting exercise, our aim was to estimate the critical infection parameters for each herd, so that we can suggest herd specific intervention strategies.

312

313 Table 4. The calculation of apparent prevalence and apparent incidence and the tracking of the animals in the next testing in bi-

314 annually phase for three farms (2004-2010).

Year	**2004**		**2005**		**2006**		**2007**		**2008**		**2009**		**2010**	
Test phase	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Herd A														
Total positive cows [a]	14	25	34	28	34	32	21	23	24	17	14	9	7	0
Animals tested	315	330	364	349	354	364	338	332	337	347	341	347	296	239
Apparent prevalence	4.4	7.6	9.3	8.0	9.6	8.8	6.2	6.9	7.1	4.9	4.1	2.6	2.4	0.0
New cases[b]	14	16	18	12	13	13	7	10	11	6	5	2	0	0
Cow-years at risk[c]		239		293		284		272		276		280		198
Apparent incidence[d]		0.13		0.10		0.09		0.06		0.06		0.02		0
Herd B														
Total positive cows	9	8	6	3	6	5	4	3	3	2	5	5	1	0
Animals tested	106	122	128	128	113	113	115	114	111	109	113	109	82	1
Apparent prevalence	8.5	6.6	4.7	2.3	5.3	4.4	3.5	2.6	2.7	1.8	4.4	4.6	1.2	0.0
New cases	9	1	2	0	5	4	1	0	1	1	4	1	0	0
Cow-years at risk		72		99		95		94		93		83		37
Apparent incidence		0.14		0.02		0.10		0.01		0.02		0.06		0

Herd C														
Total positive cows	0	17	26	23	19	22	18	20	18	15	13	8	7	0
Animals tested	0	121	145	149	178	161	145	155	157	145	142	117	102	0
Apparent prevalence	NA	14.0	17.9	15.4	10.7	13.7	12.4	12.9	11.5	10.3	9.2	6.8	6.9	NA
New cases	0	17	9	7	5	9	5	4	4	5	2	2	1	0
Cow-years at risk		13		114		123		110		117		108		33
Apparent incidence		1.27		0.14		0.11		0.08		0.08		0.04		0.03

315 [a]Test positive cows by considering enzyme-linked immunosorbent assay (ELISA) testing, fecal testing and tissue testing.

316 [b]Number of cows tested positive for the first time

317 [c]Observation time (in years) from entry in the study (at the first testing) until each cow tested positive or left the study (by culling, i.e.

318 the infection status of cow is right censored)

319 [d]New cases per year / cow-years at risk

Fig3. The fitting results of three *in silico* herds A (top), B (middle), and C (bottom) compared to the observed apparent prevalence for 7 years by biannual sampling. The shaded region shows the 95% confidence interval of the best 1% simulation runs.

Estimated parameters

Table 5 provides the best fit estimates of herd-specific infection parameters for three northeastern dairy herds. Among the three herds, the model suggested that dam-to daughter transmission routes were the major transmission routes with the coefficient (β_a) values of 0.4046, 0.1781 and 0.825 for farm A, B and C respectively. Environmental contamination was the second major transmission routes while adult-to-adult transmission route was ranked third. Interestingly, we found that the importance of adult-to-calf transmission was highest in herd C, in which the initial number of latent animals were highest in numbers among the three herds. Based on the best 1% parameter sets, herd C again had the highest number of latent animals present (shown in supplementary table S1).

Table 5. The values of fitted parameters for three farms A, B and C.

Parameters	**Herd A**	**Herd B**	**Herd C**
Adult to adult transmission coefficient (β_A)	0.0069	0.0023	0.0005
Adult to calf transmission coefficient (β_a)	0.4046	0.1781	0.825
Environmental transmission coefficient ($\beta_{environment}$)	0.0869	0.0711	0.0162

Calf to calf transmission coefficient (β_c)	5.3×10^{-06}	3.61×10^{-06}	5.2×10^{-06}
Heifer to heifer transmission coefficient (β_h)	4.36×10^{-06}	1.18×10^{-06}	1.98×10^{-06}
Initial Latent animals (H_i)	18	12	81
Initial low shedding animals (Y_{1_i})	15	2	12
Initial high shedding animals (Y_{2_i})	22	8	9

It is also noticeable that herd A has the highest adult-to-adult transmission probability among the three farms. Also, the initial starting distribution of the infected animals was very important for the fitting. It is seen that herd C start with the highest proportion of latent (73%) and low shedding (31%) animals among the three farms. The best-fitted parameters set is shown in the supplementary table (shown in supplementary table S1).

Intervention strategies

Once the three *in silico* herds were obtained from the fitting exercises, our next aim was to test the risk-based test and culling policy for each farm. The risk-based intervention was implemented after 4 years of the initially fitted model to see the efficacy of the intervention strategy. Fig4 presents the summary of the pre-intervention, post-intervention and extended intervention results to the three fitted dairy herds. The results clearly show that the suggested intervention policy reduces the overall apparent prevalence for three herds, but it is noticeable that for high endemic herds the risk-based culling was comparatively less effective than the low endemic herds. To investigate further, we extended our intervention 2 years beyond the

observations, but we did not see any elimination of MAP infection for the risk-based culling policy with control II. Culling red animals immediately (control I) was the best policy for all herds to decrease prevalence. Furthermore, we also calculated the number of years taken by the model to reduce the prevalence by 25% and 5% while two control programs were implemented after the pre-intervention period for three farms (shown in FigS2).

Fig4. The apparent prevalence during the pre- and post-intervention period during the simulation of three *in silico* herds with two control strategies. Control I: culling red animals immediately and control II: culling only red animal with a delay of 305 days in milk. The two control measures are simulated in separating runs of the three *in silico* herds.

Discussion and conclusion

Currently, only imperfect intervention strategies are available for PTB in the US. Therefore, there is a need to develop more effective control strategies to facilitate elimination of this disease from dairy herds. To enhance this effort, the mathematical modeling can play an important role, but the models can only provide realistic results when built from real herd data, to estimate the herd and infection-specific parameters and to test different intervention strategies prior to implementation in real herds. This paper presents an IBM modeling framework of MAP where simulation prediction was fitted and validated using datasets from a longitudinal study conducted in three northeastern dairy herds. The fitting exercise shows that the IBM is capable to reproduce the observed milk yield of each of the three herds separately and estimate key herd-related parameters. Next, the model results show the best fit to the observed apparent prevalence and estimate critical transmission parameters for three herds. Ultimately, the best fitted *in silico* herd models were simulated using risk-based test and culling intervention

375 strategies, showing that these strategies may be more beneficial for low prevalence herds than
376 for moderately endemic herds.
377
378 The epidemiology of MAP is difficult to study due to the slow progressing nature of MAP,
379 insufficient testing methods, intermittent shedding of MAP and lack of clinical signs. Many
380 infected animals are only detected years after initial infection or are actually never detected.
381 However, precise information on the infection status of animals is valuable for implementing
382 control strategies. Furthermore, specific information about the animal's daily life events in the
383 herd (such as age, milk yield, parity status, clinical signs and adult, calf and heifer rearing
384 management policies) may assist in designing real-world control strategies. To this purpose, our
385 IBM approach introduced a closed dairy herd model validated with longitudinal datasets
386 (43,46,50). The basic herd fitting results suggest that we were able to create three *in silico* farms
387 where the animal distribution was similar to the real herds (shown in table 1). This fitting exercise
388 suggests that our base dairy herd model is capable of producing stable closed *in silico* dairy
389 herds, with similar milk yield based on herd-specific milk yield parameters. This kind of features
390 is very important to evaluate the economic efficacy of the implemented interventions (51).
391 Moreover, often milk yield gets ignored from the MAP infection model, but accumulatively lower
392 milk yield influences the culling of animals which is not normally marked that the animal was
393 culled due to Paratuberculosis symptoms. Similar picture was seen in our data analysis of
394 RDQMA herds where we found there were only 0.01% times where the animal was culled due
395 to Paratuberculosis. Our previous study shows that low- and high-path animals produced more
396 milk before their first positive test than always-negative animals, especially high-path animals.
397 Although mean production decreased after a first positive test, low-path animals were shown to
398 recover some productivity(50,52). To account the overall impact of milk yield on culling, we used

399 threshold values of milk yield for parity 1 and 2 for each farm by calculating median milk yield
400 values for each parity from observed data.
401
402 Next, we fitted three *in silico* herds to the apparent prevalence of the RDQMA herds. The 95%
403 prediction interval shows that our model captured the trends of the apparent prevalence for three
404 farms (shown in Fig3). Here we used antemortem ELISA and fecal testing and postmortem tissue
405 and fecal testing results to determine the test positive animals in our model. In reality,
406 determining the prevalence is a complex process and such fine-grained detail is rarely available.
407 For antemorterm fecal culture test the sensitivity is determined 23-29% and 70-74% for infected
408 cattle and infectious cattle respectively while at the slaughter house culture of tissue and fecal
409 results 50% and 100% sensitivity and specificity, respectively. On the contrary for ELISA test
410 our RDQMA suggests 20% and 96% sensitivity and specificity, respectively and these numbers
411 are aligned with the previous reports by Nielsen and Toft(53) . To avoid this complexity, we have
412 chosen a range of 25-35% sensitivity for infected animals and 96% specificity. Recently, an
413 adaptive test scheme was suggested from a simulation model simulated on the standard Danish
414 dairy herd (8).In another study, test-records from 18,972 Danish dairy cows with MAP specific
415 IgG antibodies on their final test-record were used to estimate age-specific sensitivities (54). It
416 is a critical decision for a farm owner to choose one of the antemorterm test as the outcome of
417 the fecal culture results can be delayed while ELISA test is also imperfect. Moreover, it also
418 depends of the testing practices and recommendations varied in different geographical regions
419 while strategies like adaptive test scheme, age-specific sensitivities and frequent testing can
420 provide us optimal solution. But, care should be taken whether using frequent testing strategies
421 may pick the false positive animal.
422

In order to control an infectious disease, it is important to determine which transmission routes are playing a major role in persistence of the pathogen on the farm. Traditionally, the dam-to-daughter route is considered the primary route for transmitting MAP, but it can vary due to herd management policy. It is very difficult to estimate this parameter directly from the epidemiological data due to imperfect testing, misidentification of super-shedders and management policies. The parameter value range estimated here suggests that dam-to-daughter transmission was indeed the primary transmission route with environmental transmission played as a secondary role. The role of environmental contamination is also difficult to measure from the epidemiological data, as MAP is pervasive within a dairy herd. A recent effort was made to quantify the environmental contamination through fecal-culture and mathematical studies (55). In our longitudinal data, the environmental samples were collected quarterly from several locations from farms. The cultures results suggest that manure storage areas and shared alleyways were most likely to be positive for three herds (56), but no relationship was found between non-pen environmental sample status and the distance between shedding animals and the sample's location, and neighboring pens did not significantly affect the results of the pen-level analysis. In our model, we modeled $\beta_{environment}$ in a crude way using a probability distribution for the sake of simplicity. To precisely quantify the role of different environments, further investigation into infection sources may be needed, potentially by examining the pathogens' genomic sequencing data.

To date, the best-suggested control strategies against MAP is test and cull strategies. Previously, several compartmental models were used to test different testing and culling strategies, providing the average impact of the testing and culling strategies. However, targeted test and cull requires combining information from each individual animal with farm management and hygiene policy. Recently, an IBM model suggested that a new ethanol vortex ELISA (EVELISA) could be cost-effective and that quarterly test-and-cull control was able to

448 significantly reduce the prevalence (41). Another model, SimHerd, developed by Kudahl et al.
449 required fecal culture confirmation of ELISA-positive cows before culling, and relied on repeated
450 testing to find the most infectious animal. Neither of these two models were validated and fitted
451 to real dairy herd data (57). A recent mechanistic bio-economic model showed that MAP can be
452 eradicated, although the control strategy necessary was economically unattractive (27,31). That
453 model was parameterized specifically for Danish conditions, which are different from the US. In
454 a previous effort, we suggested risk-based culling strategies with four different options:
455 aggressive culling, culling open red cows after 305 DIM, culling dam and offspring and culling
456 dam but not the offspring and we tested these intervention strategies along with different hygiene
457 conditions on hypothetically endemic herds. For this study, we implemented two risk-based
458 control strategies: aggressive culling and culling open red cows after 305 DIM on three pre-fitted
459 herds. We found that aggressive culling resulted in the elimination of 24% and 47% of iterations
460 after three years and extended intervention, respectively, for a very low endemic herd (farm B).
461 We also found a probability of elimination 0.11 and 0.24 using culling of open red cows after 305
462 DIM in three years and extended intervention, respectively. However, it is expected to predict
463 elimination in very low endemic herds and previously it was seen in a few studies
464 (26,27,42,43,58). On the contrary, we found elimination in only 6% times after 5 years extended
465 intervention using aggressive culling in case farm A, which has considerably higher prevalence,
466 and we did not predict any elimination while culling open red cows after 305 DIM for farms C
467 and A in long run. However, in terms of moderate and higher prevalence most of the cases, the
468 farmers want to reduce the prevalence and it is important to simulate how long it takes to reduce
469 the prevalence at a certain level. From FigS2, it can be said that low endemic herd is more likely
470 to reach 5% of initial prevalence by less than 2 years while high endemic herd needs extended
471 time to reach to that point, but it may take more than 10 years in some cases. This suggests that
472 culling high shedding animals may not provide elimination in high endemic herds, although it can

lower the prevalence. The study by Kirkby et al. serial testing along with hygiene play a critical role in the elimination process in Danish dairy herds, but these may not be economically justifiable (58). Caution should be taken in transferring conclusions from Denmark to the US, as both systems are different in many factors. Control activities are not uniformly coordinated nationally and internationally due to the variation in different farm management policies and government programs.

MAP is endemic in the bovine population in the US, which makes elimination unlikely at this time. When elimination is not possible, we have to rely on implementing the best herd-specific control strategies. Previous compartmental models have shown variable results for investigating infection dynamics(23,25,27) , test-and-culling strategies(25,59) , vaccination(24,60,61) , and intermittent MAP shedding(30,43). None of these combined the individual animal's information with herd management policy while fitting the model to real herd data, however. In this regard, the IBM paradigm should provide more effective approaches to test the intervention by considering information about the individual animal and overall population. Before using the insights of any IBM, very careful consideration should be given how the model was parameterized and validated. In this current study, we developed a fitting framework where an existing IBM model was fit against a longitudinal field study on three northeastern dairy herds to create the real herd's condition in the *in silico* platform. The fitting exercises provide estimates of the critical parameters related to an infection whose transmission is herd-specific. Like all models, our model is limited by its assumptions. First, the current model fitting exercise only included combined testing efficacy, whereas in reality the observed herds used three different testing strategies (fecal culture, ELISA, and tissue culture). Second, the current model modeled the role of environmental contamination crudely, but the model is adaptive in nature, allowing for a more rigorous assessment of environmental contamination once data become available. Third,

497 our current model did not include any economic justification of the suggested control strategies,
498 but the same base model has previously been used to show the economic justification of culling
499 in case of the MAP in a separate study (51).

500 This modeling and fitting exercise presented in this paper open multiple doors of further
501 investigations in future. One extension of model can include the impact of MAP infection on milk
502 yield while including the economics of milk production for these three farms. Previously, it shown
503 that the mean milk production decreases after a first positive test, non-progressing animals were
504 shown to recover milk productivity while progressing animals continue to exhibit a decrease in
505 milk production, especially after their first high-positive fecal culture (52). This indicates there
506 needs more investigation how to relate milk production loss as a function of MAP infection
507 progression and testing results. Another extension of the model may include the clinical and
508 molecular data of the infected animals. But adding molecular data will require more investigation
509 how to find who infects whom parameters from the phylogenetic analysis (62,63). The current
510 model is adaptive in nature to add strain specific data for each individual animal.

511 In conclusion, an important aspect of model building is to perform validation of the models to
512 the real-life data. In this study, we developed an IBM framework for validating a dairy herd model
513 and infection dynamics of the MAP to a longitudinal dataset. The assessment of model results
514 leads us to the conclusion that the evaluation of model results is still a combination of intuitive
515 model results, validation of the model with the quality data, assumptions that integrated into the
516 modeling process and estimation of key critical parameters along with true biologics. This
517 framework can be used in any infectious disease scenario to quantify the importance of key
518 transmission routes, mapping individual-level data to population-level phenomena and decision
519 making based on implemented intervention policies while considering between host
520 transmission mechanisms within a closed population. In summary, the quality of the conclusions

drawn from model studies is closely linked to the quality of the data used for estimation of the parameters and model validation. Models that have been validated with real-world data are more likely to produce useful and valid results.

Acknowledgments

The authors gratefully acknowledge funding provided by the National Institute of Food and Agriculture of the United States Department of Agriculture through NIFA Award No. 2014-67015-2240.

Author Contributions

Conceived and designed the experiments: MAM, Performed the experiments: MAM, Analyzed the data: MAM, RLS, and AN, Wrote the paper: MAM, Editing and reviewing: RLS, AN, and YTG. and YHS.

References

1. BROUGHAN JM, JUDGE J, ELY E, DELAHAY RJ, WILSON G, CLIFTON-HADLEY RS, et al. A review of risk factors for bovine tuberculosis infection in cattle in the UK and Ireland. Epidemiol Infect [Internet]. 2016 Oct 25 [cited 2018 Jun 14];144(14):2899–926. Available from: http://www.journals.cambridge.org/abstract_S095026881600131X
2. Rathnaiah G, Zinniel DK, Bannantine JP, Stabel JR, Gröhn YT, Collins MT, et al. Pathogenesis, Molecular Genetics, and Genomics of Mycobacterium avium subsp. paratuberculosis, the Etiologic Agent of Johne's Disease. Front Vet Sci [Internet]. 2017 Nov 6 [cited 2018 Jun 14];4:187. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/29164142

3. Brooks-Pollock E, Roberts GO, Keeling MJ. A dynamic model of bovine tuberculosis spread and control in Great Britain. Nature [Internet]. 2014;511(7508):228–31. Available from: http://www.nature.com/doifinder/10.1038/nature13529

4. NAHMS. National Animal Health Monitoring System (2007) Dairy 2007-Johne's Disease on U.S. Dairies, 1991–2007,National Animal Health Monitoring Systems. Fort Collins, CO; 2007.

5. Behr M, Collin D. Paratuberculosis:Organism, Disease, Control. CABI. 2010.

6. Woolhouse M, Ward M, van Bunnik B, Farrar J. Antimicrobial resistance in humans, livestock and the wider environment. Philos Trans R Soc Lond B Biol Sci [Internet]. 2015 Jun 5 [cited 2018 Jun 14];370(1670):20140083. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25918441

7. Ott SL, Wells SJ, Wagner BA. Herd-level economic losses associated with Johne's disease on US dairy operations. Prev Vet Med [Internet]. 1999 Jun 11 [cited 2018 Jun 14];40(3–4):179–92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10423773

8. Kirkeby C, Græsbøll K, Nielsen SS, Christiansen LE, Toft N, Halasa T, et al. Adaptive Test Schemes for Control of Paratuberculosis in Dairy Cows. Sreevatsan S, editor. PLoS One [Internet]. 2016 Dec 1 [cited 2017 Jan 31];11(12):e0167219. Available from: http://dx.plos.org/10.1371/journal.pone.0167219

9. Whittington RJ, Windsor PA. In utero infection of cattle with Mycobacterium avium subsp. paratuberculosis: A critical review and meta-analysis. Vet J [Internet]. 2009 Jan 1 [cited 2018 Jun 14];179(1):60–9. Available from: https://www.sciencedirect.com/science/article/pii/S1090023307003061?via%3Dihub

10. van Roermund HJW, Bakker D, Willemsen PTJ, de Jong MCM. Horizontal transmission of Mycobacterium avium subsp. paratuberculosis in cattle in an experimental setting:

567 Calves can transmit the infection to other calves. Vet Microbiol. 2007;122(3–4):270–9.
568 11. Mitchell RM, Schukken Y, Koets A, Weber M, Bakker D, Stabel J, et al. Differences in
569 intermittent and continuous fecal shedding patterns between natural and experimental
570 Mycobacterium avium subspecies paratuberculosis infections in cattle. Vet Res
571 [Internet]. 2015;46(1):66. Available from:
572 http://veterinaryresearch.biomedcentral.com/articles/10.1186/s13567-015-0188-x
573 12. Nielsen SS. Use of diagnostics for risk-based control of paratuberculosis in dairy herds.
574 Pr [Internet]. 2009;31:150–4. Available from:
575 http://inpractice.bvapublications.com/cgi/content/abstract/31/4/150
576 13. Aly SS, Anderson RJ, Adaska JM, Jiang J, Gardner IA. Association between
577 Mycobacterium avium subspecies paratuberculosis infection and milk production in two
578 California dairies. J Dairy Sci [Internet]. 2010 Mar [cited 2018 Jun 14];93(3):1030–40.
579 Available from: http://www.ncbi.nlm.nih.gov/pubmed/20172223
580 14. Nielsen SS, Krogh MA, Enevoldsen C. Time to the occurrence of a decline in milk
581 production in cows with various paratuberculosis antibody profiles. J Dairy Sci [Internet].
582 2009 Jan 1 [cited 2018 Jun 14];92(1):149–55. Available from:
583 http://www.ncbi.nlm.nih.gov/pubmed/19109273
584 15. Sorge US, Kelton DF, Lissemore KD, Sears W, Fetrow J. Evaluation of the Dairy Comp
585 305 Module “Cow Value” in Two Ontario Dairy Herds. J Dairy Sci [Internet].
586 2007;90(12):5784–97. Available from:
587 http://www.sciencedirect.com/science/article/pii/S0022030207720556%5Cnhttp://pdn.s
588 ciencedirect.com/science?_ob=MiamiImageURL&_cid=279785&_user=789722&_pii=S0
589 022030207720556&_check=y&_origin=article&_zone=toolbar&_coverDate=31-Dec-
590 2007&view=c&originContent
591 16. Lombard JE, Garry FB, McCluskey BJ, Wagner BA. Risk of removal and effects on milk

592 production associated with paratuberculosis status in dairy cows. J Am Vet Med Assoc
593 [Internet]. 2005 Dec 15 [cited 2018 Jun 14];227(12):1975–81. Available from:
594 http://www.ncbi.nlm.nih.gov/pubmed/16379637
595 17. Tiwari A, VanLeeuwen JA, Dohoo IR, Stryhn H, Keefe GP, Haddad JP. Effects of
596 seropositivity for bovine leukemia virus, bovine viral diarrhoea virus, Mycobacterium
597 avium subspecies paratuberculosis, and Neospora caninum on culling in dairy cattle in
598 four Canadian provinces. Vet Microbiol [Internet]. 2005 Aug 30 [cited 2018 Jun
599 14];109(3–4):147–58. Available from:
600 https://www.sciencedirect.com/science/article/pii/S0378113505001872?via%3Dihub
601 18. Kudahl AB, Nielsen SS. Effect of paratuberculosis on slaughter weight and slaughter
602 value of dairy cows. J Dairy Sci. 2009;92(9):4340–6.
603 19. Marcé C, Beaudeau F, Bareille N, Seegers H, Fourichon C. Higher non-return rate
604 associated with Mycobacterium avium subspecies paratuberculosis infection at early
605 stage in Holstein dairy cows. Theriogenology [Internet]. 2009 Mar 15 [cited 2018 Jun
606 14];71(5):807–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19117602
607 20. Smith RL, Strawderman RL, Schukken YH, Wells SJ, Pradhan a K, Espejo L a, et al.
608 Effect of Johne's disease status on reproduction and culling in dairy cattle. J Dairy Sci
609 [Internet]. 2010;93(8):3513–24. Available from: http://dx.doi.org/10.3168/jds.2009-2742
610 21. Garcia, A. B., Shalloo L. Invited review: The economic impact and control of
611 paratuberculosis in cattle. J Dairy Sci [Internet]. 2015;98(8):5019–39. Available from:
612 http://www.ncbi.nlm.nih.gov/pubmed/26074241
613 22. Kirkeby C, Græsbøll K, Halasa T, Toft N, Nielsen SS. Mean effective sensitivity for
614 Mycobacterium avium subsp. paratuberculosis infection in cattle herds. BMC Vet Res
615 [Internet]. 2015;11:190. Available from:
616 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4529712&tool=pmcentrez&re

ndertype=abstract

23. Mitchell RM, Whitlock RH, Stehman SM, Benedictus A, Chapagain PP, Grohn YT, et al. Simulation modeling to evaluate the persistence of Mycobacterium avium subsp. paratuberculosis (MAP) on commercial dairy farms in the United States. Prev Vet Med. 2008;83(3–4):360–80.

24. Lu Z, Schukken YH, Smith RL, Grohn YT. Stochastic simulations of a multi-group compartmental model for Johne's disease on US dairy herds with test-based culling intervention. J Theor Biol [Internet]. 2010;264(4):1190–201. Available from: http://dx.doi.org/10.1016/j.jtbi.2010.03.034

25. Lu Z, Mitchell RM, Smith RL, Van Kessel JS, Chapagain PP, Schukken YH, et al. The importance of culling in Johne's disease control. J Theor Biol. 2008;254(1):135–46.

26. Kudahl AB, Nielsen SS, Østergaard S. Economy, efficacy, and feasibility of a risk-based control program against paratuberculosis. J Dairy Sci [Internet]. 2008;91(12):4599–609. Available from: http://www.sciencedirect.com/science/article/pii/S0022030208709263

27. Kudahl AB, Østergaard S, Sørensen JT, Nielsen SS. A stochastic model simulating paratuberculosis in a dairy herd. Prev Vet Med. 2007;78(2):97–117.

28. Kudahl a B, Nielsen SS, Ostergaard S. Strategies for time of culling in control of paratuberculosis in dairy herds. J Dairy Sci [Internet]. 2011;94(8):3824–34. Available from: http://dx.doi.org/10.3168/jds.2010-3933

29. Smith RL, Al-Mamun MA, Gröhn YT. Economic consequences of paratuberculosis control in dairy cattle: A stochastic modeling study. Prev Vet Med [Internet]. 2017;138:17–27. Available from: http://dx.doi.org/10.1016/j.prevetmed.2017.01.007

30. Smith RL, Schukken YH, Gröhn YT. A new compartmental model of Mycobacterium avium subsp. paratuberculosis infection dynamics in cattle. Prev Vet Med [Internet]. 2015;122(3):298–305. Available from: http://dx.doi.org/10.1016/j.prevetmed.2015.10.008

642 31. Kirkeby C, Græsbøll K, Nielsen SS, Christiansen LE, Toft N, Rattenborg E, et al.
643 Simulating the Epidemiological and Economic Impact of Paratuberculosis Control
644 Actions in Dairy Cattle. Front Vet Sci [Internet]. 2016 Oct 10 [cited 2017 Jan 31];3:90.
645 Available from: http://journal.frontiersin.org/article/10.3389/fvets.2016.00090/full
646 32. Al-Mamun MA, Smith RL, Schukken YH, Gröhn YT. Use of an Individual-based Model to
647 Control Transmission Pathways of Mycobacterium avium Subsp. paratuberculosis
648 Infection in Cattle Herds. Sci Rep [Internet]. 2017 Dec 19 [cited 2018 Jun 14];7(1):11845.
649 Available from: http://www.nature.com/articles/s41598-017-12078-z
650 33. Marcé C, Ezanno P, Weber MF, Seegers H, Pfeiffer DU, Fourichon C. Invited review:
651 modeling within-herd transmission of Mycobacterium avium subspecies
652 paratuberculosis in dairy cattle: a review. J Dairy Sci [Internet]. 2010;93(10):4455–70.
653 Available from: http://www.ncbi.nlm.nih.gov/pubmed/20854979
654 34. Magombedze G, Ngonghala CN, Lanzas C. Evalution of the “Iceberg Phenomenon” in
655 Johne’s Disease through Mathematical Modelling. Rao C V., editor. PLoS One [Internet].
656 2013 Oct 22 [cited 2018 Jun 18];8(10):e76636. Available from:
657 http://dx.plos.org/10.1371/journal.pone.0076636
658 35. Liu L, Oza S, Hogan D, Perin J, Rudan I, Lawn JE, et al. Global, regional, and national
659 causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an
660 updated systematic analysis. Lancet [Internet]. 2015 Jan 31 [cited 2018 Jun
661 18];385(9966):430–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25280870
662 36. Carslake D, Grant W, Green LE, Cave J, Greaves J, Keeling M, et al. Endemic cattle
663 diseases: comparative epidemiology and governance. Philos Trans R Soc Lond B Biol
664 Sci [Internet]. 2011 Jul 12 [cited 2018 Jun 18];366(1573):1975–86. Available from:
665 http://www.ncbi.nlm.nih.gov/pubmed/21624918
666 37. Kirkeby C, Græsbøll K, Nielsen SS, Toft N, Halasa T. Epidemiological and economic

667 consequences of purchasing livestock infected with Mycobacterium avium subsp.
668 paratuberculosis. BMC Vet Res. 2017;13(1):1–9.
669 38. Chi J, VanLeeuwen JA, Weersink A, Keefe GP. Direct production losses and treatment
670 costs from bovine viral diarrhoea virus, bovine leukosis virus, Mycobacterium avium
671 subspecies paratuberculosis, and Neospora caninum. Prev Vet Med [Internet]. 2002 Sep
672 30 [cited 2018 Jun 18];55(2):137–53. Available from:
673 https://www.sciencedirect.com/science/article/pii/S0167587702000946?via%3Dihub
674 39. Al-Mamun MA, Smith RL, Schukken YH, Gr??hn YT. Modeling of Mycobacterium avium
675 subsp. paratuberculosis dynamics in a dairy herd: An individual based approach. J
676 Theor Biol. 2016;
677 40. Al-Mamun MA, Grohn YT. MABSDairy: A Multiscale Agent Based Simulation of a Dairy
678 Herd. Proc 50th Annu Simul Symp [Internet]. 2017;8:1--8:12. Available from:
679 http://dl.acm.org/citation.cfm?id=3106388.3106396
680 41. Robins J, Bogen S, Francis A, Westhoek A, Kanarek A, Lenhart S, et al. Agent-based
681 model for Johne's disease dynamics in a dairy herd. Vet Res [Internet]. 2015;46:68.
682 Available from: http://dx.doi.org/10.1186/s13567-015-0195-y
683 42. Kirkeby C, Græsbøll K, Halasa T. Evaluating the impact of transmission mode,
684 calibration level and farmer compliance in simulation models of paratuberculosis in dairy
685 herds. Sci Rep [Internet]. 2018 Dec 14 [cited 2018 Jun 18];8(1):9100. Available from:
686 http://www.nature.com/articles/s41598-018-27518-7
687 43. Mitchell RM, Whitlock RH, Gröhn YT, Schukken YH. Back to the real world: Connecting
688 models with data. Prev Vet Med. 2015;118(2–3):215–25.
689 44. WOOD PDP. Algebraic Model of the Lactation Curve in Cattle. Nature [Internet]. 1967
690 Oct 14 [cited 2018 Jun 18];216(5111):164–5. Available from:
691 http://www.nature.com/doifinder/10.1038/216164a0

692 45. Dematawewa CMB, Pearson RE, VanRaden PM. Modeling Extended Lactations of
693 Holsteins. J Dairy Sci [Internet]. 2007 Aug 1 [cited 2018 Jun 18];90(8):3924–36. Available
694 from:
695 https://www.sciencedirect.com/science/article/pii/S0022030207718490?via%3Dihub
696 46. Pradhan a K, Van Kessel JS, Karns JS, Wolfgang DR, Hovingh E, Nelen K a, et al.
697 Dynamics of endemic infectious diseases of animal and human importance on three
698 dairy herds in the northeastern United States. J Dairy Sci [Internet]. 2009;92(4):1811–25.
699 Available from: http://dx.doi.org/10.3168/jds.2008-1486
700 47. Pradhan AK, Mitchell RM, Kramer AJ, Zurakowski MJ, Fyock TL, Whitlock RH, et al.
701 Molecular epidemiology of Mycobacterium avium subsp. paratuberculosis in a
702 longitudinal study of three dairy herds. J Clin Microbiol. 2011;49(3):893–901.
703 48. Al-Mamun MA, Grohn YT. MABSDairy: A Multiscale Agent Based Simulation of a Dairy
704 Herd. In: Proceedings of the 50th Annual Simulation Symposium [Internet]. San Diego,
705 CA, USA: Society for Computer Simulation International; 2017. p. 8:1--8:12. (ANSS '17).
706 Available from: http://dl.acm.org/citation.cfm?id=3106388.3106396
707 49. Lagarias JC, Reeds JA, Wright MH, Wright PE. Convergence Properties of the Nelder--
708 Mead Simplex Method in Low Dimensions. SIAM J Optim [Internet]. 1998 Jan 31 [cited
709 2018 Jun 18];9(1):112–47. Available from:
710 http://epubs.siam.org/doi/10.1137/S1052623496303470
711 50. Smith, R L, Grohn, Y T, Pradhan, A K, Whitlock, R H, Van Kessel, J S, Smith, J M,
712 Wolfgang, D R, Schukken YH. A longitudinal study on the impact of Johne's disease
713 status on milk production in individual cows. J Dairy Sci [Internet]. 2009;92(6):2653–61.
714 Available from: http://www.sciencedirect.com/science/article/pii/S0022030209705818
715 51. Verteramo Chiu LJ, Tauer LW, Al-Mamun MA, Kaniyamattam K, Smith RL, Grohn YT. An
716 agent-based model evaluation of economic control strategies for paratuberculosis in a

dairy herd. J Dairy Sci [Internet]. 2018 Apr 26 [cited 2018 Jun 18]; Available from: https://www.sciencedirect.com/science/article/pii/S0022030218303746

52. Smith RL, Gröhn YT, Pradhan AK, Whitlock RH, Kessel JS Van, Smith JM, et al. The effects of progressing and nonprogressing Mycobacterium avium ssp . paratuberculosis infection on milk production in dairy cows. J Dairy Sci [Internet]. 2016;99:1383–90. Available from: http://dx.doi.org/10.3168/jds.2015-9822

53. Nielsen SS, Toft N. Ante mortem diagnosis of paratuberculosis: A review of accuracies of ELISA, interferon-γ assay and faecal culture techniques. Vet Microbiol [Internet]. 2008 Jun 22 [cited 2018 Jul 12];129(3–4):217–35. Available from: https://www.sciencedirect.com/science/article/pii/S0378113507006451?via%3Dihub

54. Nielsen SS, Toft N, Okura H. Dynamics of Specific Anti-Mycobacterium avium Subsp. paratuberculosis Antibody Response through Age. PLoS One. 2013;8(4).

55. Davidson RS, McKendrick IJ, Wood JC, Marion G, Greig A, Stevenson K, et al. Accounting for uncertainty in model-based prevalence estimation: paratuberculosis control in dairy herds. BMC Vet Res [Internet]. 2012;8:159. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3544565&tool=pmcentrez&rendertype=abstract

56. Smith RL, Schukken YH, Pradhan AK, Smith JM, Whitlock RH, Van Kessel JS, et al. Environmental contamination with Mycobacterium avium subsp. paratuberculosis in endemically infected dairy herds. Prev Vet Med [Internet]. 2011;102(1):1–9. Available from: http://dx.doi.org/10.1016/j.prevetmed.2011.06.009

57. Kudahl AB, Østergaard S, Sørensen JT, Nielsen SS. A stochastic model simulating paratuberculosis in a dairy herd. Prev Vet Med [Internet]. 2007 Feb 16 [cited 2018 Jun 18];78(2):97–117. Available from: https://www.sciencedirect.com/science/article/pii/S0167587706002212

742 58. Kirkeby C, Græsbøll K, Nielsen SS, Christiansen LE, Toft N, Rattenborg E, et al.
743 Simulating the Epidemiological and Economic Impact of Paratuberculosis Control
744 Actions in Dairy Cattle. Front Vet Sci [Internet]. 2016 Oct 10 [cited 2017 Feb 1];3:90.
745 Available from: http://journal.frontiersin.org/article/10.3389/fvets.2016.00090/full
746 59. Smith RL, Schukken YH, Lu Z, Mitchell RM, Grohn YT. Development of a model to
747 simulate infection dynamics of Mycobacterium bovis in cattle herds in the United States.
748 J Am Vet Med Assoc. 2013;243:411–23.
749 60. Lu Z, Schukken YH, Smith RL, Mitchell RM, Gröhn YT. Impact of imperfect
750 Mycobacterium avium subsp. paratuberculosis vaccines in dairy herds: A mathematical
751 modeling approach. Prev Vet Med [Internet]. 2013;108(2–3):148–58. Available from:
752 http://dx.doi.org/10.1016/j.prevetmed.2012.08.001
753 61. Lu Z, Schukken YH, Smith RL, Gr??hn YT. Using vaccination to prevent the invasion of
754 Mycobacterium avium subsp. paratuberculosis in dairy herds: A stochastic simulation
755 study. Prev Vet Med [Internet]. 2013;110(3–4):335–45. Available from:
756 http://dx.doi.org/10.1016/j.prevetmed.2013.01.006
757 62. Crispell J, Zadoks RN, Harris SR, Paterson B, Collins DM, de-Lisle GW, et al. Using
758 whole genome sequencing to investigate transmission in a multi-host system: bovine
759 tuberculosis in New Zealand. BMC Genomics [Internet]. 2017 [cited 2018 Jul
760 20];18(1):180. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28209138
761 63. Zimpel CK, Brandão PE, de Souza Filho AF, de Souza RF, Ikuta CY, Ferreira Neto JS, et
762 al. Complete Genome Sequencing of Mycobacterium bovis SP38 and Comparative
763 Genomics of Mycobacterium bovis and M. tuberculosis Strains. Front Microbiol
764 [Internet]. 2017 Dec 5 [cited 2018 Jul 20];8:2389. Available from:
765 http://journal.frontiersin.org/article/10.3389/fmicb.2017.02389/full
766

767

768

769 Table S1. The best 1% parameter sets were ranked from the parameter searching space.

Parameters	**Herd A** **Mean (95% CI)**	**Herd B** **Mean (95% CI)**	**Herd C** **Mean (95% CI)**
Adult to adult transmission coefficient (β_A)	**0.0033 (0.00062-0.0069)**	**0.0046 (0.0017-0.0075)**	**0.0041 (0.00047-0.0065)**
Adult to calf transmission coefficient (β_a)	0.54 (0.11-0.96)	0.37 (0.064-0.079)	0.63 (0.055-0.9)
Environmental transmission coefficient ($\beta_{environment}$)	**0.053 (0.0089-0.090)**	**0.05 (0.0056-0.078)**	**0.046 (0.0036-0.087)**
Calf to calf transmission coefficient (β_c)	0.69×10^{-6} (0.4×10^{-6}-1.2×10^{-5})	6.5×10^{-6} (0. 69×10^{-6}-1.2×10^{-5})	8.4×10^{-6} (0.18×10^{-6}-1.2×10^{-6})
Heifer to heifer transmission coefficient (β_h)	0.54×10^{-6} (0.53×10^{-6}-0.1×10^{-4})	4.9×10^{-6} (0.29×10^{-6}-0.11×10^{-5})	3.9×10^{-6} (0.35×10^{-6}-0.1×10^{-4})
Initial latent	30 (3-75)	10 (5-19)	73 (52-85)
Initial low shedding animals	18 (4-35)	12 (2-36)	31 (4-49)
Initial high shedding animals	16 (7-23)	11 (2-22)	13 (2-23)

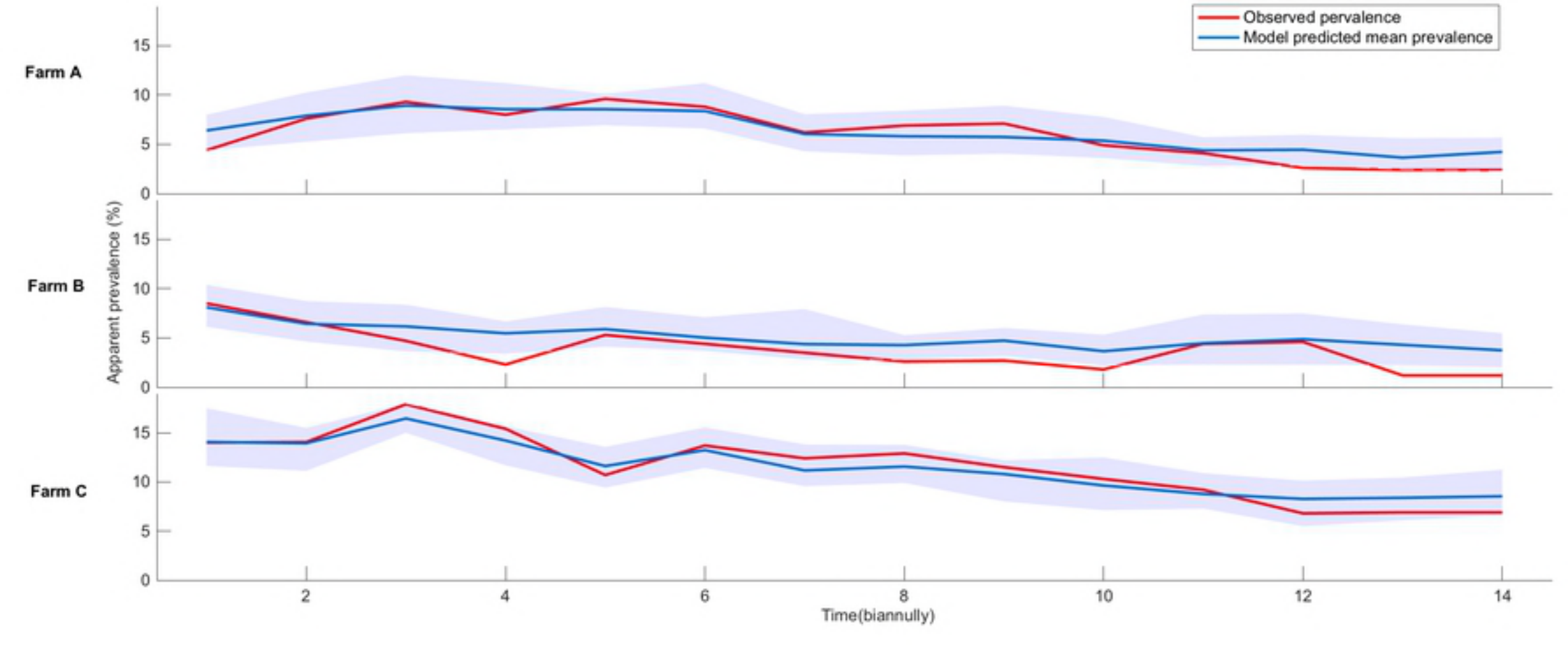
770

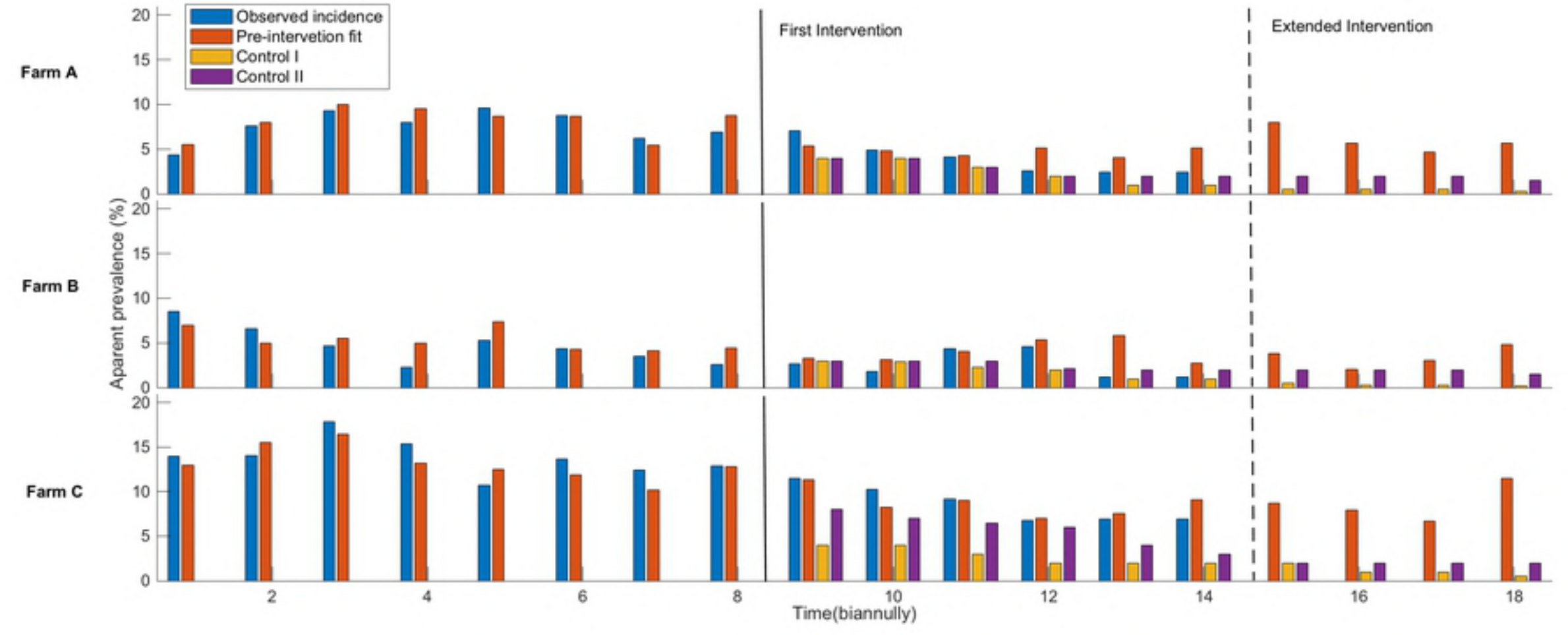
771 FigS1. The model predicted fitted to the observed milk yield for 360 days in milk for Farm A, B

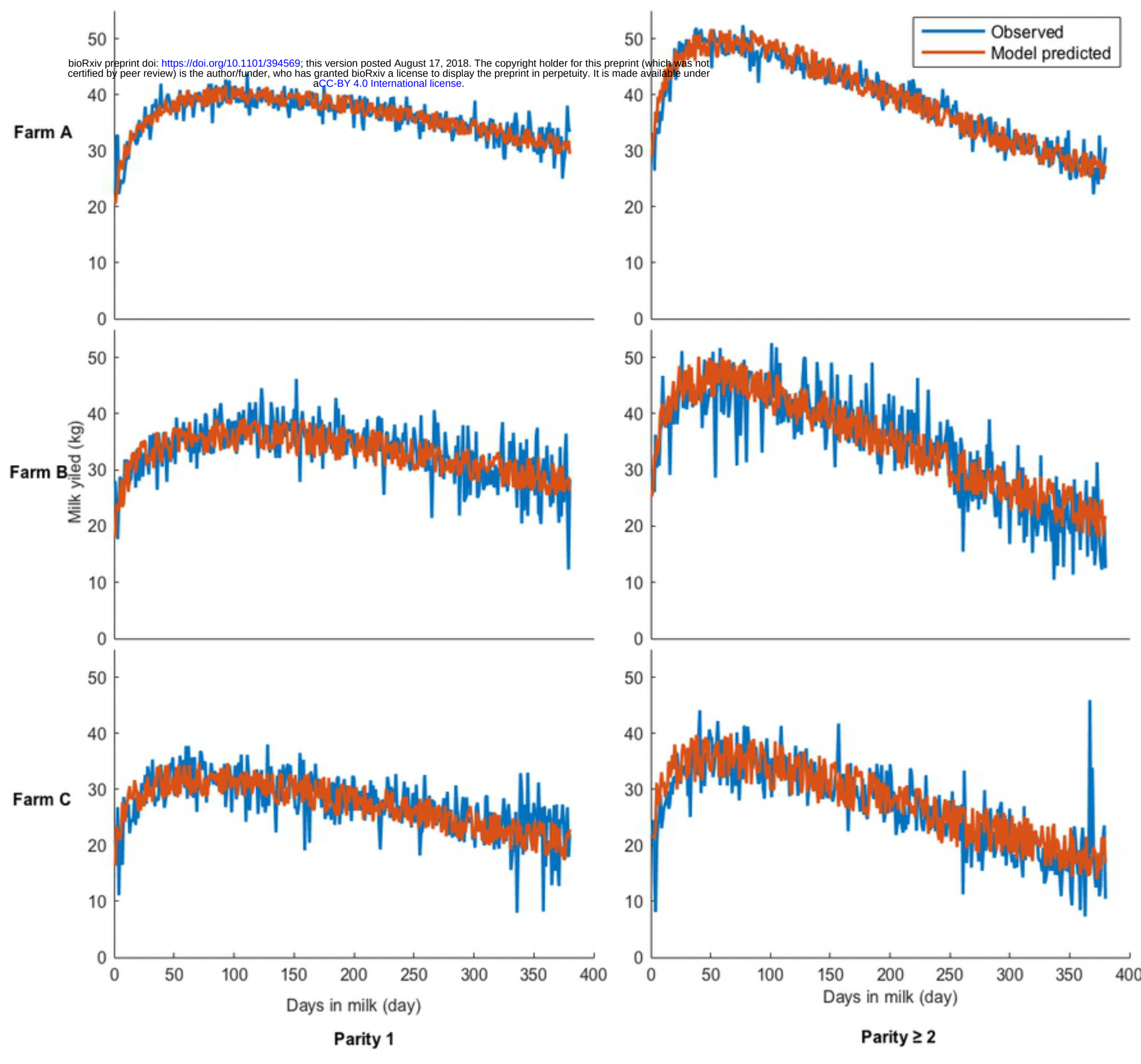
772 and C. The milk yield was calculated using equation shown in the method section.

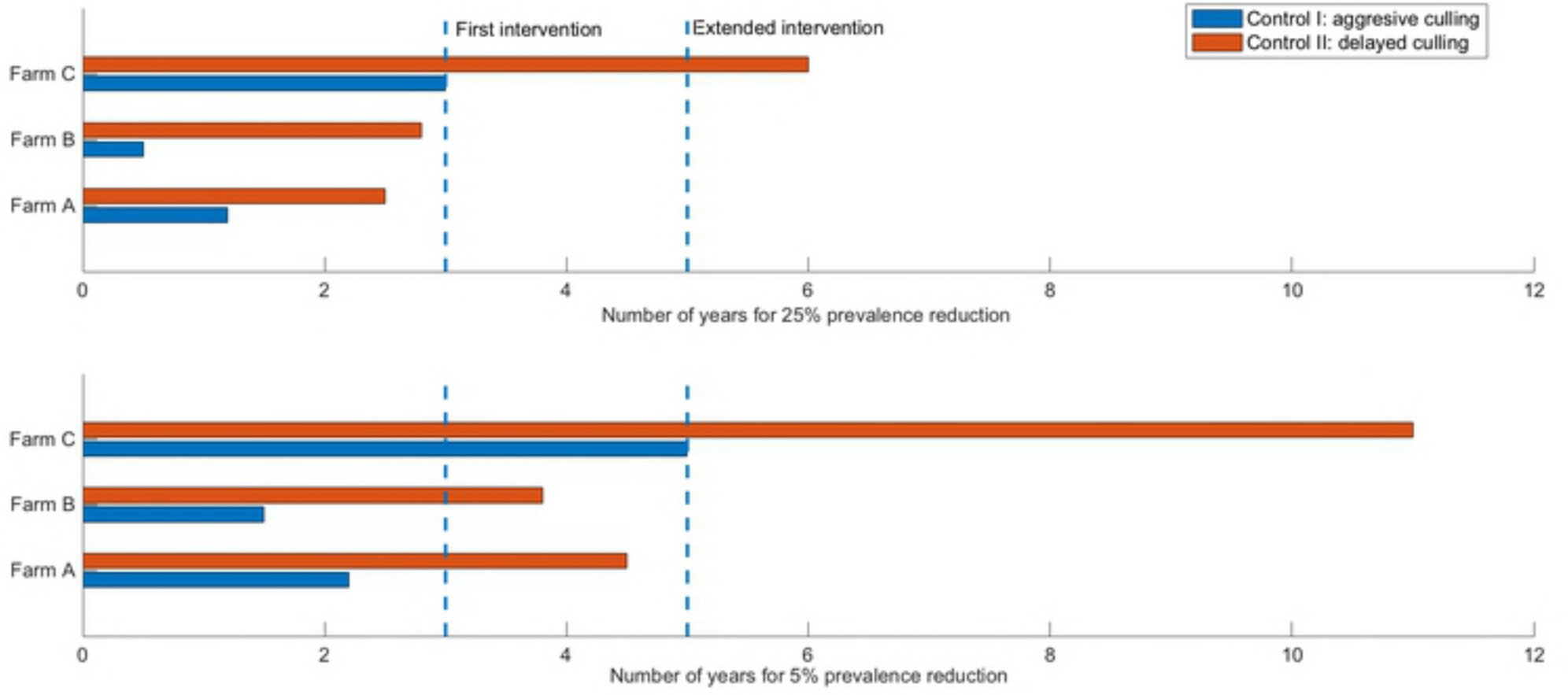
773 FigS2. The model predicted median number of years to reduce the apparent prevalence by 25%

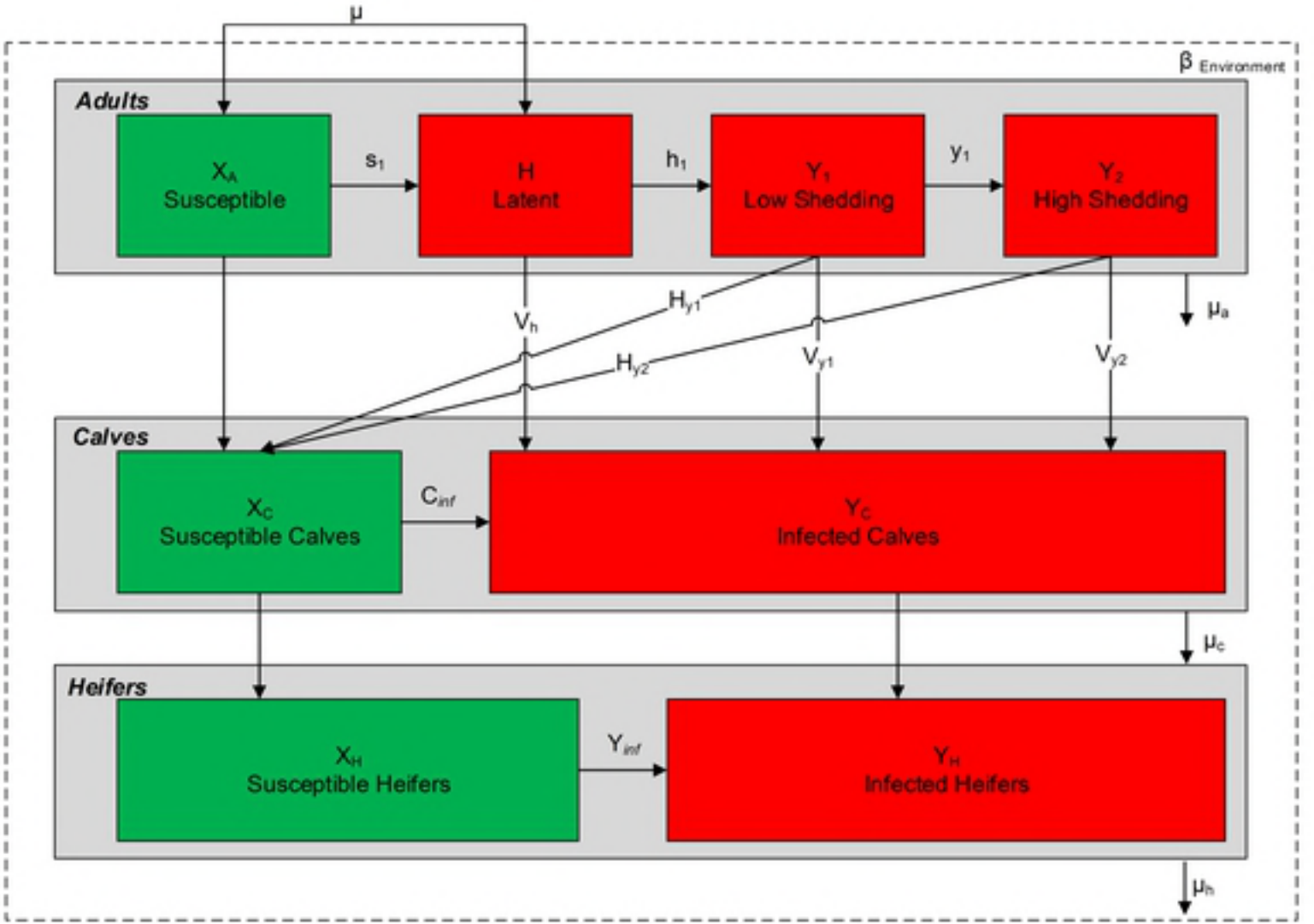
774 (top panel) and 5% (bottom panel) calculated from top 1% simulations with best set of
775 parameters while implementing two control scheme I: aggressive culling and control II: delayed
776 culling after the pre-intervention fit for the farms A, B and C.


Observed pervalence
Model predicted mean prevalence
Farm A
Farm B
Farm C
Apparent prevalence (%)
0
5
10
15
2
4
6
8
10
12
14
Time(biannully)


Observed incidence
Pre-intervetion fit
Control I
Control II
First Intervention
Extended Intervention
Farm A
Farm B
Farm C
Aparent prevalence (%)
Time(biannully)
0
5
10
15
20
2
4
6
8
10
12
14
16
18




First intervention
Extended intervention
Control I: aggresive culling
Control II: delayed culling
Farm C
Farm B
Farm A
0
2
4
6
8
10
12
Number of years for 25% prevalence reduction
Farm C
Farm B
Farm A
0
2
4
6
8
10
12
Number of years for 5% prevalence reduction


μ
β Environment
Adults
X_A
Susceptible
s_1
H
Latent
h_1
Y_1
Low Shedding
y_1
Y_2
High Shedding
V_h
H_{y1}
H_{y2}
V_{y1}
V_{y2}
μ_a
Calves
X_C
Susceptible Calves
C_{inf}
Y_C
Infected Calves
μ_c
Heifers
X_H
Susceptible Heifers
Y_{inf}
Y_H
Infected Heifers
μ_h