Preventive Veterinary Medicine 118 (2015) 215–225

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Back to the real world: Connecting models with data

Rebecca M. Mitchell[a,b], Robert H. Whitlock[c], Yrjö T. Gröhn[a], Ynte H. Schukken[a,d,*]

[a] Department of Population Medicine and Diagnostic Sciences, Cornell University, Ithaca, NY 14853, USA
[b] Centers for Disease Control and Prevention, Division of Parasitology and Malaria, GA, USA
[c] New Bolton Center, University of Pennsylvania, Kennett Square, PA, USA
[d] GD Animal Health, Deventer, The Netherlands

ARTICLE INFO

Article history:
Received 8 March 2014
Received in revised form
30 November 2014
Accepted 6 December 2014

Keywords:
Mathematical modeling
Observational studies
Mycobacterium avium subspecies *paratuberculosis*

ABSTRACT

Mathematical models for infectious disease are often used to improve our understanding of infection biology or to evaluate the potential efficacy of intervention programs. Here, we develop a mathematical model that aims to describe infection dynamics of *Mycobacterium avium* subspecies *paratuberculosis* (MAP). The model was developed using current knowledge of infection biology and also includes some components of MAP infection dynamics that are currently still hypothetical. The objective was to show methods for parameter estimation of state transition models and to connect simulation models with detailed real life data. Thereby making model predictions and results of simulations more reflective and predictive of real world situations. Longitudinal field data from a large observational study are used to estimate parameter values. It is shown that precise data, including molecular diagnostics on the obtained MAP strains, results in more precise and realistic parameter estimates. It is argued that modeling of infection disease dynamics is of great value to understand the patho-biology, epidemiology and control of infectious diseases. The quality of conclusions drawn from model studies depend on two key issues; first, the quality of biology that has gone in the process of developing the model structure; second the quality of the data that go into the estimation of the parameters and the quality and quantity of the data that go into model validation. The more real world data that are used in the model building process, the more likely that modeling studies will provide novel, innovative and valid results.

1. Introduction

Johne's disease (JD), or paratuberculosis, is a chronic enteric disease of cattle and other ruminants due to an infection with *Mycobacterium avium* subspecies *paratuberculosis* (MAP) (Behr and Collins, 2010). Herd-level MAP infection prevalence has gradually increased in the past decade; in a recent survey, it was found that 68% of US dairy herds have apparently at least one cow that is infected with MAP. This estimate was obtained from a survey published by the USDA's National Animal Health Monitoring System in 2007 (NAHMS, 2007). The economic impact of MAP infections to the dairy industry in the United States varies but the cost to dairy producers was estimated to be more than $200 million per year (Ott et al., 1999). More recently the concerns that MAP may pose a threat to public health, as MAP infections in humans have been associated with an increased risk of Crohn's disease in humans (Behr and Collins, 2010).

* Corresponding author at: GD Animal Health, Arnsbergstraat 7, 7418 EZ Deventer, The Netherlands. Tel.: +31 570 660564; fax: +31 570 634104.
E-mail address: y.schukken@gdanimalhealth.com (Y.H. Schukken).

http://dx.doi.org/10.1016/j.prevetmed.2014.12.009
0167-5877/

The epidemiology of MAP in dairy herds is difficult to study as the infection shows a very slow progression from initial infection to clinical disease. Many infected animals never show clinical signs and, even more, many infected animals are only detected using diagnostic tests a few years after initial infection or are actually never detected (NRC, 2003). Even more, under commercial farm circumstances, diagnostic testing is infrequent and there is a low diagnostic test sensitivity for animals shedding either intermittently or low levels of MAP (Nielsen and Toft, 2008). Hence, precise information on the infection status of animals is difficult to obtain. The best data will likely be obtained from animals that are studied during their complete lifetime on commercial dairy farms. Particularly if lifetime follow-up is followed by obtaining relevant tissues at slaughter and bacteriological culture of these tissues with a known predilection of MAP infection (Whitlock et al., 2005b).

Other methods to try and understand the infection biology and epidemiology of MAP infection include experimental challenge studies, in vitro studies, model studies or mathematical modeling of population infection dynamics of MAP (Collins and Morgan, 1992). Mathematical models for JD in dairy cattle have been developed to understand MAP transmission dynamics and to predict the effectiveness of recommended control programs (Kudahl et al., 2007; Lu et al., 2008, 2010; Mitchell et al., 2008; Marce et al., 2010). Mathematical models have the advantage of being able to study infectious dynamics under a wide range of scenarios, without necessarily having to obtain perfect data for all these scenarios. Of course, mathematical models are limited by their imperfect ability to provide a realistic representation of the true infection dynamics in herds (Pouillot et al., 2004). The more realistic a model as far as the infection biology and epidemiology of MAP, the more realistic the obtained observations and drawn conclusions. Both these components, infection biology and epidemiology, in the case of MAP infections are not fully known and understood and therefore models may contribute to the study of MAP. Care should be taken with regard to the conclusions drawn from such models. Clearly, models are only as good as the input of biology and epidemiology that goes into model (Dorshorst et al., 2006). Particularly, the parameter estimates that are used in the modeling need to be based on relevant knowledge and data (Whitlock et al., 2005b). In the current paper, we will assume that the models we develop are sufficiently general and flexible that they may accommodate our current knowledge of MAP biology and infection pathways. This model was initially developed and described in detail by Mitchell et al. (2008) and subsequently used in many stochastic and deterministic simulation studies (Lu et al., 2008, 2009, 2010, 2013). The model was also reviewed by Marce et al. (2010) and identified as one of the more versatile and complete MAP models, with the possible exception of an external environmental contribution to MAP infection transmission. The objective of this paper is to present parameter estimates of existing state transition models obtained from real life data and to compare results of simulation models with detailed real life data. 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.

2. Model description

The state-transition model depicting MAP infection dynamics as shown in Fig. 1, is a state transition model with three age groups. These age groups are animals from birth to 1 year of age, from 1 to 2 years of age and finally adult animals aged 2 or more years. Transitions between age groups (vertical arrows) happen at a rate of aging, essentially 1 step per year until the animals are in the adult age group. Infection takes place mostly in the youngest age group, but is not excluded in older age groups either. Initially animals when infected at young age, enter into a transient shedding state. From transient shedding they move to a latent state where no shedding takes place. Eventually animals may exit the latent state and move to low shedding and then progress to high shedding. These states and transitions are in line with our current understanding of MAP infection biology (Behr and Collins, 2010; Windsor and Whittington, 2010; Harris and Barletta, 2001), and allow some further infection mechanisms that are currently only hypothesized. Examples of such currently hypothesized transmission routes are adult infection, transient shedding in one year olds, and the existence and importance of super shedders (Whitlock et al., 2005a; VanRoermund et al., 2007).

As indicated, assuming that the model structure may be a reasonable reflection of biological reality, the key will be to parameterize this model with parameters that are grounded in reality. Particularly, it is essential that point estimates and estimates of variability in these parameters are based on true infection situations that were observed under realistic circumstances. Among the biologically and epidemiologically important parameters are γ and λ for new infections, and σ and ν for progression of existing infections (see Fig. 1 and Table 1).

In the past, infection dynamics of MAP in herds was essentially assumed to be clonal, meaning that all infections were due to a single strain of the MAP bacteria (Collins and Morgan, 1992; Groenendaal et al., 2003). In more recent studies it has become clear that this is not always the case (Pradhan et al., 2011). Here, we will discuss the value of using strain specific estimates of transition. When no distinction is made between strains in observational studies, we observe the average transition of isolates. This average transition may be a combination of strains that are not very contagious and a few strains that are more contagious. This indicates that multiple transmission systems may be going on in the same herd. The most frequent method for MAP strain typing is based on sequencing of multilocus short-sequence-repeats (MLSSR). This sequence based method is a highly discriminatory method that has been used for typing MAP isolates and many other bacteria (Harris et al., 2006). While only a limited number of cross-sectional studies have used this method, and with a restricted set of isolates, it has been recognized that the use of well-designed longitudinal studies using several herds in multiple states is essential for applying the MLSSR sequencing technique to