

Mathematical modelling in veterinary epidemiology: why model building is important

Mart C.M. de Jong

*Department of Pathobiology & Epidemiology, Institute for Animal Science & Health (ID-DLO), P.O.Box 65,
8200 AB Lelystad, Netherlands*

Accepted 27 April 1995

Abstract

Some consider modelling to be very important for (veterinary) epidemiology, others severely criticise the use of modelling. Before joining this heated debate it is worthwhile to reflect on the role of mathematical modelling. Mathematical modelling is useful for the study of complex phenomena, like the population dynamics of infectious agents, because models show how separate measurements can be seen as manifestation of the same underlying processes. To build models that can act as connecting theories, careful model building is very important. It is shown how modelling helped to understand how transmission depends on underlying factors. Through a process of careful model building and comparisons of different model assumptions and model predictions with data one hypothesis was falsified and therewith the plausibility of another strengthened. In conclusion, the gain of the modelling was not the resulting model, but instead the insight into the population dynamics of infectious agents that was obtained in the process of model building and model analysis on the one hand, and interpreting experimental and observational data on the other.

Keywords: Infectious diseases; Mass action; Reproduction ratio; Transmission

1. Introduction

Modelling receives much criticism. This criticism is partly due to the naive promotion of 'realistic' simulation models. Promotion of realistic models often prompts the kind of criticism as worded by Bart et al. (1983) with regard to measles models:

The resulting equations sometimes mirror the observed events, but to date had little impact on disease control and preventive practice

The criticism may also be due to a lack of understanding by the critics of how models can be useful. Take for example the criticisms in the editorial by Stille and Gersten (1978) for the *Journal of Infectious Diseases*. They state while discussing mass action models:

... we must regard these models as tautologies or self-contained arguments and not scientific theories.

Certainly, it is true that these models are tautologies, but tautologies are very useful as scientific theories. Rather than starting a philosophical discussion on the relation between models and the world as we experience it, I will only discuss why model building is important for the study of the population dynamics of infections. For an interesting discussion of mathematical models in relation to the observable world, the reader is referred to Schwartz (1992).

Parasitologists and microbiologists are involved in the study of the population dynamics of infectious agents, as they study the characteristics of these agents. In their studies, they focus on characteristics that are important in the agent's interaction with its host. These characteristics are, however, not sufficient to understand the population dynamics of the agent in the host population, because the behaviour of the host population is not taken into account. Moreover, the characteristics that are studied need not to be those characteristics that are important for the population dynamics. For example, in vaccine testing the protection offered to the individual is measured by a vaccination-challenge experiment. However, characteristics as infectivity, susceptibility and host behaviour that determine the transmission of the agent are not measured and therefore these experiments give little information on the effectiveness of the vaccine in a population.

As an alternative to the vaccine testing approach epidemiologists used methods, originally developed for the study of chronic diseases, to study the effect of control measures on the population dynamics of infections. For instance, vaccine efficacy is then calculated from the attack rates observed in vaccinated and non-vaccinated subjects in the same population. These attack rates can be measured as incidence of clinical disease or incidence of a positive test result. Also this epidemiological approach has met with criticism, the main one being that the statistical approaches rely on the assumption that observations are independent. However, incidences in different groups within the same population are dependent. Dependence of incidence is the very property that defines infectious diseases. Therefore, as Koopman et al. (1991) summarized it:

However, the usual formulation of risk assessment parameters, ..., which are so useful in chronic disease epidemiology, do not provide stable assessment of risk for factors that affect contagion.

Of course, if the vaccine is applied to protect a single individual against clinical symptoms, the problems with interdependence still occur but these problems may be solved by using other statistical methods (Haber et al., 1991; Halloran et al., 1992). It is then not necessary to look to estimate other parameters like transmission rates.

If, however, transmission is important mathematical models can be used as a theory on how transmission of infectious agents occurs, and experimental data (microbiology and parasitology) and the observational data (statistical epidemiology) can then be interpreted in the light of this theory. Hurd and Kaneene (1993) concluded that for the estimation of factors, that affect transmission, process models are needed. However, Hurd and Kaneene (1993) also concluded that these process models have to be based on a systems approach to modelling. I wish to refute this conclusion. The problem with the systems approach is that the importance of careful model building is neglected. Model building should be based on measurable quantities and thus arrive at testable theories on how transmission occurs.

Insight is gained by comparing predictions from simple models to observable data in a comparative manner (Dye, 1992). Here, the approach based on model building and model analysis on the one hand and the analysing of experimental and observational data on the other is illustrated with an analysis of some of the factors that influence how an infectious agent is transmitted in a population.

2. Model building

When we study the dynamics of the transmission of an infectious agent, we want to know which factors cause the infection rates (i.e. the number of infections that occur per unit of time) to differ. Mathematical models can be used to unravel how different factors contribute to the observed differences. If the contribution of a particular factor is linear and additive to all other factors the relationship will be relatively easily unravelled even without mathematical models.

However, the main factors affecting infection rates are not additive but multiplicative and thus mathematical models are very much needed. The effects of this multiplicative or non-linear dependence are seen when the underlying factors are varied: in one part of the range the infection rate changes dramatically whereas outside that range little change is observed. Also in different realisations of the process with the same starting conditions the effect of the non-linearity is seen: observed infection rates do not vary in a simple unimodal way.

These non-linearities can be demonstrated with a simple model for which analytic solutions can be found. Furthermore, even this relatively simple model can explain much of the observed variability in infection rates.

The simple model that is used here is the stochastic susceptible–infectious–removed (SIR) model, very much in the spirit of Becker (1989) but with an important modification. As variables we have: S , the number of susceptible animals; I , the number of infectious animals; R , the number of immune animals and N , the total number of animals. There are two events each with its own rate at which it occurs (Table 1).

The model in Table 1 differs from Becker's models (Becker, 1989) in that the transmission rate is written as $\beta SI/N$, instead of βSI . The expression chosen to represent the rate at which transmission occurs, is derived from chemical reaction kinetics. In chemical reaction kinetics the reaction rate is proportional to the product of the concentrations of the two substances that react (law of mass action). The parallel in the case of an infection in a group of animals is: the density of new infections per unit of time is proportional to the product of the density of susceptible animals (s) and the density of infectious animals (i). If the density of all animals in the group (susceptible, infectious and immune) is constant, the

Table 1

Description of the stochastic susceptible–infectious–recovered model: in continuous time events occur with rates (probability per unit of time) that depend on the current state (S, I) of the population

Event	Symbolic representation	Rate
Infection	$(S, I) \rightarrow (S-1, I+1)$	$\beta SI/N$
Recovery	$(S, I) \rightarrow (S, I-1)$	αI

area on which the animals are kept is proportional to the population size N . The rate of change in I , rather than in i , is then proportional to SI/N . To distinguish between the two formulations the formulation SI/N will be called true mass action and the formulation SI will be called pseudo mass action. (see De Jong et al., 1995; Diekmann et al., 1995; Bouma et al., 1995).

The same is true for the Reed–Frost formulation used in the paper by Hurd and Kaneene (1993):

$$C_{t+T} = S_t(1 - q^{C_t}) \quad (1)$$

where, C_t is the number of cases at moment t , S_t the number of susceptible animals at moment t , q the probability that a susceptible will not become infected in an time interval of length T when there is one infectious individual present. If one takes q to be independent of population size, one is not using the true mass action formulation. The true mass action formulation of the Reed–Frost model is:

$$C_{t+T} = S_t(1 - e^{-\beta C_t I / N}) \quad (2)$$

and thus the infection rate does depend on the population size.

Thus, model building can lead to two different model formulations: the true mass action formulation and the pseudo mass action formulation. From the parallel to the chemical reaction kinetics the true mass action formulation seems the more plausible one. However, the most plausible formulation is not necessarily the one that is in agreement with the observable reality of real infections in real host populations. To answer the question which formulation is applicable to the population dynamics of real agents in real populations a two step approach will be used: first it is shown by model analysis that the formulations lead to different predictions and second these predictions are compared with data.

3. Model analysis

Analysing models means establishing a relation between the assumptions and the predictions. This relation is necessarily a tautology, i.e. given that the assumptions are true the predictions are necessarily true. (This is all that mathematics does: produce tautologies.) As a consequence when the predictions do not correspond to the observations also (some of) the underlying assumptions are not true. In the case of the population dynamics in a closed group of animals the total number of animals becoming infected (final size of the outbreak) can be derived from the model. The final size for the different model formulations can be shown to differ, and thus by comparison with data conclusions regarding the correct formulation can be drawn. It is because the derived relations are tautologies that analytic mathematical methods are so very useful in scientific theories.

The full solution to the stochastic SIR model described above, namely for each moment in time a list of all possible combinations of S and I (S, I) and their probabilities, is not yet known. To calculate these probabilities the parameter values and the start condition have to be given. The start condition is, for example, that in a group of ten animals one is infected or in symbols that at $t=0$, $(S, I) = (9, 1)$ with probability one (1) and all other (S, I) combinations with probability zero. Although the full solution is not known several other results have been obtained regarding this model:

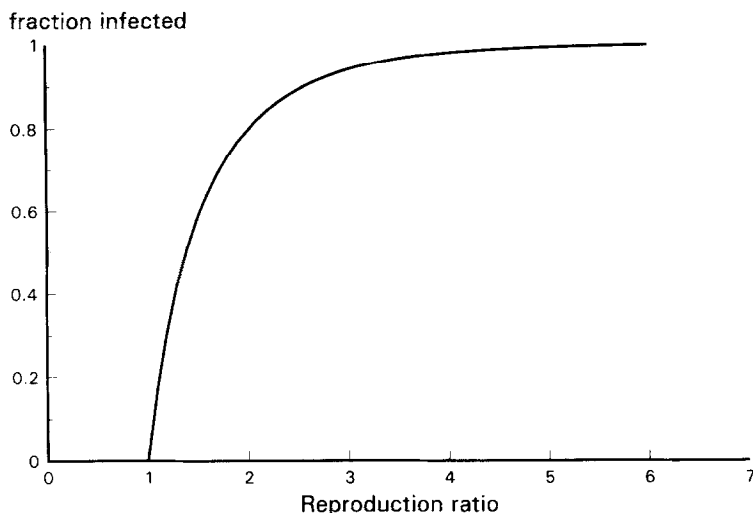


Fig. 1. The famous threshold theorem: the final size, fraction eventually infected animals in a population plotted against the reproduction ratio R_0 .

(i) For S and I both large (mathematically the conclusions are derived for infinitely large populations; in practice the conclusions hold for moderately high values of S and I), a deterministic approach is valid and the final size (q = fraction of the original population that eventually becomes infected) is given implicitly as:

$$q = 1 - e^{-R_0 q} \quad (3)$$

with:

$$R_0 = \beta / \alpha \quad (4)$$

The expression (4) for R_0 is true for the true mass action formulation; for the pseudo mass action formulation R_0 increases linearly with population size N .

The relationship (3) between average final size and parameters has the striking non-linearity discussed previously (see Fig. 1): for $1 \leq R_0 \leq 4$ there is a dramatic change in final size and outside this range either the final size is zero or it is approximately equal to one. Clearly there is a threshold ($R_0 = 1$) in the parameter space: for $R_0 \leq 1$ the infection does not spread in the populations whereas for $R_0 > 1$ the infection will spread in the population. This deterministic threshold condition was first derived, for a more general model, by Kermack and McKendrick (1927; see also Diekmann et al., 1990 and Diekmann et al., 1995).

(ii) For S large, a branching process approximation is valid. (The conclusion is derived for infinite S , but the conclusion is a good approximation for moderately high values of S .) The branching process is again a stochastic process with the number of infectious animals as the single variable. The result from the analysis is that the number of infectious animals will either become zero, and nothing can happen further, or the number of infectious animals will continue to grow indefinitely. The first set of realisations, when I becomes zero,

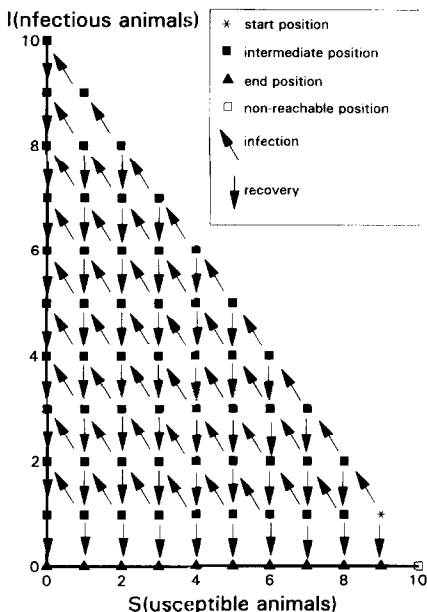


Fig. 2. The stochastic SIR model: each node is a population state. Here all states are shown that can occur after $(S, I) = (9, 1)$ at $t = 0$. At $t \rightarrow \infty$ each population will be on the x-axis (states indicated by triangles). Hence, the stochastic model is characterised by a final size distribution, that is the probability distribution over these states.

correspond in a finite population to minor outbreaks, and the other realisations, where I grows indefinitely, correspond in a finite population to major outbreaks. The probability of a minor outbreak (p), assuming that the number of infectious individuals caused by one infectious individual is Poisson distributed with expected value R_0 , is given by the smallest non-negative root of:

$$p = e^{-R_0(1-p)} \tag{5}$$

For further explanation see Cox and Miller (1965, pp. 104–105). Compare this to Eq. (3) using $p = 1 - q!$

As a result the stochastic threshold condition is:

$R_0 \leq 1 \quad p = 1$ i.e. only minor outbreaks occur;

$R_0 > 1 \quad p < 1$ i.e. minor and major outbreaks occur.

This forms a nice parallel to the deterministic threshold condition.

(iii) If S is not large, the above approximations do not suffice and an algorithm (computer programme) can be derived to calculate the exact final size distribution (De Jong and Kimman, 1994), that is the probabilities for all possible final sizes for a small group. These final outcomes are the number of animals that have become infected when the outbreak has ended, i.e. when again $I = 0$ (see Fig. 2). With the algorithm, one can easily calculate the distributions for values of R_0 .

In Fig. 3 examples of final size distributions for three different values of R_0 are shown. It shows that the conclusions from the deterministic approximation and from the stochastic branching process approximation are still qualitatively valid: for $R_0 < 1$ only minor outbreaks

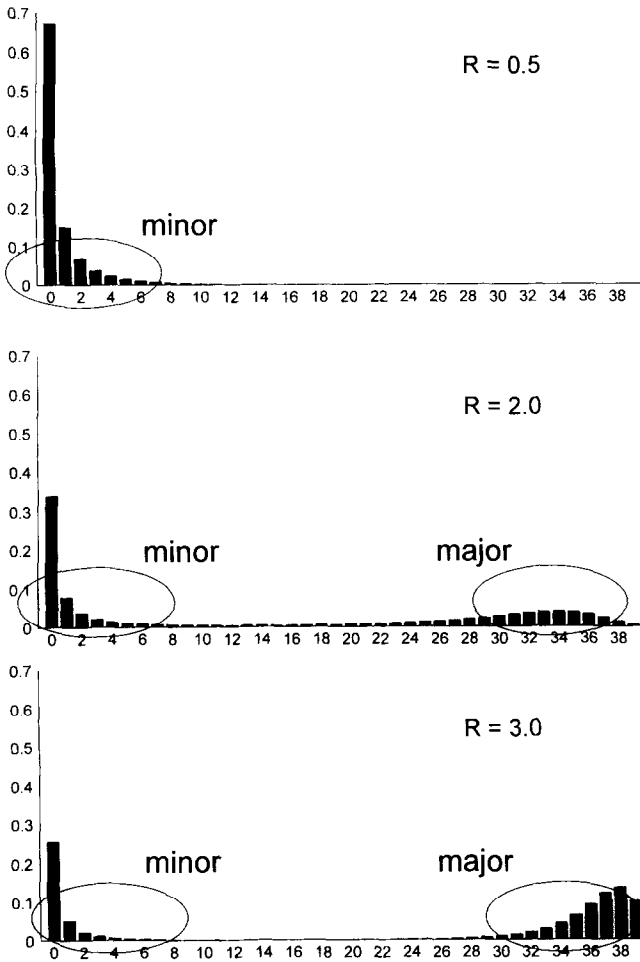


Fig. 3. The final size distribution of the stochastic SIR model for three values of the reproduction ratio R .

are possible; for $R_0 > 1$ both minor and major outbreaks are possible and moreover the larger R_0 the higher the probability on a major outbreak and the higher the number affected by a major outbreak.

The two formulations of the model, i.e. the true mass action formulation and the pseudo mass action formulation differ not regarding these conclusions, but they differ in how R_0 depends on the population size. According to the pseudo mass action assumption formulation R_0 increases linearly with the population size whereas according to the true mass action assumption R_0 is independent of the population size.

4. Comparison to data

Thus to make inferences on which formulation is 'correct' the final sizes for different population sizes have to be observed. Because of the relations established in the previous

section it is clear that according to pseudo mass action the final size as a fraction of the total population increases with increasing population size and according to the true mass action formulation it remains constant. We now seek to falsify either one of these formulations.

Agreement between theory and data is in itself not proof of the validity of the theory. One can only arrive at conclusions by making comparisons and thus falsifying alternative hypotheses. Take, for example, the classical paper by Anderson and May (1979). In this paper a pseudo mass action assumption (βSI) is used. Often (e.g. Smith and Grenfell, 1990; Heide-Jørgensen and Härkönen, 1992) it is thought that such models necessarily imply that there exists a threshold population size, i.e. a number of animals, below which the infection will not spread. De Jong et al. (1995) showed that an equally good fit is obtained by using the true mass action assumption ($\beta SI/N$) in the Anderson and May model. A very nice aspect of Anderson and May's approach was the idea to use the equilibrium as basis for comparison: using the equilibrium allowed the use of the deterministic approximations.

An alternative to the use of the equilibrium as basis for comparison is the use of the final size as basis for comparison. When comparing the final size of different experimental or observational situations the above results of the model analysis are useful.

For example De Jong and Kimmman (1994) compare the final size in experimental groups of vaccinated and non-vaccinated pigs in order to compare transmission of pseudorabies virus with and without vaccination. This comparison utilises the non-linearities in the relation between final size and R_0 . Firstly, of all a marked difference in final size was expected because it was likely that the R of the vaccinated group would be below threshold and the R_0 of the non-vaccinated above threshold. Secondly, the probability that minor outbreaks would occur while the R or R_0 would be still above threshold was minimised by starting each experiment with a 50%–50% mix of infectious and contact-exposed animals. Thus, with the stochastic SIR model as basis of statistical comparison, a high power was obtained in spite of the relatively small experiments.

Heide-Jørgensen and Härkönen (1992) studied phocine distemper virus in seals comparing the outcome in groups with different sizes (Fig. 4). For their purpose the deterministic approximation was valid, because the groups were reasonably large (S was large) and reintroduction of phocine distemper virus in the groups was very common (I was large). The latter could be concluded because no minor outbreaks were observed. Note for the comparison that the final size in Fig. 1 is given as a fraction and thus the true mass action model leads to an average final size that is a linear function (through the origin) of the population size (see Fig. 4). In contrast, for the pseudo mass action model, the following transformation of the fraction infected would be a linear function of the population size (as it is a deterministic estimate of R_0):

$$\frac{-\ln(1-q)}{q} \quad (6)$$

Estimates of R_0 for phocine distemper virus in seal populations of different sizes are shown in Fig. 5. Clearly, R_0 does not show a linear increase with population size as predicted from models with the pseudo mass action formulation (see also Diekmann et al., 1994).

In the case of the seals we assumed that introduction of the virus was frequent: i.e. no minor outbreaks were detected. However, when introduction is not frequent both minor and

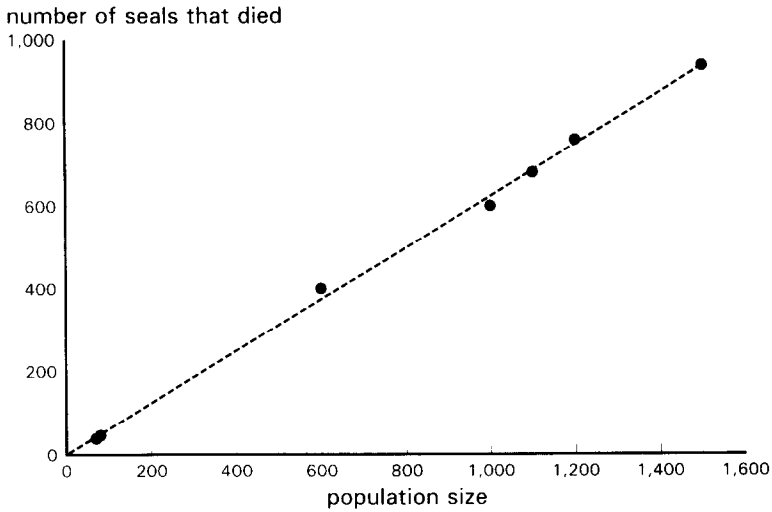


Fig. 4. The number of seals that died (data of Heide-Jørgensen and Härkönen, 1992) plotted against the population size. As was already noted by the authors, there is a clear linear relationship. That relationship is in agreement with true mass action.

major outbreaks may be detected and then estimation of transmission cannot be based upon a deterministic model. For instance Stegeman et al. (1995) compared transmission in groups of pigs: in one stable there were pigs that were vaccinated once and in the other stable there were pigs vaccinated twice against pseudorabies virus. The comparison was made on 20 farms and from the resulting final sizes it could be seen that in both treatment groups there

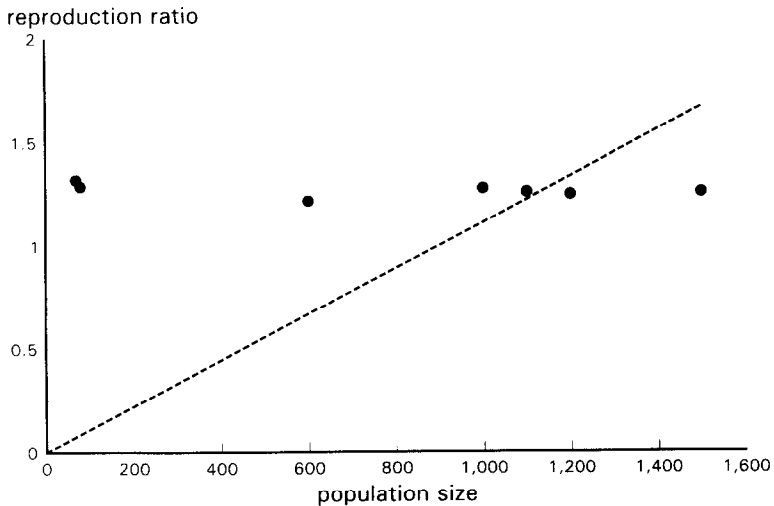


Fig. 5. The estimated R_0 (from data of Heide-Jørgensen and Härkönen, 1992) against population size. Contrary to the prediction based on the pseudo mass action assumption there is no linear relationship.

were minor and major outbreaks. As expected from the model analysis the final sizes in groups of pigs that were vaccinated once, compared with the final sizes in groups that were vaccinated twice, showed major outbreaks more often and these major outbreaks were larger. Stegeman et al. (1995) concluded that transmission was significantly higher in the groups that were vaccinated once.

5. Conclusions

Modelling, model building and model analysis, were discussed to illustrate the role modelling can play in veterinary epidemiology. The model that was discussed was a simple model for the population dynamics of infections and with this simple model the question of how transmission depends on population size was discussed. Many aspects that are important for the study of the dynamics of infectious agents have not been discussed: how to model contact structures and heterogeneity, which methods can be used for estimation and how can the role of control measures be evaluated.

All modelling has to start with model building which requires a proper appraisal of the basic tenets of the process that is studied. Subsequently, the model that was build has to be analysed and its predictions compared with data. It is, however, not sufficient to show that a model with a particular set of assumptions gives correct predictions. Other models, based on very different assumptions may give equally good predictions. A careful comparative approach to evaluate the main factors, both implicitly and explicitly included in the model, is necessary. Here, the example of how transmission depends on population size was discussed. It was shown that a good-fitting model by Anderson and May (1979) was probably based on the wrong assumption regarding transmission. In another paper a comparison of different assumptions was made using experiments (Bouma et al., 1995), and the same can be done using the data of Heide-Jørgensen and Härkönen (1992) on phocine distemper virus in seals. From all these analyses it followed that the true mass action formulation of transmission gives a better approximation to reality than the often used pseudo mass action formulation.

Furthermore, it was discussed how transmission involves essentially non-linear processes. Thus, unravelling which factors are important from experimental and observational data is very difficult and it is necessary to use mathematical models to assist the unravelling. It is even possible to exploit these non-linearities: the effect on transmission of vaccines that do sufficiently limit the transmission of the agent can be detected even in small-scale experiments (De Jong and Kimman, 1994). This approach may be extended to field observations to detect even more subtle effects of different vaccination schemes on transmission (Stegeman et al., 1995).

There is a clear opportunity to further develop methods to study infectious agents in populations. After the realisation that the risk assessment measures, developed for non-infectious disease in humans, cannot be used, the need to develop new tools is obvious. A good starting point to do this is the book by Becker (1989). Methods to evaluate the effect of control measures are especially important.

In addition, a careful examination of the assumptions underlying models is necessary. In all these aspects veterinary epidemiologists may make a contribution both to improve animal health and to improve our understanding of the epidemiology of infectious diseases.

Acknowledgements

The above ideas on modelling and its role in veterinary epidemiology have been provided from the cooperation with many modellers, microbiologists and epidemiologists, of whom I would like to thank especially Odo Diekmann, Hans Heesterbeek, Tjeerd Kimman, Arjan Stegeman, Annemarie Bouma and Arie van Nes.

References

- Anderson, R.M. and May, R.M., 1979. Population biology of infectious diseases: Part I. *Nature*, 280: 361–367.
- Bart, K.J., Orenstein W.A., Hinman, A.R. and Amler, R.W., 1983. Measles and models. *Int. J. Epidemiol.*, 12: 263–266.
- Becker, N.G., 1989. *Analysis of infectious disease data*. Chapman and Hall, London, 224 pp.
- Bouma, A., De Jong, M.C.M. and Kimman, T.G., 1995. Transmission of pseudorabies virus within pig populations is independent of the size of the population. *Prev. Vet. Med.*, 23: 163–172.
- Cox, D.R., Miller, H.D., 1965. *The theory of stochastic processes*. Chapman and Hall, London, 398 pp.
- De Jong, M.C.M. and Kimman, T.G., 1994. Experimental quantification of vaccine-induced reduction in virus transmission. *Vaccine*, 8: 761–766.
- De Jong, M.C.M., Diekmann, O. and Heesterbeek, J.A.P., 1995. How does transmission depend on population size? In: D. Mollison (Editor), *Epidemic Models: Their Structure and Relation to Data*. Cambridge University Press, Cambridge, pp. 84–94.
- Diekmann, O., Heesterbeek, J.A.P. and Metz, J.A.J., 1990. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.*, 28: 365–382.
- Diekmann, O., De Jong, M.C.M., De Koeijer, A.A. and Reijnders, P., 1994. The force of infection in populations of varying size: a modelling problem. *Proceedings of the 2nd European Conference on Mathematics Applied to Biology and Medicine*, Lyon 1993.
- Diekmann, O., Heesterbeek, J.A.P. and Metz, J.A.J., 1995. The legacy of Kermack and McKendrick. In: D. Mollison (Editor), *Epidemic Models: Their Structure and Relation to Data*. Cambridge University Press, Cambridge, 95–115.
- Dye, C., 1992. The analysis of parasite transmission by bloodsucking insects. *Annu. Rev. Entomol.*, 37: 1–19.
- Haber, M., Longini, I.M. and Halloran, E., 1991. Measures of the effects of vaccination in a randomly mixing population. *Int. J. Epidemiol.*, 20: 300–310.
- Halloran, M.E., Haber, M. and Longini, I.M., 1992. Interpretation and estimation of vaccine efficacy under heterogeneity. *Am. J. Epidemiol.*, 136: 328–343.
- Heide-Jørgensen, M.P. and Härkönen, T., 1992. Epizootiology of the seal disease in the eastern North Sea. *J. Appl. Ecol.*, 29: 99–107.
- Hurd, H.S. and Kaneene, J.B., 1993. The application of simulation models and systems analysis in epidemiology: a review. *Prev. Vet. Med.*, 15: 81–99.
- Kermack, W.O. and McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. A*, 115: 700–721.
- Koopman, J.S., Longini, I.M., Jacquez, J.A., Simon, C.P., Ostrow, D.G., Martin, W.R. and Woodcock, D.M., 1991. Assessing risk factors for transmission of infection. *Am. J. Epidemiol.*, 133: 1199–1209.
- Schwartz, J., 1992. *Creative moment: how science made itself alien to modern culture*. Cape, London, 252 pp.
- Smith, G. and Grenfell, B.T., 1990. Population biology of pseudorabies in swine. *Am. J. Vet. Res.*, 51: 148–155.
- Stegeman, A., Van Nes, A., De Jong, M.C.M. and Bolder, F.W.M.M., 1995. Assessment of the effectiveness of vaccination against pseudorabies virus in finishing pigs. *Am. J. Vet. Res.*, 56: 573–578.
- Stille, W.T. and Gersten, J.C., 1978. Tautology in epidemic models. *J. Infect. Dis.*, 138: 99–101.