

# Risk analysis in relation to the importation and exportation of animal products

D.J. Vose

Risk Analysis Services, 18 Mill Street, Wincanton, Somerset BA9 9AP, United Kingdom

## Summary

The design of a quantitative risk analysis model has to be dictated by the questions it seeks to answer. The model should also be as objective as the available data will allow. Animal and animal product import risks usually have three characteristics which make the design of a good quantitative risk analysis model quite difficult, namely:

- the probabilities of the steps leading to the undesired outcome are frequently inter-related
- the probability of the undesired outcome itself is in many cases very small, making direct simulation impractical
- important variables within the model often cannot be quantified through analysis of data, thus these variables must be modelled with probability distributions to reflect the degree of uncertainty, usually determined by expert opinion.

This paper provides a tutorial on some modelling techniques which are essential to the risk assessment of animal and animal product imports and which help overcome these problems. A number of probability distributions, their uses and inter-relationships, are examined. The application of these distributions, coupled with some general modelling techniques, is then demonstrated to produce rigorous and transparent animal import risk analyses.

## Keywords

Model - Monte Carlo simulation - Quantitative risk assessment - Risk analysis - Scenario pathway - Simulation - Techniques.

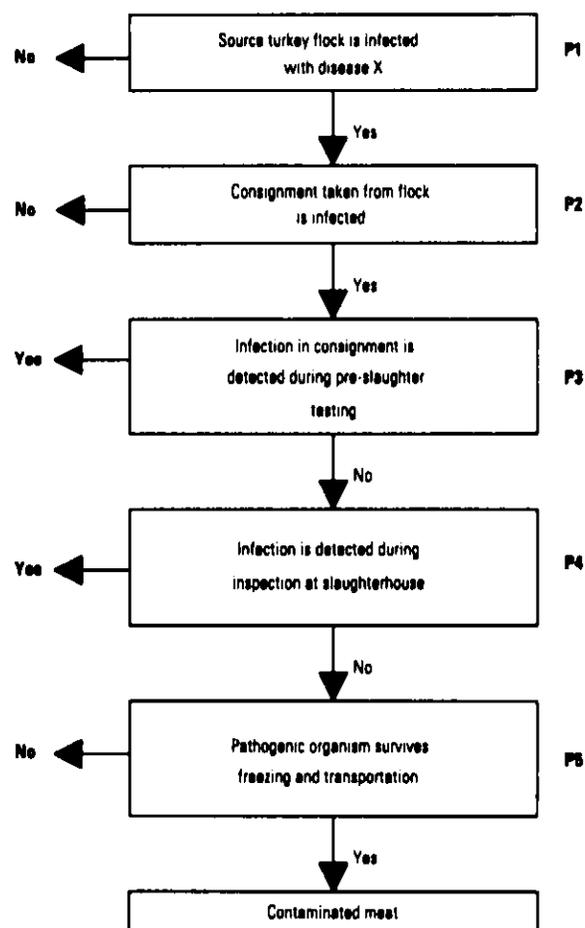
## Introduction

Two closely related quantitative risk analysis modelling techniques lend themselves to estimating the risks associated with importing animal products, namely: scenario pathway and simulation. The scenario pathway approach is the most commonly seen in published work to date. It seeks to determine a pathway of events which would ultimately lead to the importation of an infected product. A typical example is shown in Figure 1. The pathway begins by looking at the probability that an animal or a group of animals is infected at source. Then one calculates:

- the probability that the infected animal(s) will not be identified as being infected
- the probability that the infection will remain undetected at each of a number of (valid) inspection points

- the probability that the infecting organism will survive processing, storage and distribution.

Each probability along the chain is a conditional probability, i.e. the probability of the occurrence of that event given that all the previous events have occurred. So, for example, the probability of detection at pre-slaughter testing (P3 in our example) is the probability that the consignment of turkeys will be identified during pre-slaughter testing as being infected given that it is infected. The analysis should only include steps in the chain of events which have some procedural basis. For example, it would be meaningless to put in steps like 'Truck driver does not spot infection whilst loading turkeys' and 'Truck driver again does not spot infection whilst unloading turkeys' since there is no call on the truck driver to do so in his or her job description. Adding such steps would, of course, reduce the probability of the adverse final outcome, but without any logical justification.



**Fig. 1**  
Pathway of the animal product import risk analysis modelled in this paper

The second quantitative risk analysis technique is that of simulation. In this case, instead of modelling the probability of a cow being infected, for example, the number of cows which would be infected by randomly sampling from probability distributions is simulated. This way of thinking can be continued right the way through the model. Thus, whilst the scenario pathway technique explicitly calculates the probability of each possible scenario, the simulation approach lets each possible pathway be generated as a natural consequence of the random simulation. The two techniques are very closely related and have their own strengths and weaknesses. The strengths of the scenario tree approach are that it requires relatively little computing time and it can easily calculate extremely low probabilities. The weaknesses of the scenario tree approach are that it is not very intuitive and it is therefore easy to make a mistake, and the approach cannot always easily incorporate certain types of important correlation between distributions.

The problem shown schematically in Figure 1 will be modelled here using both techniques for comparison. However, it will be very useful to look first at a number of important probability distributions – the tools of the

quantitative risk analyst – which are used in some form in both the scenario pathway and the simulation techniques.

## Probability distributions: a tutorial

This section reviews a number of probability distributions which are of particular value in animal product import risk assessments. They are equally applicable to plant and plant product import problems and to food safety issues. This section begins by looking at two random processes: binomial and Poisson, from which the binomial and Poisson distributions and others are derived. The probability distributions which describe these processes form the majority of the technically based distributions which are needed in animal import risk analysis. After reviewing these two processes, the normal and hypergeometric distributions which are also very useful in particular circumstances will be discussed. This tutorial is, by necessity, a rather brief introduction but more detailed descriptions and examples can be found in a guide published by Vose (2).

### Poisson and binomial processes

Two stochastic processes, binomial and Poisson, form a large part of the structure of nearly all animal product import risk analyses. The binomial process describes a system where there are a definable number of trials ( $n$ ), a probability of success of that trial ( $p$ ) and a consequent number of successful trials ( $s$ ). The assumption of a binomial process is that all trials are independent, i.e. that each trial has the same probability of 'success' as the trial before it, regardless of the outcome of previous trials. Examples of binomial processes are tossing a coin ten times and seeing how many 'heads' come up, testing a group of infected animals and seeing how many show up positive using the test (assuming a test sensitivity of neither zero nor one) or randomly selecting a certain number of chicken wings from the supermarkets of a particular country and seeing how many are contaminated with *Salmonella*.

The Poisson process describes a system where there is a continuum of opportunity of an event occurring (as opposed to the  $n$  distinct trials for the binomial process). The Poisson process has one descriptive parameter,  $\beta$ , which is the mean number of occurrences of the event per unit of exposure. The  $\beta$  from a Poisson process is analogous to the probability  $p$  from the binomial process and the period of exposure  $t$  of the Poisson process is analogous to the number of trials  $n$  of the binomial process. Examples of Poisson processes are how many *Giardia* cysts the population of a city consumes from its water supply in a year, how many fish one catches in a day of fly-fishing, and the number of times a person is mugged in one year on the streets of a particular city. However, these would only remain Poisson processes if the city did not

improve its water treatment after any occurrence of *Giardia*, the fisherman got no better (or worse) at catching the fish during the day, and the mugged person kept following his or her same habits no matter how many times he or she was mugged.

### The distributions of the binomial process

The binomial process is characterised by the probability  $p$  of an event occurring at each trial. Once  $p$  has been estimated, it is a simple process to calculate other variables associated with the binomial process. The distributions are presented with their parameters in the standard form used by most commercial spreadsheet and Monte Carlo simulation software products.

- The distribution of the number of events that occur in  $n$  trials = Binomial( $n, p$ ).
- The number of trials needed for the event to occur for the first time = 1 + Geometric( $p$ ).
- The number of trials needed for  $s$  events to have occurred =  $s$  + Negative Binomial( $s, p$ ).

### Estimating probability $p$ from an observed number of events in a specific number of trials

Suppose that there is a need to determine the probability of occurrence of a specific event. An observation has been made that the event has actually happened  $r$  times out of a possible  $n$ . Its true probability of occurrence ( $p$ ) is modelled using the Beta distribution as:

$$p = \text{Beta}(\alpha_1, \alpha_2)$$

where:

$$\alpha_1 = r + 1$$

$$\alpha_2 = n - r + 1.$$

This use of the Beta distribution is, in fact, an application of Bayes' Theorem where no prior knowledge of  $p$  is assumed, i.e. the prior distribution of  $p$  is Uniform(0,1) – meaning  $p$  is equally likely to be any value between 0 and 1.

### Example

100 ( $n$ ) birds were randomly selected from a very large flock of turkeys, 17 ( $r$ ) were determined to be infected with *Salmonella*. An assumption is made that the sensitivity and specificity of the test are 100%. The true prevalence of *Salmonella* within the flock ( $p$ ) can be estimated as:

$$p = \text{Beta}(17+1, 100-17+1) = \text{Beta}(18, 84).$$

Figure 2a illustrates how in practical terms  $p$  is tightly contained between 0.08 and 0.30 with a strong peak at 0.17, though in theory the value ranges between 0 and 1. The more tests that are performed, the narrower the distribution would become, i.e. the more accurately  $p$  will be determined. Figure 2b shows how the distribution of  $p$  would narrow with

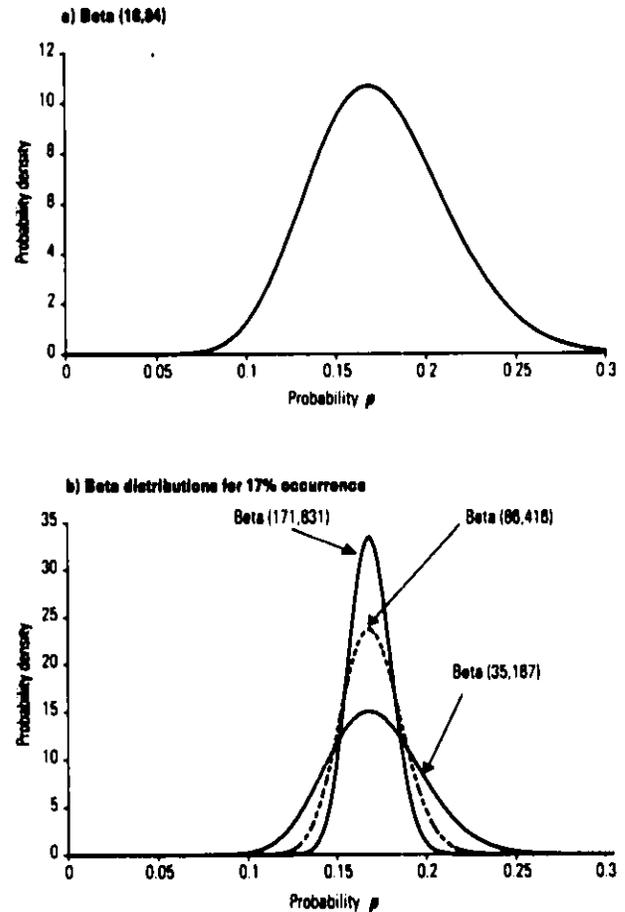


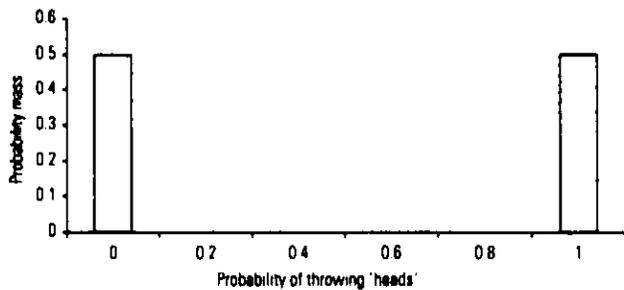
Fig. 2  
Distributions of probability of occurrence  $p$  given that 17% of trials succeeded and the number of trials varies

increasing number of trials (using the same percentage of occurrences). Understanding this behaviour can be very useful, for example, in planning future tests: the predictable reduction of uncertainty can be balanced against the extra cost and time required to complete any additional tests.

### The meaning of a distribution of probability

This use of the Beta distribution introduces the concept of a probability distribution of a probability – perhaps not something that is immediately intuitive and which deserves further explanation. Consider the following example:

A coin will be tossed. What is the probability of 'heads' showing? Assuming that the coin is fair, the probability should be exactly 50%. Before answering the question, however, the reader should also know that the coin has the same face on both sides. Now, since there is no indication as to whether both sides are 'heads' or 'tails', assigning equal probability to each would be reasonable. The probability of a 'heads' could still be described as 50%. Alternatively, the reader may state that the probability has equal chances of being 0% or 100%, which is the distribution shown in Figure 3 – a probability distribution of a probability. The



**Fig. 3**  
Distribution of the probability of throwing 'heads' when both sides of the coin are the same (but unknown)

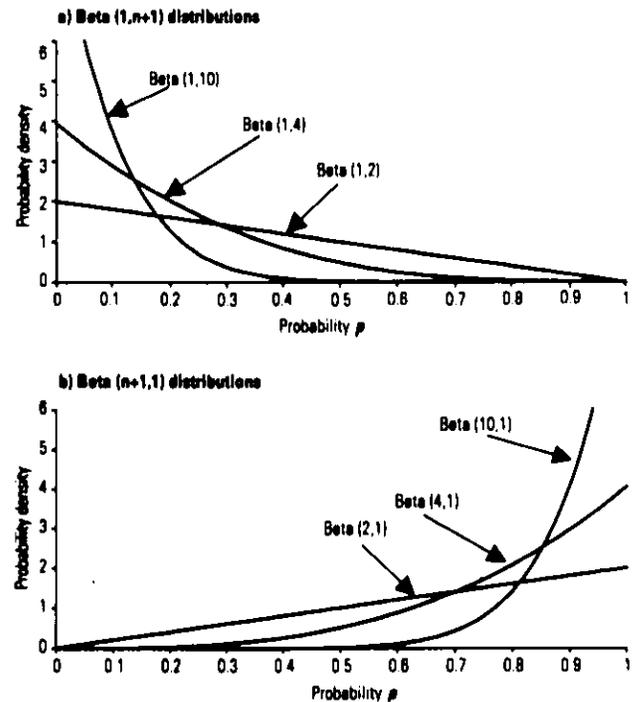
mean (average) of the distribution is equal to 50% – the first figure quoted, though the true probability in this case could never actually be that value. A probability may be described by a distribution, rather than a single value, only where there is a lack of knowledge of what that true probability actually is. The uncertainty described by a distribution of a probability never comes from the variability in the stochastic process itself. A failure to appreciate the differences between uncertainty from lack of knowledge and that from randomness has many implications in risk analysis modelling.

There are thus two forms of uncertainty in a risk analysis: the inherent uncertainty of the stochastic process being modelled, often described as variability; and the lack of exact knowledge we have of that problem, often described as uncertainty. So, for example, in the tossing of a fair coin we have precise knowledge of the system (the probability of 'heads' is exactly 50%): there is no uncertainty, but there still remains the inherent variability of what the outcome of a toss might be.

#### Estimating probability $p$ when there are no occurrences in a specific number of trials

Imagine, in the above experiment, that of the 100 tested turkeys, none tested positive, i.e.  $r = 0$ . To state categorically that there was no chance of a turkey being infected with *Salmonella* in the flock is untrue: in fact if  $p$  was less than or equal to  $1/101$ , zero positives would be more likely to be observed than any other number. However, a distribution of the probability of an infected turkey  $p$  can still be defined in the same manner as before using a Beta distribution. This obviously produces a pessimistic estimate since it assumes that the possibility of infection does exist and that, before the experiments, the prior opinion said that the prevalence could equally likely be anywhere between zero and one. The Beta distribution can be used again with  $\alpha_1 = r + 1 = 1$  and  $\alpha_2 = n - r + 1 = n + 1$ . Figure 4a shows how the distribution progressively favours a probability near zero with increasing number of tests  $n$ .

Exactly the same principle applies if a positive result has been obtained for every test, i.e.  $r = n$ :  $\alpha_1 = n + 1$  and  $\alpha_2 = 1$ . Figure 4b shows how this distribution is a mirror



**Fig. 4**  
Estimates of the probability of occurrence  $p$  of an event where  
(a) there were no successes, and  
(b) all trials were successful for 1, 3 and 9 trials

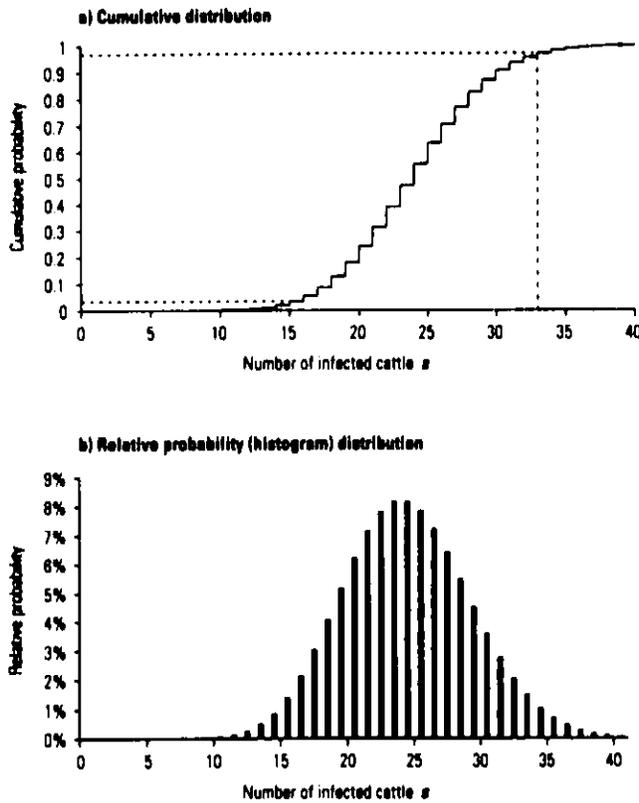
image of Figure 4a, progressively favouring a probability near 1 with increasing number of all positive tests  $n$ .

#### Estimating the probability of the occurrence of several events in a set of trials

The binomial( $n, p$ ) distribution calculates the number of events that will occur in  $n$  trials where there is a probability  $p$  of success in each trial.

#### Example

The quarantine officials of an importing country know that there is a 2% prevalence  $p$  of Johne's disease within a country from which an entrepreneur seeks to import cattle. The entrepreneur is intending to import 1,200 cattle  $n$  from this country. How many will be infected  $s$  with Johne's disease? Answer: binomial(1200,2%). Figure 5a shows the cumulative distribution function  $F(x)$  of this probability distribution: a very common representation of a probability distribution. It can be seen that there is about 3% chance that  $s$  will be less than 15 and 97% chance that  $s$  will be less than 33. These two values represent roughly the lower and upper 95% (roughly 97%–3%) confidence limits respectively. Figure 5b shows the relative probability distribution function  $f(x)$  for the same distribution. This type of plot allows one to see the relative likelihood of each allowable value. It is useful for offering a 'feel' of the uncertainty, though it is very limited in providing quantitative information.



**Fig. 5**  
**The binomial(1200,2%) distribution**

**Estimating the probability of at least one event in a set of trials**

The probability of no events in a set of  $n$  trials is  $(1 - p)^n$  i.e. the probability that trial 1 fails  $(1 - p) \times$  the probability trial 2 fails  $(1 - p) \times \dots \times$  the probability the  $n$ th trial fails  $(1 - p)$ . The probability  $P_1$  of at least one event in a set of  $n$  trials is therefore  $1 - (1 - p)^n$ . Where  $n$  is large ( $> 30$  or so) and  $p$  is sufficiently small such that  $np < 1$ , the approximation  $P_1 \approx np$  can be used. However, the accuracy of this approximation is entirely dependent on the values of  $n$  and  $p$  and it is better practice to use the full  $1 - (1 - p)^n$  formula, especially as modern computers make this very easy.

**Example**

Boneless beef portions imported from a particular country are estimated to have a  $1:10^6$  chance of being infected with foot and mouth disease (FMD) and getting past all screening tests. A supermarket chain wishes to import 2,000 of these portions a year. What is the probability that FMD will enter the country with those portions, if this import is permitted?

$$P_1 = 1 - (1 - 10^{-6})^{2,000} = 1.9998 \times 10^{-3}$$

The  $P_1 \approx np$  approximation would give a value of  $2 \times 10^{-3}$ .

**Estimating the number of trials until a specific number of events occur**

The Geometric( $p$ ) distribution estimates the number of unsuccessful trials which will have to be completed before the

first success occurs. In other words, it is a distribution of the number of unsuccessful trials before the first success. Thus the number of trials required for the first occurrence of an event equals  $(1 + \text{Geometric}(p))$ .

The Negative Binomial( $s,p$ ) distribution estimates the number of unsuccessful trials one will have to complete before  $s$  successes occur. Thus, in the same manner as the Geometric distribution, the total number of trials required for  $s$  occurrences of an event equals  $(s + \text{Negative Binomial}(s,p))$ .

**Example**

A veterinarian knows that, on average, one in every 11 pigs she tests will be infected with a particular disease. How many pigs will she have to test before she tests an infected pig and how many before she would have tested 25 infected pigs?

The prevalence of this disease  $p = 1/11 = 0.091$ . The number of pigs  $N_1$  she must test before she will have tested an infected pig can be estimated as:

$$N_1 = 1 + \text{Geometric}(0.091)$$

The number of pigs she will have to test  $N_{25}$  before she tests 25 infected pigs can be estimated as:

$$N_{25} = 25 + \text{Negative Binomial}(25,0.091)$$

However, the success or failure of the last trial is often unknown. For example, imagine that a flock of 1,000 chickens was tested for a particular disease, and 24 infected birds were identified. Imagine also that the test sensitivity is 75%. The probability that the test has failed to identify a number of infected birds remains: the best estimate might be that eight were missed. However, the Negative Binomial distribution can be used to give a better estimate. There is a temptation to model the number of infected birds  $N$  that were not detected as:

$$N = \text{Negative Binomial}(24,75\%)$$

However, this would be assuming that the last infected bird was detected (the last trial was a success), whereas clearly the last few infected chickens tested might easily not have been detected. It turns out that, in situations like this, one can use the formula:

$$\text{Number of failures} = \text{Negative Binomial}(s+1,p)$$

So, in the example above:  $N = \text{Negative Binomial}(25,75\%)$ .

**Distributions of the Poisson process**

The Poisson process is characterised by the mean interval between events (MIBE)  $\beta$ . Once the MIBE has been estimated, it is a simple process to calculate other probability measures. The assumption of the Poisson process is that the probability of an event occurring per unit interval (e.g. per hour, per metre, per kg) is constant and independent of however many events have occurred before, or how recently they occurred.

Once the MIBE is determined, other variables can easily be found:

- the distribution of the number of events  $s$  that occur in interval  $t = \text{Poisson}(t/\beta)$
- the time until the next event  $t_1 = \text{Exponential}(\beta)$
- the time until  $s$  events have occurred  $t_s = \text{Gamma}(s, \beta)$

where  $t$  and  $\beta$  are measured in the same units (e.g. days, kg, tonnes).

### Determining the mean interval between events $\beta$ from an observed number of events over a continuous interval

The MIBE is the average interval between  $n$  observed occurrences of an event. Its true value can be estimated from the observed occurrences using central limit theorem:

$$\text{MIBE } \beta = \text{Normal}\left(\bar{t}, \frac{\sigma}{\sqrt{n-1}}\right)$$

where  $\bar{t}$  is the average of the  $n-1$  observed intervals  $t_i$  between the  $n$  observed contiguous events and  $\sigma$  is the standard deviation of the  $t_i$  intervals ( $\sigma$  should be almost the same value as  $\bar{t}$  for a Poisson process). The larger the value of  $n$ , the narrower will be the distribution of  $\beta$ , i.e. the more confidence can be placed in knowing its true value. Care should be taken when  $n$  is small (< approximately 10) because the distribution will have a tail with significant probability of being negative and will therefore have to be truncated.

Sometimes the values of the intervals  $t_i$  are not known, but only the number of events  $n$  that occurred in a total interval  $T$ . A conservative (i.e. pessimistic if the event is not desired) estimate of the MIBE  $\beta$  is:  $\beta = T/(n+1)$ .

### Estimating a minimum $\beta$ where there are no observed events over a continuous interval

The Exponential distribution can be used to estimate at least a lower bound for the MIBE, given that no occurrences of the event have been observed in time  $X$ :

$$\beta = 1/\text{Exponential}(1/X).$$

Since the lower the MIBE, the more frequently the event occurs, a lower bound for the MIBE is equivalent to providing an estimate of the highest possible frequency of the event. This provides a minimum estimate of  $\beta$  since it assumes that: 1) the event is possible; and 2) it will occur for the first time immediately after the last moment of observation.

#### Example

In the sixteen years of monitoring turkeys for a particular disease there has never been an observation of that disease. What is its minimum MIBE?

Lower bound for the MIBE is calculated as:

$$\text{Minimum MIBE } (\beta_{\min}) = 1/\text{Exponential}(1/16) \text{ years.}$$

### Probability of the occurrence of several events in an interval

The Poisson( $t/\beta$ ) distribution calculates the distribution of the number of events that will occur in an interval  $t$ .

#### Example

Outbreaks of disease Z appear to occur in wild ponies in a certain area. Records for the last 36 years show five outbreaks. A conservative (upper bound) estimate is needed of how many outbreaks could occur in the next 10 years.

This can be estimated using the formula  $X/(n+1)$  where  $X$  is the time interval during which  $n$  observations of an event have occurred. Thus:  $\beta = 36/(5+1) = 6$  years. A Poisson( $t/\beta$ ) distribution estimates the number of occurrences in an interval  $t$ . Then, the number of outbreaks  $N$  in the next ten years is modelled by:

$$N = \text{Poisson}(t/\beta) = \text{Poisson}(10/6) = \text{Poisson}(1.66666).$$

### Probability of at least one event in an interval

The probability that no event will occur in an interval of length  $x$  is  $\exp(-x/\beta)$ . The probability of at least one event in a single unit interval is therefore  $1 - \exp(-x/\beta)$ .

$$\exp(x) = e^x$$

#### Example

Government veterinarians know that a cattle disease breaks out on average once every 3.6 years. The government faces a general election in 6 months and has drastically cut the disease eradication part of its agricultural regulatory budget. What is the probability of getting through the next election before another outbreak of the disease?

The country would, on average, expect to experience an outbreak every 3.6 years. Thus:

$$\beta = 3.6 \text{ years.}$$

The probability  $P_{\text{ok}}$  of no outbreaks in the next six months is then:

$$P_{\text{ok}} = \exp(-0.5/3.6) = 87\%.$$

### Other distributions in common use in animal health risk analysis

#### Hypergeometric distribution

Consider a herd of  $M$  cows that is known to include  $D$  cows which are infected with a particular virus. If  $n$  cows are selected from this herd, the hypergeometric ( $n, D, M$ ) distribution returns the number of cows in that group of  $n$

that could be infected. The hypergeometric distribution models a type of sampling without replacement. As each of the  $n$  cows from the group of  $M$  is selected, the probability that the next cow is infected changes. (If each selected cow were to be put back into the herd before the next cow were taken out, the probability of an individual cow being infected would remain the same [i.e.  $D/M$ ] and a binomial distribution could have been used to model the number of infected cows in the sample.) In general, if  $M > 20n$ , the binomial distribution is a good approximation of the hypergeometric distribution.

So, for example, imagine a herd of 20 cows of which three are known to be infected and from which four cows will be selected at random. The probability that the first cow is infected is  $3/20$ . The probability that the second cow is infected is either  $3/19$  if the first cow selected was not infected or  $2/19$  if it was – the probability does not remain the same for each selected cow.

### Normal distribution

The normal( $\mu$ ,  $\sigma$ ) distribution is often used in animal health risk analysis, either as a consequence of applying central limit theorem or because the variable is known to be roughly normally distributed. The latter is commonly the case for natural measurements such as the weight of an adult of a particular species.

## Comparison of event tree and simulation modelling using a worked problem

A hypothetical model will now be explored to see how some of the distributions described above can be put together to produce a useful risk analysis model.

### The problem

An entrepreneur wishes to import packets of 100 turkey drumsticks from free range farms of a particular country. Slaughterhouse records from 300 turkey farms showed that during the past year disease X was found on 34 farms. A serological survey of 100 turkeys on seven of these positive farms revealed the following number of positives: 4, 6, 2, 5, 8, 3, 1. The tests were performed using a procedure with an 85% sensitivity and almost 100% specificity.

Turkeys are sent from each free-range farm to the specialist slaughterhouse in batches of 50. Two birds from any batch at the farm are tested for disease X prior to transportation to the slaughterhouse. The serological test is the same as the one used for the above survey. A turkey infected with disease X will have no external signs of disease, though there is a 10% to

40%, most probably a 30% chance of discoloration of the muscle tissue, which would certainly be spotted by the meat inspectors at the slaughterhouse.

The packaged drumsticks will be exported frozen. It is estimated that there is a probability of between 20% and 50%, most likely 40%, that the pathogenic organism, if present, will survive this freezing. The authorities for the importing country have been asked to grant a licence to an entrepreneur to import packets of these turkey drumsticks. It is understood that identifying infection at any stage will result in the rejection of only the affected package(s) of drumsticks.

The licensing authority wishes to determine the probability that this licence will introduce disease X into the country. What is the distribution of the number of infected drumsticks in any one packet of 100 drumsticks that passes all inspection? What is the probability that an accepted packet has at least one infected drumstick at import?

### The model

This problem has been modelled in two ways for comparative purposes. The first method is a simple simulation model, which is very easy to construct and will provide the mean of the distributions of the probabilities in question. However, this model does not easily lend itself to constructing the distributions of the uncertainties of these probabilities, which arise from lack of precise knowledge about any input probabilities. An excellent example in the animal product import area of this method of modelling has been developed by Van der Logt *et al.* at the Ministry of Agriculture, New Zealand (1). The paper follows a very similar presentation to that shown for Model 1 below.

The second method calculates the probabilities directly. It is mathematically more complex than the first model and the method is less flexible. However, one can arrive at distributions of the probabilities one is being asked to determine. A very good model has been developed for Agriculture Canada using this approach by M.H. Cassin, C.D. Todd, W. Ross and R.S. McColl (personal communication) to assess the risk of *Escherichia coli* in beefburgers.

Both models use the Excel® spreadsheet application as the modelling environment and the @RISK® risk analysis add-in to give Excel the ability to generate Monte Carlo sampling from probability distributions. The extra functions in Excel provided by @RISK® are characterised by starting with the letters 'Risk', e.g. RiskBeta. The reader should use the following descriptions of the models in conjunction with the spreadsheet printouts (Figs 6 and 7) and formulae tables (Tables I and II). Where one formula is shown for a range of cells in these tables, the formula has been given for the first cell in the range. Formulae for the other cells in the range would be obtained by copying this first formula into all the other cells using the Copy-Paste or Autofill spreadsheet features.

<b>Flock prevalence calculation</b>					
Flocks tested			300		
Flocks infected			34		
True flock prevalence $P_f$			13.12%		
<b>Within flock prevalence calculation</b>					
Test sensitivity			85%		
	Turkeys tested	Positives $s_i$	Number missed $m_i$	Total infected	Prevalence estimate $p_i$
	100	4	1	5	4.61%
	100	6	1	7	8.38%
	100	2	0	2	0.75%
	100	5	1	6	4.82%
	100	8	1	9	9.56%
	100	3	0	3	6.45%
	100	1	0	1	3.30%
Within flock prevalence estimate $P_a$			4.61%		
<b>Estimate of No. of infected turkeys in a consignment</b>					
Drumstick consignment size			100	Infected?	Prob.
Is flock infected? (0=no, 1=yes)			1	0	86.88%
No. of infected turkeys $N_i$			2	1	13.12%
<b>Pre-slaughter testing</b>					
No. of turkeys tested				2	
No. of tested turkeys that are infected $N_{ti}$				0	
No. of positives $N_p$				0	
<b>Inspection at slaughterhouse</b>					
Probability that flesh is discoloured if infected $P_d$				27.05%	
No. infected turkeys that are detected $N_{in}$				0	
<b>Pathogen surviving freezing</b>					
Probability that a drumstick will remain infected				34.76%	0 65.24%
No. of infected drumstick pairs being imported $N_{ii}$				2	2 34.76%

Fig. 6  
Spreadsheet printout of Model 1

## Model 1

### Flock prevalence calculation

Three hundred flocks have been tested and 34 were found to be infected. The distribution of the true flock prevalence  $P_f$  can therefore be estimated as  $\text{RiskBeta}(34+1, 300-34+1) = \text{RiskBeta}(35, 267)$ . The assumptions here are that there are many more than 300 flocks; that the 300 selected flocks can be considered a random sample; and that prior to this testing there was no knowledge of the level of flock prevalence (a prior distribution of  $\text{Uniform}(0,1)$  as discussed above).

### Within flock prevalence

One hundred turkeys were tested from each of seven infected flocks for each of which we know the number  $s_i$  of turkeys which gave positive results to serological tests. The test sensitivity is 85%, so it is quite possible that a few infected tested birds  $m_i$  were not identified. The number of birds which were missed  $m_i$  can be estimated using a negative binomial distribution as:

$$m_i = \text{RiskNegBin}(s_i + 1, 85\%).$$

<b>Flock prevalence calculation</b>					
Flocks tested		300			
Flocks infected		34			
Flock prevalence		11.72%	Event 1: flock is infected		
<b>Within flock prevalence calculation</b>					
Test sensitivity	85%				
	Turkeys tested	Positives $s_i$	Number missed $m_i$	Total infected	Prevalence estimate $p_i$
	100	4	0	4	6.15%
	100	6	0	6	8.19%
	100	2	0	2	2.51%
	100	5	1	6	4.77%
	100	8	0	8	7.31%
	100	3	1	4	4.42%
	100	1	0	1	1.74%
Within flock prevalence estimate	8.00%				
<b>Input variables</b>					
Consignment size (drumsticks: turkeys)		100	50		
Number of turkeys tested			2		
Prob. flesh is discoloured if infected			32.23%		
Prob. a drumstick will remain infected after freezing			39.94%		
<b>Determining distribution of infected drumsticks in a consignment given the flock is infected</b>					
	Event 2	Event 3	Event 4	Event 5	
Number of infected turkeys $x$	Probability that $x$ is infected in a consignment from an infected flock	Probability that pre-slaughter testing will not detect infection (sample = 2)	Probability that infection is not detected at slaughterhouse	Probability that drumsticks remain infected after freezing	Probability of events 1 to 5 together
1	6.7%	96.8%	67.8%	39.9%	0.206%
2	14.3%	93.3%	45.9%		0.287%
3	19.9%	90.0%	31.1%		0.261%
48	0.0%	3.3%	0.0%		0.000%
49	0.0%	2.8%	0.0%		0.000%
50	0.0%	2.3%	0.0%		0.000%
sum	98.453%			sum	1.079%
$P(A)$	1.079%	$P(B)$	1.62%	$P(C)$	88.5%
<b>Probability that an accepted consignment contains viable pathogenic organisms</b>					1.18%

Fig. 7 Spreadsheet printout of Model 2

The assumption behind this formula is that each infected bird has equal probability of being detected (i.e. a binomial process) and that  $(m_i + s_i)$  is a lot less than the total number of birds tested (100 in this case).

The true prevalence  $p_i$  of flock  $i$  can then be estimated as:

$$p_i = \text{RiskBeta}(m_i + s_i + 1, 100 - (m_i + s_i) + 1)$$

in a similar fashion to the flock prevalence above.

There are now seven within-flock prevalences, which can be combined to produce a distribution of within-flock prevalence  $P_a$  for all other infected flocks. One method of combining these seven  $p_i$  distributions is to use the

RiskDunifom( $y$ ) distribution: a discrete distribution where all values within its parameter array ( $y$ ) have equal probability. This method of combining distributions is also very useful for combining dissimilar expert opinions (2).

#### Estimate of number of infected turkeys in a consignment

A consignment is considered to be one packet of 100 drumsticks. Cell E21 toggles between 0 (source flock is not infected) and 1 (source flock is infected), the probability of generating a value of 1 being the flock prevalence (Cell E5).

The number  $N_i$  of infected turkeys contributing to a consignment is modelled in Cell E22 as

**Table I**  
Formulae table for Model 1 and Figure 6

Cell address	Formula
E5	=RiskBeta(E4+1,E3-E4+1)
E10:E16	=RiskNegbin(D10+1,\$D\$8)
F10:F16	=E10+D10
G10:G16	=RiskBeta(F10+1,C10-F10+1)
E17	=RiskDuniform(G10:G16)
H22	=E5
H21	=1-H22
E21	=RiskDiscrete(G21:G22,H21:H22)
E22	=RiskBinomial(E20/2,E17)*E21
F26	=IF(E22=0,0,RiskHypergeo(F25,E22,E20/2))
F27	=IF(F26=0,0,RiskBinomial(F26,D8))
F30	=RiskPert(10%,30%,40%)
F31	=IF(E22=0,0,RiskBinomial(E22,F30))
F34	=RiskPert(20%,40%,50%)
H35	=F34
H34	=1-H35
G35	=IF(F31+F27=0,E22,0)
F35	=RiskDiscrete(G34:G35,H34:H35)

RiskBinomial(50,Pa) × 21. Again, the assumption here is that the flock size is many more than 50 birds (and that each turkey has two drumsticks).

### Pre-slaughter testing

Two turkeys are to be tested out of the 50 that make up a consignment. The number of these tested turkeys  $N_{ti}$  which are infected can be modelled using a hypergeometric distribution as:

$$N_{ti} = \text{RiskHypergeo}(2, N_i, 50).$$

**Table II**  
Formulae table for Model 2 and Figure 7

Cell address	Formula
E20	=D20/2
C28:C77	=BINOMDIST(B28,\$E\$20,\$D\$17,FALSE)
D28	=HYPGEOMDIST(0,\$E\$21,B28,\$E\$20)+HYPGEOMDIST(1,\$E\$21,B28,\$E\$20)*(1-\$C\$8)
D29:D75	=HYPGEOMDIST(0,\$E\$21,B29,\$E\$20)+HYPGEOMDIST(1,\$E\$21,B29,\$E\$20)*(1-\$C\$8)+HYPGEOMDIST(2,\$E\$21,B29,\$E\$20)*(1-\$C\$8)^2
D76	=HYPGEOMDIST(1,\$E\$21,B76,\$E\$20)*(1-\$C\$8)+HYPGEOMDIST(2,\$E\$21,B76,\$E\$20)*(1-\$C\$8)^2
D77	=HYPGEOMDIST(2,\$E\$21,B77,\$E\$20)*(1-\$C\$8)^2
E28:E77	=(1-\$E\$22)^B28
F28	=E23
G28:G77	=\$D\$5×C28×D28×E28×F\$28
C78:G78	=SUM(C28:C77)
C80	=G78
E80	=G78×(1-F28)/F28
G80	=1-(C78×D5)
G82	=C80/(C80+E80+G80)

An Excel® IF function is wrapped around this distribution to ensure that it returns a zero when  $N_i$  is zero, rather than an error.

The number  $N_p$  of the  $N_{ti}$  birds which test positive with the 85% test sensitivity is modelled in Cell F27 as RiskBinomial( $N_{ti}$ ,85%). Again, an IF statement is wrapped around this distribution to ensure it returns a zero if  $N_{ti}$  is zero, rather than an error.

### Inspection at slaughterhouse

The probability  $P_d$  of an infected bird having discoloured meat and therefore being spotted by the meat inspector is modelled as RiskPert(10%,30%,40%). The Pert distribution is similar to the Triang(ular) distribution frequently used in these types of models, but has the advantages over the Triang distribution of being more naturally shaped and of being less sensitive to the estimation of the minimum and maximum values (2). The number  $N_{in}$  of infected birds in the consignment which would be detected at the slaughterhouse is therefore:

$$N_{in} = \text{RiskBinomial}(N_i, P_d).$$

An IF statement is wrapped around this distribution to ensure that the cell returns a zero in the event that  $N_i$  is zero, where it would obviously be inappropriate to attempt to generate a value from the binomial distribution.

The distribution of the number of pairs of infected drumsticks  $N_{if}$  in a consignment which gets through these tests is then calculated in Cell G35 using the equation:

$$\text{IF}(N_{in}+N_p=0, N_i, 0).$$

### Pathogen surviving freezing

The probability of the pathogen surviving freezing is assumed to apply to the whole packet of 100 drumsticks. It is assumed that if the pathogen survives in one infected drumstick in a packet it will survive in all the other infected drumsticks in that packet. However, the pathogen will not spread to the other uninfected drumsticks. The probability  $P_s$  of survival is modelled in the same way as  $P_d$ :

$$P_s = \text{RiskPert}(20\%, 40\%, 50\%).$$

The number  $N_{ii}$  of infected drumstick pairs being imported in one packet of 100 drumsticks from this source is then modelled in Cell F35 as:

$$N_{ii} = \text{RiskDiscrete}(\{0, N_{if}\}, \{1-P_s, P_s\}).$$

This model was run for 100,000 Latin Hypercube iterations (2). A histogram of the resultant distribution for  $N_{ii}$  is shown in Figure 8.

### Model 2

This second model, shown in Figure 7, calculates the same problem as described above, except that the probability

masses for the output distribution are calculated directly rather than determined through a frequency analysis of the simulation results. So, for example, instead of noting that Model 1 produced a zero in 98.5% of its iterations (Fig. 8), the true value of that percentage is actually calculated. Since calculus is being used to determine this figure, the simulation capabilities can also be used to show the certainty of that value.

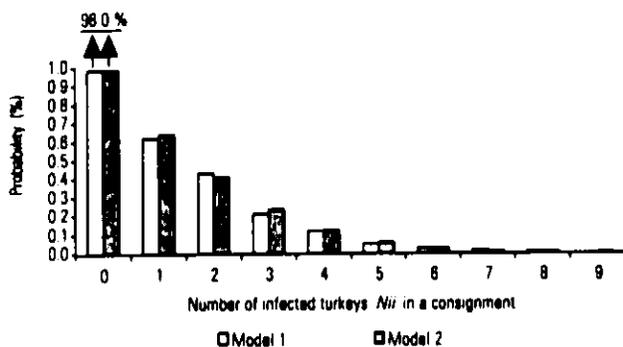


Fig. 8 Comparison of output distributions for Models 1 and 2

Model 2 is precisely the same as Model 1 up to the end of the calculations for the flock and within-flock prevalences, including the modelling of P<sub>s</sub> and P<sub>d</sub>, since these parts only produce distributions arising from our lack of precise knowledge. Thereafter, the approach of calculating the probability masses of the resultant distributions at each stage will be taken.

The table of cells from B28 to G77, partially illustrated in Figure 7, performs these calculations. The table uses the BINOMDIST and HYPERGEOMDIST functions of Excel. Column B, ranging from 1 to 50 (the number of turkeys comprising one consignment) denotes the number of infected turkeys x in a consignment. Column C returns the probability masses for the Binomial(50, P<sub>a</sub>) against these x values. This is the distribution of the number of infected turkeys in a consignment, given that the source flock is infected. So, for example, the iteration presented in Figure 7 shows that there is a 19.9% chance there will be drumsticks from three infected turkeys included in this consignment, given that the source flock is infected.

Column D calculates the probability that 0, 1 or 2 infected turkeys will be selected for the pre-slaughter tests, given that x infected turkeys are in the consignment, and that the testing then fails to detect each one of these infected birds. Note that the formulae for x = 1, 49 and 50 are a little different from the rest of the column. This is because, if x = 1, there could not be two infected turkeys in the sample, and if x = 49 or 50, there must be at least one and two infected birds respectively in the test group.

Column E calculates that, given x infected birds in the consignment, none will be detected at the slaughterhouse due to discoloration of the flesh. Column F is simply replicating the probability that the pathogenic organism would not survive freezing. Only one cell is used in this column for computational efficiency: all cells in column G refer to this one used cell (F28). Finally, column G puts the whole calculation together. It returns the probability of the source flock being infected, of having x infected turkeys in the consignment, that those infected turkeys were then not picked up during pre-slaughter inspection or during slaughtering, and that the pathogenic organism residing in the meat also remained viable (Events 1 to 5). The sum of column G, shown in Cell G78, is the probability that viable pathogenic organisms are in a particular consignment and that the consignment has been accepted. The mean of the distributions for each of the cells in column G are equivalent to the probability masses calculated in Figure 8 for Model 1. However, these means are considerably more accurate. Figure 8 plots the results from Models 1 and 2 together with the results of 3000 iterations of Model 2. They arrive at the same result but Model 2 will more reliably and more quickly reach the theoretical answer than Model 1.

With Model 2, the analysis can be taken one step further. In general, it is of far more interest to look at the probability that an accepted consignment contains viable pathogenic organisms, i.e. P(viable|accepted) rather than the probability that a consignment containing viable pathogenic organisms will be accepted, i.e. P(accepted|viable). Figure 9 shows an event tree of how the former probability could be calculated. For the probabilities P(A), P(B), P(C) labelled in this figure:

$$P(\text{viable}|\text{accepted}) = \frac{P(A)}{P(A) + P(B) + P(C)}$$

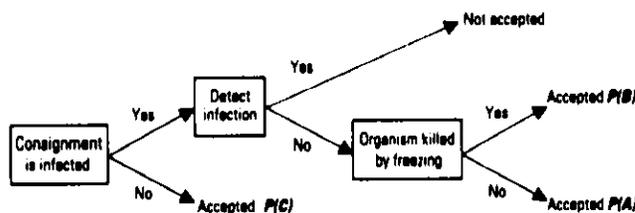


Fig. 9 Simplified event tree of the problem modelled, used to calculate P(infected | accepted)

This is the probability that the consignment contains viable pathogenic organisms given it has been accepted, divided by the sum of the probabilities of all paths leading to the consignment being accepted. P(A), P(B) and P(C) are calculated in Cells C80, E80 and G80 respectively. P(viable|accepted) is then calculated in Cell G82.

## Conclusions

It has been shown that, with a good understanding of a few basic distributions, a risk analysis model can be constructed which is transparent and provides both measures of the probabilities of outcomes and the degree of uncertainty one may have about these probabilities. Performed correctly, quantitative risk analysis is a powerful tool which will guide the decision-maker towards a better understanding of the risks being faced, the effectiveness of current and planned risk

management strategies and of the value of further research to reduce any uncertainty in the model.

## Acknowledgement

I am extremely grateful to Paul Suttmoller for his care and valuable comments in proof-reading this text.

## Analyse des risques liés à l'importation et à l'exportation de produits d'origine animale

D.J. Vose

### Résumé

Un modèle d'analyse quantitative des risques doit être conçu en fonction des questions auxquelles il est censé répondre. Il doit, en outre, être aussi objectif que les données disponibles le permettent. La difficulté d'élaborer un bon modèle d'analyse quantitative des risques tient à ce que les risques liés à l'importation d'animaux et de produits d'origine animale présentent, en général, les trois caractéristiques suivantes :

- les probabilités des étapes conduisant à des résultats indésirables sont fréquemment liées entre elles ;
- la probabilité d'aboutir à un résultat indésirable est en soi, dans bien des cas, très faible, ce qui rend les simulations directes peu pratiques à réaliser ;
- il arrive souvent que d'importantes variables au sein du modèle ne puissent pas être quantifiées par l'analyse des données ; aussi doivent-elles être modélisées en distribuant des probabilités qui reflètent le degré d'incertitude, habituellement déterminé sur avis d'expert.

Cet article constitue une initiation à certaines techniques de modélisation, essentielles à l'évaluation des risques liés à l'importation d'animaux et de produits d'origine animale, et permettant de résoudre ces problèmes. L'auteur examine un certain nombre de distributions de probabilités, leurs utilisations et leurs relations. Il montre ensuite comment l'application de ces distributions et de quelques techniques de modélisation générales permettent d'aboutir à des analyses des risques liés à l'importation d'animaux, à la fois rigoureuses et transparentes.

### Mots-clés

Analyse des risques - Arbre de probabilités - Evaluation quantitative des risques - Modèle - Simulation - Simulation de Monte-Carlo - Techniques.

# Análisis de los riesgos asociados a la importación y exportación de productos de origen animal

D.J. Vose

## Resumen

La elaboración de un modelo de análisis cuantitativo de riesgos viene determinada por los interrogantes a los que dicho modelo debe dar respuesta. El modelo debe ser además tan objetivo como vayan a permitirlo los datos disponibles. Los riesgos asociados a la importación de animales y productos de origen animal suelen poseer tres características que dificultan la elaboración de un buen modelo de análisis cuantitativo de riesgos, a saber:

- las probabilidades de los pasos que conducen a un resultado indeseado guardan con frecuencia relaciones recíprocas;

- la probabilidad del propio resultado indeseado es en muchos casos extremadamente baja, hecho que casa mal con la práctica de simulaciones directas;

- a menudo, el análisis de datos no permite cuantificar una serie de variables importantes para la aplicación del modelo. Por esta razón es preciso modelar dichas variables con distribuciones de probabilidad para reflejar el grado de incertidumbre, que en general se determina mediante una opinión de experto.

El autor proporciona una guía para la aplicación de algunas técnicas de modelización, esenciales para evaluar los riesgos asociados a la importación de animales y productos animales y de gran ayuda para salvar estos escollos. En este sentido se examinan una serie de distribuciones de probabilidad, así como sus usos e interrelaciones. La aplicación de estas distribuciones, junto con la de ciertas técnicas de modelización, se demuestra un método riguroso y transparente para el análisis de los riesgos asociados a la importación de animales.

## Palabras clave

Análisis de riesgos - Árbol de probabilidades de riesgo - Evaluación cuantitativa de riesgos - Modelo - Simulación - Simulación de Montecarlo - Técnicas.

■

## References

1. Van der Logt P.B., Hathaway S.C. & Vose D.J. (1997). - Risk assessment model for human infection with the cestode, *Taenia saginata*. *J. Food Protec.* (in press).
2. Vose D.J. (1996). - Quantitative risk analysis: a guide to Monte Carlo simulation modelling. John Wiley & Sons, Ltd., Chichester, United Kingdom, 317 pp.