The analysis of complex surveillance systems using stochastic scenario tree modelling

Guide to the Methodology

Tony Martin, Angus Cameron, Jenny Hutchison, Evan Sergeant, Nigel Perkins

Supported by:

Introduction

The SPS agreement of the WTO requires that, in international trade, measures taken to protect animal, plant or human health should be based on scientific principles and not maintained in the absence of sufficient evidence. Countries support such measures by using science-based risk analysis, which in turn demands science-based assessment of the disease status (free or infected) of each of the trading partners. If it can be demonstrated that an exporting country is free from a disease, risk analysis is straightforward; the likelihood of importation of the disease is zero. Similarly, an importing country wishing to impose quarantine measures must provide evidence for its own freedom from disease.

In the past, two approaches have been commonly used to provide the required evidence. The first is a structured, representative survey of the relevant population. The data are used to estimate the probability that the negative results of the survey would be achieved if the disease were present at a specified level (the design prevalence). If the probability is less than an agreed level, the population is considered free.

The second approach is qualitative assessment, by a panel of experts, of multiple complex sources of evidence. These may include sources of non-representative data such as laboratory records, abattoir sampling, notifiable disease databases, etc, as well as the results of structured surveys of the type referred to above. Such qualitative assessments are more subjective and generally provide a dichotomous outcome – free from infection or not. This approach has been applied by the OIE and regional animal health

organisations to assess claims of freedom at the completion of disease eradication programs.

Each of these approaches suffers from significant weaknesses. Structured surveys using representative sampling are expensive, difficult to implement, and ephemeral in their applicability. Reliance solely on the results of such surveys ignores the potential value of all other sources of evidence. On the other hand, a qualitative assessment may consider all sources of evidence, but the outcome is influenced by the assessors involved, and it is extremely difficult to achieve a transparent and repeatable process.

This manual presents methods for combining the advantages of both of these approaches, by enabling multiple sources of surveillance data (both random and non-random) to be used to develop a quantitative probability estimate to support claims of freedom from disease or infection. The methodology is presented as one possible framework for further development of appropriate methods in this important arena.

The methodology presented has been developed in the context of using the negative outcomes of surveillance activities as evidence of freedom from livestock diseases, and consequently the terminology and examples are derived from this field. It is apparent that there are many other potential applications, in the arenas of animal health, plant health and human health; notably evaluation of the efficacy of different surveillance activities for disease detection, by estimating their sensitivities and quantifying the benefits of targetting.

List of abbreviations and notation

Symbol	Term	Definition and explanations
$\Pr(A)$	Probability of outcome A	Probability of an outcome. Sometimes referred to as "marginal probability" of the outcome.
$\Pr(A\|B)$	Conditional probability of outcome A given outcome B	Probability of an outcome given preconditions. In the context of the scenario tree, the preconditions are the outcomes of each of the higher level nodes on the limb.
$\Pr(A,B)$	Joint probability of outcomes A and B	Probability of two or more outcomes occurring together.
D	disease or infection	The specified disease, for which the country is claiming freedom. NB: "Disease" is used as a generic term. In most cases, D can be interpreted in terms of "infection"
$D+$	diseased or infected	The true status of a unit of interest is infected.
$D-$	non-diseased or uninfected	The true status of a unit of interest is uninfected.
T	diagnostic test	A single diagnostic test or a diagnostic testing procedure designed to classify units of interest as test positive or test

		negative. A diagnostic procedure may include more than one test.
$T+$	test positive	Positive outcome as defined by specific criteria of a diagnostic procedure for one unit of interest.
$T-$	test positive	Negative outcome as defined by specific criteria of a diagnostic procedure for one unit of interest.
$S+$	Surveillance positive	Positive outcome from a surveillance system (or component)
$S-$	Surveillance negative	Negative outcome from a surveillance system or component
α	alpha	Probability of type I error.
β	Beta	Probability of type II error.
PA	within herd prevalence	Proportion of infected animals within a group, usually a herd. This equates to the probability that any animal chosen at random from this herd will be infected.
PH	among herd prevalence	Proportion of herds with one or more infected animals in a specified population of herds. This equates to the probability that any herd chosen at random will be infected.
P^*	Design prevalence	Fixed value for prevalence, at the unit or group level, used as the basis for declaring the herd or country free from the disease in question. The subscripts H and U are used to indicate the unit of concern.
P^*_H	Among-herd/group design prevalence	Probability that a herd is infected, given that the country (or population subgroup) is infected. The hypothetical proportion of herds that are infected for evaluation of SSC sensitivity
P^*_U	Within-herd/group design prevalence	Probability that a unit/animal is infected within an infected herd/group. The hypothetical proportion of units that are infected within infected herds/groups, for evaluation of SSC sensitivity
Se	Sensitivity	The probability that an infected unit (or group or population) will be correctly identified as infected by a test or process applied to the unit (or group or population). Subscripts U, H etc. used to specify the infected entity which is tested.
Sp	Specificity	The probability that an uninfected unit (or group or population) will be correctly identified as uninfected by a test or process applied to the unit (or group or population). Subscripts U, H etc. used to specify the infected entity which is tested.

SeH	Herd sensitivity	The probability that an infected herd will be correctly identified as infected by the herd-level test or process applied to the herd.
SeU	Unit sensitivity	The probability that an infected surveillance unit will give a positive result in the surveillance process.
SpH	Herd specificity	The probability that an uninfected herd will be correctly identified as uninfected by the herd-level test or process applied to the herd.
SR	Sensitivity Ratio	Ratio of actual *CSe* to that of a hypothetical representative standard for the SSC
PrP	Population proportion	Proportion of units or groups that fall into a category (of units or groups) in the SSC reference population
PrSSC	SSC Proportion	Proportion of units or groups processed in the SSC which fall into a category (of units or groups).
SSC	Surveillance System Component	One component of a surveillance system for a specified disease or infection in a population.
CSe	Surveillance Component sensitivity	Sensitivity of a SSC.
CSeU	Surveillance Component Unit sensitivity	The sensitivity of a a randomly selected unit from the SSC for detection of disease in the population (the probability that a unit selected randomly from those sampled will give a positive result, given that the population is infected at the design prevalence).
SSe	System sensitivity	Sensitivity of an entire surveillance system.
RR_i	Relative risk	The relative risk that the section of the population represented by the *i*th branch of a risk category node will be infected (at P^*) relative to the population section represented by the lowest risk branch of the node.
AR_i	Adjusted relative risk	RR_i adjusted to ensure population average relative risk is 1.

List of Tables

List of Formulae

Main Concept

Proof of "disease freedom" in the strict sense is impossible. This is due to practical constraints:

- Data to support such claims are not available for *all* animals of the population of interest.
- Methods for disease detection such as clinical, pathological, virological or serological diagnostic methods inherently provide limited certainty. In this context, false negative test results are of concern.
- The time period of the observations and records must be considered in the analysis of survey or surveillance data.

This introduction reviews and summarises the fundamental concepts and theories behind "disease freedom" that will be applied in later sections of this document. The learning objective of this section is to develop a solid understanding of:

- the notion of disease freedom, incidence, prevalence, apparent prevalence;
- the impact of diagnostic test properties (sensitivity and specificity) and the importance of unbiased, empirical estimates of diagnostic parameters;
- the Bayesian approach applied at the level of an individual animal, a herd of animals or a population of animals;
- the statistical and biological concepts of the design prevalence;
- the statistical and biological concepts of differential risks of disease; and
- the statistical and biological considerations in connection to the time period of data collection for disease freedom.

Condition of Interest

The term "disease" is often used as a generic expression for any condition to be detected in animals or other units of concern. In these notes, we shall follow this "loose" terminology. However, the methodology of "disease freedom" is particularly relevant for infectious, contagious animal diseases. Of course it is necessary to consider the important difference between "infection" and "disease". Infections of animals follow a typical course. Incubating or immune animals appear clinically normal, i.e. they are not diseased however the disease agent is harboured. As a principle, infected-non-diseased animals are not acceptable in a country that claims freedom from the disease. From the disease-detection point of view, such animals are problematic because they can't be identified clinically.

The immune status and other factors determine the chance that an infected animal can be diagnosed using direct or indirect detection methods such as virus detection or demonstration of specific antibodies, respectively.

In conclusion, the technical term "free from disease" should be translated as "free from infection" in the context of the infectious diseases. Inherent limitations of the detection principles for the infection must be considered. The different statuses of animals relevant to the disease-freedom problem are shown in Figure 1.
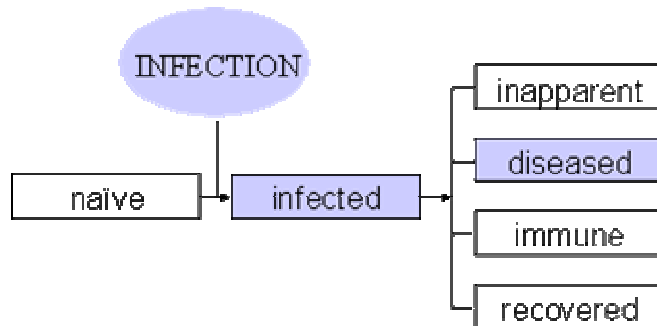
Figure 1. Statuses of a (live) animal relevant for the "disease-freedom" problem.

Infection leads to the status "infected". Depending on immune status and other factors, an infected animal may be in inapparent, diseased, etc. Infected animals are not acceptable in a country or region that claims "disease-freedom" (recovered and immune animals may be an exception).

Measures of disease occurrence

Incidence and prevalence are two different concepts of measuring the frequency of diseases in a population. Incidence is a measure of the occurrence of cases over a defined period of time. Under the simplistic scenario, a number of n animals is followed over the defined time period (d). An incidence rate based on count data IR is given as the number of new cases X during the period over the total of observed animals,

$$IR(d) = \frac{X}{n}.$$

In observational studies, the observation period is typically different among animals due to late entries, drop-outs, losses, etc. The time under risk for each animal can be established and totalled (total animal time, AT). Using the number of new cases observed in AT, the incidence density (ID) is just

$$ID = \frac{X}{AT}.$$

A measure of disease frequency in cross-sectional studies is the prevalence

$$P = \frac{K}{n},$$

where K and n are the number of animals found diseased and the sample size, respectively. The prevalence provides a snapshot of the disease situation, whereas incidence tells something about the disease dynamics. Incidence and prevalence are mathematically related. The prevalence odds, $P/(1-P)$, is equal to the product of ID and the mean duration of the disease (Rothman and Greenland, 1998). Further elaboration of these concepts and more refined formulas (point estimates, variance estimates for different sampling designs) can be found in standard epidemiology textbooks.

Investigations for documenting disease-freedom are special cases of incidence or prevalence studies because a zero incidence or prevalence is expected. For this reason, standard study designs for estimation of these parameters are not suitable. Nevertheless, there is an obvious analogy. A routine monitoring or surveillance programme shares the longitudinal aspect of an incidence study; a structured survey for documentation of disease freedom resembles a prevalence study.

What we mean by "free"

From a hypothesis testing point of view, one could define a null hypothesis

H_0: The country/region is not free of the disease.

Based on data collected to document disease freedom, one usually wishes to reject H_0 in favour of the alternative hypothesis

H_A: The country/region is free of the disease.

However, it is practically impossible to prove that a whole country is "free of the disease". A single infected animal in the population classifies the country or region as "not free". To "prove freedom" would require that the whole population is investigated for the presence of the disease ("tested") using a perfect diagnostic method. Clearly, this is an unfeasible task. The proof of "absolute disease freedom" seems not even necessary if one considers the expected spread of contagious diseases in a naïve population. "Freedom from circulating virus" in the context of viral diseases might be a useful interpretation of the concept.

Later in the guide, the concept of the design prevalence (P^*) will be described in detail. Briefly, the design prevalence is an alternative way of specifying the hypotheses mentioned above. Rather than using zero for incidence (over the observation period) or prevalence, one would use a small value for P^* and the hypotheses become

H_0: The country/region is infected at a level at or above P^*

H_A: The country/region is free of the disease or infection level is below P^*

More about the concept of design prevalence follows in later sections.

Probability and distributions

We consider here the basic laws for dealing with probabilities (Pr). Common sense and experience tells us that the probability to throw a "6" with a fair die is exactly 1/6. But what is the probability that an animal selected at random from the population has the disease in question? Or what is the probability that a diseased animal is detected with a diagnostic test? No theory or laboratory experiment tells us this. We have to conduct a survey to find it out.

Prevalence

Let Y be a dichotomous variable with the two levels $Y = 0$ for non-diseased ($D–$) and $Y = 1$ for diseased animals ($D+$). The probability that the randomly chosen animal is $D+$ is called the prevalence of D,

$$P = Pr(D+)$$

or, equivalently, $Pr(Y = 1)$. P is an unknown quantity and can only be estimated. Assume we have a randomly chosen sample of n animals from the population and K are diseased (note that K is the sum of Y in the sample). We assume here that a perfect test is used to classify D. The estimate of the disease prevalence is of course

$$P = \frac{K}{n}.$$

We omit the "hat" on the P, which would correctly indicate that this is an estimate, not the true prevalence. "Estimates" are the numerical results obtained when we evaluate a formula, such as K/n. The formula is called the "estimator".

Diagnostic tests are used to estimate the prevalence and such tests are not free of errors. Therefore, the prevalence, as defined above, should be clearly distinguished from the probability of a positive test result. The latter is also called *apparent prevalence* (see below).

Quality of estimates

The quality of estimates has two key aspects from which other useful criteria can be derived: bias and variance. The bias tells us something about the central tendency (expected value) of the estimator. If the expected value (E) is equal to the population parameter, the estimator is unbiased.

Examples for biased estimators occur in analysis of survey data. The simple proportion $P=K/n$ may be a biased (pooled) point estimator of the population prevalence if the observations are grouped in herds or other primary sampling units (two-stage cluster sampling) and have unequal sampling weights. The latter typically occurs if the herd sizes are unequal, ie, if it is not possible to sample the same proportion from each herd. For such aggregated observations, the estimator of the variance of P, $\text{var}(P)=P(1–P)/n$, is also typically biased, because the observations within the primary sampling units are not independent. Such biases should be clearly distinguished from biases due to the study design or lack of randomness in the process (e.g., selection bias). The general definition of a bias is

$$\text{bias} = E(\text{estimator}) – P.$$

The variance (var) tells us something about the expected sampling variation of the estimator in terms of the expected value of the squared difference between $E(P)$ and P,

$$\text{var} = E[(E(\text{estimator}) – P)^2].$$

Sometimes it is desirable to consider both bias and variance at the same time because it might be advantageous to work with a biased estimator with small variance compared to an unbiased estimator with large variance. A combined quality criterion of bias and variance is the mean square error (mse),

$$\text{mse} = \text{var} + \text{bias}^2.$$

Conditional probability

Let's consider the disease prevalence in male and female animals as a simple illustrative example. The notation we can use to describe this is $P_1 = \text{Pr}(D+|\text{male})$, read "probability of $D+$ given the animal is male", and $P_2 = \text{Pr}(D+|\text{female})$, "prevalence of D given the animal is female". After the vertical bar, the conditional statement follows. Both statuses, disease and gender have two possible outcomes. If we sample n animals randomly from a population we can arrange the observed data in a 2x2 table (Table 1).

Table 1. 2x2 table with hypothetical data.

	Diseased ($D+$)	Non-diseased (D–)	Total
male	$a = 10$	$b = 90$	$n_1 = 100$
female	$c = 40$	$d = 360$	$n_2 = 400$
	$m_1=50$	$m_2 = 450$	$n = 500$

The estimate of the prevalence in male animals is $P_1 = a/n_1 = 0.1$ or 10%. The sample estimate of the prevalence is $P = m_1/n = 50/500 = 0.1$ or 10%. We could also say

something about the gender distribution. The observed proportion of male animals is $P_M = n_1/n = 100/500 = 0.2$ or 20%. The proportion of female animals is 80%.

Both the disease status D and the gender are mutually exclusive categories. The statuses are also jointly exhaustive. If one animal is not $D+$, it must be $D-$, and so on. For this reason, we can now state that

$Pr(D+) = 1 - Pr(D-)$; and $Pr(M) = 1 - Pr(F)$.

If the probability of disease is the same in male and females; that means if

$Pr(D+) = Pr(D+|M) = Pr(D+|F)$

holds, we can say the disease status and gender are independent. In practice, it involves a statistical test (for example chi-square) to investigate whether the observed differences between P_1 and P_2 can be explained just by chance (sampling fluctuation). The point to make here is the general notion of independence. The prevalence P and the probability $Pr(M)$ are called the marginal distributions of Table 1. If the two variables are independent of each other, it follows that the probability of selecting a male diseased animal can be found simply by the law of multiplication of the marginal probabilities,

$Pr(D+, M) = Pr(D+) Pr(M)$.

In this notation, the comma has the meaning of "and". In the example, the estimated marginal probability of disease (unconditional on gender) is 10%, and the marginal probability of male is 20%. The probability of both features applying jointly is 10% of 20% (or 20% of 10%), equal to 2%. The Venn diagram below (Figure 2)is a graphical representation of the data in Table 1.
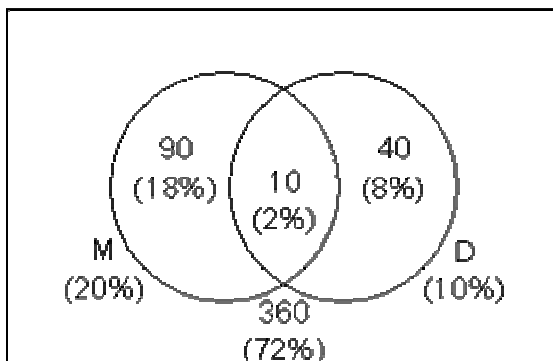

Figure 2. Venn diagram for data in Table 1.

The cells within the table reflect the joint distribution of disease and gender. Under the same marginal distribution, we could have $a = 50$, $b = 50$, $c = 0$ and $d = 400$, reflecting a very strong dependence between D and gender. The link between conditional and unconditional probabilities is given by Bayes' theorem.

Bayes' theorem

Bayes' Theorem is a procedure for revising the probability of some event in the light of new evidence. Assume B_j is the event of interest and B_1, B_2, …, B_k are mutually exclusive and exhaustive outcomes. We wish to revise the probability $Pr(B_j)$ using the observation that event A has occurred. Bayes Theorem can be written as

$$Pr(B_j|A) = \frac{Pr(A|B_j)Pr(B_j)}{\sum_{j=1}^{k} Pr(A|B_j)Pr(B_j)}$$

The quantity $Pr(B_j)$ is called the *prior probability* and $Pr(B_j|A)$ is the *posterior probability*. The term $Pr(A \mid B_j)$ is equivalent to the so-called *likelihood*. It is given by the conditional probability of A (due to the observed result of A), given B_j. The likelihood and the prior must be derived from independent sources of evidence. The denominator is a scaling factor that assures that the posterior probability totals 1 over all categories of B_j.

Using the example above as illustration, we could be interested in the probability of disease (event B_1 is now $D+$) in a female (event A is now F) animal,

$$Pr(D+|F) = \frac{Pr(F|D+)Pr(D+)}{Pr(F|D+)Pr(D+) + Pr(F|D-)Pr(D-)}$$

Bayes Theorem is applicable if we have a prior estimate of the disease prevalence $Pr(D+)$ = $Pr(B_j)$ and if we know that the likelihood of disease depends on the gender; i.e. we have *independent* information on the likelihood $Pr(F \mid D+)$ and $Pr(F \mid D-)$. Note that the posterior $Pr(D+ \mid F)$ and the conditional probability $Pr(F \mid D+)$ are two different things. The latter is required to estimate the posterior probability; it has no other immediate relevance in this context. If a female animal is encountered, Bayes Theorem can be used to find the posterior probability $Pr(D+|F)$. In the case that gender is no explanatory factor, we have $Pr(F|D+) = Pr(F|D-)$ and the formula above reduces to $Pr(D+|F) = Pr(D+)$.

Bayes Theorem cannot be applied if all information stems from one single source (in our case from one single 2x2 table) since it requires prior knowledge. Expressing all quantities with the cell frequencies *a*, *b*, *c* and *d* leads to $Pr(D+|F) = c / (c+d)$. This shows that estimation of the conditional probability $Pr(D+|F)$ from the 2x2 table is a simple proportion rather than a posterior probability. Therefore, other sources of data are required to apply Bayes' theorem. Further detailed description of such sources will be presented below ([data sources](#)). Bayes' theorem will be further elaborated in the context of [diagnostic testing](#).

Probability density function (PDF)

Some PDFs, such as the normal, are widely known and applied. If we deal with categorical data, we need different types of PDFs. A basic idea in statistics is that we can use *models* to describe what we observe. Let's use the term "outcome" for anything we could be interested in observing (whether one animal is diseased, how many animals are diseased, how many kilograms one animal weighs, etc). The models help us to extrapolate from our limited data to the real world. PDFs are distribution models and share three characteristics.

1. They indicate the possible outcomes of an event (yes/no for the outcome "animal diseased"; 0, 1, 2, …, n for the outcome "how many diseased animals"; from 10 to 500kg for the outcome "weight").
2. They define how likely each of the possible outcomes is using a formula that consists of constants and *parameters*. The sum of all these probabilities is exactly 1.
3. They can be used (in most cases) to derive the expected value (mean value) and the expected variation (variance).

The mean of the PDF is the expected value of the random variable,

$m = \text{E}(X)$.

The variance is the expected squared deviation of X from its mean (see above),

$var = E[\ (X - m)2]$

PDF of a dichotomous variable (Y)

For example, the diagnosis of one animal with observed outcomes $Y = 0,1$ to denote the disease status.

1. Possible outcomes: $y = 0, 1$.
2. The model is called a Bernoulli distribution with the single parameter P:

   PDF: $\Pr(Y = y) = P^{y}(1-P)^{1-y}$ where $0^0 = 1$.

   The probability of disease is $\Pr(D+) = \Pr(Y = 1) = P$

   The probability of no disease is $\Pr(D\text{-}) = \Pr(Y = 0) = 1-P$

   The sum is $P + 1 - P = 1$.

3. The expected value (mean) is P; the variance is $P(1-P)$.

The PDF of a binomial variable (K)

A sample of $n = 10$ animals was investigated for the disease, K were found positive.

1. Possible outcomes: $k = 0, 1, 2, \ldots, n$.
2. The model is called a binomial distribution with the parameter P and the constant n:

   PDF[1]: $Pr(K = k) = \binom{n}{K} P^k (1-P)^{n-k}$

   The probability of the outcome 0 is $Pr(k = 0) = (1 - P)^n$
   The probability of the outcome K = 1 is $Pr(k = 1) = nP(1 - P)^{n-1}$
   and so on for $K = 2, 3, \ldots$
   The probability of the outcome $K = n$ is $Pr(k = n) = P^n$
   The sum of all probabilities (for $k = 0$ to n) is 1.

3. The expected value (mean) is nP; the variance is $nP(1-P)$.

[1] The binomial coefficient $\binom{n}{K} = \dfrac{n!}{k!(n-K)!}$ indicates the number of ways that K items can be selected from n items.

Diagnostic test

For the purpose of this course, any procedure used to classify a unit as either positive or negative with respect to the infection or disease of concern is referred to as a diagnostic test. This definition covers any device or process designed to detect a sign, substance, tissue change or host response. To be a test, the procedure must be better than a random process (e.g. flipping a coin); the test must help distinguish between affected and non-affected individuals. Tests can be measured on dichotomous (+/-), ordinal (ordered response) and continuous scales. In order to derive a diagnostic decision from ordinal or continuous test results, a *cut-off value* is used to define test positive ($T+$) and test negative ($T-$) outcomes. Multiple tests and herd tests will be described in later sections.

Test results should reflect the true state of disease (infection) as closely as possible but diagnostic misclassifications, i.e. false positive and false negative results, occur in most (if not all) diagnostic test procedures. Some reasons for this are given below. A comprehensive discussion of this topic is beyond the scope of this course.

Technical variability: The results of repeated testing of the same sample are usually not identical. The degree of inherent variability can be measured within one laboratory (repeatability concept) or among laboratories (reproducibility concept). Quality assurance methods can help to keep a diagnostic process stable within accepted measurement errors.

Biological variability: Quantitative diagnostic test results vary among animals with the same true infection or disease status as a result of biological factors (age, immune status, physical condition, etc.). One example is the stage of infection. Depending on the test

principle used, infected animals are not detectable by the test during periods of incubation or antibody latency. The range of observed test values in truly infected and non-infected animals is typically overlapping. This leads to diagnostic errors and imperfect sensitivity and specificity (see below).

Evaluation of diagnostic tests

The diagnostic performance measures can be derived from the results of a test evaluation study. This requires that the test results and the outcomes of a reference test ("gold standard") diagnosis are available, matched on individuals. The results of an evaluation study can be summarised in a 2x2 table as below (Table 2).

Table 2. Result of an evaluation study displayed in a 2x2 table

		True disease state		
		$D+$	$D-$	
Test	$T+$	a	b	n_1
	$T-$	c	d	n_2
		m_1	m_2	n

The following typical study designs are encountered in test evaluation.

Cross-sectional
> A total of n animals is randomly selected from the target population and subjected to the new test and the reference method. Advantage: sensitivity, specificity and predictive values can be estimated as simple proportions as indicated below. Disadvantage: Not very efficient for rare diseases or costly reference method.

Pre-stratified
> A number of m_1 and m_2 animals are selected separately from two sampling frames of truly infected and uninfected animals, respectively. Advantage: Efficient for rare diseases and suitable for experimental design. Disadvantage: Pre-stratification often introduces a selection bias.

Partial verification
> The new test is performed first. Different fractions of test positive and test negative animals are subjected to the reference test. Advantage: Efficient if reference method is costly or invasive. Disadvantage: Selection of individuals for confirmation may introduce bias.

Complex survey design
> Cross-sectional or partial verification design combined with two-stage cluster sampling and/or stratification. This design occurs naturally, when populations of farming animals are sampled.

The design must be considered for correct statistical inference (see details in Greiner and Gardner, 2000a). The basic measures of diagnostic performance are described in the following section.

The data in the 2x2 table shown above can be described with two different prevalence measures.

$P = (a+c)/n$

is an unbiased estimate of the true prevalence $Pr(D+)$ in the cross-sectional study design.

$AP = (a+b)/n$

is an unbiased estimate of the apparent prevalence $Pr(T+)$ in the cross-sectional and the partial verification design (also see below).

The diagnostic accuracy has two components: Sensitivity and Specificity.

Sensitivity (*Se*)

Is the probability of a positive test result ($T+$) given the disease is present ($D+$) (diagnostic sensitivity),

$Se = Pr(T+|D+)$.

It can be estimated as relative frequency of positive test results in infected individuals,

$Se = a/(a+c)$.

Specificity (*Sp*)

Is the probability of a negative result ($T-$) given the disease is not present ($D-$) (diagnostic specificity),

$Sp = Pr(T-|D-)$.

It can be estimated as the relative frequency of negative test results in non-infected individuals,

$Sp = d/(b+d)$.

Se and *Sp* are important, since they are related to all other measures of the diagnostic performance, as we shall see in later sections. We note at this stage that *Se* and *Sp* do not change with prevalence [2].

Variance of accuracy measures

Often *D+* and *D–* animals are sampled independently (stratified sampling, see section on study design). In this case, *Se* and *Sp* are simple proportions (*p*), established using a sample size n. The variance is then given as

$$\text{var}(p) = p(1\text{-}p)/n$$

where *p* = *Se*, *Sp*.

Apparent prevalence

The apparent prevalence (*AP*) denotes the probability of an animal having a positive test result, *AP* = Pr(*T+*). A positive test result can be due to a correctly classified diseased animal or a misclassified non-diseased animal,

$$\Pr(T+) = P\ Se + (1{-}P)\ (1\text{-}Sp).$$

Operational diagnostic test parameters

Sensitivity and specificity are essential for a proper interpretation of test results. If unbiased estimates of these error rates are known, one can adjust diagnostic interpretations (predictive values), prevalence estimates (Rogan and Gladen, 1978), or risk factor estimates (Greiner and Gardner, 2000b) for diagnostic misclassification. A critical requirement is that sensitivity and specificity estimates are actually valid for the population of the intended use.

In the context of disease freedom, reliable estimates of *Se* and *Sp* are required for two different purposes. For the planning of monitoring or surveillance systems, sample sizes must be established. Typically, a lack of sensitivity must be compensated with increased sample sizes. In later sections of these notes, existing surveillance schemes will be evaluated in terms of the sensitivity of the whole system. As diagnostic testing is part of the system, realistic estimates of the diagnostic parameters must be available. Evidence from multiple validation studies can be summarised by meta-analysis (Irwig et al., 1995). Latent class models were described to address the problem of a lack of reference methods (Enoe et al., 2000).

Testing systems

Diagnostic testing in practice often involves the use of more than one test before a final diagnosis is made.

Multiple diagnostic tests

Two principal versions of multiple diagnostic tests (T_M) are used. In a *parallel testing* scheme, the sample is subjected to more than one test and the final diagnosis is reached by summarising the test results ([Figure 3](#)).
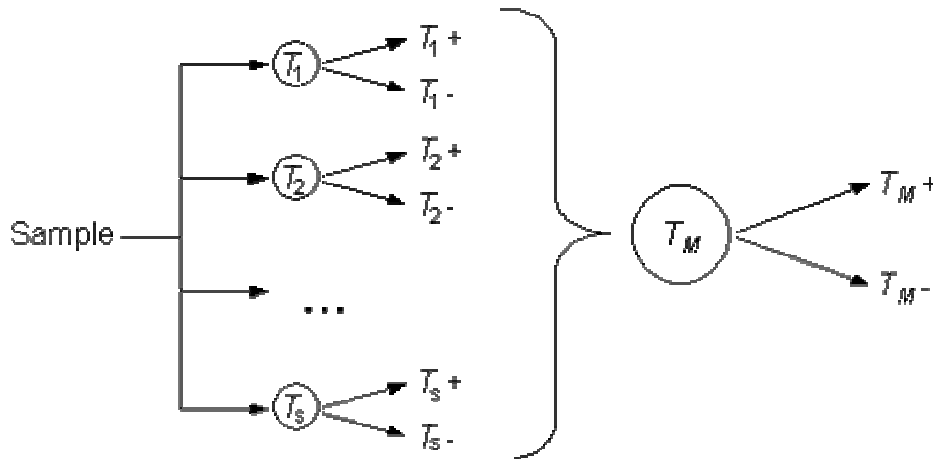
Figure 3. Parallel performance of multiple diagnostic tests.

This summary often follows a cut-off rule: "T_M+ if at least *c* out of *s* tests are positive". More complicated rules could be useful if one of the tests is extremely specific but not very sensitive. The cut-off *c*=1 leads to the most sensitive parallel test with sensitivity (under assumption of independent test errors)

$$Se_M = 1 - \prod_{i=1}^{s}(1 - Se_i)$$

and specificity

$$Sp_M = \prod_{i=1}^{s} Sp_i$$

where Se_i and Sp_i denote the sensitivity and specificity of the *i*th test, respectively. However, a possible correlation of test errors must be taken into account. For example, if two serological tests are combined, it is likely that both fail in animals that have low antibody levels for biological reasons. Likewise, false positive results can be dependent. The sensitivity and the specificity covariances can be quantified as

$$\gamma_{Se} = Pr(T_1+,T_2+|D+) - Se_1 Se_2$$

$$\gamma_{Sp} = Pr(T_1-,T_2-|D-) - Sp_1 Sp_2$$

and be used to establish Se and Sp of the parallel test (with cut-off $c=1$) as

$$Se_M = 1 - (1-Se_1)(1-Se_2) - \gamma_{Se}$$

$$Sp_M = Sp_1 Sp_2 + \gamma_{Sp}$$

(Gardner et al., 2000). In sequential testing, the result of a preceding test determines whether or not another test is done. In disease surveillance, one would typically encounter positive-sequential tests: the testing is continued after a positive result of the first screening test is obtained (Figure 4).
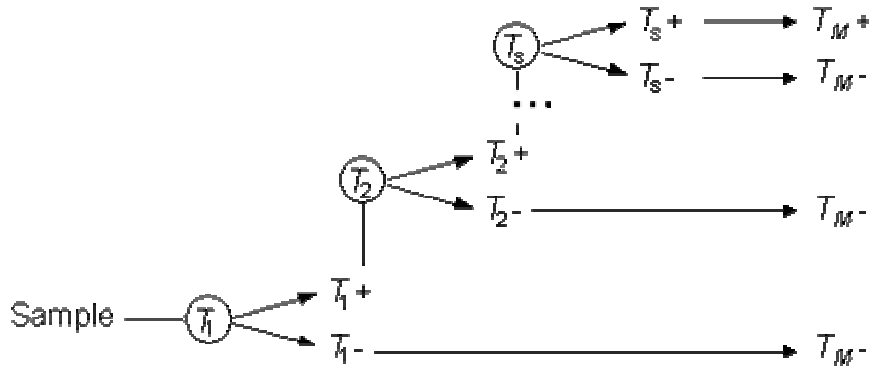
Figure 4. (Positive) sequential performance of multiple diagnostic tests.

The positive sequential testing strategy is typically chosen to increase the specificity of the multiple tests. A stopping rule (number of tests) must be defined. The positive sequential test has the sensitivity

$$Se_M = \prod_{i=1}^{s} Se_i$$

and specificity

$$Sp_M = 1 - \prod_{i=1}^{s} (1 - Sp_i)$$

under the assumption of independent test errors. If the correlations are nonzero, one should use

$$Se_M = Se_1 Se_2 + \gamma_{Se}$$

$$Sp_M = 1 - (1 - Sp_1)(1 - Sp_2) - \gamma_{Sp}$$

"Diagnostic testing systems" can be complex combinations of serial and parallel diagnostic tests, even in combinations with pooled tests and herd tests (see Herd Testing).

Herd testing

Test results of individual animals of one herd are often summarised to obtain a classification of a herd as test positive or negative. The performance of the herd classification process is dependent on

- Se and Sp of the diagnostic test used;
- n, the sample size for the herd testing;
- N, the herd size;
- c, the minimum number of positive individual tests to declare the herd as positive (herd cut-off);
- P_A, the within herd prevalence.

Under a binomial distribution model, which is valid for small sample sizes from large herds, one can use the apparent prevalence

$$AP = P_A Se + (1 - P_A)(1 - SP)$$

to derive the herd-level sensitivity

$$SeH = 1 - \sum_{y=0}^{c-1} \binom{n}{y} AP^y (1 - AP)^{n-y}$$

and herd-level specificity

$$SpH = \sum_{y=0}^{c-1} \binom{n}{y} (1 - Sp)^y Sp^{n-y}$$

The choice of the cut-off $c=1$ cannot universally be recommended but leads to simplified formulas

$$SeH = 1 - (1 - AP)^n$$

$$SpH = Sp^n$$

In practice, the choice of the cut-off and sample size should be optimised with regard to some criterion derived from *SeH* and *SpH*.

Bayes theorem and predictive values

The predictive value of a diagnostic test is a classical application of Bayes Theorem. Assume an animal is to be diagnosed for a disease with prior probability $Pr(D+) = P$. The prior probability in this context is often called the *pretest* probability. Its value is derived from the prevalence and other indicators of the disease (clinical signs, other test results) except the diagnostic test result.

Positive predictive value

The predictive value of a positive test result is the probability of disease ($D+$) given a positive test result ($T+$). Assume an animal tests positive and we wish to use this observation to find the posterior probability $Pr(D+|T+)$. According to Bayes Theorem, we need the likelihood (probability of observing $T+$ given $D+$) in order to find the posterior. According to Bayes,

$$Pr(D+|T+) = \frac{Pr(T+|D+)Pr(D+)}{Pr(T+|D+)Pr(D+)+Pr(T+|D-)Pr(D-)}$$

The posterior probability $Pr(D+|T+)$ is also called the positive predictive value (*PPV*). Replacing the conditional probability of the last equation with known symbols yields

$$PPV = \frac{SeP}{SeP+(1-Sp)(1-P)}$$

The interpretation of the *PPV* can be demonstrated using the 2x2 table shown before. We are only referring to the first row of the table (*T+* results). What proportion of *T+* animals (in the population) is actually diseased? The answer is clearly: *PPV = a/(a+b)*. The notation in Table 3 confirms the given formula.

Note: The data in a 2x2 table can be used to estimate the *PPV* as simple proportion *a/(a+b)* only if the prevalence in the study data reflects the population prevalence. This can be presupposed for the cross-sectional and partial verification design but not for the pre-stratified design.

Table 3. 2x2 table with expected cell probabilities

		Disease	
		+	–
Test	+	*Se P*	(1-*Sp*) (1-*P*)

	−	(1-*Se*) *P*	*Sp* (1-*P*)

Negative predictive value

The predictive value of a negative test result is the probability of "no disease" (*D*–) given a negative test result (*T*–). The *NPV* is the post–test probability of no disease given a negative test result, *NPV* = Pr(*D*–|*T*–). In formula,

$$Pr(D-|T-) = \frac{Pr(T-|D-)Pr(D-)}{Pr(T-|D-)Pr(D-)+Pr(T-|D+)Pr(D+)}$$

or using the simplified notation

$$NPV = \frac{Sp(1-P)}{Sp(1-P)+(1-Se)P}$$

It can be seen that the *NPV* also depends on prevalence. The negative predictive value is also used in an expression of the probability that a negative test result is a false negative

$$Pr(D+|T-) = 1-NPV$$

which is of great practical importance.

The link between pre- and post-test probability of disease

The odds of the post-test probability of disease, given a positive test result can be written as *POST(T+)=PPV/(1–PPV)* or, equivalently,

$$POST(T+) = \frac{Pr(D+|T+)}{Pr(D-|T+)}$$

Using Bayes' theorem, the numerator and denominator can be re-written as

where the second factor on the right hand side cancels out. The term

$$\frac{Pr(T+|D+)}{Pr(T+|D-)} = \frac{Se}{1-Sp} = LR(T+)$$

in the last equation is called the *likelihood ratio* of a positive test result *LR(T+)* and can be expressed in terms of *Se* and *Sp*. The Likelihood ratio of a negative test result is

$$\frac{Pr(T-|D+)}{Pr(T-|D-)} = \frac{1-Se}{Sp} = LR(T-)$$

More general, the likelihood ratio of any outcome X of a diagnostic procedure is

$$LR(X) = \frac{Pr(X|D+)}{Pr(X|D-)}$$

All this can be summarised into an important result: The post-test odds of disease for a given diagnostic outcome X is the product of the likelihood ratio of X and the prior odds of disease ($PRET$),

$$POST(T+) = LR(T+)PRET$$

$$POST(T-) = LR(T-)PRET$$

$$POST(T+) = LR(T+)PRET$$

Application of predictive values in disease surveillance

The predictive values can be used on the animal-level as shown above. For example, the positive predictive value can be written (normalising constant is omitted)

Pr(animal diseased | animal tests positive) $\propto$
Pr(animal tests positive | animal diseased) $\times$ Pr(animal diseased)

$PPV_{animal} \propto Se\ P$

Likewise, the positive predictive value of a herd test is

Pr(herd infected| herd tests positive) $\propto$
Pr(herd tests positive | herd is infected) ´ Pr(herd infected)

$PPV_{herd} \propto SeH\ P_H$

and the positive predictive value of a country region tested is

Pr(region is not free | surveillance positive) $\propto$
Pr(surveillance positive | region is not free) $\times$ Pr(region not free)

$PPV_{region} \propto Se_{surveillance}\ P_{region}$

Mathematically, it is not difficult to handle these three different levels of predictive values. Care must be taken to distinguish the unit level to which parameters refer. P and Se are on animal-level, SeH and P_H are on herd-level and $Se_{surveillance}$ and P_{region} are on

country or region level. The question is whether the necessary data and information exist in the context of documenting disease freedom.

Se can be estimated from standard test evaluation studies
P should be zero; a small design prevalence is assumed
SeH can be derived from *Se*, *Sp* and the sampling design
P$_H$ should be zero; a small design prevalence is assumed
Se$_{surveillance}$ can be estimated from the design of the surveillance system

A critical parameter is *P$_{region}$*, the prior probability of the region being not free from the disease. This issue is discussed in more detail under [Calculation of the Probability of Country Freedom](#).

Definition and legal issues

Rinderpest is one of the few animal diseases, where internationally accepted guidelines on the prevalence level exist. The Terrestrial Animal Health Code 2003 states in Appendix 3.8.2 in connection with the [Recommended Standards for Epidemiological Surveillance Systems for Rinderpest](#):

"Annual sample sizes shall be sufficient to provide 95% probability of detecting evidence of rinderpest if present at a prevalence of 1% of herds or other sampling units and 5% within herds or other sampling units."

In these notes, the term design prevalence denotes a set of fixed values concerning

P^*_H, the prevalence of infected herds or other sampling units in a country or zone (among herd prevalence)

P^*_A, the prevalence of infected animals within infected herds or sampling units (within herd prevalence).

Depending on the epidemiology of the infection, a non-homogeneous distribution of within herd prevalences may be relevant. The cited OIE guidelines only specify $P^*_H = 0.01$ and $P^*_A = 0.05$. The design prevalence must be found by international agreement. In the given context, these are hypothetical values.

Biological issues

In an outbreak situation of rinderpest in a naïve population, the design prevalences will soon be reached if no control measures are taken. In endemic foci, however, the within herd prevalence may be substantially lower ([James, 1998](#)). Setting a low value for the design prevalences will allow that the infection is detected soon after the introduction

into the non-endemic area. In an endemic area, low design prevalences will be useful to increase the chance of detection.

Statistical issues and application

The prevalence level to be detected determines the necessary sample size of a surveillance system or survey. The lower this level, the higher the sample size will be (Cannon and Roe, 1982). This will be demonstrated using the animal-level design prevalence P^*_A, which applies to infected herds. All formulas below are approximations.

Assuming a perfect diagnostic test and a large herd size compared to the sample size (ie, the binomial distribution assumption holds), the necessary sample size to detect the herd as infected with probability $1-\alpha$ is the smallest integer, greater than the logarithm to the base $(1-P^*)$ of alpha,

$$n > \alpha \log_{1-P^*_A}$$

The effect of diagnostic misclassification can be taken into account using the apparent (design) prevalence $AP^* = Se\,P^*_A + (1-Sp)(1-P^*_A)$. The sample size is

$$n > \alpha \log_{1-AP^*}$$

Mainly a lack of sensitivity is considered in this context. For imperfect specificity, a false positive classification of animals would help to correctly classify infected herds. Since positive test results are followed-up with highly specific tests, the overall specificity of the testing system can be assumed 100%.

For small herd sizes N, a hypergeometric distribution model should be used. The required expected number of diseased animals is established as $D^*=PN$, and the sample size is

$$n > (1-\alpha^{1/D^*})(N-D^*/2)+1$$

Under a misclassification model, the expected number of test positive cases is used, $T^*=AP^*N$, and the sample size is

$$n > (1-\alpha^{1/T^*})(N-T^*/2)+1$$

The effect of the design prevalence is demonstrated in Table 4 using the following values: $P^*_A = (0.005, 0.01, 0.05)$ and $Se = (1, 0.95)$. The calculations can be done using the programme FreeCalc (version 1.0 beta; Cameron and Baldock, 1998a,b).

Table 4. Sample sizes required to document disease freedom in a population (or herd) of 200 animals with probability of 95% for three different levels of animal-level design prevalence

	binomial model		hypergeometric model	
$P*$	Se=0.95	Se=1	Se=0.95	Se=1
0.005	630	598	200	190
0.01	314	299	164	156
0.05	62	59	55	52

The design prevalence is not only required for sample size calculations. In later sections, the concept of design prevalence will be used to estimate the sensitivity of a surveillance system.

Describing differential risk

It may be thought that the quantities P^*_H and P^*_A should not be the same for the whole country. If areas are known to be at higher risk (ie. close to endemic areas or entry points for imported animals), the expected prevalence in case of an outbreak may even be higher than the design prevalence fixed by international agreement. However, the high expected prevalence should not be used to replace the design prevalence for sample size calculations. It may be even relevant to over-sample risk areas in routine monitoring or surveillance. The reason is that this would allow an early detection of an outbreak, when prevalences are still low. Differential risk can also be associated to other factors such as age/birth cohort, production type and trade practices, etc. Thus, there is a need to recognize the clustering of a disease due to specific biological or other logical factors.

The identification of differential risk is based on conventional epidemiological risk factor studies and analyses when observed data is available. Such risk factors include the production type but also spatial or temporal (e.g. seasonal) factors. However, observed data are typically not available for exotic diseases in non-endemic countries. In such situations, differential risks may be quantified using methods of import risk analysis. Geographical and other risk factors can also be studied using the information of past outbreaks. For example, differential geographical risks for contagious bovine pleuropneumonia (CBPP) were investigated in Italy using data from sporadic outbreaks between 1990 and 1993 (Giovanni et al., 2000). On a higher scale, geographical risk analyses can be used to classify countries into risk groups, as for example in the context of BSE. This information can also be used for differential risk estimates within countries, based on closeness to borders with a high-risk region, or trade relationships.

Internal time period for analysis

Structured surveys and ongoing monitoring/surveillance activities differ importantly with regard to the internal time period for data collection and analysis. In both cases, the date of sampling and the characteristic course of the infection in affected animals are important. The latter can be characterised with at least two phases.

Latent period
> Time period between infection and the first possible detection. In case of clinical diagnosis, it is equivalent to the incubation period. In case of antibody serology, it is equivalent to the time until a detectable level of specific antibody is mounted.

Apparent period
> Time period between onset and termination of detectability. The termination can be due to self-cure or (disease-related) death.

Surveys

A survey produces estimates of proportions. Disease events are expressed as prevalence. Survey results are usually considered valid for the given time point of sampling. The sampling period is ignored in the analysis. However, the dynamics of the infection and the characteristics of the diagnostic methods applied must be considered for the interpretation of the results. The situation is represented schematically below (latent and apparent periods are indicated with dashed and solid lines, respectively, Figure 5).

Figure 5. Observations obtained by a survey in relation to time.

At the time point of the survey, only animals in the apparent period can be diagnosed. A non-detected outbreak or introduction of the disease in the past is not relevant in the situation of apparent disease freedom. On the other hand, the possibility of not detecting a current infection could be of concern. The probability of this false negative survey result depends on factors such as

- *Infection lag time*. The time between infection and survey. If this lag time is too long or too short, the outbreak will not be detected because of the apparent period or the latent period, respectively. An optimal lag time identifies a calendar time period, about which the strongest inferences can be made. The "optimal" lag time in terms of detectability is when the expected proportion of apparent animals is at its maximum. In practice, this period could be found by simulation modeling.
- *Latent period*. If the latent period is very long, current non-apparent infections may occur and the infection remains undetected.
- *Apparent period*. If the apparent period is short, infected animals will be difficult to detect. This period may be short for biological or other reasons, i.e. animals may die or be removed from the population.

The rate at which the value of information of surveillance data declines depends on the characteristics of the disease in question. An extreme example for a situation where surveillance data can be accumulated over long periods of time is bovine spongiform encephalopathy (BSE). Any BSE surveillance should account for the long incubation period and for the possibility that exposure and infection happened at an early life stage. The potential public health risk is associated with the event that infected animals enter the food chain. Therefore, evidence for freedom from BSE in selected birth cohorts can be accumulated over the entire survival time of the birth cohort in the population. A recent study has been conducted to establish the confidence for BSE freedom in selected birth cohorts in Denmark[4]. The following factors were found to be major determinants for the magnitude of confidence reached with the surveillance.

- **Number of cattle subjected to BSE testing**. In Denmark as in all European countries, the majority of tests are conducted on healthy slaughters (HS). Note that the value of testing HS, in terms of contribution to confidence, is low compared to risk surveillance streams (fallen stock, suspects, etc.).
- **Design prevalence**. Only animal-level design prevalence was considered as no clustering of BSE within herds occurs. The design prevalence for risk animals was 15 times the value chosen for HS, based on empirical rate ratios in European countries.
- **Diagnostic sensitivity**. In case of BSE, *Se* is a function of the age at exposure/infection and incubation period. Both determinants are unobserved in the animals tested. The stochastic modelling of *Se* introduced uncertainty in the outcome of confidence estimates.

The Danish case study suggests that the confidence of freedom from BSE in Danish cattle born after March 1999 is at the order of 85 to 87%. These results are based on a design prevalence of 1/10,000 in HS and 15/10,000 in risk animals and age-specific *Se*.

[4] Böhning and Greiner, 2005. Report of project P12 at the International EpiLab (full report available on request; paper published 2006).

Ongoing surveillance

Surveillance produces a stream of observations. The sampling can be described using rates (animals/time). The dynamic changes in the population and herds must be considered. Disease events can be described in terms of incidence density (cases per animal time). A surveillance activity is represented schematically below (latent and apparent periods are indicated again with dashed and solid lines, respectively, Figure 6).
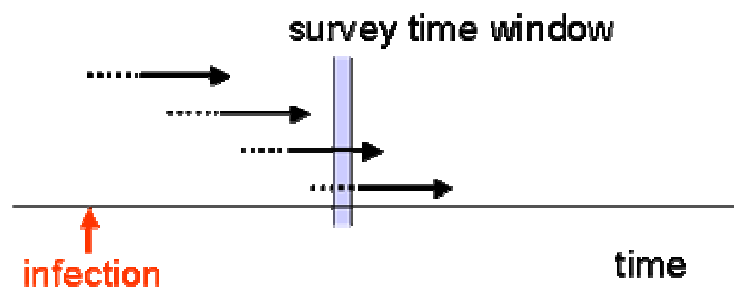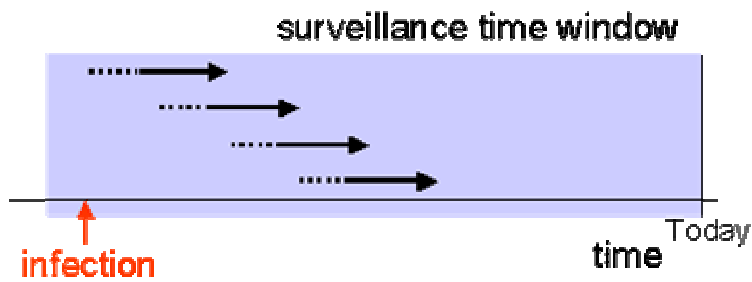
Figure 6. Observations obtained by an ongoing surveillance in relation to time.

The size of the time window for the analysis of surveillance data (grey area in the graphic) indirectly determines the sample size and thus the probability to detect infections occurring during that time. The chronologically last time window contains the most recent observations but also earlier time windows may contribute valuable information. The value of such "historical" information depends on the disease in question. Methods were suggested to discount such observations (Schlosser and Ebel, 2001). In an ongoing surveillance, infections can be detected as soon as infected animals become apparent. Therefore, ongoing surveillance is required when there is continuous risk of outbreaks. The factors that determine the probability of not detecting an outbreak in a surveillance program are similar to those given for the survey. An important difference is that the *infection lag time* is here defined as the time between analysis of the surveillance results and the infection.

Statistical evidence from surveys and surveillance

The statistical evidence from surveys can be established using standard methods developed for this purpose. The statistical evidence from ongoing surveillance activities are usually analysed with statistical methods designed for surveys. For example, the cumulated sample size over a time window is treated in the same way as a survey sample size.

For illustration, we assume that the sampling rate (i.e. animals tested per week) is constant over time[5]. The choice of a time window is therefore linked to the choice of the sample size for statistical analysis. It would be tempting to declare long time periods, e.g. one year, as time window in order to attain a large power. The choice of an optimal time window for analysis of surveillance data is an interesting but yet unresolved methodological problem. The followings aspects should be considered when time windows are to be defined.

- The question of surveillance must be addressed clearly. Is it required to make a statement about freedom form disease at one given time point (e.g. date of request)? Or is it required to make a statement about disease freedom over a

longer period (which would be sensible for infections with long incubation period)?
- What is the relative value of negative surveillance results obtained in the past for the current probability of being disease free? It seems logical that the latent period and apparent period are important factors here.
- What minimum window size should be used to obtain sufficient statistical power?

[5] If the sampling rate is not constant over time, this should be considered in the analysis.

Documenting evidence for disease freedom in small herds

Standard statistical methods for documenting disease freedom presuppose that the population and herd sizes are large. However, a problem to reach acceptable confidence may occur when herd sizes or population sizes are small. Obviously, typical values for animal-level design prevalence are not directly applicable for small herds. Consider for example a herd of size 10. The smallest possible non-zero prevalence is 1/10 or 10%, which may well be greater than the nominal design prevalence. In the following, we briefly outline the special situation encountered when freedom from disease is to be documented for small herds[6]. We assume throughout this section that all animals of such a small herd are selected for diagnostic classification.

We consider first the herd-level sensitivity, which is given as

$SeH = 1 - (1-Se)^m \geq Se,$

where m denotes the number of truly diseased animals in the herd. It can be seen that the animal-level Se is the lower bound of the herd sensitivity and applies if there is exactly $m=1$ infected animal present in the herd. A useful working definition of a small herd is that the expected number of infected animals is less than 2, given such a herd is infected. Equivalently, herds smaller than $2/P^*_A$. For example, using the within-herd design prevalence 5%, herds smaller than 40 can be considered small with the justification that infected herds would contain only a single infected animal. From the definition follows also that $Se_H = Se$ for small herds.

The confidence about freedom from diseases in the stratum of small herds will be based on a sample of h small herds and can be given using a binomial (bin) or alternatively by a Poisson (poi) model:

$$C_{bin} = 1 - (1 - P^*_H Se)^h$$

$$C_{poi} = \sum_{j=0}^{h} [1 - (1-Se)^j] e^{(-\tau)} \tau^j / j!$$

where $r = hP^*_H$ is the Poisson rate parameter. For the scenario of $Se = 0.8$ and $P^*_H = 0.02$, we obtain a confidence of $C_{bin} = 0.918$ and $C_{poi} = 0.916$. It can be shown by simulation that the binomial model fits better than the Poisson model (smaller deviations between predicted and true confidence), whereas the Poisson model is more conservative (it fails less frequently to yield the nominal confidence level).

The required sample size h can be established for the binomial model as the smallest integer greater than or equal to

$\log(1-C) / \log(1-P^*_H Se)$.

The sample size according to the Poisson model can not be given in a closed form but can be computed using numerical optimisation ([EXCEL spreadsheet available from the authors](#)). For the scenario of $Se = 0.8$ and $P^*_H = 0.02$ and a required confidence of 95%, the sample size should be 186 and 188 according to the binomial and Poisson model, respectively.

[6] The topic is elaborated in more detail by [Greiner and Dekker (2005)](#).

Bayesian inference

Bayes theorem can be used for statistical inference. This will be shown using estimates of proportions. A typical situation is as follows. We have a prior assumption or knowledge about a parameter P

$$Pr(P)$$

This implies that we regard P quasi as a random variable with a probability distribution attached to it. We observe some data X that can be described with a probability model involving the parameter P. The likelihood L

$$L = Pr(X|P)$$

indicates the probability of observing X given the parameter P. The probability of P given the prior and the data is now called the posterior distribution of P,

$$Pr(X|P) = \frac{Pr(X|P)Pr(P)}{Pr(X,P)}$$

The denominator is often omitted and therefore ($\square$ means "proportional to")

$$Pr(P|X) \propto Pr(X|P)Pr(P)$$

Example

Assume 100 swine from four small herds are gathered at a slaughterhouse. The number of swine from the herds i=1,…,4, is n_1=10, n_2=50, n_3=30 and n_4=10, respectively. Unfortunately, the swine are not labelled and cannot be assigned to one of the four herds. Assume, we know that the disease prevalence in the four herds is P_1=0, P_2=.2, P_3=.4 and P_4=1. The first swine is slaughtered and found to be diseased (X=1). What is the prior and posterior probability that this swine originates from farm 1, 2, 3 or 4?

The prior probability is given by the fractions $\Pr(P_i) = n_i/100$. The likelihood is given by the Bernoulli density

$$Pr(X|P_i) = P_i^X(1-P_i)^{1-X}$$

For X=1 (one data point), the likelihood reduces to $\Pr(X \mid P_i) = P_i$. The posterior distribution of P is

$$Pr(P_i|X) = \frac{Pr(X|P_i)Pr(P_i)}{\sum_i Pr(X|P_i)Pr(P_i)}$$

The denominator on the right hand side is the summation of the likelihood over the complete range of P_i, in this case a list of four discrete values. Table 5 below shows that this denominator just serves as a scaling factor to assure the posterior distribution sums to 1 as in all well-behaved PDFs. The important information is the product of the likelihood and the prior (column 4). We can see how X has changed our prior assessment (column 3: pig is most likely from farm 2 with P_2=0.1) to the posterior assessment (column 5: pig is most likely from farm 3 with P_3=0.5).

Table 5. Hypothetical data for demonstration of Bayesian inference.

Farm i	P_i	Prior $\Pr(P_i)$	Likelihood x Prior $\Pr(X=1 \mid P_i)$	Posterior $\Pr(P_i \mid X)$
1	0	0.1	0.00	0.000
2	0.1	0.5	0.05	0.167
3	0.5	0.3	0.15	0.500
4	1	0.1	0.10	0.333

It makes sense that the posterior for farm 1 is zero. If we know farm 1 is disease-free, the probability that the diseased animal is from farm 1 is zero. The example of a discrete prior is somewhat artificial because it is a very strong assumption that P can only assume four discrete values. Usually one would rather use a continuous prior. This approach is central to the application for diagnostic tests and will be described in the next section.

The beta prior for proportions

We assume that we have a prior distribution of the binomial parameter P. This prior PDF should have the following properties

- The support range should be from 0 to 1 (usually; but not wider).
- It should be continuous.
- It should be possible to obtain this PDF based on study data such as K/n.
- It should be possible to obtain this PDF using expert opinion.
- It should be possible to update the prior PDF with new data or new expert opinions.
- It follows that the posterior and the prior PDF should be from the same family of distributions.

All these requirements are met by the beta distribution.

$$Pr(P) \sim Beta(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} P^{a-1}(1-P)^{b-1}$$

where the fraction, which is the inverse beta function, is sometimes omitted because it doesn't include P. For the beta distribution we have the mean

$m = a/(a + b)$

and variance

$s^2 = ab/((a + b + 1)(a + b)^2)$.

How to get the prior?

Option 1
> We can use a "conservative" approach and assume that we don't know anything about P. An uninformative ("flat") prior with the coefficients $a=b=1$ (prior 1 in graph below) reflects this situation. Note that the Beta(1,1) is identical to the continuous uniform distribution over the range [0,1]. In a sense, this is a strong assumption, because we are saying that every value for P between 0 and 1 has exactly the same plausibility.

Option 2
> We could use study data such as K/n to obtain the coefficients
> $a = K + a'$
> $b = n - K + b'$
> where a' and b' denote the coefficients of the "prior of this prior", usually $a'=b'=1$. Taking the pilot study results $K/n=2/10$ as informative prior, we would obtain the coefficients $a=3$ and $b=9$ (prior 2).

Option 3

Assume only the parameter estimate $P=0.4$ and the standard error $s=0.07$ is given in a study. We can use the formulae for mean (m) and standard error (s) as above and obtain the corresponding

$$a = \frac{m(m - 2m^2 + m^3 - s^2 + s^2 m)}{s^2(1 - m)}$$

$$b = \frac{m - 2m^2 + m^3 - s^2 + 2s^2 m}{s^2}$$

in our case, $a=19.2$ and $b=28.8$ (prior 3).

Option 4

Using the additive properties of the beta distribution, we can combine the priors 2 and prior 3 into $a=22.2$ and $b=37.8$ (prior 4).

Option 5

We can translate expert opinion about the probability of P into a beta prior. Since the beta is defined by two parameters, we only need two orientation points to derive the PDF. For example, we could ask

-What is the most likely value (m) of the parameter P?

-What are the limits of the interval in which the P is located with 95% certainty?

Example

Our expert says, the most likely value for P is 30%. She is 95% sure that P is within the range of 20% to 40%. Recovering the parameters a and b from the 2.5[th], 50th and 97.5[th] percentile can be achieved numerically. For example, one can find values a and b such that the difference between the given percentiles and the corresponding values of the cumulative PDF becomes minimal. This methods yields $a=23.7$ and $b=54.9$ (prior 5). See also Suess et al. (2002) for details of deriving beta distributions for a given mode and single percentile (eg 95%).

Option 6

Other approaches to summarise and quantify prior study data can be developed from the principles of meta-analysis. The combination of expert opinions can be done by weighting, iterative review to consensus (delphi-technique) or discussion. One can also refrain from solving discrepancies and use the range of responses in a sensitivity analysis (how much changes the posterior depending on the choice of the expert). In our example, we can seek to combine the results of the two studies with Beta (22.2,37.8) (prior 4: Beta) with the expert opinion (prior 5). We can express the latter prior in terms of K and n as

$K = a$-1
$n = a+b$-2.

This yields the values *K*=22.7 and *n*= 76.6, which reflects the expert certainty. Updating yields the new coefficients *a*= 22.7+22.2=44.9 and *b*=53.9+37.8=91.7 (prior 6). This is a new prior that summarises all available information described above.

Bayesian inference provides a flexible tool to incorporate various sources of evidence and is appealing – at least from the pragmatic point of view. However, it should be noted that the approach is sometimes criticised (the conflict between "frequentist" and Bayesian statistics). The selection of the prior is in principle subjective. Any use of priors (from experts or based on data) should be governed by the principles of

- science-based assessments (coherence with scientific knowledge);
- transparency and full documentation of methods including reasoning about best-case or worst-case scenarios; and
- using reference priors (agreed value or even a non-informative prior, best-case, worst-case) for reporting likelihood-based results and analysis of their impact by sensitivity analysis.

Figure 7. Six candidates for a prior distribution for the binomial parameter P

(see text).

Prior 4 — a=22.2; b=37.8

Prior 5 — a=23.7; b=54.9

Prior 6 — a=44.9; b=91.7
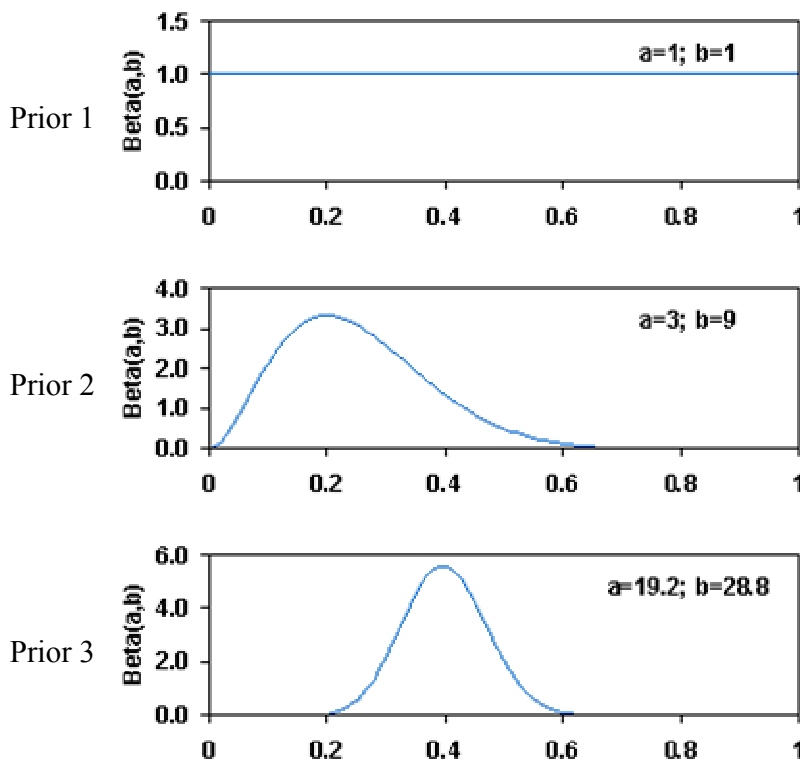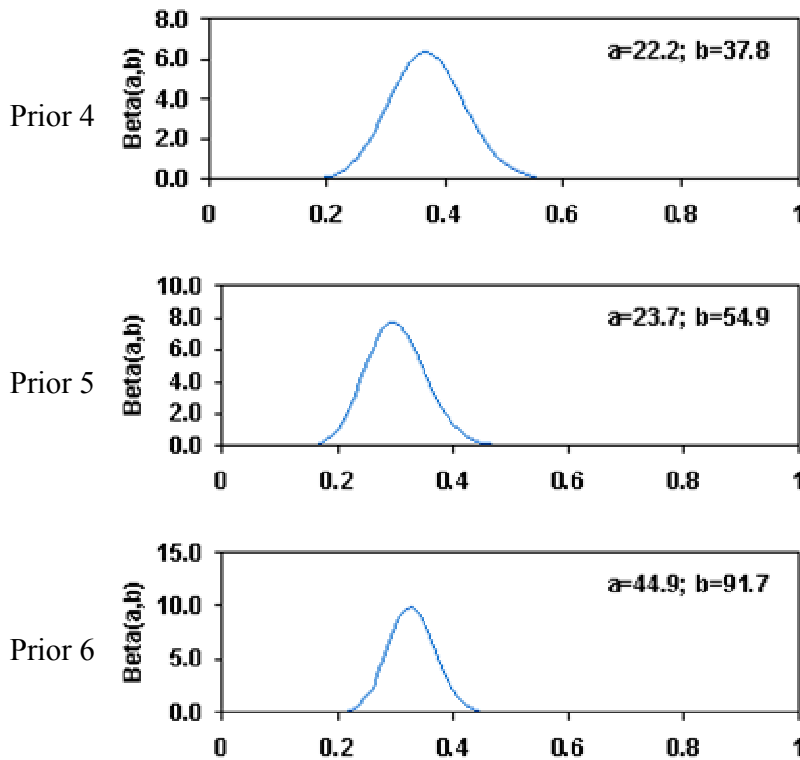
Obtaining the posterior

The prior (what we know before) times the likelihood (what the data tell us) gives us the posterior (prior knowledge updated with the new information). We consider again the binomial parameter (proportion) P. The likelihood of the data K/n is given by the binomial distribution (see previously). If the prior is given in terms of a beta distribution, this will lead (as can be shown by algebra) to a posterior distribution which is also beta. This is meant by the expression that the beta is the conjugate prior to the binomial. It is not a must to select a beta prior but it is quite nice if the posterior remains a beta and can be used as prior for the next step of updating.

$\Pr(P \mid X) \square \Pr(P) \Pr(X \mid P)$

Posterior $\square$ Prior Likelihood

$\text{Beta}(K+a, n-K+b) \square \text{Beta}(a,b) \text{Bin}(n,K)$

Example

Consider the prior that summarises all the evidence described above (prior 6), which is expressed as a Beta(44.9, 91.7). Assume we have now the study data $K/n= 21/50$. Updating the coefficients yields $a=21+44.9=65.9$ and $b=29+91.7=120.7$. The posterior is

now given as Beta(65.9,120.7). See Figure 8 for graphic representations of these distributions.
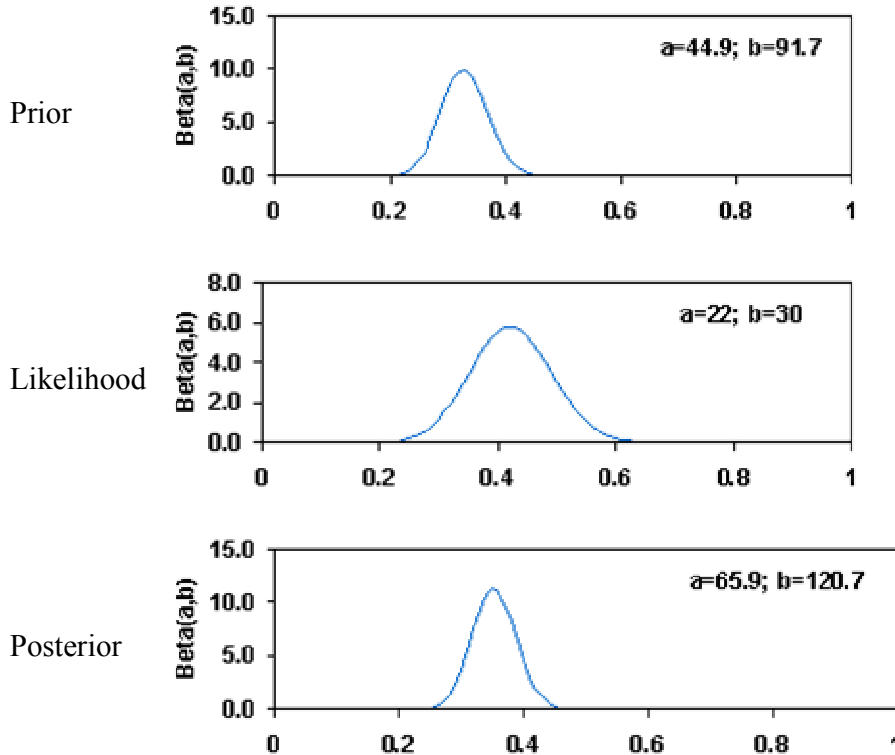
Figure 8. Beta distributions for prior, likelihood and posterior

In this example the prior has been very influential. The way, the prior is generated should always be reviewed very critically.

The posterior distribution can be summarised in many ways, e.g. using the mean, mode, percentiles, etc. In Bayesian inference it is legitimate to state that the true parameter is within the range of the (1-p)th and (p)th percentile with probability 1-2p, which is not possible for the confidence interval based on frequentist analysis. The Bayesian version of the CI is therefore called "credibility interval" to avoid any confusion between the two concepts.

Markov chain Monte Carlo (MCMC)

Practical applications of Bayesian inference can be much more complex than the example above. One may want to work with non-conjugate priors and with combinations of likelihood and priors that cannot be easily solved analytically. A practical problem is often the integration of the joint probability of the data and prior (denominator of Bayes formula). MCMC is a very powerful and flexible solution to such problems. With the

(currently) free software WinBUGS ([Gilks et al., 1996](#); [Spiegelhalter et al. 1999](#)) the application of MCMC is relatively straightforward. However, it is strongly recommended to seek support by a statistician when using the technique. BUGS stands for "Bayesian inference using Gibbs sampling" and it is the most commonly used platform for MCMC analysis. A Windows-based version is also available free of charge.

Surveillance, Monitoring and Surveys

Disease monitoring describes the ongoing efforts directed at assessing the health and disease status of a given population. Sampling of individuals from the population to assess disease or health status may be ongoing or repeated. The disease may be specific infectious diseases, specific production diseases, or disease/health in general. The population may be defined at the national, regional, or herd level. For an alternative definition see [Table 6](#).

Disease surveillance is used to describe a more active system and implies some form of directed action will be taken if the data indicate a disease prevalence or incidence above a certain threshold. Similar to disease monitoring, sampling of individuals from the population to assess disease or health status may be ongoing or repeated and the population may be defined at the national, regional, or herd level. Surveillance is usually directed at a specific disease. Disease surveillance systems require three components:

- defined disease monitoring system
- defined threshold for disease level. (pre-defined critical level at which action will be taken)
- pre-defined directed action (interventions)

The term "surveillance" was first used during the French Revolution and it meant "to keep watch over a group of persons thought to be subversive". The term has been used extensively by epidemiologists and other animal health professionals in the context of monitoring and controlling health-related events in animal populations. Disease surveillance is the key to early warning of a change in the health status of any animal population. It is also essential to provide evidence about the absence of diseases or determine the extent of a disease that is known to be present. The two terms "surveillance" and "monitoring" are often used interchangeably in animal health programs. Animal disease surveillance is watching an animal population closely to determine if a specific disease or a group of diseases makes an incursion. Monitoring of animal diseases focuses on identifying a disease or a group of diseases to ascertain changes in prevalence, and determine the rate and direction of disease spread. Therefore, monitoring by definition lacks action to prevent or control a health problem. Surveillance, on the other hand, includes an action to prevent or control the health problem that is being monitored. In actual field situations, monitoring usually follows early reaction should surveillance activities indicate introduction or spread of a disease. Many of the approaches used to implement monitoring can be used for surveillance and vice versa. In

practical terms, the distinction between these two terms often becomes blurred. The differentiation, however, pertains more to the objectives than the approaches applied.

The term "survey" is used to indicate an investigation or a study in which information is systematically collected for a specific aim or conceptual hypothesis. The time frame for this type of investigation is a specific and usually short period of time. This is in contrast to surveillance and monitoring which involve the on-going systematic collection of data and information. Surveys are more frequently used to answer a specific research question oriented toward a scientific and exploratory purpose. Approaches used for survey studies are similar to those used for surveillance and monitoring. In concept, a series of surveys can be considered as a monitoring system that may transition into a surveillance system if action is taken to prevent or control the disease. Therefore, the three terms "surveillance", "monitoring", and "survey" share several common components and hence, it is logical to consider them as a single topic for the purpose of these notes.

Some authors have proposed the use of the term "monitoring and surveillance system" (MOSS) to summarize the concepts and approaches (Stärk, 1996; Noordhuizen et al., 1997, Doherr & Audigé, 2001). In that context, "monitoring" describes a continuous, adaptable process of collecting data about disease and their determinants in a given population, but without any immediate control activities. "Surveillance" is a specific case of monitoring where control or eradication measures are implemented whenever certain threshold levels related to the infection or disease status have been exceeded. By definition, surveillance is therefore part of any disease control program (Noordhuizen et al., 1997; James, 1998).

Table 6. Definitions of monitoring and surveillance from three textbooks on veterinary epidemiology.

Textbook	Monitoring	Surveillance
Martin et al. 1997 (page 259)	Animal disease monitoring describes the ongoing efforts directed at assessing the health and disease status of a given population.	The term "disease surveillance" is used to describe a more active system and implies that some form of directed action will be taken if the data indicate a disease level above a certain threshold.
Thrusfield, 1995 (page 22)	Monitoring is the making of routine observations on health, productivity and environmental factors and the recording and transmission of these observations.	Surveillance is a more intensive form of data recording than monitoring.
(Page 358 and 360)	The routine collection of information on disease, productivity, and other characteristics possibly related to	An intensive form of monitoring (q.v.) Designed so that action can be taken to improve the health status of a population, and therefore

	them in a population.	frequently used in disease control campaigns.
[Noordhuizen et al. 1997](#) (page 379)	Monitoring refers to a continuous, dynamic process of collecting data about health and disease and their determinants in a given population over a defined time period (descriptive epidemiology).	Surveillance refers to a specific extension of monitoring where obtained information is utilized and measures are taken if certain threshold values related to disease status have been passed. It, therefore, is part of disease control programs.

Data collection for Surveillance

One of the main components for any surveillance is the collection of data, which can be classified as either passive or active. Unfortunately, some authors have generalized these terms as labelling surveillance as passive vs. active ([Lilienfeld & Stolley 1994](#)). A surveillance system cannot be passive if an action is part of its definition.

An active collection of data for surveillance or survey is referred to as the systematic or regular recording of cases of a designated disease or a group of diseases for a specific goal of monitoring or surveillance. A population by specific location and/or time period is usually defined for the system. This should provide each individual within the defined population with a known and often equal chance of being selected. The identification of such appropriate population depends on the event of interest, its expected prevalence, and the available diagnostic tests.

Information about the health-related event might be collected from owners by interview or mail. Biological samples might be collected during farm visits, at abattoirs, knackeries or carcass rendering plants. In addition, the screening of animal medical records, either the files or electronic databases, for specific entries, or biological sample banks for specific pathogens or lesions, can be considered part of the active collection of data for a surveillance system. Examples of such a system include the tuberculosis and brucellosis programs that is routinely performed in several countries of the world, infectious bovine rhinotracheitis (IBR) and enzootic bovine leucosis (EBL) sero-surveys in Switzerland ([Stärk 1996](#)), abattoir screening for contagious bovine pleuropneumonia (CBPP) in Switzerland ([Stärk 1996](#)), BSE screening of fallen stock and emergency slaughtered cattle in Switzerland and Europe ([Doherr et al. 1999](#) & [2001](#)) and of "downer cows" in the United States (http://www.aphis.usda.gov/oa/bse). Other examples would be the Scrapie surveillance in the United Kingdom ([Simmons et al. 2000](#)), and postal surveys for scrapie in the UK, the Netherlands and in Switzerland ([Morgan et al. 1990](#); [Schreuder et al. 1993](#); [Hoinville et al. 1999](#) & [2000](#); [Baumgarten et al. 2002](#)). Some national surveillance system includes mail or interview questionnaires as well as the collection of biological samples for laboratory testing ([Traub-Dargatz et al. 2000a & 2000b](#); [Kane et al. 2000](#); [Wagner et al. 2000](#)).

A major disadvantage of the active data collection for surveillance is that it is very costly when the occurrence of the target disease is rare. The lower the disease prevalence, the larger the sample size required for detection. Once the prevalence becomes very low (< 0.1%), it often is not feasible to further increase the sample size due to funding constraints, limitations in the working capacity of diagnostic laboratories or simply because of limitations of the chosen test system: the tests are not sensitive and specific enough to distinguish between zero and very low prevalence levels. The situation changes from low prevalence to the probability of disease freedom. Instead of prevalence estimation, the focus is now on the identification of a health-related event if it occurs in the defined population above the design prevalence. An example where all animals in a defined population are tested is the mandatory fallen stock surveillance for BSE in Europe. Within this program, due to the expected very low prevalence of detectable cases of < 0.1%, all fallen cattle older than 24 months have to be examined. Between January 2001 and April 2002, the average prevalence in this "high-risk" target population was approximately 0.05% - or one case per 2000 samples tested (http://europa.eu.int/comm/food/fs/bse/testing/bse_results_en.html).

The passive collection of data involves the reporting of clinical or subclinical suspect cases to the health authorities by health care professionals *at their discretion* (Lilienfeld & Stolley 1994). Therefore, the validity of the system depends solely on the willingness of these professionals to secure the flow of data. In veterinary medicine, the passive collection of data can be influenced by the awareness and level of knowledge of a particular disease among veterinary practitioners and producers or owners of animals. Another important component for this type of data collection is the availability of a diagnostic laboratory scheme to support and confirm cases. The main limitation of passive data collection is inconsistency in the data collection for different diseases and among communities that provide the data. Thus, a comparison of various passively collected surveillance data should be approached with caution. Disease awareness, educational level of the surveillance data providers (practitioners, regulatory veterinarians, and owners/producers), and the nature of the disease under the surveillance are the major elements in the effectiveness of the surveillance. For instance, a disease with a high case-fatality rate may be reported more frequently than a disease with a low case-fatality rate. A disease with more public awareness (for example, that has had extensive advertising or educational programs) may be more likely to be reported as compared to a disease with less awareness, even though its true prevalence and incidence are lower. It should be also noted that the use of the passive collection of data would not ensure the early detection of a disease.

Passive collection of data for surveillance can identify a change in a pattern that may warrant further investigation. Typically then an active method of collection of data can be implemented. For instance, the first few BSE cases found in UK at the initial epidemic were reported using the passive collection of data for surveillance that was not designed specifically for collection of BSE cases. Then, a surveillance program was implemented to actively collect data for BSE.

Some countries have used the term "notifiable animal diseases" for those diseases that are required by law to be reported. Most of the OIE List A and specific zoonotic diseases fit the criteria to be on the notifiable list. Although these notifiable diseases by definition should require active collection of data for surveillance, most countries have used passive collection of data for surveillance. The main reason for this is the lack of a well-planned study design to maintain and actively detect cases for these diseases.

Other authors (Dufour & Audigé, 1997; Doherr & Audigé, 2001) have classified surveillance activities by the method of data collection into three classes (passive, active and sentinel networks). Baseline data collection was considered a subcategory of passive collection. In our opinion, a disease trend which is determined by surveillance is different from baseline data. Disease trends can change overtime and the use of the term baseline data in this context may be misleading. The term "sentinel networks" is a method to actively collect data for surveillance using a selected sample to represent the population.

Targeted Surveillance

The term "targeted surveillance" is becoming popular and it principally refers to focusing the sampling for the surveillance on high-risk population (i.e., targeted population) in which specific commonly known risk factors exist. An example of a target population is fallen cattle stock in Europe because this high risk group of cattle has more BSE than otherwise healthy cattle. Another target population is the specific hamburger meat processed in large quantities, which is associated with a greater risk of *Escherichia coli* O157:H7 than in unprocessed meat.

The main purpose of implementing this surveillance approach is to increase the efficiency of the system. This design is appropriate when the following two conditions exist: the disease under consideration is less common in the general population than in the targeted group and specific risk factors are established or known. Therefore, prior knowledge about the disease and its epidemiology is required before this design can be considered. Occasionally, targeted surveillance is used to ensure the absence of a specific disease from a highly susceptible population. For instance, the purpose of the surveillance of downer cows and cattle with suspected neurological signs in the USA for BSE is mainly to provide evidence of the absence of BSE.

Targeted surveillance is an effective design to purposely implement an action that can reduce the impact of a disease rapidly. An example of this approach is nosocomial infection surveillance in a veterinary teaching hospital in which equine colic cases are targeted for Salmonella surveillance. This is due to the fact that these cases are more susceptible to this infection than other hospital admitted cases (Tillotson K et al., 1997 and Kim et al. 2001).

The impact of the change in trade regulations on surveillance planning and implementation

In a country, the demand for scientifically reliable surveillance system has coincided with a reduction in budgetary and human resources among the government veterinary services. Several countries therefore, have attempted to identify the most efficient methods to satisfy the national and international requirements for animal health. During the last decade, numerous methods and approaches for surveillance in animal health programs have been discussed or proposed. The most important outcome from this type of exploration is the determination of absence of the disease or its agent from a country i.e., when prevalence of a disease is at or near zero. The objective of this type of surveillance is to provide evidence (with known confidence) that a disease or pathogen, if present in a zone or country, is present at or below an acceptably low (practically undetectable) prevalence. While it will probably continue to be commonly used, the term 'freedom from disease' is potentially misleading. 'Freedom' implies complete absence, which is analogous to the now unacceptable concept of 'zero risk'.

Current approaches generally involve the compilation of evidence from a range of sources, and the use of this evidence to put forward a convincing argument about a country's disease status. One source of evidence that is commonly used or demanded is a structured statistically valid survey. The primary advantages of the use of surveys are that well-established theory and methodologies exist, and they are able to produce a quantifiable probability estimate for the presence of disease. International regulations increasingly demand that the level of proof of disease status meets quantitative standards, e.g. that the probability of the presence of disease at a prevalence in animals of 0.2% or greater is less than 1%. Other sources of evidence that may be used include passively collected data, an assessment of the quality of the veterinary services, livestock movement history, geographical and environmental factors, abattoir monitoring, sentinel herds, etc.

It has become clear that there are a number of problems with this approach. Structured surveys are often too expensive or impractical to achieve the level of proof required. This is due to the very large sample sizes necessary when the prevalence is very low, and when applied tests do not have very high sensitivity and specificity. This is further complicated by variability in sensitivity and specificity, and a lack of reliable estimates of these test accuracy parameters for the population under study.

As a result, true disease status cannot always be determined through the use of surveys alone. It is necessary to combine all the different sources of evidence available to assess the overall probability that a disease does not exist or is below the design prevalence.

It is proposed that these problems may now be overcome through the use of a range of different analytical methods, including:

- A standardized approach to scenario tree analysis and stochastic simulation to estimate the power of complex surveillance/survey outcomes. These notes address this approach.
- Improved use of techniques to elicit and combine expert opinion as additional information to data generated by surveillance outcome (K. Stark, Personal communication),
- Methods to adjust the value of data sources for surveillance based on the time that has passed since their generation (Schlosser and Ebel, 2001),
- Bayesian approaches to the combination of data from multiple sources of surveillance system (Suess et al., 2002).

Regardless of whether one or a combination of the above approaches is used, there is a need to ensure that the principles behind it, and the tools required to implement it, are sound and made widely available to those who need it. The use of these approaches would require specific tasks:

1. Identify all possible sources of evidence for the absence of disease.
2. Analyze each source independently through the construction of a scenario tree, to estimate the probability that an infected animal, if present, would be identified by the surveillance. At each branch of the tree, probability estimates and ranges are required. These should be derived from reliable data sources, if available, or formally structured expert opinion methods, if not.
3. Use stochastic methods to determine a point estimate of the probability of detecting disease based on a scenario tree, as well as the probability distribution around that estimate (to provide measures of confidence).
4. Adjust all values for the time elapsed since data collection.
5. Combine the estimates from all different sources of evidence to provide an overall probability and confidence level.
6. If the resultant probability is inadequate to meet international standards, either a) use sensitivity analysis to determine which method may be most effective at increasing the level of confidence, or b) conduct a (relatively small) structured survey to fill the 'probability gap'.

Identifying potential data sources

In the following section the term *data* is used in its broadest sense, covering all factual information which can be used for analysis or as a basis for reasoning or decisions.

In the process of analysing surveillance data to estimate the sensitivity of a surveillance system, we find that we need data for three different purposes:

- describing the detailed structure of components of the system for drawing up scenario trees
- estimating branch probabilities and proportions
- analysing the results of the system as it has been applied

There are many, varied SSCs in operation around the world, and the data sources needed to model them are extremely numerous, so it is not possible to give exhaustive lists of potential data sources. Here we suggest the types of data source which will be commonly available to draw on, and illustrate with a few specific examples.

Drawing up a scenario tree

When constructing a scenario tree for a component of the surveillance system, we are attempting to create a model of the process. For this we need detailed information on the structure of the process:

- the sequence of events involved in the process
- structure of the livestock production system in the country
- epidemiology of the disease including likely risk factors
- sampling / testing strategy
- practical implementation details

Take as an example the Australian export testing programme in relation to Western Australian freedom from bovine Johne's disease (paratuberculosis; BJD). Many trading partners require that all cattle imported from Australia should have a negative serological test for BJD before shipment. How can Western Australia use the data generated by these tests to support its claim to freedom from BJD?

To draw up a scenario tree we must know the structure of the cattle export industry. What cattle are exported, and to which countries? Which importing countries require BJD testing, and what type of cattle do they import? Are they beef cattle for breeding, dairy cattle, beef cattle for slaughter, or perhaps cull cows? What breed, age and sex are they? Where do these cattle originate? We must ensure that our analysis only includes cattle from the state of Western Australia.

We need to have a good grasp of the epidemiology of BJD:

- factors affecting the probability that a herd will be infected
- factors affecting the probability that cattle will be infected within an infected herd
- factors affecting the probability that the serological test will give a positive result in infected cattle

Clearly, we need good knowledge of how animals for shipment are managed, including criteria for selection (on the farm, in sales, and from the exporter's stock), when they are tested, and any potential or motivation for misclassification in the records.

Where will all this information come from? A team of experts with collective knowledge covering the disease and the process should be good enough for drawing up the scenario tree, but it might be necessary also to look at cattle export records, references on BJD and possibly laboratory submission records.

Estimating branch probabilities and proportions

Examples of the data needed are given here under headings relating to the 3 node types. See the following section on Development of scenario trees for descriptions of node types.

- *category node* proportions
  Most commonly these relate to the structure of the livestock production system under consideration, and the data needed are such things as:
    - records from/of
        - markets
        - industry administrative bodies
        - livestock movements
        - farm and animal productivity
        - identification schemes
        - farmer organisations
        - farm registrations
        - abattoirs
    - results of surveys yielding industry/farming statistics
        - census data related to livestock production systems
    - spatial data
        - GIS databases
        - map references
        - postcodes
        - latitude and longitude
- *risk category node* relative risks
  What are needed here are realistic estimates of relative probabilities of infection among branches of the node, in the hypothetical scenario of the country being infected. Potentially valuable data sources include
    - prevalence surveys
    - literature (text books; journal articles)
    - reports from countries where the disease is endemic
    - climatic / environmental data
        - for the country / region
        - for areas from which prevalence data are available
- *infection node* probabilities
  Design prevalences are not estimated from data, and are set separately (see under Design prevalence)

- *detection node* probabilities
  There are many potential sources of data depending on the SSC being modelled.
  Common ones include
    - laboratory test performance data
      - literature
      - laboratory conducting the test
    - laboratory records
      - submissions
      - number and nature of tests performed
      - results
    - sensitivity of inspection or examination by veterinarians
      - literature
      - records
        - laboratory submissions
        - veterinary clinic records
        - prescription records
        - drug sales
      - sensitivity of inspection or examination by meat inspectors
        - literature
        - records of meat inspection findings and inspections conducted
        - surveys / trials
      - sensitivity of inspection or examination by abattoir-based surveillance program inspectors
        - literature
        - trials
      - sensitivity of inspection or examination by farmers
        - production records
        - veterinary clinic records
        - prescription records
        - laboratory records
        - quality assurance programme records
        - farm records

As an example of data required for estimating branch probabilities and proportions, we will consider a scenario tree for the Danish diagnostic system applied to poultry. A tree has been drawn up as shown in Figure 9. Nodes, branch probabilities to be estimated, and data sources used for each are shown in Table 7

The surveillance unit in this case is the batch of broilers, the house of layers or breeders, or the flock of backyard birds. Danish poultry are divided into four industry sectors (broilers; layers; breeders; backyard) using the INDUSTRY SECTOR category node. Within each industry sector there are nodes describing the probability that a farm will be infected, the probability that a unit will be infected within an infected farm, and a series of detection nodes:

- Farmer consults veterinarian
- Vet sends samples to lab
- Lab performs test
- Lab isolates agent

The branch proportions for the INDUSTRY SECTOR node are based on the number of farms in each sector. These data come from the Danish central husbandry register (which lists all livestock producers who supply animals or produce to other people, along with the type of enterprise and numbers of livestock on the farm); the Danish Poultry Council (an industry body which maintains records of commercial poultry producers); and results of a door-to-door census of backyard (i.e. unregistered) flocks in designated infected areas during a recent disease outbreak.

The unit (within farm) standard design prevalence is of interest since there are so few units per farm (average of 2.25 for broiler farms; 1.4 for layers) and for backyard flocks the unit is the whole flock. Clearly it is not possible for a farm to be infected with less than one unit infected, so a unit-level design prevalence of one infected unit per farm is used. The design prevalence is thus the reciprocal of the average number of units per farm, and this varies from sector to sector. The data source for the numbers of units per commercial poultry farm was the Danish Poultry Council databases.

Branch probabilities for detection nodes are estimated using data from many sources, as listed below. For example, the *Farmer Sensitivity of detection*, or Pr(Farmer calls vet | infected house) is estimated from records of disease (high mortality) events in poultry houses (denominator) and matched records of veterinary consultations for the numerator. High mortality in poultry houses is derived from weekly mortality figures recorded by the Danish Poultry Council. Records of veterinary consultations are found in veterinary clinic records, laboratory submission records, and the national veterinary drug prescription database.

Table 7 Data sources for branch probabilities and proportions for simplified scenario tree of Danish poultry diagnostic system.

Node	Branch	Probability/Proportion	Data sources
INDUSTRY SECTOR	Broilers	*PrBroilers*	1. Central Husbandry Register 2. Danish Poultry Council databases 3. Limited census of backyard flocks
	Layers	*PrLayers*	
	Breeders	*PrBreeders*	
	Backyard	*PrBackyard*	
FARM STATUS	infected	P^{*}_{H}	Design prevalence (not based on data)

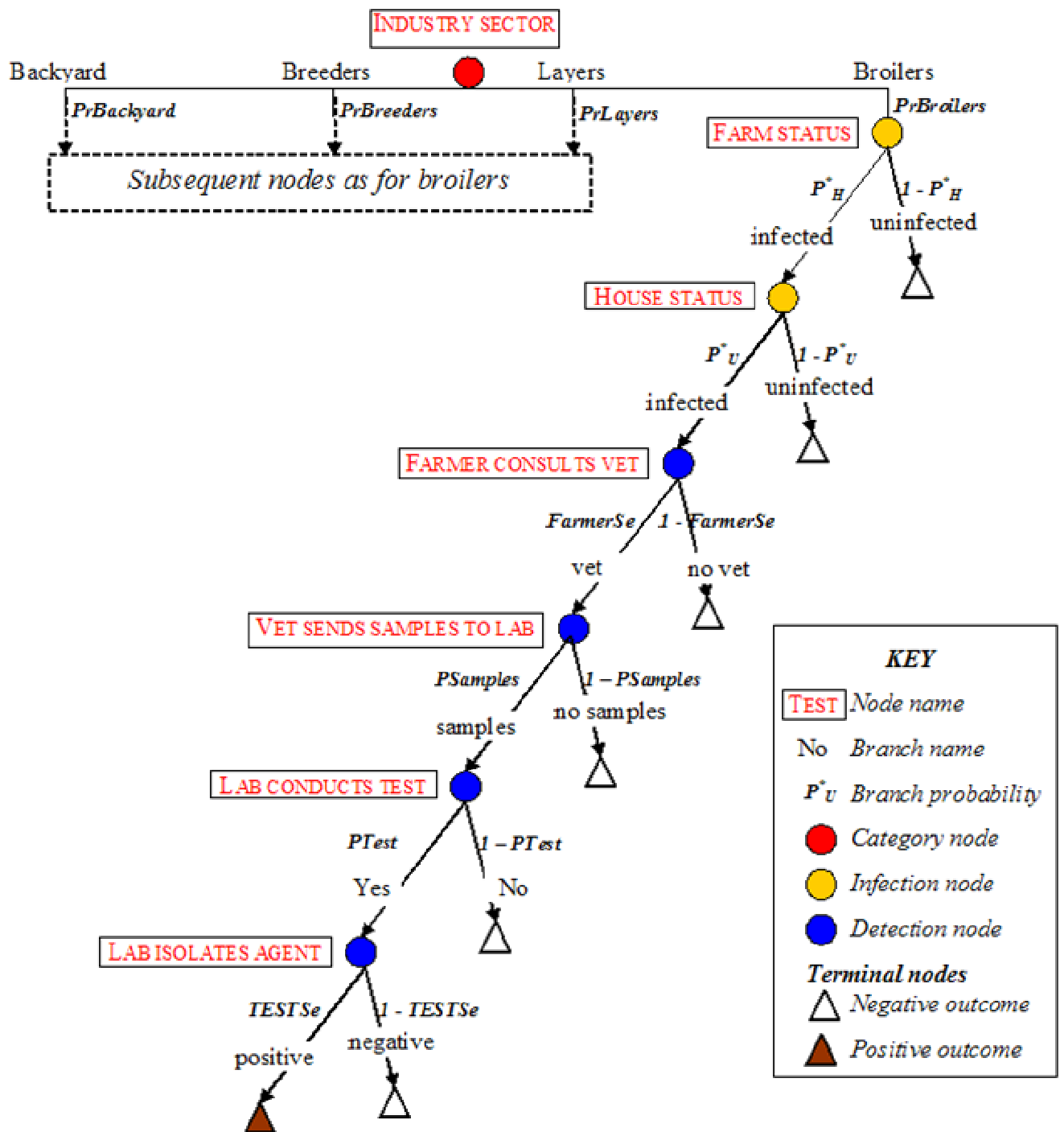|			
	uninfected	$1 - P^*_H$	
HOUSE STATUS	infected	P^*_U	Design prevalence (Danish Poultry Council databases)
	uninfected	$1 - P^*_U$	
FARMER CONSULTS VET	Vet	$FarmerSe$	1. Danish Poultry Council mortality records 2. Poultry veterinary clinic records 3. Laboratory submission records 4. Veterinary prescription database
	No vet	$1 - FarmerSe$	
VET SENDS SAMPLES TO LAB	Samples	$PSamples$	1. Poultry veterinary clinic records 2. Laboratory submission records 3. Veterinary prescription database
	No samples	$1 - PSamples$	
LAB CONDUCTS TEST	Test	$PTest$	1. Laboratory submission records 2. Laboratory internal records
	No test	$1 - PTest$	
LAB ISOLATES AGENT	Positive	$TESTSe$	1. Literature 2. Expert opinion
	Negative	$1 - TESTSe$	

Figure 9. Simplified scenario tree for Danish poultry diagnostic system.

Analysing results of application of the SSC

For calculating the probability that the SSC would have detected the disease if it were present, we need data describing the units actually processed by the surveillance system. Essential data are

- characteristics of each unit processed with regard to each of the factors included in the model;
- herd (or other group of units) from which each unit comes
- number of units processed, and
- their dates of processing.

Potential sources include

- laboratory records
- surveillance database
- abattoir records
- population description data
  - census
  - survey
  - central registers
  - industry organisations

(Where some or all these essential data are not available, it is possible to simulate them from summary statistics for the SSC and the livestock industry involved; see section below on Simulated data.)

We will take the example of the Danish serological surveillance program for classical swine fever (CSF). CSF sero-surveillance is based on an ongoing program of blood sample collection from pigs at slaughter and testing for CSF antibodies using ELISA. Samples tested for CSF are a sub-sample of a larger abattoir-sampling program aimed at Aujeszky's disease surveillance. The protocol dictates that all culled boars and 10% of culled sows in Southern Jutland and 10% of culled boars and 5% of culled sows in other parts of Denmark are tested. These specimens are not collected according to a specified protocol, but on an ad hoc basis. The total number of specimens collected for 1998, 1999 and 2000 were 28,073, 27,255 and 20,142 respectively. Two databases, recording the results of testing are available, one at the individual animal level, and one providing summary totals of the results for batches of specimens sent in by abattoirs. Factors (and their levels) affecting probabilities of infection which are included as nodes in the tree are

- COUNTY (South Jutland; other)
- FARM TYPE (breeder; slaughter)
- AGE (adult; grower)
- SEX (male; female).

The CSF sero-surveillance databases supply

- Farm ID

- Sex
- Date of receipt of samples

Through farm ID further data are available from the Central Husbandry Register, giving

- COUNTY
- FARM TYPE

Simulated data

Where data are available for analysis of the tree, they should be used. If using a spreadsheet for calculation of the model, cross tabulation of the data is necessary at the start, to establish how many units go with which group, etc. Group-level sensitivities can then be calculated (see Analysing the tree), and the analysis is completed straightforwardly.

When these data are not available, for example if animals are not identified, or if there is no way of matching animals to herds at the abattoir, then the only useful data available may well be one figure: the number of units processed in the time frame of interest. If there is also no information on the proportion of processed units falling into each level of an important grouping factor, then you have a problem! To analyse the tree successfully, it is necessary to simulate the missing data. This is simple if statistics are available to form the basis of such simulations, e.g.

- number of herds
- herd size
- regional distribution of herds
- proportions of animals processed coming from different areas
- proportions of animals processed associated with different industry sectors or production systems or farm types
- other descriptive statistics for the population of processed units.

In a stochastic simulation model the necessary proportions and numbers may be simulated from such summary statistical data, giving an end result very close or identical to that which would have been derived from real data.

The one figure which cannot be simulated, and which is crucial to the analysis, is the number of units processed.

To analyse the tree assuming that all units are independent when this is not the case (see Analysing the tree) should be a last resort, since it will result in an over-estimate of confidence in the SSC.

Describing surveillance system components

A surveillance system component (*SSC*) is a surveillance activity which, in itself, can contribute evidence to disease freedom; it has the capacity to detect disease if / when it occurs. This includes both general and targeted surveillance activities; both "active" and "passive".

How does our selected *SSC* detect disease? What is the step-by-step process which must occur for a surveillance unit to be infected and detected? The starting point is always that the country is infected: disease/infection is present in the country.

Scope of the model

Immediately it is apparent that we must define clearly

- whether we are talking about disease or infection;
- which agent or syndrome is under consideration;
- which livestock species we are covering;
- exactly what geographical area does our *SSC* cover?

Such questions are generally easily answered, but not always. If we are lucky the International Animal Health Code will give us clear guidelines.

Unit of analysis

The unit of analysis is that unit for which results are generated by the *SSC*. This will often be an individual animal, but may be a sample, a pooled sample, a group of animals, a batch of animals, a house, a farm, etc. For each unit passing through the surveillance process an outcome is recorded, and every outcome of the process applies to a single unit.

Coverage

The population "covered[7]" by the *SSC* is called the *SSC reference population*. This is the population about which statements will be made concerning sensitivity of the surveillance system, or probability of freedom.

Comprehensive vs incomplete coverage

Some *SSC*s may be said to have "comprehensive" coverage of the *SSC reference population*. The obvious example is the clinical diagnostic system, in which all animals

have a probability of being observed with disease, and a diagnosis being pursued. Probabilities may vary from one sub-population to another, but all animals are covered by the system to a greater or lesser degree. Where the *SSC* has comprehensive coverage of its reference population, this must be borne in mind when defining tree structure, and there will be no need for records of the numbers of animals processed by the system – they are all processed, and what are needed are data describing the population. Thus, where a system has comprehensive coverage, the number of units for calculating *CSe* is the entire population.

Where a component has comprehensive population coverage, inclusion of risk nodes will have no effect on the resulting *CSe* value unless there is also a difference in detection probabilities between risk groups. This is because if there is comprehensive coverage it is not possible to target surveillance according to risk, because the entire population is included in the surveillance.

Where the *SSC* does not have comprehensive coverage of the *SSC reference population* (i.e. "incomplete" coverage – only a restricted number of units are processed by the *SSC*), the coverage of the surveillance process being analysed is determined by the units actually processed, which may be limited by various factors including:

- location
- management / production system
- industry sector
- species
- age
- sex
- temporal constraints, e.g. seasonal incidence of disease or seasonal application of the surveillance process

Where a surveillance component has incomplete coverage, *CSe* calculation is based on the actual numbers of units sampled, rather than the whole population, as units not sampled do not contribute to knowledge of the population status. Where coverage of a component is incomplete it may be either "representative" of the population (for example a random survey) or "biased" (targeted to specific risk groups).

[7] The term *coverage* is used here to refer to the extent to which the *SSC* processes units from each level of each relevant factor dividing up the population (as in the list above); the term *representativeness* is used to refer to the extent to which the number of samples processed in each relevant population subgroup is proportional to the size of the subgroup.

Accounting for lack of coverage

One of the key benefits of a scenario-tree approach is that is allows the analysis of data from biased surveillance activities, particularly where sampling is specifically targeted at high-risk sub-populations. Targeting sampling at higher-risk sub-populations provides an equivalent *CSe* for a smaller overall number of samples compared to representative

sampling, improving surveillance efficiency and reducing costs. In fact, from a purely statistical perspective, for a given number of units sampled, targeting exclusively high-risk groups will maximise *CSe*, whereas each additional sample taken from a low-risk group instead of a high-risk group will reduce *CSe*.

A secondary result of targeting sampling in this way is that some sub-populations (usually low-risk ones) will be under-represented in the data. This is an appropriate and expected outcome if targeting of sampling is to provide any benefit. However, this could lead to concerns of inadequate coverage of the *SSC reference population* if some sub-populations are severely under-represented or not represented at all.

One approach to address this issue would be to undertake a mathematical adjustment of *CSe* for coverage in the various sub-populations, using weightings based on their proportional representation. This approach effectively negates any benefit of targeting for risk, and is therefore not recommended. The recommended alternative is to recognise that coverage is incomplete and not representative and to include this when reporting the results of the analysis.

Whether or not the coverage of the various sub-populations is considered adequate will depend on the level of coverage, relative risks and contributions to *CSe* for each sub-population. For example, a low-risk group with poor representation would be of less concern than a high-risk group with comparably poor representation. It is therefore essential when reporting results to provide an assessment and comparison of the level of coverage of the various sub-populations and discuss the potential impact of lack of representativeness on the results.

Where one or more sub-populations are under-represented it would be appropriate to report the assumed relative risk values, and contribution to *CSe*, along with the level of coverage, and some justification for why this is thought to be reasonable and appropriate. In addition, where a sub-population is not included in the surveillance at all it might be appropriate to explicitly exclude this group from the *SSC reference population* and the surveillance system, depending on the importance of this group in the disease epidemiology and tree structure.

Temporal applicability

The time period to which a single analysis of the surveillance data applies must be clearly defined. Factors which will influence the time period for analysis include

- Epidemiology of the disease: speed of spread; seasonal spread; seasonal occurrence; duration of disease in a unit
- Time course of the surveillance process; e.g. monthly disease reporting system; on-the-spot serological testing; mycobacterial culture
- Economic and political importance of the disease; surveillance for a disease whose presence will result in loss of trade or severe economic hardship should

probably be assessed repeatedly over short time frames, compared with that for a disease whose presence is of marginal concern.

- Time course of the production system(s) under surveillance: e.g. units or groups which are isolated from the rest of the population for substantial periods; units with fixed production periods (e.g. tanks of fish; batches of broilers).
- maximum time an infection can remain undetected in the population despite the disease awareness of the professional animal health workers

In general, unless there are good reasons for selecting a different time period for analysis, we suggest:

- one month for rapidly spreading, high consequence diseases
- one year for slow moving diseases

The *SSC*'s time period for analysis is only relevant when considering an ongoing surveillance activity, and in this case the data from successive time periods will be used sequentially to derive an updated estimate of the probability that the population is free of disease (at the design prevalence), as described under Temporal Discounting. If this process is followed, the choice of time period is not crucial from the perspective of the end result, since the probability of disease introduction over time is included in the analysis, and in general the final estimate of Pr(freedom) does not vary much with length of time period; so the required frequency of reporting (and associated factors as outlined above) is the major driving force behind choice of time period.

Outcomes

In the context of scenario tree analysis, a surveillance process has two possible outcomes: positive and negative. A positive outcome occurs when a unit in the surveillance process is identified as being diseased, and a negative outcome occurs when a unit in the process is not identified as being diseased.

The meanings of the terms *positive* and *negative outcomes* should be clearly defined, and are usually obvious, e.g. isolation of virus. Since the surveillance process includes all follow-up testing to resolve initial positive laboratory test results, a positive result in a serological test is generally not the definition of a positive outcome from the surveillance process. Follow-up testing may well not be confined to the unit which generated the suspicious test result.

Specificity

*SSC*s used to provide evidence that a country is free from disease have, by definition, perfect specificity. In a disease-free environment, any positive surveillance process result will inevitably be investigated fully until a clear decision can be made on whether this was a true or false positive result. If it was a true positive, the population is infected, and

any analysis of *SSC*s to develop confidence in disease freedom is no longer relevant. If it is subsequently identified as a false positive test result, it is then a further negative outcome from the surveillance process, and can be used along with the others towards *SSC* sensitivity. The *SSC* should be seen to encompass all necessary follow-up testing to resolve potential false positive outcomes. It can then be said to have perfect specificity. Other authors have taken a similar approach (Cannon 2002; Dufour *et al* 2001).

For some diseases, the occasional positive surveillance outcome may be within some allowable prevalence limit. While this may be acceptable within certain specific trading contexts or definitions, these notes address only the disease-free scenario, and the question of how to quantify confidence derived from negative surveillance findings. It should be noted, however, that the scenario tree methodology is theoretically perfectly capable of analysing data from a surveillance process that also yields false positive results.

Development of scenario trees

A scenario tree models the process of disease detection by a *SSC*, starting from the country being infected. It includes all factors affecting the probability that a surveillance unit will be infected, and all those affecting the probability it will be detected. The tree may be thought of as tracing the probabilities that a single unit, chosen at random from the population, will fall into one of these groups:

- positive outcome
- negative outcome.

In this context, a negative outcome includes

- unit is tested with negative result
- unit is not processed.

Functions of a scenario tree in analysis of a *SSC*

1. to visualise and document the logical and practical structure of the surveillance process
2. to define the interrelationships of factors affecting
   o Pr(infection)
   o Pr(detection)
3. to clarify and describe the steps involved in analysis of the *SSC*
4. to summarise or average probabilities of infection and detection across factor levels where factor level data are not available for units processed.

The tree defines the probabilities for each individual unit to be infected and detected. If factor level data are available for all units, the tree simply serves to describe the process and logical structure for analysis (1 – 3 above); each processed unit's and group's Pr(detection) and Pr(infection) are calculated individually in analysis. If data are missing, the tree is used to calculate weighted averages for Pr(detection) and effective Pr(infection) across factor levels.

Figure 10 shows a generalised scenario tree that will be used for examples and exlanation of the methodology throughout the rest of this guide.
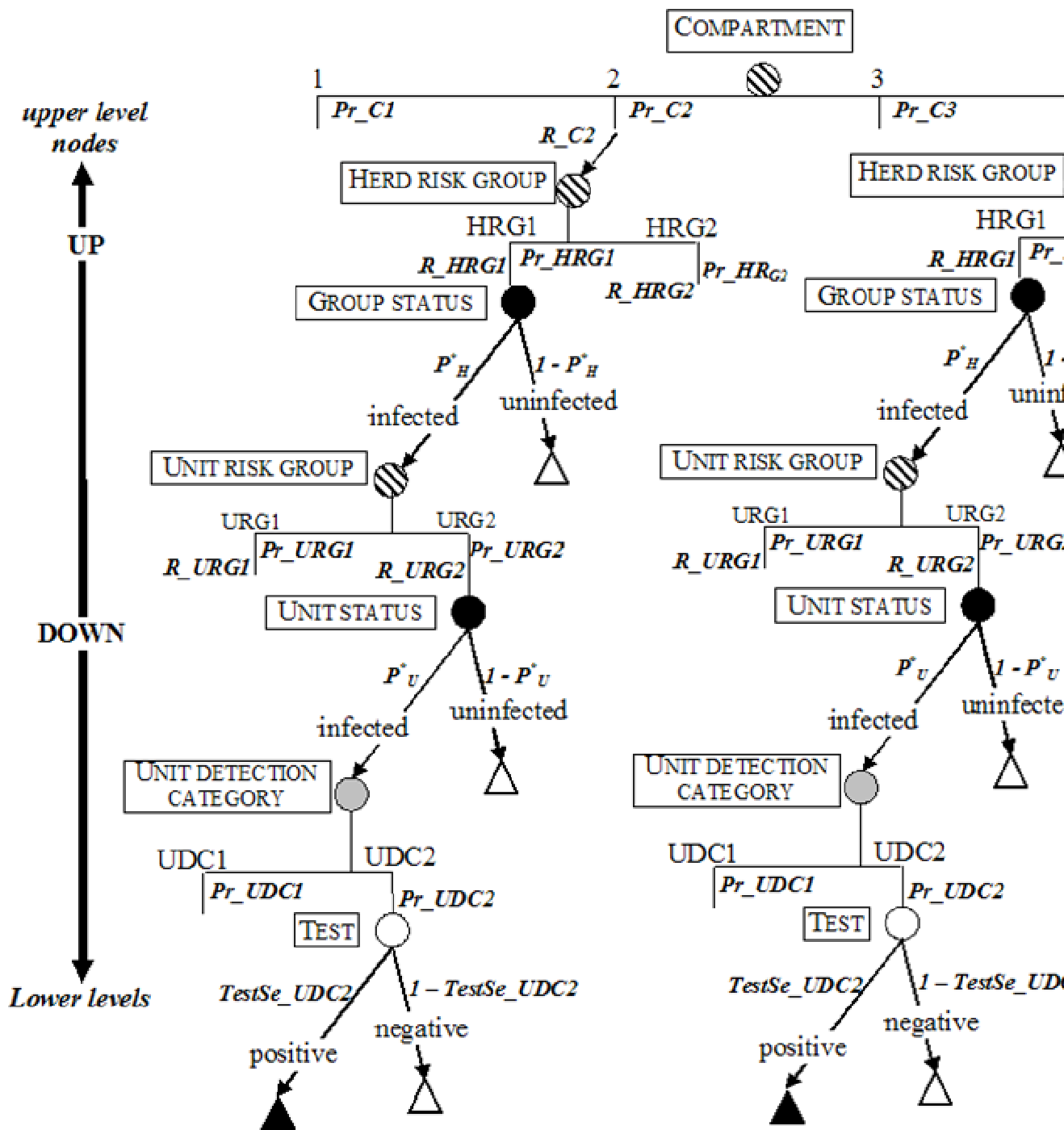
Figure 10. Stylised scenario tree (only 2 of four main branches completed; assume other 2 identical).

Node types

There are three main node types, each of which is illustrated in Figure 10:

- Infection nodes
- Detection nodes
- Category nodes, which are subdivided into
  - risk category nodes (risk nodes); and
  - detection category nodes

Rules for inclusion, exclusion and ordering of nodes follow.

Scenario tree structure is similar to decision tree structure, except that all nodes (excluding *terminal* (or *end*) nodes) are *probability* (or *chance*) nodes, and have no *value* (or *utility*) associated with them. Infection and detection nodes are effectively chance nodes, and the sum of their branch probabilities is 1. Category nodes are similar to chance nodes, with population proportions replacing branch probabilities, such that the sum of their branch proportions is 1. Risk category nodes also have a risk (of infection) associated with each branch. How each node type is handled during analysis is described in the following sections.

Infection nodes

An infection node specifies the infection status of a unit or group of units. It has 2 branches; *infected* and *uninfected*. Examples include

- ANIMAL INFECTION STATUS
- HERD INFECTION STATUS

Detection nodes

A detection node represents any event, action, choice, procedure or test which contributes to an infected unit's detection in the surveillance process being modelled. In other words, detection nodes all contribute to the *unit sensitivity* of the process. Typically, such nodes placed sequentially describe the process of detection of infection, and have two branches, representing positive and negative node outcomes. Examples include

- FARMER CALLS VETERINARIAN
- BLOOD SAMPLE TAKEN
- ELISA TEST CONDUCTED
- ELISA TEST RESULT

Category nodes

A category node simply defines categories into which units or groups of units fall. Rather than probabilities, each branch has a *proportion* attached to it; the proportion of groups (farms) in each COMPARTMENT in Figure 10, for example. Category nodes are included because they represent factors which affect the probabilities that a unit will be infected or detected, or because their branches represent essential segments of the *SSC reference population*, which must be covered by a surveillance system supporting disease freedom. Category nodes are divided into risk category nodes and detection category nodes.

The population proportion attached to a branch of a category node is the proportion of units / groups *processed* in the *SSC* (*PrSSC*). If the SSC has comprehensive coverage of its reference population, this is the same thing as the proportion of units / groups falling into that category in the reference population (*PrP*). As will be seen below, these branch proportions are used for various purposes, and it is important to ensure that the proportion used is appropriate for the circumstances. (There are circumstances in which the *population* proportion of units / groups is needed.)

A category node has two or more branches, one for each level of the factor represented by the node. Within each branch of a category node the units (or groups, as appropriate) are homogeneous with regard to risk of infection or probability of detection; if they are not, category nodes to subdivide the population further should be inserted. Examples include

- administrative subdivisions such as state, province or county (one branch per subdivision, with the proportion attached to each branch being the proportion of groups (or units, as appropriate) which originate in the subdivision, among those processed in the *SSC*)
- sex (one branch per sex, with branch proportions representing the proportions of male and female animals among those processed in the *SSC*).

At each category node, branch proportions must sum to 1.

Risk category nodes

Risk category nodes (risk nodes) represent factors dividing the *SSC* reference population into subsets with different risks of being infected (at the design prevalence). Each branch has an associated differential risk (*R*), as well as a branch proportion (*PrSSC*). (Refer to Incorporating differential risk for estimation of *R* values.)

Detection category nodes

These represent factors affecting the probability of detection, and each branch has an associated branch proportion, as described under Category nodes.

Grouping levels

When the tree is analysed for multiple units assessed by the surveillance process, sensitivities of detection must be calculated at different *grouping levels* within the tree structure. In the sample tree shown in Figure 10, the process has a *unit sensitivity*. From this, the probability that a unit will be infected, and the number of units processed from the group is calculated the *group sensitivity*, and from the group sensitivity for each group processed from the compartment and the probability that a group is infected, is calculated the *compartment sensitivity*. The GROUP and COMPARTMENT "levels" of the tree are referred to as *grouping levels*.

While on the subject of levels, COMPARTMENT is a "higher level" node than GROUP STATUS.

Ordering of nodes

The first rule for ordering of nodes is that there are no hard-and-fast rules for ordering nodes. We have found, however, that the tree will be easier to design and comprehend, and more importantly, easier to analyse, if the following guidelines are followed. The *order* of nodes in a scenario tree is the same as the *sequence* in which they are placed. The first node is the *root node* (generally COUNTRY INFECTION STATUS in disease freedom scenario tree models) and the last nodes are *terminal nodes*, representing the outcomes (positive or negative) of each limb of the tree. Between the root and the terminal nodes we place our infection, category and detection nodes in an order or sequence, running from high to low levels. Each path from the root node to a terminal node is referred to as a *limb* of the tree.

A few principles first:

- Each node is conditional on all previous (higher level) nodes in that limb of the tree
- Order has no effect on the total number of nodes
- The ordering of nodes influences the nature of the conditional relationships; some conditional relationships reflect the form in which probability estimates are available, while the probabilities of the inverse relationships are difficult to calculate or estimate. For example, the probability that a herd is infected given that it is in region 1 is more easily grasped than the probability that it is in region 1 given that it is infected. As another example, if information on cattle farms is collected and held at the regional government level, the proportion of cattle herds in each region that are dairy herds will be more readily available than the proportions of dairy herds in the country that are located in individual regions.

And now some rules and guidelines:

1. Nodes should be ordered in decreasing size of groupings of units, as one goes down the tree.
2. Keep nodes in chronological order of events within the surveillance process; don't have the ELISA being done before the sample has been collected.
3. Place infection nodes above detection nodes. This is essential for easy analysis where it is necessary to account for clustering of disease within groups (most analyses).
4. Category nodes may be placed among infection nodes or among detection nodes; a risk category node above the infection node to which it applies, and a detection category node above the detection node(s) to which it applies.

Factors to include

The scenario tree is a model of a single unit in the *SSC* (selected at random from those processed in the SSC), from which may be calculated its probability of a positive outcome in the surveillance process. This unit is already selected for inclusion in the *SSC*. Bearing this in mind, the factors that should be included in the scenario tree are those that influence the probability that a unit will be:

- infected or
- detected.

These factors are represented in the tree by nodes of the types described above.

The set of factors included should be the minimum required to describe the system accurately. Factors not to include are all those which have no significant effect on the probability that a unit will be infected or detected. In a *SSC* with comprehensive coverage of the *SSC reference population*, factors affecting probability that the unit will be sampled or tested are all factors affecting the probability of detection. In a *SSC* which does not have comprehensive coverage, factors affecting the probability of selection for inclusion in the *SSC* should not be included, since the tree is modelling units which have already been selected.

It is important not to include more factors than are necessary in the model, for the usual modelling reasons:

- unnecessarily complex models are less easily understood by the modeller, and particularly by others
- the more variables (with associated uncertainty distributions) that are included, the more uncertain the model outputs, so there need to be significant benefits from including any additional factor
- given the tendency of disease to cluster in groups, it is usually necessary to analyse the model taking this into account, generally at each grouping level in the tree. For every grouping level included there is therefore an accompanying loss of sensitivity of (or confidence in) the surveillance process.

Clearly all factors directly describing the probability of detection or the probability of infection will be included; such things as

- HERD INFECTION STATUS
- UNIT INFECTION STATUS
- FARMER CALLS VET
- LABORATORY CONDUCTS TEST

These factors will be modelled as infection nodes and detection nodes. Factors describing characteristics of different sections of the population of units or groups, which need to be included because of the effects of their different levels on the probability of infection or detection, will be modelled as category nodes. Examples are

- LOCATION when modelling a surveillance process for a vector-borne disease with patchy distribution of the vector;
- INDUSTRY SECTOR when modelling a surveillance process for a disease spread mainly to back yard poultry by wild birds;
- FARM TYPE when modelling an abattoir monitoring process in which only cull sows are sampled;
- ADMINISTRATIVE REGION or LABORATORY when modelling a diagnostic surveillance process handled by regionally based laboratories with different diagnostic capacities.

Detection is the identification of an infected unit as positive. A unit cannot be identified as positive if there is no chance it will be included in the surveillance process, due to systematic exclusion of its category of units. Thus a factor which results in systematic exclusion of a significant number of units from the process is a factor affecting the probability of detection. This includes all issues of the coverage of the surveillance process. One aim of analysing the model is to account for detection of disease in all sectors of the population in which disease might occur, and it is important that all such sectors for which the probability of detection is significantly different should be identified and explicitly included in the model as branches of the tree, with associated branch probabilities of detection. For example, the following factors affect the probability of detection, and should be included:

- INDUSTRY SECTOR (*beef* or *dairy*) when modelling a surveillance process for IBR which only tests milk samples
- ABATTOIR or perhaps REGION when modelling an abattoir sampling surveillance process which is only operated in certain abattoirs.

Both of these factors may also need including for possible effects on probability of *infection* as well as probability of *detection*. If not, they should be inserted in the tree among the detection nodes, below the unit infection node (see rule 4 above). If they do also affect probability of infection they should be placed among the infection nodes, probably above the herd grouping level. In addition, if they are considered to define different compartments or major groupings of the *SSC reference population*, within each

of which surveillance is required for adequate coverage of the population, they should probably be placed high up the tree as risk category nodes (see [Application in relation to zoning and compartmentalisation](#) under [Incorporating differential risk](#)).

Selection is unit-by-unit selection of units for inclusion in the surveillance process, for example in a sampling system such as

- sampling every fifth carcase on the line;
- sampling every 10th animal through the race;
- testing a random sample of animals

Factors affecting only the probability of selection need not be included, for example:

- DAY OF SLAUGHTER when modelling an abattoir sampling program and the sampler only works 3 days a week (unless, of course, INDUSTRY SECTOR SLAUGHTERED or REGION SLAUGHTERED or SPECIES SLAUGHTERED are also arranged by the day);
- SEX when modelling a serological surveillance process in which only females are bled. While this results in systematic exclusion of males, the two sexes are generally present in most groups (as defined by grouping factors), and there is no need to include SEX in the model unless it also influences probability of infection (e.g. mastitis) or detection (e.g. sampling only male poultry excludes all layer farms).

Estimation of branch probabilities

Every branch of every detection and infection node in the tree must have a branch probability assigned to it. Branch probabilities may, in theory, be qualitative, semiquantitative or quantitative in nature, and methods have been described for calculation of wholly qualitative or semiquantitative scenario trees in the related field of risk analysis. Such methods are perfectly valid, and we hope that in time similar methods will be developed for use in the context of scenario tree modelling for evaluation of surveillance for disease freedom. It is important that the modelling methodology should be generally applicable if it is to be of use in the arena of international trade, and in data-poor environments qualitative or semiquantitative methods may be the sensible approach.

The analysis method described here is quantitative, however, and requires all branch probabilities to be expressed quantitatively. We also recommend stochastic modelling rather than deterministic, and suggest that each branch probability estimate should incorporate the uncertainty associated with the estimate, in the form of an appropriate probability distribution.

Using available data

Whether quantitatively or qualitatively expressed, it is always important that branch probabilities should be based as far as possible on data. The required conditional probabilities are often very specific in nature, and require detailed analysis of available data sources. Examples of data which may be used for estimation of branch probabilities are given in [Identifying potential data sources](#).

Since the tree models surveillance for an exotic disease, prevalence data and diagnostic data are generally not available in the country conducting the surveillance. In these circumstances estimates of probabilities for detection node branches may need to be made from records involving clinically or epidemiologically similar diseases. Estimates of relative risks for branches of risk nodes will probably have to be derived from data from other countries, or from records of past outbreaks, together with climatic and geographical knowledge of the country and epidemiological knowledge of the disease.

Expert opinion

Where adequate data are not available for estimation of branch probabilities, and this is almost certain to be the case for some probabilities in all models, the generally accepted approach is to estimate the probabilities from expert opinion. Various authors have described procedures for harnessing expert opinion for use in the related field of risk analysis. Relevant guidelines include

- Gather opinion from a panel of experts. The panel should include relevant scientists and representatives of stakeholders as appropriate
- Capture the uncertainty of the experts as well as their best estimates
- Combine disparate opinions into a single probability distribution (e.g. [Vose 2000](#); [Stärk *et al*. 2000](#))

Design Prevalence (P*)

Detection of disease at a high prevalence is much easier than detection at low prevalence, so it is necessary to specify the prevalence(s) at which we will determine the sensitivity of our surveillance system component.

This prevalence, set for the purpose of drawing conclusions about the effectiveness of the *SSC* against an agreed standard, is termed the *design prevalence*. Design prevalence forms part of the design of the model, and is not related to any actual prevalence of disease in the population under study (which, since we are considering issues of freedom from disease, is most likely to be zero). It is not subject to uncertainty or variability, and need not be described by a distribution.

A number of different terms have been used to describe the concept of design prevalence, including minimum detectable prevalence, maximum acceptable or permissible prevalence, and minimum expected prevalence. These describe different aspects of the

design prevalence, and relate to ways in which a value can be selected. The term *design prevalence* is consistent with that used by Cannon (2002).

When multiple surveillance systems (or other sources of evidence for freedom from disease) are compared, it is important that the comparison be based on the same design prevalence assumptions for all systems.

Typically, design prevalence refers to the animal-level prevalence of disease (or, equivalently, the proportion of animals in the population that are diseased, or the probability that an animal selected at random from the population will be diseased). This tends to imply that the probability of infection is relatively constant across the entire population, which is usually untrue due to the clustering of disease. Most diseases cluster in groups, which means that the probability that an animal is diseased is much higher if other animals in the same group are infected, compared to a group where no other animals are infected. In order to capture the concept of clustering at the herd level, two levels of design prevalence may be used, one at the animal level and one at the herd level. The design prevalence at the herd level (P^*_H) is the proportion of herds in the country that are infected. Where P^*_H is specified, the design prevalence at the animal level (P^*_U) is the proportion of animals (or units) that are infected *within infected herds*.

The concept of multiple levels of design prevalence may be extended to account for further levels of disease clustering, such as the design prevalence of infected chicken houses on an infected farm (where the population is grouped into 1) farms, and 2) houses. Despite this, the most common approach is to consider clustering only at one level, and therefore include only two design prevalence levels in the tree definition.

It is worth noting that this approach is consistent with statistical approaches to the analysis of survey data to demonstrate freedom from disease (Cameron & Baldock 1998a).

Determining design prevalence(s)

The design prevalence is set by the person undertaking the analysis of the surveillance system, and guidance is therefore required as to the appropriate figures to use. Approaches that should be used, in order of preference, are:

International Standards
>Where international standards exist, these should be used. It may be reasonable to use a less rigorous standard (*ie* higher design prevalences) than those set by international standards when a trading partner is willing to accept such a level. It is not reasonable to require a standard more stringent than those set by international standards. The OIE has established standard design prevalence levels for surveillance for rinderpest, contagious bovine pleuropneumonia, bovine spongiform encephalopathy (less clearly), enzootic bovine leukosis, and others.

Trading partner requirements

Where international standards do not exist, and the purpose of demonstrating freedom from disease is for international trade, the requirements of trading partners should be taken into consideration. These may be subject to negotiation, based on the following points.

Biological plausibility and production system parameters

If no existing guidelines, standards or requirements exist, you (the analyst) must establish suitable values. The best guide for this is an understanding of the biology of the disease, and the structure of the production system.

- The design prevalence at the animal level (P^*_U) may be thought of as the minimum prevalence that would occur if the disease were present in a herd. If the disease is highly contagious and rapidly spreading, and the test being used is serological examination for antibodies, it may be expected that a large proportion of an initially naïve population would seroconvert to the disease, if it were present. This assumes that all animals in the group (*eg* herd) are in reasonably close contact with each other (hence the need for a consideration of clustering). For a disease such as FMD, HPAI or CSF, it is reasonable to expect that about 80% of survivors within a herd would seroconvert. While this may be the expected prevalence, it is possible that, under certain circumstances, the prevalence may be significantly lower than this. In order to be conservative, a lower figure should be chosen, such as 20%. This reflects the fact that the disease spreads rapidly, but is lower than any actual prevalence that we would normally expect.
- For diseases that are not rapidly spreading, the prevalence may be much lower. For instance, it is conceivable that for a disease like bovine paratuberculosis or bovine spongiform encephalopathy a single infected animal could persist in a herd for a significant time. The minimum expected prevalence is therefore extremely low, in which case the next point (Resources and political considerations) must be considered.
- When establishing P^*_H, the same problem is encountered. It is conceivable, especially in industries with good farm level biosecurity, that a single farm may be infected.

Resources and political considerations

When consideration of the biological behaviour of the disease indicates that it may be present at very low prevalence levels, it is often not feasible to detect the disease at these low levels. The development of a survey or surveillance process to detect extremely low levels of disease may be too costly and require more resources than are available. It is therefore often necessary to decide on design prevalence levels that are based more on what is feasible and can reasonably be demanded, than solely on biological grounds. This is in fact the way that international and trading partner standards are generally set. Political considerations dictate that a balance be established between the interests of the country demonstrating freedom (desiring that the design prevalence be as high as possible in order to maximise the reported level of confidence in the surveillance process) and a trading partner (aiming for as low a design prevalence as possible

to give the strongest evidence that disease is not present). The most common result of this compromise is to select a design herd prevalence of 1%, although values of 0.1%, 0.5% and 5% are also used.

Small herds

When a low P^*_U is specified, this can have little meaning in a small herd – in a herd of 10 animals, how many are infected with $P^*_U = 0.05$? The problem lies in the logic of the tree, which requires that this herd is definitely infected. In these circumstances (i.e. when the units processed are known to have come from a small herd) it is appropriate to assume that a minimum of one animal is infected in an infected herd. In the formulae presented below for calculation of herd-level sensitivity, it will then be appropriate to use hypergeometric probability formulae (e.g. Cameron & Baldock, 1998a and b) rather than the binomial probabilities presented.

Incorporating differential risk

Some subgroups of the population may be considered to be at higher risk of being infected (at the design prevalence) than others. This forms the basis of targeted surveillance, and the advantages of targeting high-risk groups will be lost if such differential risk is not incorporated in the analysis, since P^* is fixed across the population. For example, surveillance to demonstrate freedom from infection for a disease with an insect vector (*eg* bluetongue) may cover all areas of the free zone. However, risk of disease is much greater in areas closer to a known infected zone. Demonstration that this border area is free from disease may be considered more convincing than testing done in areas which are a great distance from any infected area. The way in which these differences in disease risk, and consequently differences in the value of information derived from different subpopulations, are captured, is through the use of risk nodes.

Risk nodes are category nodes representing risk factors for infection, and may apply at the unit level (e.g. AGE in a *SSC* detecting presence of antibodies) or at the group level (e.g. HERD TYPE (dairy or beef) in a *SSC* for bovine Johne's disease, where intensive dairy herds are more likely to be infected than extensive beef herds).

Differential risks are applied to the branches of a risk node such that the average risk for the *SSC reference population* is 1 (see below). This ensures that the calculated sensitivity of the *SSC* is not altered by the differential risk values used, when there is representative sampling of the *SSC reference population*; any effect on *SSC* sensitivity associated with differential risks is due to biased sampling (i.e. differences between proportions of the units actually processed in the *SSC*, to which the differential risks are applied, and proportions of the *SSC reference population*), which may well be deliberate, as in targeted surveillance processes. When a disproportionately high number of high-risk units has been processed, differential risks will raise the calculated sensitivity of the *SSC* by applying a higher effective probability of a positive outcome (Pr(unit positive) =

Pr(unit infected) x Pr(detected|infected)) to each high-risk unit than to each low-risk unit processed. The *effective probability of being infected* (*EPI*) for any unit or group is the product of the *design prevalence* and any applicable *differential risk*(s). So for the tree in Figure 10, the *effective probability* of a herd in HERD RISK GROUP 1 in COMPARTMENT 4 being infected is $R_{C4} \times R_{HRG1} \times P_H^*$, and the *effective probability* of a unit in UNIT RISK GROUP 2 being infected within an infected group (herd) is $R_{URG2} \times P_U^*$.

Differential risks are derived from relative risks estimated for each branch of the node. Relative risks for each branch may be estimated from data (historical observations or observations from countries / zones / compartments where disease is present) or from expert opinion, based on the epidemiology of the disease and the characteristics of the environment and population involved. Relative risks are specified relative to the lowest risk branch, and are then adjusted to ensure that the (weighted) average risk for the section of the *SSC reference population* represented by the risk node is 1.

$$\sum_{l=1}^{L}(AR_l \times PrP_l)=1 \tag{1}$$

where the node has *L* branches, and *PrP$_l$* is the proportion of units or groups (as appropriate) in the *SSC reference population* that falls into the *l*[th] branch[8].

It is important to note that the proportions used to adjust relative risks (*PrP$_l$*) are proportions of the *SSC reference population*. The proportions applied to risk node (and all category node) branches for calculation of actual system sensitivity are proportions of groups or units actually processed in the *SSC* (*PrSSC$_l$*).

When the units processed come from a known group (*e.g.* animals from a known herd) and there is, as in Figure 10, a UNIT RISK GROUP node (*e.g.* AGE (Adult / Juvenile)), then the population proportions needed to adjust the relative risk for each branch of the risk node are the proportions of Adults and Juveniles in that herd; these are herd-specific proportions. Often, these data will not be available, and it will be necessary to estimate an average proportion of Adults / Juveniles for all herds in this limb of the tree (i.e. this COMPARTMENT and this HERD RISK GROUP in Figure 10).

Differential risks must further be constrained such that, for the highest risk branch, $R_l P^*$ is no greater than 1. In practice, *R* is multiplied by P^* to give the effective probability that the unit or group is infected, given that the population is infected at P^*. Clearly, this probability should not be allowed to exceed 1. If this constraint results in downwards adjustment of *R* values (such that average population risk is then <1), the end result will be underestimation of actual system sensitivity adjusted for targeting. Such scenarios are rare, given the generally low values used for P^*.

Multiple risk nodes

Where there are multiple risk nodes preceding an infection node (i.e. there is more than one factor affecting the effective probability of infection for a unit or group) it is conceptually possible to calculate the EPI in two ways. For a group infection node with two preceding risk nodes (nodes A & B) with branches 1, 2, 3 and 1, 2, there are 6 possible risk configurations for any given group, namely A1B1; A1B2; A2B1; A2B2; A3B1; A3B2. Possible ways of calculating values for EPIH for a A2B1 group are:

1. calculate adjusted risks for each risk node separately, based on applicable reference population proportions (*PrP*s), and then

    $EPIH\_A2B1 = ARH\_A2 \times ARH\_A2\_B1 \times P^*_H$

    In this case, the *PrP* specified for each branch of node B (the second, lower, node of the two) is conditional on the branch of node A in which it is situated. So *PrPH_B1* (the proportion of groups in branch 1 of node B in the reference population) varies with the branch of node A, and is therefore designated *PrPH_A2_B1* (etc.) for branch 2 (etc.) of node A. This notation is also used for the relative (*RR*) and adjusted (*AR*) risk variable names.

2. define 6 risk configurations (limbs of the tree) as above (A1B1 etc.), then
    o calculate reference population proportions for each limb by multiplying together the *PrP*s for the relevant branch of each node on the limb (eg *PrPH_A2* × *PrPH_A2_B1*);
    o calculate among-limb relative risks (*RR*) by multiplying together *RR*s for each limb's contributing branches of each of the 2 risk nodes (*RRH_A2B1* = *RRH_A2* × *RRH_A2_B1* in this case);
    o calculate a single adjusted risk for each configuration using the calculated *PrP*s and *RR*s for each limb;
    o then

    $EPIH\_A2B1 = ARH\_A2B1 \times P^*_H$

Only the first of these two possible methods is correct, and method (2) above should **not** be used. This is because of the conditionality of relative risks and population proportions on the branches of the preceding nodes in which they are situated. Multiplying them all out before calculating the adjusted risk (as in method (2) above)

leads to a loss of the appropriate relativity of proportions and risks. In summary, always specify population proportions and relative risks conditional on all preceding nodes, and calculate adjusted risks separately for each risk node, before calculating the effective probability of infection as in the equation given in (1) above.

Building of scenario trees

Once the types of nodes, factors for inclusion and order of nodes have been established, the scenario tree can be built. The tree starts with a <u>root node</u>, and continues through branches down to the terminal nodes of the tree. These represent the final outcome, i.e. recognition of a unit as positive or negative.

Root node

The root of the tree may be thought of as COUNTRY (or other population) INFECTED. The purpose of the model is to estimate confidence in the surveillance process's ability to detect disease when it is present in the country (at the design prevalence), and to do this all factors influencing a unit's probability of giving a positive outcome, from the country level down to the unit level, must be included. The highest level grouping factor (e.g. REGION or INDUSTRY SECTOR) then forms the first or root node (in these examples a category node).

If the analyst wishes to estimate a posterior probability that the country is free of disease, given the negative surveillance results and a prior estimate of the probability of freedom, then the root node will be an infection node, COUNTRY STATUS. The rest of the tree is identical, and grows from the *infected* branch of the COUNTRY STATUS node, while the *uninfected* branch represents a negative process outcome, and is immediately followed by a terminal node. The tree is then calculated back to the *infected* branch of the root node to estimate the sensitivity of the process, and finally the extra step is taken of making a Bayesian estimate of the probability that the country is free of disease.

Completion and truncation

For the sake of completeness, and for calculation of a *<u>sensitivity ratio</u>* (*SR*), trees modelling surveillance processes with incomplete coverage of the *SSC reference population* must also be calculated for the scenario of complete, representative coverage, and therefore it is not possible to ignore branches of the tree which are not represented in the units processed in the *SSC*; all category branches must be followed to the bitter end. Naturally, this is also true for trees modelling systems with complete (and/or comprehensive) coverage of all units in the population.

However, one branch of each infection node (and many detection nodes) represents a negative outcome of the process, and so should be terminated at this point. In <u>Figure 11</u>,

for example, every node except the REGION category node has only one branch which needs to be followed further. The tree must be followed through for each REGION branch, but all other nodes have one of their two branches meeting a terminal node immediately.
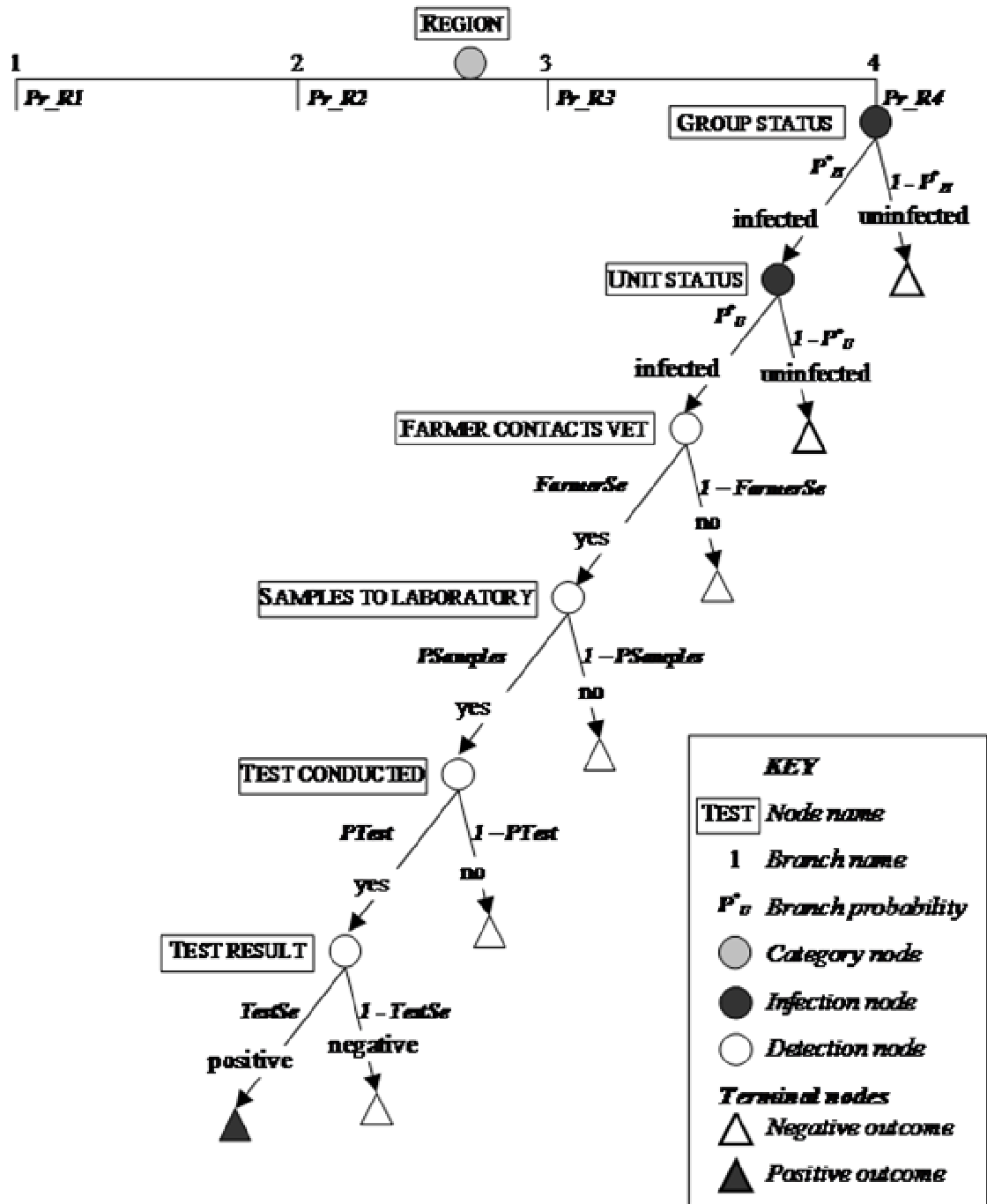
Figure 11 Scenario tree for a typical on-farm clinical diagnostic system (only 1 of four main branches completed; assume others identical in structure)

Analysing the tree assuming independence

Analysis of the tree is the calculation of the sensitivity of the *SSC*; the probability that the *SSC* would detect infection if it were present at the design prevalence(s). In situations where

- all units (generally animals) can be considered independent of each other with regard to probability of being infected, and
- the units processed are representative of the population, i.e. the proportions of units processed (*PrSSC*) falling into each of the categories specified in the tree are the same as their proportions in the SSC reference population (*PrP*),

then the process is simple. First, calculate the probability that any randomly selected unit in the population will give a positive outcome (*CSeU*, the *SSC (component) unit sensitivity*). This is done be calculating the overall limb probability for each limb of the tree (i.e. for each outcome / terminal node), and summing these limb probabilities for all limbs with *positive* outcomes. The limb probability for any limb's outcome is simply the product of all branch probabilities (for infection and detection nodes), branch proportions for all category nodes, and branch differential risks for risk category nodes, along the limb. For the tree in Figure 11 (with regions $k = 1 \ldots 4$) this will give

$$CSeU = \sum_{k=1}^{4} Pr\_R_k \times P_H^* \times P_U^* \times FarmerSe_k \times PSamples \times PTest_k \times TestSe \quad (2)$$

if *FarmerSe* and *PTest* vary among regions and *PSamples* and *TestSe* do not. There is only one limb with a positive outcome for each region.

The calculation for the tree in Figure 10, which includes risk factor nodes at 2 levels, as well as detection factor nodes, is

$$(3)$$

where UDC_i denotes levels of the UNIT DETECTION CATEGORY node; URG_j denotes levels of the UNIT RISK GROUP node; HRG_k denotes levels of the HERD RISK GROUP node; and C_l denotes levels of the COMPARTMENT risk node.

Suppose our surveillance process tests *n* units, all with negative results. Since these units are independent of each other, each has the same value, or contribution towards our confidence in the process, and the overall component sensitivity (*CSe*) is the probability that one or more positive units will be detected, given that the country is infected.

CSe = Pr(≥ 1 positive unit | country infected), or Pr(S+ | D+)

Now Pr($\geq$ 1 positive unit | D+) = 1 – Pr(all units negative | D+),

and Pr(all units negative | D+) = (Pr(1 unit negative | D+))^n

= (1 – Pr(1 unit positive | D+))^n

So $CSe = 1 - (1 - CSeU)^n$.

Advanced approach - accounting for lack of independence among units

The assumption of independence is generally unreasonable, although circumstances in which it is appropriate include

- genuinely representative sampling of units in which disease is evenly spread throughout the population of units;
- a surveillance process with comprehensive coverage of all units in the population, with negligible clustering of disease.

In general however, disease occurs in clusters or groups of susceptible animals. If a herd is infected, the probability that an individual animal is infected is P^*_U, while the probability that an animal in the uninfected herd next-door is infected is zero. This clustering effect is dealt with in analysis of our scenario tree by specifying separate values for P^*_H (herd-level design prevalence) and P^*_U (animal-level design prevalence given that the herd is infected).

We test for the presence of infection in a herd by testing a suitable number of animals, and if they are all negative, we conclude with a certain level of confidence that the herd is not infected at a prevalence of P^*_U. If we now test another animal from the same herd, the additional information we get about the absence of infection in the herd from another negative result is minimal. If we instead test an animal in the herd next-door, we will get considerably more information about the probability that this new herd is infected when we get our first negative result; each additional negative sample we take from a herd gives us incrementally less information about the likelihood that the herd is infected. We can revise our estimate of the effective probability that the herd is infected (i.e. differential risk(s) multiplied by herd-level design prevalence) at the animal-level design prevalence after every negative sample we process, using Bayesian revision:

(4)

where we have perfect specificity; $PH_{h,i}$ is our estimate of the effective probability that the hth herd is infected after the ith negative unit; P^*_U is the unit (within herd) design prevalence; SeU is the sensitivity of detection for each unit; and assuming that the number of diseased animals in an infected herd is binomially distributed (i.e. units within

herds are independent of each other with respect to disease state). Where there are unit-level (i.e. within herd) risk nodes, P^*_U will be replaced in this formula by the relevant *EPIU*.

Diseased animals will often be found clustered in infected herds, and diseased herds clustered in an infected locality. This clustering effect must be dealt with in analysis of our scenario tree (see <u>Stepwise calculation of system sensitivity</u>).

Calculation of unit sensitivity (SeU)

The tree comprises, in its lower branches, steps relating to the detection, or lack thereof, of an infected unit. For example, given that an animal is diseased, there may be several steps in the detection process (farmer notices, farmer calls vet, vet takes samples, samples tested for disease, test gives positive result). Whatever the number and nature of these steps, they may be combined to give the probability that the infected unit will not be detected, and likewise the probability that it will be detected. This latter probability is the unit sensitivity (*SeU*) of the surveillance process.

In the simplest case in which there are no factors in the model which affect the probability of detection (for example as in <u>Figure 11</u>), *SeU* is the same for all infected units. All detection nodes should have been placed below the unit infection node (<u>Figure 11</u>), and the same limb structure for detection nodes is to be found below each unit infection node in the tree. *SeU* is then calculated for a single unit infection node by multiplying together detection node branch probabilities for each limb with a positive outcome arising from the unit infection node, then summing these products across all positive outcomes arising from the unit infection node. The tree in <u>Figure 11</u> gives

$$SeU = FarmerSe \text{ x } PSamples \text{ x } PTest \text{ x } TestSe \qquad (5)$$

When there are factors affecting probability of detection, these will be represented by category nodes, which may have been placed below the unit infection node, or, if the factor also affects the probability of infection, above the unit infection node. In these circumstances, any value calculated for *SeU* will either be unit-specific or an average figure for all units or some subset of units. This will depend on the data available on units processed. If factor (category) levels are available for each unit processed, then unit-specific *SeU* can be calculated. Where such data are not available, unit sensitivity of detection must be averaged across levels of factors for which the data are missing. *SeU* is thus calculated as a weighted average of subgroup sensitivities, the weights being subgroup category proportions (of units processed); so in this case limb probabilities are the product of all constituent branch probabilities and category proportions. As an example, suppose that in the *SSC* represented by the tree in <u>Figure 10</u>, data for some units do not include their UNIT DETECTION CATEGORY (denoted by UDC_i; this might be, for example, the age of the animal). For these units

$$SeU = PrSSC\_UDC_1 \text{ x } TestSe\_UDC_1 + PrSSC\_UDC_2 \text{ x } TestSe\_UDC_2 \qquad (6)$$

where each of the values for these variables may vary among levels of other factors in the tree, and will thus need calculating separately for units in different limbs of the tree.

Stepwise calculation of system sensitivity

Units are generally grouped by some grouping factor which affects the probability that an individual will be infected. In general this is a matter of animals being grouped in herds. Within an infected herd, individual animals are considered to have a certain constant probability of being infected (the within-herd or unit prevalence, P^*_U). The herd is either infected or it isn't; if it is, then the probability that any animal will be infected is P^*_U. If this is not true – i.e. if there is another level of grouping, such as the management group within the herd – then a risk category node must be included for this level too. The tree structure is flexible, but the essential principle is that at any node all units (or groups, depending on the grouping level) going down a given branch have equal probability of being infected or detected, depending on the node type.

Calculation of group (herd) level sensitivity

At the level of the first (lowest level) grouping node (for example, herd), results from multiple units are aggregated to give a group-level sensitivity (*SeH*) of detection. There are three different methods for calculating *SeH*, depending on the proportion of the population sampled:

- where the population (group) is large and sample size is small (<10%) compared to group size a <u>Binomial approach</u> should be used.
- where the population (group) is small and sample size is large (>10%) compared to group size a <u>Hypergeometric approach</u> should be used.
- where the entire population (group) is included in the surveillance an <u>Exact method</u> should be used.

Binomial approach

The Binomial method is simplest and should be used where sample size is small (<10%) relative to population (group) size. The general formula for *SeH* in the hth group is:

$$SeH_h = 1 - \prod_{j=1}^{J} (1 - AR\_URG_j \times P^*_U \times SeU)^{n_j}$$

(7)

where:

- there are J branches to the UNIT RISK GROUP node,
- n_j units are processed in the jth UNIT RISK GROUP,
- AR_URG_j is the adjusted risk for the jth UNIT RISK GROUP,
- P^*_U is the unit-level design prevalence and
- SeU is the overall unit level sensitivity of the diagnostic process (i.e. the combined sensitivity of all detection nodes).

(Throughout this section refer to Figure 10 for node and variable names.)

Alternatively, if there are no risk nodes associated with the unit infection node this collapses to

$$SeH_h = 1 - \left(1 - P^*_U \times SeU\right)^{n_h} \tag{8}$$

where we have negative outcomes from n_h units in the group.

This is the probability that one or more of the units processed from group h will have a positive outcome, given that the group is infected; one minus the probability that all units will give negative results. This is the standard formula for the probability of obtaining one or more successes in a binomial process. The probability that any one unit will give a negative result is (1 – Pr(positive result)). Pr(positive result) for each unit is the probability that it is infected (P^*_U in equation 8) multiplied by the probability that it will be detected if it is infected (SeU). So Pr(all n_h tested units give negative results) is (1 - P^*_U x SeU) raised to the power of the number tested in herd h (n_h).

Hypergeometric approach

The Hypergeometric method should be used where sample size is large (>10%) relative to population (group) size. This method is based on the binomial approximation to the hypergeometric distribution (Cameron and Baldock, 1998a). For this method, the general formula for SeH in the hth group is:

$$SeH_h = 1 - \left(1 - SeUAv_h \times \frac{n_h}{N_h}\right)^{P^*_U \times N_h} \tag{9}$$

where:

- $SeUAv_h$ is the average unit sensitivity for herd h,
- n_h is the number of animals sampled from herd h,
- N_h is the total number of animals in herd h and
- P^*_U is the unit-level design prevalence.

Exact method

The Exact method should be used where the entire population (group) is part of the surveillance system (i.e. comprehensive coverage at the group level). This method calculates the exact probability of one or more positives being detected, given the estimated number infected in the group and the probability of detection. For this method, the general formula for SeH in the hth group, assuming there are no risk nodes associated with the unit infection node is:

$$SeH_h = 1 - (1 - SeU_{Av})^{d_h} \qquad (10)$$

where:

- $d_h = P^*_U \times n_h$ is the estimated number of infected individuals in the group of size n_h,
- P^*_U is the unit-level design prevalence and
- SeU_{Av} is the average sensitivity of the detection process at the unit level (combined sensitivity across all detection nodes averaged across all units and risk groups).

Calculating sensitivity at higher grouping levels

Moving up to the next grouping level (if there is one; we will follow the example shown in Figure 10), we again need to aggregate the herd sensitivities of detection to give a sensitivity of detection at the HERD RISK GROUP level. As at the herd-level, we should again use Binomial, Hypergeometric or Exact methods, depending on the proportion of herds in the group that have been sampled.

Binomial approach

As at the primary grouping level, the Binomial method should be used where sample size is small (<10%) relative to group size. Using the Binomial approach in a similar way as for the herd level, Se_HRG for the kth HERD RISK GROUP, Se_HRG_k, is

$$Se\_HRG_k = 1 - \prod_{h=1}^{N_k} (1 - AR\_HRG_k \times P^*_H \times SeH_h) \qquad (11)$$

where

- there are N_k is the number of herds in the kth HERD RISK GROUP,
- AR_HRG_k is the adjusted risk for the kth HERD RISK GROUP,
- P^*_H is the herd/cluster-level design prevalence and
- SeH_h is the herd/cluster-level sensitivity for the hth HERD RISK GROUP.

Hypergeometric approach

As at the group or herd level, the Hypergeometric approach should be used where a large (>10%) proportion of herds are included in the surveillance component. For this method, the general formula for Se_HRG in the kth group is:

$$\tag{12}$$

where:

- $SeHAv_j$ is the average herd-sensitivity for the jth risk group of J risk groups in the kth HERD RISK GROUP,
- n_j is the number of herds sampled from the jth risk HERD RISK GROUP,
- N_j is the number of herds in the jth risk HERD RISK GROUP,
- AR_j is the adjusted risk value for the jth risk HERD RISK GROUP,
- P^*_H is the herd-level design prevalence,
- N is the total number of herds in the population, and
- PrP_j is the proportion of the population in the jth risk HERD RISK GROUP.

Equation 12 allows for multiple risk groups within the group for which the sensitivity is being calculated. If there are no risk groups within the main grouping, the formula simplifies to:

$$Se\_HRG_k = 1 - (1 - SeHAv\tfrac{n_k}{N_k})^{AR\_HRG_k \times P^*_H \times N \times PrP_k} \tag{13}$$

where:

- $SeHAv$ is the average herd-sensitivity for herds in the kth HERD RISK GROUP,
- n_k is the number of herds sampled from the kth HERD RISK GROUP,
- N_k is the number of herds in the kth HERD RISK GROUP,
- AR_k is the adjusted risk value for the kth HERD RISK GROUP,
- P^*_H is the herd-level design prevalence,
- N is the total number of herds in the population, and
- PrP_k is the proportion of the population in the kth HERD RISK GROUP.

Exact method

As at the primary grouping level, the Exact method should be used where the entire population (group) is part of the surveillance system (i.e. comprehensive coverage at the group level). This method calculates the exact probability of one or more positives being detected, given the estimated number of infected clusters in the group and the probability of detection. For this method, the general formula for Se_HRG in the kth group is:

$$Se\_HRG_k = 1 - (1 - SeH_k)^{d_k} \tag{14}$$

where:

- $d_k = AR\_HRG_k \times P^*_H \times n_k$ (rounded up to the next whole number) is the number of infected clusters in the kth group,
- SeH_k is the average sensitivity for clusters in the kth HERD RISK GROUP and
- P^*_H is the cluster-level design prevalence.

Approaches to calculating component sensitivity

As for herd or group level calculations, component sensitivity (*CSe*) can be calculated in a number of ways. If there are several grouping levels, the methods described in the section Calculating sensitivity at higher grouping levels can be used. However, where there is only a single grouping level (for example herd or farm level), *CSe* can be calculated directly from *SeH* values using Binomial, Hypergeometric or Exact methods, depending on coverage and sampling proportions:

Binomial

The Binomial approach should be used where a small (<10%) proportion of herds are included in the surveillance component, with *CSe* calculated as:

$$CSe = 1 - \prod_{i=1}^{I} (1 - EPI\_H_i * SeH_i) \tag{15}$$

where:

- EPI_H_i is the effective probability of infection for the ith herd from a population of I herds,
- SeH_i is the herd (or cluster) -level sensitivity for the ith herd

Hypergeometric

The Hypergeometric approach should be used where a large (>10%) proportion of herds are included in the surveillance component, with *CSe* calculated as:

$$CSe = 1 - \prod_{j=1}^{J} \left(1 - SeHAv_j \frac{n_j}{N_j}\right)^{AR_j \times P^*_H \times N \times PrP_j} \tag{16}$$

where:

- $SeHAv_j$ is the average herd-sensitivity for the $_j$th risk group of $_J$ risk groups in the population,
- n_j is the number of herds sampled from the $_j$th risk group,
- N_j is the number of herds in the $_j$th risk group,
- AR_j is the adjusted risk value for the $_j$th risk group,
- P^*_H is the herd-level design prevalence,
- N is the total number of herds in the population, and
- PrP_j is the proportion of the population in the $_j$th risk group.

Exact

The Exact method should be used where all herds are part of the surveillance component (comprehensive coverage), with *CSe* calculated as:

$$CSe = 1 - (1 - SeH_{Av})^D \qquad (17)$$

where:

- SeH_{Av} is the average herd-sensitivity for all herds in the population, and
- D is the estimated number of infected herds ($P^*_H \times N$)

Calculating Country-level (system) sensitivity

This needs to be re-done

This process may be continued up through any number of grouping levels, progressively aggregating the results and calculating group-level sensitivity estimates, accounting for any clustering within groups. At the top (country) level, the system sensitivity (*SSe*) is often given by an aggregation of compartment-level sensitivities thus:

$$SSe = 1 - \prod_{l=1}^{L} \left(1 - R_{C_l} \times Se_{C_l}\right) \qquad (18)$$

is this right? Shouldn't it be adjusted risk, not R_C?

where L is the number of compartments in the country, and Se_{C_l} is the compartment-level sensitivity for the lth compartment. *SSe* is then our estimate of sensitivity for one component of the surveillance system.

Category proportions

The observant reader will have noted that the branch proportions at category nodes were used in analysis under the assumption of independence, but seemed to be ignored in the clustering model.

Category node branch proportions are only used in analysis of the tree where data on factor levels are not available for individual units processed by the *SSC*. Where the factor levels are known for each category node in the tree, we can apply the appropriate probabilities of infection and detection to each unit in our calculation of system sensitivity. Where we do not have these data for each unit, we need to apply *average* probabilities of infection and detection to units and groups in the calculations. Average probabilities are calculated as weighted averages, using branch proportions of category nodes as weights.

For example, suppose Pr(detection) for an infected beef herd is *SeH_B* and Pr(detection) for an infected dairy herd is different (*SeH_D*). The proportion of dairy herds in our reference population is *PrP_D*, and *PrP_B* is (1 – *PrP_D*). Now if we know for each herd processed whether it is a dairy or beef herd, we can apply the relevant *SeH_D* or *SeH_B*. But if we do not know, we will use the average probability of detection for a herd selected randomly from the *SSC reference population*, which is

$$SeH = PrP\_D \times SeH\_D + PrP\_B \times SeH\_B \tag{19}$$

It is possible that, although we do not know the herd type for each herd processed, we do know the proportion of the processed herds that were dairy herds (*PrSSC_D*). In this case it would be preferable to use the proportions of units processed as weights for calculation of the average *SeH*, rather than *PrP_D* and *PrP_B*:

$$SeH = PrSSC\_D \times SeH\_D + PrSSC\_B \times SeH\_B \tag{20}$$

This is clear for actual *SSC* data. How to apply the same rules for calculation of the fully representative system is not immediately obvious, but follows the same principles. With the representative sampling system, we know the branch proportions for all category nodes from industry statistics or expert opinion. Armed with these figures we can calculate expected probabilities of infection or detection as above, effectively applying the category proportions as weights. Alternatively, we can simulate data using the category proportions, and generate a data set which is representative of the population, and in which factor levels for each unit are known. This can then be analysed by applying specific Pr(detection) and Pr(infection) figures to each unit, in just the same way as was done with the actual *SSC* data for which factor levels were all known for each unit.

Sensitivity ratio (*SR*)

This is the ratio of the calculated sensitivities of two *SSC*s, or two surveillance systems. It is used to evaluate the relative sensitivity of the *SSC*, independent of assumptions about the absolute values of design prevalences. It is calculated by comparing the sensitivity of the *SSC* being analysed (including differential risks among sub-populations) with a hypothetical version of the same *SSC* which uses fully representative sampling from the entire population (and the same differential risks among sub-populations). The latter represents the 'representative standard' approach to surveillance and is a useful standard for comparison.

The process for calculating the *SR* is as follows:

1. Calculate the sensitivity of the *SSC* using a stochastic scenario tree model, as described above, using actual data on units processed. The result is the actual system sensitivity (CSe_{Actual}).
2. Recalculate the sensitivity using a "data" set prepared with the same total number of units processed, but with these units distributed representatively over all population subgroups (*i.e.* proportions of units etc. falling into each population subgroup are the same in the "representative data" set as in the reference population). In this representative data set every unit in the population would have an equal chance of detection. This gives the hypothetical fully representative *SSC* sensitivity ($CSe_{Representative}$).

3. Then $$SR = \frac{CSE\_Actual}{CSe\_Standard}$$ (21)

The *SR* measures the performance of the actual *SSC* being modelled relative to the hypothetical representative system. A *SR* could also be calculated for the whole surveillance system (combination of multiple SSCs) if required. A *SR* of 1 indicates that the *SSC* being studied is equally effective at detecting disease as one using representative sampling. A *SR* greater than one indicates that the *SSC* is better than one using simple random sampling, due to the fact that it targets animals at greater risk of disease. Conversely, a *SR* of less than one indicates that the *SSC* is less effective than one using simple random sampling, due either to biased sampling of animals with lower risk of disease, or to a failure to sample units in one or more sub-populations (ie inadequate *coverage* of the *SSC reference population*).

As an output of the stochastic scenario tree model, *SR* will, in practice, be estimated as a frequency distribution.

A sensitivity ratio can also be calculated to compare the efficacies of two *SSC*s, assuming the same values have been used for inputs common to both scenario trees.

Combination of data from multiple sources

In order to gain the full benefit of the analysis of multiple sources of evidence for freedom from disease, it is necessary that the results of these analyses be able to be combined into a single estimate of the confidence in the combined surveillance systems.

Cannon (2002) described two techniques for the combination of levels of confidence from multiple sources of evidence. The first uses simple combination of probabilities, based on the following formula:

$$Se_{combined} = 1 - \prod_{j=1}^{J} (1 - Se_j) \tag{22}$$

where j denotes the component of the total surveillance system. The second is a qualitative point system, which is targeted at application in situations without access to quantitative analytical resources (e.g. herd classification schemes). The former approach is used in the case study presented here, but both suffer from the problem that they assume that the surveillance systems being analysed are independent. However, this is often not the case.

Accounting for lack of independence

This section proposes a method of combining the sensitivity of different SSCs to provide an overall measure of sensitivity that takes into account lack of independence among SSCs, and is based on the scenario-tree methodology already described.

Estimating system overlap

Consider an analysis of two SSCs, using two different scenario trees. Using the approach of step-wise calculation of sensitivities at different grouping levels, the sensitivity of, for example, each farm, and each county is calculated. Not all farms, nor all counties, may be represented in the first SSC. The second SSC follows a similar grouping structure but with a range of different nodes and probabilities defining that SSC. During stand-alone analysis of each scenario tree (as described above), the prior effective probability of a group being infected is the group-level design prevalence P^*_H multiplied by any applicable differential risk AR, and a posterior estimate of this probability (P_H) may be calculated (using Bayes theorem; equation 4), given the number of negative units processed from that group; P^*_U; any differential risks applied at the unit level; and SeU.

Once the analysis of the first SSC is complete, PH is available for each group at each grouping level. For independent analysis of the second scenario tree, we start with the same prior effective probabilities of infection as were used for the first (i.e. P^* and applicable differential risk values). However, to account for the information already gathered from the first SSC, group-level posterior probabilities of infection from the first SSC are used as priors in the second tree, replacing P^* values. We thus account for the fact that we already have some information about the status of certain farms and counties

(for example) from the first *SSC*, so information gathered in the second *SSC* does not contribute as much to our knowledge. (Refer to [Advanced approach – accounting for lack of independence among units](#) above.)

Some farms in the second *SSC* may not have been examined in the first SSC, and for such groups there is no prior information for analysis of the second *SSC*, and P^* values are used as priors. Where there is no overlap in the farms or other groups examined, there is clearly no prior information available from the first *SSC*, and the second *SSC* can be considered to be independent.

This procedure can be continued for multiple *SSC*s. Where a farm (for example) appears in a third *SSC*, and is also present in the second, its *PH* from the second is used as the prior for the third. Where it appears in the first and third, but not the second, then the posterior from the first is used as the prior for the third.

Once posterior estimates of Pr(Infected) have been inserted as priors into each successive model in a chain-like fashion, the sensitivity of each *SSC* can be calculated. Because the resulting sensitivities have already been adjusted to take into account their lack of independence, the results of analysis of each scenario tree can then be considered independent of each other, and combined using [equation 22](#).

Incorporating data from random surveys

In some instances, it may be desirable to combine data from targeted surveillance based on risk with data from a representative survey, to improve the overall sensitivity of the surveillance system (*SSe*) and the probability of freedom for the population. For example, a large volume of data may be available from export testing for a particular disease and is supplemented by a properly structured representative survey of the population. In this situation, analysis of the biased sampling requires a scenario-tree approach, whereas analysis of the representative survey data could be achieved much more simply using standard methods.

The simplest way to combine multiple data sources is therefore to calculate a component sensitivity for each data set and combine them assuming independence. For this approach, *CSe* for the targeted surveillance component would be calculated using a scenario tree approach to account for biasing of sampling and variations in risk and detection between sub-groups of the population, while *CSe* for the representative sample could be calculated using the standard methods that would normally be used for such a survey. The two component sensitivities are then combined to calculate system sensitivity (*SSe*) using [Equation 22](#).

This approach is appropriate when there is reasonable independence between the two samples, but will overestimate the combined *CSe* if there is substantial overlap between them (for example the same farms are represented in both groups). Where this is the case (significant overlap between samples) the preferred approach is to construct a second

scenario tree for the representative sampling and to calculate the additional sensitivity generated by the biased sampling after adjusting for overlap between the two samples (see previous section [Accounting for lack of independence](#)). The two *CSe* estimates are then combined, again using [Equation 22](#).

Calculation of the probability of country freedom

To this point, the purpose of the analysis has been to calculate the sensitivity of the surveillance system, or the probability that the system would be able to detect disease if it were present (at the design prevalence). This may be expressed as $\Pr(S+ \mid D+)$. As has been shown, the system sensitivity is a valuable tool in assessing the performance of the system. However, it is based on a hypothetical assumption – the probability of detecting disease *if it were present in the population.* This is a measure of the quality of the surveillance system, but it does not directly answer the question most trading partners are likely to be asking – is the country actually free from disease. Expressed in probability notation, this is $\Pr(D- \mid S-)$ or the probability that the country is free from disease, given that the surveillance system has failed to find disease.

This question is analogous to those posed in the interpretation of diagnostic tests. Given that a test *T* has produced a negative result, what is the probability that that the animal does not have the disease? This is **negative predictive value** (*NPV*) of the test, and is calculated using Bayes' theorem. Intuitively, in the diagnostic testing framework, the probability of an animal being truly negative if it tests negative is the proportion of all negative results (true and false) that are truly negative:

$$NPV = Pr(D- T-) = \frac{true\ negatives}{true\ negatives + false\ negatives}$$

In terms of sensitivity and specificity this can be expressed as:

$$NPV = \frac{Sp(1-P)}{Sp(1-P)+(1-Se)P}$$

In this case, the prevalence of disease in the population (*P*) is used as an estimate of the prior probability of any particular animal being diseased. Adjusting this prior probability with new evidence (the test result) allows a posterior probability to be calculated.

The same principles can be applied at the country level, using the same formula. Using the scenario tree, we have calculated the sensitivity of the *SSC*.

The negative predictive value, or the probability that the country is free from disease, given that the surveillance system did not detect disease, can be calculated using the same formula, by substituting system sensitivity for individual animal test sensitivity, 1 for specificity, and prior probability of disease being present in the country for prevalence:

$$Pr(D-|S-) = NPV = \frac{1-prior}{(1-prior)+prior(1-SSe)} = \frac{1-prior}{1-prior \times SSe} \qquad (23)$$

This is a simple approach to calculating the value of interest. As with the other calculations presented, it is shown here in a deterministic form, but can be implemented within the model to provide a stochastic version yielding a probability distribution.

Selection of a prior

A problem with this approach is the difficulty in selecting an appropriate value for the prior probability of the presence of disease in a country. The choice of prior value has a significant impact on the final estimate of the probability of country freedom, as shown in Table 8.

Table 8 Probability of country freedom from disease (at P*)calculated using Bayes' theorem, using a system specificity of 1, and varying system sensitivity (*SSe*) and prior probability of the country being infected (*Prior*)

Prior	*SSe*												
	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.850	0.900	0.950	0.990	0.999
0.990	0.011	0.012	0.014	0.017	0.020	0.025	0.033	0.048	0.063	0.092	0.168	0.503	0.910
0.950	0.055	0.062	0.070	0.081	0.095	0.116	0.149	0.208	0.265	0.345	0.513	0.840	0.981
0.900	0.110	0.122	0.137	0.156	0.182	0.217	0.270	0.357	0.426	0.526	0.690	0.917	0.991
0.800	0.217	0.238	0.263	0.294	0.333	0.385	0.455	0.556	0.625	0.714	0.833	0.962	0.996
0.700	0.323	0.349	0.380	0.417	0.462	0.517	0.582	0.681	0.741	0.811	0.896	0.977	0.998
0.600	0.426	0.455	0.488	0.526	0.571	0.625	0.690	0.769	0.816	0.870	0.930	0.985	0.999
0.500	0.526	0.556	0.588	0.625	0.667	0.714	0.769	0.833	0.870	0.909	0.952	0.990	0.999
0.400	0.625	0.652	0.682	0.714	0.750	0.789	0.833	0.882	0.909	0.937	0.968	0.993	0.999
0.300	0.722	0.745	0.769	0.795	0.824	0.854	0.886	0.921	0.940	0.959	0.979	0.996	1.000
0.200	0.816	0.833	0.851	0.870	0.889	0.909	0.930	0.952	0.964	0.976	0.988	0.998	1.000

0.100	0.909	0.918	0.928	0.937	0.947	0.957	0.968	0.978	0.984	0.989	0.994	0.999	1.000
0.050	0.955	0.960	0.964	0.969	0.974	0.979	0.984	0.990	0.992	0.995	0.997	0.999	1.000
0.010	0.991	0.992	0.993	0.994	0.995	0.996	0.997	0.998	0.998	0.999	0.999	1.000	1.000
0.001	0.999	0.999	0.999	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Extremely high system sensitivities provide a high probability of freedom regardless of the prior probability of infection, while low prior probabilities of disease provide a high probability of freedom regardless of the system sensitivity.

A number of approaches have been suggested for selection of the prior probability for this calculation:

1. The prior probability in this calculation is analogous to the prevalence (proportion of diseased animals in the population) in the *NPV* calculation. The same approach could be used to calculate the proportion of infected countries in the population of countries. The difficulty is identifying the appropriate population of countries. Using the whole world is probably not appropriate, as the risk of disease in different countries varies considerably, both for geographical and developmental reasons. Narrowing the group to 'comparable' countries such as a region is also problematic, as the definition of what is comparable will greatly influence the prior, and in most cases when demonstrating disease freedom, comparable countries will be considered as those that are free from disease, giving a prior of 0. This approach may therefore not be considered to be particularly valid or practical.
2. Estimation of a prior based on expert opinion. This is likely to be very difficult to get a reliable an internationally acceptable estimate (or distribution) using this approach, but it may be considered.
3. The use of a 'neutral' prior, such as 50% (point value). In essence, this assumes no useful prior information about the disease status. This is a very conservative approach, and naturally ignores any historical information on freedom that may be available.
4. The use of estimates of disease probability gained from previous analysis of surveillance. This approach allows incorporation of historical surveillance findings, and a link with import risk analysis, and is the approach we recommend (see next section).

Temporal discounting of historical surveillance data

When a country is free from a disease and is conducting ongoing surveillance for that disease, there is a continuous stream of negative surveillance data, perhaps supplemented by periodic additional negative surveillance information such as surveys. Each "survey", or temporally contained *SSC*, can be analysed as described above to give a *CSe* for the *SSC*. The continuous stream of data, however, must be divided into appropriately sized slices for analysis, the size of the slice being defined temporally by the length of the surveillance time period. We have already discussed how to determine the surveillance time period for analysis (TP), and suggested that in most cases either one month or one year will be appropriate. So our continuous stream of surveillance data in an ongoing *SSC* will be divided into sequential slices for analysis, each of length *TP*.

When assessing (subjectively) whether a population is free from a disease we do not consider only the last TP, for which we have calculated the *CSe*; we also look at the history. How long has this negative surveillance been going on? When was the disease last seen? Have there been any other historical surveillance activities to add support to the claim to freedom? In other words, we value past negative surveillance findings. When looking at the most recent TP's findings, we therefore make a mental assessment of what we believe to be the "background" status of the population, based on its history; onto which we will add the evidence of the most recent TP. In calculating the (posterior) probability that the population is free from the disease, we add the evidence from the current surveillance findings (the most recent TP) to our prior belief, and this prior belief is based on the posterior probability from the previous TP, if we assume that this recalculation is done at the end of every TP.

Another important component of this assessment is to ask the question: what biosecurity measures are in place to prevent the disease getting into the population?

Method

If we have an ongoing surveillance system which delivers negative results continuously over time, we divide the surveillance data into equal time periods of an appropriate length. At the end of each TP the surveillance data gathered during that TP are used to calculate a sensitivity (SSe_{tp}) for detection of the presence of the disease under consideration at the design prevalence P^*. The probability that the country is free from disease (at P^*) at the end of the TP ($PostPFree_{tp}$) can then be calculated as the negative predictive value of the surveillance system, PR(D-|S-):

$$PostPFree_{tp} = \frac{(1 - PriorPInf_{tp})SSp}{(1 - PriorPInf_{tp})SSp + PriorPInf_{tp}(1 - SSe_{tp})} \qquad (24)$$

where $PriorPInf_{tp}$ is an estimate of the probability that the country (or zone or compartment) is infected at P^* or greater, at the end of the TP, prior to application of the surveillance results (and $PriorPInf_{tp} = (1 - PriorPFree_{tp})$). Our surveillance system has perfect specificity, so this simplifies to

$$PostPFree_{tp} = \frac{1 - PriorPInf_{tp}}{1 - PriorPInf_{tp} \times SSe_{tp}} \qquad (25)$$

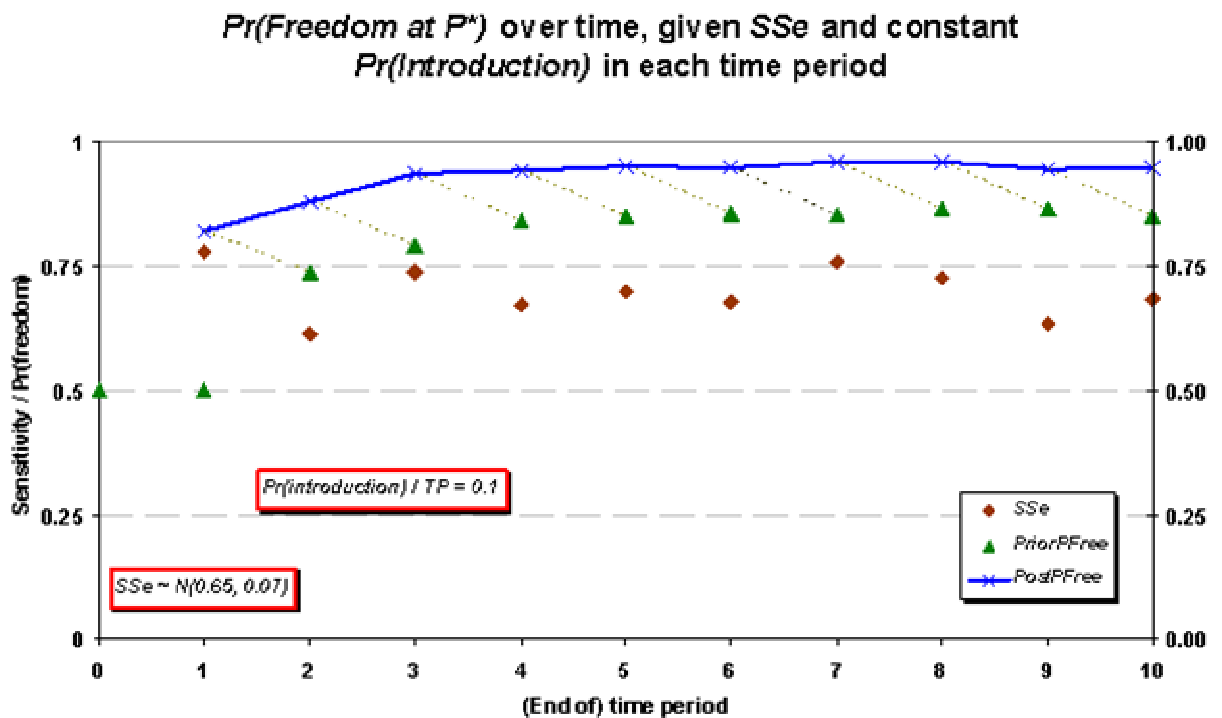which is the same result as we obtained in (24) above

Figure 12 Probability of freedom at P* over time

If we select a point in time which marks the beginning of our ongoing surveillance system (delivering negative results) and refer to it as the start of TP 1 ($tp = 1$), when we come to calculate $PostPFree_1$ we might choose to be very conservative in estimating $PriorPInf_1$ and say that it is 0.5, as in Figure 12. At the end of the TP we calculate SSe_1, and use it to estimate $PostPFree_1$. Using equation 25 $PostPFree_1$ will have a value greater than or equal to (1 - $PriorPInf_1$), since in the worst case scenario (*ie* absence of surveillance; $SSe_1 = 0$) our estimate of the probability of freedom will be unchanged. Any (negative) surveillance evidence at all can only increase our estimate of $PostPFree_1$.

At the end of TP 2 we again calculate SSe_2, and we now need a value for $PriorPInf_2$ in order to apply equation 25 again. It is reasonable to suppose that our estimate of $PriorPInf_2$ should be less than $PriorPInf_1$, since we concluded after TP 1 that the probability the country was infected was (1 - $PostPFree_1$), which was less than $PriorPInf_1$, as a result of our surveillance during TP 1. However, in estimating $PriorPInf_2$ we should take account of the fact that disease could have entered the country during TP 2. This is the only possible reason for the probability that the country is infected (at P*) to

have increased since we calculated $PostPFree_1$[9], and forms the basis of the adjustment that must be made to $PostPFree_1$ in estimating $PriorPInf_2$.

[9] It is also possible that the probability that the country is infected has decreased, due either to spontaneous disappearance of infection from the population (including inadvertent culling), or to an eradication program. Since the country is believed to be free, an active eradication program is highly unlikely, and for most diseases of significance the probability of spontaneous disappearance is probably negligible.

Probability of disease introduction – interface with IRA

The probability that infection has been introduced into the population during TP 2 is, in import risk analysis (IRA) terminology, the probability of release and exposure during the TP; here we will refer to $PIntro_{tp}$. Figure 13 shows three arbitrary introductions of infection (A, B & C).
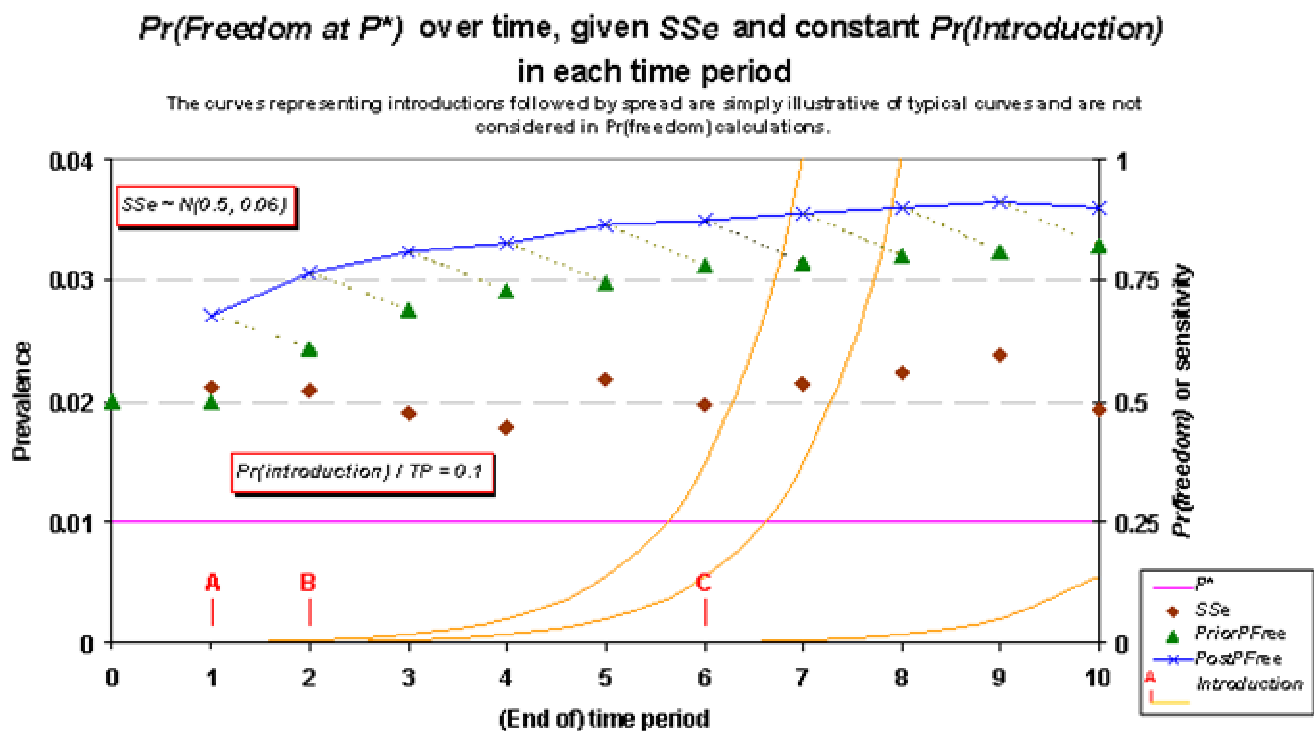

Figure 13 Pr(Freedom at P*) over time with disease introductions

Typically, disease spread follows an exponential curve initially, and this is illustrated in Figure 13. We are assessing the probability of country freedom at the design prevalence P^*, and not making any assertions about disease present at lower prevalence[10]. As Figure

illustrates, the interval from introduction of disease to the prevalence P^* being reached may be several TPs[11], and the necessary adjustment to *PostPFree$_1$* should actually involve the probability that previously introduced infection reached a level of P^* during TP 2. However, if we assume that

a) the likely spread curve (the dynamics of potential disease spread in the population) remains the same over time; and

b) the probability of introduction is constant;

then the probability of disease reaching P^* during TP 2 is equal to *PIntro$_2$*[12]. In determining the appropriate adjustment to *PostPFree$_1$* to derive *PriorPInf$_2$* we must take into account the possibility that disease (at $\geq P^*$) was in fact present at the end of TP 1; a possibility represented by the probability (1 - *PostPFree$_1$*). So it is possible for disease to be introduced (or cross the P^* threshold) during TP 2 (with probability *PIntro$_2$*) whether or not it was present (at $\geq P^*$) at the end of TP 1. The appropriate adjustment is therefore

$$PriorPInf_{tp} = PostPInf_{tp-1} + PIntro_{tp} - PostPInf_{tp-1} \times PIntro_{tp} \tag{26}$$

or

$$PriorPInf_{tp} = (1 - PostPFree_{tp-1}) + PIntro_{tp} - PIntro_{tp}(1 - PostPFree_{tp-1}) \tag{27}$$

This is the formula used to generate the values for *PostPFree$_{tp}$* and *PriorPInf$_{tp}$* in Figure 12 & Figure 13, with a constant *PIntro* as shown. In these examples *SSe$_{tp}$* is allowed to vary over time ($\sim N(SSeMean, SSeSD)$) with the values for *SSeMean* and *SSeSD* as shown on the figures. The value selected for *PIntro* is also shown on each figure. A range of summary curves for *PostPFree* is also shown in Figure 14.
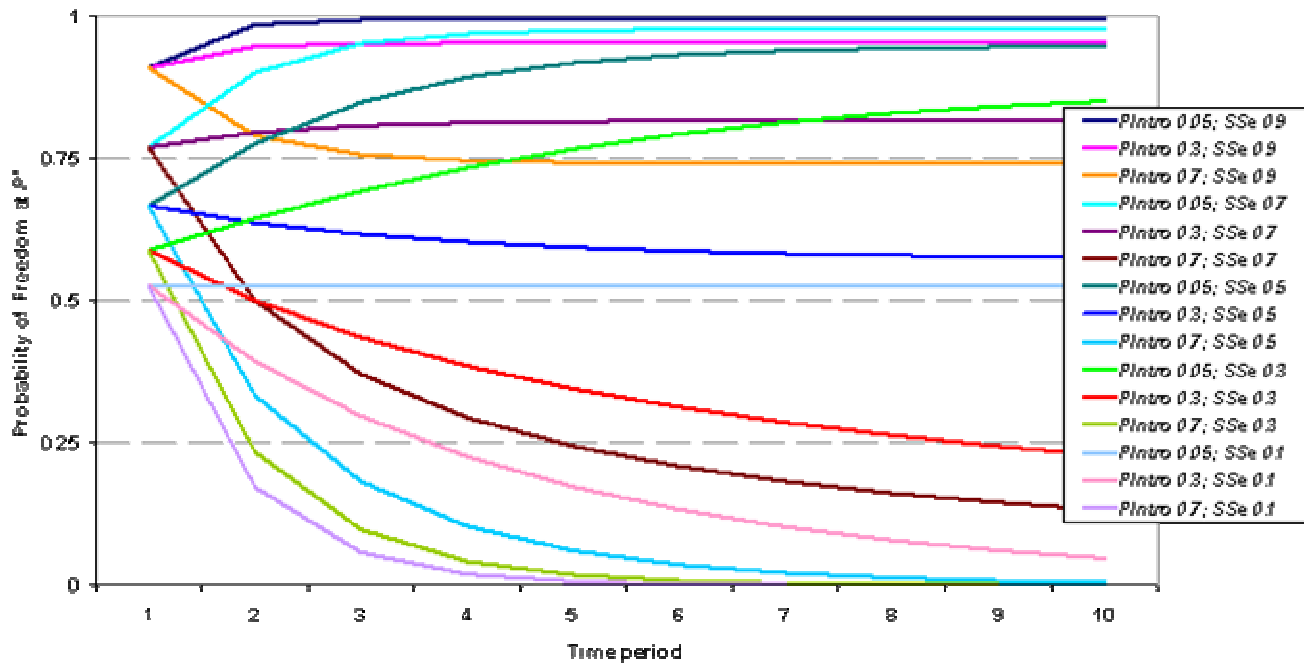
Figure 14 Pr(Freedom at P*) for different *SSe/PIntro* combinations

[10] If disease is detected the country is pronounced infected, whatever the prevalence; the arguments here apply only to situations involving negative surveillance findings.

[11] Length of TP for analysis should be determined with likely spread curves in mind.

[12] If these assumptions cannot reasonably be made, adjustments to *PostPFree1* in order to derive *PriorPInf2* will need to be based on TP-by-TP modelling of disease introduction and spread.

Acceptable Level of Protection (ALOP)

It will often be the case that the results of a quantitative IRA are not available for estimating *PIntro*. Given that conscientious members of the WTO will always manage risks according to the SPS agreement, the risk associated with introduction of the disease in question may be supposed always to be consistent with the country's ALOP, so this may be used as the basis for a default value for *PIntro*. We can estimate roughly the probability of release and exposure (*ie PIntro*) as the maximum level of acceptable risk consistent with the ALOP, divided by the average consequences of introduction of the disease. In situations where specific IRA results are not available, this estimation of *PIntro* will certainly be a qualitative process. Biosecurity Australia's risk estimation matrix (Figure 15) is useful to visualise how it works. The maximum acceptable risk

level is shown in white; *Very Low* for Australia. The column for the appropriate consequence category (say *Extreme* for FMD) is followed to the white cell, and the likelihood category associated with this is read off at the left end of the relevant row – *Negligible* in this case. For a disease with *Low* consequences, the acceptable *PIntro* is *Low*.

If we then convert this qualitative category into the quantitative probability range that it represents, we have a numerical estimate for *PIntro* derived in a reasonably transparent manner.

	Consequences					
Likelihood of release and exposure per time period	**Negligible**	**Very low**	**Low**	**Moderate**	**High**	**Extreme**
High	Negligible	Very low	Low	Moderate	High	Extreme
Moderate	Negligible	Very low	Low	Moderate	High	Extreme
Low	Negligible	Negligible	Very low	Low	Moderate	High
Very low	Negligible	Negligible	Negligible	Very low	Low	Moderate
Extremely low	Negligible	Negligible	Negligible	Negligible	Very low	Low
Negligible	Negligible	Negligible	Negligible	Negligible	Negligible	Very low

Figure 15 Biosecurity Australia's risk estimation matrix for qualitative categories. Australian ALOP is represented by Very Low risk.

Another potential approach to developing an on-going level of confidence in freedom based on continuing negative surveillance is to determine the desired probability that the country is free at P^* (presumably as prescribed by the Terrestrial Animal Health Code) and the sensitivity of the ongoing surveillance system. From these two we can calculate the maximum level of *PIntro* which will result in maintenance of the desired *PostPFree*, using a simple model based on (2) and (4) above. This potentially gives a transparent way of calculating the probability component of acceptable risk consistent with ALOP, and even of determining ALOP, for those motivated by a simple, pragmatic interpretation of the SPS agreement.

Stochastic modelling in the context of disease freedom - Principles

In its simplest form, a scenario tree is a branching series of probabilities. Calculation of the probability of an event involves the multiplication of each probability down the branches of the tree, and summing the result for each branch that will produce the specified event. All relevant outcomes/probabilities to produce such events should be

considered. This simple form of the tree is purely deterministic, in that it will produce the same result every time you evaluate the same probability.

Unfortunately, many of the probabilities commonly required for inclusion in the scenario tree are not known fixed values. This is because there is either some uncertainty about their true value, or the true value is not fixed, but varies. Stochastic (or 'Monte Carlo') modelling is the technique used to capture this uncertainty and/or variability. It is important to differentiate between the two concepts because uncertainty can be reduced through generation of more information, whereas variability is a biological fact. This must also be considered when interpreting the output of such stochastic models.

In stochastic modelling, the probabilities in the scenario tree are defined not as fixed values, but as distributions (described by specified parameters). When the probability of the model outcome is calculated, values for each of the branch probabilities used in the model are selected at random from the distributions specified for each of them. This is done repeatedly (multiple "iterations"), and for each iteration a different value is selected from the specified distribution for each branch probability. The result is therefore different for every iteration, depending on the values of the different branch probabilities chosen.

For a single iteration, this provides little benefit. A new result is obtained, somewhat different to that obtained by using fixed probabilities. This result is just as valid as the result based on fixed probabilities, as it is based on the possible range of values for the branch probabilities. However, it is virtually impossible to interpret such output on its own. Instead, the value of stochastic modelling which uses distributions as inputs, is that it can produce a distribution as its output. This is a achieved by running the model many times until the range of different results that are produced build up to form a distribution. The distribution indicates the range and frequency of different results that may occur in the system, given the variability and uncertainty in the input probability estimates.

Practical implementation

The process of implementing a stochastic scenario tree model is relatively straightforward. It is made vastly easier by the existence of dedicated stochastic modelling software such as PopTools or @RISK, which work with existing spreadsheet software such as Microsoft Excel.

Broadly speaking, the steps are:

1. Build a deterministic model using a spreadsheet. Fixed probabilities (the most likely values) for each branch of the tree will be used at this stage. These can be used to check the plausibility of the model. Refine the model accordingly and assign a range of values for each probability.

2. Decide on the appropriate type of distribution to describe more accurately each of these assigned values to the probabilities. Note that some of the input probabilities should not be described by a distribution, but a fixed value, because either:
   - They define the assumptions upon which the model is based (for instance, the design prevalence values); or
   - The probability or proportion is in fact known with certainty and non-varying. This may be because exhaustive data (*eg* census data on production types) is available to calculate directly the exact probability/proportion.
3. For each distribution, determine or estimate the appropriate parameters to describe the distribution.
4. Enter the parameters in a list on your model spreadsheet.
5. Modify your model by replacing the fixed branch probabilities with the appropriate probability distribution function (many are available in both PopTools and @RISK, describing a wide range of distributions), referencing the parameter list already drawn up.
6. Set up the necessary simulation options, such as naming outputs, report types, number of iterations etc.
7. Run the model and examine the output distribution that is produced. The output distribution of the probability of the event of interest can be summarised using standard statistics (mean, mode, percentiles).

Most of these steps are straightforward once one is familiar with the modelling software. The most challenging step is determining the appropriate distribution (and associated parameters) to describe input probabilities and proportions. A full discussion of the best way to go about this process is beyond the scope of this course. For a more detailed discussion of distributions in stochastic modelling, see Vose (2008). The following rules of thumb may be useful.

- Probabilities derived from expert opinion: use **PERT** (minimum, most likely, maximum) or **BETA**(a_1, a_2), where a_1 and a_2 are derived from the mode (most likely) and 95 (or other) percentile supplied by the expert(s), using software such as *Betabuster* or the online calculator http://www.ausvet.com.au/pprev/content.php?page=BetaParams
- Probabilities based on observed distributions: use non-parametric functions including **Histogram**, **General**, **Discrete**, **Cumulative.**
- Distribution of the time between randomly occurring events: **Exponential** (mean time between events).
- Distribution of the time taken for a number of events to occur: **Gamma** (number of events, mean time between events).
- Distribution of the number of successes from a number of independent trials with equal probability of success: **Binomial**(number of trials, constant probability of success).
- Distribution of the number of independent trials (x) with equal probability of success (p) to observe a given number of successes (s): x=s + **NegBin**(s,p). The

geometric distribution is the special case of the negative binomial, where the number of successes is one.

- Estimation of population prevalence from results of a representative survey which found s infected among n sampled: **Beta**$(s + 1, n - s + 1)$

If the model is correctly set up, and you have specified your calculated system sensitivity as the output variable, when the simulation is complete, you will be presented with a histogram showing the frequency distribution of the calculated system sensitivity.

In calculating the sensitivity of a SSC using a stochastic scenario tree model, probability distributions are used primarily to represent the uncertainty of model parameters. If "data" are simulated, however (e.g. in calculating $SSe_{Representative}$ for sensitivity ratios) they may also be used to model variability. In this situation it may be considered important to separate the effects of uncertainty and variability in the analysis – see Vose (2008) for guidance on how to achieve this.

Data from multiple sources

The combination of data from multiple surveillance components is useful in practice because of the gain in sensitivity of the whole surveillance system. However, this gain may be less than expected due to the dependence among the components. In Combination of data from multiple sources, we showed a simple formula that allows the combination of evidence under the premise of independence. In the same section, an approach was sketched to solve this problem. In the next few sections, we shall focus more on the reasons of dependence among surveillance components.

Sampling probability on animal level

On the animal level, observations may be dependent because they were made on the same animal (eg testing of live animals and testing at slaughter).

The sampling probability can be established if the sampling coverage (empirical sampling fractions) is known. A dependency can occur as a result of a correlation of the sampling probabilities among the surveillance components and implies also a correlation of the probability of not being selected for testing. Surveillance components may have specific *escape routes*, *ie* missing value processes that are correlated with some informative factors. Missing values for risk animals could occur simultaneously in different surveillance components. In the most extreme case, the disease of interest could be the exclusion factor.

If the animals of the population are individually registered and the surveillance data are indexed with the animal numbers, the degree of overlap between the sampling coverage of the two (or more) surveillance components can be assessed quantitatively. The joint sampling coverage on animal-level could be directly estimated and used to adjust for this

type of lack of independence. Reasons for correlated sampling probabilities should be investigated to rule out substantial biases.

Sampling probability on herd level and for population strata

The principles described above apply also to herds or other primary sampling units as well as to relevant strata of the population. Examples for the latter are production types with less intensive surveillance coverage. Backyard or pet animals are typically excluded from any regular surveillance. Intensive production types may be over-represented in many of the surveillance components, which would lead to correlation of the sampling probabilities on herd-level. This could potentially lead to a bias towards high values of sensitivity. Use of risk category nodes representing such strata can deal with such biases within a *SSC*.

Base-line dependence among surveillance components

Even when the sampling probabilities of the surveillance components are independent, a certain overlap of testing will occur just by chance at both animal-level and herd-level. This overlap of sampling can be estimated using the marginal sampling probabilities of the single components.

Dependence among diagnostic methods

False negative results of the diagnostic tests used in different components may be correlated (see section on Multiple diagnostic tests). If this is the case, the gain by using multiple surveillance components is less that expected. The estimation of error correlation of diagnostic tests is not trivial but possible if gold standard information is available (Hanson et al., 2000).

Estimating the surveillance system's sensitivity under dependence

The different sources of dependence among *SSC*s can be analysed separately. Methods to establish the surveillance system's sensitivity that account simultaneously for different reasons for dependence are still lacking.

Current issues

In this section, we would like to share our view of priority areas for research into the methods in the area of disease freedom.

Independence of tests

The estimation of error correlations for diagnostic tests used in surveillance systems is an important milestone. Methods that require a gold standard could eventually be supplemented with latent class analysis approaches.

Independence of surveillance system components

The dependence structure among the outcomes of surveillance components should be thoroughly investigated for some well-documented systems. Specialised statistical methods should be developed that allow unbiased estimates of the systems' sensitivity. These methods should also be suitable to investigate the surveillance coverage of the population.

Eliciting and combining expert opinion

If data are to be replaced with expert opinion, the methodology of eliciting and summarising those opinions becomes a central quality criterion. These methods should take into account the size of the expert panels (usually small), personal agendas, levels of expertise and sociological group phenomena. Approaches to resolve conflicting opinions should be critically assessed.

Non-quantitative analyses

The methods developed in this project and applied in the various case studies for analysis of scenario trees are strictly quantitative in nature. They rely on quantitative values derived either empirically or from expert opinion as inputs for probabilities, proportions and relative risks in the scenario trees. These inputs are expressed as either fixed values or as specified probability distributions. Similarly, model outputs such as component sensitivity and probability of freedom are also expressed quantitatively, usually as probability distributions.

One objective of this project was to investigate the potential for use of qualitative methods as an alternative to quantitative methods when quantitative values are not

available. In this context qualitative methods are ones where input values are expressed as qualitative statements (words) such as "low", "moderate", "high", etc, rather than as quantitative values. The conclusion from this investigation is that qualitative methods are not appropriate or applicable for this methodology. This conclusion is based on the fact that the methodology relies on complex calculations for the calculation of adjusted risks, expected probability of infection and unit and component sensitivities which would not be possible using qualitative values. Similarly, semi-quantitative methods based on an arbitrary scale or raking would also not be possible with this method. In addition, the method used is flexible enough to allow for uncertainty and imperfect knowledge about parameters through the use of expert opinion and appropriate probability distributions to express the associated uncertainty, so that non-quantitative methods are not needed.

One possible alternative for incorporation of non-quantitative values would be to use an arbitrary scale of probability and risk values to convert qualitative terminology into quantitative ranges, as has been used in import risk analysis. In fact this approach is likely to be more difficult and prone to error because no single conversion scale is likely to adequately express the appropriate ranges of values for different parameters such as population proportions, test sensitivities and relative risk values, resulting in an excessively complex and cumbersome approach. This approach is therefore not recommended and is unnecessary, given the ability to directly express values as probability distributions.

References & Bibliography

Audigé, L.; Beckett, S. A quantitative assessment of the validity of animal-health surveys using stochastic modelling. *Prev. Vet. Med.* 1999**,** 38, 259-276.

Baumgarten, L.; Heim, D.; Fatzer, R.; Zurbriggen, A.; Doherr, M. G. Assessment of the Swiss approach to Scrapie surveillance. *Vet. Rec.* 2002**,** 151, 545-547.

Cameron, A. R.; Baldock, F. C. A new probability formula for surveys to substantiate freedom from disease. *Prev. Vet. Med.* 1998a**,** 34, 1-17.

Cameron, A. R.; Baldock, F. C. Two-stage sampling in surveys to substantiate freedom from disease. *Prev. Vet. Med.* 1998b**,** 34, 19-30.

Cannon, R. M. Demonstrating disease freedom -- combining confidence levels. *Prev. Vet. Med.* 2002**,** 52, 227-249.

Cannon, R. M.; Roe, R. T. Livestock Disease Surveys: a Field Manual for Veterinarians; Australian Government Publishing Service: Canberra, 1982.

Doherr, M. G.; Oesch, B.; Moser, M.; Vandevelde, M.; Heim, D. Targeted surveillance for bovine spongiform encephalopathy. *Vet. Rec.* 1999**,** 145, 672.

Doherr, M. G.; Heim, D.; Fatzer, R.; Cohen, C. H.; Vandevelde, M.; Zurbriggen, A. Targeted screening of high-risk cattle populations for BSE to augment mandatory reporting of clinical suspects. *Prev. Vet. Med.* 2001**,** 51, 3-16.

Doherr, M. G.; Audigé, L. Monitoring and surveillance for rare health-related events: a review from the veterinary perspective. *Philos. Tr. R. Soc. London* 2001**,** 356, 1097-1106.

Dufour, B.; Audigé, L. A proposed classification of veterinary epidemiosurveillance networks. *Revue Scientifique et Technique Office International des Epizooties* 1997**,** 16, 746-758.

Dufour, B.; Pouillot, R.; Toma, B. Proposed criteria to determine whether a territory is free of a given animal disease. *Veterinary Research* 2001**,** 32, 545-563.

Enoe, C.; Georgiadis, M. P.; Johnson, W. O. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Prev. Vet. Med.* 2000**,** 45, 61-81.

Farrington, C. P.; Andrews, N. J.; Beale, A. D.; Catchpole, M. A. A statistical algorithm for the early detetction of outbreaks of infectious diseases. *Journal of the Royal Statistical Society (Series A, Statistics in Society)* 1996**,** 159, 547-563.

Gardner, I. A.; Stryhn, H.; Lind, P.; Collins, M. T. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Prev. Vet. Med.* 2000**,** 45, 107-122.

Gilks, W. R.; Richardson, S.; Spiegelhalter, D. J. Markov Chain Monte Carlo in Practice; Chapman & Hall: London, 1996.

Giovannini, A.; Bellini, S.; Salman, M. D.; Caporale, V. Spatial risk factors related to outbreaks of contagious bovine pleuropneumonia in northern Italy (1990-1993). *Revue Scientifique et Technique Office International des Epizooties* 2000**,** 19, 764-772.

Greiner, M.; Gardner, I. A. Epidemiologic issues in the validation of veterinary diagnostic tests. *Prev. Vet. Med.* 2000a**,** 45, 3-22.

Greiner, M.; Gardner, I. A. Application of diagnostic tests in veterinary epidemiologic studies. *Prev. Vet. Med.* 2000b**,** 45, 43-59.

Greiner, M.; Dekker, A. On the surveillance for animal diseases in small herds. *Prev. Vet. Med.* 2005, 70, 223-234.

Hanson, T. E.; Johnson, W. O.; Gardner, I. A. Log-linear and logistic modeling of dependence among diagnostic tests. *Prev. Vet. Med.* 2000**,** 45, 123-137.

Hoinville, L.; McLean, A. R.; Hoek, A.; Gravenor, M. B.; Wilesmith, J. Scrapie occurrence in Great Britain. *Vet. Rec.* 1999**,** 145, 405-406.

Hoinville, L. J.; Hoek, A.; Gravenor, M. B.; McLean, A. R. Descriptive epidemiology of scrapie in Great Britain: results of a postal survey. *Vet. Rec.* 2000**,** 146, 455-461.

Irwig, L.; Macaskill, P.; Glasziou, P.; Fahey, M. Meta-analytic methods for diagnostic test accuracy. *J. Clin. Epidemiol.* 1995**,** 48, 119-130.

James, A. D. Guide to epidemiological surveillance for rinderpest. *Revue Scientifique et Technique Office International des Epizooties* 1998**,** 17, 796-824.

Jordan, D.; McEwen, S.A. Herd-level test performance based on uncertain estimates of individual test performance, individual true prevalence and herd true prevalence. *Prev. Vet. Med.* 1998, 36, 187-209.

Kane, A. J.; Traub-Dargatz, J. L.; Losinger, W. C.; Garber, L. B.; Wagner, B. A.; Hill, G. W. A Cross-Sectional Study of Lameness and Lamenitis in U.S. Horses. In *Electronic Proceedings of the 9th Symposium of the International Society for Veterinary Epidemiology and Economics, August 6-12, 2000, Breckenridge, USA*; Salman, M. D., Morley, P. S., Ruch-Galle, R., Eds.; Colorado State University: Fort Collins, 2000; Abstract 441.

Kim, L. M.; Morley, P. S.; Traub-Dargatz, J. L.; Salman, M. D.; Gentry, W. C. Factors associated with Salmonella shedding among equine colic patients at a veterinary teaching hospital. *J. Am. Vet. Med. Assoc.* 2001**,** 218, 740-748.

Lilienfeld, D. E.; Stolley, P. D. Foundations of Epidemiology; Oxford University Press: New York, Oxford, 1994.

Martin, P. A. J.; Cameron, A. R.; Greiner, M. Demonstrating freedom from disease using multiple complex data sources 1: A new methodology based on scenario trees. *Prev. Vet. Med.* 2007**,** 79, 71-97.

Martin, P. A. J.; Cameron, A. R.; Barfod, K.; Sergeant, E. S. G.; Greiner, M. Demonstrating freedom from disease using multiple complex data sources 2: Case study - Classical swine fever in Denmark. *Prev. Vet. Med.* 2007**,** 79, 98-115.

Martin, S. W.; Meek, A. H.; Willeberg, P. Veterinary Epidemiology. Principles and Methods; Iowa State University Press: Ames, Iowa, United States, 1986.

Mintiens, K.; Verloo, D.; Venot, E.; Laevens, H.; Dufey, J.; Dewulf, J.; Boelaert, F.; Kerkhofs, P.; Koenen, F. Estimating the probability of freedom of classical swine fever virus of the East-Belgium wild-boar population. *Prev. Vet. Med.* 2005**,** 70, 211-222.

Morgan, K. L.; Nicholas, K.; Glover, M. J.; Hall, A. P. A questionnaire survey of the prevalence of scrapie in sheep in Britain. *Vet. Rec.* 1990**,** 127, 373-376.

Noordhuizen, J. P. T. M.; Frankena, K.; van der Hoofd, C. M.; Graat, E. A. M. Application of Quantitative Methods in Veterinary Epidemiology; Wageningen Press: Wageningen, 1997.

OIE (2000) Recommended standard for epidemiological surveillance systems for Rinderpest. Office International des Epizooties, International Animal Health Code Part 3, Section 3.8, Appendix 3.8.1.

Rogan, W. J.; Gladen, B. Estimating prevalence from the results of a screening test. *Am. J. Epidemiol.* 1978**,** 107, 71-76.

Rothman, K. J.; Greenland, S. Measures of Disease Frequency. In *Modern Epidemiology*; Rothman, K. J., Greenland, S., Eds.; Lippincott–Raven Publishers: Philadelphia, 1998; Chapter 3.

Salman, M.D. (*ed.*) Animal disease surveillance and survey systems: methods and applications. Iowa State Press, 2003.

Schlosser, W.; Ebel, E. Use of a Markov-chain Monte Carlo model to evaluate the time value of historical testing information in animal populations. *Prev. Vet. Med.* 2001, 48, 167-175.

Schreuder, B. E. C.; Jong, M. d.; Pekelder, J. J.; Vellema, P.; Broker, A. J. M.; Betcke, H.; de Jong, M. C. M. Prevalence and incidence of scrapie in the Netherlands: a questionnaire survey. *Vet. Rec.* 1993**,** 133, 211-214.

Simmons, M. M.; Ryder, S. J.; Chaplin, M. C.; Spencer, Y. I.; Webb, C. R.; Hoinville, L. J.; Ryan, J.; Stack, M. J.; Wells, G. A. H.; Wilesmith, J. W. Scrapie surveillance in Great Britain: results of an abattoir survey, 1997/98. *Vet. Rec.* 2000**,** 146, 391-395.

Spiegelhalter, D. J.; Myles, J. P.; Jones, D. R.; Abrams, K. R. Methods in health service research. An introduction to bayesian methods in health technology assessment. *Br. Med. J.* 1999, 319, 508-512.

Stärk, K. D. C. Animal health monitoring and surveillance in Switzerland. *Aust. Vet. J.* 1996**,** 73, 96-97.

Suess, E. A.; Gardner, I. A.; Johnson, W. O. Hierarchical Bayesian model for prevalence inferences and determination of a country's status for an animal pathogen. *Prev. Vet. Med.* 2002**,** 55, 155-171.

Thrusfield, M. V. Veterinary Epidemiology; Blackwell Science: London, 1995.

Tillotson, K.; Savage, C. J.; Salman, M. D.; Gentry-Weeks, C. R.; Rice, D.; Fedorka-Cray, P. J.; Hendrickson, D. A.; Jones, R. L.; Nelson, A. W.; Traub-Dargatz, J. L. Outbreak of Salmonella infantis infection in a large animal veterinary teaching hospital. *J. Am. Vet. Med. Assoc.* 1997**,** 211, 1554-1557.

Traub-Dargatz, J. L.; Garber, L. B.; Hill, G. W.; Wagner, B. A.; Losinger, W. C.; Seitzinger, A. H.; Rodriguez, J. M.; Stanton, N. G. Overview of the Initial Phase of the National Animal Health Monitoring Systems (NAHMS) Equine '98 Study. In *Electronic Proceedings of the 9th Symposium of the International Society for Veterinary Epidemiology and Economics, August 6-12, 2000, Breckenridge, USA*; Salman, M. D., Morley, P. S., Ruch-Galle, R., Eds.; Colorado State University: Fort Collins, 2000; Abstract 61.

Traub-Dargatz, J. L.; Garber, L. P.; Fedorka-Cray, P. J.; Ladely, S.; Ferris, K. E. Fecal shedding of Salmonella spp by horses in the United States during 1998 and 1999 and detection of Salmonella spp in grain and concentrate sources on equine operations. *J. Am. Vet. Med. Assoc.* 2000**,** 217, 226-230.

Wagner, B. A.; Wise, D. J.; Khoo, L. H. Health Monitoring in the U.S. Catfish Industry. In *Electronic Proceedings of the 9th Symposium of the International Society for Veterinary Epidemiology and Economics, August 6-12, 2000, Breckenridge, USA*; Salman, M. D., Morley, P. S., Ruch-Galle, R., Eds.; Colorado State University: Fort Collins, 2000; Abstract 358.

Vose, D. Risk Analysis: A Quantitative Guide, 3rd edition; John Wiley and Sons: Chichester, 2008.

Ward, M. P.; Carpenter, T. E. Analysis of time-space clustering in veterinary epidemiology. *Prev. Vet. Med.* 2000**,** 43, 225-237.

Zepeda, C.; Salman, M.; Ruppanner, R. International trade, animal health and veterinary epidemiology: challenges and opportunities. *Prev. Vet. Med.* 2001**,** 48, 261-271.

Fischer, E. A. J.; van Roermund, H. J. W.; Hemerik, L.; van Asseldonk, M. A. P. M.; de Jong, M. C. M. Evaluation of surveillance strategies for bovine tuberculosis (Mycobacterium bovis) using an individual based epidemiological model. *Prev. Vet. Med.* 2005**,** 67, 283-301.

Hadorn, D. C.; Rüfenacht, J.; Hauser, R.; Stärk, K. D. C. Risk-based design of repeated surveys for the documentation of freedom from non-highly contagious diseases. *Prev. Vet. Med.* 2002**,** 56, 179-192.

Hueston, W. D. Science, politics and animal health policy: epidemiology in action. *Prev. Vet. Med.* 2003**,** 60, 3-12.

Stärk, K. D. C.; Horst, H. S.; Kelly, L. Combining expert opinions: a comparison of different approaches. *Proc. 9th ISVEE*, Breckenridge 2000.

Stärk, K. D. C.; Regula, G.; Hernandez, J.; Knopf L.; Fuchs K.; Morris R.S.; Davies P. Concepts for risk-based surveillance in the field of veterinary medicine and veterinary public health: Review of current approaches. *BMC Health Services Research*. 2006**,** 6:20

Meats, A.; Clift, A. D. Zero catch criteria for declaring eradication of tephritid fruit flies: the probabilities. *Australian J of Experimental Agriculture*. 2005**,** 45, 1335-1340.

Vourc'h, G.; Bridges, V. E.; Gibbens, J.; De Groot, B. D.; McIntyre, L.; Poland, R.; Barnouin, J. Detecting emerging diseases in farm animals through clinical observations. *Emerging Infectious Diseases*. 2006**,** 12, 204- 210.

Böhning, D.; Greiner, M. Evaluation of cumulative evidence for freedom from BSE in birth cohorts. *Eur. J. Epidemiol.* 2006**,** 21, 47-54.

Project Credits

A wide range of scientists, projects and funding have contributed to the methodology described in this document. The main participants are listed here, but the authors wish to make special thanks to the participants of training courses and workshops, too numerous to mention, but who, through their keen engagement, questionning and suggestions have all strengthened and extended the framework described here.

Contributing Authors

Tony Martin, BA VetMB MPVM
>   Department of Agriculture & Food Western Australia,
>   PO Box 1231, Bunbury, WA 6231, Australia
>   TMartin@agric.wa.gov.au

Angus Cameron, BVSc, MVS, PhD MACVSc
>   AusVet Animal Health Services,
>   140 Falls Road, Wentworth Falls, NSW 2782, Australia
>   angus@ausvet.com.au

Matthias Greiner, PD Dr. med. vet., MSc, Dipl. ECVPH
>   Bundesinstitut für Risikobewertung (BfR), Wissenschaftliche
>   Querschnittsaufgaben
>   Fachgruppe 33 - Epidemiologie, Biometrie und mathematische Modellierung,
>   Alt-Marienfelde 17-21, D-12277 Berlin, Germany
>   m.greiner@bfr.bund.de

Jenny Hutchison, BVetBiol(Dist) BVSc(Hons I) MS PhD Dip ACVIM(LAM)
>   AusVet Animal Health Services,
>   124 Perry Drive, Chapman ACT 2611 Australia
>   jenny@ausvet.com.au

Evan Sergeant, BVSc (Hons II), MACVSc, PhD
>   AusVet Animal Health Services,
>   69 Turner Cr, Orange NSW 2800, Australia
>   evan@ausvet.com.au

Nigel Perkins, BVSc(Hons I), MS, PhD, Dip ACT, FACVSc
>   AusVet Animal Health Services,
>   30 Plant Street, Toowoomba 4350 QLD, Australia
>   nigel@ausvet.com.au

Mo Salman, BVMS, MPVM, PhD DACVPM F.A.C.E.
>   Professor of Epidemiology
>   Animal Population Health Institute (APHI),
>   Colorado State University, Ft. Collins, CO 80523-1681, USA
>   M.D.Salman@colostate.edu

Supporting Organisations

Australian Biosecurity Cooperative Research Centre for Emerging Infectious Disease
>   The AB-CRC funded the project under which further research was undertaken, a number of case studies conducted, and this web site developed. The project was led by Tony Martin. Click here for more information.

Danish International EpiLab
>   The methodology was first developed in projects funded by the Disease Freedom theme of the Danish International EpiLab in the Danish Institute for Food and Veterinary Research, undertaken by Tony Martin and Angus Cameron, with the assistance of Matthias Greiner. Click here for more information.

Zero Prevalence Workshop
>   The original ideas for this methodology were initially crystalised at a workshop convened by Mo Salman in Fort Collins, Colorado, immediately after the 2000 ISVEE meeting, and attended by many leading epidemiological researchers.

AusVet Animal Health Services
>   In addition to providing four of the leading scientist involved in the methodologies' development, AusVet also developed the software and provides the web server and web space to host this web site.

Suggested notation

In this document:

- node names are shown in small capitals (e.g. HOUSING)
- branch names are shown in italics (e.g. *Dairy*)
- variable names are shown in italics (e.g. *PrSSC_Dairy*)

Variable names
Design prevalence

Variable names

- Variable names to be based on abbreviations of descriptions of the quantities to which they apply, e.g. *Se* for sensitivity; *PrP* for Population Proportion.
- The conditionality of all variables on all preceding nodes in the tree is assumed, and not explicitly included in their names, except where such distinctions need to be made, and in these cases:
    - the specific limb of the tree in which a variable occurs is described in the variable name, not in subscripts
    - the variable name should include all relevant node branch names
    - node branch names included in the variable name should be ordered as in their sequence in the limb of the tree, e.g. *PrP_SJ_Breeder_Adult* for the population proportion of Adults in a Breeder herd in South Jutland (where the sequence of nodes in the tree is COUNTY, FARM TYPE, AGE); NOT *PrP_Adult_Breeder_SJ* or any other sequence
    - underscores should be used to separate branch names within a variable name

    Examples of variable names which assume conditionality are:

    - *SeU* to denote unit-level sensitivity in a general discussion of how this variable should be used (for example, in calculating *SeH*); It is understood that different limbs of the tree may well have different values for *SeU*
    - *PrSSC_Dairy* to denote the proportion of herds processed in the SSC that are of type *Dairy*; it is understood that this may be different from one limb to another in the tree, the potential differences being apparent from the structure of the tree
- The grouping level at which a variable applies is incorporated in the variable name, NOT as a subscript, e.g. *SeH* for herd-level sensitivity; this precedes any applicable names of node branches in the variable name, e.g. *EPIH_Layer_Caged* for the effective probability that a farm is infected for *Caged* HOUSING in the *Layer* COMPARTMENT
- **Subscripts** are numerical indices, and are used for three purposes, depending on context:
    - as an index number for the individuals to which a variable name is applicable, e.g. SeH_i for the herd-level sensitivity in the ith herd, in a context where multiple herds are being considered
    - as an index for the branch of a node to which a node-level variable name is applicable, e.g. PrP_j for the population proportion of (herds) in county j, in a context where multiple branches of the COUNTY node are being considered

- as an index denoting the time period to which the variable relates, in analysis of ongoing surveillance data split into successive time periods, e.g. $PIntro_{tp}$ for the probability of infection being introduced during time period tp.
  - a non-numeric subscript is also used in design prevalence notation (see below)
- Double (or occasionally more) subscripts may be used, but only in contexts (formulae) where it aids comprehension to represent variables this way; e.g:

$$SeR_j = 1 - \prod_{s=1}^{S} \prod_{t=1}^{n_s} (1 - ARH\_HT_s \times P^*_H \times SeH_{i,s})$$

where SeR_j is the region-level sensitivity for the jth region; ARH_HT_s is the adjusted risk for the sth herd type at the HERD TYPE risk node in the limb; P^*_H is the herd-level design prevalence; $SeH_{i,s}$ is the herd-level sensitivity in the ith herd of type s; there are S herd types; there are n_s herds of the sth herd type.

In these situations, subscripts are separated by commas, and the first mentioned applies to the lowest-level unit or group. It is unlikely that time period indices will need to be mixed with branch or unit indices.

Only occasionally is such notation going to aid comprehension, I think – it may make for conciseness, but not necessarily comprehension!

Design prevalence

Design prevalence is denoted P^* with a subscript letter denoting the level to which it applies; typically P^*_U for unit-level design prevalence, and P^*_H for the herd-level design prevalence. P^*_U, when specified as the only design prevalence, is the proportion of units in the population which are infected. When P^*_H is also specified, P^*_U is the proportion of units in an infected herd which are infected.

Sensitivity notation

- Within a component of a surveillance system (SSC), sensitivity at different grouping levels within the tree is denoted as explained above for variables in general (i.e. *SeU*; *SeH*; *SeR* etc., where U, H and R refer to different grouping levels, Unit, Herd and Region).
- The sensitivity of a diagnostic test should be given an explanatory name (i.e. not just *Se*), for example *ELISASe*, to avoid confusion.
- The sensitivity of the SSC is denoted *CSe*, the Component Sensitivity. Where multiple components are referred to, their sensitivities need to be named as in the variable name conventions above, e.g. *CSeSERO* for a SSC denoted SERO
- The sensitivity of the whole surveillance system (potentially with multiple components) is denoted *SSe*, the System Sensitivity.

- An actual *CSe* (i.e. based on the units actually processed) is generally referred to simply as *CSe* or *CSeSERO*, etc.; but when it needs to be distinguished from the *CSe* for a representative standard of the SSC, it is called *CSe_Actual* or *CSeSERO_Actual*, etc., and the representative standard *CSe* is then *CSe_Representative*, or *CSeSERO_Representative*, etc..
- The same applies to actual and representative standard *SSe*.

Probabilities

Variables representing probabilities are named as described above (Variable Names), based on:

- *P* (Probability), e.g. *PSamples* for the probability that samples are submitted; *PIntro* for the probability of infection being introduced.
- *PriorP* for a Bayesian prior probability
- *PostP* for a Bayesian posterior probability
- *EPI* for an Effective Probability of Infection (i.e. design prevalence multiplied by applicable adjusted risks), e.g. *EPIUAdult* for the effective probability that an *Adult* unit will be infected given that the herd/group is infected.

Branch proportions

Variable names are assigned as described above, based on:

- *Pr* for Proportion
- *PrP* for a proportion as found in the SSC reference population.
- *PrSSC* for a proportion of units processed in the SSC.

Risk nodes

Relative risks as specified by the user (i.e. unadjusted for population proportions) are named as described above (Variable Names), based on *RR* (Relative Risk):

- *RRH* for a relative risk applying to a herd-level infection node
- *RRU* for a relative risk applying to a unit-level infection node
- *etc*.

Adjusted risks (weighted by population proportions) are named based on *AR* (Adjusted Risk), but otherwise exactly as for *RR* (i.e. *ARH*, etc.).

Each branch of a risk category node used to capture population coverage will have associated with it a sensitivity weighting (probability of the branch being infected given that the population is infected) and these should be named (as above) based on *SW* (e.g. *SWDairy* for the weighting applied to the sensitivity of detection in the Dairy branch of a COMPARTMENT node).

Stochastic modelling using @RISK

In this project we are using Microsoft *Excel* and Palisade *@RISK*[13] as the modeling software. The following notes are supplied for reference for those who are not familiar with the software.

A few suggestions for spreadsheet modelling before we start:

- Make your spreadsheet easy to follow, for yourself and others
  - formatting
  - labelling
  - comments
- It is important to identify clearly model inputs and outputs
  - different colours are effective
- name cells and ranges wherever appropriate
  - use *Add output* to name *@RISK* outputs
  - use *RiskName()* to name *@RISK* inputs
  - use *Excel /Insert/Names/…* to name all relevant cells and ranges, (including *@RISK* inputs and outputs) so that formulae are easily read.

[13] Palisade Corporation: www.palisade.com.

For those not familiar with *@RISK*:

(This is not an *@RISK* manual – you will find one in your *@RISK* installation.) *@RISK* is an add-in for Microsoft *Excel*, providing the following additional functions within the Excel spreadsheet:

- entry of probability distributions into spreadsheet cells. This is done using a range of functions giving access to about 40 probability distributions. With each recalculation of the spreadsheet a new value is drawn from each distribution in the spreadsheet, using Monte Carlo sampling.
- simulation: you can perform simulations consisting of multiple iterations of the spreadsheet, where each iteration represents a recalculation of the spreadsheet, with associated resampling from probability distribution functions. *@RISK* will store all sampled values along with all calculated values for outputs, allowing analysis and presentation of simulation results, including graphing and sensitivity analysis.

@RISK provides new menus in *Excel*, new functions available through *Excel*'s *Insert / Paste function* dialog, and 2 new windows:

- the *@RISK* model window is accessed via the *Show @RISK Model window* button on the *@RISK* toolbar

- the *@RISK* results window is accessed via the *Show @RISK Results window* button on the *@RISK* toolbar.

To get back to your spreadsheet from one of these windows, use Alt-Tab or the *Show Excel window* button on the toolbar; closing the Model window will shut down *@RISK*.

Inputs and Outputs

@RISK refers to any cell containing an *@RISK* distribution function as an **input**. Such functions may or may not be what you regard as inputs to your model, so you need to be careful of how this word input is used.

@RISK outputs, on the other hand, must be specified. Unless you specify them, your model will have no outputs. You make a cell (or range) an output by selecting it and then clicking on the *Add Output* button on the *@RISK* toolbar. During simulation *@RISK* will then store values of outputs at each iteration, present you with a variety of statistics for the output's frequency distribution, and give you access to a range of analytical options for the output.

@RISK can also collect and store values for inputs from each iteration, but only when you tell it to. This is done in one of 2 ways:

- On the *Sampling* tab of the *Simulation settings* dialog box, check *All* under *Collect Distribution Samples*. Data for all inputs will now be collected at each simulation. This is fine, but it slows things down and uses a lot of memory for large simulations.

In general it is preferable to

- check *Inputs marked with Collect* in the same dialog. Now you can specify which inputs you want data stored for. Hit the *Display list of Inputs and Outputs* button on the toolbar, and on the right hand side of the *@RISK* Model window check the *Collect* box for the inputs you want collected during simulation.

Distribution functions

There are many to choose from, and they are accessed through the *Insert function* dialog box in *Excel* (*Paste function* in some versions of *Excel*), by selecting the *@RISK Distributions* category of functions. These functions require you to specify standard distribution parameters (mean, standard deviation, binomial probability, etc.), but you can also use the functions found in the category *@RISK distrib (Alt parms)* . Here you will be asked for such things as percentiles of the distribution, allowing you to specify probability distributions by using these alternative parameters.

To visualise sampling from the distributions you have specified each time you recalculate the spreadsheet, check *Monte Carlo* under *Standard Recalc* on the *Sampling* tab of the *Simulation Settings* dialog.

@RISK statistics functions

These are useful when you want to embed some summary statistic for a distribution (input or output) in your spreadsheet (e.g. the mean or 95th percentile of an output). You'll find them through *Insert/Paste function*, under *@RISK statistics*. You can watch them converge during simulation if you check *Update display* on the *Iterations* tab of the *Simulation Settings* dialog. This setting updates the display for all cells at every iteration, and can be fun the first time you run a simulation; but thereafter it serves mainly to slow down the calculations. In general, leave it turned off.

When you have run the simulation and have the required statistic in your spreadsheet cell, copy its value into the spreadsheet, otherwise you will lose it when you close the simulation file.

Simulation files

These files store all data and statistics etc from a simulation. In general there is no need to save such data unless the simulation took a long time to run, since it is easy to rerun the simulation. Of course the results will be slightly different unless you use the same random number seed (see *Sampling* tab of *Simulation Settings* dialog).

Saving simulation results

Click the *Report Settings* button on the *@RISK* toolbar to select a wide range of options for saving reports to Excel worksheets. A useful one-page summary for an output is obtained using *Quick Output Report*. Note the button in the bottom right of the dialog box giving the option to *Generate Reports Now*.

Simulation settings

For getting started:

- Use 5,000 iterations
- Use Latin Hypercube sampling.
- Use a *Fixed* random generator seed. (Choose your own – default is 1.)

Sensitivity analysis

See the manual for how to use the *Advanced sensitivity analysis* facility in *@RISK 4.5.2*. Standard sensitivity analysis (less flexible, but useful and quick) ranks inputs marked for

*collecti*on by their relative impact on the output, based on regression analysis or rank correlation. The tornado plot and sensitivities report are the standard output, and these are available after you have run your simulation (see *Report Settings*) as long as you marked relevant inputs with *RiskCollect()* as outlined above (*Inputs and Outputs*).

Correlated inputs

Where variables in your model are correlated and they are specified as *@RISK* input distributions, you can set appropriate levels of correlation among multiple inputs using the *Define Correlation* button on the toolbar in the *@RISK* Model window. First select the inputs to be correlated in the *Explorer* pane (left hand side) of the Model window, then hit the *Define Correlation* button. You are presented with a correlation matrix for your selected inputs, which will be saved into a new worksheet in your active workbook, called *@RISK Correlations*. New values for correlations may be entered into the *Excel* worksheet or into the Model window matrix. New inputs may be included in the matrix by dragging them across from the *Explorer* list in the Model window.

Using PopTools for Stochastic Spreadsheet Simulation

What is PopTools?

From the [PopTools web site](): "PopTools is a versatile add-in for PC versions of Microsoft Excel (97, 2000 or XP) that facilitates analysis of matrix population models and simulation of stochastic processes. It was originally written to analyse ecological models, but has much broader application. It has been used for studies of population dynamics, financial modelling, calculation of bootstrap and resampling statistics, and can be used for preparing spreadsheet templates for teaching statistics.

"When installed, PopTools adds a new menu item to Excel's main menu (see the slightly outdated screenshot), and also adds over a hundred new worksheet functions. The routines include array formulas for matrix decompositions (Cholesky, QR, singular values, LU), eigenanalysis (eigenvalues and real eigenvectors of square matrices) and formulas for generation of random variables (eg, Normal, binomial, gamma, exponential, Poisson, logNormal).

"Also included in PopTools are routines for iterating spreadsheets. These make it possible to run Monte Carlo simulations, conduct randomisation tests (including the Mantel test) and calculate bootstrap statistics.

"PopTools requires no programming knowledge, but to fully utilise the package you need some knowledge of matrix algebra, and some understanding of probability and statistics. It is therefore most suitable for those who have done some undergraduate statistics."

Download PopTools

PopTools can be downloaded here.

Installing PopTools

- Install the SOLVER add-in in Excel (Tools | Add-ins | Solver Add-in)
- run the setup file downloaded from the web site. It will install the add-in and demonstration files, and register the PopTools.xla file with Excel. At the end of the installation process an XLS readme file will be launched.

Setting up a scenario tree using PopTools

To be completed

Summary of relevant formulae

To be completed

The Australian Biosecurity CRC

The Australian Biosecurity Cooperative Research Centre (ABCRC) supports research in three programmes:

- Technologies to enhance detection
- Ecology of emerging infectious diseases
- Advanced surveillance systems

Project 3.010R in the *advanced surveillance systems* program is entitled *Quantification of confidence in disease freedom*. The material presented in this booklet has been developed within this project, and the associated training workshops are presented by research workers in this project.

Case Studies

In addition to development and refinement of the methodology for evaluation of surveillance for disease freedom, four case studies are being conducted:

1. Enzootic bovine leukosis in the Australian dairy industry (completed)
2. Human poliomyelitis in Australia
3. Bluetongue in the Australian free zone
4. Bovine Johne's disease in Western Australia

Software

Software for conducting scenario tree analysis of surveillance system components for disease freedom is being produced, and will be demonstrated during the course. The software will be freely accessible over the internet, and performs the following functions

- Constructs a scenario tree from the user's inputs
    - node types and names, branches and names, and sequence
    - branch probabilities, proportions and relative risks
- Presents the tree in an expandable / collapsible and readily edited format
- Presents graphical illustration of input and output probability / frequency distributions
- Calculates SSC sensitivity using the stochastic modelling process described in section 9 above
- Reports results of the analysis

Extensions to the methodology

In order to make the methodology as generally applicable as possible, the following are also addressed in the project:

- Methods for eliciting, capturing and modelling expert opinion
- Methods for dealing with semiquantitative and qualitative inputs

International EpiLab in Denmark

The methodology presented here was partly developed during the project visits of AC, TM and Mo Salman to the International EpiLab in Denmark. This Appendix gives a brief overview about this research centre and presents executive summaries of the research projects relevant to the topic of this course.

General information

The International EpiLab was established as an international research platform in veterinary epidemiology in Denmark in co-operation of the Danish Veterinary Institute with the Danish Veterinary and Food Administration, The Royal Veterinary and Agricultural University, the Danish Institute of Agricultural Sciences, the Danish Bacon and Meat Council, the Danish Cattle Federation and The Danish Poultry Council. The research network of International EpiLab also includes international guest scientists and an international advisory committee. Current research areas are documenting disease freedom, risk assessment for exotic diseases, epidemiological research in antimicrobial resistance, improved methods for herd classification, population demographics and disease transmission and animal health, animal welfare and medicine use. More on International EpiLab can be found under http://www.dfvf.dk/Default.asp?ID=9406.

International EpiLab project 3 (CSF; Cameron)

Documenting disease freedom in swine by combination of surveillance programmes using information from multiple non-survey-based sources.

Project period

1 July 2002 to 31 January 2003

Funding

Funded by the Directorate for Food, Fisheries and Agri Business under the Danish Ministry of Food, Agriculture and Fisheries, innovation law programme (93S-2465-Å02-01358). Co-funded by Danish Bacon and Meat Council.

Project leader

Kristen Barfod, Danish Bacon and Meat Council

Project partners

- Guest scientist: Angus Cameron , Australia
- DBMC: Kristen Barfod
- DVFA: Sten Mortensen
- DVI: Sven Erik Jorsal, Mette M. Larsen, René Bødker, Anne Bruun, Matthias Greiner
- Further co-workers: Evan Sergeant, Australia; Tony Martin (c/o International EpiLab)

Executive summary

The Agreement on Sanitary and Phytosanitary Measures (SPS agreement) of the World Trade Organisation requires that, in international trade, measures taken to protect animal, plant or human health should be based on scientific principles and not maintained in the absence of sufficient evidence. Countries support such measures by using science-based risk analysis, which in turn demands science-based assessment of the disease status (free or infected) of each of the trading partners. Traditionally, national disease status has been determined using structured cross-sectional surveys, which are generally difficult and expensive to implement. On-going surveillance may also be assessed by expert panels, but there are no accepted methods for quantifying either confidence in the surveillance process, or the probability of national disease freedom demonstrated thereby. This report presents a proposed framework and detailed methods for quantitative assessment of complex surveillance data from multiple sources, and an illustrative case study using evidence from three surveillance systems to demonstrate Denmark's freedom from

classical swine fever. The framework and its methodology have been developed jointly with other EpiLab theme 1 projects.

Framework for analysis
The scenario tree is proposed as the modelling format for analysis of surveillance systems under a null hypothesis of the country being infected at a level equal to or greater than specified design prevalences. A scenario tree is developed to represent all known significant factors influencing the probability that a unit in an infected population will be detected as infected. The conditional probabilities associated with each limb of the tree are then multiplied together to give the overall probability of each limb's outcome, and these are summed for all branches with positive outcomes to give the probability that the whole surveillance process will have a positive outcome for a randomly chosen population unit, given that infection is present in the country (the system unit sensitivity).

Independence and clustering models are described for analysis. Under the independence model, overall system sensitivity of detection is derived directly from system unit sensitivity, as the probability that one or more of the independent units processed would have positive surveillance outcomes, given an infected country. Under the clustering model, animals (and disease) are assumed to cluster in groups, and surveillance system sensitivity is calculated taking this into account, by stepwise aggregation of sensitivity at each grouping level in the tree.

Surveillance processes give either complete or incomplete coverage of the population, and the sensitivity of a process with incomplete coverage must be adjusted for its representativeness of the population. This is achieved through calculation and use of a sensitivity ratio for the process; the ratio of its sensitivity to that of a truly representative surveillance process.

The surveillance process's sensitivity, $Pr(\geq 1$ positive unit | country infected), is the confidence level for the statistical test of the null hypothesis. If one has a prior estimate of P(country is free of disease), one can then use Bayesian inference to calculate a posterior estimate of this probability, given the negative surveillance results.

Where multiple surveillance systems are available, the results of the analysis of each (whether they be survey-based or the result of scenario tree analysis) may be combined to produce an overall estimate of the confidence of the combined surveillance system.

While this research has developed the framework for a practical methodology to analyse complex surveillance data sources, it has also identified a number of areas of further research which would enhance the methodology. These include 1) standardised, transparent and acceptable methods for eliciting expert opinion, 2) methods to adjust the value of information based on the time of collection, and 3) methods to account for the lack of independence between surveillance systems when calculating the combined confidence that surveillance systems provide.

Case study: Classical Swine Fever in Denmark
The methodology described above was used to analyse three different surveillance systems that provide evidence of Danish freedom from classical swine fever. The surveillance systems examined were:

1. A structured CSF sero-surveillance system, based on the collection of blood samples at abattoirs. Sampling was targeted at adult animals, with differential sampling pressures for boars compared to sows, and for South Jutland compared to the rest of the country;
2. Abattoir inspections (ante-mortem and post-mortem) routinely carried out at all abattoirs, primarily for food safety purposes; and
3. Clinical surveillance based on farmer observation, and routine visits by veterinarians to farms.

Each surveillance system was modelled using separate scenario trees, and estimates of the system confidence generated using stochastic modelling. Data sources used in the analysis included the Central Husbandry Register database, results of serological analysis of blood samples, abattoir slaughter records, and the VetStat drug prescription database (used as a proxy for veterinary visits). A number of parameters in each model were provided either by an expert informant, or through educated guesses.

Analyses were performed using a number of different design prevalence combinations, to examine the impact of the assumptions under the null hypothesis. In addition, for each surveillance system, a parallel analysis was conducted based on a hypothetical fully representative system using the same surveillance approach. For instance, in the case of sero-surveillance, this involved conceptually sampling from the farm population (rather than targeted sampling from the abattoir population). For meat inspections, it was based on the theoretical examination of animals selected from the farm population.

The results of analysis indicated that (not surprisingly) the estimated system sensitivity (or equivalently, confidence in the surveillance system) was very sensitive to the design prevalence assumptions under the null hypothesis. When reduced to a common period of one month's worth of surveillance, and based on those values used in the study, the sensitivity of the sero-surveillance system was estimated as 26.37% with a 5th to 95th percentile range of 23.44% to 27.87%. The sensitivity for the meat inspection system was 67.80% (39.46% to 90.34%) and for the clinical surveillance system was 93.80% (90.77% to 96.43%).

The sensitivity ratio is the ratio of the sensitivity of the actual system, to the sensitivity of a theoretical fully representative system. It indicates the effect of targeting the system, and indicates if a system is more or less effective than random selection. The sensitivity ratio for the sero-surveillance system was 3.73, for the meat inspection system was 0.998 and for the clinical surveillance system was 0.991. This indicates that the sero-surveillance system was very well targeted and much more efficient that simple population sampling.

The other two systems were essentially equivalent to representative population sampling. The combined sensitivity of the three surveillance systems was calculated providing a monthly confidence of 98.53%. If surveillance data over the period of one year were considered, the confidence would increase to essentially 100% $(1 - (1'10-22))$.

The strength of evidence for freedom from CSF is undeniable, and sensitivity analysis shows that even if the confidence in one or more systems is greatly overestimated, the annual confidence in the combined surveillance system well exceeds international requirements. Nevertheless, it is recommended that further research be undertaken in this area, including the use of more formal methods to generate estimates from expert opinion, and the application of a proposed methodology to account for the lack of independence between surveillance systems.

International EpiLab project 4 (HPAI; Martin)

Documenting freedom from Highly Pathogenic Avian Influenza (HPAI) in Danish poultry.

Project period

1 July 2002 to 30 November 2002

Funding

Funded by the Directorate for Food, Fisheries and Agri Business under the Danish Ministry of Food, Agriculture and Fisheries, innovation law programme (93S-2465-Å02-01359). Co-funded by The Danish Poultry Council.

Project leader

Poul H. Jørgensen, DVI

Project partners

- Guest scientist: Tony Martin, Australia
- DPC: Thorkil Ambrosen, Jacob Bo Christensen
- DVFA: Hanne M. Hansen
- DVI: Poul H. Jørgensen, Vibeke F. Jensen, Mette M. Larsen, René Bødker, Anne Bruun, Matthias Greiner
- Further co-workers: Angus R. Cameron (c/o International EpiLab)

Executive summary

The SPS agreement of the WTO requires that, in international trade, measures taken to protect animal, plant or human health should be based on scientific principles and not

maintained in the absence of sufficient evidence. Countries support such measures by using science-based risk analysis, which in turn demands science-based assessment of the disease status (free or infected) of each of the trading partners. Traditionally, national disease status has been determined using structured cross-sectional surveys, which are generally difficult, expensive, and ephemeral in their applicability. On-going surveillance may also be assessed by expert panels, but there are no accepted methods for quantifying either confidence in the surveillance process, or the probability of national disease freedom demonstrated thereby. This report presents a proposed framework and detailed methods for quantitative assessment of complex surveillance data from multiple sources, and an illustrative case study of the diagnostic surveillance process for HPAI in Denmark. The framework and its methodology have been developed jointly with other EpiLab theme 1 projects.

Framework for analysis
The scenario tree is proposed as the modelling format for analysis of surveillance systems under a null hypothesis of the country being infected at a level equal to or greater than specified design prevalences. A tree is developed using infection, category and detection nodes to represent all factors influencing the probability that a population unit will lead to a positive outcome of the surveillance process. The conditional probabilities associated with each limb of the tree are then multiplied together to give the overall probability of each limb's outcome, and these are summed for all branches with positive outcomes to give the probability that the whole surveillance process will have a positive outcome for a randomly chosen population unit, given that infection is present in the country (the system unit sensitivity).

Independence and clustering models are described for analysis. Under the independence model, overall system sensitivity of detection is derived directly from system unit sensitivity, as the probability that one or more of the independent units processed would have positive surveillance outcomes, given an infected country. Under the clustering model, animals (and disease) are assumed to cluster in groups, and surveillance system sensitivity is calculated taking this into account, by stepwise aggregation of sensitivity at each grouping level in the tree.

Surveillance processes give either complete or incomplete coverage of the population, and the sensitivity of a process with incomplete coverage must be adjusted for its representativeness of the population. This is achieved through calculation and use of a sensitivity ratio for the process; the ratio of its sensitivity to that of a truly representative surveillance process.

The surveillance process's sensitivity, P($\geq$1 positive unit | country infected), is the confidence level for the statistical test of the null hypothesis. If one has a prior estimate of P(country is free of disease), one can then use Bayesian inference to calculate a posterior estimate of this probability, given the negative surveillance results.

Case study – avian influenza in Denmark
A case study is presented, which analyses the Danish poultry diagnostic surveillance

process, applied to highly pathogenic avian influenza, which has never been recorded in Denmark. This surveillance process has complete coverage of the population. Existing data sets covering broiler batch mortality, dispensing of pharmaceuticals, veterinary consultation records and diagnostic laboratory records, are used to estimate values for key branch probabilities in the scenario tree model. These are supplemented where necessary with expert opinion. The tree is modelled stochastically to incorporate uncertainty of parameter estimates. The model's surveillance unit is a house of commercial birds, or a flock of backyard birds. Information on numbers of backyard flocks was obtained from surveillance activities conducted during the Danish 2002 Newcastle disease outbreak.

Nodes included in the tree are Industry sector; Farm infected; House infected; High mortality; Farmer seeks diagnosis; Veterinarian sends samples to laboratory; Laboratory looks for viruses; Laboratory finds AI virus. Isolation of HPAI virus represents a positive outcome of the surveillance process.

Given the comprehensive nature of the data available for broilers, the broiler industry section of the model is examined in greatest detail. It is analysed under assumptions of both independence (of units) and clustering, with only minor differences in results. Sensitivity of the diagnostic process applied to a single infected broiler house was 0.09 (mean value). Applying the process to one rotation of broilers through each house in the country (663 batches) using the independence model, the probability that one or more will give a positive surveillance outcome (under the null hypothesis with design herd prevalence of 1%) was 0.24. Using the clustering model it was 0.23. Clustering was therefore ignored in calculation of the tree for all industry sectors. The unit sensitivity of the whole diagnostic surveillance process (i.e. the probability that a randomly chosen chicken flock will test positive when the country is infected at a design herd prevalence of 1%) was 0.00002. When applied to all 51,000 chicken flocks in Denmark, the probability that one or more will test positive is 0.71.

Based on this result alone we do not have sufficient confidence to reject the null hypothesis that Denmark is infected with HPAI at a herd prevalence of 1% or more. However, continuous negative outcomes over time lead to accumulated confidence, and this can be encapsulated in a prior estimate of the probability that the country is free of HPAI (although the procedures for doing this rigorously are not defined here). If this is set at 90%, the calculated posterior estimate from this analysis of the diagnostic surveillance process is 97%.

Sensitivity analysis demonstrates that system sensitivity of detection is most sensitive to the probability that the laboratory will look for viruses in submitted samples, and the probability that they will find HPAI if it is there. Next in importance is the probability that a veterinary consultant will send samples from investigations of high mortality to the laboratory for testing. This has clear but predictable implications for HPAI surveillance in Denmark: improvements in confidence are most easily obtained by conducting virological investigations on more samples from cases of high mortality.

International EpiLab project 5 (IBR; Salman)

Targeted sampling for disease surveillance in Danish cattle at the example of infectious bovine rhinotracheitis (IBR)

Project period

1 July 2002 to 30 November 2002

Funding

Project leader

Mariann Chriél, DCF

Project partners

- Guest scientist: Mo Salman, USA
- DCF: Mariann Chriél,
- DVFA: Hanne M. Hansen
- DVI: Poul H. Jørgensen, Vibeke F. Jensen, Mette M. Larsen, René Bødker, Anne Bruun, Matthias Greiner
- Further co-workers: Bruce Wagner, USA

Executive summary

The International EpiLab in Denmark has initiated and funded a "disease freedom" research theme. The theme consists of three projects; each is related to a single livestock species: cattle, swine, and poultry. The aim of this theme is to develop and implement a standardized approach to assessing freedom from disease, using information from multiple non-survey-based sources. Three research teams were composed to address the specific objectives of this project. Each team was focusing on specific livestock species with the intention to combine the approaches and alternatives to address the national interest. This report is the first phase of the accomplishments of the bovine research team. Further consolidation and discussion of the other teams' accomplishments will be presented in the near future.

Data related to IBR surveillance were used in the first phase of this project. The evaluation of surveillance for IBR in Denmark was approached from two perspectives. First, the system was assessed relative to international requirements. Secondly, the system was examined for meeting the national needs for rapid detection of infected herds.

The assessment involved the determination of the sensitivity of the surveillance system for detecting infected herds. The model can be expanded to include surveillance for other diseases in near future. Furthermore, the assessment methods and the application can be expanded to include certification for disease freedom.

The probability of detecting at least one infected herd in the country if the herd prevalence for IBR is greater than or equal to 2 per 1000 is calculated for the international requirement using specific assumptions. The detailed methods are described in the project report. The numbers of herds that were used in this calculation were for beef herds 24355 (year 2000) and 25233 (year 2001), while the numbers of dairy herds were 13034 (year 2000) and 12003 (year 2001). The results indicate that the existing sampling scheme for dairy and beef populations is adequate to satisfy the international requirements. National disease detection needs (i.e. early detection of the infection) exceed the international requirements and require more intensive sampling. Therefore, no further sampling scheme options were explored for the international requirement.

The surveillance to meet national requirements was evaluated under current sampling conditions and three alternative scenarios. The national objective for detecting a single infected herd as quickly as possible requires a much more intensive approach. The current implementation of the system can identify dairy herds within a reasonable period of time with desired accuracy largely because of the test characteristics and the number of bulk tank milk samples. The system is less likely to detect infected beef herds since surveillance in those herds depends solely on slaughter serological testing.

The system can be adjusted to improve the efficiency of the surveillance. Modelling demonstrated that the efficiency of surveillance in dairy herds, which depends on bulk tank milk testing, would not be substantially decreased if the slaughter surveillance component was dropped. Beef surveillance can only be improved by increasing the number of herds that are tested. Modelling showed that targeted sampling during the critical winter season could increase the likelihood of detecting disease.

Contents