Handbook on Import Risk Analysis for Animals and Animal Products

Volume 2
1st Edition, 2004
Quantitative risk assessment

Published by
The World Organisation for Animal Health
(OIE)

World organisation for animal health
12, rue de Prony, 75017 Paris, France
Tel.: 33-(0)1 44 15 18 88
Fax: 33-(0)1 42 67 09 87
http://www.oie.int

Volume I, First Edition (2004), ISBN : 92-9044-629-3
Second Edition (2010), ISBN:978-92-9044-807-5
Reprinted (2012):

Volume 2, First Edition (2004): ISBN 92-9044-626-9 (2004)
Reprinted (2010 & 2012): ISBN: 978-92-9044-626-2

Cover: © P. Blandín, OIE

Contents

Volume 2. Quantitative risk assessment

Author

Noel Murray
Ministry of Agriculture and Forestry
New Zealand

Editorial assistance

Stuart C. MacDiarmid
Ministry of Agriculture and Forestry
New Zealand
Marion Wooldridge
Veterinary Laboratories Agency (Weybridge)
United Kingdom
Bruce Gummow
Faculty of Veterinary Science
University of Pretoria
South Africa
Randall S. Morley
Canadian Food Inspection Agency
Canada
Stephen E. Weber
Centers for Epidemiology and Animal Health
Fort Collins
United States of America
Armando Giovannini
Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise,
Italy
David Wilson
Head
International Trade Department
OIE, France

Acknowledgments

This second volume of the OIE ***Handbook on Import Risk Analysis: Animals and Animal Products*** draws heavily on David Vose's ***Risk Analysis: A Quantitative Guide*** (John Wiley & Sons, Chichester. 2000). Vose's book is an indispensable reference text for both the student and practitioner of animal health risk analysis. The text also draws on the book ***Import Risk Analysis: Animals and Animal Products*** (2002) by Noel Murray, published by the Biosecurity Authority, Ministry of Agriculture and Forestry, New Zealand.

Various people have offered critical comment on all or part of the modified text. In particular the Author and Editors wish to acknowledge:

David Vose
David Vose Consultancy
www.risk-modelling.com
Michael Roberts
Neil Cox
AgResearch Ltd
New Zealand
Lisa Gallagher
Tracey England
Louise Kelly
Rowena Jones
Veterinary Laboratories Agency (Weybridge)
United Kingdom
Sanping Chen
Carleton Quantitative Research
Ottawa
Canada

Anna Maria Conte
Istituto Zooprofilattico Sperimentale
dell'Abruzzo e del Molise,
Italy
Ziad A. Malaeb
Centers for Epidemiology and Animal Health
Fort Collins
United States of America

The Author and Editors thank Professor Vincenzo Caporale, Director of the **OIE Collaborating Centre for Epidemiology and Organization of Veterinary Services in Developing Countries**, Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise, Teramo, Italy, for hosting their meetings and providing secretarial services during the drafting of the text.

Chapter 1
An introduction to quantitative risk analysis[1]

Introduction

In Volume 1 of this Handbook we stated that no single method of import risk assessment has proven applicable in all situations, and different methods may be appropriate in different circumstances[2]. In qualitative assessments, the likelihood the release and subsequent exposure to a hazard and the magnitude of the resulting consequences are expressed using non-numerical terms such as high, medium, low or negligible, and the qualitative approach has so far proved suitable for the majority of import risk assessments. However, in some circumstances it may be desirable to undertake a quantitative analysis, for example, to gain further insights into a particular problem, to identify critical steps or to compare sanitary measures.

The terms 'parameter', 'variable', 'input' and are often used interchangeably in quantitative risk assessments. In this Handbook, these terms are used as follows:

– **Parameter**

In experimental statistics the term parameter represents a numerical descriptive measure that characterises a population, for example the population mean (μ), the population standard deviation (σ) and the binomial proportion (p). In spread sheet computer software, it is often used to represent the arguments of mathematical, statistical or probability distribution functions such as the values required to define the shape of a Beta distribution or the mean and standard deviation of a normal distribution.

– **Variable**

A variable is any characteristic that has a different value for different subjects or objects. If it can take on a different value as a result of a random process it is called a random variable. It can either be discrete, where it can only take on a limited number of values, or continuous, where it can take on any value within a given range. Examples of discrete variables include the number of infected animals, the number of test positive animals or the number of piglets in a litter, while examples of continuous variables include bodyweight or blood copper levels.

– **Inputs**

An input is any information that is fed into a model. As a result parameters and variables, together with data and distributions, can be considered as inputs as they provide information that is used in a quantitative risk assessment model.

– **Model**

A model is a simplified representation of the real world. Most models are symbolic because symbols represent properties of the system. In this handbook, a 'model' is a representation of an importation scenario in graphical or mathematical form where

---

1 The general reference for this chapter is Vose D. (2000). – Risk Analysis, A Quantitative Guide. John Wiley & Sons Chichester.

2 *Terrestrial Animal Health Code*, Article 1.3.1.1

equations are used to simulate the biological processes under study and the impact of risk management options.

– **Quantitative risk assessment**

A quantitative risk assessment is a mathematical model where the inputs and outputs are expressed numerically. In its simplest form, commonly referred to as a deterministic or point estimate analysis, both the inputs and outputs are expressed as single numbers or point values. These may represent a 'best guess', the 'average' or 'expected case' or perhaps the 'worst case'. When one wants to determine the impact of one or more of the input values on the output, one simply substitutes a new value into the model. This is effectively a 'what if', or scenario, analysis. For simple models with few inputs, this type of analysis can be easily undertaken using a calculator.

For more complex models, or in situations where one has more data to work with, probabilistic risk assessments are preferable. In these, inputs are described as probability distributions and a computer is essential for constructing the risk assessment model.

Deterministic (point estimate) risk assessment

Quantification of risk begins with considering an experiment, or trial with only two possible outcomes: success or failure. The trial may be repeated a number of times. For example, a trial may be a single embryo transfer from an infected animal to a susceptible recipient. A 'success' in this case would be where the infection is transmitted while a 'failure' would be a transfer where infection is not transmitted. If we observe no successes after ten transfers (trials) we may begin to suspect that the probability of transmitting infection by embryo transfer is low. As more transfers are undertaken without transmitting infection, the more confident we become that transmission is unlikely. This is shown in Table I, where confidence intervals[3] have been determined by consulting the statistical tables presented in Appendix 1.

Table I
Probability of transmitting infection following embryo transfer from a viraemic donor

Number of transfers (*n*)	Number of infected recipients I	Probability of transmitting infection $p_t = \left(\frac{r}{N} \times 100\right)$	Lower 95% confidence limit	Upper 95% confidence limit
10	0	0.00	0.00	30.85
20	0	0.00	0.00	16.84
30	0	0.00	0.00	11.57
40	0	0.00	0.00	8.81
100	0	0.00	0.00	3.62
1,000	0	0.00	0.00	0.37

If 100 experimental transfers were undertaken without transmitting infection, we could reasonably conclude, using the upper 95th percent confidence interval, that the probability

[3] A confidence interval is a range of numbers believed to include an unknown quantity with a specified level of confidence. For example, if we weighed 10 sheep we could calculate their average weight and the associated confidence intervals. If the average weight is 50 kg and the 95% confidence interval is ± 2.5 kg, this indicates that we could be 95% confident that the true average weight of all sheep in the flock lies somewhere within the interval bounded by 47.5 kg and 52.5 kg

of transmitting infection for each embryo transferred from an infected donor is 'at worst' 3.62%.

If we plan on undertaking an embryo transfer program we might like to estimate the probability that at least one recipient becomes infected or, alternatively, the average number of infected recipients we could expect.

To calculate the probability that at least one recipient becomes infected we proceed as follows:

– the probability of transmitting infection (a success) is p_t the probability of not transmitting infection (a failure) is 1- p_t

– the probability that none of the recipients become infected is $(1- p_t)^e$, where e refers to the number of recipients (trials)

– so, the probability that at least one recipient becomes infected is $1-(1- p_t)^e$

– the probability is expressed in mathematical notation as $P(x \geq 1)$, where P refers to probability and x refers to the outcome, that is, an infected recipient

– and the final equation is then written as:

$$P(x \geq 1) = 1-(1- p_t)^e \qquad \text{Equation 1}$$

To calculate the expected number of infected recipients we multiply the probability of transmitting infection p_t, by the number of recipients e:

$$\text{expected number of infected recipients} = p_t \times e \qquad \text{Equation 2}$$

If we assume a situation where the probability of transmission equals 3.62% (n=100) and the number of embryos transferred equals 30, we could determine the probability that at least one recipient becomes infected (Table II). For simplicity, we will assume that each recipient is implanted with only one embryo and that each donor produces a single transferable embryo. As a result the number of recipients equals 30.

$P(x \geq 1) = 1-(1-0.0362)^{30} = 0.6692 = 66.92\%$
expected number of infected recipients = $0.0362 \times 30 = 1.086$

This scenario is essentially a 'worst case' as we have assumed that all the donors are infected. If we had some information on the prevalence of disease among the donors we could incorporate this into the model. Suppose a survey had been recently undertaken in a donor flock of sheep and 5 I animals out of 100 (n) tested were found to be infected. By consulting the statistical tables in Appendix 4 we could estimate that the true disease prevalence, with a 95% level of confidence, is likely to be between 1.64% (lower 95% confidence limit) and 11.28% (upper 95% upper confidence limit) with an expected value of 5%. We could include these estimates of disease prevalence in the model to determine three possible outcomes (Table II) using the following formulae:

$$P(x \geq 1) = 1 - \left(1 - p \times p_t\right)^e \qquad \text{Equation 3}$$

$$\textit{expected number of infected recipients} = p \times p_t \times e \qquad \text{Equation 4}$$

where: p = prevalence,
p_t = probability of transmitting infection and
e = number of recipients.

Table II
Probability of transmitting infection to at least one recipient and the expected number of infected recipients if thirty embryos are transferred

Scenario	p = prevalence in the flock of origin	p_t = probability of transmitting infection via embryo transfer	Probability ≥ 1 recipient infected (Equation 3)	Expected number of infected recipients (Equation 4)
Minimum	1.64% (lower 95% CL*)	3.62% (upper 95% CL)	1.77%	0.017 (17 out of every 1,000)
Most likely	5% (expected value)		5.28%	0.054 (54 out of every 1,000)
Worst case	11.28% (upper 95% CL)		11.55%	0.122 (122 out of every 1,000)

* CL = confidence limit

After considering the probabilities that one or more recipients would become infected, we might consider that the likelihood is too high and that some risk management measure is desirable. So, we might then decide to test the donors and discard any that are positive. If we test a potential donor, chosen at random, we could calculate the probability that it is infected D^+, given that it is test negative T^-. This is a conditional probability, which is expressed as $P(D^+|T^-)$. For a perfect test, this probability would be zero. However, since all tests are imperfect (with a sensitivity[4] of less than 1), we can expect that the test will fail to detect some infected animals. In addition, some uninfected animals will be incorrectly classified as positive, since the specificity[5] will also be less than 1. In these circumstances we calculate the $P(D^+|T^-)$, by firstly determining the predictive value of a negative test *NPV* as outlined in Chapter 4 and then calculate its complementary probability (1-*NPV*). This represents the prevalence of infection within the group of donor animals we accept. That is, the prevalence of infection amongst the test negative animals as a result of discarding test positive animals. From Equation 40 in Chapter 4 the *NPV* is calculated as:

$$NPV = P(D^-|T^-) = \frac{Sp(1-p)}{p(1-Se)+(1-p)Sp} \qquad \text{Equation 5}$$

where: p = the prevalence of infection in the flock of sheep
Se = test sensitivity
Sp = test specificity

So the prevalence of infection within the test negative group is calculated as:

$$P(D^+|T^-) = 1 - NPV \qquad \text{Equation 6}$$

If we use a test with a sensitivity of 90% and specificity of 99% and reject any positive animals, we could calculate the probability of infection for a test negative animal by substituting these values into Equation 6 (Table III):

[4] Sensitivity of a test is its ability to correctly classify an infected animal as test positive. It is calculated as the proportion of infected animals that yield a positive test result $P(T^+|D^+)$

[5] Specificity of a test is its ability to correctly classify an uninfected animal as test negative. It is calculated as the proportion of uninfected animals that yield a negative test result $P(T^-|D^-)$

Table III
Prevalence of infection among test negative donors

Scenario	p = prevalence in the flock of origin	Se = test sensitivity	Sp = test specificity	Prevalence among test negative donors (Equation 6)
Minimum	1.64% (lower 95% CL*)	90%	99%	0.17%
Most likely	5% (expected value)			0.53%
Worst case	11.28% (upper 95% CL)			1.27%

* CL = confidence limit

Since 1-*NPV* is the prevalence of infection within the test negative group, we can replace '*p*' in Equation 3 with '1-*NPV*' to determine the probability of transmitting infection to at least one recipient:

$$P(R^+ \geq 1) = 1 - (1 - (1 - NPV) \times p_t)^e \qquad \text{Equation 7}$$

where: R^+ = infected recipient
and the expected number of infected recipients:

$$(1 - NPV) \times p_t \times e \qquad \text{Equation 8}$$

The results of these calculations are shown in Table IV.

Table IV
Probability of transmitting infection to at least one recipient and the expected number of infected recipients if thirty embryos are transferred

Scenario	(1-NPV) = prevalence in the group of test negative donors (from Table III)	P_t = probability of transmitting infection via ET	Probability ≥ 1 recipient infected (Equation 7)	Expected number of infected recipients (Equation 8)
Minimum	0.17%	3.62% (upper 95% CL*)	0.18%	0.002 (2 out of every 1,000)
Most likely	0.53%		0.57%	0.006 (6 out of every 1,000)
Worst case	1.27%		1.37%	0.014 (14 out of every 1,000)

* CL = confidence limit

So, by making use of a statistical table and a calculator, we have been able to undertake a simple deterministic or point estimate analysis that has given us a very good idea of the risks we face. We could go on adding to this model, for example by including an estimate of the probability that a randomly chosen flock is actually infected and the effect of quarantining and testing recipients to screen out positive animals.

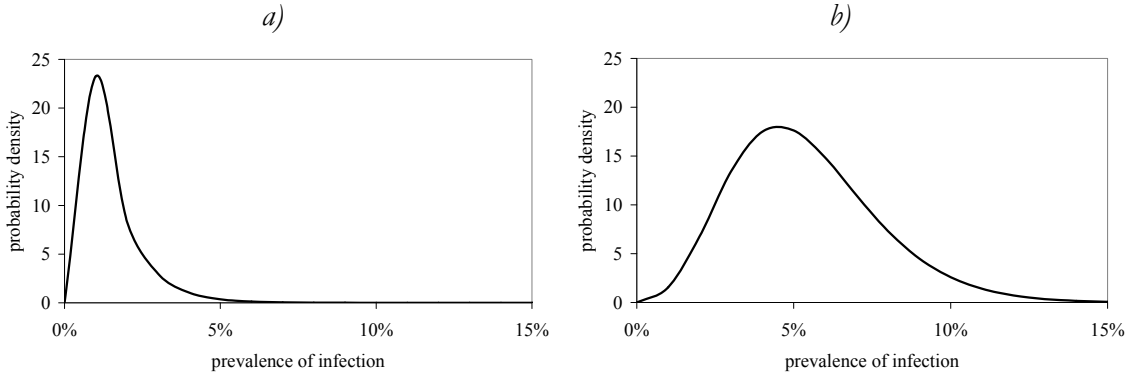
Probabilistic risk assessment (Monte Carlo simulation)

The embryo transfer model under discussion could be refined further. Just as we have estimated the probability of transmitting infection by embryo transfer, and the prevalence of infection within the flock of origin, we could include confidence intervals of the estimates of sensitivity, specificity and the probability that the flock of origin is infected.

However, as the number of such variables[6] increases there will be a rapid escalation in the number of potential combinations or 'what if' scenarios. For example, if we had four variables, each with a mean and upper and lower 95th percent confidence limits, we would have 3^4, or 81 possible scenarios. Such an approach has significant drawbacks. It can rapidly become impractical to analyse the results. In addition there is no weighting for each of the values chosen. For example, our 'best guess' might be far more likely to happen than the 'worst case'.

If we had information about the range of values and the likelihood of each value, we could assign a probability distribution to each variable, which we can now describe as random variables as they can take on a different value as a result of a random process. In our embryo transfer example we could use the Beta distribution (Chapter 4) to define a probability distribution for each input variable (Fig. 1). Such a model is called a stochastic model and we can calculate the combined impact of the variation in each of the model's input distributions to determine a probability distribution of the possible model outcomes. The simplest way to do this is to perform a simulation. This involves randomly sampling values from each distribution and combining the values generated, according to the mathematical logic of the model, to produce a result for that particular scenario. This process is repeated many times and the results from each scenario, which are also known as iterations, trials or realisations, are combined to produce a probability distribution of possible model outcomes.

Throughout this text, probability distributions will be described in terms of functions used in the risk assessment computer software @RISK[7] and the spreadsheet software Microsoft Excel[8]. For example, the notation Binomial( ) is an @RISK function while BINOMDIST( ) is a Microsoft Excel function and is distinguished by capital letters.



a) a Beta distribution of the probability of transmitting infection by ET if 100 transfers from infected donors to susceptible recipients were undertaken without transmitting infection: Beta (0+1,100-0+1)

b) a Beta distribution of the prevalence of infection if 5 infected animals were detected in a sample of 100: Beta(5+1,100-5+1)

Figure 1
An example of two probability distributions that could be assigned to the input variables in the embryo transfer quantitative risk assessment example

An ascending cumulative frequency plot (Fig. 2a) is often used to display the results of a simulation. It shows the probability of being equal to or less than a certain value.

---

6 A variable is any characteristic that has a different value for different subjects or objects

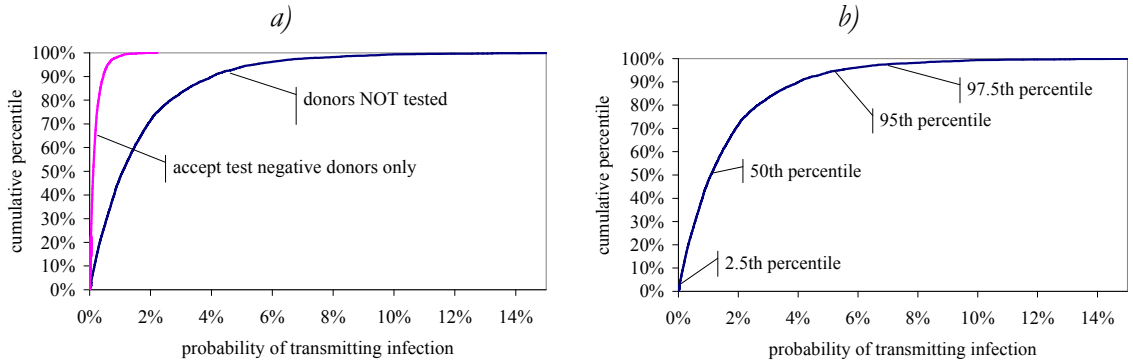
7 Palisade Corporation, Newfield, New York

8 Microsoft Inc., Redmond, Washington

For example, we could report the results as follows, by reading from the 95th cumulative percentile:

In 95% of iterations, the probability of transmitting infection to at least one recipient is equal to or less than 5.4% if test positive donors are not rejected and less than 0.61% if test positive donors are rejected.

Alternatively, we might choose to report the median result (50th percentile) and the associated 95% confidence intervals. In the case of testing and rejecting positive donors the median is 0.12% with lower and upper 95% confidence limits of 0.004% and 0.8% respectively. It is important to note that the 95th percentile does not represent the upper 95% confidence limit. The upper and lower 95% confidence limits about the 50th percentile are represented by the 97.5th and 2.5th percentiles respectively (Fig. 2b). The area under the curve embraced by these percentiles is equal to 95% of the total area, which is the relevant area for the 95% confidence interval.



a) with and without testing donors

b) percentiles for the probability without testing

Figure 2
Ascending cumulative frequency plots of the probability of transmitting infection to at least one recipient if thirty embryos are transferred

Sampling values from a probability distribution

Sampling values from probability distributions is most commonly undertaken by either Monte Carlo or Latin hypercube sampling. The Monte Carlo method is based on simple random sampling from the entire distribution, which represents the sampling frame for each iteration. It is sampling with replacement, as it is possible for the same values to be selected more than once. Latin hypercube sampling, on the other hand, involves stratified sampling without replacement. The range of the distribution is divided up into a number of intervals, equal to the number of iterations to be performed and a simple random sample is then chosen from within each interval. Each interval is only selected once during a simulation. As a result, Latin hypercube sampling ensures that values from the entire range of the distribution will be sampled proportional to the probability density of the distribution. Fewer samples are usually required to reproduce the probability distribution so it is more efficient than Monte Carlo sampling for the same number of iterations. It is generally the preferred method of numerical simulation since fewer iterations are required for a particular level of accuracy.

Differentiating variability and uncertainty

The way in which variability and uncertainty have been described by risk analysts has led to a degree of confusion. To understand what is meant by these terms, it is important to

appreciate that risk assessment is essentially a tool aimed at predicting the probability of an outcome of a particular action or actions. For example, we might want to predict the likely height of a person chosen at random. We know from our own observations that there is a great deal of natural variation among individuals in the population. While we might have a good 'feel' for its range and what the average might be, it is only by measuring several people that we can begin to make some accurate predictions about the heights of people in the general population. As more measurements are collected, more knowledge is acquired. We can begin to describe the variation in people's heights with increasing certainty, enabling us to be more and more confident in our predictions. If we measured everybody in the population, we would have a perfect understanding and we would be able to state exactly what the population parameters, such as the average height and standard deviation (a measure of the amount of variation that exits), were. Obviously, this is impractical and we need to achieve a balance between acquiring perfect knowledge and obtaining reasonable estimates upon which we can base our predictions with a reasonable level of confidence.

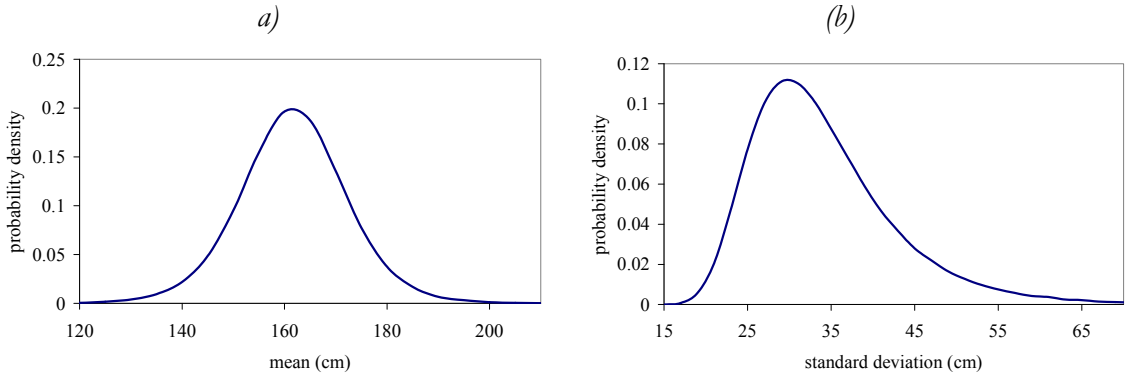
Table V
A hypothetical example of the height of ten adults chosen at random and the associated statistics

Height in centimetres (x_i)									
152.3	118.4	158.5	168.8	163.4	162.9	180.7	99.5	188.9	198.5
Sample average = $(\bar{x}) = \frac{\sum_{i=1}^{n} x_i}{n}$ = 159.2									
Sample standard deviation = $s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n}}$ = 30.3									
Standard error of the mean = $s_{\bar{x}} = \frac{s}{\sqrt{n}}$ = 9.6									
t value with (n –1) degrees of freedom = 2.262 (from the student's t distribution)									
Confidence interval = $\pm t \times s_{\bar{x}} = \pm 2.262 \times 9.6 = \pm 21.7$									
Upper 95% confidence limit = $\bar{x} + t \times s_{\bar{x}}$ = 159.2 + 2.262 × 9.6 = 180.9									
Lower 95% confidence limit = $\bar{x} - t \times s_{\bar{x}} = 159.2 - 2.262 \times 9.6 = 137.5$									

Note: sample statistics are represented by $\bar{x}$ (average) and s (standard deviation) while the corresponding population parameters are represented by μ and σ

If we choose ten adults at random and measure them, we can calculate their average height and standard deviation. These are actually *sample statistics*, rather than *population* parameters because we have collected data from a subset of the population only (Table V). If we deduce, from previous observations, that height is a normally distributed variable, we could use these sample statistics in a normal distribution function (Chapter 3) to enable us to describe the distribution of height in the general population and make some predictions. However, because of the small sample size we might be concerned that these sample statistics do not adequately reflect the population parameters. That is, the population parameters are uncertain. As shown in Figure 3 we could develop a sampling distribution for both the mean and standard deviation (see Chapter 6 for details).

A sampling distribution enables us to capture the uncertainty associated with the estimate of a population parameter based on the data we have collected. For example, we can calculate confidence intervals, which allow us to determine how confident we can be that the true population parameter lies within so many units either side of the corresponding sample statistic. Confidence intervals are determined from the area under the curve surrounding the average value of the distribution. The 95% confidence interval, for example, corresponds to ± 47.5% of the area under the curve either side of the average value. In our case the 95% confidence interval is ± 21.7 cm about the sample average of 159.2 cm (Table V). This indicates that we could be 95% confident that the true population average lies somewhere within the interval bounded by 137.5 cm to 180.9 cm.



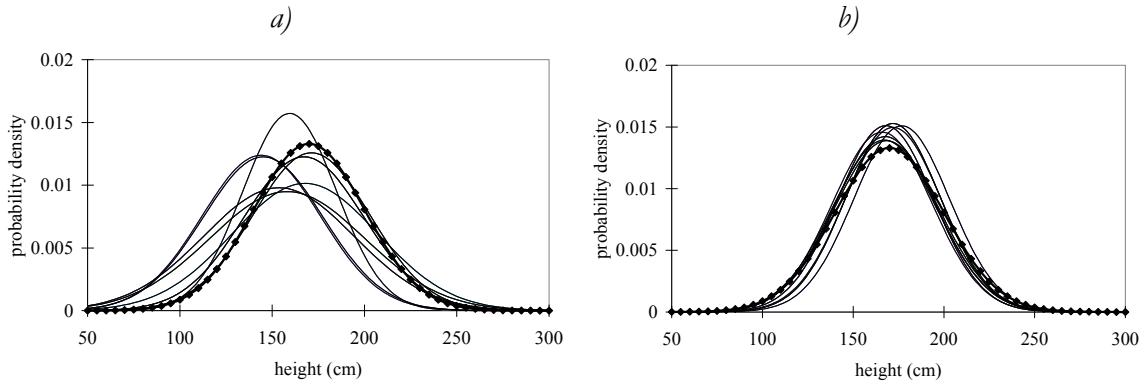
a) hypothetical sampling distributions of the mean

b) standard deviation

Figure 3
Hypothetical sampling distributions of the mean and standard deviation based on the data in Table V

If we randomly select a value from each sample distribution of the mean and standard deviation in Figure 3 and insert them into a normal distribution function, plot its graph and repeat this exercise a number of times, we could build up a picture of possible distributions of height (Fig. 4a). Each of these distributions separately represents a *first order distribution*, while together they form a *second order distribution*. These distributions, which enable variability and uncertainty to be modelled separately, are explored in more detail in Chapter 7. The thick black line in Figure 4a represents the hypothetical situation where we have perfect knowledge. It can be seen that there is a certain degree of uncertainty associated with the small sample size, because there are a number of different possible distributions.

What happens if we increase the sample size to 100 adults? By repeating the exercise just outlined, we can see from Figure 4b, that by collecting some additional information we have reduced the uncertainty considerably as the range of possible distributions is very close to the distribution representing perfect knowledge. We appear to have achieved a good balance between acquiring perfect knowledge and obtaining reasonable estimates.



a) sample size = 10

b) sample size = 100

Figure 4
A hypothetical normal distribution of the height of adults in Great Britain.
The thick line represents perfect knowledge where the average height of all adults is 170 cm with a standard deviation of 30 cm. Each thin line represents one possible distribution of height

Uncertainty, then, may be thought of as a measure of the incompleteness of one's knowledge or information about an unknown quantity. It is important to remember that even with perfect knowledge variability still exists.

As was observed in Volume 1, even though quantitative risk assessments involves numbers, they are not necessarily more objective, nor are the results necessarily more 'precise' than with qualitative assessments. Choosing an appropriate model structure, which pathways to include or exclude, the level of aggregation or disaggregation, the actual values used for each input variable and the type of distribution applied to them, all involve a degree of subjectivity. Further, because data are often lacking, models may need to incorporate expert opinion, which by its very nature is subjective.

The means by which this inherent subjectivity is countered in a good risk assessment is by ensuring that it is *transparent*. All the information, data, assumptions, uncertainties, methods and results must be comprehensively documented and the discussion and conclusions supported by a reasoned and logical discussion. The assessment should be fully referenced and subjected to peer review.

Chapter 2
Probability and probability distributions

Defining probability

Probability describes the likelihood of something happening and can be expressed using words such as low, medium or high, or as a number between 0 and 1 or as a percentage between 0% and 100%. Numerically, there are several ways of defining probability including:

Classical probability

The probability of a particular event is the number of ways the event can occur divided by the total number of possible outcomes, for example, the probability of event A, written as P(A) is:

$$P(A) = \frac{the\ \ number\ \ of\ \ ways\ \ event\ \ A\ \ can\ \ occur}{the\ \ total\ \ possible\ \ number\ \ of\ \ outcomes}$$

If we had a flock of 100 lambs consisting of 65 ewe lambs and 35 ram lambs we could determine the probability that a lamb chosen at random will be a ram lamb as follows:

- the event of interest (A) is a ram lamb
- there are 35 ram lambs so, the number of ways event A can occur is 35
- the total possible number of outcomes is 100 since there is a total of 100 lambs in the flock.

$$P(A) = \frac{35}{100} = 0.35$$

Empirical probability (relative frequency)

The number of events of interest *x*, that occur in a number of identical and repeatable trials *n*, is expressed as a ratio (fraction or proportion) of the total number of events that occurred. Under this definition probability is a measurable property of the physical world and can never actually be observed. It is expressed as the limit of the ratio:

$$\frac{number\ \ of\ \ events}{number\ \ of\ \ trials} = \frac{x}{n}$$

as *n* approaches infinity this ratio would converge to:

$$p = \lim_{n \longrightarrow \infty} \frac{x}{n}$$

Subjective probability

Under a subjective, or Bayesian, definition of probability, an individual's state of knowledge or degree of belief about the occurrence of an event is captured. For example, a farmer may estimate that a particular cow has a 60% chance of calving tonight. As a result, probability is a function of the estimator's knowledge of the event itself. Furthermore, it may change over time as new information becomes available.

The rules of probability

Independence

The probability of event A can be written as $P(A)$. If two events, A and B, are independent then the occurrence of event A has no effect on the occurrence of event B and vice versa. In these circumstances the probability of event A occurring at the same time as, or immediately followed by, event B is the product of these two probabilities, which is written as $P(A \cap B)$:

$$P(A \cap B) = p(A) \times p(B)$$

This concept can be extended to several events, for example the probability of obtaining a head H, followed by a tail T, then a head H, when tossing a coin:

$$P(H \cap T \cap H) = P(H) \times P(T) \times P(H)$$

Suppose we have a very large herd of cows with a disease prevalence of 30%. The probability that a cow chosen at random will be infected can be expressed as $P(D^+) = 0.3$. If we purchased four cows at random from this herd we might want to determine the probability that all four cows are infected. We will assume that the infection is static within the herd. That is, there is no transmission occurring. Since the disease status of any cow is independent of the status of any other, the probability that all four cows are infected is:

$$P(D^+ \cap D^+ \cap D^+ \cap D^+) = P(D^+) \times P(D^+) \times P(D^+) \times P(D^+) = 0.3 \times 0.3 \times 0.3 \times 0.3 = 0.0081$$

We can extend this calculation as far as we like, for example, to the random selection of 10 cows. However if we express the calculation as shown above it would become rather tedious. A simpler way of writing the probability that all ten cows are infected is $P(D^+)^{10}$ which can be generalised to n cows as $P(D^+)^n$, provided n is much smaller than the herd size.

If we wanted to calculate the probability that there is at least one infected cow among a group of size n, selected at random from the herd, then we proceed as follows:

- the probability that all n cows are infected is $P(D^+)^n$
- the probability that none of the n cows is infected is $(1 - P(D^+))^n$
- the probability that at least one of the n cows is infected is $1 - (1 - P(D^+))^n$

Conditional probability

The probability that B will occur given A has already occurred is a *conditional probability*, which is written as $P(B \mid A)$. If A and B are independent then the occurrence of B is not influenced by the occurrence of A. As a result $P(B \mid A) = P(B)$ and similarly $P(A \mid B) = P(A)$ and the probability that event A is followed by event B is simply the probability of both events occurring, $P(A \cap B) = P(A) \times P(B)$. If, however, the occurrence of B is dependent on A having already occurred, then the probability that event A is followed by event B is:

$$P(A \cap B) = P(A) \times P(B \mid A).$$

Continuing with the cattle example from the previous section, suppose we test one of the cows and want to know the probability that it is test positive (T^+) given that it is infected (D^+). This is a conditional probability that the cow is test positive given it is infected, and is written as $P(T^+ \mid D^+)$. It is worth noting that this particular probability is commonly referred

to as *test sensitivity*. To determine the probability that the cow is both infected and test positive, we need to multiply these two probabilities together. Assuming that the probability that a cow is infected is 0.3 and that we use a test with a sensitivity of 0.9 then:

$$P(D^+ \cap T^+) = P(D^+) \times P(T^+ \mid D^+) = 0.3 \times 0.9 = 0.27$$

To calculate the probability that there is at least one test positive and infected cow amongst a group of size *n* selected at random:

$$P((D^+ \cap T^+) \geq 1) = 1 - (1 - P(D^+) \times P(T^+ \mid D^+))^n$$

Mutually exclusive events

If two or more independent events cannot happen together they are said to be mutually exclusive. For example, if we have a flock of sheep consisting of Merinos (M), Suffolks (S) and Romneys (R) we could represent their respective probabilities of occurrence using a Venn diagram (Fig. 5). The probability of selecting either a Merino or a Suffolk, written as $P(M \cup S)$ is:

$$P(M \cup S) = P(M) + P(S) = 0.5 + 0.15 = 0.65$$

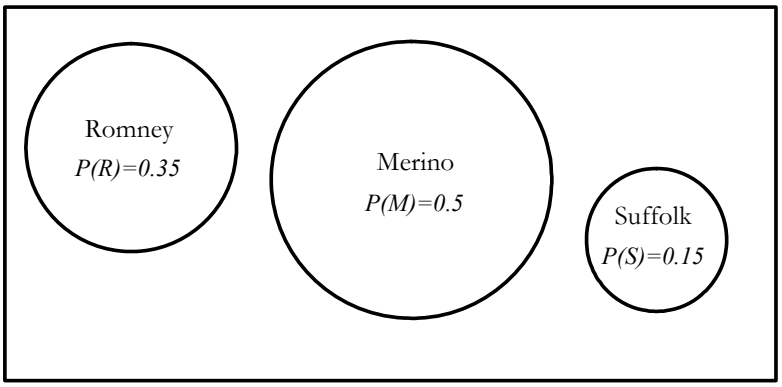


Figure 5
A Venn diagram of a flock of sheep consisting of three breeds. Each breed represents a mutually exclusive event

A particular subset of mutually exclusive events is where the events are complementary, that is, either one or the other occurs. In such cases $P(A \cup B) = P(A) + P(B) = 1$ so that $P(A) = 1 - P(B)$. An example of a complementary event is pregnancy. An animal is either pregnant or it is not.

Independent events that can occur simultaneously

Suppose that there has been an outbreak of both foot rot and lice in our flock of sheep, each of which can independently affect any of the three breeds of sheep (Fig. 6). This time we want to determine the probability that a sheep chosen at random will have either foot rot *or* lice. Since some sheep may have both diseases, we need to adjust the estimate by subtracting the probability that some sheep have both foot rot *and* lice:

$$P(L \cup F) = P(L) + P(F) - P(L \cap F) = 0.25 + 0.6 - 0.15 = 0.7$$

where $P(L \cap F) = P(L) \times P(F)$, since L and F are independent.

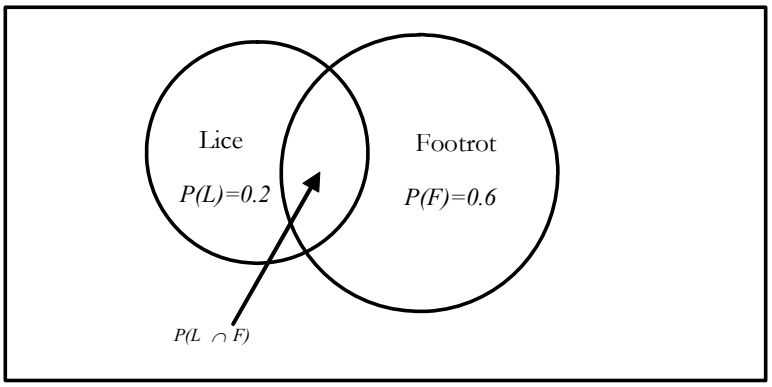


Figure 6
A Venn diagram of the probability that sheep within a flock are affected by footrot and/or lice
An assumption is made that the occurrence of footrot does not influence the occurrence of the lice and vice versa. That is, the events are independent

Probability distributions

Random variables

A variable is any characteristic that has a different value for different subjects or objects. If it can take on a different value as a result of a random process it is called a random variable. It can either be discrete, where it can only take on a limited number of values, or continuous, where it can take on any value within a given range. Examples of discrete variables include the number of infected animals, the number of test positive animals or the number of piglets in a litter, while examples of continuous variables include bodyweight or blood copper levels.

Discrete distributions

If a random variable can only take on a limited number of values it is classified as a discrete variable and its corresponding distribution will also be discrete. Suppose we collect some data on the size of litters born to sows over a one-year period in a particular pig herd. We could summarise the data as has been done in Table VI. In this hypothetical example there are 500 observations. From this we could determine the relative frequency of the different litter sizes and plot these results on a bar graph (Fig. 7). Since litter size is a discrete variable the resulting distribution is a discrete distribution and the relative frequency is the actual probability of occurrence. This probability is referred to as the *probability mass* and all the individual probabilities must add up to one.

If we were interested in determining the probability that a sow has a litter less than or equal to a certain value we need to calculate the cumulative probability (Table VI), for example, the probability that a sow has a litter size less than or equal to 3. We do this by adding up the respective probabilities for each value up to and including 3, that is 0.02+0.08+0.15, to determine the cumulative probability of 0.25.

We could also calculate the expected value or mean of the distribution by multiplying litter size by its respective probability and adding all the results together. This is essentially a weighted average.

Table VI
Some hypothetical observations on the size of litters born to sows

Litter size x_i	Number of litters l_i	Probability $p(x_i) = \frac{l_i}{n}$	Cumulative probability $P(X \le x_i) = \sum_{i=1}^{n} p(x_i)$
1	10	0.02	0.02
2	40	0.08	0.10
3	75	0.15	0.25
4	175	0.35	0.60
5	125	0.25	0.85
6	50	0.10	0.95
7	25	0.05	1.00
Total	n **= 500**	**1.00**	–

Mean (expected value) $= \sum_{i=1}^{n} x_i \times p(x_i) = 4.23$

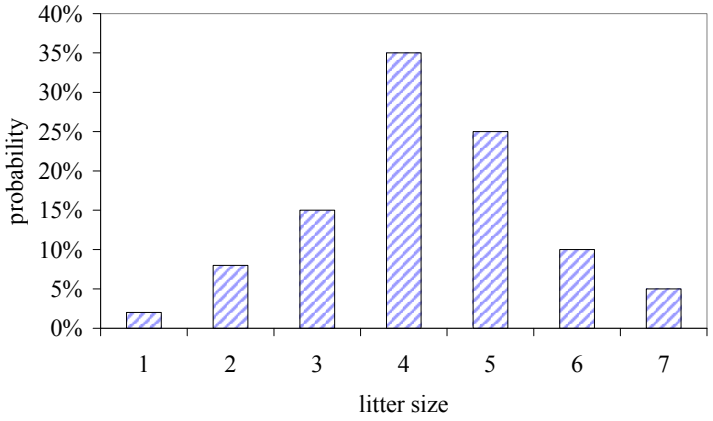


Figure 7
A hypothetical discrete probability distribution of litter size based on the data in Table VI

Continuous distributions

If a random variable can take on any value within a given range it is classified as a continuous variable and its corresponding distribution will also be continuous. For example, the weight of cattle is a continuous variable since the weight of a particular cow can be measured to the nearest kilogram, gram, milligram and so on. We can continue this process, dividing the scale into smaller and smaller units. In fact, it is infinitely divisible.

Suppose we weigh 100 cattle and record each one's weight to the nearest kilogram (Table VII). Since weight is on a continuous scale we cannot divide the data into discrete values to develop a distribution. Instead, we need to divide the weights into convenient, non-overlapping classes with no gaps between each class (Table VIII). The resulting graph is a special case of a bar chart called a histogram (Fig. 8). In this case the relative frequency is the probability that the weight of a cow falls within a particular class interval. For example, from Table VIII the probability that a cow weighs between 460 kg and 480 kg is 0.19. Alternatively, we might be interested in knowing the probability that the weight of a cow is less than or equal to a certain class interval. In this case we need to determine the cumulative probability, for example the probability that a cow weighs less than or equal to the 460 kg < 480 kg interval. We do this by adding up the respective probabilities in

Table VIII for each class up to and including the 460 kg < 480 kg interval, that is 0.01+0.06+0.09+0.19, to determine the cumulative probability of 0.35.

Table VII
Bodyweights (kg) of 100 cattle (hypothetical example)

411	423	425	428	433	435	437	444	444	445
452	456	456	456	457	459	460	462	463	463
464	464	464	468	470	470	470	472	472	475
478	478	479	479	479	482	484	485	487	487
488	488	489	489	491	491	493	495	495	496
496	497	500	500	500	501	502	503	503	505
505	508	509	509	510	511	511	512	512	514
514	515	515	515	517	519	520	520	523	525
525	527	528	530	530	531	533	537	537	538
538	538	540	553	560	562	568	569	581	587

Table VIII
Distribution of the bodyweight of 100 cattle from Table VII

Weight class (kg) x_i	**Number of cattle in each class** c_i	**Relative frequency of each class** $p(x_i)=\frac{c_i}{n}$	**Cumulative probability** $P(X \le x_i)=\sum_{i=1}^{n} p(x_i)$
400 < 420	1	0.01	0.01
420 < 440	6	0.06	0.07
440 < 460	9	0.09	0.16
460 < 480	19	0.19	0.35
480 < 500	17	0.17	0.52
500 < 520	25	0.25	0.77
520 < 540	15	0.15	0.92
540 < 560	3	0.03	0.95
560 < 580	3	0.03	0.98
580 < 600	2	0.02	1.00
Total	***n* =100**	**1.00**	–

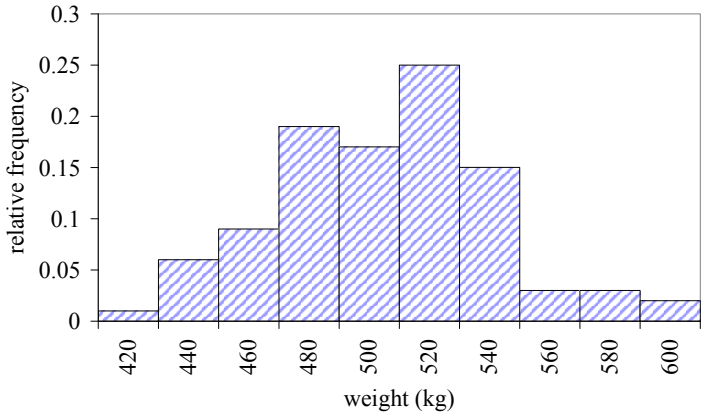
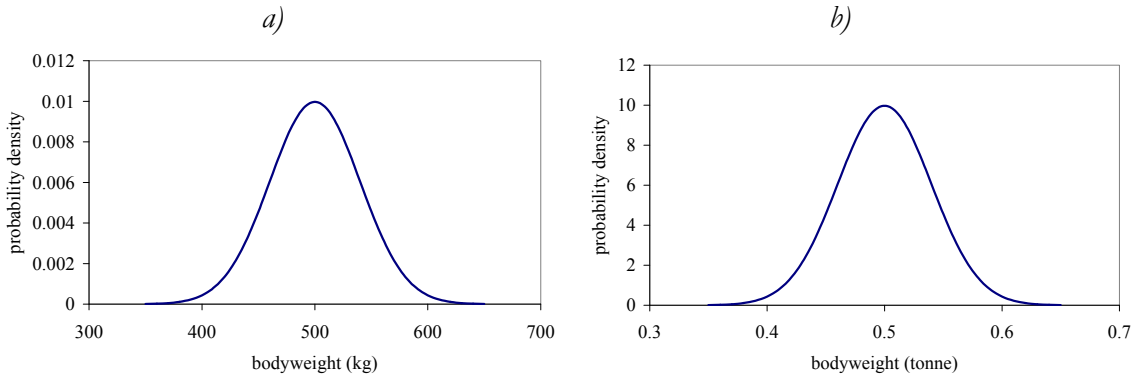


Figure 8
A continuous distributions (histogram plot) of cattle bodyweight based on the data in Table VIII. The class interval is 20 kg

Instead of defining a continuous distribution directly from the data, as we have just done, we could use a mathematical function, such as the normal distribution function. As discussed in Chapter 3, this function is characterised by two parameters, the mean (μ) and standard deviation (σ) which are estimated from a data set. While it provides quite a good approximation for many biological observations, such as weight, you should satisfy yourself that it is, in fact, an appropriate distribution to use for your data. For example, a normal distribution is unimodal, symmetrical about the mean (the mean, median and mode are all equal) and 95% of the distribution lies within ± 1.96 standard deviations of the mean.

A normal distribution, based on the data in Table VII, is presented in Figure 9a. In contrast to a discrete distribution, where the relative frequency is the actual probability of occurrence of the discrete variable, the relative frequency of a continuous variable refers to an interval rather than an exact value. Since a continuous scale is infinitely divisible any value chosen simply reflects an interval. For example, the *NORMDIST* function in Excel calculates the probability of a cow weighing 500 kg as 0.011. The correct interpretation is that the probability of a cow weighing 500 kg, plus or minus a tiny amount, is 0.011. For a continuous variable, probability is correctly referred to as *probability density* and the area under the curve must add up to one. It is important to note that the vertical scale (*y*-axis) changes according to the units used on the *x*-axis. This is demonstrated in Figure 9 where bodyweight is expressed in either kilograms or tonnes.



a) cattle bodyweight in kg: normal (496,35)

b) cattle bodyweight in tonnes: normal (0.496,0.035)

Figure 9
Two normal distribution plots of cattle bodyweight based on the data in Table VII

Some discrete variables, such as bacterial cell counts or faecal egg counts, can be conveniently treated as continuous variables, where the gap between allowable values is considered to be insignificant in comparison to the magnitude of the values.

Table IX provides some examples of discrete and continuous distributions. These are discussed in Chapter 4.

Table IX
Some examples of discrete and continuous distributions

Discrete distributions	Continuous distributions
Binomial	Beta
Discrete	Cumulative
Discrete uniform	Exponential
Hypergeometric	Gamma
Negative binomial	General
Poisson	Histogram
	Lognormal
	Normal
	PERT
	Triangular
	Uniform

Chapter 3
Theorems providing a basis for probabilistic risk assessment

There are three important theorems which provide a basis for probabilistic risk assessment: the binomial theorem, the central limit theorem and Bayes' theorem.

Binomial theorem

The binomial theorem provides a formula that enables us to easily calculate the probability of x successes in n trials where each trial has the same probability (p) of success. This idea was introduced in Chapter 1 of this volume.

To understand the binomial theorem we will start with a Bernoulli trial, which is one of the simplest, yet most important, random processes in probability. The classic example of a Bernoulli trial is tossing a coin. There are two possible outcomes, either a head or a tail. If the coin is fair, the probability of obtaining either a head or a tail after a single toss is 0.5 or 50%. If we toss the coin again, the probability of obtaining a head or a tail does not change. That is, the results of the two trials are independent. As discussed in Chapter 2 the probability of obtaining a head followed by a tail then a head is calculated by multiplying the respective probabilities:

$$P(H \cap T \cap H) = P(H) \times P(T) \times P(H) = 0.5 \times 0.5 \times 0.5 = 0.125$$

A binomial process is a collection of such Bernoulli trials, and satisfies three assumptions:

– each trial has two possible outcomes, called a success or a failure

– the trials are independent. That is, the outcome of one trial has no influence over the outcome of another trial

– each trial has the same probability of success (p). The probability of failure is 1-p.

The binomial process can be easily applied to a situation where we choose animals from an infected herd. In this case, the two possible outcomes are that an animal is either infected or uninfected. Provided the herd is sufficiently large, we can reasonably assume that the probability that an animal is infected remains constant. This means that the disease status of an individual animal selected at random is independent of the disease status of all the other animals chosen beforehand. It also assumes that transmission of infection does not occur during the sampling period.

Suppose we want to determine the probability of obtaining x infected animals in a sample of size n drawn from a herd where the disease prevalence is p. If we select three animals ($n = 3$) we can see from Figure 10 that there is/are:

– one way of obtaining 3 infected animals, $p \times p \times p$

– three ways of obtaining 2 infected animals, $p \times p \times (1-p)$, $p \times (1-p) \times p$, $(1-p) \times p \times p$

– three ways of obtaining 1 infected animal, $p \times (1-p) \times (1-p)$, $(1-p) \times p \times (1-p)$, $(1-p) \times (1-p) \times p$

– one way of obtaining 0 infected animals, $(1-p) \times (1-p) \times (1-p)$.

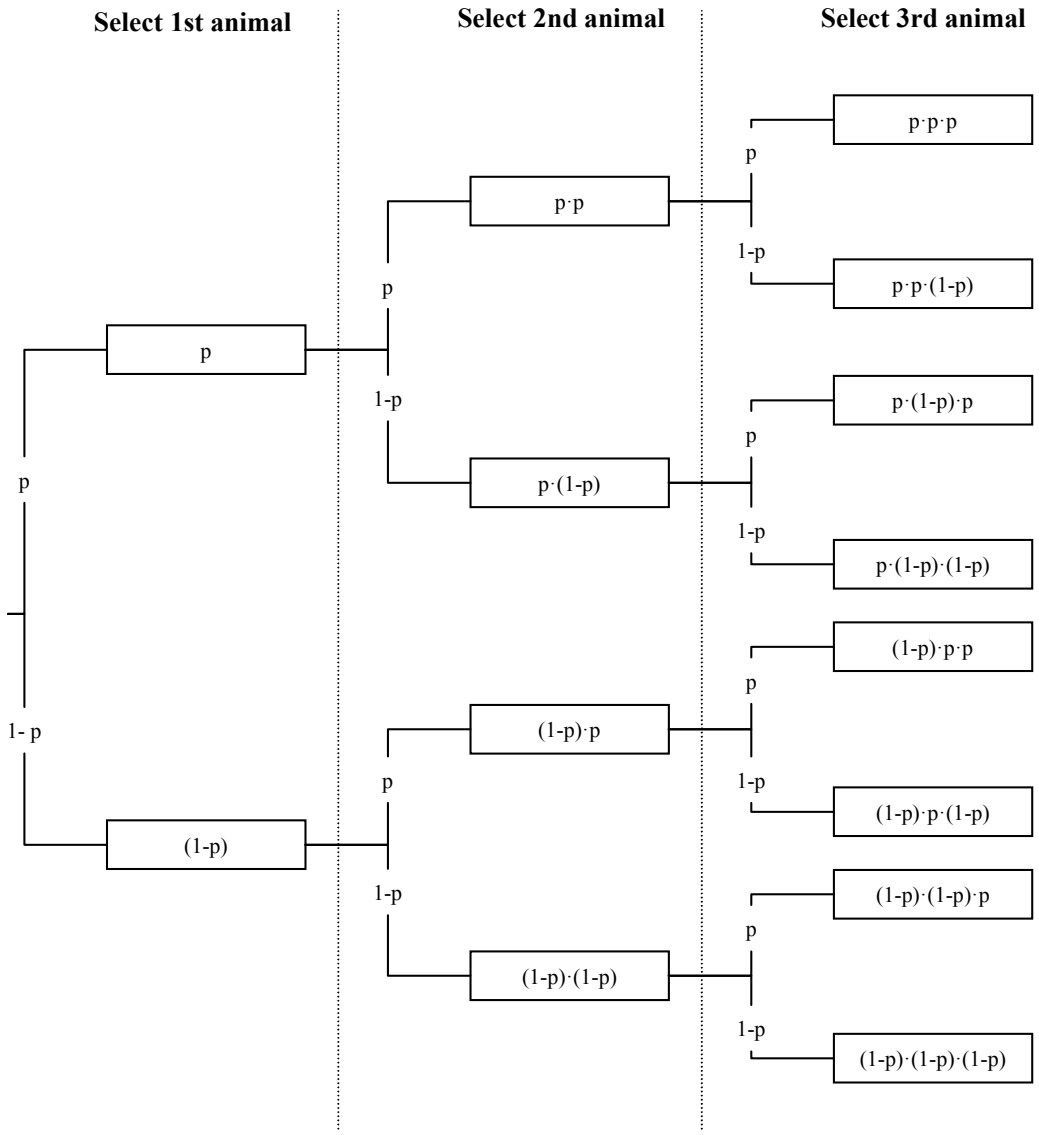


Figure 10
A probability tree outlining the ways of obtaining *x* infected animals if three animals are selected, where *p* is the probability that an animal is infected and (1-*p*) is the probability that an animal is not infected

Each of these probabilities can be represented as $p^x \times (1-p)^{n-x}$. By substituting the appropriate values for x and n into this formula we can calculate the probability of a particular outcome. For example, the probability that all three animals are infected is given by $p^3 \times (1-p)^0 = p^3$, while each of the probabilities of obtaining two infected animals is given by $p^2 \times (1-p)^1$. Since we are interested in the probability of obtaining x infected animals, regardless of the order in which they were selected, we need to multiply these results by the number of ways we can obtain x infected animals, for example $3 \times p^2 \times (1-p)^1$. As the number of animals sampled increases, it rapidly becomes impractical to work out the number of ways of obtaining x infected animals by drawing a probability tree and determining the number of ways that lead to a particular outcome. Fortunately, there is an easy way to do this. The number of combinations by which x successes in n trials may be obtained is calculated as $\frac{n!}{x!(n-x)!}$. This is known as the *binomial coefficient* and in

mathematical notation is expressed as either $\binom{n}{x}$ or ${}^{n}C_{x}$, which is read as '*n* combination *x*'. In the preceding example the corresponding binomial coefficients are:

$$\binom{3}{3} = {}^{3}C_{3} = \frac{3!}{3!(3-3)!} = 1$$

$$\binom{3}{2} = {}^{3}C_{2} = \frac{3!}{2!(3-2)!} = 3$$

$$\binom{3}{1} = {}^{3}C_{1} = \frac{3!}{1!(3-1)!} = 3$$

$$\binom{3}{0} = {}^{3}C_{0} = \frac{3!}{0!(3-0)!} = 1$$

Note: the factorial of a number, $n! = 1 \times 2 \times 3 \ldots \times n$ and $0! = 1$

It can be seen that the generic formula to calculate the probability of exactly *x* successes in *n* trials is:

$$P(X = x) = \binom{n}{x} p^{x} (1-p)^{n-x} \qquad \text{Equation 9}$$

If we add up all the possibilities, that is from $x = 0$ to $x = n$, we obtain the binomial distribution:

$$\sum_{x=0}^{n} \binom{n}{x} p^{x} (1-p)^{n-x} = 1 \qquad \text{Equation 10}$$

Continuing with our example, if we just choose three animals, the corresponding binomial distribution is:

$$\sum_{x=0}^{3} \binom{3}{x} p^{x} (1-p)^{3-x} = p^{3} + 3p^{2}(1-p) + 3p(1-p)^{2} + (1-p)^{3} = 1 \qquad \text{Equation 11}$$

Rather than having to laboriously work out each of the terms in the binomial distribution, one may use a spreadsheet package such as Excel which provides a binomial distribution function enabling an individual binomial term to be calculated by entering the appropriate values for *x*, *n*, and *p:*

$$P(X = x) = BINOMDIST(x, n, p, 0) = \binom{n}{x} (p)^{x} (1-p)^{n-x} \text{ [9]} \qquad \text{Equation 12}$$

The general solution to raising a binomial to an integral power, for example $(a+b)^{n}$, is provided by:

$$(a+b)^{n} = \sum_{x=0}^{n} \binom{n}{x} a^{x} b^{n-x} \qquad \text{Equation 13}$$

where *n* is a positive integer.

Equation 13 can be manipulated to obtain a particular result. For example, the most common situation is likely to be where we want to determine the probability that at least one infected animal will be present in a sample drawn from an infected herd. In this case

[9] The last parameter 0 in the BINOMDIST is a switch to make the function return either the binomial probability mass (switch = 0) or the cumulative binomial probability (switch = 1)

we sum the distribution from $x = 1$ to $x = n$. Alternatively we could subtract the term $(b)^n$ from $(a+b)^n$ to obtain:

$$P(x \geq 1) = (a+b)^n - (b)^n \qquad \text{Equation 14}$$

We can then replace a and b with p and $(1-p)$:

$$P(x \geq 1) = (p + (1-p))^n - (1-p)^n = 1 - (1-p)^n \qquad \text{Equation 15}$$

We can also determine the probability that all the animals in our sample are test negative given there is at least one infected animal among them. In this case we replace a with $p \times (1 - Se)$, the probability that an animal is test negative given it is infected and replace b with $(1-p) \times Sp$, the probability that an animal is test negative given it is not infected:

$$P(T^- = 0 | \geq 1 D^+) = (p \times (1 - Se) + (1-p) \times Sp)^n - ((1-p) \times Sp)^n \qquad \text{Equation 16}$$

where: T^- = test negative
D^+ = infected
p = prevalence
Se = test sensitivity
Sp = test specificity.

An alternative derivation using Equation 13 is to imagine that we randomly select n animals from a population with prevalence p and we test all these animals using a test with sensitivity Se and specificity Sp. The probability that there are x infected animals in our sample is:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{Equation 17}$$

The probability that all these animals also test negative is $(1 - Se)^x Sp^{n-x}$. So the probability that the sample tests negative is these two probabilities multiplied together and the result summed for all x:

$$P(allT^-) = \sum_{x=0}^{n} \binom{n}{x} p^x (1 - Se)^x (1-p)^{n-x} Sp^{n-x} \qquad \text{Equation 18}$$

Comparing Equation 18 with Equation 13, we can set $a = p(1-Se)$ and $b = (1-p).Sp$, so Equation 18 reduces to:

$$P(allT^-) = (p(1 - Se) + (1-p)Sp)^n \qquad \text{Equation 19}$$

Central limit theorem

The normal distribution

The normal distribution (Fig. 11) is characterised by two parameters, the mean (μ), and standard deviation (σ).

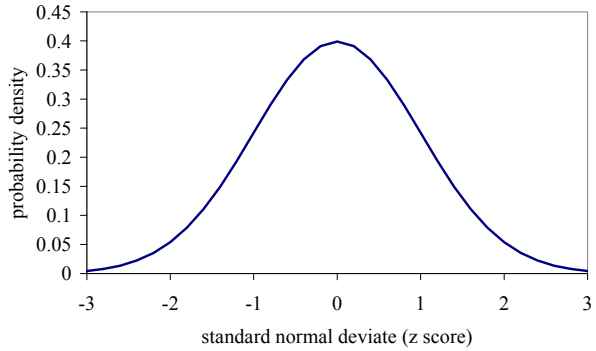


Figure 11
A standardised normal probability distribution where the mean (μ) equals zero and the standard deviation (σ) equals 1

The mean or average is calculated by summing all the values in the population and dividing by the number of values:

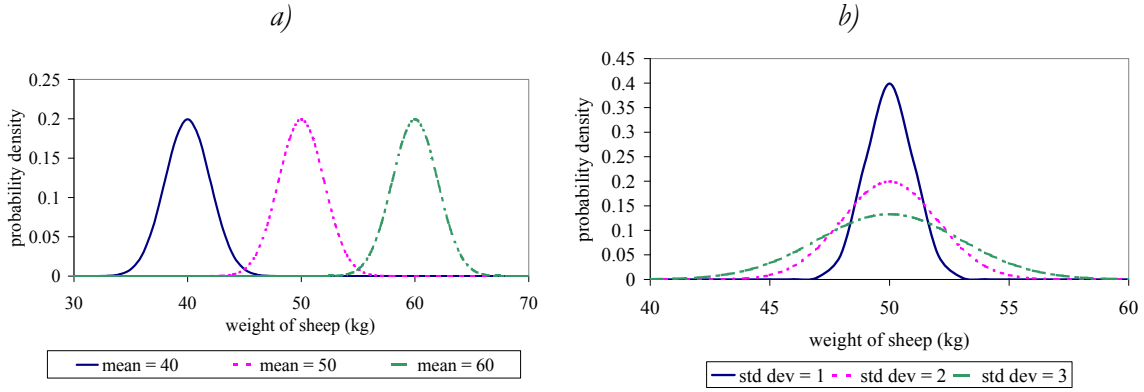
$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \text{Equation 20}$$

The standard deviation is a measure of the amount of variation about the mean. It is calculated by summing the square of the difference from the mean for each value in the population, dividing the sum by the number of values and finally taking the square root of this result:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}} \qquad \text{Equation 21}$$

The square of the standard deviation is known as the variance (σ^2).

The normal distribution is an unbounded continuous distribution that extends from minus infinity to plus infinity and has a bell shaped curve. It is symmetrical about its mean so that the area under the curve to either the left or right of the mean is 50%. Ninety-nine percent of its values lie within ± 2.58 standard deviations of the mean. The mean locates the distribution on the *x*-axis (Fig. 12a) and the standard deviation determines its spread (Fig. 12b). As the standard deviation increases, the height of the distribution decreases and its spread increases.



a) mean or location parameter changes *b)* standard deviation or scale parameter changes

Figure 12
A series of normal distributions demonstrating influence of changes in either the mean (μ) or standard deviation (σ)

The standardised normal distribution has a mean of zero and a standard deviation of 1. It permits the use of the same table of probabilities for all normal distributions. A normal random variable (x) can be standardised by expressing its value as the number of standard deviations (σ), to the left or right of the mean (μ). The resulting value is known as the standard normal deviate or z score:

$$z = \frac{x - \mu}{\sigma} \qquad \text{Equation 22}$$

The percentage of the area under the standard normal curve between two z scores represents the probability $P(z_1 \leq z_i \leq z_2)$. For example, if we weighed all the sheep in a flock and calculated the average body weight (μ = 50 kg) and standard deviation (σ = 4 kg), we could determine the probability that a sheep chosen at random weighs between 40 kg and 45 kg by referring to Figure 13 and:

– determining the z scores for 40 kg and 45 kg: $z_1 = \frac{40-50}{4} = -2.5 \quad z_2 = \frac{45-50}{4} = -1.25$
– looking up the percentage of the area under the curve between the mean and each respective z score from a normal probability table: for z1 49.4% of the total area falls between the mean and –2.5 while z2 encompasses 39.4% of the total area.
– subtracting the area encompassed by z2 from z1: 49.4%-39.4% = 10.0%

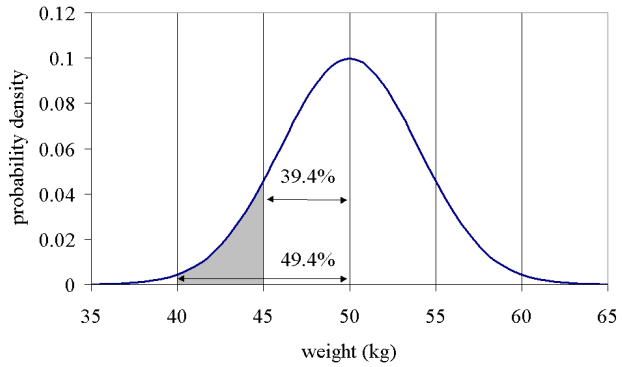


Figure 13
A normal probability distribution of the bodyweight of sheep with a mean of 50 kg and a standard deviation of 4 kg

The answer to this example can also be obtained directly from the cumulative probability distribution by simply subtracting the cumulative probability for 40 kg from the cumulative probability for 45 kg (Fig. 14).

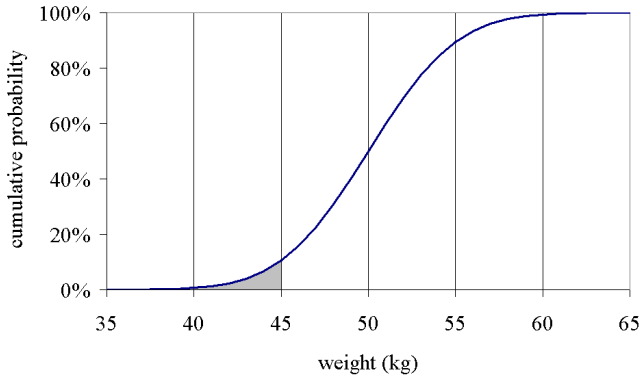


Figure 14
A cumulative normal probability distribution of the bodyweight of sheep with a mean of 50 kg and a standard deviation of 4 kg

Specific z scores, associated with a particular area under the curve relative to the mean, are often used in descriptive statistics and statistical inference. For example, 90% of the total area under the curve lies between $z = \pm 1.64$, 95% lies between $z = \pm 1.96$ and 99% lies between $z = \pm 2.58$. It is important to remember that the z score relates to the distance from the mean. As a result, a z score of +1.96 represents 47.5% of the area under the curve to the right of the mean a z score of –1.96 represents 47.5% of the area under the curve to the left of the mean. So, if from the previous example, we are interested in describing the range in which we expect 95% of sheep bodyweights to lie we calculate it as $\mu \pm 1.96 \times \sigma = 50 \pm 7.8$, which equals a range of 42.2 kg to 58.8 kg.

Instead of weighing all the sheep in the flock we might just weigh some of them and want to make an inference about the average bodyweight of all the sheep in the flock. Since we are dealing with a sample of the population, the notation for the mean and standard deviation changes to reflect this. The sample mean is expressed as $\bar{x}$, while the sample standard deviation is *s*. In this case, since the population parameters (μ and σ) are unknown we need to calculate the standard deviation of the sample mean. This is more commonly referred to as the standard error of the sample mean, which is written as $s_{\bar{x}}$ and calculated as: $s_{\bar{x}} = \frac{s}{\sqrt{n}}$ Equation 23

If we weighed 30 sheep and calculated their average weight as 48.5 kg and the standard deviation as 3.5 kg, the standard error of the sample mean, from Equation 23, is $\frac{3.5}{\sqrt{30}} = 0.64$. We can now calculate a 95th percent confidence interval about our sample mean, $\bar{x} \pm 1.96 \times \frac{s}{\sqrt{n}} = 48.5 \pm 1.96 \times 0.64 = 48.5 \pm 1.25$ and conclude that, at the 95th percent confidence level, the average weight of all the sheep in the flock is likely to be between 47.25 kg and 49.75 kg.

Defining the central limit theorem

The central limit theorem defines a relationship between the sampling distribution of the mean and the population distribution. A sampling distribution of the mean is obtained by repeatedly collecting *n* samples, calculating the mean of each of the *n* samples, determining how frequently we obtain each mean value and plotting the results on a graph. For example, we could weigh the amount of de-boned beef derived from each of five carcasses chosen at random and calculate the mean weight of the batch. If we repeated this exercise 100 times we would gradually build up a distribution of mean weights (Fig. 15). Instead of weighing the amount of de-boned beef from five carcasses, we could increase our sample size to 10, or 25, or 100. As the sample size increases the sampling distribution of the mean looks more and more like a normal distribution (Fig. 16). This is, in fact, the relationship described by the central limit theorem, which can be formally stated as:

If samples of size n, where n is large (usually greater than 30), are repeatedly taken from any population, regardless of the shape of its distribution, then the means of each of the samples are approximately normally distributed with a mean of μ *and standard deviation of* $\frac{\sigma}{\sqrt{n}}$,

where: μ *is the mean and*
σ *is the standard deviation of the population from which the samples are taken.*

The distribution of the sample means $\bar{x}$ is modelled by:

$$\bar{x} = Normal\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \qquad \text{Equation 24}$$

Each sampling distribution of the means in Figure 16 is overlaid with a normal distribution curve modelled by this function. As can be seen from these plots it does not provide a good approximation for small sample sizes.

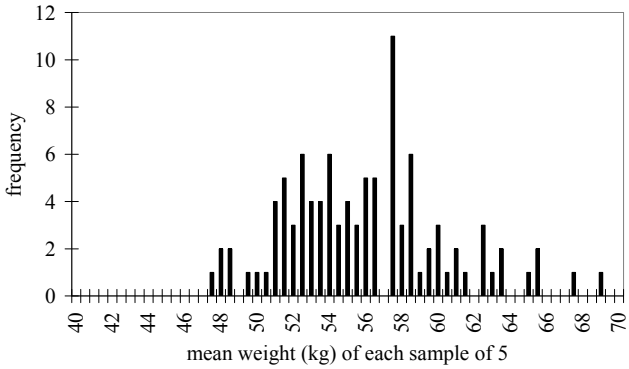


Figure 15
A sampling distribution of the mean of the weight of de-boned beef derived from five carcasses
Five samples were randomly selected from a *PERT*(40,50,90) distribution, the mean for each sample calculated and the results plotted on a frequency graph

The fact that the central limit theorem does not depend on the shape of the distribution from which the samples are drawn is amply demonstrated in Figure 17, which is the original distribution upon which the examples in Figures 15 and 16 are based.

The standard deviation of the sampling distribution of the mean, $\left(\frac{\sigma}{\sqrt{n}}\right)$, is more commonly referred to as the standard error of the mean. It enables us to obtain a measure of the extent we can expect the means from different samples of size *n* to vary.

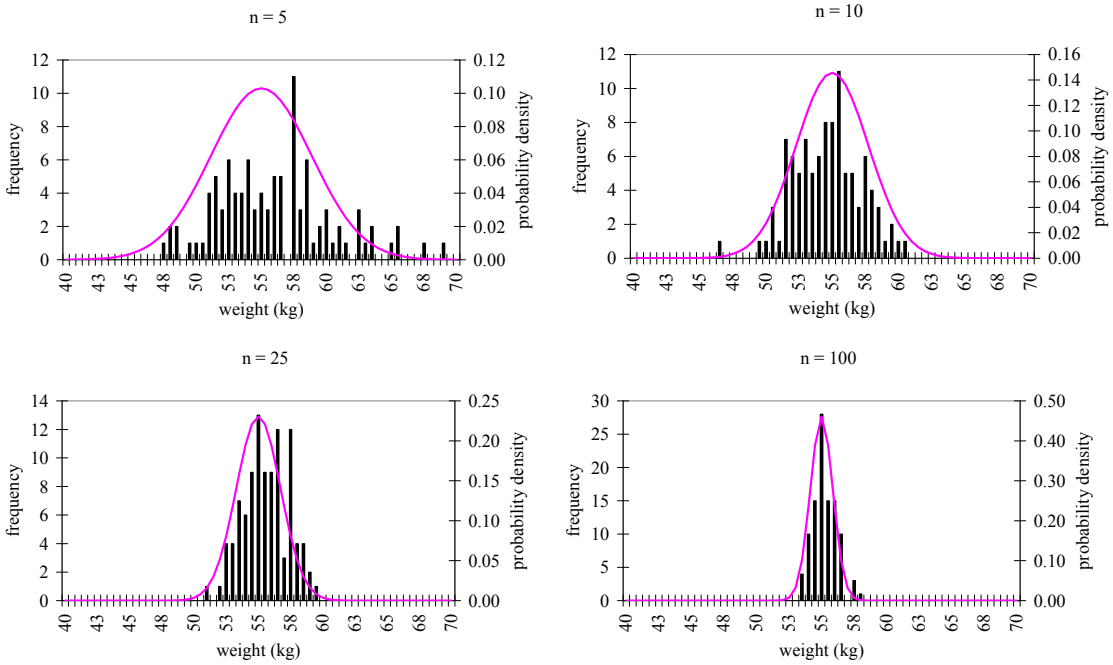


Figure 16
Sampling distribution of the mean of 100 samples of size *n* drawn from a *PERT*(40,50,90) distribution

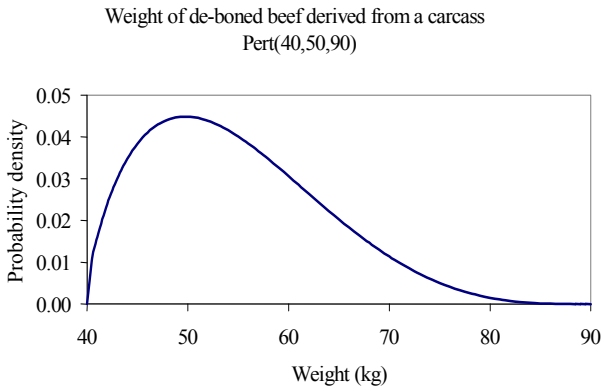


Figure 17
The original distribution upon which the examples in Figures 15 and 16 are based

Since the mean of a sample of size *n* drawn from a population is given by $\bar{x} = Normal\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, it follows that the distribution of the sum of these *n* independent samples is obtained by multiplying by *n*, that is:

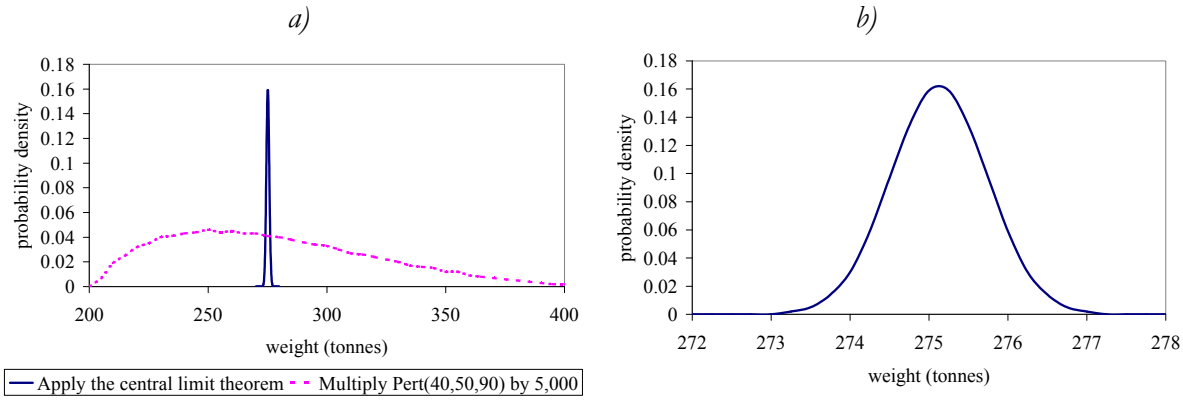
$$\sum x_i = n\bar{x} = Normal\left(n\mu, \sigma\sqrt{n}\right) \qquad \text{Equation 25}$$

Population mean (μ) and population standard deviation (σ) are known

In some circumstances, we might reasonably assume that the population mean (μ) and population standard deviation (σ) are known. For example, if we have abundant representative data from which to derive the parameters of a distribution. Following on from the de-boned meat example in the preceding section, we will assume that a *PERT*(40,50,90) distribution accurately reflects the amount of de-boned beef derived from a carcass. Suppose we want to determine how much de-boned meat would be obtained if 5,000 cattle were slaughtered. A common mistake would be to multiply a random value from the distribution of weight (*PERT*(40,50,90)) by 5,000 in a simulation. This results in a grossly exaggerated and incorrect distribution (Fig. 18a). Such a calculation fails to appreciate that the amount of meat derived from each animal is a random sample from the *PERT*(40,50,90) distribution. That is each animal is independent of the next. It models implausible scenarios by assuming that each of the 5,000 animals contributes exactly the same amount of meat. By applying the central limit theorem we can avoid this mistake. In this example the correct way to model the amount of de-boned meat in tonnes derived from 5,000 cattle using Equation 25 is (Fig. 18b):

$$\sum x_i = \frac{Normal\left(5000\times 55, 8.7\times\sqrt{5000}\right)}{1000}$$

where 55 kg and 8.7 kg are the mean and standard deviation of the PERT(40, 50, 90) population distribution respectively.



a) a comparison of the probability distributions obtained by a commonly used incorrect method of calculating the weight of de-boned meat derived from 5,000 cattle with the correct result obtained by applying the central limit theorem. *Note*: the area under the curve for both distributions equals 1

b) the correct distribution for the amount of de-boned beef derived from 5,000 cattle

Figure 18

If each tonne of beef results in $6,000 worth of export revenue, what is the likely return from 5,000 cattle? It should be apparent that the correct way to model this question is:

$$\$6{,}000\times\frac{Normal\left(5000\times 55,8.7\times\sqrt{5000}\right)}{1000}$$

Suppose the revenue per tonne is modelled by a distribution, such as *PERT*($420,$550,$780). In this case consideration will need to be given as to how the meat is sold. That is, what is the smallest independent unit by which the meat is sold? For instance, is it per tonne, per container or per shipment? If the meat is sold in lots of one tonne then we might assume that the price received for each tonne is a random sample from the *PERT*($420,$550,$780) distribution. In this instance, the output from the formula predicting the number of tonnes of de-boned meat derived from 5,000 cattle, $tonnes = \frac{Normal\left(5000\times 55,8.7\times\sqrt{5000}\right)}{1000}$, is used as an input into a formula modelling the return per tonne:

$$Normal\left(tonnes\times\$\mu,\$\sigma\times\sqrt{tonnes}\right) \qquad \text{Equation 26}$$

where $\$\mu$ and $\$\sigma$ represent the mean and standard deviation of the *PERT*($420,$550,$780) distribution modelling the expected return ($) per tonne.

If, however, the meat from 5,000 cattle is sold as one unit then the revenue generated is modelled by:

$$Pert(\$420,\$550,\$780)\times\frac{Normal\left(5000\times 55,8.7\times\sqrt{5000}\right)}{1000}$$

Population mean (μ) and population standard deviation (σ) are not known

In many situations we do not know the value of the population mean (μ) or the population standard deviation (σ) or the shape of the underlying population distribution. We may only have one sample available to estimate these values. In these circumstances, provided we have more than thirty random samples, and provided the population distribution is not highly skewed, the central limit theorem will allow us to use the sample mean ($\bar{x}$) and the sample standard deviation (s) to make inferences about the population mean (μ).

A probability distribution for the uncertain parameter, the population mean (μ) can be modelled by: $\mu = Normal\left(\bar{x}, \frac{s}{\sqrt{n}}\right)$ Equation 27

We can see from Figure 16 that as the sample size (n) increases we can be increasingly confident that we can make use of the central limit theorem. With larger sample sizes the distribution of the sample means is subject to less variability, as the estimate for each mean is less influenced by particular values that may dominate the estimate. Sample sizes of thirty or more are usually regarded as adequate. If the sample size is less than thirty and the underlying population distribution is not normal or approximately symmetrical, the central limit theorem may not provide a valid approximation. In this case, more samples could be collected. If this is not possible, the techniques outlined in Chapter 6 for developing probability distributions for uncertain parameters such as the mean, where there are few representative data, could be used.

*Estimating the number of individuals (*n*) required to achieve a fixed total quantity*

Continuing with the de-boned meat example we might want to determine how many cattle contribute to a tonne of de-boned meat. Simply dividing 1,000 kg by a random value from the distribution of weight (*PERT*(40,50,90)) does not take into account that the amount of de-boned meat derived from each animal is a random sample from the *PERT*(40,50,90) distribution. That is, each animal is independent of the next. It models implausible scenarios by assuming that each animal contributes exactly the same amount of meat. Once again we need to apply the central limit theorem. In this case, however, it is not quite so straightforward. We need to set up a spreadsheet model that accounts for independence between animals as outlined in Table X. Each cell in column B independently samples from a *PERT*(40,50,90) distribution while column D adds up the cumulative results. Column E determines when the cumulative sum reaches one tonne and how many animals are needed by referring to column B. The model is then run on the output (cell E34), the results collected and the mean and standard deviation determined. From the discussion in the previous section, it should be apparent that we need to run the model for at least thirty iterations to estimate the sampling statistics. Several hundred iterations, as shown in Figure 19, should suffice. We can now use these results to determine how many cattle contribute to, for example, 20,000 tonnes of meat in the formula $Normal\left(20000\mu, \sigma\sqrt{20000}\right)$. Figure 20 demonstrates the consequence of ignoring the fact that each animal is independent.

Table X
An extract from a spreadsheet model calculating the number of cattle equivalents in one tonne of de-boned meat

	A	B	C	D	E
1	Weight of meat derived from an animal	(kg)	40	50	90
2	*Formulae:*	Number of	Meat	Cumulative	Number
3	C4:C33 {PERT($C$1,$D$1,$E$1)}	cattle	(kg)	sum	per tonne
4	D4 {=$C$4}	1	52	52	0
5	D5 {=D4+C5}, D6 {=D5+C6} etc	2	49	101	0
6	E4 {=0}	3	55	156	0
7	E5 {=IF(D5<1000,IF(D6>1,000,B6,0),0)} etc.	4	49	206	0
...	E34 {SUM($E$4:$E$33)}	…	…	…	…
33		30	53	1,651	0
34	**Number of cattle represented in a tonne**				**19**

Note: lines 8-32 not shown

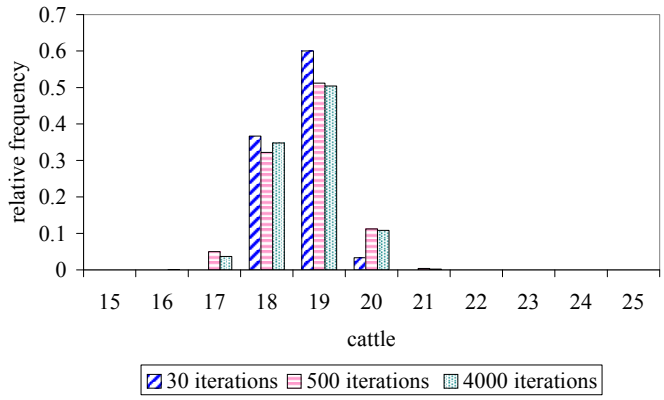


Figure 19
The results obtained from running the model outlined in Table X for different numbers of iterations to obtain a distribution of the number of cattle equivalents per tonne of de-boned beef

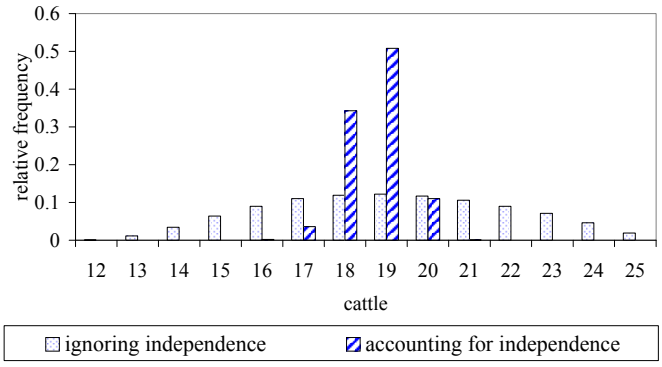


Figure 20
A comparison of the probability distributions of the number of cattle equivalents per tonne of de-boned beef when independence between cattle is either ignored or taken into account

Bayes' theorem

Bayes' theorem is a fundamental probability law governing the process of logical inference based on the information available. Suppose a herd of cows has a disease prevalence (*p*) of 30% and we have a test available that has a sensitivity (*Se*) of 90% and a specificity (*Sp)* of 98%. From this information we can directly determine the probabilities, P1 to P12, outlined in Figure 21 and Table XI. These probabilities allow us to answer questions such as 'What is the probability that a cow is infected and is test negative?' or 'What is the probability that a cow chosen at random will be test negative?' P10 and P12 give the answers to these particular questions. We might also want to determine the probability that a cow is infected, given that it is test negative. To do this we need to work out the proportion of a negative test result that is attributable to a false negative result by dividing the probability that a cow is infected and is test negative by the probability that a cow will be test negative irrespective of its disease status. This is Bayes' theorem in operation. It allows us to revise our original probability estimate that a randomly chosen cow is infected (30%), in light of the new information we obtained by testing the cow.

Bayes' theorem can be more formally expressed as:

$$P(A \mid B) = \frac{P(A) \times P(B \mid A)}{P(B)} \qquad \text{Equation 28}$$

where *P(A)* represents our existing knowledge and is referred to as the prior probability.

- *P(B|A)* is a conditional probability expressing the likelihood that *(B)* will be observed given our prior knowledge of event *(A)*
- *P(B)* is the probability of event *(B)* irrespective of the status of event *(A)*
- *P(A|B)* is the revised or conditional probability of event *(A)* given the new information we have obtained, *P(B)*.

Using the cow example described above we could calculate the probability that the cow is infected given that it is test negative:

$$P(D^+ \mid T^-) = \frac{P(D^+) \times P(T^- \mid D^+)}{P(T^-)} = \frac{p(1-Se)}{p(1-Se) + (1-p)Sp} \qquad \text{Equation 29}$$

$$P(D^+ \mid T^-) = \frac{0.3 \times (1-0.9)}{0.3 \times (1-0.9) + (1-0.3) \times 0.98} = 0.04$$

We could also calculate this probability as: $P(D^+ \mid T^-) = 1 - P(D^- \mid T^-) = 1 - NPV$ Equation 30

where: $NPV = \frac{(1-p)Sp}{(1-p)Sp + p(1-Se)}$

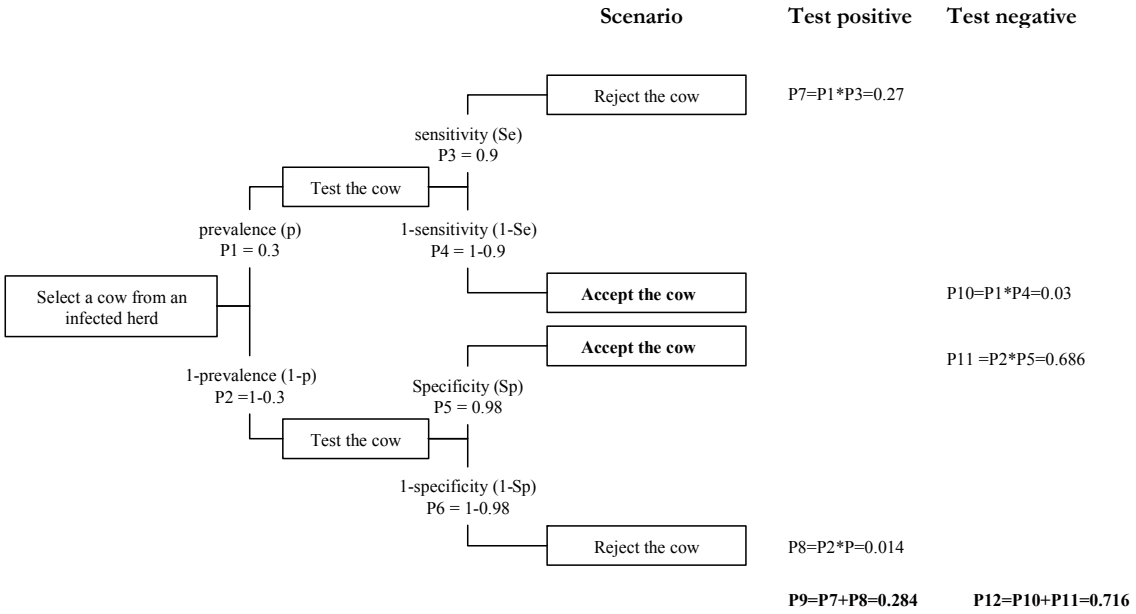


Figure 21
A scenario tree outlining the pathways whereby a cow, selected at random from an infected herd, is accepted or rejected after it is tested

Table XI
A table of probabilities related to animal disease prevalence and testing

Note: probabilities in the form of $P(T^+ \mid D^+)$ are conditional probabilities. This example is for the sensitivity of a test, that is the probability of a cow being test positive is conditional on the cow being infected

P1.	**Prevalence:** the probability that a cow is infected	$P(D^+) = p$
P3.	**Sensitivity:** the probability that the test will yield a positive result if the cow is infected	$P(T^+ \mid D^+) = Se$
P5.	**Specificity:** the probability that the test will yield a negative result if the cow is <u>not</u> infected	$P(T^- \mid D^-) = Sp$
1. Complementary probabilities $P(B) = 1 - P(A)$		
P2.	The probability that a cow is not infected	$P(D^-) = 1 - P(D^+) = 1 - p$
P4.	The probability that the test will yield a negative result if the cow is infected	$P(T^- \mid D^+) = 1 - P(T^+ \mid D^+) = 1 - Se$
P6.	The probability that the test will yield a positive result if the cow is <u>not</u> infected	$P(T^+ \mid D^-) = 1 - P(T^- \mid D^-) = 1 - Sp$
2. Joint probabilities under statistical dependence i.e. where the probability of one event is dependent on another $P(A \cap B) = P(A) \times P(B \mid A)$		
P7.	**True Positive:** the probability that the cow is infected and yields a positive test result	$P(D^+ \cap T^+) = P(D^+) \times P(T^+ \mid D^+) = p \times Se$
P8.	**False Positive:** the probability that the cow is <u>not</u> infected and yields a positive test result	$P(D^- \cap T^+) = (1 - P(D^+)) \times (1 - P(T^- \mid D^-)) = (1 - p) \times (1 - Sp)$
P10.	**False negative:** the probability that the cow is infected and yields a negative test result	$P(D^+ \cap T^-) = P(D^+) \times (1 - P(T^+ \mid D^+)) = p \times (1 - Se)$
P11	**True negative:** the probability that the cow is <u>not</u> infected and yields a negative test result	$P(D^- \cap T^-) = (1 - P(D^+)) \times P(T^- \mid D^=) = (1 - p) \times Sp$
3. Mutually exclusive events $P(A \cup B) = P(A) + P(B)$		
P9.	**Test positive:** the probability of a positive test result irrespective of the disease status of the cow	$P(T^+) = P(D^+ \cap T^+) + P(D^- \cap T^+) = p \times Se + (1 - p) \times (1 - Sp)$
P12.	**Test negative:** the probability of a negative test result irrespective of the disease status of the cow	$P(T^-) = P(D^+ \cap T^-) + P(D^- \cap T^-) = p \times (1 - Se) + (1 - p) \times Sp$
4. Conditional probability under statistical dependence (Bayes' theorem) $P(A \mid B) = \frac{P(A) \times P(B \mid A)}{P(B)}$		
P13	**True positive or PPV (positive predictive value):** the probability that the cow is infected given the test result is positive	$P(D^+ \mid T^+) = \frac{P(D^+ \cap T^+)}{P(T^+)} = \frac{p \times Se}{p \times Se + (1 - p) \times (1 - Sp)}$
P14	**False positive:** the probability that the cow is <u>not</u> infected given the test result is positive	$P(D^- \mid T^+) = \frac{P(D^- \cap T^+)}{P(T^+)} = \frac{(1 - p) \times (1 - Sp)}{pSe + (1 - p) \times (1 - Sp)}$
P15	**True negative or NPV (negative predictive value):** the probability that the cow is <u>not</u> infected given the test result is negative	$P(D^- \mid T^-) = \frac{P(D^- \cap T^-)}{P(T^-)} = \frac{(1 - p) \times Sp}{p \times (1 - Se) + (1 - p) \times Sp}$
P16	**False negative:** the probability that the cow is infected if the test result is negative	$P(D^+ \mid T^-) = \frac{P(D^+ \cap T^-)}{P(T^-)} = \frac{p \times (1 - Se)}{p \times (1 - Se) + (1 - p) \times Sp}$

Note: p = prevalence, Se = test sensitivity, Sp = test specificity

Chapter 4
Useful probability distributions[10]

A large number of probability distributions is available to the risk analyst, but their inappropriate use may lead to important flaws in the analysis. A relatively small number of probability distributions has proven useful and appropriate in import risk analysis and will be discussed in this chapter. The distributions discussed include those based on a binomial process (binomial, Beta, negative binomial) and those based on a Poisson process (Poisson, Gamma, exponential). Also covered are the cumulative, discrete, general, histogram, normal and lognormal, *PERT* (Beta PERT), triangular and uniform distributions.

Distributions used to model a binomial process

A binomial experiment or process has five characteristics:

– the experiment consists of n identical trials

– each trial results in one of two possible outcomes, either a success or a failure

– the probability of a success on a single trial is equal to p and remains the same from trial to trial

– the trials are independent, that is they are not influenced by the results of any previous trails

– the interest is in x, the number of successes observed in n trials, for x = 0, 1, 2, ..., n.

The binomial process can be characterised by two parameters; the number of trials (*n*) and the probability (*p*) that each trial is successful. The outcome is expressed as the number of successes (*x*). Tossing a coin or selecting a card from a pack are classic examples of a binomial process. It can also be applied to some situations where all the assumptions are not strictly met but which, for all practical purposes, approximate a binomial process. For example, suppose we choose animals from an infected herd. In this case, the two possible outcomes are that an animal is either infected or uninfected. If the herd is sufficiently large, we can reasonably assume that the probability that an animal is infected remains constant. This means that the disease status of an individual animal selected at random is independent of the disease status of all the other animals chosen beforehand. It also means that transmission of infection does not occur during the sampling period.

Provided we can satisfy the assumptions of the binomial process, once two of the values *n*, *p* or *x* are known the third one can be estimated from the following distributions (Fig. 22):

– binomial distribution is used to model the number of successes x:

$$x = Binomial(n, p)$$

– Beta distribution is used to model the probability of success p:

$$p = Beta(x+1, n-x+1)$$

– negative binomial distribution is used to model the number of trials n, undertaken before x successes have occurred:

$$n = x + Negative\ binomial(x, p)$$

[10] The general reference for this chapter is Vose D. Risk Analysis, A Quantitative Guide. John Wiley & Sons Chichester, 2000

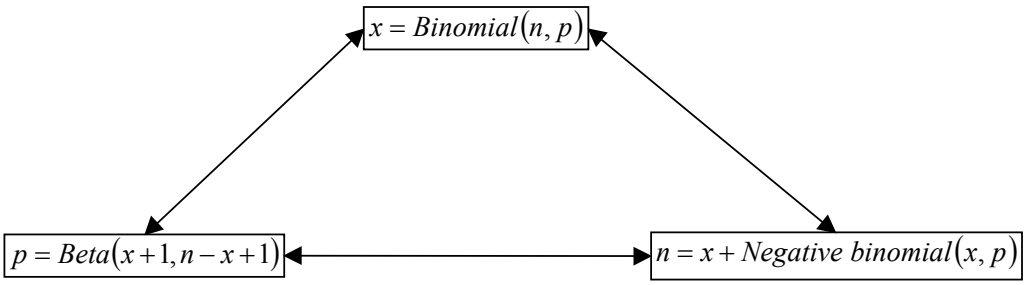
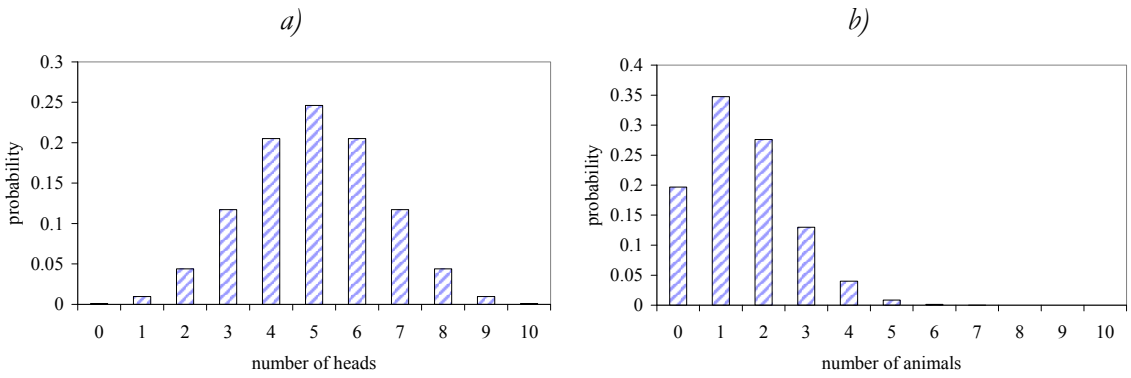


Figure 22
Three distributions used to model the parameters of a binomial process

Binomial distribution

$$x = Binomial(n, p)$$

The binomial distribution is used to model the variation in the number of successes (x), that occur when n trials, each with a probability p, of success are undertaken. For example, if we toss a coin ten times, how many heads are we likely to obtain? If the coin is fair there will be an equal chance of obtaining either a head or a tail on each toss, that is, a probability of 0.5. By repeating the experiment a number of times and recording how many heads there are in each group of ten tosses, we will get a pretty good idea of the probability of a certain number of heads arising from 10 tosses. Rather than undertaking a large number of experiments we can use a binomial distribution function to model a distribution of x successes in n trials (Fig. 23a). We can extend this idea to choosing animals from an infected herd where we would like to determine the likely number of infected animals in the group selected (Figure 23b).



a) a binomial distribution of the probability of obtaining *x* heads when tossing a coin 10 (*n*) times. Tossing a coin, Binomial (10,0.5)

b) selecting *x* diseased animals when selecting 10 (*n*) animals from an infected herd. Selecting a group of animals from an infected herd with 15% prevalence, Binomial (10,0.15)

Figure 23

Since the binomial coefficient, $\frac{n!}{x!(n-x)!}$, involves calculating factorials, a computational limit is imposed by various software packages. For example, *n* must be less than or equal to 32,767 in the *Binomial(n,p)* function in @RISK. When the number of trials is larger than this the function returns an error message. If *n* is very large and *p* is very small, the mean *np*, of the binomial distribution will be approximately equal to the variance *npq*, that is $np \approx npq$. Since the mean and the variance of a Poisson distribution are equal we can use

the Poisson distribution as a convenient approximation, $Binomial(n, p) \approx Poisson(np)$. Where *n* is very large and *p* is neither very small nor very large, a normal approximation can be used, $Binomial(n, p) \approx Normal(np, \sqrt{npq})$.

Beta distribution

$$p = Beta(\alpha_1, \alpha_2)$$

The Beta distribution is characterised by two shape parameters, α_1 and α_2, which do not have any particular intuitive meaning in themselves other than to define the shape of the distribution. Since its domain is between zero and one inclusive, the Beta distribution provides a convenient way of modelling uncertainty about the parameter p, the probability of success, in a binomial process.

According to an empirical definition of probability (Chapter 2) the exact value of a probability can never actually be observed unless an infinite number of trials are undertaken. However, we can be increasingly certain of what its 'true' value is by undertaking a number of trials and observing how many successes there are. For example, if nine out of ten rams, known to be infected with *Brucella ovis* were positive to a serological test, we could estimate that the sensitivity of the test is 90%, that is, the probability that the test is positive given that a ram is infected $P(T^+ | D^+)$. But, how confident could we be that this is a reasonable estimate, particularly considering that there were only ten rams in the trial? We could use the Beta distribution to model the uncertainty surrounding the parameter p, by replacing α_1 with (x+1) where x is the number of successes and α_2 with (n-x+1) where n is the number of trials:

$$p = Beta(x+1, n-x+1) = Beta(9+1, 10-9+1)$$

This distribution is actually the posterior distribution that arises from using a particular Beta distribution (*Beta* (1,1)) as a non-informative conjugate prior to a binomial likelihood function in Bayesian inference (Chapter 6).

Figure 24 depicts a distribution modelling the uncertain parameter *p*, representing test sensitivity in our example. It also demonstrates that as more information is gathered by testing more animals we would be increasingly confident that the 'true' sensitivity is, in fact, around 90%. In the end there is always a trade-off between obtaining a reasonable level of confidence and the cost and effort needed to acquire additional information.

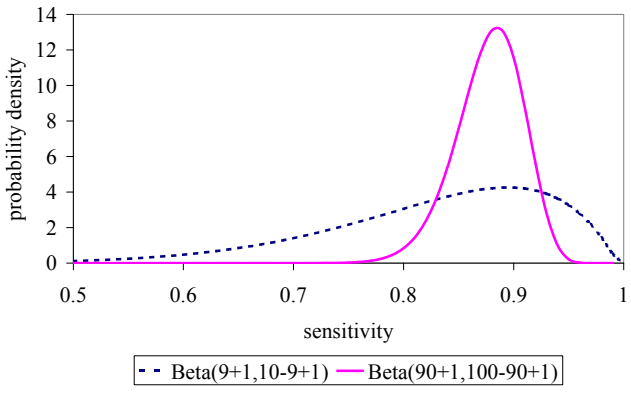


Figure 24
Using the Beta distribution function to model an uncertain parameter (*p*) of a binomial distribution. In this case *p* represents test sensitivity

Table XII lists some useful applications of the Beta distribution in animal health risk analyses.

Table XII
Some applications of the Beta distribution (*Beta*(*x*+1, *n*-*x*+1))

Application	*n*	*X*
Test sensitivity	Number of diseased animals	Number of diseased animals that are test positive
Test specificity	Number of non-diseased animals	Number of non-diseased animals that test negative
Prevalence	Number of animals, herd, flocks etc	Number of diseased animals, herd, flocks etc.
Estimating a probability when there are no 'successes', e.g. estimating prevalence when none of the animals sampled are found to be infected	Number of animals sample, units imported, etc.	Zero

Negative binomial distribution

$$n = s + Negbin(s, p)$$

The negative binomial distribution is characterised by two parameters, the number of successes (x) and the probability of a success (p). The outcome is expressed as the number of failures there will be before x successes have occurred. We can use this distribution to estimate the number of trials (n) that are likely to be undertaken before x successes have occurred by adding the number of successes (x) and the number of failures (Negbin(x, p)). For example, we might like to know how many animals we could select from an infected herd, which has a disease prevalence of 10%, before including an infected animal in the group. In this case the formula is Negbin(1,0.1) since we are interested in the number of uninfected animals ('failures') before obtaining the first infected one (Fig. 25). If we were interested in the number of animals we need to select to include one infected animal the formula would be $1 + Negbin(1,0.1)$.

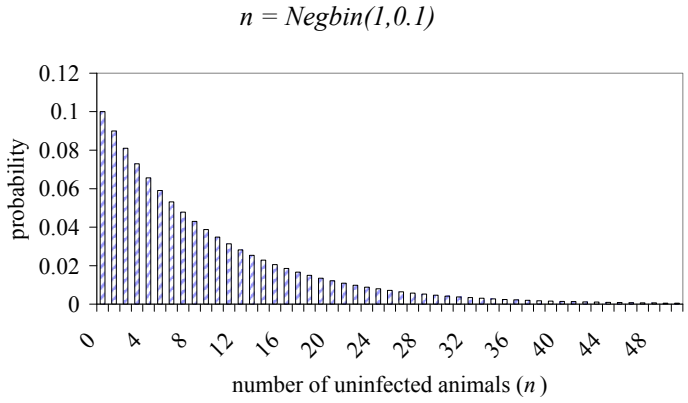


Figure 25
A negative binomial distribution of the number of uninfected animals likely to be selected from a herd with a disease prevalence of 10% before including an infected animal in the group

Distributions used to model a Poisson process

A Poisson process has four characteristics:

- it models the number of events (x) that occur in an interval (t) of space or time
- it is characterised by one parameter lambda (λ), the average number of events per unit interval of space or time
- there is a constant and continuous probability of an event occurring per unit interval
- the number of events that occur in any one interval is independent of the number that occur in any other interval. It does not matter how far apart the events are in space or time. For instance, an event may have only just been observed or there may have been a considerable interval between them.

The interval (*t*) is measured in either space (per litre, per kilogram, per kilometre, etc) or time (per second, per hour, per day, per year, etc). The mean number of events per *unit* interval (λ) can also be expressed as $\frac{1}{\beta}$ where β is the mean interval between events.

It should be noted that there are some differences in the terminology used in Excel and @RISK. The Poisson function in Excel is expressed as *POISSON*(*x*, expected value,0)[11], where *x* is the number of events and the expected value is the expected number of events in the interval under study (*t*), which is calculated as either $\lambda \times t$ or $\frac{t}{\beta}$. In @RISK the Poisson function is expressed as *Poisson*(*lambda*) where lambda actually equals either $\lambda \times t$ or $\frac{t}{\beta}$, not just simply λ, unless of course *t* equals one.

Three distributions are used to model the Poisson process (Fig. 26):

a) Poisson distribution is used to model the number of events (x), in an interval (t):

$$x = Poisson(\lambda \times t)$$

b) Gamma distribution is used to model:

- a distribution of λ, the average number of events per unit interval

$$\lambda = Gamma\left(x, \frac{1}{t}\right)$$

- a distribution of the time until the next x events have occurred

$$t_x = Gamma\left(x, \frac{1}{\lambda}\right)$$

c) exponential distribution is used to model:

- a distribution of the time until the next event has occurred

$$t_{next} = Expon\left(\frac{1}{\lambda}\right) = Gamma\left(1, \frac{1}{\lambda}\right)$$

- a distribution of a lower bound for β, the mean interval between events, when no events have been observed:

$$\beta_{min} = \frac{1}{Expon\left(\frac{1}{t}\right)}$$

[11] '0' is a switch that determines whether the result is returned as a Poisson probability mass function, that is, the probability that number of events will be exactly equal to *x*. If it is set to '1' it returns the cumulative Poisson probability, that is, probability that the number of events will be between zero and *x* inclusive

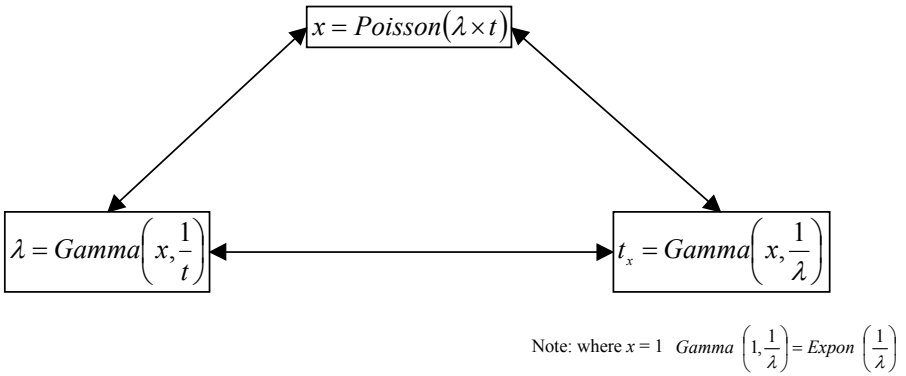


Figure 26
Distributions used to model the parameters of a Poisson process

Poisson distribution

$x = Poisson(\lambda \times t)$, or alternatively, this equation may be expressed as $x = Poisson\left(\frac{t}{\beta}\right)$

The Poisson distribution is used to model the variability in the number of events (α) in an interval (t). Examples for which the Poisson process provides a very good approximation for estimating the number of events in an interval are the number of bacteria per litre of water, the number of outbreaks of a disease per year and the number of earthquakes per decade. Although, theoretically, there can be anything between zero and an infinite number of events in a specific interval, this is almost never a restriction in practice. For example, if there are four *Giardia* cysts per millilitre of water on average, Figure 27 demonstrates that the probability of more than 20 cysts is vanishingly small.

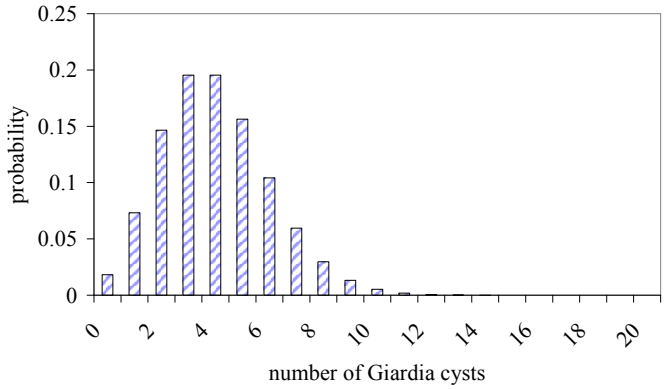


Figure 27
A Poisson probability distribution of the number of *Giardia* cysts in a volume (t) of water where λ = 4 cysts/ml, t = 1ml

We can estimate the number of disease outbreaks expected during the next six months, given that historical information indicates an outbreak occurs on average every 24 months. In this situation the mean interval between events (β) is 24 months and λ is 1/24 (0.04) outbreaks per month. The number of outbreaks in the next 6 months is then modelled as *Poisson*(6/24) as shown in Figure 28.

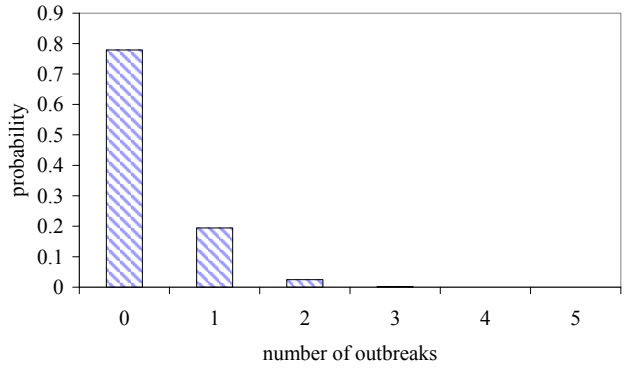


Figure 28
A Poisson probability distribution of the number of disease outbreaks expected during the next time interval (t) = six months and the mean interval between events (β) =24 months: $Poisson\left(\frac{6}{24}\right)$ = Poisson(0.25)

Gamma distribution

Estimating λ, the average number of events per unit interval

$$\lambda = Gamma\left(x, \frac{1}{t}\right)$$

The Gamma distribution can be used to model uncertainty about the parameter λ, the mean number of events per *unit* interval in a Poisson process. Just as with the binomial probability (p) if we adopt an empirical definition of probability (Chapter 2), λ can never actually be observed unless our observations extend over an infinite interval. However, we can be increasingly confident of what its true value is by collecting more data. For example, if we observed three disease outbreaks over a period of 18 months we could estimate that the mean number of outbreaks per month is 0.17, provided, of course, that we can reasonably assume a Poisson process applies. In other words, has there been a continuous and constant probability of an outbreak arising throughout the period of observation and are the outbreaks we observed independent of one another? If we are satisfied that a Poisson process is applicable, how confident can we be that this is a reasonable estimate? We can use the Gamma distribution to model the uncertainty surrounding the parameter λ as shown in Figure 29. If we extended the period of observation and found there were seven outbreaks in 42 months we would be increasingly confident that the 'true' mean number of outbreaks per month is 0.17.

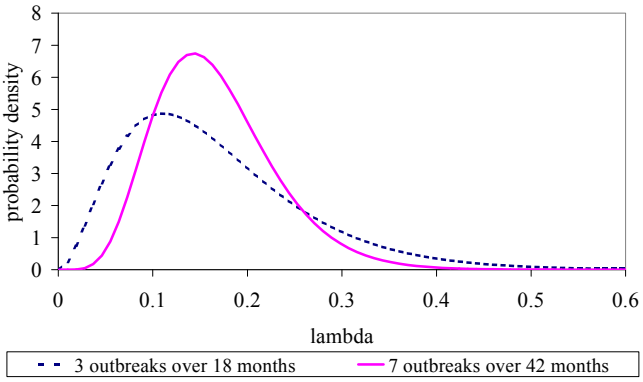


Figure 29
Estimates of the mean number of disease outbreaks per month (λ) using the Gamma distribution, $Gamma\left(x, \frac{1}{t}\right)$, where x = the number of outbreaks observed, t = the period of observation. A Poisson process is assumed to apply

The $Gamma\left(x, \frac{1}{t}\right)$ distribution is, in fact, the posterior distribution obtained by assuming an uninformed prior with a Poisson likelihood function (Chapter 6). If we can reasonably describe our prior opinion with a $Gamma(a,b)$ distribution and we then observe x events in an interval t, the posterior distribution for λ is given by:

$$\lambda = Gamma\left(a + x, \frac{b}{1 + b \times t}\right)$$

where: a = the number of events and
b = the mean interval between events.

Estimating the time until the next x events have occurred

$t_\alpha = Gamma\left(\alpha, \frac{1}{\lambda}\right)$ or alternatively, this equation may be expressed as $t_\alpha = Gamma(\alpha, \beta)$

The Gamma distribution can be used to model the variation in the time until the next x events have occurred. If the mean interval between outbreaks of a particular disease (β) is 24 months, we can estimate the length of time that is likely to elapse before we observe α disease outbreaks. Figure 30 plots a distribution of the length of time that is likely to elapse before we experience four disease outbreaks.

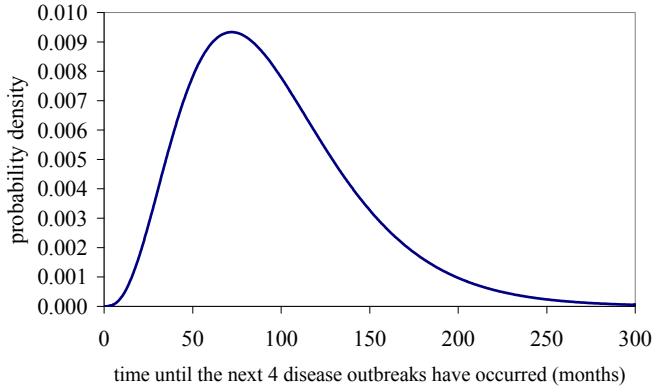


Figure 30
The length of time that is likely to elapse before four disease outbreaks have occurred if the mean interval between events (β) = 24 months, t_4 = *Gamma*(4, 24). This may also be expressed in terms of the mean number of events per interval where $\lambda = \frac{1}{24} = 0.042$ and t_4 = *Gamma*(4, $\frac{1}{0.042}$)

Exponential distribution

Estimating the time until the next event

$$t_{next} = Expon\left(\frac{1}{\lambda}\right) = Gamma\left(1, \frac{1}{\lambda}\right)$$

Alternatively, these equations may be expressed as $t_{next} = Expon(\beta) = Gamma(1, \beta)$.

Both the exponential and Gamma distributions can be used to model the variation in the time until the next event (the time between events). The *Expon*(β) is equivalent to a *Gamma*(α, β) where the number of events (α) is equal to 1. For example, if we can assume a

Poisson process and estimate that the mean interval between outbreaks of a particular disease is 24 months, we could define a distribution of how long it is likely to be before we can expect the next outbreak (Fig. 31).

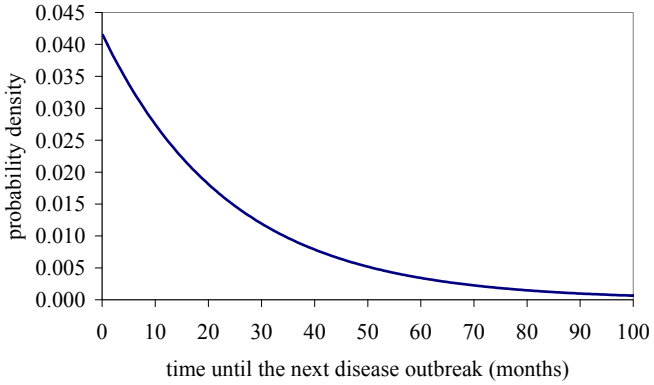


Figure 31
The time until the next disease outbreak if the mean interval between events (β) = 24 months, t_{next} = *Expon*(24) or alternatively t_{next} = *Gamma*(1,24)

Estimating a lower bound for β, the mean interval between events, when no events have been observed

$$\beta = \frac{1}{Expon\left(\frac{1}{t}\right)}$$

The exponential distribution can be used to estimate the lower bound for the mean interval between events (β) given that no events have been observed during an interval (t). There are several important assumptions in this estimate. It is assumed that the event is possible, that it occurs for the first time immediately after the last observation and that it follows a Poisson process. Since we are dealing with only one event, we can estimate λ as $Gamma\left(1,\frac{1}{t}\right) = Expon\left(\frac{1}{t}\right)$ and since $\beta = \frac{1}{\lambda}$ it follows that $\beta = \frac{1}{Expon\left(\frac{1}{t}\right)}$.

The assumption of a Poisson process is particularly important and should be considered when determining an appropriate interval. For example, if foot and mouth disease (FMD) has not occurred in a particular country during the last ten years, can it be reasonably assumed that there has been a continuous and constant probability of an outbreak arising during that period? It may turn out that, as a result of increasing political instability over the last three years, border controls are not as effective as they once were and there is an increasing incidence of animals and meat being smuggled from a neighbouring country that has regular outbreaks of FMD. Obviously, the probability of an outbreak may have changed significantly. For this reason it may be inappropriate to estimate β based on a ten year interval and perhaps a two or three year interval should be chosen.

Estimating the probability of at least one event in an interval

$$P(x \geq 1) = 1 - EXP\left(\frac{-t}{\beta}\right)$$

We can use the exponential function, for example the *EXP*() function[12], to estimate the probability of at least one event in an interval. The probability that no events will occur in

12 *Note*: this is an Excel function and should not be confused with the @RISK *Expon*() function

an interval *t*, is $EXP\left(\frac{-t}{\beta}\right)$. This is equivalent to $e^{-t/\beta}$ and should not be confused with the exponential function in @RISK. It follows that the probability of at least one event in an interval is $1-EXP\left(\frac{-t}{\beta}\right)$. Alternatively, this equation may be expressed in Excel as:

$$1-EXP(-t\times\lambda), \text{ or } 1-POISSON\left(0,\frac{t}{\beta},0\right), \text{ or } 1-POISSON(0,t\times\lambda,0)$$

The following example calculates the probability of at least one disease outbreak during the next six months given that the mean interval between disease outbreaks (β) is 24 months:

$$P(x\geq 1)=1-Exp\left(\frac{-6}{24}\right)=0.22$$

This could also be determined by summing the individual probabilities for 1, 2, 3, ..., *n* events in Figure 28.

Cumulative distribution

Cumul (minimum, maximum,{x_i},{p_i}), where i = 1 to n

The cumulative distribution can be used to convert a set of data into an empirical distribution provided the data are continuous and cover a reasonable range. For example, Melville and colleagues, 1996[13] reported the duration of viraemia in cattle naturally infected with bluetongue virus (Table XIII). Such data can be modelled using either the cumulative (Fig. 32) or histogram (Fig. 33) distribution function:

Cumul (minimum, maximum,{x_i},{p_i}) = Cumul (0,13,{B3:B14},{E3:E14})
Histogrm (minimum, maximum,{p_i}) = Histogrm (0,13,{D3:D14})

Table XIII
The duration of viraemia in cattle naturally infected with bluetongue virus (adapted from Melville *et al.*, 1996[1])

	A	B	C	D	E
1	Weeks		Number of cattle	Histogram probability	Cumulative probability
2	from	to			
3	0	1	53	0.111	0.111
4	1	2	124	0.260	0.371
5	2	3	148	0.310	0.681
6	3	4	83	0.174	0.855
7	4	5	36	0.075	0.931
8	5	6	14	0.029	0.960
9	6	7	10	0.021	0.981
10	7	8	5	0.010	0.992
11	8	9	2	0.004	0.996
12	9	10	0	0.000	0.996
13	10	11	0	0.000	0.996
14	11	12	2	0.004	1.000

[13] Melville L.F., Weir P., Harmsen M., Walsh S., Hunt N.T. & Daniels P.D. (1996). – Characteristics of naturally occurring bluetongue viral infections of cattle. *In* Bluetongue Disease in Southeast Asia and the Pacific (St George TD, Peng Kegao, eds). Proceedings No. 66, ACIAR, Canberra, 245-250.

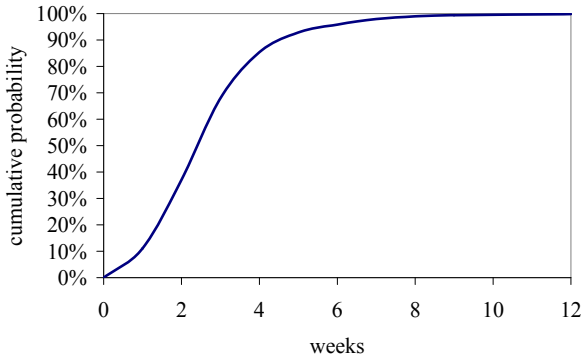


Figure 32
A cumulative probability distribution of the duration of viraemia in cattle naturally infected with bluetongue virus

Discrete and discrete uniform distributions

Discrete($\{x_i\},\{p_i\}$), where $i = 1$ to n
Duniform($\{x_i\}$), where $i = 1$ to n

The discrete distribution has no theoretical basis and can be used as a general distribution function to describe a variable that can have one of several discrete values (x_i), each with a weight (p_i) which specifies the value's probability of occurrence. The probabilities (p_i) do not have to add up to 1 as the @RISK function automatically normalises them. It can be used to model a posterior distribution in a Bayesian inference calculation, to model expert opinion (Chapter 6) where there are divergent views, or to construct a composite distribution. The discrete uniform distribution is a particular form of the discrete distribution that can have one of several discrete values (x_i), each with an equal probability of occurrence.

General distribution

General($\{x_i\},\{p_i\}$), where $i = 1$ to n

The general distribution produces a generalised probability curve based on a density curve created using the specified x, p pairs. As with the discrete distribution, the probability densities (p_i) do not have to add up to 1 as the @RISK function automatically normalises them. It can be used to model a posterior distribution in a Bayesian inference calculation where the parameter being estimated is continuous and to produce a fairly detailed distribution that reflects an expert's opinion (Chapter 6).

Histogram distribution

Histogrm(minimum,maximum, $\{p_i\}$), where i = 1 to n

The histogram distribution is closely related to the general distribution and is used for continuous data to specify a distribution with a specified number of equal length classes. The range, defined by the minimum and maximum values, is divided into n classes, with each class having a probability p_i, of occurrence. As with the discrete and general distributions the probabilities (p_i) do not have to add up to one as the @RISK function automatically normalises them. It is useful for replicating the distribution shape of a set of data.

As discussed in the example of a cumulative distribution, the data reported by Melville and colleagues (1996) on the duration of viraemia in cattle naturally infected with bluetongue virus may also be modelled using a histogram distribution (Fig. 33).

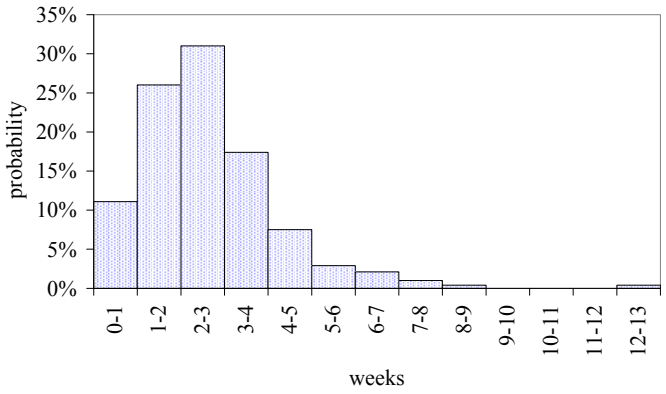


Figure 33
A histogram probability distribution of the duration of viraemia in cattle naturally infected with bluetongue virus, *Histogrm* (minimum, maximum,{p_i}) = *Histogrm* (0,13,{D3:D14})

Hypergeometric distribution

x = *Hypergeo*(n, D, M)

A hypergeometric process is characterised by three parameters; the sample size (n), the number of individuals with the characteristic of interest (D), and the population size (M). The outcome is expressed as number of successes (x) in the sample. As discussed in Chapter 5, the probability of success in a hypergeometric process changes each time an individual is selected and removed from the population. It is effectively modelling sampling without replacement. For example, if the herd consists of 100 animals (M = 100) and there are five infected animals (D = 5), $\frac{D}{M}$ is initially 0.05. If the first animal selected is infected then $\frac{D}{M} = \frac{4}{99} = 0.04$ whereas, if it is uninfected, $\frac{D}{M} = \frac{5}{99} = 0.051$. As a result the probability, measured by $\frac{D}{M}$, changes depending on whether the previous animal was infected or not. That is, p is no longer independent of the outcome of the previous trial. This is in contrast to the binomial process where the probability of success (p) remains constant and the result of each trial is independent of the results of any of the previous trials. As a result, the binomial process is effectively modelling sampling with replacement. When the sample size (n) is small, (less than one tenth) compared to the population size (M), the binomial distribution closely approximates the hypergeometric distribution.

Figure 34 provides a series of probability distributions of the number of diseased animals (x) in a sample of size n, selected from a herd of size M, where there are a number of infected animals (D). It compares the hypergeometric to the binomial distribution for decreasing values of the ratio of herd size to the number of animals in the group selected (M/n). The herd size (M) and number of infected animals (D) are held constant throughout the series. It can be seen that once the ratio of M/n falls below ten there is an increasing disparity between the two distributions. In fact, as M/n approaches 1 the binomial distribution predicts that there is a reasonable chance of having more infected animals in the group selected than actually exist in the herd. Depending on the particular

situation being modelled, these differences may not be very important. However, they should be kept in mind and the impact of low values of M/n investigated.

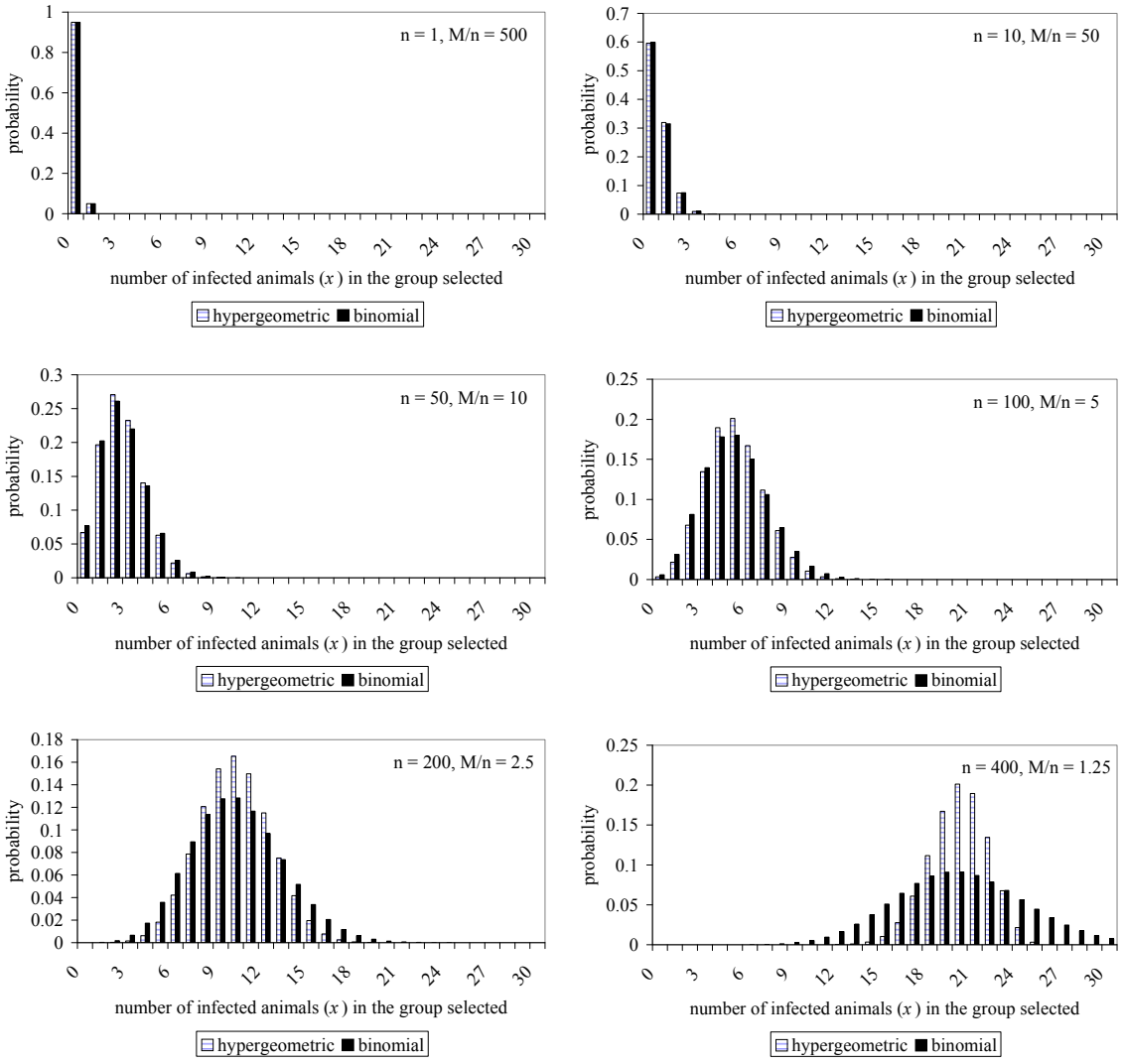


Figure 34
A series of probability distributions comparing the hypergeometric and binomial estimates of the number of infected animals (*x*) in a group (*n*) selected from a herd (*M*) with a number of infected animals (*D*).
The variables for herd size (*M*) and infected animals (*D*) are fixed at 500 and 25 respectively. For the hypergeometric distribution, $x = Hypergeo(n,D,M)$, while for the binomial distribution prevalence is calculated as D/M and $x = Binomial(n,D/M)$

Lognormal distribution

Lognorm(μ, σ)
Tlognorm(*μ, σ, minimum, maximum*) – truncated lognormal distribution

The lognormal distribution is characterised by two parameters; the mean (μ) and standard deviation (σ). It is an unbounded, continuous distribution extending from zero to plus infinity that is used to model a variable (x) the natural log of which (ln(x)) is normally distributed. The parameters μ and σ are the actual mean and standard deviation of the lognormal distribution. Alternatively, the lognormal distribution may be specified by the mean and standard deviation of the normal distribution of ln(x). The lognormal

distribution is one of the most widely used distributions in probabilistic risk assessment. It often provides a good representation for data that extend from zero and are positively skewed. That is, data which have a longer right hand tail, such as herd and flock sizes, weight of processed ham, carcass weights and disease incubation periods. In addition, the outputs from computer simulations involving the multiplication of two or more distributions are often distributed lognormally.

Since the lognormal distribution extends from zero to plus infinity we may need to constrain it to avoid implausible values. For example, we could model the *incubation* period for a particular disease as *Lognorm*(5,3), as has been done in Figure 35. If this disease had a minimum and maximum incubation period of two and fourteen days respectively, there is a reasonable chance that some random samples drawn from the distribution would fall outside this range. For this reason, we need to truncate the distribution so that sensible values are sampled .

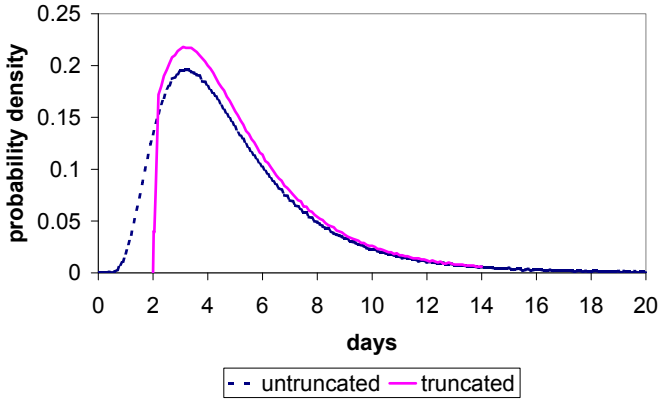


Figure 35
A lognormal distribution modelling a disease incubation period.
The untruncated distribution = *Lognorm* (5,3), while the truncated distribution with a minimum of two days and a maximum of fourteen days = *Tlognorm* (5,3,2,14)

Normal distribution

Normal(μ, σ)
Tnormal(μ, σ, *minimum, maximum*) – truncated normal distribution

The normal distribution is characterised by two parameters; the mean (μ) and standard deviation (σ). It is an unbounded continuous distribution that extends from minus infinity to plus infinity and has a bell shaped curve (Fig. 36). It is symmetrical about its mean with 99.9% of its values lying within ± 3 standard deviation of the mean. Many naturally occurring variables such as weight, height, viral titre in tissues, physiological characteristics, pH of tissues and fluids, and milk and egg production are normally distributed. Others are normally distributed following some transformation of the data; for example, a log transformation of a set of data on the incubation period of a disease. The normal distribution has an extensive variety of applications ranging from the central limit theorem (Chapter 3) to statistical theory where it is widely used in statistical inference and hypothesis testing.

Since the normal distribution is unbounded, we may need to constrain it if we are to avoid implausible values. Figure 36 illustrates the impact of different values of the coefficient of variation, $\left(\frac{s\tan dard\ \ deviation}{mean}\right)$, on the spread of the distribution and the likelihood of randomly sampling 'unrealistic' values. It is certainly worthwhile checking, particularly if the

coefficient of variation is reasonably large. If it is likely that unrealistic values will be sampled we need to truncate the distribution using the *Tnormal(μ,σ,minimum, maximum)* function where *minimum* and *maximum* define the minimum and maximum of the plausible range of values.

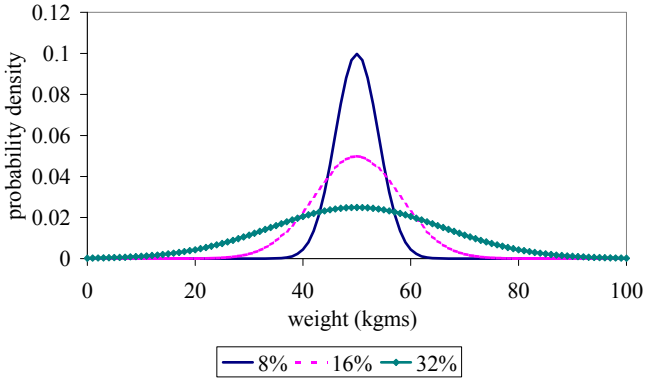


Figure 36
Three normal probability plots.
All with the same mean (μ = 50), demonstrating the impact that different values of the coefficient of variation have on the spread of the distribution and the likelihood of randomly sampling 'unrealistic' values

If the normal distribution is used to model a discrete variable, such as the number of animals in a herd, we may need to consider correcting for continuity. This is easily achieved by applying a *ROUND*(…,0) function to the distribution *ROUND*(*Normal*(μ, σ),0). Alternatively, suppose we set a spreadsheet up to calculate the probability of obtaining a herd of x animals. Rather than simply calculating the probability using the probability mass function, *NORMDIST*(x,μ,σ,0), we need to add and subtract 0.5 to each value of x. Then we can use the cumulative density function to calculate the probability associated with the interval bounded by $x \pm 0.5$, that is, *NORMDIST*(x+0.5,μ,σ,1)- *NORMDIST*(x-0.5,μ,σ,1).

PERT (Beta PERT) distribution

PERT(*minimum, most likely, maximum*)

A PERT distribution is a modification to the Beta distribution that enables a continuous smooth distribution to be defined by its minimum, most likely and maximum values:

$$PERT(a,b,c) = Beta(\alpha_1,\alpha_2)\times(c-a)+a$$

where: a = minimum
b = most likely
c = maximum

$$\alpha_1 = \frac{(\mu-a)\times(2b-a-c)}{(b-\mu)\times(c-a)}$$

$$\alpha_2 = \frac{\alpha_1\times(c-\mu)}{(\mu-a)}$$

$$\mu \text{ (mean)} = \frac{a+4b+c}{6}$$

The PERT distribution provides a more natural shape than the triangular distribution and is not influenced as much by the extreme (minimum and maximum) values, particularly when the distribution is skewed (Fig. 38). It is a useful distribution for modelling expert opinion (Chapter 6).

In the standard PERT distribution a weight of four is applied to the mean so that the mean is four times more sensitive to the most likely value than it is to the minimum or maximum values. This can be manipulated by incorporating a weighting factor (γ) into the formula calculating the mean enabling various shapes to be generated using the same values for the minimum, most likely and maximum:

$$\mu = \frac{a + \gamma \times b + c}{\gamma + 2}, \text{ where } \gamma \text{ is the weight.}$$

Figure 37 shows an example of the age at which chickens are likely to become infected with infectious bursal disease (IBD) virus prior to being slaughtered at 49 days of age. Initially there was a great deal of uncertainty, so a uniform distribution, *Uniform*(1,49) was used[14]. Later some information became available indicating that they were most likely to become infected around three weeks of age. This was modelled as a *PERT*(1,21,49). After further enquiries the estimate was refined to 'most chickens become infected between 14 and 28 days of age'. This was interpreted as 90% of chickens being likely to become infected during this period[15]. A modified PERT, with a weight of 28.2, determined by using the Solver function in Excel (Table XIV), was used to model this new information. The same estimates for the minimum, most likely and maximum values were used as in the original PERT distribution (Fig. 37).

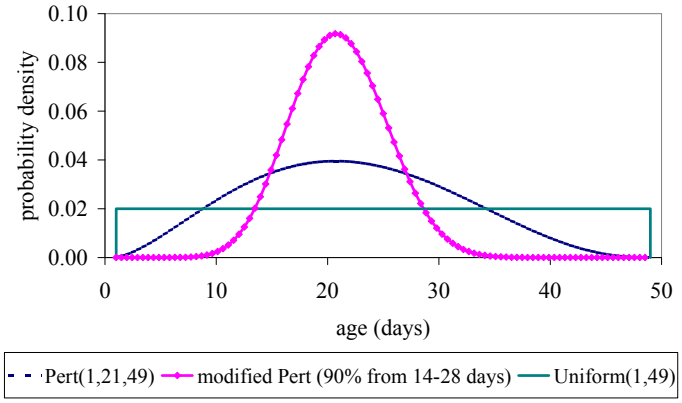


Figure 37
A comparison of a uniform distribution, a standard PERT distribution and a modified PERT distribution of the age when a chicken is likely to become infected with IBD virus prior to slaughter at 49 days of age

[14] MAF Regulatory Authority. Import Risk Analysis: chicken meat and chicken meat products; Bernard Matthews Foods Ltd turkey meat preparations from the United Kingdom. Wellington, New Zealand, 1999

[15] MAF Regulatory Authority. Revised Quantitative Risk Analysis on Chicken Meat from the United States. Wellington, New Zealand, 2000

Table XIV
A spreadsheet model to calculate the weight (γ) for a modified PERT distribution using the Solver function in Excel
The weight (γ) is set as the cell to change (B8) and the target is set to cell B9 with a value equal to 0.9, which represents the area under the curve that falls between fourteen and twenty-eight days. The *BETADIST* function calculates the cumulative Beta probability density

	A	B
1	**Inputs**	
2	Minimum (*a*)	1
3	Most likely (*b*)	21
4	Maximum (*c*)	49
	Calculated values	
5	Mean (μ)	$\mu = \frac{a + \gamma \times b + c}{\gamma + 2}$
6	Alpha 1	$\alpha_1 = \frac{(\mu - a) \times (2b - a - c)}{(b - \mu) \times (c - a)}$
7	Alpha 2	$\alpha_2 = \frac{\alpha 1 \times (c - \mu)}{(\mu - a)}$
8	Cell to change (γ = weight)	28.2
9	Target cell (area under the curve)	*BETADIST*(28, α_1, α_2, *a*, *c*)- *BETADIST*(14, α_1, α_2, *a*, *c*)

Triangular distribution

Triang(minimum, most likely, maximum)

The triangular distribution is a continuous distribution that has been extensively used to model expert opinion. It is defined by minimum, most likely and maximum values. Its main drawback is its unnatural shape, which rarely, if ever, provides a reasonable description of a biological process. It tends to overemphasise the tails and underestimate the shoulders of a distribution compared to more naturally curved distributions, such as the PERT (Fig. 38).

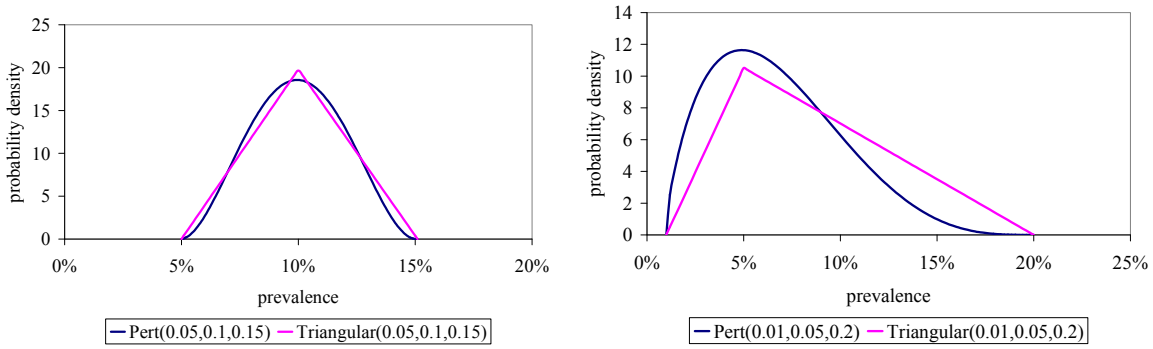


Figure 38
Comparing a triangular and PERT distribution

Uniform (rectangular) distribution

Uniform(minimum, maximum)

The uniform distribution, which is also known as a rectangular distribution, is a simple continuous distribution that only requires an assumption about the range of possible values. All values within the range have an equal probability of occurrence. It is used mostly when there is very little information available, other than a range of possible values. For instance, we might have some information that IBD virus has been isolated from muscle tissue between two and six days following infection[16]. This information can be usefully incorporated into a model as a *Uniform*(2,6) distribution as shown in Figure 39.

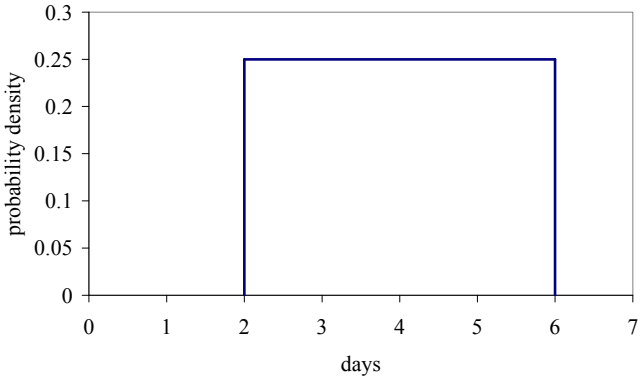


Figure 39
A uniform distribution of the duration of tissue infectivity in chicken muscle following infection with IBD virus

[16] MAF Regulatory Authority. Import Risk Analysis: chicken meat and chicken meat products; Bernard Matthews Foods Ltd turkey meat preparations from the United Kingdom. Wellington, New Zealand, 1999

Chapter 5
Probability processes and calculations

Expressing probability: the binomial versus the hypergeometric process

The outcome of interest in many quantitative risk assessments is a binary response. That is, there are only two possible outcomes. For example, an animal is infected or it is not; a test is positive or it is not; a disease outbreak occurs or it does not. In these circumstances probability may be expressed in one of two ways. Using disease prevalence as an example we might estimate that the probability that an animal from a herd of size *M* is infected is 0.05, or we could assume that we know exactly how many infected animals (*D*) there are in the herd, in which case the prevalence is exactly $\frac{D}{M}$. What is the difference? In the first case we assigned a probability, $(p = 0.05)$, that a randomly selected animal within the herd will be infected. In this case a prevalence of *p* does not mean that there are exactly $p \times M$ infected animals. Rather, it means that each animal has the same probability (*p*) of being infected and that we would expect, on average, there to be $p \times M$ infected animals in the herd. This situation is akin to each animal within a herd being selected from a hypothetical 'super-herd' with an infinite number of animals of which proportion *p* are infected. In such a herd, the number of infected animals (*D*) is binomially distributed, $D = Binomial(M, p)$. Similarly, if a sample of *n* animals is randomly chosen, the number of infected animals (*x*) in the sample and the probability (*P*) that the sample contains *x* infected animals are both binomially distributed:

$$x = Binomial(n, p)$$
$$P(X = x) = BINOMDIST(x, n, p, 0)$$

A fundamental property of the binomial process is that each trial has the same probability of success. That is, the probability (*p*) that an animal is infected remains constant for each animal. This means that the disease status of an individual animal selected at random is independent of the disease status of all the other animals chosen beforehand. As a result the binomial process is effectively modelling sampling with replacement. While we would not obviously do this in practice, it is reasonable to assume that the probability remains constant from animal to animal provided the population from which we are sampling from in relation to the sample size is large. As a guide, if the size of the population is at least ten times the sample size such an assumption is adequate.

If we model prevalence as a binomial process (Chapters 3 and 4), then in herds where the expected number of infected animals ($p \times M$) is low, there is a reasonable chance that some herds may not contain any infected animals (Fig. 40). This may seem unreasonable at first, since we have stated that the herd has a prevalence of infection equal to *p*, and so it is natural to assume that the herd actually contains at least one infected animal. Rather than thinking of infected and uninfected herds we could, instead, think of risk herds and non-risk herds. Risk herds are the proportion of herds where it is expected that, on average, there will be $p \times M$ infected animals. Non-risk herds, on the other hand, are the proportion of herds where the prevalence of infection (*p*) is zero and, as a result, the expected number of infected animals, ($p \times M$) is zero.

If we assume we know exactly how many infected animals there are in the herd the binomial process no longer applies. In this case the number of infected animals (*x*) in a

sample of animals (n) chosen at random, and the probability (P) that the sample contains x infected animals are both distributed hypergeometrically:

$$x = Hypergeo(n, D, M)$$
$$P(X = x) = HYPGEOMDIST(x, n, D, M)$$

The hypergeometric process models sampling without replacement and enables us to relax the assumption of the binomial process that the probability (p) is constant. Each time an animal is chosen for inclusion in a sample the probability that the next animal selected will be infected changes. That is, the probability of selecting an infected animal does not remain constant. While modelling prevalence as a hypergeometric process has a certain appeal there are, unfortunately, some significant drawbacks. We are rarely, if ever, in a position to know precisely how many infected animals there are in a herd. More importantly though, as discussed below, because the probability does not remain constant, the mathematics involved quickly become unwieldy.

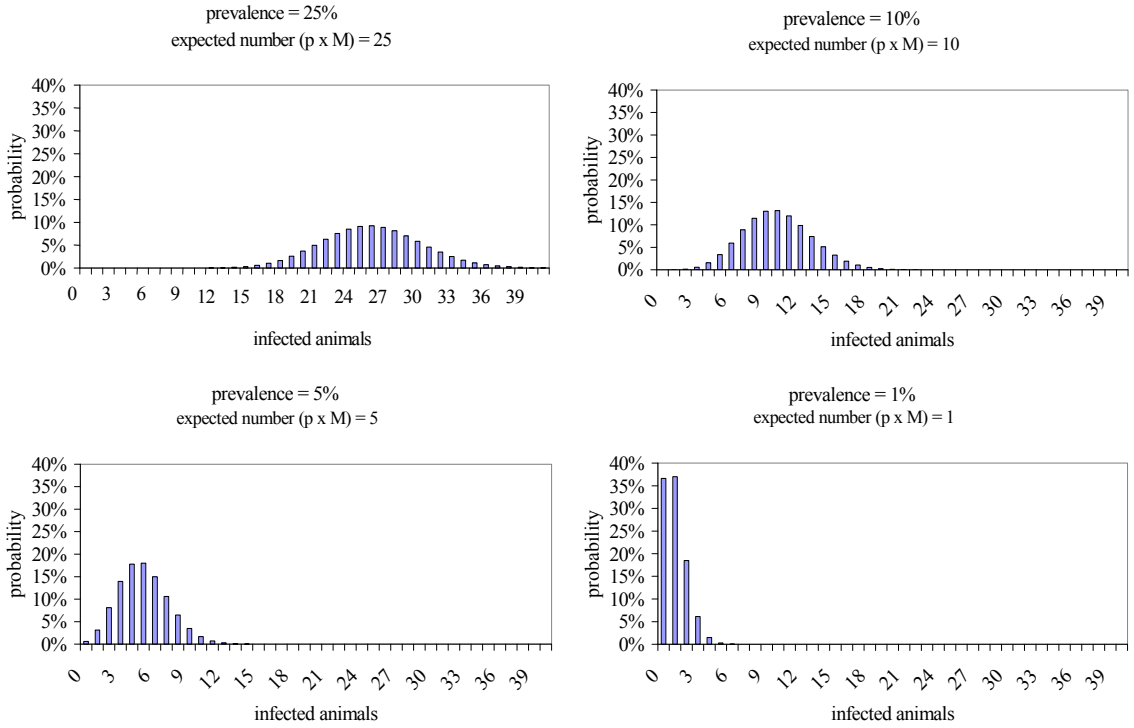


Figure 40
A series of binomial probability distributions modelling the probability of (D) infected animals in a herd of 100 animals with a prevalence of infection ranging from 1% to 25% where $D = Binomial(M, p)$

Binomial probability calculations

At this stage we will assume that a binomial process (Chapters 3 and 4) is applicable. Therefore the probability (p) of an event means that we would expect there to be, on average, $p \times n$ events, if we undertook n trials. So, if we assume the prevalence of infection is p, we would expect there to be $p \times M$ infected animals on average in a herd of size M. In addition, the probability that the next animal selected is infected remains constant, no matter how many animals have already been selected. We will consider later the alternative scenario, where we assume that we know exactly how many infected animals there are in a herd. In this case we expect the probability of an animal being infected to change, depending on the disease status of those selected previously, as outlined below.

The probability of including at least one infected animal in a consignment

No sanitary measures

A sanitary measure is one that is applied to manage risks posed to animal or human health. In this section, however, we will consider the likelihood of importing an infected animal if no sanitary measures are applied. The following calculations provide an estimate of the unrestricted or unmitigated risk. The likelihood following the application of sanitary measures is considered below.

a) Animals are selected at random from a population (e.g. a herd, region or country) with a prevalence of infection p

As discussed previously, if the probability that an animal is infected is *p* then, if *n* animals are selected at random, the probability that:

- all *n* animals are infected is *pn*
- none of the *n* animals is infected is (1- *p*)*n*
- at least one of the *n* animals is infected is 1-(1- *p*)*n*.

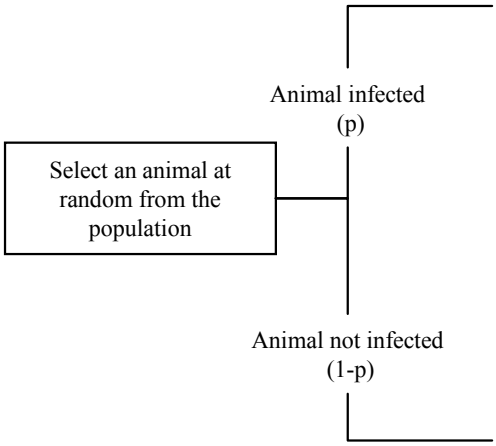


Figure 41
A scenario tree outlining the pathways whereby an animal selected at random from a population is infected or not

We can also derive this result from the binomial function (Chapter 3). The probability of including exactly 0, 1, 2, 3 …, *x* infected animals amongst a group of *n* randomly selected animals from a particular population with a prevalence of infection *p* is calculated by the binomial function:

$$P(X = x) = BINOMDIST(x, n, p) = \binom{n}{x}(p)^x(1-p)^{n-x} \qquad \text{Equation 31}$$

where: x = number of infected animals
n = the number of animals in the consignment and
p = the prevalence of infection in the population.

Each of these terms can be added together. For example if we want to determine the probability of including at least one infected animal (D^+) in a consignment of *n* animals we would add the terms from $x = 1$ to $x = n$:

$$P(D^+ \geq 1) = \sum_{x=1}^{n}\binom{n}{x}(p)^x(1-p)^{n-x} \qquad \text{Equation 32}$$

Alternatively, the solution may be obtained by subtracting the complementary probability of at least one infected animal. That is 1 minus the probability of no infected animals in a sample *n:*

$$P(D^+ \geq 1) = 1 - \binom{n}{0}(p)^0(1-p)^{n-0} \qquad \text{Equation 33}$$

Since both $\binom{n}{0}$ and $(p)^0 = 1$ this equation can be more conveniently expressed as:

$$P(D^+ \geq 1) = 1 - (1-p)^n \qquad \text{Equation 34}$$

b) Animals are selected from subsets of the population. That is, disease clustering is taken into account

In this case the animals will be selected in two stages. Initially a herd will be selected and then the animals will be chosen from within that herd (Fig. 42). The simplest case is where we assume that the prevalence of infection within each infected herd and the number of animals chosen is the same for each herd. If *h* herds are selected and *n* animals chosen from each herd then the probability of at least one infected animal being included in the consignment is:

$$P(D^+ \geq 1) = 1 - \left[1 - HP\left(1 - (1-p)^n\right)\right]^h \qquad \text{Equation 35}$$

where: *HP* = the herd level prevalence (proportion of infected herds)
n = number of animals chosen from a herd
h = number of herds.

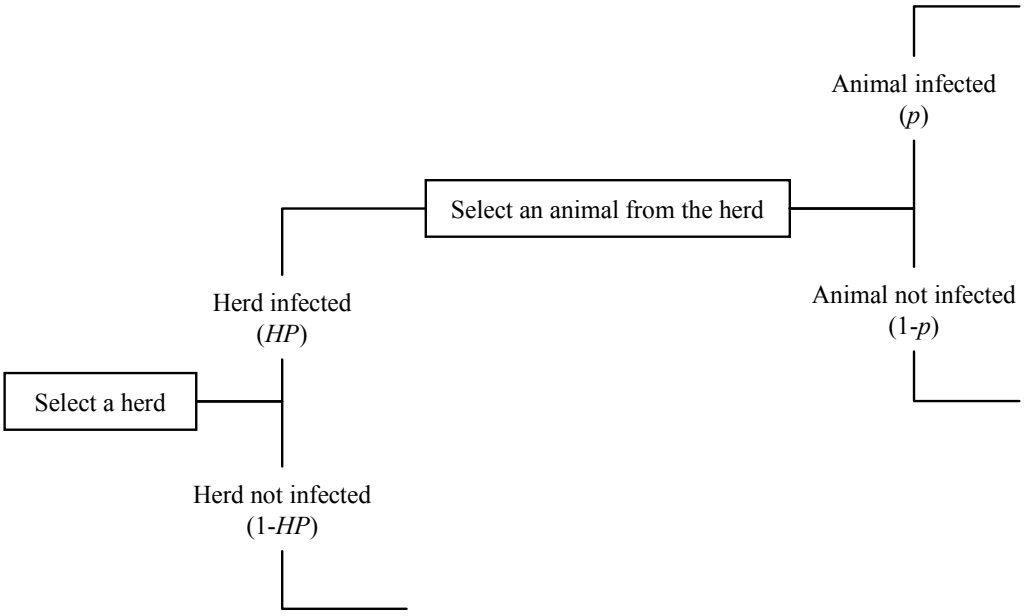


Figure 42
A scenario tree outlining the pathways whereby an animal selected at random from a herd is infected or not

If the number of animals (*n*) selected and/or the prevalence (*p*) varies from one herd to another, then this formula will need to be modified accordingly:

$$P(D^+ \geq 1) = 1 - \prod_{i=1}^{h}\left[1 - HP\left(1 - (1-p_i)^{n_i}\right)\right] \qquad \text{Equation 36}$$

In this case we need to calculate the probability that there are no infected animals from each of *i* herds, $\left[1-HP\left(1-\left(1-p_i\right)^{n_i}\right)\right]$, multiply all these results together, $\prod_{i=1}^{h}$ and subtract the answer from one to determine the probability of at least one infected animal in a consignment derived from *h* herds.

Sanitary measures are applied

A sanitary measure is one that is applied to manage risks posed to human or animal health. Measures include tests, inspections, treatment, quarantine, etc. This section examines the effect of applying tests as sanitary measures. The calculations thus provide an estimate of the risk remaining after sanitary measures are applied.

a) Test positive animals are rejected

– *Animals are selected at random from a population (e.g. a herd, region or country) with a prevalence of infection =p, and tested with a test with sensitivity = Se and specificity = Sp.*

Figure 43 outlines how we can determine if there is at least one infected animal present amongst all the animals we accept; that is, amongst the test negative animals. We need to calculate the proportion of false negatives (*B*) among the animals accepted (*B*+*C*). That is, $\frac{B}{B+C}$. More formally, we divide the probability that an animal is both infected and test negative by the probability that it is either infected and test negative or uninfected and test negative (Bayes' theorem, Chapter 3):

$$P\left(D^+\middle|T^-\right)=\frac{P\left(D^+\right)\times\left(1-P\left(T^+\middle|D^+\right)\right)}{P\left(D^+\right)\times\left(1-P\left(T^+\middle|D^+\right)\right)+\left(1-P\left(D^+\right)\right)\times P\left(T^-\middle|D^-\right)} \qquad \text{Equation 37}$$

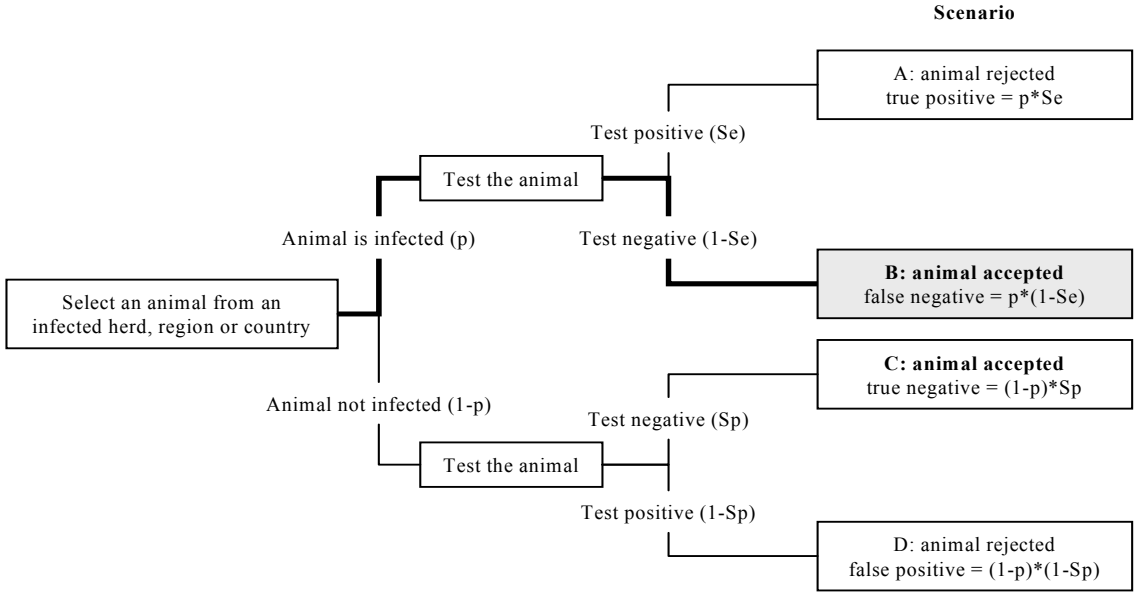


Figure 43
A scenario tree outlining the pathways whereby an animal, selected at random from a population and tested, is infected or not

Equation 37 can also be expressed as:

$$P\left(D^+\middle|T^-\right)=\frac{p\left(1-Se\right)}{p\left(1-Se\right)+\left(1-p\right)Sp} \qquad \text{Equation 38}$$

and scaled up to determine the probability that there will be at least one infected animal in the consignment:

$$P\left(D^{+} \geq 1 \middle| allT^{-}\right) = 1 - \left(1 - \frac{p(1-Se)}{p(1-Se)+(1-p)Sp}\right)^{n} \qquad \text{Equation 39}$$

In Equation 39 it is assumed that all the test results are independent for all animals tested. If this is not true, then the protection afforded by the sanitary measure may be over-estimated.

Alternatively we could determine the negative predictive value (NPV) of the test and calculate its complementary probability to determine the probability of at least one infected animal in our consignment. The NPV ($P\left(D^{-} \middle| T^{-}\right)$) is calculated as:

$$NPV = P\left(D^{-} \middle| T^{-}\right) = \frac{Sp(1-p)}{p(1-Se)+(1-p)Sp} \qquad \text{Equation 40}$$

and the probability of there being at least one infected animal amongst the group accepted is:

$$P\left(D^{+} \geq 1 \middle| allT^{-}\right) = 1 - \left(\frac{Sp(1-p)}{p(1-Se)+(1-p)Sp}\right)^{n} = 1 - NPV^{n} \qquad \text{Equation 41}$$

Note that Equations 39 and 41 give identical answers.

An alternative formula can be derived from the binomial function (Chapter 4). The probability that all the animals in a group of size *n* are test negative, if there are 0, 1, 2, 3, …, *x* infected animals in the group, is calculated by extending Equation 31 to include the probability that the infected animals (*x*) and uninfected animals (*n-x*) are all test negative and summing the individual binomial terms from $x = 0$ to $x = n$:

$$P\left(allT^{-}\right) = \sum_{x=0}^{n} \binom{n}{x} (p)^{x} (1-p)^{n-x} (1-Se)^{x} Sp^{n-x} \qquad \text{Equation 42}$$

Equation 42 can also be expressed as:

$$P\left(allT^{-}\right) = \left(p(1-Se)+(1-p)Sp\right)^{n} \qquad \text{Equation 43}$$

To calculate the probability that there will be at least one infected animal in a consignment, given that all the animals we accept are test negative, we need to first calculate the probability that all the animals are test negative, given there is at least one infected animal among them. To do this we need to sum the binomial probabilities in Equation 42 for $x = 1$ to $x = n$:

$$P\left(allT^{-} \cap D^{+} \geq 1\right) = \sum_{x=1}^{n} \binom{n}{x} (p)^{x} (1-p)^{n-x} (1-Se)^{x} Sp^{n-x} \qquad \text{Equation 44}$$

Equation 44 can also be expressed as:

$$P\left(allT^{-} \cap D^{+} \geq 1\right) = \left(p(1-Se)+(1-p)Sp\right)^{n} - \left((1-p)Sp\right)^{n} \qquad \text{Equation 45}$$

Equation 15 represents the binomial summation from ($x = 0$ to n) minus ($x = 0$). Next we need to determine the proportion that these animals represent out of all the test negative animals. This is an application of Bayes' theorem and involves dividing Equation 26 by Equation 24:

$$P\left(D^{+} \geq 1 \middle| allT^{-}\right) = \frac{\left(p(1-Se)+(1-p)Sp\right)^{n} - \left((1-p)Sp\right)^{n}}{\left(p(1-Se)+(1-p)Sp\right)^{n}} \qquad \text{Equation 46}$$

Note that Equations 39, 41 and 46 give identical answers. They are simply alternative forms of calculating the probability in which we are interested.

– *Animals are selected from subsets of the population. That is, herd level effects are taken into account.*

In this case the animals will be selected in two stages. Initially a herd will be selected and then test negative animals will be chosen from within that herd (Fig. 44).

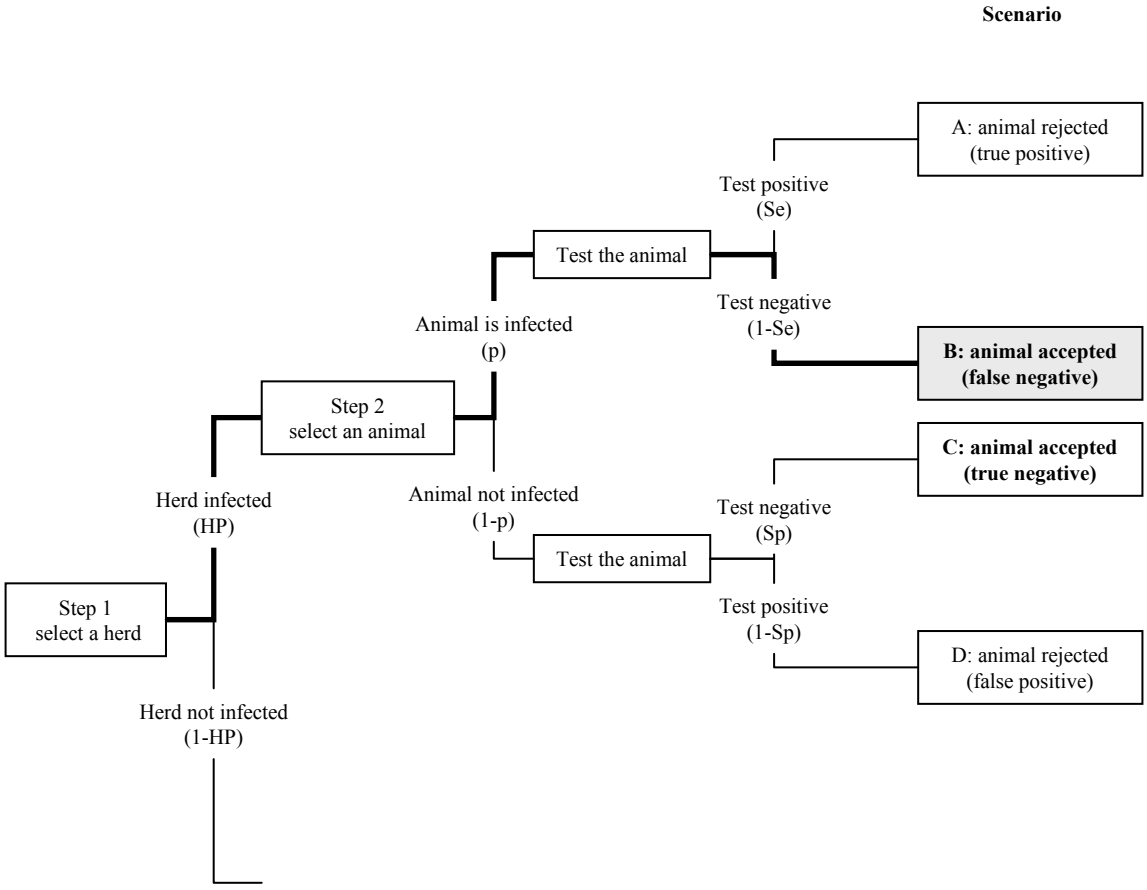


Figure 44
A scenario tree outlining the biological pathways leading to an animal, selected from an infected herd, being either accepted or rejected after it has been tested

The simplest case is where we assume that the prevalence of infection within each herd, the number of animals chosen and the test sensitivity and specificity are the same for each herd. If h herds are selected and n test negative animals chosen from each herd then the probability of at least one infected animal being included in the consignment, given that they are all test negative, can be calculated by any one of Equations 47, 48 or 49:

$$P\left(D^{+} \geq 1 \middle| allT^{-}\right) = 1 - \left[1 - HP\left(1 - \left(1 - \frac{p(1-Se)}{p(1-Se)+(1-p)Sp}\right)^{n}\right)\right]^{h} \qquad \text{Equation 47}$$

$$P\left(D^{+} \geq 1 \middle| allT^{-}\right) = 1 - \left[1 - HP\left(1 - NPV^{n}\right)\right]^{h} \qquad \text{Equation 48}$$

$$P\left(D^{+} \geq 1 \middle| allT^{-}\right) = 1 - \left[1 - HP\left(\frac{\left(p(1-Se)+(1-p)Sp\right)^{n} - \left((1-p)Sp\right)^{n}}{\left(p(1-Se)+(1-p)Sp\right)^{n}}\right)\right]^{h} \qquad \text{Equation 49}$$

These equations are derived from Equations 39, 41 and 46 respectively.

If the number of animals selected (n), the prevalence (p), the test sensitivity (Se) or the test specificity (Sp) vary from herd to herd, then these equations will need to be modified accordingly:

$$P\left(D^{+} \geq 1 \middle| allT^{-}\right) = 1 - \prod_{i=1}^{h}\left[1 - HP\left(1 - \left(1 - \frac{p_i(1-Se_i)}{p_i(1-Se_i) + (1-p_i)Sp_i}\right)^{n_i}\right)\right] \quad \text{Equation 50}$$

$$P\left(D^{+} \geq 1 \middle| allT^{-}\right) = 1 - \prod_{i=1}^{h}\left[1 - HP\left(1 - NPV_i^{\,n_i}\right)\right] \quad \text{Equation 51}$$

$$P\left(D^{+} \geq 1 \middle| allT^{-}\right) = 1 - \prod_{i=1}^{h}\left[1 - HP\left(\frac{(p_i(1-Se_i) + (1-p_i)Sp_i)^{n_i} - ((1-p_i)Sp_i)^{n_i}}{(p_i(1-Se_i) + (1-p_i)Sp_i)^{n_i}}\right)\right] \quad \text{Equation 52}$$

b) Test positive groups are rejected

– *Animals are selected at random from a population (e.g. a herd, region or country) with a prevalence of infection p, and tested with a test with sensitivity Se, and specificity Sp.*

Rather than rejecting individual test positive animals, we could reject all animals from a particular group if at least one of them were test positive. If we simply selected another group of animals from the same population, then the probability of including at least one infected animal in the consignment would be the same as calculated from Equations 39, 41 or 46. In other words, a group selection strategy in these circumstances offers no advantage over a strategy where we reject test positive individuals. The disadvantage, of course, is that there will be a considerable wastage of animals.

– *Animals are selected from subsets of the population. That is herd level effects are taken into account.*

In this case the group will be selected in two stages. Initially a herd is selected and a group of animals chosen from within that herd. If at least one of the animals is test positive the group, and therefore the herd, is rejected and another herd will need to be chosen (Fig. 45).

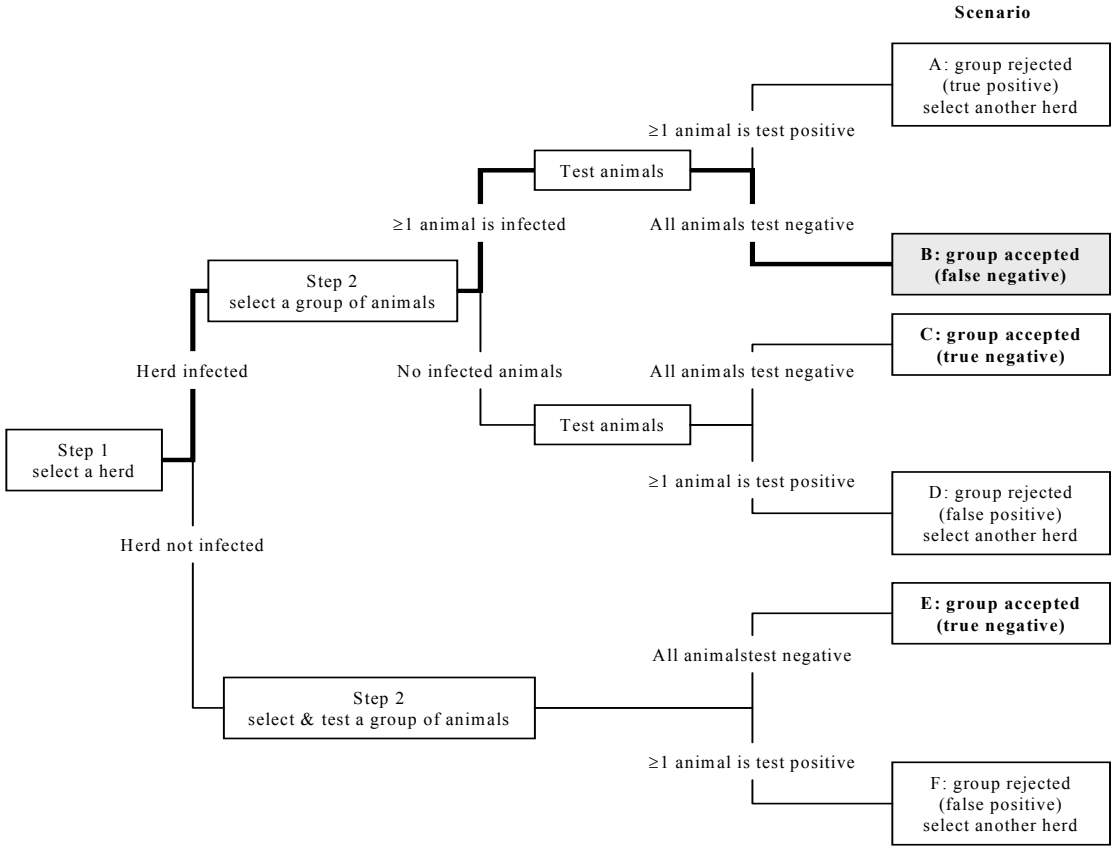


Figure 45
A scenario tree outlining the biological pathways leading to group of animals selected from a herd, being either accepted or rejected following testing. One or more test positive animals in the group results in the entire group being rejected and another herd being selected

To account for disease clustering within herds we need to modify Equations 43 and 45 accordingly:

$$P(groupT^-) = HP(p(1-Se)+(1-p)Sp)^n + (1-HP)Sp^n \quad \text{Equation 53}$$

$$P(groupT^- \cap D^+ \geq 1) = HP\left((p(1-Se)+(1-p)Sp)^n - ((1-p)Sp)^n\right) \quad \text{Equation 54}$$

where: HP = the herd level prevalence (proportion of infected herds).

The simplest case is where we assume that the prevalence of infection within each herd, the number of animals chosen, and the test sensitivity and specificity are the same for each herd. If h herds are selected and there are n animals in the group chosen from each herd then the probability of at least one infected animal being included in the consignment is:

$$P(D^+ \geq 1 | all\_groups T^-) = 1 - \left[1 - \frac{HP\left((p(1-Se)+(1-p)Sp)^n - ((1-p)Sp)^n\right)}{HP\left((p(1-Se)+(1-p)Sp)^n\right)+(1-HP)Sp^n}\right]^h \quad \text{Equation 55}$$

If the number of animals selected (n), the prevalence (p), the test sensitivity (Se) or the test specificity (Sp) vary from herd to herd, then this formula will need to be modified accordingly:

$$P(D^+ \geq 1 | all\_groups T^-) = 1 - \prod_{i=1}^{h}\left[1 - \frac{HP\left((p_i(1-Se_i)+(1-p_i)Sp_i)^{n_i} - ((1-p_i)Sp_i)^{n_i}\right)}{HP\left((p_i(1-Se_i)+(1-p_i)Sp_i)^{n_i}\right)+(1-HP)Sp_i^{n_i}}\right] \quad \text{Equation 56}$$

Hypergeometric probability calculations

As noted previously, a fundamental property of the binomial process is that probability remains constant. This means that the disease status of an individual animal selected at random is independent of the disease status of all the other animals chosen beforehand. The binomial process is effectively modelling sampling with replacement. In other words, the proportion of infected animals remaining in the herd can be considered constant; the binomial process is a good approximation if the number of animals being sampled is much less than (usually less than a tenth of) the herd size.

In some situations, an assumption of constant probability may no longer be reasonable as, for example, when sampling from small populations. For example, if the herd consists of 100 animals (M = 100) and there are five infected animals (D = 5), $\frac{D}{M}$ is initially 0.05. If the first animal selected is infected, then the fraction of animals infected becomes $\frac{4}{99} \cong 0.04$ whereas if the selected animal is uninfected this fraction becomes $\frac{5}{99} \cong 0.051$. As a result the probability that a selected animal is infected changes depending on whether the previous animal was infected or not. That is, p is no longer independent of the outcome of the previous trial.

The hypergeometric probability of obtaining exactly x infected animals in a sample of size s from a herd of M animals where there are exactly D infected animals is:

$$P(X = x) = HYPGEOMDIST(x, n, D, M) = \frac{\binom{D}{x}\binom{M-D}{n-x}}{\binom{M}{n}} \quad \text{Equation 57}$$

The numerator $\binom{D}{x}\binom{M-D}{n-x}$ calculates the total number of combinations for obtaining x infected and n-x uninfected animals in a sample n, from a herd of M animals of which D are infected. This is divided by the denominator $\binom{M}{n}$, which calculates the number of combinations for obtaining a sample of n animals from a herd of size M.

The probability that all the animals in a group of size n are test negative, if there are 0, 1, 2, 3, …, x infected animals in the group, is calculated by extending Equation 57 to include the probability that the infected animals (x) and uninfected animals (n-x) are all test negative and summing the individual hypergeometric terms from $x = 0$ to $x = n$:

$$P(all\ T^- \mid D\ infected\ animals\ in\ the\ herd) = \sum_{x=0}^{n} \frac{\binom{D}{x}\binom{M-D}{n-x}}{\binom{M}{n}} (1-Se)^x\, Sp^{n-x} \qquad \text{Equation 58}$$

To calculate the probability that there will be at least one infected animal in a consignment, given that all the animals we accept are test negative, we need to firstly calculate the probability that all the animals are test negative, given there is at least one infected animal among them. To do this we need to sum the hypergeometric probabilities in Equation 58 for $x = 1$ to $x = n$:

$$P(all\ \ T^- \cap D+ \geq 1\ \ in\ \ sample) = \sum_{x=1}^{n} \frac{\binom{D}{x}\binom{M-D}{n-x}}{\binom{M}{n}} (1-Se)^x\, Sp^{n-x} \qquad \text{Equation 59}$$

Next we need to determine the proportion that these animals represent out of all the test negative animals. This is an application of Bayes' theorem and involves dividing Equation 59 by Equation 58:

$$P(D^+ \geq 1 \mid all\ \ T^-) = \frac{\sum_{x=1}^{n} \frac{\binom{D}{x}\binom{M-D}{n-x}}{\binom{M}{n}} (1-Se)^x\, Sp^{n-x}}{\sum_{x=0}^{n} \frac{\binom{D}{x}\binom{M-D}{n-x}}{\binom{M}{n}} (1-Se)^x\, Sp^{n-x}} \qquad \text{Equation 60}$$

Unlike the binomial calculations discussed previously, this equation cannot be simplified because the probability changes as each animal is sampled. For this reason, the hypergeometric distribution is cumbersome to work with and the calculations quickly become unwieldy. We need to set up a spreadsheet to either calculate (Table XV) or simulate (Table XVI) the required probability. The series of graphs depicting the probability of obtaining test negative results (Fig. 46) shows, as expected, that as the prevalence of infection and the number of animals sampled increases, the probability that an infected herd is missed, or of obtaining a group of test negative animals, becomes increasingly unlikely. However, of those groups that are accepted, there is an increasing chance that there will be at least one infected animal amongst them.

Table XV
A spreadsheet model to *calculate* the probability that there is at least one infected animal in a sample of *n* test negative animals selected from a herd of size *M* in which there are *D* infected animals [17]

	A	B
1		**Input variables**: *M* = herd size *N* = sample size (set to 100 in this example) *D* = number of infected animals in the herd *Se* = test sensitivity *Sp* = test specificity
2	Number of infected animals in sample *x*	$P\left(D \geq 1 \cap allT^{-}\right)$
3	0	*HYPGEOMDIST*(A3,*n*,*D*,*M*)*(1-*Se*)^A3**Sp*^(*n*-A3)
4	1	*HYPGEOMDIST*(A4,*n*,*D*,*M*)*(1-*Se*)^A4**Sp*^(*n*-A4)
5	2	*HYPGEOMDIST*(A5,*n*,*D*,*M*)*(1-*Se*)^A5**Sp*^(*n*-A5)
...	...	...
103	100	*HYPGEOMDIST*(A103,*n*,*D*,*M*)*(1-Se)^A103**Sp*^(*n*-A103)
104	Probability of having at least one infected animal in the test negative group	*SUM*(B4:B103)/*SUM*(B3:B103) *Note*: this formula calculates the sum from *x* = 1 to *n*

Table XVI
A spreadsheet model to *simulate* the probability that there is at least one infected animal in a sample of *n* test negative animals selected from a herd of size *M* in which there are *D* infected animals

	A	B
1		**Input variables:** *M* = herd size *N* = sample size *D* = number of infected animals in the herd *Se* = test sensitivity *Sp* = test specificity
2	Number of infected animals in the sample	*Hypergeo*(*n*,*D*,*M*)
3	Number of uninfected animals in the sample	*n*-B2
4	Number of test positive animals	IF(B2=0,0, *Binomial*(B2,Se))+IF(B3=0,0, *Binomial*(B3,1-Sp))
5	Probability of having at least one infected animal in the test negative group	IF(B4>0,NA(),IF(B2>0,1,0))

Note: The mean of the output of cell B5 over all iterations is the estimated probability

[17] An 'IF' statement, IF(*x*>D,",HYPGEOMDIST(*x*,*n*,*D*,*M*)(1-Se)^{x}Sp$^{(n-x)}$) needs to be wrapped around the formulae in cells B3:B103 to ensure that *x* is not greater than *D*, otherwise an error will be returned

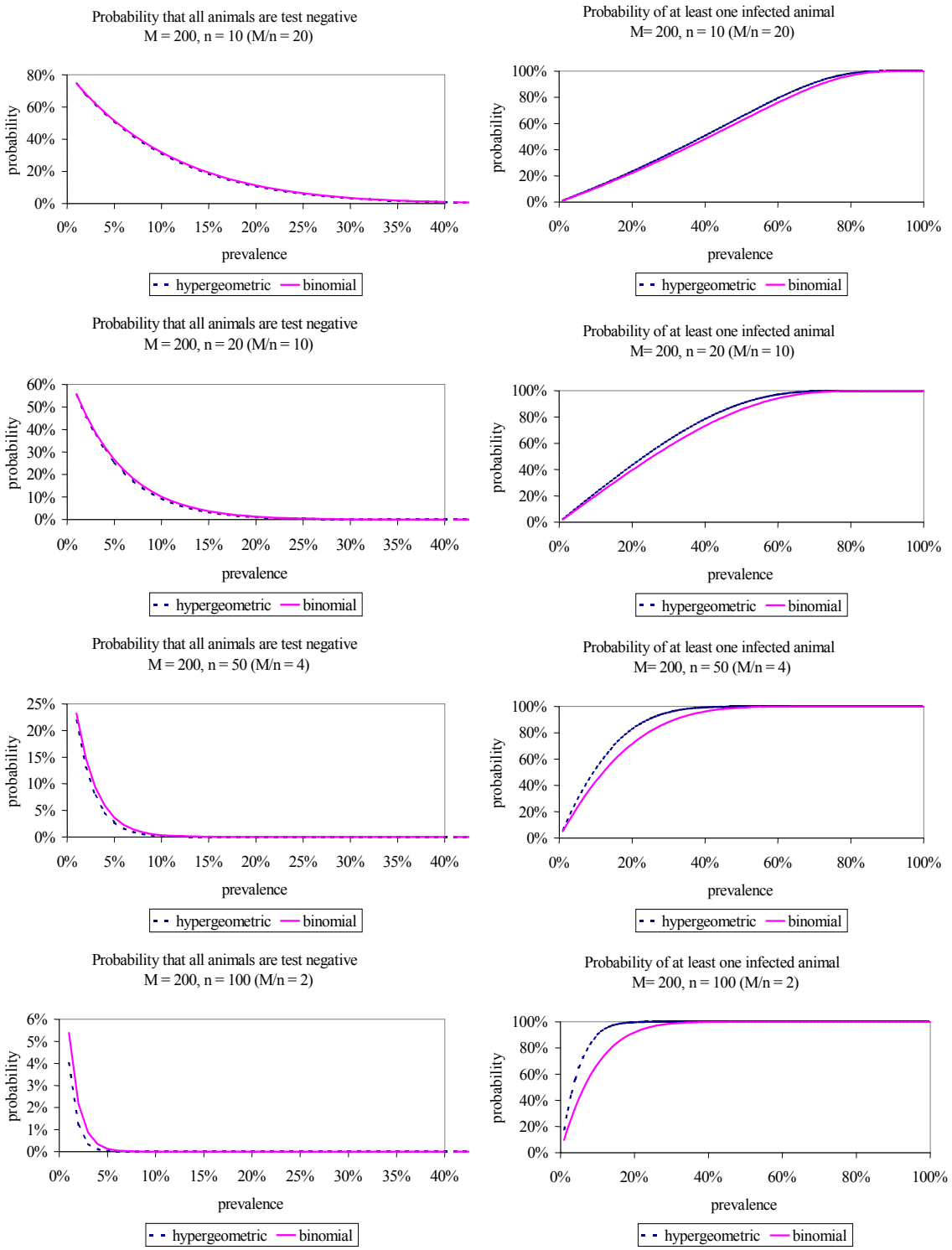


Figure 46
A comparison of the hypergeometric and binomial methods of calculating the probability of obtaining negative test results and the probability of at least one infected animal being included in the group of test negatives
Four different sample sizes (n) are compared. Herd size (M) = 200, test sensitivity (Se) = 90% and test specificity (Sp) = 98%

Figure 46 compares the results obtained from the hypergeometric and binomial methods of calculating the probability of obtaining negative test results and the probability of at least one infected animal being included in the group of test negatives for four different sample sizes. For small sample sizes, where the herd size is at least ten times the sample size, there

is good agreement between the hypergeometric and binomial calculations. As the ratio *M*:*n* becomes smaller, the binomial calculation becomes less accurate. However, given the level of imprecision that is nearly always present in risk assessment models, the magnitude of the difference may not be significant. Considering the simplicity of calculating binomial probabilities, there is considerable appeal in avoiding hypergeometric calculations wherever it is reasonable to do so.

Chapter 6
Determining a distribution to represent a variable[18]

Sources of information

There are essentially two sources of information from which a distribution can be developed to represent a variable in a risk assessment model:

– available empirical data
– expert opinion.

Three approaches are available to develop a distribution from these sources of information:

– fitting empirical data to a distribution using either a parametric or a non-parametric approach
– a purely subjective approach using expert opinion
– a combined approach that incorporates empirical data and expert opinion using Bayes' theorem.

Before deciding on which approach to use, it is important to consider the amount and relevance of the available information:

– if there are abundant representative data then either parametric or non-parametric techniques can be used to develop a distribution to model variability. This type of distribution is called a first order distribution (Chapter 7)

– if there are few representative data then either parametric or non-parametric techniques, which account for uncertainty in the parameters of the distribution, can be used to generate a second order distribution (Chapter 7)

– if there is a complete absence of data or the data that exist are either scarce or not representative, a subjective approach utilising expert opinion is appropriate

– if there are abundant non-representative data, for example experimental results from a different species, then a mixed approach may need to be employed whereby expert opinion is used to modify the distribution.

Even for data that meet stringent assumptions regarding independence and random sampling, there may be random fluctuations (non-systematic errors) in the sample data that make it difficult to select the 'true' distribution that best represents the data. Interpreting data inevitably requires subjective input such as, for example, assuming that the data represent a random sample from some probability distribution.

Determining a distribution where there are abundant representative data

Where there are abundant representative data a probability distribution that models variability can be specified from either the parameters, which are derived by fitting empirical data to a theoretical distribution using parametric techniques, or directly from the data using non-parametric techniques.

---

[18] The general reference for this chapter is Vose D. Risk Analysis, A Quantitative Guide. John Wiley & Sons Chichester, 2000

Parametric techniques

Parametric techniques involve fitting empirical data to a theoretical distribution, such as the normal or Poisson, which is then used to represent the data in a risk assessment model. Software packages are available, for example BestFit[19], that make distribution fitting 'easy'. Since it is assumed that the parameters derived from the data for the 'best fitting distribution' are the population parameters, the distribution chosen to model the data will be a first order distribution, which models variability (Chapter 7). However, if these software packages are not used with due care, an inappropriate distribution may be chosen. A distribution should not be selected arbitrarily from those that best 'fit' the data. Rather, careful consideration needs to be given to the underlying phenomena that generated the data so that the distribution chosen is both plausible and provides a good fit to the data. Several techniques are available to assist in selecting an appropriate distribution including goodness-of-fit statistics and probability plotting.

Goodness-of-fit statistics

Goodness-of-fit is the degree to which a fitted distribution matches the observed data. There is a wide variety of goodness-of-fit statistics, the most common of which are discussed below. The parameters for a fitted distribution under investigation are determined most commonly by the maximum likelihood estimators (MLEs). The MLEs are the parameter values that would give the highest probability of producing the observed data, given that the distribution type is correct. Goodness-of-fit statistics are calculated and compared for appropriate distribution functions to find the best fitting distribution.

a) Chi-squared test

The Chi-squared test is very flexible and can be used to test any assumption about a distribution. However, its major limitation is that it requires the data to be grouped so that some of the information from the original data is lost. As a guide at least twenty-five data points should be available before applying a Chi-squared test.

b) Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test compares a stepwise empirical cumulative density function (CDF) with the CDF of the hypothesised distribution. It identifies the maximum discrepancy between the two distributions, but takes no account of how the distribution fits the rest of the data. As a result, it may indicate a poor fit for an empirical distribution that generally fits the hypothesised distribution, apart from a single large discrepancy. On the other hand, it may indicate a good fit for a distribution that is a poor fit overall, but does not have such a large discrepant single value.

c) Anderson-Darling test

The Anderson-Darling test is a sophisticated version of the Kolmogorov-Smirnov test that assesses the discrepancy between the empirical and theoretical cumulative distributions over the entire distribution range. As a result, it is influenced much less by one large discrepant value and is generally more useful than the Kolmogorov-Smirnov test.

[19] Palisade Corporation, Newfield, New York

Probability plotting

Probability plotting is a subjective technique that is relatively easy and intuitive to use. It is advantageous in situations where one wishes to obtain a good fit for a particular portion of a distribution, such as the upper tail, and to retain a reasonable fit for the remainder of the distribution. Probability plotting involves graphing the data against a transformation of the data appropriate to the particular theoretical distribution under investigation. If there is a good fit there will be a straight line, although there will inevitably be some deviation. The extent of the deviation that constitutes a rejection of the proposed distribution is purely subjective.

Non-parametric techniques

Non-parametric techniques involve fitting data to an empirical distribution. They offer a number of advantages, as they are intuitive and simple to use. A particular form or shape of a distribution does not need to be assumed and inappropriate or confusing theoretical (parametric) distributions can be avoided. Empirical distributions can be defined for either continuous or discrete data:

Continuous data

Either the cumulative or histogram distribution function (Chapter 4) can be used to convert a set of data into an empirical (non-parametric) distribution, provided the data are continuous and cover a reasonable range:

Cumul(minimum, maximum,$\{x_i\}$*,*$\{p_i\}$*)*
Histogrm(minimum, maximum,$\{p_i\}$*)*

Discrete data

A discrete distribution can be defined for small data sets by using the data points themselves in a discrete or general distribution. For large data sets it might be more convenient to arrange the data into a histogram format and use either a histogram function or a cumulative function. The following distributions are discussed in Chapter 4:

Cumul(minimum, maximum,$\{x_i\}$*,*$\{p_i\}$*)*
Discrete($\{x_i\}$*,*$\{p_i\}$*)*
Duniform($\{x_i\}$*)*
General(minimum, maximum,$\{x_i\}$*,*$\{p_i\}$*)*
Histogrm(minimum, maximum,$\{p_i\}$*)*

Determining a distribution where there are few representative data

Where there are few representative data the parameters used to specify a distribution will be uncertain. Since the data are representative, the source of the uncertainty is random sampling error, which can be quantified by a sampling distribution. Confidence intervals for the parameters that specify a distribution, such as the mean and standard deviation, can then be estimated from the appropriate sampling distribution. Two approaches will be discussed in this section to derive a distribution for an uncertain parameter. These are classical statistical techniques and bootstrap simulation. A third approach, Bayesian inference is also discussed in this chapter.

Classical statistics

Classical statistical techniques usually assume that the underlying distribution of a data set is binomial or normal. For example, if we can assume that the underlying distribution is normal, the sampling distribution for the mean is the Student's *t* distribution and for the standard deviation, the Chi-squared distribution. The uncertainty associated with the estimate of the population mean, μ, is modelled by:

$$\mu = Student(n-1)\left(\frac{s}{\sqrt{n}}\right) + \bar{x}$$

where *Student*(*n*-1) is the Student's *t* distribution with *n*-1 degrees of freedom, *n* is the sample size and $\bar{x}$ and *s* are the sample mean and sample standard deviation respectively.

As the number of samples increases, the Student's *t*-distribution approaches the normal distribution. For sample sizes greater than thirty, the normal distribution function can be used to estimate the uncertainty in the estimate of the population mean:

$$\mu = Normal\left(\bar{x}, \frac{s}{\sqrt{n}}\right)$$

The uncertainty associated with the estimate of the population standard deviation is modelled by:

$$\sigma = \sqrt{\frac{(n-1)\times s^2}{Chisq(n-1)}}$$

where *Chisq*(*n*-1) is the Chi-squared distribution with *n*-1 degrees of freedom.

These sampling distributions enable us to capture the uncertainty associated with the estimates of the population mean (μ) and standard deviation (σ). They are used as inputs into a normal distribution function, which is then used to specify a second order normal distribution that enables us to encode and propagate variability and uncertainty separately (Chapter 7):

$$\underline{\underline{X}} = Normal(\underline{\mu}, \underline{\sigma})$$

where a single underscore denotes a first order random variable with constant parameters and the double underscore denotes a second order random variable with uncertain parameters.

Suppose we want to estimate the average weight of sheep in a flock but we only have information on the weights of ten sheep chosen at random. Since the sheep were randomly selected, we could be confident that the data are representative, albeit scarce. In addition, based on past observations, it is reasonable to assume that the weight of sheep is distributed normally. Table XVII is a spreadsheet model set up to derive a sampling distribution for the mean and standard deviation of the weight of all sheep in the flock, while Figure 47 depicts their respective sampling distributions. These distributions were obtained by running a simulation of 4,000 iterations on cells B14 and B15 in Table XVII. We are now in a position to define a second order distribution of sheep bodyweight (Fig. 48a). This is done by randomly selecting a value from each sampling distribution, inserting each value into a normal distribution function, plotting its graph and repeating this exercise a number of times. This enables us to build up a picture of possible distributions of weight, each of which represents a first order distribution, while together they form a second order distribution (Chapter 7).

Table XVII
A spreadsheet model to determine a sampling distribution for the population mean and standard deviation based on the weight of ten sheep collected at random

	A	B
1		Weight (kg)
2		50
3		54
4		46
5		47
6		44
7		52
8		47
9		56
10		41
11		48
12	Sample mean $\bar{x}$	*AVERAGE*(B2:B11)
13	Sample standard deviation *s*	*STDEV*(B2:B11)
14	Sampling distribution of the population mean μ	*Student*(9)*(B13/*SQRT*(10))+B12
15	Sampling distribution of population standard variation σ	*SQRT*(9*(B13^2)/*Chisq*(9))

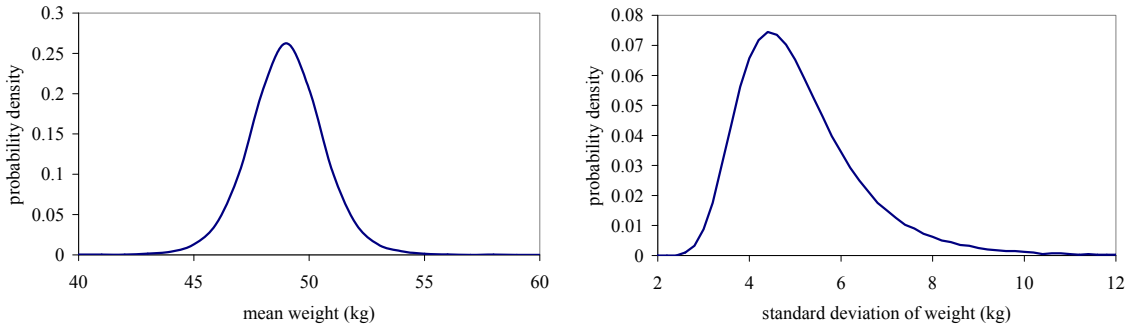


Figure 47
Hypothetical sampling distributions of the mean and the standard deviation of the weight of sheep in a flock based on ten samples collected at random

How important is the overall impact of the uncertainty associated with the mean and standard deviation which results from a small sample size? Is it important to model variability and uncertainty separately or is it reasonable to ignore the impact of uncertainty and model them together? Accounting for uncertainty can be complex and time consuming, so we need a relatively straightforward method to help us decide if it is really worth the effort. We can get a good idea by setting the model up and, initially, running it as if it were a first order model, where the sampling distributions for the mean and standard deviation are set to their expected values:

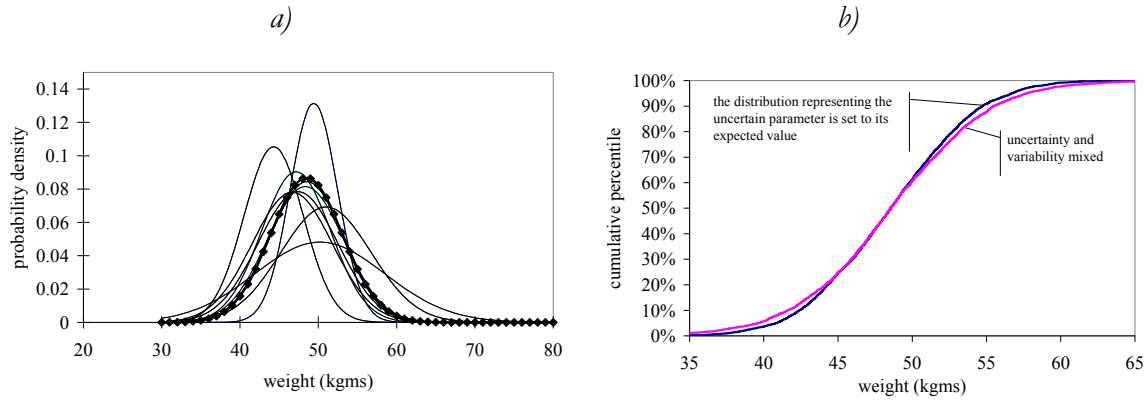
$X = Normal(48.5,4.58)$

Then we need to run it as a 'mixed' model where the uncertain and variable components are simulated together:

$X = Normal(\mu,\sigma)$

Where: $\mu = Normal\left(\bar{x}, \frac{s}{\sqrt{n}}\right)$ and $\sigma = \sqrt{\frac{(n-1)\times s^2}{CHISQ(n-1)}}$

We can then compare the results, as has been done in Figure 48b, which indicates that the impact of the uncertainty appears to be insignificant, as there is very little difference between these distributions. The means are identical and, while there is some inflation in the variance in the 'mixed' model, as reflected in the tails of the distribution, the magnitude of the difference is minor. In this particular case we could reasonably ignore the impact of uncertainty in modelling the bodyweight of sheep in the flock based on a random sample of ten animals. If we chose to use the 'mixed' distribution we would need to truncate it to avoid unrealistically low or high values. For example, we could use *Tnormal*($\underline{\mu}$,$\underline{\sigma}$,35,70) where 35 and 70 represent the minimum and maximum bodyweights we would expect in the flock.



a) a series of possible probability density curves of the bodyweight of sheep with the expected value curve plotted as a thick line with diamond markers

b) a cumulative probability plot investigating the impact of the uncertainty associated with random sampling error

Figure 48

Bootstrap simulation

Bootstrap simulation can be used to derive a distribution to represent an uncertain parameter, such as the mean or standard deviation. The general approach is to define a sampling distribution for the data, which consist of *n* samples, then collect *n* random samples with replacement from the data set and calculate the parameter of interest. This process is repeated many times and the results from each iteration are combined to produce a sampling distribution for the parameter. Depending on the circumstances either a non-parametric or a parametric approach may be used.

Non-parametric bootstrap simulation

Non-parametric bootstrap simulation is a powerful technique that does not require any *a priori* assumptions regarding the shape of the sampling distribution. Two options are available. We can either use the actual data and resample them, or fit an empirically based cumulative distribution to the data and sample from this empirical distribution.

Using the sheep bodyweight example, from the preceding section, we can set up a spreadsheet model to perform a non-parametric bootstrap simulation to define a sampling distribution for the uncertain parameter. We can either use the data directly by employing a discrete uniform function, where each data point has an equal probability of occurrence (Table XVIII), or develop a cumulative distribution (Table XIX) to define a sampling distribution of data. We then replicate the appropriate function *n* times, where *n* equals the number of samples in the original data set, so that we obtain *n* random samples from the sampling distribution. This is effectively sampling with replacement and constitutes a bootstrap replicate from which the mean and standard deviation are calculated.

A simulation is run and the results from each iteration are combined to produce a sampling distribution for each uncertain parameter. Figure 49 compares the outputs from these two models for the sampling distribution of the mean after running a simulation on cells C12 (Table XVIII) and D12 (Table XIX).

Table XVIII
A non-parametric bootstrap simulation model that uses a discrete uniform function to define a sampling distribution of the data which is then used to define a sampling distribution for the uncertain parameters, the population mean (μ) and standard deviation (σ)

	A	B
1	Weight (kg)	Non-parametric function
2	50	Duniform(A2:A11)
3	54	Duniform(A2:A11)
4	46	Duniform(A2:A11)
5	47	Duniform(A2:A11)
6	44	Duniform(A2:A11)
7	52	Duniform(A2:A11)
8	47	Duniform(A2:A11)
9	56	Duniform(A2:A11)
10	41	Duniform(A2:A11)
11	48	Duniform(A2:A11)
12	Replicate mean = AVERAGE(B2:B11)	
13	Replicate standard deviation = STDEV(B2:B11)	

Table XIX
A non-parametric bootstrap simulation model that fits an empirically based cumulative distribution (CDF) to the data. Samples are then drawn from the CDF to derive a sampling distribution for the uncertain parameters. The cumulative percentile is calculated by $\frac{i}{n+1}$ for $I = 1, 2, ..., n$, where n = sample size

	A	B	C
1	Cumulative percentile	Weight (kg)	Non-parametric function
2	9%	41	*Cumul*(41,56,B2:B11,A2:A11)
3	18%	44	*Cumul*(41,56,B2:B11,A2:A11)
4	27%	46	*Cumul*(41,56,B2:B11,A2:A11)
5	36%	47	*Cumul*(41,56,B2:B11,A2:A11)
6	45%	47	*Cumul*(41,56,B2:B11,A2:A11)
7	55%	48	*Cumul*(41,56,B2:B11,A2:A11)
8	64%	50	*Cumul*(41,56,B2:B11,A2:A11)
9	73%	52	*Cumul*(41,56,B2:B11,A2:A11)
10	82%	54	*Cumul*(41,56,B2:B11,A2:A11)
11	91%	56	*Cumul*(41,56,B2:B11,A2:A11)
12	Replicate mean = *AVERAGE*(C2:C11)		
13	Replicate standard deviation = *STDEV*(C2:C11)		

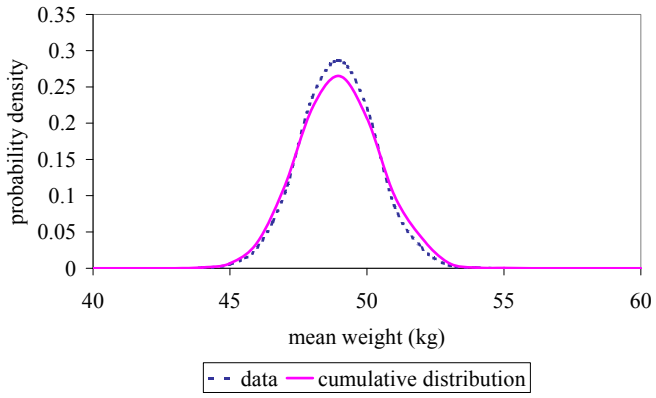


Figure 49
A comparison of the output from two non-parametric bootstrap simulations where the sampling distribution for the uncertain parameter, the mean (μ) was derived either directly from the data using a discrete uniform function or from a cumulative distribution of the data

Parametric bootstrap simulation

In some circumstances, we may be able to reasonably assume that the underlying distribution from which the data are derived belongs to a particular family of distributions, for example the normal, Poisson, exponential or binomial distributions. For this reason, we can use the relevant distribution to define a sampling distribution to represent the data. Table XX outlines a parametric bootstrap simulation model for the sheep bodyweight example from the preceding section. As already noted, we can reasonably assume that the weight of sheep is distributed normally. For this reason, the normal distribution function can be used to determine a sampling distribution. Firstly, we need to calculate the mean and standard deviation of the data and use these results in a normal distribution function, $Normal\left(\bar{x}, s\right)$. This function is replicated n times, where n equals the number of samples in the original data set, so that we obtain n random samples. This is effectively sampling with replacement and constitutes a bootstrap replicate from which the mean and standard deviation are calculated. A simulation is run and the results from each iteration are combined to produce a sampling distribution for each uncertain parameter. Figure 50 compares the results obtained by running a parametric bootstrap simulation for the sampling distribution of the mean on cell C12 (Table XX) with the results obtained from the classical statistical approach (Table XVII, Figure 47). The distribution derived from bootstrap simulation is somewhat narrower than the distribution derived from classical statistical methods and tends to underestimate the uncertainty. There are a number of techniques available to correct for this 'bias', such as the bias corrected and accelerated method. Discussion of such methods is beyond the scope of this text and the reader is referred to such books as Vose (2000)[20] or Cullen and Frey (1999)[21].

[20] Vose D. (2000). – *Risk Analysis: A Quantitative Guide.* John Wiley & Sons, Chichester.

[21] Cullen A.C. & Frey H.C. (1999). – Probabilistic Techniques in exposure Assessment. A Handbook for dealing with Uncertainty in Models and Inputs. Plenum Press, New York.

Table XX
A parametric bootstrap simulation model that uses the normal distribution function to define a sampling distribution for the data which is then used to define a sampling distribution for the uncertain parameters, the population mean (μ) and standard deviation (σ)

	A	B	C
1		Weight (kg)	Parametric function
2		50	*Normal*($B$12,$B$13)
3		54	*Normal*($B$12,$B$13)
4		46	*Normal*($B$12,$B$13)
5		47	*Normal*($B$12,$B$13)
6		44	*Normal*($B$12,$B$13)
7		52	*Normal*($B$12,$B$13)
8		47	*Normal*($B$12,$B$13)
9		56	*Normal*($B$12,$B$13)
10		41	*Normal*($B$12,$B$13)
11		48	*Normal*($B$12,$B$13)
12	Replicate mean	*AVERAGE*(B2:B11)	*AVERAGE*(C2:C11)
13	Replicate standard deviation	*STDEV*(B2:B11)	*STDEV*(C2:C11)

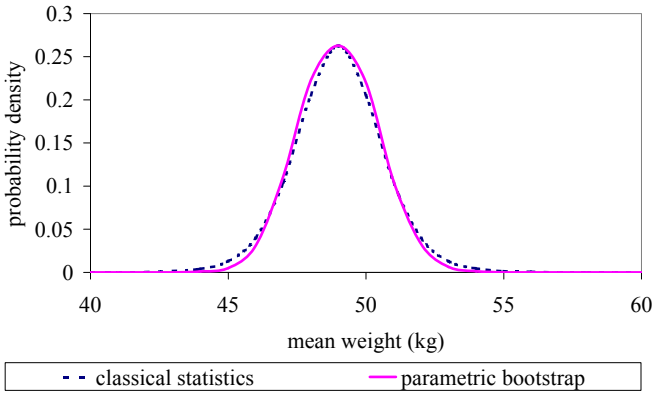


Figure 50
A comparison of the output from a parametric bootstrap simulation, where the sampling distribution for the uncertain parameter is derived using a normal distribution function, with classical statistical methods. The inherent assumption is that the data are derived from a normal distribution

Using expert opinion to determine a distribution where data are non-existent, scarce or not representative

If there is a complete absence of data, or the data that exist are either scarce or not representative, then a subjective approach utilising expert opinion is appropriate. Potential sources of bias, dealing with disagreement among experts, eliciting expert opinion and choosing an appropriate distribution are important issues that need to be considered carefully.

Bias

There are a number of human behaviours that lead to biases in one's judgements estimating a distribution's parameters or selecting an appropriate probability distribution. For example, individuals have a tendency to weight information which comes readily to mind; to be strongly influenced by small unrepresentative sets of data with which they are familiar; to be overconfident and estimate uncertainty too narrowly; to resist changing their mind in the face of new information; to try to influence decisions and outcomes by casting their beliefs in a particular direction; to state their beliefs in a way that favours their own performance or status; to knowingly suppress uncertainty in order to appear knowledgeable; and, to persist in stating weakening views to simply remain consistent over time (Cullen and Frey, 1999).

Expert disagreement

In cases of expert disagreement, it is usually best to explore the implications of the judgements of different experts separately to determine whether substantially different conclusions are likely. If the conclusions are not significantly affected, we can conclude that the results are robust despite the disagreement among experts. In some cases, experts may not disagree about the body of knowledge; rather, they may draw different inferences from an agreed body of knowledge. In such cases we need to make a judgement about which expert is more authoritative for the problem under scrutiny (Cullen and Frey, 1999).

Eliciting expert opinion

Psychological research has shown that accurate subjective probability judgements cannot be elicited simply by asking an individual to provide a probability. The methods of reasoning, or heuristics, employed when generating subjective estimates consistently introduces biases, which can be quite large, regardless of whether the individual is experienced in providing estimates and is familiar with probability theory or is a novice in this field (Merkhofer, 1987). Biases may also be introduced by the methodology used to elicit the opinion and then by the means in which it is modelled.

To minimise the impact of bias in eliciting expert opinion, a workshop method (Panel 1) has been developed jointly by the Veterinary Laboratories Agency in the United Kingdom and the Food and Agriculture Organization of the United Nations[22]. The method is based on a modified Delphi technique and is preferably carried out over two to three day period. Experts are gathered in one room and given a questionnaire that is answered individually and anonymously without discussion among them. The anonymity of the answers allows biases introduced through group discussion to be reduced.

The answers are then analysed and the results presented, followed by a facilitated discussion.

The questionnaire is given again, preferably the following day under the same conditions. This allows sufficient time for the experts to think about the points raised in the discussion and, if appropriate, take the opportunity to amend their answers given in the first questionnaire.

[22] Lisa Gallagher, Veterinary Laboratories Agency, Weybridge, United Kingdom, 2001. Personal communication with Noel Murray

Panel 1: The workshop method[23]

Introduction
– Explain the background to the work and aims of the workshop
– Briefly introduce risk analysis, the use of expert opinion and probability theory
– Explain the questions to be asked, definitions used within the questions and assumptions made

Conditioning the experts
– Explain the importance of accurate estimates, emphasising that this is an elicitation of knowledge, not a test of knowledge
– Provide any data that may be available associated with the question(s) being asked in an easily understood format

Questionnaire 1
– Conduct a pilot questionnaire prior to the workshop with a different group of individuals to insure that each question is clear and to gauge how long it will take to answer
– Allow the questionnaire to be answered individually and anonymously
– Ensure that the questionnaire is clear, easy to understand and not too long. Use decomposition – break the questions down into parts
– Ask experts to provide estimates for the maximum and minimum values followed by a most likely value for each question. Asking for estimates in this order reduces anchoring bias
– Ask the experts to provide percentage estimates rather than probabilities. Percentages are conceptually easier to estimate
– Provide aids such as computer software, graph paper or pie charts to help experts visualise percentages
– Allow enough time during the workshop to complete the questionnaire

Analysis 1
– Produce Beta-PERT distributions describing each expert's uncertainty around each question using the minimum, most likely and maximum values elicited
– Combine the distributions from each expert regarding a particular question, for example, using a discrete distribution, using appropriate weighting for each expert

The discussion
– Use a facilitator to ensure that all experts are equally included in the discussion to allow a free flow of information between experts
– Discuss the combined distribution for each question in turn

Questionnaire 2
– Present the questionnaire to the experts again to allow them to amend their previous answers, if appropriate

Analysis 2
– Analyse the answers to Questionnaire 2 as previously described
– Answers from subsidiary experts may not be included, depending on their degree of expertness. This should be determined before the commencement of the workshop

Results 2
– Provide experts with preliminary results as soon as possible after the workshop and send out a validation questionnaire to ensure results are reproducible
– Provide experts with the final results as soon as possible
– Invite feedback on the usefulness of the results and the process itself

A maximum of twenty experts is suggested as a manageable number for a workshop. The choice of expert is crucial. Each one should be impartially selected through a consultative process based on of their knowledge for the given subject. Biases may be introduced if the choice of expert is motivated by, for example, political or commercial reasons. The experts should also come from a variety of disciplines concerned with the subject in question, for example, veterinarians, scientists and policymakers. However, the inclusion of subsidiary experts who may not be quite as expert as a selected group of *core* experts can be useful. Subsidiary experts may provide extreme values in their estimates, which can be used to generate discussion and provide evidence of overconfidence, overestimation or underestimation. Discussion of these extreme values can be used to reduce biases and obtain more accurate estimates from the second questionnaire. Although it may not be

[23] Lisa Gallagher, Veterinary Laboratories Agency, Weybridge, United Kingdom, 2001

appropriate to include the estimates of subsidiary experts in final analysis, such a decision should be made prior to the workshop.

Choosing an appropriate distribution to model expert opinion

The choice of an appropriate distribution depends on the nature of the problem, the type of information available and whether its parameters are intuitive. Table XXI provides some examples of the most useful distributions used for modelling expert opinion, together with their parameters. These distributions have been discussed in detail in Chapter 4.

Table XXI
Examples of some useful distributions used to model expert opinion

Distribution	Parameters
Cumulative	*Minimum, maximum,*$\{x_i\}$, $\{p_i\}$
Discrete	$\{x_i\}$, $\{p_i\}$
General	*Minimum, maximum,*$\{x_i\}$, $\{p_i\}$
PERT	*Minimum, most likely, maximum*
Triangular	*Minimum, most likely, maximum*
Uniform	*Minimum, maximum*

It should be noted that small changes made to a cumulative plot may result in significant distortions to its corresponding relative frequency plot (Fig. 51). For this reason, the cumulative distribution should be used with caution when modelling expert opinion and the impact of any changes made directly to it to reflect an expert's opinion should be investigated by examining its corresponding probability density plot.

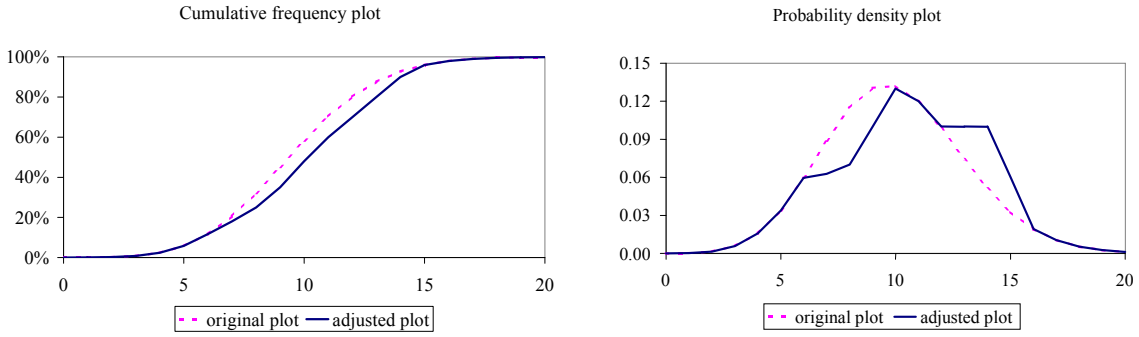


Figure 51
An example of how small changes in a distribution's cumulative plot can dramatically distort the shape of its corresponding probability density plot

Determining a distribution by combining empirical data and expert opinion

Bayesian inference

Bayesian inference is a useful, powerful technique whereby newly acquired empirical data can be combined with existing information, whether that information is itself based on pre-existing empirical data or on expert opinion, to improve an estimate of the parameter(s) used to characterise a distribution.

While a Bayesian approach is sometimes criticised as being subjective, it should be remembered that classical methods are also unavoidably subjective in so far as the choices made for using particular distributions, the confidence intervals and P-values chosen, and in accepting the assumptions inherent in the underlying statistical models. P-values, for instance, provide an indirect measurement of the evidence for or against a particular hypothesis and should not be interpreted as a final probability. They rely on proof by contradiction with high probability and indicate whether the hypothesis being investigated is sufficiently unlikely. P-values can be very deceptive as they involve probabilities of unobserved data that are more extreme than the observed data. On the other hand, a Bayesian approach provides direct evidence as it shows how the initial probability estimate is altered by the data.

Bayesian inference is a natural extension of Bayes' theorem (Chapter 3) and provides a powerful and flexible means of learning from experience. As new information becomes available it enables our existing knowledge to be easily and logically updated. It explicitly acknowledges subjectivity and describes the learning process mathematically. We begin with an opinion, however vague, and modify it as new information becomes available. Bayesian inference involves three steps:

a) determining a prior estimate of a parameter in the form of a probability distribution that expresses our state of knowledge (or ignorance) before any observations are made. The prior distribution is not necessarily dependent on data and may be purely subjective.

b) finding an appropriate likelihood function for the observed data. The likelihood function calculates the probability of observing the data for a given value of the prior estimate of the parameter. The shape of the likelihood function embodies the amount of information contained in the data. If the information is limited, the likelihood function will be broadly distributed, whereas if the information is significant, the likelihood function will be tightly focused around a particular parameter value.

c) calculating the posterior (i.e. revised) estimate of the parameter by multiplying the prior estimate and the likelihood function, then normalising the result so that the area under the curve sums to one.

The posterior distribution describes our state of knowledge of the parameter after we have acquired additional information. If it is similar to the prior distribution, then the information gained will confirm our pre-existing belief or state of knowledge. On the other hand, if it is significantly different to the prior we will have acquired some important new information. In fact, in many situations, as more information becomes available it is likely that the influence of the prior distribution will wane.

Bayesian inference can be helpful in elucidating the influence of different prior assumptions relative to the information available, and to determining the amount and quality of data necessary for convergence to the same posterior distribution. It offers a transparent means of modelling expert opinion, which is explicitly acknowledged in the prior distribution.

Prior distributions

As discussed above, a prior distribution expresses our state of knowledge before any new observations are made. Depending on the circumstances there are several options available:

Uninformed priors

An uninformed prior does not provide any additional information to a Bayesian inference other than establishing a possible range. For example, in some circumstances we may not have any information about the likely prevalence of infection within a herd. We might assume that, for a particular disease, the prevalence is likely to range from 0% to 30% and that any value within this range is equally as likely as any other value. This constitutes a uniform prior, *Uniform*(0,0.3), and has no influence on the Bayesian inference calculation, apart from establishing a range (Fig. 52).

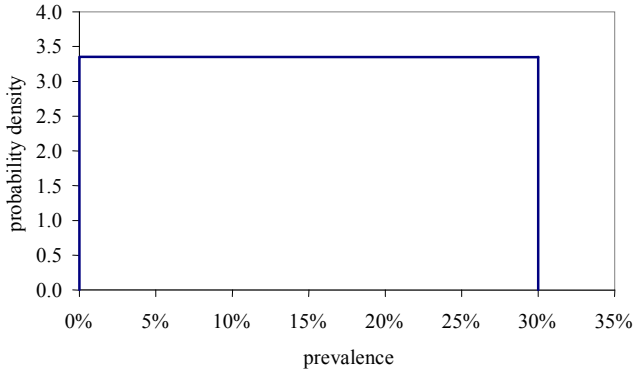


Figure 52
A uniform distribution to represent an uninformed prior

In some circumstances the parameter we are estimating may be expressed (re-parameterised) in a number of different ways. For example, we might want to estimate the average number of disease outbreaks per year (λ). If we assume that each outbreak is independent of every other outbreak and that there is a constant and continuous probability of a disease outbreak occurring throughout the year, then the outbreaks follow a Poisson process. The average number of outbreaks per year can also be expressed as $\frac{1}{\beta}$, where β is the mean interval between events. We might think it is reasonable to assign an uninformed prior in the form of a uniform distribution, *Uniform*(0,*x*), to λ. However, we could have just as easily parameterised the problem in terms of β. Since $\beta = \frac{1}{\lambda}$ our prior distribution would be $\frac{1}{Uniform(0,x)}$ which, as Figure 53 shows, is clearly not uninformed with respect to β. A useful technique in these circumstances to minimise the effects of re-parameterisation is to set up the prior distribution for λ as $\frac{1}{\lambda}$ and for β as $\frac{1}{\beta}$, that is we are using β as a prior for λ and vice versa. As a result, the prior distribution is transformation invariant. While such a distribution still does not appear to be uninformed (Fig. 54) it is the best that can be achieved in the circumstance and gives the same answer whether we undertake an analysis from the point of view of λ or β.

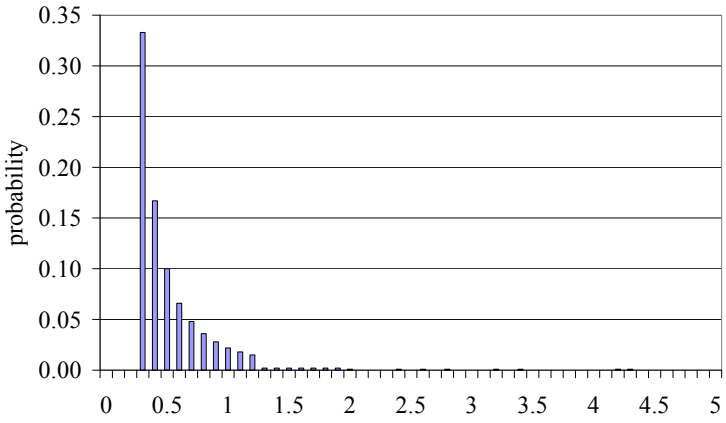


Figure 53
A prior distribution for β expressed as
$$\frac{1}{\lambda} = \frac{1}{Uniform(0,5)}$$
(simulation results over 1,000 iterations)

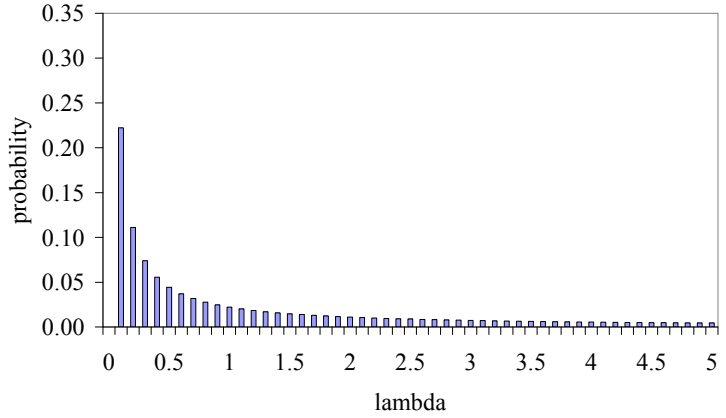


Figure 54
A prior distribution for lambda
$$prior(\lambda) \propto \frac{1}{\lambda}$$

Informed priors

An informed prior may be based on actual data or be purely subjective.

A conjugate prior has the same functional form as the likelihood function and leads to a posterior distribution belonging to the same distribution as the prior. Conjugate priors are often called convenience priors as we can determine the posterior distribution directly without having to construct a model as, for example, in Table XXIV. Table XXII lists two useful conjugate priors and their associated likelihood and posterior distributions.

Table XXII
Conjugate priors and their associated likelihoods and posterior distributions

Parameter to estimate	Conjugate prior	Likelihood	Posterior
Probability p	Uninformed prior: *Beta*(1,1) *note*: this is equal to *Uniform*(0,1)	Binomial	*Beta*(x+1,n-x+1)
	Informed prior: *Beta*(α_1,α_2)		*Beta*(α_1+x,α_2+n-x) where: x = number of successes, e.g. test positive animals n = number of trials, e.g. animals tested
Mean number of events per unit interval λ	Uninformed prior: $prior(\lambda) \propto 1/\lambda$	Poisson	$Gamma(x, \frac{1}{t})$
	Informed prior: *Gamma*(a,b) where a = the number of events, e.g. disease outbreaks b = the mean interval between events, e.g. years between outbreaks		$Gamma(a+x, \frac{b}{1+b\times t})$ where x = the number of events observed (e.g. outbreaks) in the interval, t t = the unit interval (time, litres, kilograms etc.)

Likelihood functions

The likelihood function calculates the probability of observing the data for a given value of the prior estimate of a parameter. The shape of the likelihood function embodies the amount of information contained in the data. Suppose we sample n animals from a herd, test them and find that there are x reactors. To determine the likelihood that there are x reactors, given a prior estimate of the prevalence of infection, p, we could use the binomial distribution function in Excel:

$$P(X = x) = BINOMDIST(x, n, p \times Se + (1-p) \times (1-Sp), 0)$$

where: p = the prevalence of infection,
n = the number of animals tested,
Se = test sensitivity and
Sp = test specificity.

There are a number of other useful probability distribution functions that can be used as likelihood functions, depending on the circumstances. These include the Poisson, hypergeometric and negative binomial.

Posterior distributions

The posterior distribution is the revised estimate of the parameter we are investigating and is obtained simply by multiplying the prior distribution and the likelihood function. Since the individual probabilities calculated by the likelihood function are independent of each other, the resulting posterior probabilities need to be normalised. This ensures that the area under the curve of a continuous distribution equals one and that the probabilities for a discrete distribution all add up to one. Two functions provided in @RISK enable normalisation to be carried out automatically. The *Discrete({x},{p})* function is used for discrete distributions and the *General(min,max,{x},{p})* function is used for continuous distributions.

An example of a Bayesian inference calculation: developing a distribution for an uncertain parameter p*, the prevalence of infection in a chicken flock*

Uninformed prior

The simplest situation is where we have no information about the disease status of a chicken flock. That is, our prior opinion about the flock's disease status is uninformed. Suppose the flock consists of several thousand chickens and we decide to test some of them using a test with a sensitivity of 80% and a specificity of 98%. What could we say about the disease status of the flock if we tested thirty chickens and they were all negative? Table XXIII outlines a spreadsheet model to derive a posterior distribution for the uncertain parameter *p*, the prevalence of infection. The resulting distribution is shown in Figure 55. In this case the posterior distribution is equal to the likelihood function, as the prior has no influence, apart from establishing a range.

Table XXIII
A spreadsheet model for a Bayesian inference calculation

	A	B	C	D	E
1	*n* = chickens tested	30	*Formulae:*		
2	*Se* = test sensitivity	80.0%	A8:A258 {0.0, 0.001, 0.002, 0.003,…., 0.25}		
3	*Sp* = test specificity	98.0%	B8:B258 {1}		
4			C8:C258 {BINOMDIST(0,*n*,A8**Se*+(1-A8)*(1-*Sp*),0)}		
5			D8:D258 {B8*C8}		
6			E8:E258 {D8/SUM(D8:D258)}		
7	Prevalence (p)	Prior probability density	Likelihood P(T-|D+)	Posterior	Normalised posterior probability
8	0.00%	1	5.45E-01	5.45E-01	2.44E-02
9	0.10%	1	5.33E-01	5.33E-01	2.38E-02
10	0.20%	1	5.20E-01	5.20E-01	2.33E-02
…	…	…	…	…	…
258	25%	1	7.02E-04	3.07E-11	3.14E-05

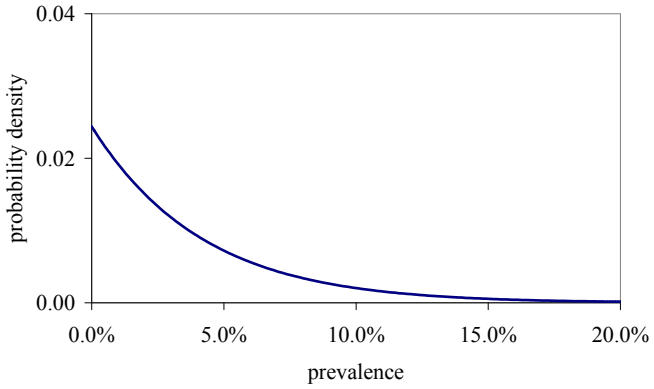


Figure 55
A posterior distribution for an uncertain parameter *p*, the prevalence of infection in a flock. The prior is uninformed and thirty chickens were tested with negative results

Informed prior

Suppose we have some information that suggests that the prevalence of disease in the flock is between 1 and 10% with a most likely value of 5%. If we apply the same test as in the previous example and tested the same number of chickens, what could we say about the disease status of the flock if they were all negative? In this example, as with the uninformed prior example, we will make the binomial approximation to the hypergeometric sample, as we will assume that the flock (*M*) is large in relation to the sample size (*n*), so that *n* is less than 10% of *M*. The calculations become more complex if a sample larger than 0.1x*M* is taken (see Chapter 6). Since we now have an informed prior we need to replace the prior distribution in column B of Table XXIII with a distribution to reflect this. We can use either a triangular or a PERT distribution as a prior (as outlined in Tables XXIV and XXV respectively). The main advantage of the triangular distribution is that it is easy to apply as we can readily obtain the appropriate densities in just one step. Although several steps are required to obtain the densities for the PERT distribution it offers significantly greater flexibility in modelling expert opinion, as discussed in Chapter 6.

As shown in Figures 56a and 57a, the likelihood function indicates that the information gained by testing thirty chickens is limited. This is reflected in the respective posterior distributions, which have not changed much from the prior. In contrast, if we test 100 chickens, (Figures 56b and 57b), we gain a lot more information. This is reflected in the likelihood function and consequently the posterior distribution.

Table XXIV
A prior probability density distribution (triangular) of the prevalence of infection in a flock of chickens where the probability that a chicken is infected is thought to be between 1% and 10%, with a most likely value of 5%

	A	B
1	Prevalence (p_i)	Probability density *f(x)**
2	1.0%	IF(A2<=ML,2*(A2-min)/((ML-min)*(max-min)),2*(max-A2)/((max-min)*(max-ML)))
3	1.1%	IF(A3<=ML,2*(A3-min)/((ML-min)*(max-min)),2*(max-A3)/((max-min)*(max-ML)))
4	1.2%	IF(A4<=ML,2*(A4-min)/((ML-min)*(max-min)),2*(max-A4)/((max-min)*(max-ML)))
5	…	…
92	10%	IF(A92<=ML,2*(A92-min)/((ML-min)*(max-min)),2*(max-A92)/((max-min)*(max-ML)))

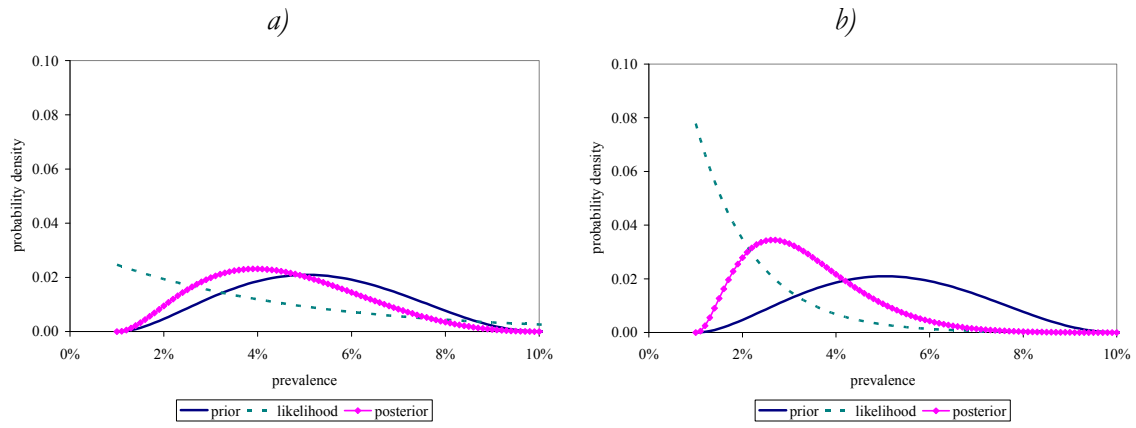
* The triangular distribution (*Triang*(a,b,c)) has a density $f(x) = \frac{2(x-a)}{(b-a)(c-a)}$ for values of x less than or equal to b, and a density $f(x) = \frac{2(c-x)}{(c-a)(c-b)}$ for values greater than b, where: a = minimum
b = most likely
c = maximum.

Table XXV
A prior probability density distribution (PERT) of the prevalence of infection in a flock of chickens where the probability that a chicken is infected is thought to be between 1% and 10%, with a most likely value of 5%

	A	B	C
1	Prevalence (p_i)	Cumulative probability density	Probability density $f(x)$ [a]
2	1.0%	*BETADIST*(A2,alpha_1,alpha_2,min,max) [b]	0
3	1.1%	*BETADIST*(A3,alpha_1,alpha_2,min,max)	(B3-B2)/(A3-A2)
4	1.2%	*BETADIST*(A4,alpha_1,alpha_2,min,max)	(B4-B3)/(A4-A3)
5	…	…	…
92	10%	*BETADIST*(A92,alpha_1,alpha_2,min,max)	(B92-B91)/(A92-A91)

a) The probability density for the prior distribution is obtained by differentiating the cumulative probability density curve: $\frac{dy}{dx} = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}$

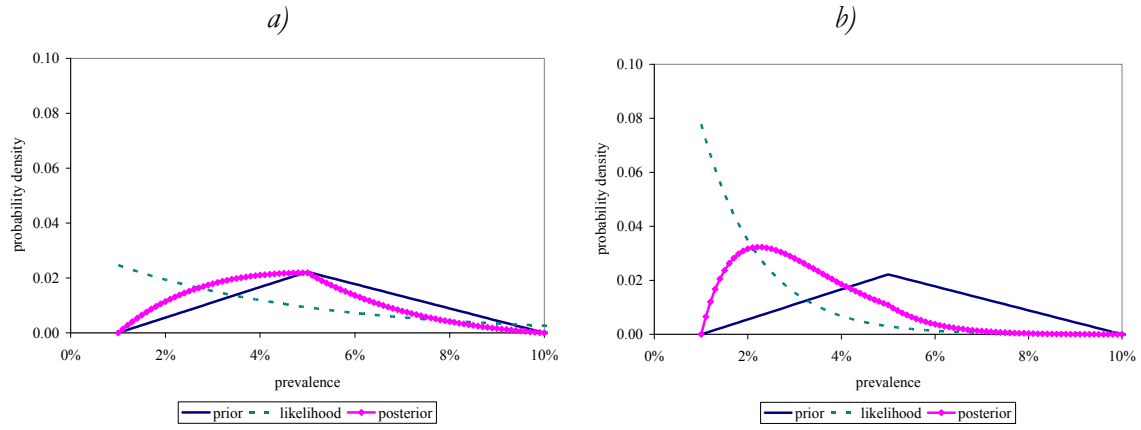
b) The parameter values for alpha 1 and alpha 2 for the cumulative Beta probability density function in (*BETADIST*) are calculated as outlined in Chapter 6 (PERT distribution). Min and max are the respective minimum and maximum values of the PERT distribution (*PERT*(a,b,c))



a) a sample of 30 chickens is tested

b) a sample of 100 chickens is tested

Figure 56
A distribution for an uncertain parameter *p*, the prevalence of infection, in a large chicken flock where the prior is informed
(densities obtained from a PERT distribution)



a) a sample of 30 chickens is tested

b) a sample of 100 chickens is tested

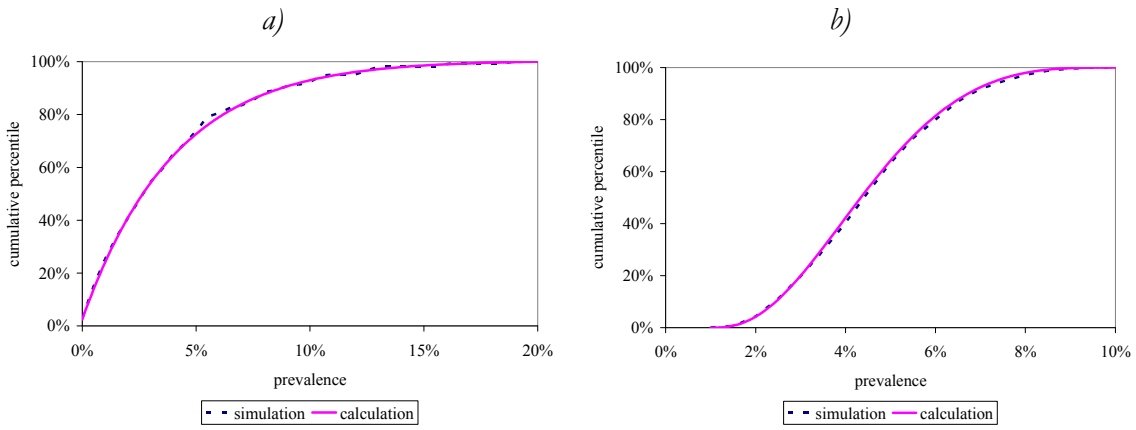
Figure 57
A distribution for an uncertain parameter *p*, the prevalence of infection, in a large chicken flock where the prior is informed
(densities obtained from a triangular distribution)

An example of a Bayesian inference simulation

Instead of calculating a Bayesian inference analysis, we could simulate it. Simulation models are arguably easier to construct and certainly more intuitive. For example, the model outlined in the preceding section for the Bayesian inference calculation could be replaced by the one outlined in Table XXVI. The output in cell B8, the posterior, determines whether the result is accepted or not. If one or more of the animals is test positive, the result is rejected and an error, NA(), generated. If a sufficiently large number of iterations is run, then a suitable number will be accepted, enabling a distribution to be plotted. The major drawback is that for low probability events, a large number of iterations are required. However, with a sufficiently fast computer, simulation is a reasonable alternative to calculation, as the simulated results will eventually converge with calculated values. Figures 58a and b compare the output distributions from a Bayesian inference calculation with those generated by simulation for both an uninformed and informed prior. In this case, the posterior distribution for the Bayesian simulation is based on 10,000 iterations. For the uninformed prior only 228 (about 2%) of all iterations were accepted, while for the informed prior, 1,658 (about 17%) were accepted. Although there is some minor disparity between the calculated and simulated results the difference is not important.

Table XXVI
A spreadsheet model for a Bayesian inference simulation

	A	B
1	**Input variables:** *M* = flock size = 1,000 *N* = sample size = 30 *Se* = test sensitivity = 80% *Sp* = test specificity = 98%	
2	**Prior for p_I** Prevalence of infected chickens p, in the flock M,	Uninformed p_1 =IntUniform(0,M)/M Informed p_2 = PERT(0.01,0.05,0.1)
3	**Likelihood**	IF(p_i=0,0,Binomial(n,p_i))
	a) Number of infected chickens in the group selected n	
4	*b)* Number of infected chickens that test positive	IF(B3=0,0,Binomial(B3,Se))
5	*c)* Number of uninfected chickens in the group selected	n-B3
6	*d)* Number of uninfected chickens that test positive	IF(B5=0,0,Binomial(B5,1-Sp))
7	*e)* Number of test positives	B4+B6
8	**Posterior for p_I**	IF(B7=0,p_i,NA())



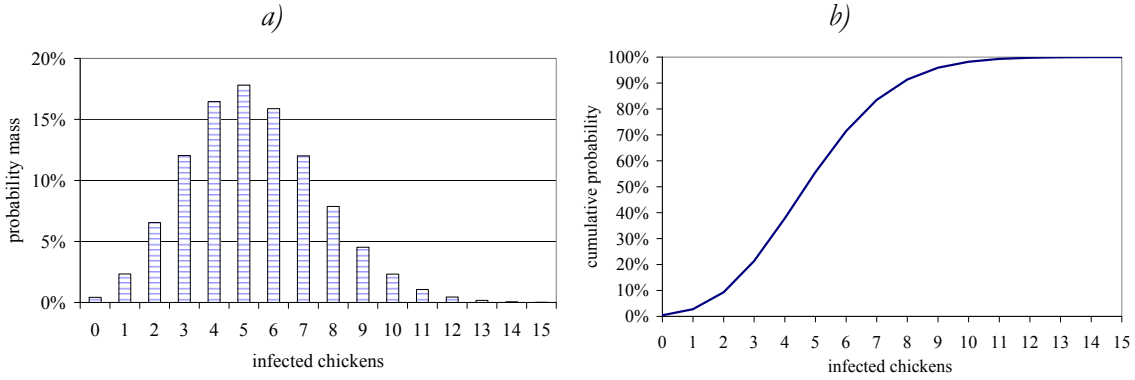
a) uninformed prior

b) informed prior

Figure 58
A cumulative distribution for an uncertain parameter *p*
The prevalence of infection in a chicken flock obtained by either a Bayesian inference calculation (Table XXV) or a Bayesian inference simulation (Table XXVI)

Chapter 7
An introduction to second order modelling[24]

Most quantitative risk assessment models are a blend of uncertainty and variability. Traditionally, most models have treated all random variables as if they were either all attributable to modelling variability or all to modelling uncertainty. The models have usually not attempted to distinguish between them. However, the situation most commonly encountered is likely to be when a distribution is used to model variability but the parameters characterising the distribution are actually uncertain. For example, if we choose 100 chickens from a large flock known to be infected with a particular disease, we might want to determine how many are likely to be infected. To do this we could use the binomial function (Chapter 4), *Binomial*(*n*,*p*). The two parameters that characterise this distribution, *n* and *p*, can be used to represent the number of chickens selected and the prevalence of infection in the flock respectively. If we assume that the actual prevalence is 5% then this distribution is simply modelling the variation in the number of infected chickens in the sample as both parameters, *n* and *p*, are fixed or constant values (Fig. 59). However, if we are unsure of the actual prevalence we could seek advice from an expert and model it using a PERT distribution (Chapter 4), which offers a convenient means of representing estimates such as the minimum, most likely and maximum, and weighting the probability of their occurrence. Since the prevalence (*p*) is now an uncertain parameter, the binomial distribution is modelling both variability and uncertainty.



a) a relative frequency plot

b) a cumulative probability plot

Figure 59
A binomial distribution modelling the variation in the number of infected chickens in a group of 100 selected from a large flock which has a 5% prevalence of infection.
Binomial (100,0.05)

Separating variability and uncertainty

A quantitative risk assessment model may be represented as $R = f(V,U)$ where the output (R) is a function of the variability (*V*) and the uncertainty (*U*). If we wish to account for the impact of uncertainty separately, we need to disaggregate the model into variable and uncertain components:

[24] The general reference for this chapter is Vose D. Risk Analysis, A Quantitative Guide. John Wiley & Sons Chichester, 2000

– Variable components are those for which the exact value of the parameter(s) that characterise a distribution are known or where there are abundant representative data and it is assumed that the parameter(s) derived from these data are the population parameters. A probability distribution can be either specified from the parameter(s), which are derived by fitting empirical data to a theoretical distribution, such as the normal distribution using parametric techniques, or directly from the data using non-parametric techniques (Chapter 6). These distributions, which have fixed or constant parameter(s), are also referred to as first order distributions. They can be represented, for example, as:

$$\underline{x} = Binomial(n, p) \qquad \text{Function 1}$$

where a single underscore ($\underline{x}$) represents a first order random variable and the parameters, n and p, without underscores, are fixed or constant values.

– Uncertain components are those for which the parameter(s) that characterise a distribution are uncertain, for example where there are few representative data, where there are no actual data or where the data are unrepresentative. Since the parameter(s) that characterise these distributions are themselves uncertain, a distribution needs to be specified for each one. Several techniques are available, including classical statistics, Bayesian inference and bootstrap, discussed in Chapter 6. The resulting distribution is known as a second order distribution. It allows us to encode and propagate both variability and uncertainty separately and is represented, for example, as:

$$\underline{\underline{x}} = Binomial(n, \underline{p}) \qquad \text{Function 2}$$

where $\underline{p}$ represents the uncertain parameter, which is itself represented as a first order random variable with constant parameters, for example *PERT*(0.02,0.05,0.01). A double underscore, $\underline{\underline{x}}$ denotes a second order random variable.

Variability and uncertainty are separated by developing a second order model. Initially, the model is built around the variability of a problem and any uncertainty that might exist is then overlaid.

The two most common techniques involve:
– calculating the variability and then simulating the uncertainty
– simulating both variability and uncertainty.

Can a second order model be justified?

Is it important to model variability and uncertainty separately or is it reasonable to ignore the impact of uncertainty and model them together? Accounting for uncertainty can be complex and time consuming, so we need a relatively straightforward method to help decide if it is really worth the effort. We can get a good idea by setting a model up and running it initially as if it were a first order model where the distributions representing the uncertain parameter(s) are set to their expected values. Then we need to run it as a 'mixed' model where the uncertain and variable components are simulated together. We can then compare the results of these two models

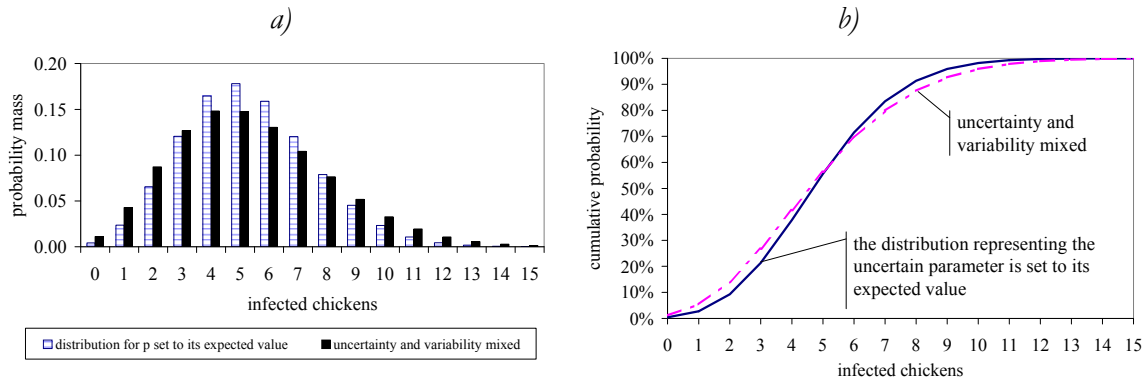
Continuing with the example introduced above, if we assigned a *PERT*(0.02,0.05,0.1) distribution to represent the uncertain parameter p, the prevalence of infection within the flock, we need to determine the expected value of this distribution and run the model as if it were a first order model:

$$\underline{x} = Binomial(n, \text{expected value for } p) = Binomial(100, 0.053) \qquad \text{Function 3}$$

Then we run it as a 'mixed' model where variability and uncertainty are simulated together:

$$\underline{x} = Binomial(100, PERT(0.02,0.05,0.1)) \qquad \text{Function 4}$$

Finally, we compare the results from these two models (Fig. 60). In this particular situation there appears to be very little difference in the plots, so we could reasonably conclude that we do not need to worry about accounting for the impact of uncertainty. The most obvious difference is in the tails of the distribution, with very little difference between the means. Mixing variability and uncertainty in a model will always inflate the variance, leading to a greater spread in the results, particularly in the tails of the distribution. However, the impact on the mean will not be as great. If you are reporting the results from the tails, for example the 95th percentile, rather than the mean, it is worth bearing in mind that a such a 'mixed' or 'hybrid' model will give a higher estimate. Of course, the magnitude of the difference may not be important.



a) a relative frequency plot

b) a cumulative probability plot

Figure 60
A comparison of the results from running a model with an uncertain parameter when:
(i) the distribution representing the uncertain parameter, the prevalence of infection, *p*, is set to its expected value (*Binomial*(100,0.053)) and
(ii) variability and uncertainty are simulated together
(*Binomial*(100,*PERT*(0.02,0.05,0.1)))

Calculating variability, simulating uncertainty

Continuing with our example: since the parameter p is uncertain, we need to define a second order distribution for the number of infected chickens, which in this case, is a second order random variable $\underline{\underline{x}}$:

$$\underline{\underline{x}} = Binomial(n, \underline{p}) = Binomial\left(100, \underline{Pert(0.02,0.05,0.1)}\right) \qquad \text{Function 5}$$

To separate variability and uncertainty in this function we first need to calculate the variability and then simulate uncertainty. We can do this by setting up a spreadsheet to calculate the variability in the number of infected chickens in the group selected by using the *BINOMDIST* function in Excel as outlined in Table XXVII, cells B4:AD104. We then sample the PERT distribution, representing prevalence, by collecting 30 Latin hypercube samples. We could, of course, collect more than 30, but this number should be sufficient for the purpose of illustration. Each sample value is used as an estimate of prevalence for a particular scenario in cells B2:AD2 . Latin hypercube sampling is preferred as it ensures that values from the entire range of the distribution will be sampled proportionally to the

probability density of the distribution. These values are then used as fixed or constant estimates of prevalence in the series of binomial calculations in columns B to AD to determine a distribution of the number of infected chickens in the sample. If we plot each of these distributions, as has been done in Figure 61, we will get a good idea of the impact of uncertainty. Figure 61 shows a second order distribution where each line is itself a first order distribution representing a particular scenario where the parameters *n* and *p* and constant. The distribution representing the combined effects of uncertainty and variability is also included in the figure and is plotted as a thick line. It clearly represents the average or mean of the first order distributions.

Table XXVII
Calculating variability, simulating uncertainty

	A	B	…	AD
1	The uncertain parameter p = *PERT*(0.02,0.05,0.1). 30 Latin hypercube samples are collected and used as constant input values (cells B2:AD2) for the *BINOMDIST* function in cells B2:B104 to AD4:AD104. Each of these groups of cells constitutes a first order distribution binomial distribution			
2	LHC samples:→	*5.05%*	…	*6.09%*
3	Infected chickens *(x)*	P(*X*=*x*)	…	P(*X*=*x*)
4	0	*BINOMDIST*($A4,100,B$2,1)	…	*BINOMDIST*($A4,100,AD$2,1)
5	1	*BINOMDIST*($A5,100,B$2,1)	…	*BINOMDIST*($A5,100,AD$2,1)
6	2	*BINOMDIST*($A6,100,B$2,1)	…	*BINOMDIST*($A6,100,AD$2,1)
…	…	…	…	…
104	100	*BINOMDIST*($A104,100,B$2,1)	…	*BINOMDIST*($A104,100,AD$2,1)

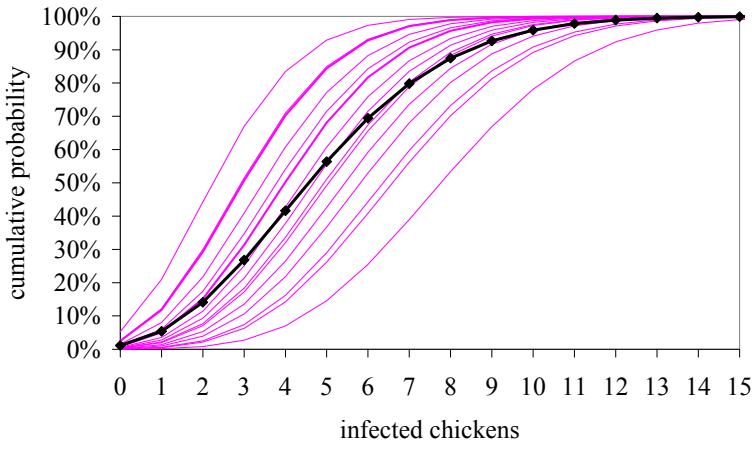


Figure 61
A second order distribution of the number of infected chickens in a sample of 100 selected from a large flock where the uncertain parameter *p*, prevalence of infection, is represented by a *PERT*(0.02,0.05,0.01) distribution.
Each fine line is a first order distribution for a particular estimate of prevalence and the distribution representing the output where uncertainty and variability are mixed is plotted as a thick line

A distribution of the 95th percentiles is plotted in Figure 62. The 95% confidence interval about the sample mean of 9 for this distribution is 8.8 to 9.5, which indicates that variability dominates uncertainty as the confidence interval is very narrow. If required, confidence limits could be determined for the entire cumulative distribution. These confidence limits are also known as an uncertainty band.

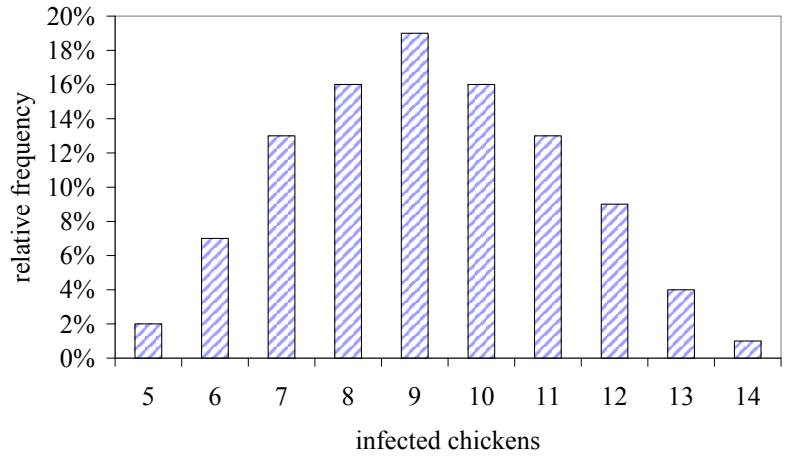


Figure 62
A frequency distribution of the 95th percentiles from a second order distribution of the number of infected chickens in a sample of 100 selected
from a large flock where the uncertain parameter p, the prevalence of infection, is represented by a PERT(0.02,0.05,0.01) distribution

Suppose we want to determine the probability of including at least one infected chicken in a group of ten chickens. In this case the second order distribution is represented by:

$$P\left(\underline{\underline{x \geq 1}}\right) = 1 - \left(1 - \underline{p}\right)^{n} \qquad \text{Equation 61}$$

Since Equation 61 explicitly calculates the variation in the number of infected chickens in the sample, we can run a simulation directly. The resulting distribution is an example of a second order distribution which consists of a single line (Fig. 63).

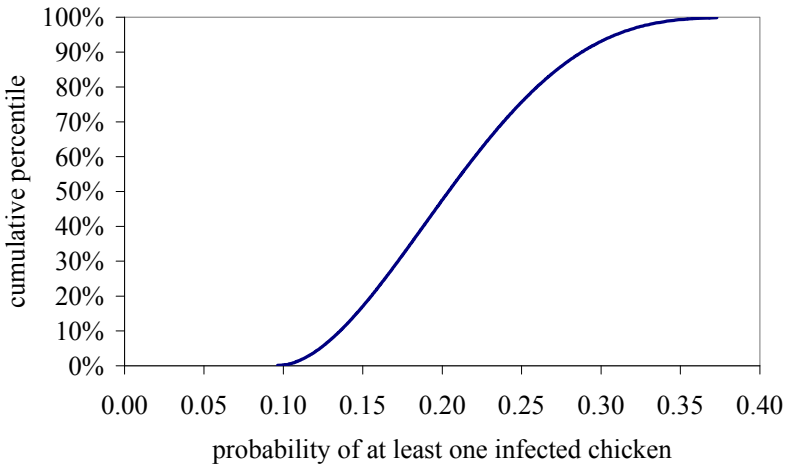


Figure 63
A second order distribution of the probability of including at least one infected chicken amongst a group of ten selected
from a large flock where the uncertain parameter *p*, the prevalence of infection, is represented by a *PERT*(0.01,0.02,0.05) distribution

Simulating both variability and uncertainty

We can also separate the effects of variability and uncertainty by simulating both. In the example being considered, we could simulate the variation in the number of infected chickens in the sample with a first order distribution function:

$$\underline{x} = Binomial(n, p) \qquad \text{Function 6}$$

However, since we are using the random sampling of simulation to model variability, it is no longer available to model uncertainty. Just as we did in the preceding section, we need to collect some Latin hypercube samples from the PERT distribution representing the uncertain parameter *p*, the prevalence of infection. Each sample value is then used as a constant value input into a first order distribution function (Function 6). The model is then run using the first Latin hypercube sample for the first simulation and so on. The Simtable function in @RISK can be used to automate the process by referencing the list of values generated by the Latin hypercube sampling of the PERT distribution (Table XXVIII). Despite its simplicity, simulating both variability and uncertainty can impose a significant computational cost as it can take considerable time to run a model, particularly if we are modelling rare events and need to undertake a large number of iterations. For example, if we collect 30 Latin hypercube samples and run 10,000 iterations per simulation, we will end up performing 300,000 iterations. Once the model has finished running, we can collect the results and plot them to produce a graph similar to that in Figure 61.

Table XXVIII
Simulating both variability and uncertainty

	A	B
1	Variability	Uncertainty
2		Prevalence = *PERT*(0.02,0.05,.1)
3	Number of infected chickens *x* = *Binomial* (100,B2)	Simtable(B4:B33)
4		5.05%
5		6.09%
6		4.4%
…	…	…
33		4.1%

Chapter 8
Guidelines for developing a quantitative risk assessment model

Regardless of whether one is developing a qualitative or quantitative risk assessment model, a number of important steps must be worked through in a systematic manner. Nevertheless, the development of a quantitative risk assessment model presents greater challenges. The steps required in developing a quantitative model include:

- state the question to be answered clearly and explicitly
- identify the populations of interest
- draw a scenario tree
- keep the model as simple as possible
- consider if you need to account for independence between units
- ensure a proper account is taken of independence and dependency or correlation between variables
- determine the type of information available for each of the model's inputs
- document the assumptions, evidence, data and uncertainties for each variable
- select an appropriate distribution for each variable
- decide if variability and uncertainty need to be separated
- ensure that each iteration of the model is biologically plausible
- verify the calculations independently
- conduct a sensitivity analysis
- consider how the results should be presented to facilitate communication
- commission a peer review of the model.

Each of these will be discussed in turn.

Determining the scope of the risk analysis

From the outset, it is essential to have a clear understanding of the question to be answered, regardless of whether one is planning a qualitative or quantitative risk assessment. The process of defining the question is known as 'determining the scope of the risk analysis' and, if this step is not carried out properly, problems will inevitably arise in interpreting and communicating the results.

When considering likelihood, the units of the numerator and denominator must be stated explicitly. For example, the numerator may be expressed as the probability of an event, several events or, more commonly, of at least one event. The denominator may be expressed per imported animal, per tonne of meat, per consignment, or per year etc. The way in which risk is expressed has an important bearing on how a model is developed and how the results are interpreted and communicated.

A question might be asked: 'What is the likelihood of introducing classical swine fever (CSF) with porcine embryos?' The imprecise phrasing of this question makes it impossible to identify clearly the exact outcome of interest. For instance, is the decision-maker interested in the probability per embryo, per donor, per recipient, per consignment, per month or per year? Is he/she interested in the probability that the embryo donors pass all the tests, despite there being at least one embryo contaminated with CSF virus ($P\left(allT^{-}\middle|D^{+}\geq 1\right)$)? Or is he/she interested in the probability that at least one embryo is contaminated with CSF virus even though all donors have passed the tests and the embryos have been accepted for importation ($P\left(D^{+}\geq 1\middle|allT^{-}\right)$)? The latter scenario

considers all the embryos, whether they are likely to be from infected donors or not, while the former scenario considers only embryos derived from infected donors.

A clear and explicit question would be:

What is the probability of at least one outbreak of CSF in (my country) each year if it is anticipated that between one and two thousand porcine embryos, that comply with the sanitary measures outlined in the *Code*, are imported annually from a country where CSF is endemic?

The population(s) of interest

When constructing your risk assessment model, you need to be specific about the population of animals you are interested in. For example, are you interested in all cattle herds in a country or region regardless of their disease status? Or are you interested in a subset of these herds with a history of disease freedom, such as those herds participating in an accreditation program? What animal and human populations may be exposed to the imported animal or animal products?

Depicting the model graphically

Whether one is planning a qualitative or quantitative risk assessment, a graphical depiction of the biological pathways provides a useful conceptual framework. It assists in conveying visually the range and types of pathways to be considered in a simple, transparent and meaningful fashion for qualitative assessments, and is an essential step if a quantitative model is to be developed. A graphical depiction provides a useful 'mind map' or visual representation to:

– identify variables
– identify relationships among variables
– identify information requirements
– ensure a logical chain of events in space and time
– provide a framework for the development of a mathematical model
– ensure the appropriate estimate is calculated
– assist with communicating the model structure
– clarify ideas and understanding of the problem.

Scenario trees are an appropriate and effective way of depicting biological pathways. A scenario tree starts with an initial event, for example selecting some animals from a herd which is potentially infected. It then outlines the various pathways that lead to different outcomes, such as accepting animals that are test negative or the outbreak of a disease. By convention events are described in boxes or nodes, while the probability of an event is described by a line or arrow emanating from the respective box or node (Fig. 64). Examples of scenario trees are presented in Figures 65 to 68.

There are alternative ways of depicting a model graphically For example, an influence diagram that shows how different variables interact with each other (Fig. 69[25]). Such diagrams may be of assistance in communicating some aspects of a model, but they usually do not provide an appropriate outline of the various pathways leading from an initial event to the outcome(s) of interest. Influence diagrams can rapidly become quite complex and difficult to follow if there are a large number of interdependent variables.

[25] MAF Regulatory Authority. Import Risk Analysis: chicken meat and chicken meat products; Bernard Matthews Foods Ltd turkey meat preparations from the United Kingdom. Wellington, New Zealand, 1999

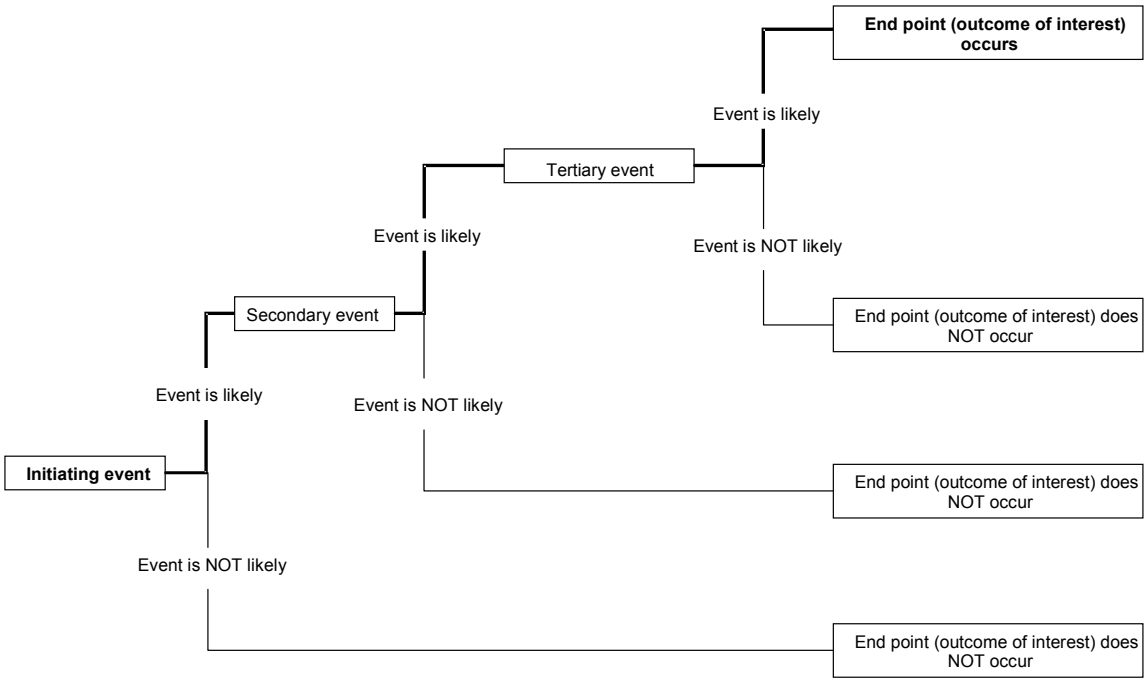


Figure 64
Generalised framework for a scenario tree where probabilities are examined

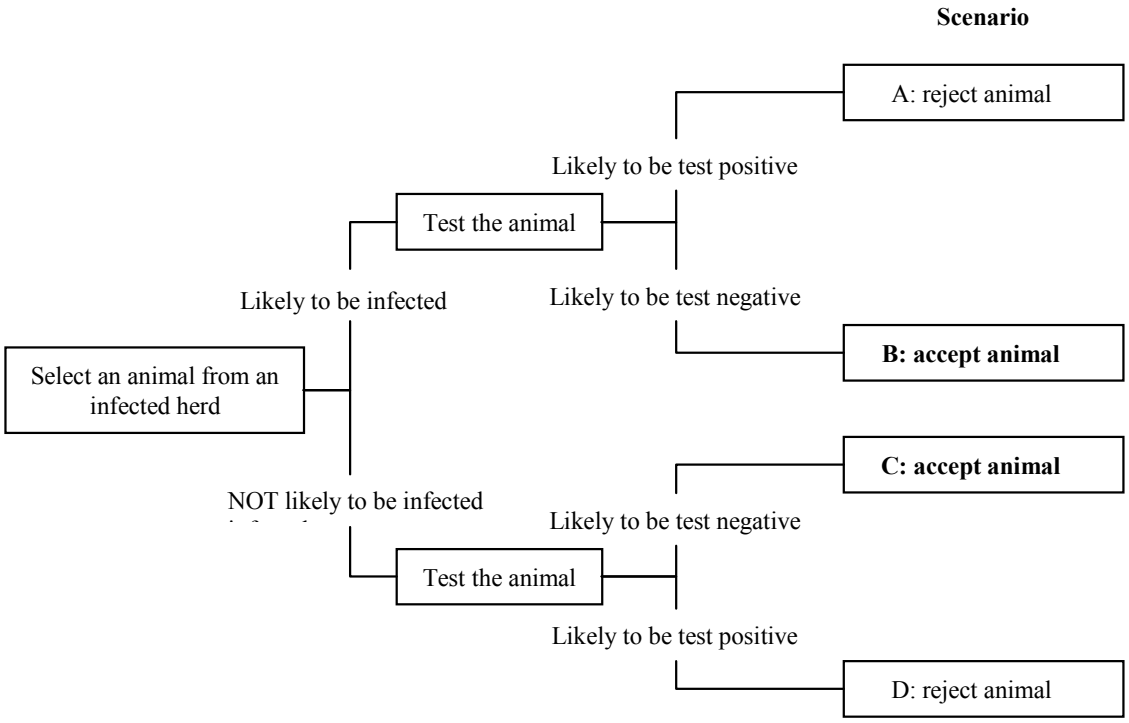


Figure 65
A scenario tree outlining the biological pathways leading to an animal, selected from an infected herd being either accepted or rejected after it has been tested

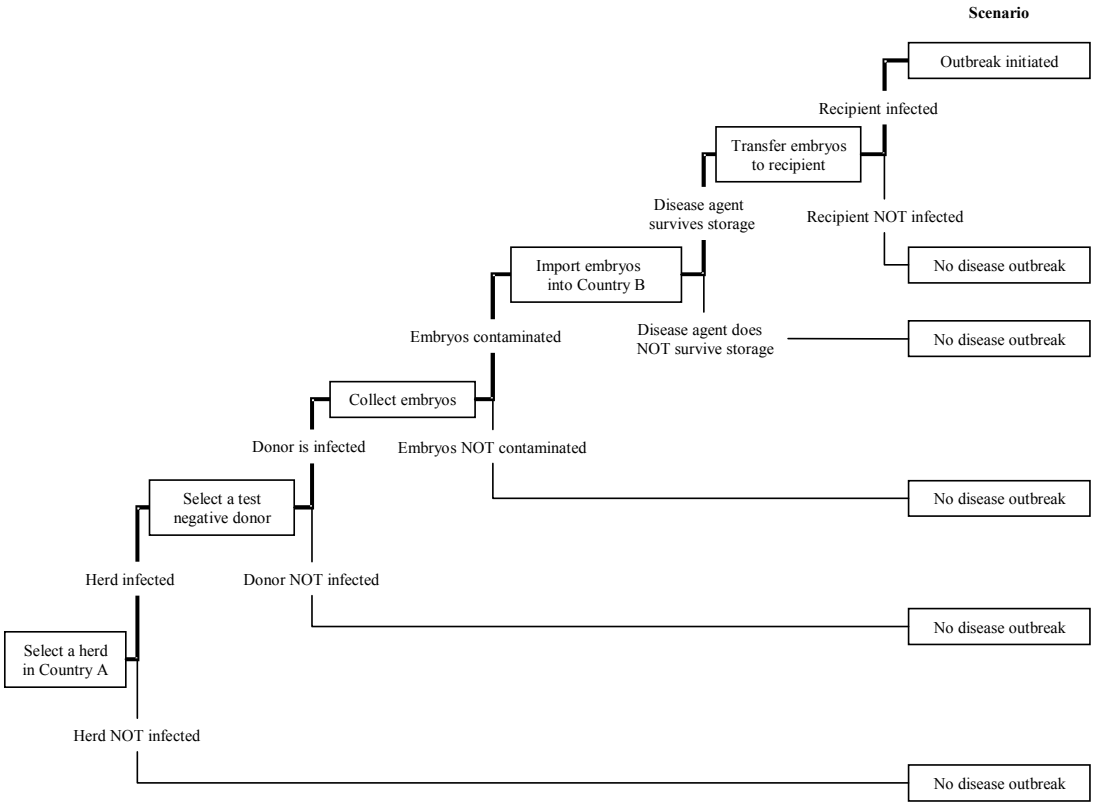


Figure 66
A scenario tree outlining some pathways leading to a disease outbreak following the importation of embryos

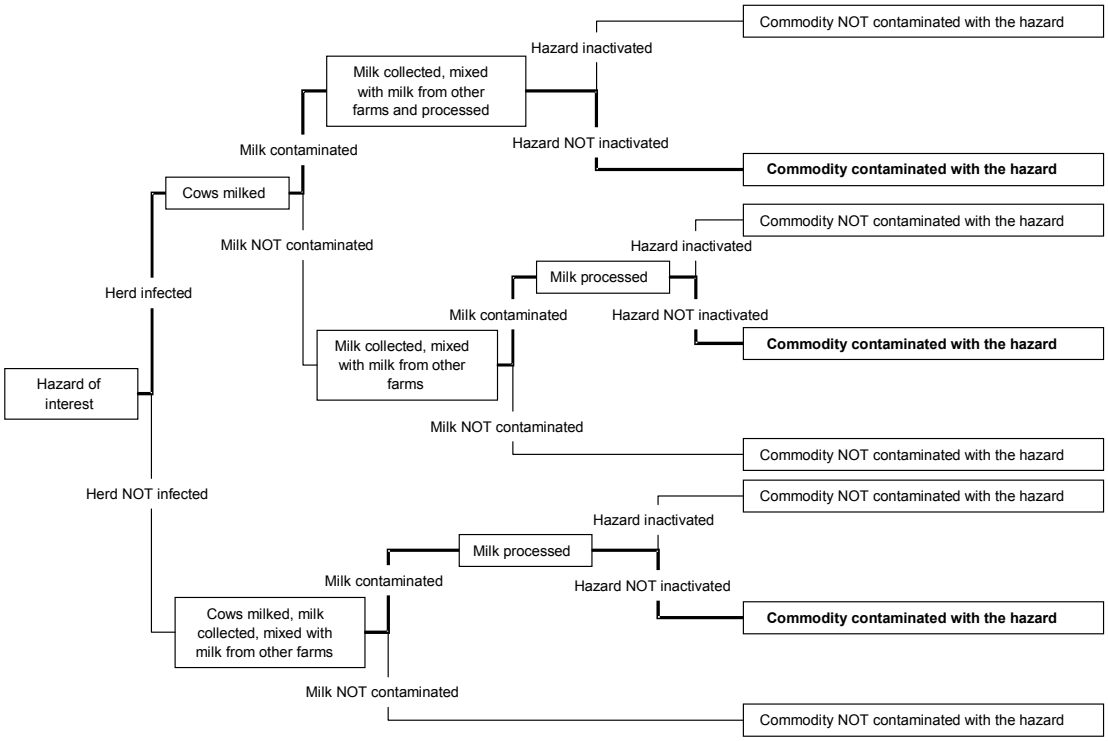
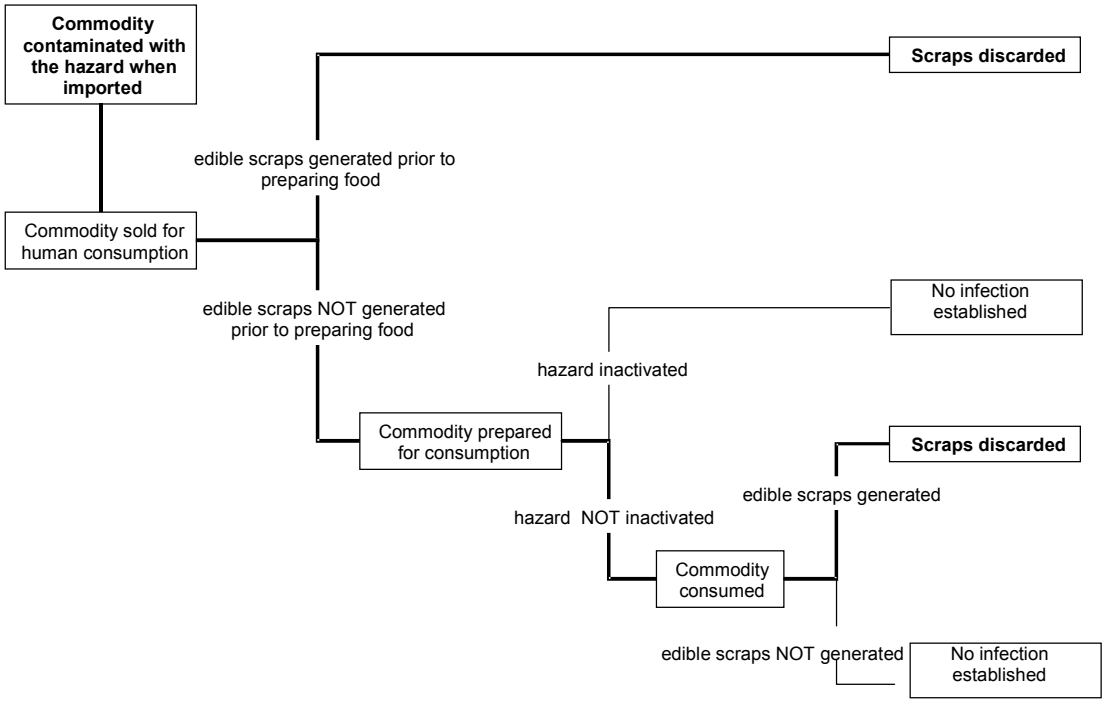


Figure 67
A release assessment scenario tree for dairy products outlining the pathways leading to contamination of an imported commodity

Exposure assessment [Part 1]



Exposure assessment [Part 2]

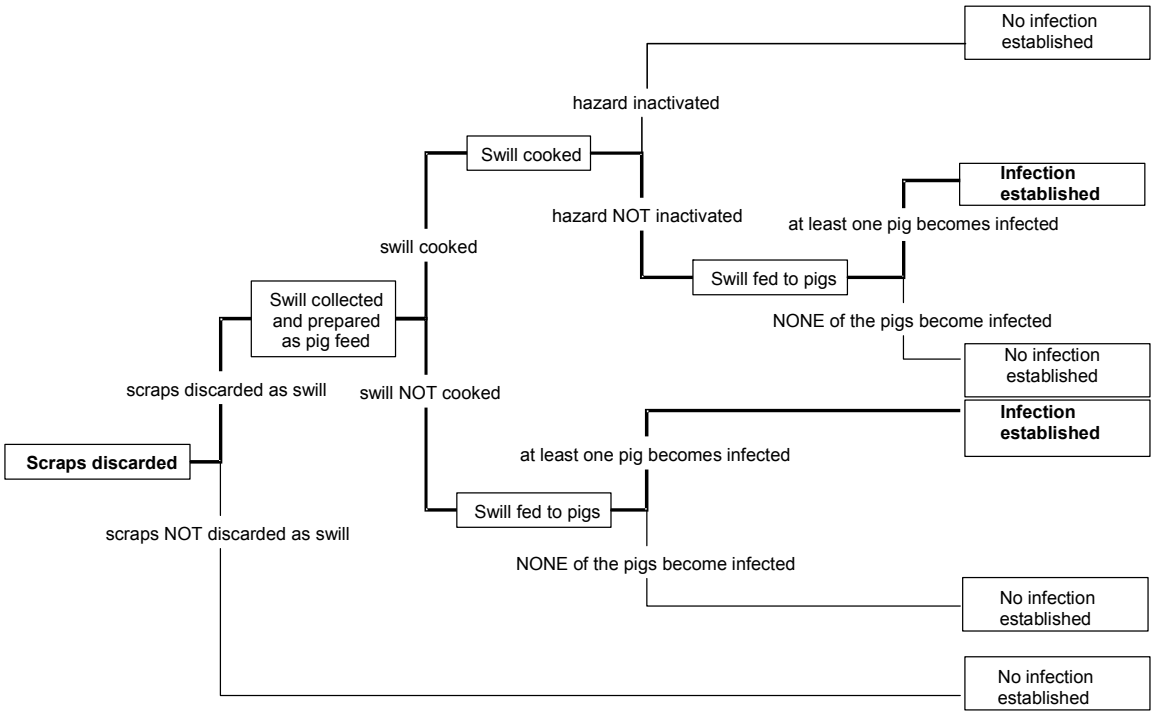


Figure 68
Exposure assessment scenario trees outlining the pathways leading to susceptible animals being exposed to contaminated imported commodity, resulting in infection

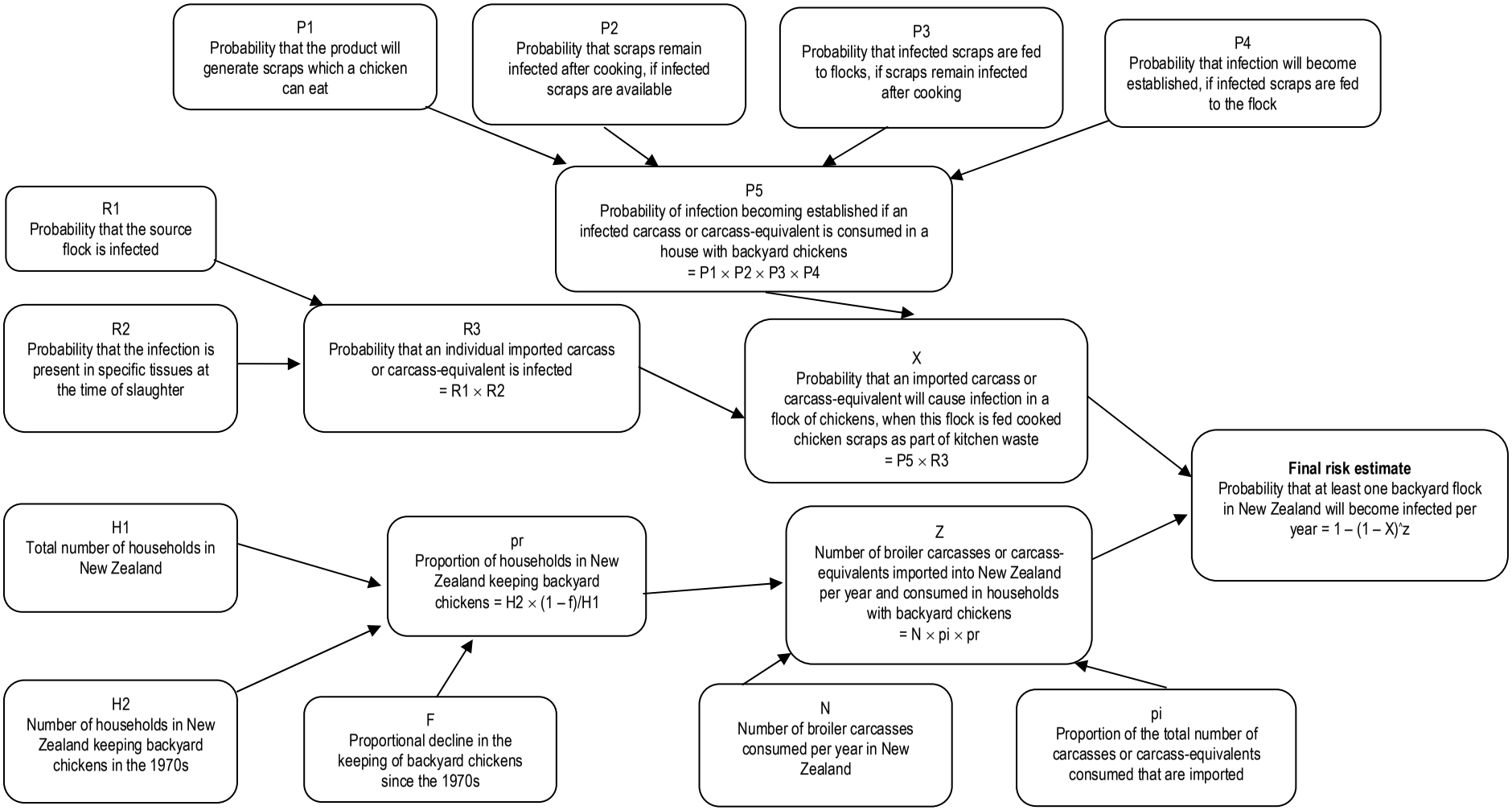


Figure 69
An example of an influence diagram
Modelling the risk of introducing infectious bursal disease virus in imported chicken meat and its establishment in backyard poultry flocks in New Zealand

Simplicity

The purpose of modelling is to represent as accurately as necessary the system of interest. Models are never more than an approximation of reality, so it is important to keep them as simple as you can without sacrificing usefulness. In most models, the outcome is driven by a handful of variables only. Simple models are more transparent, easier to use and explain to interested parties.

Accounting for independence between units

Some of the calculations used in quantitative risk assessments assume that the variables are independent. For example, Equation 62 below[26], which calculates the probability of including at least one infected animal (D^+) in a lot of n animals selected at random from a particular herd, assumes that each of the n animals is independent:

$$P(D^+ \geq 1) = 1 - (1 - p)^n \qquad \text{Equation 62}$$

where p = prevalence of infection within a herd.

If we extend this scenario to selecting k lots of the same size, each from a different herd, and apply the same method of calculation, then from Equation 63, the probability of selecting at least one infected animal (D^+) is:

$$P(D^+ \geq 1) = 1 - \left[1 - HP \times \left(1 - (1 - p)^n\right)\right]^k \qquad \text{Equation 63}$$

where HP is the herd level prevalence (proportion of infected herds).

This calculation is correct so long as we assume that the prevalence of infection within each herd is the same. If we attempt to account for a different prevalence within each herd by modelling it as a *PERT*(1%, 2%, 5%) distribution and substitute this distribution into Equation 63, we would be effectively saying that each of the k lots are selected from herds with exactly the same prevalence of infection. This would not be biologically plausible and we would be ignoring the fact that prevalence is likely to vary among herds. That is, each herd is an independent unit. Figure 70 compares the results of two models, one ignoring that each herd is an independent unit and the other that considers them to be independent. As we can see the 50th percentiles are the same, but there is a greater spread in the results when we do not consider them as independent units. If we report the results based on the tails of the distribution, for example the 95th percentile, we will overestimate the probability of importing at least one infected lot. (However, in this particular case the magnitude of the difference between the two 95th percentiles is quite small, 23% versus 19%.)

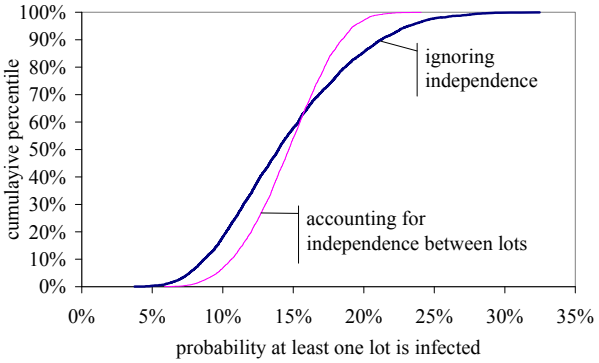


Figure 70
A comparison of a model in which independence between units is ignored with one where independence between units is taken into account

[26] This and subsequent equations have already been discussed in Chapter 5

The discussion on the central limit theorem in Chapter 7 gives other examples of where it may be important to account for independence between units.

So, how important is it to account for independence between units? The answer will depend to some extent on the estimate required. If the estimate being reported is the median (50th percentile), it may not be critical. However, if the estimate is derived from the tails of a distribution, for example the 95th percentile, it will be exaggerated, although the difference may not be great.

Independence and dependence or correlation between variables

Ideally, a quantitative risk assessment model should be structured so that the input variables are independent. If there is a dependence or correlation between two or more variables, the joint probabilities for the various combinations, when based on the product rule, will be incorrect and implausible scenarios will be generated. Two examples are provided:

– It is often assumed that the results of several tests conducted on an animal will be independent of each other. Depending on the circumstances, there may be a significant correlation between test results. For tests which measure similar biological responses to infectious agents, a positive correlation may occur in either an infected or uninfected animal. False negative results following repeated tests may occur early in the infectious process or late in the process if latent infections or intracellular infections occur.

  For example, we might decide to test some cattle from a particular farm for Johne's disease (paratuberculosis) using an ELISA, accept only those that are test negative, move them to a quarantine facility and test them again using the same test, accepting only those that test negative. Can we reasonably assume that both test results are independent? Probably not, since Johne's disease is a chronic disease. It is unlikely that it will have progressed in just a few weeks to the point where a previously negative cow has seroconverted and will thus return a positive test on the second occasion. If we calculate the probability of including at least one infected cow among the group of negative animals and do not take into account that the test results are likely to be significantly correlated, we will under-estimate the likelihood of introducing an infected animal. Of course, if the disease is an acute viral disease with a short incubation period and there is a continual opportunity of exposure during the interval between tests, it would be reasonable to assume that the test results are independent.

– As the prevalence of furunculosis (a disease that causes skin lesions in salmonid fish) increases, clinical manifestations of disease are more likely to be seen. In such circumstances it might be expected that visual inspection and grading would become more effective in identifying and excluding infected fish from the processing chain. Since there is a correlation between disease prevalence and the effectiveness of inspection and grading, the two distribution defining these variables need to be linked to ensure that sensible scenarios are modelled. In this example, one would want to avoid simulating scenarios that include inputs of low disease prevalence with a high level of effectiveness for inspection and grading.

Data and information

Whether one is conducting qualitative or quantitative risk assessments, there are many questions that need to be posed and answered in order for the correct data and information to be identified and obtained. For example:

– Are there abundant representative data covering the population of interest so that you can reasonably estimate a value for each of the variable's parameters?

– Are the data representative of the population of interest but based on a small sample size?

– In the absence of data from the population of interest, are there possibly relevant data available from other, similar populations?

In the field of import risk analysis in particular, and animal health risk analysis in general, it is common for the analyst to find that data are lacking. The analyst then has to rely on a combination of limited data and expert opinion, or perhaps just expert opinion alone when no data at all are available.

Potential sources of data and information for qualitative and quantitative risk assessments include:

– research findings published in refereed journals
– textbooks
– official reports such as the OIE's web site, *Bulletin* and *World Animal Health*
– veterinary services of trading partners
– industry sources
– expert opinion.

Modelling a variable

For each variable which is to be modelled in a risk assessment, the analyst should:

a) document the evidence, data, assumptions and uncertainties

b) decide whether to model it as a point estimate or to use a probability distribution

c) select an appropriate probability distribution to represent the variable

d) ensure the chosen distribution is biologically plausible and not simply selected because it provides a 'good fit' to the data. Careful consideration needs to be given to the underlying phenomena that generated the data. There are several techniques to assist in developing an appropriate distribution from the available information. These have been discussed in Chapter 6 where three approaches to developing a distribution were described:
   – fitting empirical data to a distribution using either parametric or non-parametric techniques
   – a purely subjective approach using expert opinion
   – a combined approach that incorporates empirical data and expert opinion using Bayes' theorem.

Separating uncertainty and variability

As discussed in Chapter 7, for some variables in a quantitative risk assessment model, the exact value of the parameter(s) that characterise a distribution may be known, or there may be abundant representative data and it can be assumed that the parameter(s) derived from these data are the population parameters. On the other hand, for some variables, the

parameter(s) that characterise a distribution will be uncertain, as for example where there are few representative data, where there are no data, or where data are not representative.

Since the separation of uncertainty and variability can be complex, it is important to investigate the potential impact of uncertainty to determine if it is reasonable to ignore it and model the uncertain and variable components together. Chapter 7 provided an approach to deciding if it is worth the effort to separate these components and develop a second order model.

Ensuring a model generates plausible scenarios

It is essential to ensure that the output from each distribution in the model, and the overall results for each and every iteration are biologically plausible. Calculations need to be checked thoroughly to ensure that unexpected results are not generated. Some distributions, such as the normal distribution, may need to be truncated to ensure that only those values within the plausible range are included. Repeated use of the recalculation key, for example the F9 key in Excel, as the model is being developed, is helpful in ensuring that each iteration is plausible.

Verifying calculations

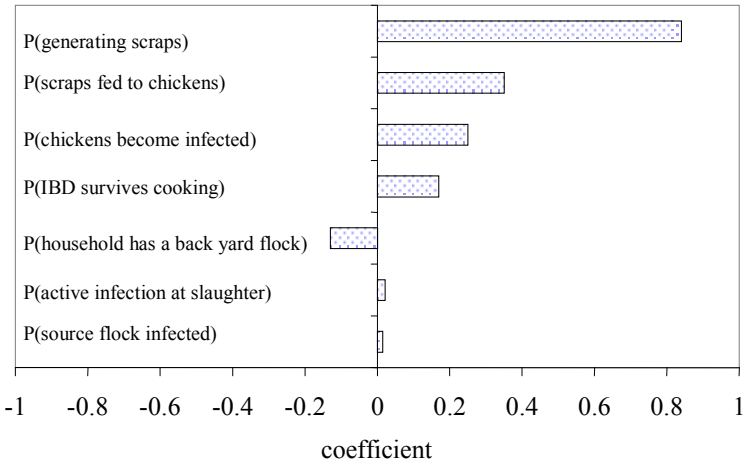
It is important to ensure that the model is mathematically correct and that the inputs are specified appropriately. Make sure that changes in inputs result in the expected changes in output. If the results are counter-intuitive, then the reasons need to be ascertained. Are they unexpected but reasonable, or do they reflect an error?

Sensitivity analysis

Sensitivity analysis is used to identify the most influential variable(s) in a quantitative model. An awareness of which inputs are most influential in determining the output may be desirable for a number of reasons. Exploring the inputs and results promotes a better understanding and interpretation of the analysis and provides a basis for gathering further information and prioritising future research. Where a correlation is believed to exist between input variables a sensitivity analysis can also help determine if its existence could affect the model's results.

There are a number of techniques available to carry out sensitivity analysis, but the most common involves determining the degree of correlation between the output variables and their associated inputs. Correlation is a quantitative measurement of the strength of the relationship. The degree of correlation can be calculated by either rank order correlation or multivariate stepwise regression. Rank order correlation is generally preferred as no assumptions are made about the nature of the relationship. In contrast, multivariate stepwise regression assumes there is a linear relationship between the variables.

The correlation coefficients calculated in a sensitivity analysis can be plotted on a tornado chart (Fig. 71). The length of the bars represents the degree of correlation between each input variable and the output. The higher the degree of correlation the more the input variable is affecting the output. A tornado chart is a useful tool for depicting the most influential variables(s). For example, in Figure 71 the probability of at least one chicken being infected with IBD is highly dependent on the probability of scraps being generated. Scatter plots also provide a way of visualising and investigating the nature of these relationships (Fig. 72).



P(generating scraps)	P(scraps fed to chickens)	P(chickens become infected)	P(IBD virus survives cooking)	P(household has a backyard flock)	P(active infection at slaughter)	P(source flock infected)
0.84	0.35	0.25	0.17	-0.13	0.02	0.01

Figure 71
A tornado chart of a rank order correlation sensitivity analysis of the probability of at least one chicken in a backyard flock becoming infected with IBD virus as a result of importing boneless chicken meat[27]

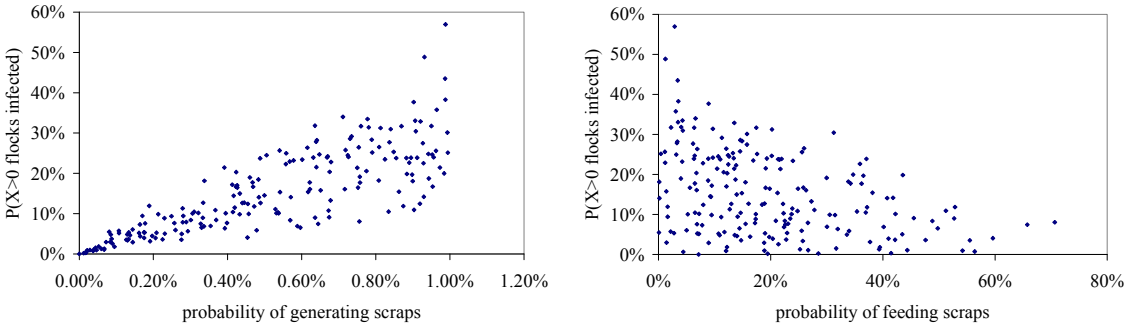


Figure 72
Two scatter plots demonstrating the effect of a model input on the model output[28]

Presenting the results

To facilitate communication of the results of a quantitative risk assessment it is important to:

– restate the question that has been asked
– explain the model's structure clearly with the aid of appropriate diagrams, such as scenario trees
– document all the evidence, data and assumptions, including their references

[27] MAF Regulatory Authority. Import Risk Analysis: chicken meat and chicken meat products; Bernard Matthews Foods Ltd turkey meat preparations from the United Kingdom. Wellington, New Zealand, 1999

[28] Based on MAF Regulatory Authority. Import Risk Analysis: chicken meat and chicken meat products; Bernard Matthews Foods Ltd turkey meat preparations from the United Kingdom. Wellington, New Zealand, 1999

– use clearly labelled, uncluttered graphs. Histograms, cumulative frequency plots, scatter plots, and tornado charts are generally the most useful
– avoid reporting results to more than one or two decimal places as reporting to several decimal places implies a level of precision that is usually unattainable. One should consider reporting the results to the nearest order of magnitude only
– ensure the report is as focused and as uncluttered as possible
– keep any statistics to a minimum
– verbal communication of the results ensures a better understanding of the problem and the outcome of the risk assessment.

Peer review

As discussed in Volume 1, peer review is important in ensuring that the risk analysis is based on the most up to date and credible information available. For a quantitative model, peer review is intended also to ensure that the distributions used and the mathematical structure are appropriate.

In some instances it will be possible to use a model which has already been subjected to a process of peer review. In such cases, only the new data inputs will need reviewing.

Appendices

Appendix 1
Table of exact binomial confidence limits

Confidence intervals (%) for the binomial distribution (N = 1-20)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p	Confidence intervals			
	$\left(\frac{r}{N}\times 100\right)$	95%		99%	
		Lower	Upper	Lower	Upper
N = 2					
0	0.00	0.00	84.19	0.00	92.93
1	50.00	1.26	98.74	0.25	99.75
2	100.00	15.81	100.00	7.07	100.00
N = 3					
0	0.00	0.00	70.76	0.00	82.90
1	33.33	0.84	90.57	0.17	95.86
2	66.67	9.43	99.16	4.14	99.83
3	100.00	29.24	100.00	17.10	100.00
N = 4					
0	0.00	0.00	60.24	0.00	73.41
1	25.00	0.63	80.59	0.13	88.91
2	50.00	6.76	93.24	2.94	97.06
3	75.00	19.41	99.37	11.09	99.87
4	100.00	39.76	100.00	26.59	100.00
N = 5					
0	0.00	0.00	52.18	0.00	65.34
1	20.00	0.51	71.64	0.10	81.49
2	40.00	5.27	85.34	2.29	91.72
3	60.00	14.66	94.73	8.28	97.71
4	80.00	28.36	99.49	18.51	99.90
5	100.00	47.82	100.00	34.66	100.00
N = 6					
0	0.00	0.00	45.93	0.00	58.65
1	16.67	0.42	64.12	0.08	74.60
2	33.33	4.33	77.72	1.87	85.64
3	50.00	11.81	88.19	6.63	93.37
4	66.67	22.28	95.67	14.36	98.13
5	83.33	35.88	99.58	25.40	99.92
6	100.00	54.07	100.00	41.35	100.00
N = 7					
0	0.00	0.00	40.96	0.00	53.09
1	14.29	0.36	57.87	0.07	68.49
2	28.57	3.67	70.96	1.58	79.70
3	42.86	9.90	81.59	5.53	88.23
4	57.14	18.41	90.10	11.77	94.47
5	71.43	29.04	96.33	20.30	98.42
6	85.71	42.13	99.64	31.51	99.93
7	100.00	59.04	100.00	46.91	100.00
N = 8					
0	0.00	0.00	36.94	0.00	48.43
1	12.50	0.32	52.65	0.06	63.15
2	25.00	3.19	65.09	1.37	74.22
3	37.50	8.52	75.51	4.75	83.03
4	50.00	15.70	84.30	9.99	90.01
5	62.50	24.49	91.48	16.97	95.25
6	75.00	34.91	96.81	25.78	98.63
7	87.50	47.35	99.68	36.85	99.94
8	100.00	63.06	100.00	51.57	100.00
N = 9					
0	0.00	0.00	33.63	0.00	44.50
1	11.11	0.28	48.25	0.06	58.50
2	22.22	2.81	60.01	1.21	69.26
3	33.33	7.49	70.07	4.16	78.09
4	44.44	13.70	78.80	8.68	85.39
5	55.56	21.20	86.30	14.61	91.32
6	66.67	29.93	92.51	21.91	95.84
7	77.78	39.99	97.19	30.74	98.79
8	88.89	51.75	99.72	41.50	99.94
9	100.00	66.37	100.00	55.50	100.00
N = 10					
0	0.00	0.00	30.85	0.00	41.13
1	10.00	0.25	44.50	0.05	54.43
2	20.00	2.52	55.61	1.09	64.82
3	30.00	6.67	65.25	3.70	73.51
4	40.00	12.16	73.76	7.68	80.91
5	50.00	18.71	81.29	12.83	87.17
6	60.00	26.24	87.84	19.09	92.32
7	70.00	34.75	93.33	26.49	96.30
8	80.00	44.39	97.48	35.18	98.91
9	90.00	55.50	99.75	45.57	99.95
10	100.00	69.15	100.00	58.87	100.00
N = 11					
0	0.00	0.00	28.49	0.00	38.22
1	9.09	0.23	41.28	0.05	50.86
2	18.18	2.28	51.78	0.98	60.85
3	27.27	6.02	60.97	3.33	69.33
4	36.36	10.93	69.21	6.88	76.68
5	45.45	16.75	76.62	11.45	83.07
6	54.55	23.38	83.25	16.93	88.55
7	63.64	30.79	89.07	23.32	93.12
8	72.73	39.03	93.98	30.67	96.67
9	81.82	48.22	97.72	39.15	99.02
10	90.91	58.72	99.77	49.14	99.95
11	100.00	71.51	100.00	61.78	100.00

r	p	Confidence intervals			
	$\left(\frac{r}{N}\times 100\right)$	95%		99%	
		Lower	Upper	Lower	Upper
N = 12					
0	0.00	0.00	26.46	0.00	35.69
1	8.33	0.21	38.48	0.04	47.70
2	16.67	2.09	48.41	0.90	57.30
3	25.00	5.49	57.19	3.03	65.52
4	33.33	9.92	65.11	6.24	72.75
5	41.67	15.17	72.33	10.34	79.15
6	50.00	21.09	78.91	15.22	84.78
7	58.33	27.67	84.83	20.85	89.66
8	66.67	34.89	90.08	27.25	93.76
9	75.00	42.81	94.51	34.48	96.97
10	83.33	51.59	97.91	42.70	99.10
11	91.67	61.52	99.79	52.30	99.96
12	100.00	73.54	100.00	64.31	100.00
N = 13					
0	0.00	0.00	24.71	0.00	33.47
1	7.69	0.19	36.03	0.04	44.90
2	15.38	1.92	45.45	0.83	54.10
3	23.08	5.04	53.81	2.78	62.06
4	30.77	9.09	61.43	5.71	69.13
5	38.46	13.86	68.42	9.42	75.46
6	46.15	19.22	74.87	13.83	81.13
7	53.85	25.13	80.78	18.87	86.17
8	61.54	31.58	86.14	24.54	90.58
9	69.23	38.57	90.91	30.87	94.29
10	76.92	46.19	94.96	37.94	97.22
11	84.62	54.55	98.08	45.90	99.17
12	92.31	63.97	99.81	55.10	99.96
13	100.00	75.29	100.00	66.53	100.00
N = 14					
0	0.00	0.00	23.16	0.00	31.51
1	7.14	0.18	33.87	0.04	42.40
2	14.29	1.78	42.81	0.76	51.23
3	21.43	4.66	50.80	2.57	58.92
4	28.57	8.39	58.10	5.26	65.79
5	35.71	12.76	64.86	8.66	72.01
6	42.86	17.66	71.14	12.67	77.66
7	50.00	23.04	76.96	17.24	82.76
8	57.14	28.86	82.34	22.34	87.33
9	64.29	35.14	87.24	27.99	91.34
10	71.43	41.90	91.61	34.21	94.74
11	78.57	49.20	95.34	41.08	97.43
12	85.71	57.19	98.22	48.77	99.24
13	92.86	66.13	99.82	57.60	99.96
14	100.00	76.84	100.00	68.49	100.00
N = 15					
0	0.00	0.00	21.80	0.00	29.76
1	6.67	0.17	31.95	0.03	40.16
2	13.33	1.66	40.46	0.71	48.63
3	20.00	4.33	48.09	2.39	56.05
4	26.67	7.79	55.10	4.88	62.73
5	33.33	11.82	61.62	8.01	68.82
6	40.00	16.34	67.71	11.70	74.39
7	46.67	21.27	73.41	15.87	79.49
8	53.33	26.59	78.73	20.51	84.13
9	60.00	32.29	83.66	25.61	88.30
10	66.67	38.38	88.18	31.18	91.99
11	73.33	44.90	92.21	37.27	95.12
12	80.00	51.91	95.67	43.95	97.61
13	86.67	59.54	98.34	51.37	99.29
14	93.33	68.05	99.83	59.84	99.97
15	100.00	78.20	100.00	70.24	100.00
N = 16					
0	0.00	0.00	20.59	0.00	28.19
1	6.25	0.16	30.23	0.03	38.14
2	12.50	1.55	38.35	0.67	46.28
3	18.75	4.05	45.65	2.23	53.44
4	25.00	7.27	52.38	4.55	59.91
5	31.25	11.02	58.66	7.45	65.85
6	37.50	15.20	64.57	10.86	71.32
7	43.75	19.75	70.12	14.71	76.38
8	50.00	24.65	75.35	18.97	81.03
9	56.25	29.88	80.25	23.62	85.29
10	62.50	35.43	84.80	28.68	89.14
11	68.75	41.34	88.98	34.15	92.55
12	75.00	47.62	92.73	40.09	95.45
13	81.25	54.35	95.95	46.56	97.77
14	87.50	61.65	98.45	53.72	99.33
15	93.75	69.77	99.84	61.86	99.97
16	100.00	79.41	100.00	71.81	100.00

r	p	Confidence intervals			
	$\left(\frac{r}{N}\times 100\right)$	95%		99%	
		Lower	Upper	Lower	Upper
N = 17					
0	0.00	0.00	19.51	0.00	26.78
1	5.88	0.15	28.69	0.03	36.30
2	11.76	1.46	36.44	0.63	44.13
3	17.65	3.80	43.43	2.09	51.04
4	23.53	6.81	49.90	4.26	57.32
5	29.41	10.31	55.96	6.97	63.10
6	35.29	14.21	61.67	10.14	68.46
7	41.18	18.44	67.08	13.71	73.44
8	47.06	22.98	72.19	17.64	78.07
9	52.94	27.81	77.02	21.93	82.36
10	58.82	32.92	81.56	26.56	86.29
11	64.71	38.33	85.79	31.54	89.86
12	70.59	44.04	89.69	36.90	93.03
13	76.47	50.10	93.19	42.68	95.74
14	82.35	56.57	96.20	48.96	97.91
15	88.24	63.56	98.54	55.87	99.37
16	94.12	71.31	99.85	63.70	99.97
17	100.00	80.49	100.00	73.22	100.00
N = 18					
0	0.00	0.00	18.53	0.00	25.50
1	5.56	0.14	27.29	0.03	34.63
2	11.11	1.38	34.71	0.59	42.17
3	16.67	3.58	41.42	1.97	48.84
4	22.22	6.41	47.64	4.00	54.92
5	27.78	9.69	53.48	6.54	60.55
6	33.33	13.34	59.01	9.51	65.79
7	38.89	17.30	64.25	12.84	70.68
8	44.44	21.53	69.24	16.49	75.26
9	50.00	26.02	73.98	20.47	79.53
10	55.56	30.76	78.47	24.74	83.51
11	61.11	35.75	82.70	29.32	87.16
12	66.67	40.99	86.66	34.21	90.49
13	72.22	46.52	90.31	39.45	93.46
14	77.78	52.36	93.59	45.08	96.00
15	83.33	58.58	96.42	51.16	98.03
16	88.89	65.29	98.62	57.83	99.41
17	94.44	72.71	99.86	65.37	99.97
18	100.00	81.47	100.00	74.50	100.00
N = 19					
0	0.00	0.00	17.65	0.00	24.34
1	5.26	0.13	26.03	0.03	33.11
2	10.53	1.30	33.14	0.56	40.37
3	15.79	3.38	39.58	1.86	46.82
4	21.05	6.05	45.57	3.78	52.71
5	26.32	9.15	51.20	6.17	58.18
6	31.58	12.58	56.55	8.95	63.29
7	36.84	16.29	61.64	12.07	68.09
8	42.11	20.25	66.50	15.49	72.60
9	47.37	24.45	71.14	19.19	76.84
10	52.63	28.86	75.55	23.16	80.81
11	57.89	33.50	79.75	27.40	84.51
12	63.16	38.36	83.71	31.91	87.93
13	68.42	43.45	87.42	36.71	91.05
14	73.68	48.80	90.85	41.82	93.83
15	78.95	54.43	93.95	47.29	96.22
16	84.21	60.42	96.62	53.18	98.14
17	89.47	66.86	98.70	59.63	99.44
18	94.74	73.97	99.87	66.89	99.97
19	100.00	82.35	100.00	75.66	100.00
N = 20					
0	0.00	0.00	16.84	0.00	23.27
1	5.00	0.13	24.87	0.03	31.71
2	10.00	1.23	31.70	0.53	38.71
3	15.00	3.21	37.89	1.76	44.95
4	20.00	5.73	43.66	3.58	50.66
5	25.00	8.66	49.10	5.83	55.98
6	30.00	11.89	54.28	8.46	60.96
7	35.00	15.39	59.22	11.39	65.66
8	40.00	19.12	63.95	14.60	70.09
9	45.00	23.06	68.47	18.06	74.28
10	50.00	27.20	72.80	21.77	78.23
11	55.00	31.53	76.94	25.72	81.94
12	60.00	36.05	80.88	29.91	85.40
13	65.00	40.78	84.61	34.34	88.61
14	70.00	45.72	88.11	39.04	91.54
15	75.00	50.90	91.34	44.02	94.17
16	80.00	56.34	94.27	49.34	96.42
17	85.00	62.11	96.79	55.05	98.24
18	90.00	68.30	98.77	61.29	99.47
19	95.00	75.13	99.87	68.29	99.97
20	100.00	83.16	100.00	76.73	100.00

Confidence intervals (%) for the binomial distribution (N = 21-36)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 21					
0	0.00	0.00	16.11	0.00	22.30
1	4.76	0.12	23.82	0.02	30.43
2	9.52	1.17	30.38	0.50	37.18
3	14.29	3.05	36.34	1.68	43.22
4	19.05	5.45	41.91	3.39	48.76
5	23.81	8.22	47.17	5.53	53.92
6	28.57	11.28	52.18	8.01	58.78
7	33.33	14.59	56.97	10.78	63.37
8	38.10	18.11	61.56	13.81	67.72
9	42.86	21.82	65.98	17.07	71.85
10	47.62	25.71	70.22	20.55	75.75
N = 22					
0	0.00	0.00	15.44	0.00	21.40
1	4.55	0.12	22.84	0.02	29.24
2	9.09	1.12	29.16	0.48	35.77
3	13.64	2.91	34.91	1.60	41.61
4	18.18	5.19	40.28	3.23	46.99
5	22.73	7.82	45.37	5.26	52.01
6	27.27	10.73	50.22	7.61	56.74
7	31.82	13.86	54.87	10.24	61.23
8	36.36	17.20	59.34	13.10	65.49
9	40.91	20.71	63.65	16.18	69.54
10	45.45	24.39	67.79	19.46	73.40
11	50.00	28.22	71.78	22.93	77.07
N = 23					
0	0.00	0.00	14.82	0.00	20.58
1	4.35	0.11	21.95	0.02	28.14
2	8.70	1.07	28.04	0.46	34.46
3	13.04	2.78	33.59	1.53	40.12
4	17.39	4.95	38.78	3.08	45.34
5	21.74	7.46	43.70	5.02	50.22
6	26.09	10.23	48.41	7.25	54.83
7	30.43	13.21	52.92	9.74	59.21
8	34.78	16.38	57.27	12.46	63.38
9	39.13	19.71	61.46	15.37	67.36
10	43.48	23.19	65.51	18.48	71.16
11	47.83	26.82	69.41	21.76	74.79
N = 24					
0	0.00	0.00	14.25	0.00	19.81
1	4.17	0.11	21.12	0.02	27.13
2	8.33	1.03	27.00	0.44	33.24
3	12.50	2.66	32.36	1.46	38.73
4	16.67	4.74	37.38	2.95	43.79
5	20.83	7.13	42.15	4.79	48.55
6	25.00	9.77	46.71	6.92	53.04
7	29.17	12.62	51.09	9.30	57.32
8	33.33	15.63	55.32	11.88	61.40
9	37.50	18.80	59.41	14.65	65.30
10	41.67	22.11	63.36	17.59	69.04
11	45.83	25.55	67.18	20.70	72.62
12	50.00	29.12	70.88	23.96	76.04
N = 25					
0	0.00	0.00	13.72	0.00	19.10
1	4.00	0.10	20.35	0.02	26.18
2	8.00	0.98	26.03	0.42	32.10
3	12.00	2.55	31.22	1.40	37.43
4	16.00	4.54	36.08	2.82	42.35
5	20.00	6.83	40.70	4.59	46.98
6	24.00	9.36	45.13	6.63	51.36
7	28.00	12.07	49.39	8.89	55.53
8	32.00	14.95	53.50	11.35	59.52
9	36.00	17.97	57.48	13.99	63.35
10	40.00	21.13	61.33	16.79	67.02
11	44.00	24.40	65.07	19.74	70.54
12	48.00	27.80	68.69	22.83	73.93
N = 26					
0	0.00	0.00	13.23	0.00	18.44
1	3.85	0.10	19.64	0.02	25.29
2	7.69	0.95	25.13	0.41	31.04
3	11.54	2.45	30.15	1.34	36.21
4	15.38	4.36	34.87	2.71	41.00
5	19.23	6.55	39.35	4.40	45.50
6	23.08	8.97	43.65	6.35	49.77
7	26.92	11.57	47.79	8.52	53.85
8	30.77	14.33	51.79	10.87	57.75
9	34.62	17.21	55.67	13.38	61.50
10	38.46	20.23	59.43	16.05	65.10
11	42.31	23.35	63.08	18.86	68.57
12	46.15	26.59	66.63	21.81	71.91
13	50.00	29.93	70.07	24.89	75.11
N = 27					
0	0.00	0.00	12.77	0.00	17.82
1	3.70	0.09	18.97	0.02	24.46
2	7.41	0.91	24.29	0.39	30.04

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 27 (continued)					
3	11.11	2.35	29.16	1.29	35.07
4	14.81	4.19	33.73	2.60	39.73
5	18.52	6.30	38.08	4.23	44.11
6	22.22	8.62	42.26	6.10	48.28
7	25.93	11.11	46.28	8.17	52.26
8	29.63	13.75	50.18	10.42	56.08
9	33.33	16.52	53.96	12.83	59.75
10	37.04	19.40	57.63	15.38	63.28
11	40.74	22.39	61.20	18.07	66.69
12	44.44	25.48	64.67	20.88	69.98
13	48.15	28.67	68.05	23.81	73.14
N = 28					
0	0.00	0.00	12.34	0.00	17.24
1	3.57	0.09	18.35	0.02	23.69
2	7.14	0.88	23.50	0.38	29.11
3	10.71	2.27	28.23	1.24	33.99
4	14.29	4.03	32.67	2.51	38.53
5	17.86	6.06	36.89	4.07	42.80
6	21.43	8.30	40.95	5.86	46.87
7	25.00	10.69	44.87	7.86	50.76
8	28.57	13.22	48.67	10.02	54.49
9	32.14	15.88	52.35	12.32	58.08
10	35.71	18.64	55.94	14.77	61.55
11	39.29	21.50	59.42	17.33	64.90
12	42.86	24.46	62.82	20.02	68.14
13	46.43	27.51	66.13	22.82	71.26
14	50.00	30.65	69.35	25.72	74.28
N = 29					
0	0.00	0.00	11.94	0.00	16.70
1	3.45	0.09	17.76	0.02	22.96
2	6.90	0.85	22.77	0.36	28.23
3	10.34	2.19	27.35	1.20	32.98
4	13.79	3.89	31.66	2.42	37.40
5	17.24	5.85	35.77	3.92	41.57
6	20.69	7.99	39.72	5.65	45.54
7	24.14	10.30	43.54	7.56	49.33
8	27.59	12.73	47.24	9.64	52.99
9	31.03	15.28	50.83	11.85	56.51
10	34.48	17.94	54.33	14.20	59.91
11	37.93	20.69	57.74	16.66	63.20
12	41.38	23.52	61.06	19.23	66.38
13	44.83	26.45	64.31	21.91	69.46
14	48.28	29.45	67.47	24.69	72.43
N = 30					
0	0.00	0.00	11.57	0.00	16.19
1	3.33	0.08	17.22	0.02	22.28
2	6.67	0.82	22.07	0.35	27.40
3	10.00	2.11	26.53	1.16	32.03
4	13.33	3.76	30.72	2.33	36.34
5	16.67	5.64	34.72	3.78	40.40
6	20.00	7.71	38.57	5.45	44.28
7	23.33	9.93	42.28	7.29	47.99
8	26.67	12.28	45.89	9.29	51.56
9	30.00	14.73	49.40	11.42	55.01
10	33.33	17.29	52.81	13.67	58.34
11	36.67	19.93	56.14	16.04	61.57
12	40.00	22.66	59.40	18.50	64.70
13	43.33	25.46	62.57	21.07	67.73
14	46.67	28.34	65.67	23.73	70.67
15	50.00	31.30	68.70	26.48	73.52
N = 31					
0	0.00	0.00	11.22	0.00	15.71
1	3.23	0.08	16.70	0.02	21.63
2	6.45	0.79	21.42	0.34	26.62
3	9.68	2.04	25.75	1.12	31.13
4	12.90	3.63	29.83	2.25	35.33
5	16.13	5.45	33.73	3.65	39.30
6	19.35	7.45	37.47	5.26	43.08
7	22.58	9.59	41.10	7.04	46.71
8	25.81	11.86	44.61	8.96	50.21
9	29.03	14.22	48.04	11.02	53.58
10	32.26	16.68	51.37	13.18	56.85
11	35.48	19.23	54.63	15.46	60.02
12	38.71	21.85	57.81	17.83	63.09
13	41.94	24.55	60.92	20.29	66.08
14	45.16	27.32	63.97	22.85	68.98
15	48.39	30.15	66.94	25.49	71.79
N = 32					
0	0.00	0.00	10.89	0.00	15.26
1	3.13	0.08	16.22	0.02	21.02
2	6.25	0.77	20.81	0.33	25.88
3	9.38	1.98	25.02	1.08	30.28
4	12.50	3.51	28.99	2.18	34.38
5	15.63	5.28	32.79	3.53	38.25

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 32 (continued)					
6	18.75	7.21	36.44	5.09	41.95
7	21.88	9.28	39.97	6.80	45.50
8	25.00	11.46	43.40	8.66	48.92
9	28.13	13.75	46.75	10.64	52.23
10	31.25	16.12	50.01	12.73	55.43
11	34.38	18.57	53.19	14.92	58.54
12	37.50	21.10	56.31	17.20	61.56
13	40.63	23.70	59.36	19.57	64.50
14	43.75	26.36	62.34	22.03	67.35
15	46.88	29.09	65.26	24.56	70.13
16	50.00	31.89	68.11	27.18	72.82
N = 33					
0	0.00	0.00	10.58	0.00	14.83
1	3.03	0.08	15.76	0.02	20.44
2	6.06	0.74	20.23	0.32	25.18
3	9.09	1.92	24.33	1.05	29.47
4	12.12	3.40	28.20	2.11	33.47
5	15.15	5.11	31.90	3.42	37.26
6	18.18	6.98	35.46	4.92	40.87
7	21.21	8.98	38.91	6.58	44.34
8	24.24	11.09	42.26	8.38	47.69
9	27.27	13.30	45.52	10.29	50.93
10	30.30	15.59	48.71	12.31	54.08
11	33.33	17.96	51.83	14.42	57.13
12	36.36	20.40	54.88	16.62	60.10
13	39.39	22.91	57.86	18.90	62.98
14	42.42	25.48	60.78	21.27	65.79
15	45.45	28.11	63.65	23.71	68.53
16	48.48	30.80	66.46	26.22	71.19
N = 34					
0	0.00	0.00	10.28	0.00	14.43
1	2.94	0.07	15.33	0.01	19.90
2	5.88	0.72	19.68	0.31	24.52
3	8.82	1.86	23.68	1.02	28.71
4	11.76	3.30	27.45	2.05	32.62
5	14.71	4.95	31.06	3.32	36.31
6	17.65	6.76	34.53	4.77	39.85
7	20.59	8.70	37.90	6.38	43.24
8	23.53	10.75	41.17	8.11	46.52
9	26.47	12.88	44.36	9.96	49.70
10	29.41	15.10	47.48	11.91	52.78
11	32.35	17.39	50.53	13.95	55.78
12	35.29	19.75	53.51	16.07	58.69
13	38.24	22.17	56.44	18.28	61.53
14	41.18	24.65	59.30	20.56	64.30
15	44.12	27.19	62.11	22.91	67.00
16	47.06	29.78	64.87	25.33	69.62
17	50.00	32.43	67.57	27.82	72.18
N = 35					
0	0.00	0.00	10.00	0.00	14.05
1	2.86	0.07	14.92	0.01	19.38
2	5.71	0.70	19.16	0.30	23.89
3	8.57	1.80	23.06	0.99	27.98
4	11.43	3.20	26.74	1.99	31.80
5	14.29	4.81	30.26	3.22	35.42
6	17.14	6.56	33.65	4.63	38.87
7	20.00	8.44	36.94	6.18	42.20
8	22.86	10.42	40.14	7.86	45.41
9	25.71	12.49	43.26	9.65	48.52
10	28.57	14.64	46.30	11.54	51.55
11	31.43	16.85	49.29	13.51	54.49
12	34.29	19.13	52.21	15.56	57.35
13	37.14	21.47	55.08	17.69	60.14
14	40.00	23.87	57.89	19.89	62.87
15	42.86	26.32	60.65	22.16	65.52
16	45.71	28.83	63.35	24.50	68.11
17	48.57	31.38	66.01	26.90	70.64
N = 36					
0	0.00	0.00	9.74	0.00	13.69
1	2.78	0.07	14.53	0.01	18.89
2	5.56	0.68	18.66	0.29	23.30
3	8.33	1.75	22.47	0.96	27.29
4	11.11	3.11	26.06	1.93	31.02
5	13.89	4.67	29.50	3.12	34.56
6	16.67	6.37	32.81	4.49	37.94
7	19.44	8.19	36.02	6.00	41.20
8	22.22	10.12	39.15	7.63	44.35
9	25.00	12.12	42.20	9.36	47.40
10	27.78	14.20	45.19	11.19	50.37
11	30.56	16.35	48.11	13.10	53.25
12	33.33	18.56	50.97	15.09	56.07
13	36.11	20.82	53.78	17.14	58.81
14	38.89	23.14	56.54	19.27	61.49
15	41.67	25.51	59.24	21.46	64.11

Confidence intervals (%) for the binomial distribution (N = 37-47)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 36 (continued)					
16	44.44	27.94	61.90	23.72	66.66
17	47.22	30.41	64.51	26.03	69.16
18	50.00	32.92	67.08	28.41	71.59
N = 37					
0	0.00	0.00	9.49	0.00	13.34
1	2.70	0.07	14.16	0.01	18.42
2	5.41	0.66	18.19	0.28	22.73
3	8.11	1.70	21.91	0.93	26.63
4	10.81	3.03	25.42	1.88	30.28
5	13.51	4.54	28.77	3.04	33.75
6	16.22	6.19	32.01	4.36	37.06
7	18.92	7.96	35.16	5.83	40.25
8	21.62	9.83	38.21	7.41	43.33
9	24.32	11.77	41.20	9.09	46.33
10	27.03	13.79	44.12	10.86	49.24
11	29.73	15.87	46.98	12.71	52.07
12	32.43	18.01	49.79	14.64	54.83
13	35.14	20.21	52.54	16.63	57.53
14	37.84	22.46	55.24	18.69	60.17
15	40.54	24.75	57.90	20.81	62.75
16	43.24	27.10	60.51	22.99	65.26
17	45.95	29.49	63.08	25.22	67.73
18	48.65	31.92	65.60	27.52	70.13
N = 38					
0	0.00	0.00	9.25	0.00	13.01
1	2.63	0.07	13.81	0.01	17.98
2	5.26	0.64	17.75	0.28	22.19
3	7.89	1.66	21.38	0.91	26.01
4	10.53	2.94	24.80	1.83	29.58
5	13.16	4.41	28.09	2.95	32.97
6	15.79	6.02	31.25	4.24	36.21
7	18.42	7.74	34.33	5.67	39.34
8	21.05	9.55	37.32	7.20	42.36
9	23.68	11.44	40.24	8.83	45.30
10	26.32	13.40	43.10	10.55	48.15
11	28.95	15.42	45.90	12.35	50.94
12	31.58	17.50	48.65	14.21	53.65
13	34.21	19.63	51.35	16.14	56.31
14	36.84	21.81	54.01	18.14	58.90
15	39.47	24.04	56.61	20.19	61.44
16	42.11	26.31	59.18	22.30	63.92
17	44.74	28.62	61.70	24.47	66.35
18	47.37	30.98	64.18	26.68	68.72
19	50.00	33.38	66.62	28.95	71.05
N = 39					
0	0.00	0.00	9.03	0.00	12.70
1	2.56	0.06	13.48	0.01	17.56
2	5.13	0.63	17.32	0.27	21.67
3	7.69	1.62	20.87	0.89	25.41
4	10.26	2.87	24.22	1.78	28.91
5	12.82	4.30	27.43	2.87	32.22
6	15.38	5.86	30.53	4.13	35.40
7	17.95	7.54	33.53	5.51	38.47
8	20.51	9.30	36.46	7.00	41.43
9	23.08	11.13	39.33	8.59	44.31
10	25.64	13.04	42.13	10.26	47.12
11	28.21	15.00	44.87	12.00	49.85
12	30.77	17.02	47.57	13.81	52.52
13	33.33	19.09	50.22	15.69	55.13
14	35.90	21.20	52.82	17.62	57.68
15	38.46	23.36	55.38	19.61	60.18
16	41.03	25.57	57.90	21.66	62.62
17	43.59	27.81	60.38	23.75	65.02
18	46.15	30.09	62.82	25.90	67.36
19	48.72	32.42	65.22	28.10	69.66
N = 40					
0	0.00	0.00	8.81	0.00	12.41
1	2.50	0.06	13.16	0.01	17.15
2	5.00	0.61	16.92	0.26	21.18
3	7.50	1.57	20.39	0.86	24.84
4	10.00	2.79	23.66	1.73	28.26
5	12.50	4.19	26.80	2.80	31.51
6	15.00	5.71	29.84	4.02	34.63
7	17.50	7.34	32.78	5.37	37.63
8	20.00	9.05	35.65	6.82	40.54
9	22.50	10.84	38.45	8.36	43.37
10	25.00	12.69	41.20	9.98	46.12
11	27.50	14.60	43.89	11.68	48.81
12	30.00	16.56	46.53	13.44	51.43
13	32.50	18.57	49.13	15.26	54.00
14	35.00	20.63	51.68	17.13	56.51
15	37.50	22.73	54.20	19.06	58.97
16	40.00	24.87	56.67	21.05	61.38
17	42.50	27.04	59.11	23.08	63.74
18	45.00	29.26	61.51	25.16	66.05
19	47.50	31.51	63.87	27.29	68.32
20	50.00	33.80	66.20	29.46	70.54
N = 41					
0	0.00	0.00	8.60	0.00	12.12
1	2.44	0.06	12.86	0.01	16.77
2	4.88	0.60	16.53	0.26	20.71
3	7.32	1.54	19.92	0.84	24.29
4	9.76	2.72	23.13	1.69	27.65
5	12.20	4.08	26.20	2.73	30.83
6	14.63	5.57	29.17	3.92	33.89
7	17.07	7.15	32.06	5.23	36.83
8	19.51	8.82	34.87	6.64	39.69
9	21.95	10.56	37.61	8.14	42.46
10	24.39	12.36	40.30	9.72	45.17
11	26.83	14.22	42.94	11.37	47.81
12	29.27	16.13	45.54	13.08	50.38
13	31.71	18.08	48.09	14.85	52.91
14	34.15	20.08	50.59	16.67	55.38
15	36.59	22.12	53.06	18.55	57.80
16	39.02	24.20	55.50	20.47	60.17
17	41.46	26.32	57.89	22.44	62.50
18	43.90	28.47	60.25	24.46	64.78
19	46.34	30.66	62.58	26.52	67.02
20	48.78	32.88	64.87	28.63	69.22
N = 42					
0	0.00	0.00	8.41	0.00	11.85
1	2.38	0.06	12.57	0.01	16.40
2	4.76	0.58	16.16	0.25	20.26
3	7.14	1.50	19.48	0.82	23.77
4	9.52	2.66	22.62	1.65	27.05
5	11.90	3.98	25.63	2.66	30.18
6	14.29	5.43	28.54	3.82	33.18
7	16.67	6.97	31.36	5.10	36.07
8	19.05	8.60	34.12	6.47	38.87
9	21.43	10.30	36.81	7.94	41.59
10	23.81	12.05	39.45	9.47	44.25
11	26.19	13.86	42.04	11.08	46.84
12	28.57	15.72	44.58	12.74	49.38
13	30.95	17.62	47.09	14.46	51.86
14	33.33	19.57	49.55	16.23	54.29
15	35.71	21.55	51.97	18.06	56.68
16	38.10	23.57	54.36	19.93	59.02
17	40.48	25.63	56.72	21.84	61.31
18	42.86	27.72	59.04	23.80	63.56
19	45.24	29.85	61.33	25.80	65.77
20	47.62	32.00	63.58	27.85	67.94
21	50.00	34.19	65.81	29.93	70.07
N = 43					
0	0.00	0.00	8.22	0.00	11.59
1	2.33	0.06	12.29	0.01	16.04
2	4.65	0.57	15.81	0.24	19.82
3	6.98	1.46	19.06	0.80	23.27
4	9.30	2.59	22.14	1.61	26.49
5	11.63	3.89	25.08	2.60	29.55
6	13.95	5.30	27.93	3.73	32.49
7	16.28	6.81	30.70	4.97	35.33
8	18.60	8.39	33.40	6.32	38.08
9	20.93	10.04	36.04	7.74	40.76
10	23.26	11.76	38.63	9.24	43.37
11	25.58	13.52	41.17	10.80	45.92
12	27.91	15.33	43.67	12.42	48.41
13	30.23	17.18	46.13	14.09	50.85
14	32.56	19.08	48.54	15.82	53.25
15	34.88	21.01	50.93	17.59	55.59
16	37.21	22.98	53.27	19.41	57.90
17	39.53	24.98	55.59	21.27	60.16
18	41.86	27.01	57.87	23.18	62.38
19	44.19	29.08	60.12	25.12	64.56
20	46.51	31.18	62.35	27.11	66.70
21	48.84	33.31	64.54	29.13	68.80
N = 44					
0	0.00	0.00	8.04	0.00	11.34
1	2.27	0.06	12.02	0.01	15.70
2	4.55	0.56	15.47	0.24	19.41
3	6.82	1.43	18.66	0.78	22.79
4	9.09	2.53	21.67	1.57	25.95
5	11.36	3.79	24.56	2.54	28.95
6	13.64	5.17	27.35	3.64	31.84
7	15.91	6.64	30.07	4.85	34.63
8	18.18	8.19	32.71	6.16	37.33
9	20.45	9.80	35.30	7.55	39.96
10	22.73	11.47	37.84	9.01	42.52
11	25.00	13.19	40.34	10.53	45.03
12	27.27	14.96	42.79	12.11	47.48
13	29.55	16.76	45.20	13.74	49.88
14	31.82	18.61	47.58	15.43	52.24
15	34.09	20.49	49.92	17.15	54.55
16	36.36	22.41	52.23	18.92	56.82
17	38.64	24.36	54.50	20.73	59.05
18	40.91	26.34	56.75	22.59	61.24
19	43.18	28.35	58.97	24.48	63.39
20	45.45	30.39	61.15	26.41	65.51
21	47.73	32.46	63.31	28.37	67.58
22	50.00	34.56	65.44	30.38	69.62
N = 45					
0	0.00	0.00	7.87	0.00	11.11
1	2.22	0.06	11.77	0.01	15.38
2	4.44	0.54	15.15	0.23	19.01
3	6.67	1.40	18.27	0.77	22.32
4	8.89	2.48	21.22	1.53	25.43
5	11.11	3.71	24.05	2.48	28.38
6	13.33	5.05	26.79	3.56	31.21
7	15.56	6.49	29.46	4.74	33.95
8	17.78	8.00	32.05	6.02	36.60
9	20.00	9.58	34.60	7.37	39.18
10	22.22	11.20	37.09	8.80	41.71
11	24.44	12.88	39.54	10.28	44.17
12	26.67	14.60	41.94	11.82	46.58
13	28.89	16.37	44.31	13.41	48.95
14	31.11	18.17	46.65	15.05	51.27
15	33.33	20.00	48.95	16.73	53.54
16	35.56	21.87	51.22	18.46	55.78
17	37.78	23.77	53.46	20.22	57.98
18	40.00	25.70	55.67	22.02	60.14
19	42.22	27.66	57.85	23.86	62.26
20	44.44	29.64	60.00	25.74	64.35
21	46.67	31.66	62.13	27.65	66.40
22	48.89	33.70	64.23	29.60	68.42
N = 46					
0	0.00	0.00	7.71	0.00	10.88
1	2.17	0.06	11.53	0.01	15.07
2	4.35	0.53	14.84	0.23	18.63
3	6.52	1.37	17.90	0.75	21.88
4	8.70	2.42	20.79	1.50	24.93
5	10.87	3.62	23.57	2.42	27.82
6	13.04	4.94	26.26	3.47	30.60
7	15.22	6.34	28.87	4.63	33.29
8	17.39	7.82	31.42	5.88	35.90
9	19.57	9.36	33.91	7.20	38.44
10	21.74	10.95	36.36	8.59	40.92
11	23.91	12.59	38.77	10.04	43.34
12	26.09	14.27	41.13	11.54	45.72
13	28.26	15.99	43.46	13.10	48.04
14	30.43	17.74	45.75	14.69	50.33
15	32.61	19.53	48.02	16.33	52.57
16	34.78	21.35	50.25	18.01	54.77
17	36.96	23.21	52.45	19.73	56.94
18	39.13	25.09	54.63	21.49	59.07
19	41.30	27.00	56.77	23.28	61.16
20	43.48	28.93	58.89	25.11	63.23
21	45.65	30.90	60.99	26.97	65.25
22	47.83	32.89	63.05	28.86	67.25
23	50.00	34.90	65.10	30.79	69.21
N = 47					
0	0.00	0.00	7.55	0.00	10.66
1	2.13	0.05	11.29	0.01	14.77
2	4.26	0.52	14.54	0.22	18.27
3	6.38	1.34	17.54	0.73	21.45
4	8.51	2.37	20.38	1.47	24.44
5	10.64	3.55	23.10	2.37	27.29
6	12.77	4.83	25.74	3.40	30.02
7	14.89	6.20	28.31	4.53	32.66
8	17.02	7.65	30.81	5.75	35.23
9	19.15	9.15	33.26	7.04	37.73
10	21.28	10.70	35.66	8.40	40.16
11	23.40	12.30	38.03	9.81	42.55
12	25.53	13.94	40.35	11.28	44.88
13	27.66	15.62	42.64	12.79	47.17
14	29.79	17.34	44.89	14.35	49.42
15	31.91	19.09	47.12	15.95	51.63
16	34.04	20.86	49.31	17.59	53.80
17	36.17	22.67	51.48	19.27	55.94
18	38.30	24.51	53.62	20.98	58.04
19	40.43	26.37	55.73	22.73	60.11
20	42.55	28.26	57.82	24.51	62.14
21	44.68	30.17	59.88	26.32	64.14
22	46.81	32.11	61.92	28.16	66.11

Confidence intervals (%) for the binomial distribution (N = 48-56)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p ($\frac{r}{N} \times 100$)	95% Lower	95% Upper	99% Lower	99% Upper
N = 47 (continued)					
23	48.94	34.08	63.94	30.04	68.05
N = 48					
0	0.00	0.00	7.40	0.00	10.45
1	2.08	0.05	11.07	0.01	14.48
2	4.17	0.51	14.25	0.22	17.91
3	6.25	1.31	17.20	0.72	21.05
4	8.33	2.32	19.98	1.44	23.98
5	10.42	3.47	22.66	2.32	26.78
6	12.50	4.73	25.25	3.32	29.46
7	14.58	6.07	27.76	4.43	32.06
8	16.67	7.48	30.22	5.62	34.58
9	18.75	8.95	32.63	6.89	37.03
10	20.83	10.47	34.99	8.21	39.43
11	22.92	12.03	37.31	9.59	41.78
12	25.00	13.64	39.60	11.03	44.08
13	27.08	15.28	41.85	12.51	46.33
14	29.17	16.95	44.06	14.03	48.55
15	31.25	18.66	46.25	15.59	50.72
16	33.33	20.40	48.41	17.19	52.86
17	35.42	22.16	50.54	18.83	54.97
18	37.50	23.95	52.65	20.50	57.04
19	39.58	25.77	54.73	22.20	59.08
20	41.67	27.61	56.79	23.93	61.09
21	43.75	29.48	58.82	25.70	63.07
22	45.83	31.37	60.83	27.50	65.01
23	47.92	33.29	62.81	29.33	66.93
24	50.00	35.23	64.77	31.18	68.82
N = 49					
0	0.00	0.00	7.25	0.00	10.25
1	2.04	0.05	10.85	0.01	14.21
2	4.08	0.50	13.98	0.21	17.58
3	6.12	1.28	16.87	0.70	20.65
4	8.16	2.27	19.60	1.41	23.53
5	10.20	3.40	22.23	2.27	26.28
6	12.24	4.63	24.77	3.25	28.92
7	14.29	5.94	27.24	4.34	31.47
8	16.33	7.32	29.66	5.50	33.95
9	18.37	8.76	32.02	6.74	36.37
10	20.41	10.24	34.34	8.03	38.73
11	22.45	11.77	36.62	9.39	41.04
12	24.49	13.34	38.87	10.79	43.30
13	26.53	14.95	41.08	12.23	45.52
14	28.57	16.58	43.26	13.72	47.70
15	30.61	18.25	45.42	15.24	49.85
16	32.65	19.95	47.54	16.81	51.96
17	34.69	21.67	49.64	18.40	54.03
18	36.73	23.42	51.71	20.03	56.07
19	38.78	25.20	53.76	21.69	58.09
20	40.82	27.00	55.79	23.39	60.07
21	42.86	28.82	57.79	25.11	62.02
22	44.90	30.67	59.77	26.86	63.95
23	46.94	32.53	61.73	28.64	65.84
24	48.98	34.42	63.66	30.45	67.71
N = 50					
0	0.00	0.00	7.11	0.00	10.05
1	2.00	0.05	10.65	0.01	13.94
2	4.00	0.49	13.71	0.21	17.25
3	6.00	1.25	16.55	0.69	20.27
4	8.00	2.22	19.23	1.38	23.11
5	10.00	3.33	21.81	2.22	25.80
6	12.00	4.53	24.31	3.19	28.40
7	14.00	5.82	26.74	4.25	30.91
8	16.00	7.17	29.11	5.39	33.35
9	18.00	8.58	31.44	6.60	35.73
10	20.00	10.03	33.72	7.86	38.05
11	22.00	11.53	35.96	9.19	40.32
12	24.00	13.06	38.17	10.56	42.55
13	26.00	14.63	40.34	11.97	44.74
14	28.00	16.23	42.49	13.42	46.89
15	30.00	17.86	44.61	14.91	49.00
16	32.00	19.52	46.70	16.44	51.08
17	34.00	21.21	48.77	18.00	53.12
18	36.00	22.92	50.81	19.59	55.14
19	38.00	24.65	52.83	21.21	57.13
20	40.00	26.41	54.82	22.87	59.08
21	42.00	28.19	56.79	24.55	61.01
22	44.00	29.99	58.75	26.26	62.91
23	46.00	31.81	60.68	27.99	64.78
24	48.00	33.66	62.58	29.76	66.63
25	50.00	35.53	64.47	31.55	68.45
N = 51					
0	0.00	0.00	6.98	0.00	9.87
1	1.96	0.05	10.45	0.01	13.68
2	3.92	0.48	13.46	0.20	16.94
3	5.88	1.23	16.24	0.67	19.90
4	7.84	2.18	18.88	1.35	22.69
5	9.80	3.26	21.41	2.18	25.35
6	11.76	4.44	23.87	3.12	27.90
7	13.73	5.70	26.26	4.16	30.37
8	15.69	7.02	28.59	5.28	32.77
9	17.65	8.40	30.87	6.46	35.11
10	19.61	9.82	33.12	7.70	37.39
11	21.57	11.29	35.32	8.99	39.63
12	23.53	12.79	37.49	10.33	41.82
13	25.49	14.33	39.63	11.72	43.98
14	27.45	15.89	41.74	13.14	46.10
15	29.41	17.49	43.83	14.59	48.18
16	31.37	19.11	45.89	16.09	50.23
17	33.33	20.76	47.92	17.61	52.25
18	35.29	22.43	49.93	19.17	54.23
19	37.25	24.13	51.92	20.75	56.19
20	39.22	25.84	53.89	22.37	58.13
21	41.18	27.58	55.83	24.01	60.03
22	43.14	29.35	57.75	25.68	61.91
23	45.10	31.13	59.66	27.37	63.76
24	47.06	32.93	61.54	29.10	65.58
25	49.02	34.75	63.40	30.84	67.38
N = 52					
0	0.00	0.00	6.85	0.00	9.69
1	1.92	0.05	10.26	0.01	13.44
2	3.85	0.47	13.21	0.20	16.63
3	5.77	1.21	15.95	0.66	19.55
4	7.69	2.14	18.54	1.32	22.29
5	9.62	3.20	21.03	2.13	24.90
6	11.54	4.35	23.44	3.06	27.41
7	13.46	5.59	25.79	4.08	29.84
8	15.38	6.88	28.08	5.17	32.20
9	17.31	8.23	30.33	6.33	34.51
10	19.23	9.63	32.53	7.54	36.76
11	21.15	11.06	34.70	8.81	38.96
12	23.08	12.53	36.84	10.12	41.12
13	25.00	14.03	38.95	11.47	43.24
14	26.92	15.57	41.02	12.86	45.33
15	28.85	17.13	43.08	14.29	47.38
16	30.77	18.72	45.10	15.75	49.40
17	32.69	20.33	47.11	17.24	51.39
18	34.62	21.97	49.09	18.76	53.36
19	36.54	23.62	51.04	20.31	55.29
20	38.46	25.30	52.98	21.89	57.20
21	40.38	27.01	54.90	23.49	59.08
22	42.31	28.73	56.80	25.12	60.93
23	44.23	30.47	58.67	26.78	62.76
24	46.15	32.23	60.53	28.46	64.57
25	48.08	34.01	62.37	30.17	66.35
26	50.00	35.81	64.19	31.90	68.10
N = 53					
0	0.00	0.00	6.72	0.00	9.51
1	1.89	0.05	10.07	0.01	13.20
2	3.77	0.46	12.98	0.20	16.34
3	5.66	1.18	15.66	0.65	19.21
4	7.55	2.09	18.21	1.30	21.90
5	9.43	3.13	20.66	2.09	24.47
6	11.32	4.27	23.03	3.00	26.94
7	13.21	5.48	25.34	4.00	29.33
8	15.09	6.75	27.59	5.07	31.66
9	16.98	8.07	29.80	6.20	33.93
10	18.87	9.44	31.97	7.39	36.14
11	20.75	10.84	34.11	8.63	38.31
12	22.64	12.28	36.21	9.92	40.44
13	24.53	13.76	38.28	11.24	42.53
14	26.42	15.26	40.33	12.60	44.59
15	28.30	16.79	42.35	14.00	46.61
16	30.19	18.34	44.34	15.43	48.61
17	32.08	19.92	46.32	16.89	50.57
18	33.96	21.52	48.27	18.37	52.51
19	35.85	23.14	50.20	19.89	54.41
20	37.74	24.79	52.11	21.43	56.30
21	39.62	26.45	54.00	23.00	58.15
22	41.51	28.14	55.87	24.59	59.99
23	43.40	29.84	57.72	26.21	61.79
24	45.28	31.56	59.55	27.86	63.58
25	47.17	33.30	61.36	29.52	65.34
26	49.06	35.06	63.16	31.21	67.07
N = 54					
0	0.00	0.00	6.60	0.00	9.35
1	1.85	0.05	9.89	0.01	12.97
2	3.70	0.45	12.75	0.19	16.06
3	5.56	1.16	15.39	0.64	18.88
4	7.41	2.06	17.89	1.27	21.53
5	9.26	3.08	20.30	2.05	24.06
6	11.11	4.19	22.63	2.94	26.49
7	12.96	5.37	24.90	3.92	28.84
8	14.81	6.62	27.12	4.97	31.13
9	16.67	7.92	29.29	6.08	33.36
10	18.52	9.25	31.43	7.25	35.55
11	20.37	10.63	33.53	8.46	37.69
12	22.22	12.04	35.60	9.72	39.78
13	24.07	13.49	37.64	11.02	41.85
14	25.93	14.96	39.65	12.35	43.87
15	27.78	16.46	41.64	13.72	45.87
16	29.63	17.98	43.61	15.12	47.83
17	31.48	19.52	45.55	16.55	49.77
18	33.33	21.09	47.47	18.00	51.68
19	35.19	22.68	49.38	19.48	53.56
20	37.04	24.29	51.26	20.99	55.42
21	38.89	25.92	53.12	22.53	57.26
22	40.74	27.57	54.97	24.09	59.07
23	42.59	29.23	56.79	25.67	60.85
24	44.44	30.92	58.60	27.27	62.62
25	46.30	32.62	60.39	28.90	64.36
26	48.15	34.34	62.16	30.55	66.08
27	50.00	36.08	63.92	32.23	67.77
N = 55					
0	0.00	0.00	6.49	0.00	9.18
1	1.82	0.05	9.72	0.01	12.75
2	3.64	0.44	12.53	0.19	15.79
3	5.45	1.14	15.12	0.62	18.56
4	7.27	2.02	17.59	1.25	21.17
5	9.09	3.02	19.95	2.01	23.66
6	10.91	4.11	22.25	2.89	26.05
7	12.73	5.27	24.48	3.85	28.37
8	14.55	6.50	26.66	4.88	30.62
9	16.36	7.77	28.80	5.97	32.82
10	18.18	9.08	30.90	7.11	34.97
11	20.00	10.43	32.97	8.30	37.08
12	21.82	11.81	35.01	9.53	39.15
13	23.64	13.23	37.02	10.81	41.18
14	25.45	14.67	39.00	12.11	43.18
15	27.27	16.14	40.96	13.45	45.15
16	29.09	17.63	42.90	14.82	47.08
17	30.91	19.14	44.81	16.22	49.00
18	32.73	20.68	46.71	17.64	50.88
19	34.55	22.24	48.58	19.10	52.74
20	36.36	23.81	50.44	20.57	54.57
21	38.18	25.41	52.27	22.07	56.39
22	40.00	27.02	54.09	23.60	58.17
23	41.82	28.65	55.89	25.15	59.94
24	43.64	30.30	57.68	26.72	61.68
25	45.45	31.97	59.45	28.31	63.40
26	47.27	33.65	61.20	29.92	65.10
27	49.09	35.35	62.93	31.56	66.78
N = 56					
0	0.00	0.00	6.38	0.00	9.03
1	1.79	0.05	9.55	0.01	12.53
2	3.57	0.44	12.31	0.19	15.52
3	5.36	1.12	14.87	0.61	18.25
4	7.14	1.98	17.29	1.23	20.82
5	8.93	2.96	19.62	1.98	23.27
6	10.71	4.03	21.88	2.83	25.63
7	12.50	5.18	24.07	3.77	27.91
8	14.29	6.38	26.22	4.79	30.13
9	16.07	7.62	28.33	5.86	32.30
10	17.86	8.91	30.40	6.98	34.42
11	19.64	10.23	32.43	8.14	36.49
12	21.43	11.59	34.44	9.35	38.53
13	23.21	12.98	36.42	10.60	40.53
14	25.00	14.39	38.37	11.88	42.50
15	26.79	15.83	40.30	13.19	44.45
16	28.57	17.30	42.21	14.53	46.36
17	30.36	18.78	44.10	15.90	48.24
18	32.14	20.29	45.96	17.30	50.10
19	33.93	21.81	47.81	18.72	51.94
20	35.71	23.36	49.64	20.17	53.75
21	37.50	24.92	51.45	21.64	55.54
22	39.29	26.50	53.25	23.13	57.31
23	41.07	28.10	55.02	24.65	59.05
24	42.86	29.71	56.78	26.18	60.77
25	44.64	31.34	58.53	27.74	62.48
26	46.43	32.99	60.26	29.32	64.16
27	48.21	34.66	61.97	30.92	65.82
28	50.00	36.34	63.66	32.54	67.46

Confidence intervals (%) for the binomial distribution (N = 57-64)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p ($\frac{r}{N} \times 100$)	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 57					
0	0.00	0.00	6.27	0.00	8.88
1	1.75	0.04	9.39	0.01	12.32
2	3.51	0.43	12.11	0.18	15.27
3	5.26	1.10	14.62	0.60	17.96
4	7.02	1.95	17.00	1.20	20.48
5	8.77	2.91	19.30	1.94	22.90
6	10.53	3.96	21.52	2.78	25.22
7	12.28	5.08	23.68	3.71	27.47
8	14.04	6.26	25.79	4.70	29.65
9	15.79	7.48	27.87	5.75	31.79
10	17.54	8.75	29.91	6.85	33.88
11	19.30	10.05	31.91	7.99	35.92
12	21.05	11.38	33.89	9.18	37.93
13	22.81	12.74	35.84	10.40	39.91
14	24.56	14.13	37.76	11.66	41.85
15	26.32	15.54	39.66	12.94	43.77
16	28.07	16.97	41.54	14.26	45.65
17	29.82	18.43	43.40	15.60	47.51
18	31.58	19.90	45.24	16.97	49.35
19	33.33	21.40	47.06	18.36	51.16
20	35.09	22.91	48.87	19.78	52.95
21	36.84	24.45	50.66	21.22	54.72
22	38.60	26.00	52.43	22.68	56.46
23	40.35	27.56	54.18	24.17	58.19
24	42.11	29.14	55.92	25.67	59.89
25	43.86	30.74	57.64	27.20	61.57
26	45.61	32.36	59.34	28.74	63.24
27	47.37	33.98	61.03	30.31	64.88
28	49.12	35.63	62.71	31.89	66.51
N = 58					
0	0.00	0.00	6.16	0.00	8.73
1	1.72	0.04	9.24	0.01	12.12
2	3.45	0.42	11.91	0.18	15.02
3	5.17	1.08	14.38	0.59	17.67
4	6.90	1.91	16.73	1.18	20.16
5	8.62	2.86	18.98	1.91	22.53
6	10.34	3.89	21.17	2.73	24.82
7	12.07	4.99	23.30	3.64	27.03
8	13.79	6.15	25.38	4.61	29.19
9	15.52	7.35	27.42	5.64	31.29
10	17.24	8.59	29.43	6.72	33.35
11	18.97	9.87	31.41	7.85	35.37
12	20.69	11.17	33.35	9.01	37.35
13	22.41	12.51	35.27	10.21	39.30
14	24.14	13.87	37.17	11.44	41.22
15	25.86	15.26	39.04	12.70	43.11
16	27.59	16.66	40.90	13.99	44.97
17	29.31	18.09	42.73	15.31	46.80
18	31.03	19.54	44.54	16.65	48.62
19	32.76	21.01	46.34	18.02	50.41
20	34.48	22.49	48.12	19.41	52.17
21	36.21	23.99	49.88	20.82	53.92
22	37.93	25.51	51.63	22.25	55.64
23	39.66	27.05	53.36	23.70	57.35
24	41.38	28.60	55.07	25.18	59.03
25	43.10	30.16	56.77	26.67	60.70
26	44.83	31.74	58.46	28.18	62.34
27	46.55	33.34	60.13	29.72	63.97
28	48.28	34.95	61.78	31.27	65.57
29	50.00	36.58	63.42	32.84	67.16
N = 59					
0	0.00	0.00	6.06	0.00	8.59
1	1.69	0.04	9.09	0.01	11.93
2	3.39	0.41	11.71	0.18	14.78
3	5.08	1.06	14.15	0.58	17.39
4	6.78	1.88	16.46	1.16	19.84
5	8.47	2.81	18.68	1.87	22.18
6	10.17	3.82	20.83	2.69	24.43
7	11.86	4.91	22.93	3.58	26.62
8	13.56	6.04	24.98	4.53	28.74
9	15.25	7.22	26.99	5.54	30.82
10	16.95	8.44	28.97	6.60	32.84
11	18.64	9.69	30.91	7.71	34.83
12	20.34	10.98	32.83	8.85	36.79
13	22.03	12.29	34.73	10.03	38.71
14	23.73	13.62	36.59	11.24	40.60
15	25.42	14.98	38.44	12.47	42.47
16	27.12	16.36	40.27	13.74	44.30
17	28.81	17.76	42.08	15.03	46.12
18	30.51	19.19	43.87	16.35	47.91
19	32.20	20.62	45.64	17.69	49.67
20	33.90	22.08	47.39	19.05	51.42
21	35.59	23.55	49.13	20.43	53.14
22	37.29	25.04	50.85	21.84	54.85
23	38.98	26.55	52.56	23.26	56.53
24	40.68	28.07	54.25	24.70	58.19
25	42.37	29.61	55.93	26.17	59.84
26	44.07	31.16	57.60	27.65	61.47
27	45.76	32.72	59.25	29.15	63.08
28	47.46	34.30	60.88	30.67	64.67
29	49.15	35.89	62.50	32.20	66.24
N = 60					
0	0.00	0.00	5.96	0.00	8.45
1	1.67	0.04	8.94	0.01	11.74
2	3.33	0.41	11.53	0.17	14.55
3	5.00	1.04	13.92	0.57	17.12
4	6.67	1.85	16.20	1.14	19.53
5	8.33	2.76	18.39	1.84	21.84
6	10.00	3.76	20.51	2.64	24.06
7	11.67	4.82	22.57	3.51	26.21
8	13.33	5.94	24.59	4.45	28.31
9	15.00	7.10	26.57	5.45	30.35
10	16.67	8.29	28.52	6.49	32.35
11	18.33	9.52	30.44	7.57	34.31
12	20.00	10.78	32.33	8.69	36.24
13	21.67	12.07	34.20	9.85	38.14
14	23.33	13.38	36.04	11.04	40.01
15	25.00	14.72	37.86	12.25	41.84
16	26.67	16.07	39.66	13.49	43.66
17	28.33	17.45	41.44	14.76	45.45
18	30.00	18.85	43.21	16.05	47.21
19	31.67	20.26	44.96	17.37	48.96
20	33.33	21.69	46.69	18.70	50.68
21	35.00	23.13	48.40	20.06	52.39
22	36.67	24.59	50.10	21.44	54.07
23	38.33	26.07	51.79	22.83	55.73
24	40.00	27.56	53.46	24.25	57.38
25	41.67	29.07	55.12	25.68	59.01
26	43.33	30.59	56.76	27.13	60.62
27	45.00	32.12	58.39	28.60	62.21
28	46.67	33.67	60.00	30.09	63.78
29	48.33	35.23	61.61	31.60	65.34
30	50.00	36.81	63.19	33.12	66.88
N = 61					
0	0.00	0.00	5.87	0.00	8.32
1	1.64	0.04	8.80	0.01	11.56
2	3.28	0.40	11.35	0.17	14.33
3	4.92	1.03	13.71	0.56	16.86
4	6.56	1.82	15.95	1.12	19.24
5	8.20	2.72	18.10	1.81	21.51
6	9.84	3.70	20.19	2.59	23.70
7	11.48	4.74	22.22	3.45	25.82
8	13.11	5.84	24.22	4.38	27.88
9	14.75	6.98	26.17	5.35	29.90
10	16.39	8.15	28.09	6.38	31.87
11	18.03	9.36	29.98	7.44	33.81
12	19.67	10.60	31.84	8.54	35.71
13	21.31	11.86	33.68	9.68	37.58
14	22.95	13.15	35.50	10.84	39.42
15	24.59	14.46	37.29	12.04	41.24
16	26.23	15.80	39.07	13.26	43.03
17	27.87	17.15	40.83	14.50	44.80
18	29.51	18.52	42.57	15.77	46.54
19	31.15	19.90	44.29	17.06	48.26
20	32.79	21.31	46.00	18.37	49.97
21	34.43	22.73	47.69	19.70	51.65
22	36.07	24.16	49.37	21.05	53.31
23	37.70	25.61	51.04	22.42	54.96
24	39.34	27.07	52.69	23.81	56.59
25	40.98	28.55	54.32	25.21	58.20
26	42.62	30.04	55.94	26.64	59.79
27	44.26	31.55	57.55	28.08	61.36
28	45.90	33.06	59.15	29.54	62.92
29	47.54	34.60	60.73	31.01	64.46
30	49.18	36.14	62.30	32.50	65.99
N = 62					
0	0.00	0.00	5.78	0.00	8.19
1	1.61	0.04	8.66	0.01	11.38
2	3.23	0.39	11.17	0.17	14.11
3	4.84	1.01	13.50	0.55	16.60
4	6.45	1.79	15.70	1.11	18.95
5	8.06	2.67	17.83	1.78	21.19
6	9.68	3.63	19.88	2.55	23.35
7	11.29	4.66	21.89	3.40	25.44
8	12.90	5.74	23.85	4.30	27.47
9	14.52	6.86	25.78	5.26	29.46
10	16.13	8.02	27.67	6.27	31.41
11	17.74	9.20	29.53	7.32	33.32
12	19.35	10.42	31.37	8.40	35.20
13	20.97	11.66	33.18	9.51	37.04
14	22.58	12.93	34.97	10.66	38.86
15	24.19	14.22	36.74	11.83	40.65
16	25.81	15.53	38.50	13.03	42.42
17	27.42	16.85	40.23	14.25	44.16
18	29.03	18.20	41.95	15.49	45.89
19	30.65	19.56	43.65	16.76	47.59
20	32.26	20.94	45.34	18.05	49.27
21	33.87	22.33	47.01	19.35	50.93
22	35.48	23.74	48.66	20.68	52.58
23	37.10	25.16	50.31	22.02	54.21
24	38.71	26.60	51.93	23.38	55.81
25	40.32	28.05	53.55	24.76	57.41
26	41.94	29.51	55.15	26.16	58.98
27	43.55	30.99	56.74	27.57	60.54
28	45.16	32.48	58.32	29.00	62.08
29	46.77	33.98	59.88	30.45	63.61
30	48.39	35.50	61.44	31.91	65.12
31	50.00	37.02	62.98	33.39	66.61
N = 63					
0	0.00	0.00	5.69	0.00	8.07
1	1.59	0.04	8.53	0.01	11.21
2	3.17	0.39	11.00	0.17	13.90
3	4.76	0.99	13.29	0.54	16.36
4	6.35	1.76	15.47	1.09	18.67
5	7.94	2.63	17.56	1.75	20.88
6	9.52	3.58	19.59	2.51	23.00
7	11.11	4.59	21.56	3.34	25.07
8	12.70	5.65	23.50	4.23	27.08
9	14.29	6.75	25.39	5.18	29.04
10	15.87	7.88	27.26	6.17	30.96
11	17.46	9.05	29.10	7.19	32.84
12	19.05	10.25	30.91	8.26	34.70
13	20.63	11.47	32.70	9.35	36.52
14	22.22	12.72	34.46	10.48	38.31
15	23.81	13.98	36.21	11.63	40.08
16	25.40	15.27	37.94	12.81	41.83
17	26.98	16.57	39.65	14.01	43.55
18	28.57	17.89	41.35	15.23	45.25
19	30.16	19.23	43.02	16.47	46.93
20	31.75	20.58	44.69	17.74	48.59
21	33.33	21.95	46.34	19.02	50.24
22	34.92	23.34	47.97	20.32	51.86
23	36.51	24.73	49.60	21.64	53.47
24	38.10	26.15	51.20	22.97	55.06
25	39.68	27.57	52.80	24.33	56.64
26	41.27	29.01	54.38	25.70	58.19
27	42.86	30.46	55.95	27.08	59.74
28	44.44	31.92	57.51	28.49	61.26
29	46.03	33.39	59.06	29.90	62.77
30	47.62	34.88	60.59	31.34	64.27
31	49.21	36.38	62.11	32.79	65.75
N = 64					
0	0.00	0.00	5.60	0.00	7.95
1	1.56	0.04	8.40	0.01	11.04
2	3.13	0.38	10.84	0.16	13.69
3	4.69	0.98	13.09	0.53	16.12
4	6.25	1.73	15.24	1.07	18.40
5	7.81	2.59	17.30	1.72	20.57
6	9.38	3.52	19.30	2.47	22.67
7	10.94	4.51	21.25	3.29	24.71
8	12.50	5.55	23.15	4.16	26.69
9	14.06	6.64	25.02	5.09	28.62
10	15.63	7.76	26.86	6.06	30.52
11	17.19	8.90	28.68	7.08	32.38
12	18.75	10.08	30.46	8.12	34.21
13	20.31	11.28	32.23	9.20	36.01
14	21.88	12.51	33.97	10.30	37.78
15	23.44	13.75	35.69	11.44	39.53
16	25.00	15.02	37.40	12.59	41.25
17	26.56	16.30	39.09	13.77	42.95
18	28.13	17.60	40.76	14.97	44.63
19	29.69	18.91	42.42	16.19	46.29
20	31.25	20.24	44.06	17.43	47.94
21	32.81	21.59	45.69	18.69	49.56
22	34.38	22.95	47.30	19.97	51.17
23	35.94	24.32	48.90	21.27	52.76
24	37.50	25.70	50.49	22.58	54.33
25	39.06	27.10	52.07	23.91	55.89
26	40.63	28.51	53.63	25.25	57.43
27	42.19	29.94	55.18	26.61	58.95
28	43.75	31.37	56.72	27.99	60.46
29	45.31	32.82	58.25	29.38	61.96

Confidence intervals (%) for the binomial distribution (N = 65-71)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 64 (continued)					
30	46.88	34.28	59.77	30.79	63.44
31	48.44	35.75	61.27	32.21	64.90
32	50.00	37.23	62.77	33.64	66.36
N = 65					
0	0.00	0.00	5.52	0.00	7.83
1	1.54	0.04	8.28	0.01	10.88
2	3.08	0.37	10.68	0.16	13.49
3	4.62	0.96	12.90	0.53	15.88
4	6.15	1.70	15.01	1.05	18.13
5	7.69	2.54	17.05	1.70	20.28
6	9.23	3.46	19.02	2.43	22.35
7	10.77	4.44	20.94	3.23	24.36
8	12.31	5.47	22.82	4.10	26.31
9	13.85	6.53	24.66	5.01	28.22
10	15.38	7.63	26.48	5.97	30.09
11	16.92	8.76	28.27	6.96	31.93
12	18.46	9.92	30.03	7.99	33.73
13	20.00	11.10	31.77	9.05	35.51
14	21.54	12.31	33.49	10.14	37.26
15	23.08	13.53	35.19	11.25	38.99
16	24.62	14.77	36.87	12.38	40.69
17	26.15	16.03	38.54	13.54	42.37
18	27.69	17.31	40.19	14.72	44.03
19	29.23	18.60	41.83	15.93	45.67
20	30.77	19.91	43.45	17.14	47.29
21	32.31	21.23	45.05	18.38	48.90
22	33.85	22.57	46.65	19.64	50.49
23	35.38	23.92	48.23	20.91	52.06
24	36.92	25.28	49.80	22.20	53.61
25	38.46	26.65	51.36	23.50	55.15
26	40.00	28.04	52.90	24.82	56.68
27	41.54	29.44	54.44	26.16	58.19
28	43.08	30.85	55.96	27.51	59.68
29	44.62	32.27	57.47	28.88	61.16
30	46.15	33.70	58.97	30.26	62.63
31	47.69	35.15	60.46	31.65	64.08
32	49.23	36.60	61.93	33.06	65.52
N = 66					
0	0.00	0.00	5.44	0.00	7.71
1	1.52	0.04	8.16	0.01	10.73
2	3.03	0.37	10.52	0.16	13.30
3	4.55	0.95	12.71	0.52	15.66
4	6.06	1.68	14.80	1.04	17.88
5	7.58	2.51	16.80	1.67	19.99
6	9.09	3.41	18.74	2.39	22.04
7	10.61	4.37	20.64	3.18	24.02
8	12.12	5.38	22.49	4.03	25.95
9	13.64	6.43	24.31	4.93	27.83
10	15.15	7.51	26.10	5.87	29.68
11	16.67	8.62	27.87	6.85	31.49
12	18.18	9.76	29.61	7.86	33.27
13	19.70	10.93	31.32	8.90	35.03
14	21.21	12.11	33.02	9.97	36.75
15	22.73	13.31	34.70	11.07	38.46
16	24.24	14.54	36.36	12.18	40.14
17	25.76	15.78	38.01	13.32	41.80
18	27.27	17.03	39.64	14.48	43.44
19	28.79	18.30	41.25	15.67	45.06
20	30.30	19.59	42.85	16.86	46.67
21	31.82	20.89	44.44	18.08	48.25
22	33.33	22.20	46.01	19.31	49.82
23	34.85	23.53	47.58	20.56	51.38
24	36.36	24.87	49.13	21.83	52.92
25	37.88	26.22	50.66	23.11	54.44
26	39.39	27.58	52.19	24.41	55.95
27	40.91	28.95	53.71	25.72	57.44
28	42.42	30.34	55.21	27.05	58.92
29	43.94	31.74	56.70	28.39	60.39
30	45.45	33.14	58.19	29.74	61.84
31	46.97	34.56	59.66	31.11	63.28
32	48.48	35.99	61.12	32.49	64.70
33	50.00	37.43	62.57	33.89	66.11
N = 67					
0	0.00	0.00	5.36	0.00	7.60
1	1.49	0.04	8.04	0.01	10.57
2	2.99	0.36	10.37	0.16	13.11
3	4.48	0.93	12.53	0.51	15.44
4	5.97	1.65	14.59	1.02	17.63
5	7.46	2.47	16.56	1.65	19.72
6	8.96	3.36	18.48	2.36	21.73
7	10.45	4.30	20.35	3.14	23.69
8	11.94	5.30	22.18	3.97	25.59
9	13.43	6.33	23.97	4.86	27.45
10	14.93	7.40	25.74	5.78	29.28

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 67 (continued)					
11	16.42	8.49	27.48	6.74	31.07
12	17.91	9.61	29.20	7.74	32.82
13	19.40	10.76	30.89	8.76	34.56
14	20.90	11.92	32.57	9.82	36.26
15	22.39	13.11	34.22	10.89	37.95
16	23.88	14.31	35.86	11.99	39.61
17	25.37	15.53	37.49	13.11	41.25
18	26.87	16.76	39.10	14.25	42.87
19	28.36	18.01	40.69	15.41	44.47
20	29.85	19.28	42.27	16.59	46.06
21	31.34	20.56	43.84	17.79	47.63
22	32.84	21.85	45.40	19.00	49.18
23	34.33	23.15	46.94	20.23	50.72
24	35.82	24.47	48.47	21.47	52.24
25	37.31	25.80	49.99	22.73	53.74
26	38.81	27.14	51.50	24.01	55.24
27	40.30	28.49	53.00	25.30	56.72
28	41.79	29.85	54.48	26.60	58.18
29	43.28	31.22	55.96	27.92	59.63
30	44.78	32.60	57.42	29.25	61.07
31	46.27	34.00	58.88	30.59	62.49
32	47.76	35.40	60.33	31.95	63.90
33	49.25	36.82	61.76	33.32	65.30
N = 68					
0	0.00	0.00	5.28	0.00	7.50
1	1.47	0.04	7.92	0.01	10.42
2	2.94	0.36	10.22	0.15	12.93
3	4.41	0.92	12.36	0.50	15.22
4	5.88	1.63	14.38	1.01	17.38
5	7.35	2.43	16.33	1.62	19.45
6	8.82	3.31	18.22	2.32	21.44
7	10.29	4.24	20.07	3.09	23.37
8	11.76	5.22	21.87	3.91	25.25
9	13.24	6.23	23.64	4.78	27.08
10	14.71	7.28	25.39	5.69	28.88
11	16.18	8.36	27.10	6.64	30.65
12	17.65	9.47	28.80	7.62	32.39
13	19.12	10.59	30.47	8.63	34.10
14	20.59	11.74	32.12	9.66	35.78
15	22.06	12.90	33.76	10.72	37.45
16	23.53	14.09	35.38	11.80	39.09
17	25.00	15.29	36.98	12.91	40.71
18	26.47	16.50	38.57	14.03	42.31
19	27.94	17.73	40.15	15.17	43.90
20	29.41	18.98	41.71	16.33	45.46
21	30.88	20.24	43.26	17.51	47.02
22	32.35	21.51	44.79	18.70	48.55
23	33.82	22.79	46.32	19.91	50.07
24	35.29	24.08	47.83	21.13	51.58
25	36.76	25.39	49.33	22.37	53.07
26	38.24	26.71	50.82	23.62	54.54
27	39.71	28.03	52.30	24.89	56.00
28	41.18	29.37	53.77	26.17	57.45
29	42.65	30.72	55.23	27.46	58.89
30	44.12	32.08	56.68	28.77	60.31
31	45.59	33.45	58.12	30.09	61.72
32	47.06	34.83	59.55	31.42	63.12
33	48.53	36.22	60.97	32.77	64.50
34	50.00	37.62	62.38	34.12	65.88
N = 69					
0	0.00	0.00	5.21	0.00	7.39
1	1.45	0.04	7.81	0.01	10.28
2	2.90	0.35	10.08	0.15	12.75
3	4.35	0.91	12.18	0.50	15.02
4	5.80	1.60	14.18	0.99	17.15
5	7.25	2.39	16.11	1.60	19.18
6	8.70	3.26	17.97	2.29	21.15
7	10.14	4.18	19.79	3.04	23.05
8	11.59	5.14	21.57	3.85	24.91
9	13.04	6.14	23.32	4.71	26.72
10	14.49	7.17	25.04	5.61	28.50
11	15.94	8.24	26.74	6.54	30.25
12	17.39	9.32	28.41	7.50	31.96
13	18.84	10.43	30.06	8.50	33.65
14	20.29	11.56	31.69	9.51	35.32
15	21.74	12.71	33.31	10.56	36.96
16	23.19	13.87	34.91	11.62	38.58
17	24.64	15.05	36.49	12.71	40.19
18	26.09	16.25	38.06	13.81	41.77
19	27.54	17.46	39.62	14.93	43.34
20	28.99	18.69	41.16	16.07	44.89
21	30.43	19.92	42.69	17.23	46.42
22	31.88	21.17	44.21	18.40	47.94
23	33.33	22.44	45.71	19.59	49.44

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 69 (continued)					
24	34.78	23.71	47.21	20.80	50.93
25	36.23	24.99	48.69	22.01	52.40
26	37.68	26.29	50.17	23.24	53.86
27	39.13	27.60	51.63	24.49	55.31
28	40.58	28.91	53.08	25.75	56.75
29	42.03	30.24	54.52	27.02	58.17
30	43.48	31.58	55.96	28.30	59.58
31	44.93	32.92	57.38	29.60	60.97
32	46.38	34.28	58.80	30.91	62.36
33	47.83	35.65	60.20	32.23	63.73
34	49.28	37.02	61.59	33.57	65.09
N = 70					
0	0.00	0.00	5.13	0.00	7.29
1	1.43	0.04	7.70	0.01	10.14
2	2.86	0.35	9.94	0.15	12.58
3	4.29	0.89	12.02	0.49	14.81
4	5.71	1.58	13.99	0.98	16.92
5	7.14	2.36	15.89	1.57	18.93
6	8.57	3.21	17.73	2.25	20.87
7	10.00	4.12	19.52	3.00	22.75
8	11.43	5.07	21.28	3.80	24.58
9	12.86	6.05	23.01	4.64	26.37
10	14.29	7.07	24.71	5.52	28.13
11	15.71	8.11	26.38	6.44	29.85
12	17.14	9.18	28.03	7.39	31.55
13	18.57	10.28	29.66	8.37	33.22
14	20.00	11.39	31.27	9.37	34.86
15	21.43	12.52	32.87	10.40	36.49
16	22.86	13.67	34.45	11.45	38.09
17	24.29	14.83	36.01	12.51	39.67
18	25.71	16.01	37.56	13.60	41.24
19	27.14	17.20	39.10	14.71	42.79
20	28.57	18.40	40.62	15.83	44.32
21	30.00	19.62	42.13	16.97	45.84
22	31.43	20.85	43.63	18.12	47.34
23	32.86	22.09	45.12	19.29	48.82
24	34.29	23.35	46.60	20.47	50.30
25	35.71	24.61	48.07	21.67	51.76
26	37.14	25.89	49.52	22.88	53.20
27	38.57	27.17	50.97	24.11	54.63
28	40.00	28.47	52.41	25.34	56.05
29	41.43	29.77	53.83	26.59	57.46
30	42.86	31.09	55.25	27.85	58.86
31	44.29	32.41	56.66	29.13	60.24
32	45.71	33.74	58.06	30.42	61.61
33	47.14	35.09	59.45	31.72	62.97
34	48.57	36.44	60.83	33.03	64.32
35	50.00	37.80	62.20	34.35	65.65
N = 71					
0	0.00	0.00	5.06	0.00	7.19
1	1.41	0.04	7.60	0.01	10.00
2	2.82	0.34	9.81	0.15	12.41
3	4.23	0.88	11.86	0.48	14.62
4	5.63	1.56	13.80	0.96	16.69
5	7.04	2.33	15.67	1.55	18.68
6	8.45	3.16	17.49	2.22	20.59
7	9.86	4.06	19.26	2.95	22.45
8	11.27	4.99	21.00	3.74	24.26
9	12.68	5.96	22.70	4.57	26.03
10	14.08	6.97	24.38	5.44	27.77
11	15.49	8.00	26.03	6.35	29.47
12	16.90	9.05	27.66	7.28	31.14
13	18.31	10.13	29.27	8.24	32.79
14	19.72	11.22	30.87	9.23	34.42
15	21.13	12.33	32.44	10.24	36.02
16	22.54	13.46	34.00	11.27	37.61
17	23.94	14.61	35.54	12.33	39.17
18	25.35	15.77	37.08	13.40	40.72
19	26.76	16.94	38.59	14.48	42.25
20	28.17	18.13	40.10	15.59	43.77
21	29.58	19.33	41.59	16.71	45.27
22	30.99	20.54	43.08	17.84	46.75
23	32.39	21.76	44.55	18.99	48.22
24	33.80	23.00	46.01	20.16	49.68
25	35.21	24.24	47.46	21.34	51.12
26	36.62	25.50	48.90	22.53	52.56
27	38.03	26.76	50.33	23.73	53.97
28	39.44	28.03	51.75	24.95	55.38
29	40.85	29.32	53.16	26.18	56.77
30	42.25	30.61	54.56	27.42	58.15
31	43.66	31.91	55.95	28.67	59.52
32	45.07	33.23	57.34	29.94	60.88
33	46.48	34.55	58.71	31.22	62.23
34	47.89	35.88	60.08	32.50	63.56

Confidence intervals (%) for the binomial distribution (N = 72-78)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 71 (continued)					
35	49.30	37.22	61.44	33.80	64.88
N = 72					
0	0.00	0.00	4.99	0.00	7.09
1	1.39	0.04	7.50	0.01	9.87
2	2.78	0.34	9.68	0.14	12.25
3	4.17	0.87	11.70	0.47	14.42
4	5.56	1.53	13.62	0.95	16.48
5	6.94	2.29	15.47	1.53	18.44
6	8.33	3.12	17.26	2.19	20.33
7	9.72	4.00	19.01	2.91	22.16
8	11.11	4.92	20.72	3.69	23.95
9	12.50	5.88	22.41	4.51	25.70
10	13.89	6.87	24.06	5.36	27.41
11	15.28	7.88	25.69	6.26	29.09
12	16.67	8.92	27.30	7.18	30.75
13	18.06	9.98	28.89	8.12	32.38
14	19.44	11.06	30.47	9.10	33.99
15	20.83	12.16	32.02	10.09	35.57
16	22.22	13.27	33.56	11.11	37.14
17	23.61	14.40	35.09	12.14	38.69
18	25.00	15.54	36.60	13.20	40.22
19	26.39	16.70	38.10	14.27	41.73
20	27.78	17.86	39.59	15.36	43.23
21	29.17	19.05	41.07	16.46	44.71
22	30.56	20.24	42.53	17.58	46.18
23	31.94	21.44	43.99	18.71	47.64
24	33.33	22.66	45.43	19.86	49.08
25	34.72	23.88	46.86	21.01	50.51
26	36.11	25.12	48.29	22.19	51.92
27	37.50	26.36	49.70	23.37	53.33
28	38.89	27.62	51.11	24.57	54.72
29	40.28	28.88	52.50	25.78	56.10
30	41.67	30.15	53.89	27.00	57.47
31	43.06	31.43	55.27	28.23	58.82
32	44.44	32.72	56.64	29.48	60.17
33	45.83	34.02	58.00	30.73	61.50
34	47.22	35.33	59.35	32.00	62.82
35	48.61	36.65	60.69	33.28	64.13
36	50.00	37.98	62.02	34.57	65.43
N = 73					
0	0.00	0.00	4.93	0.00	7.00
1	1.37	0.03	7.40	0.01	9.74
2	2.74	0.33	9.55	0.14	12.09
3	4.11	0.86	11.54	0.47	14.24
4	5.48	1.51	13.44	0.94	16.26
5	6.85	2.26	15.26	1.51	18.20
6	8.22	3.08	17.04	2.16	20.07
7	9.59	3.94	18.76	2.87	21.88
8	10.96	4.85	20.46	3.63	23.65
9	12.33	5.80	22.12	4.44	25.37
10	13.70	6.77	23.75	5.29	27.07
11	15.07	7.77	25.36	6.17	28.73
12	16.44	8.79	26.95	7.07	30.37
13	17.81	9.84	28.53	8.01	31.98
14	19.18	10.90	30.08	8.97	33.57
15	20.55	11.98	31.62	9.95	35.13
16	21.92	13.08	33.14	10.95	36.68
17	23.29	14.19	34.65	11.97	38.21
18	24.66	15.32	36.14	13.01	39.73
19	26.03	16.45	37.62	14.06	41.22
20	27.40	17.61	39.09	15.13	42.71
21	28.77	18.77	40.55	16.22	44.17
22	30.14	19.94	42.00	17.32	45.63
23	31.51	21.13	43.44	18.43	47.06
24	32.88	22.33	44.87	19.56	48.49
25	34.25	23.53	46.28	20.70	49.91
26	35.62	24.75	47.69	21.86	51.31
27	36.99	25.97	49.09	23.02	52.70
28	38.36	27.21	50.48	24.20	54.07
29	39.73	28.45	51.86	25.39	55.44
30	41.10	29.71	53.23	26.59	56.79
31	42.47	30.97	54.59	27.80	58.14
32	43.84	32.24	55.95	29.03	59.47
33	45.21	33.52	57.29	30.26	60.79
34	46.58	34.80	58.63	31.51	62.10
35	47.95	36.10	59.96	32.77	63.40
36	49.32	37.40	61.28	34.03	64.69
N = 74					
0	0.00	0.00	4.86	0.00	6.91
1	1.35	0.03	7.30	0.01	9.62
2	2.70	0.33	9.42	0.14	11.93
3	4.05	0.84	11.39	0.46	14.06
4	5.41	1.49	13.27	0.92	16.06
5	6.76	2.23	15.07	1.49	17.97
6	8.11	3.03	16.82	2.13	19.81
7	9.46	3.89	18.52	2.83	21.60
8	10.81	4.78	20.20	3.58	23.35
9	12.16	5.71	21.84	4.38	25.06
10	13.51	6.68	23.45	5.21	26.73
11	14.86	7.66	25.04	6.08	28.37
12	16.22	8.67	26.61	6.97	29.99
13	17.57	9.70	28.17	7.89	31.58
14	18.92	10.75	29.70	8.84	33.15
15	20.27	11.81	31.22	9.80	34.70
16	21.62	12.89	32.72	10.79	36.24
17	22.97	13.99	34.21	11.80	37.75
18	24.32	15.10	35.69	12.82	39.25
19	25.68	16.22	37.16	13.86	40.73
20	27.03	17.35	38.61	14.91	42.19
21	28.38	18.50	40.05	15.98	43.64
22	29.73	19.66	41.48	17.07	45.08
23	31.08	20.83	42.90	18.16	46.51
24	32.43	22.00	44.32	19.27	47.92
25	33.78	23.19	45.72	20.40	49.32
26	35.14	24.39	47.11	21.53	50.70
27	36.49	25.60	48.49	22.68	52.08
28	37.84	26.81	49.87	23.84	53.44
29	39.19	28.04	51.23	25.01	54.79
30	40.54	29.27	52.59	26.19	56.14
31	41.89	30.51	53.94	27.39	57.47
32	43.24	31.77	55.28	28.59	58.79
33	44.59	33.02	56.61	29.81	60.10
34	45.95	34.29	57.93	31.03	61.39
35	47.30	35.57	59.25	32.27	62.68
36	48.65	36.85	60.56	33.52	63.96
37	50.00	38.14	61.86	34.77	65.23
N = 75					
0	0.00	0.00	4.80	0.00	6.82
1	1.33	0.03	7.21	0.01	9.49
2	2.67	0.32	9.30	0.14	11.78
3	4.00	0.83	11.25	0.46	13.88
4	5.33	1.47	13.10	0.91	15.85
5	6.67	2.20	14.88	1.47	17.74
6	8.00	2.99	16.60	2.10	19.57
7	9.33	3.84	18.29	2.79	21.34
8	10.67	4.72	19.94	3.53	23.06
9	12.00	5.64	21.56	4.32	24.75
10	13.33	6.58	23.16	5.14	26.40
11	14.67	7.56	24.73	5.99	28.03
12	16.00	8.55	26.28	6.88	29.63
13	17.33	9.57	27.81	7.78	31.20
14	18.67	10.60	29.33	8.71	32.75
15	20.00	11.65	30.83	9.67	34.29
16	21.33	12.71	32.32	10.64	35.80
17	22.67	13.79	33.79	11.63	37.30
18	24.00	14.89	35.25	12.64	38.78
19	25.33	15.99	36.70	13.66	40.24
20	26.67	17.11	38.14	14.70	41.69
21	28.00	18.24	39.56	15.75	43.13
22	29.33	19.38	40.98	16.82	44.55
23	30.67	20.53	42.38	17.90	45.96
24	32.00	21.69	43.78	19.00	47.36
25	33.33	22.86	45.17	20.10	48.74
26	34.67	24.04	46.54	21.22	50.11
27	36.00	25.23	47.91	22.35	51.48
28	37.33	26.43	49.27	23.49	52.83
29	38.67	27.64	50.62	24.65	54.16
30	40.00	28.85	51.96	25.81	55.49
31	41.33	30.08	53.30	26.99	56.81
32	42.67	31.31	54.62	28.17	58.12
33	44.00	32.55	55.94	29.37	59.42
34	45.33	33.79	57.25	30.57	60.70
35	46.67	35.05	58.55	31.79	61.98
36	48.00	36.31	59.85	33.02	63.25
37	49.33	37.58	61.14	34.25	64.50
N = 76					
0	0.00	0.00	4.74	0.00	6.73
1	1.32	0.03	7.11	0.01	9.37
2	2.63	0.32	9.18	0.14	11.63
3	3.95	0.82	11.11	0.45	13.71
4	5.26	1.45	12.93	0.90	15.66
5	6.58	2.17	14.69	1.45	17.52
6	7.89	2.95	16.40	2.07	19.33
7	9.21	3.78	18.06	2.75	21.07
8	10.53	4.66	19.69	3.49	22.78
9	11.84	5.56	21.29	4.26	24.45
10	13.16	6.49	22.87	5.07	26.08
11	14.47	7.45	24.42	5.91	27.69
12	15.79	8.43	25.96	6.78	29.27
13	17.11	9.43	27.47	7.68	30.83
14	18.42	10.45	28.97	8.59	32.36
15	19.74	11.49	30.46	9.53	33.88
16	21.05	12.54	31.92	10.49	35.37
17	22.37	13.60	33.38	11.47	36.85
18	23.68	14.68	34.82	12.46	38.32
19	25.00	15.77	36.26	13.47	39.77
20	26.32	16.87	37.68	14.49	41.20
21	27.63	17.99	39.09	15.53	42.62
22	28.95	19.11	40.49	16.58	44.03
23	30.26	20.25	41.87	17.65	45.43
24	31.58	21.39	43.25	18.73	46.81
25	32.89	22.54	44.63	19.82	48.18
26	34.21	23.71	45.99	20.92	49.54
27	35.53	24.88	47.34	22.03	50.89
28	36.84	26.06	48.69	23.16	52.22
29	38.16	27.25	50.02	24.29	53.55
30	39.47	28.44	51.35	25.44	54.86
31	40.79	29.65	52.67	26.59	56.17
32	42.11	30.86	53.98	27.76	57.46
33	43.42	32.08	55.29	28.94	58.75
34	44.74	33.31	56.59	30.13	60.03
35	46.05	34.55	57.87	31.32	61.29
36	47.37	35.79	59.16	32.53	62.55
37	48.68	37.04	60.43	33.75	63.79
38	50.00	38.30	61.70	34.97	65.03
N = 77					
0	0.00	0.00	4.68	0.00	6.65
1	1.30	0.03	7.02	0.01	9.26
2	2.60	0.32	9.07	0.14	11.49
3	3.90	0.81	10.97	0.44	13.54
4	5.19	1.43	12.77	0.89	15.47
5	6.49	2.14	14.51	1.43	17.31
6	7.79	2.91	16.19	2.04	19.09
7	9.09	3.73	17.84	2.72	20.82
8	10.39	4.59	19.45	3.44	22.50
9	11.69	5.49	21.03	4.20	24.15
10	12.99	6.41	22.59	5.00	25.77
11	14.29	7.35	24.13	5.83	27.36
12	15.58	8.32	25.64	6.69	28.92
13	16.88	9.31	27.14	7.57	30.46
14	18.18	10.31	28.62	8.48	31.98
15	19.48	11.33	30.09	9.40	33.48
16	20.78	12.37	31.54	10.35	34.96
17	22.08	13.42	32.98	11.31	36.42
18	23.38	14.48	34.41	12.29	37.87
19	24.68	15.56	35.82	13.28	39.31
20	25.97	16.64	37.23	14.29	40.72
21	27.27	17.74	38.62	15.32	42.13
22	28.57	18.85	40.00	16.35	43.52
23	29.87	19.97	41.38	17.40	44.90
24	31.17	21.09	42.74	18.46	46.27
25	32.47	22.23	44.10	19.54	47.63
26	33.77	23.38	45.45	20.62	48.97
27	35.06	24.53	46.78	21.72	50.31
28	36.36	25.70	48.12	22.83	51.63
29	37.66	26.87	49.44	23.95	52.95
30	38.96	28.05	50.75	25.08	54.25
31	40.26	29.23	52.06	26.21	55.54
32	41.56	30.43	53.36	27.36	56.82
33	42.86	31.63	54.65	28.52	58.10
34	44.16	32.84	55.93	29.69	59.36
35	45.45	34.06	57.21	30.87	60.62
36	46.75	35.29	58.48	32.06	61.86
37	48.05	36.52	59.74	33.26	63.10
38	49.35	37.76	61.00	34.46	64.32
N = 78					
0	0.00	0.00	4.62	0.00	6.57
1	1.28	0.03	6.94	0.01	9.14
2	2.56	0.31	8.96	0.13	11.35
3	3.85	0.80	10.83	0.44	13.37
4	5.13	1.41	12.61	0.87	15.28
5	6.41	2.11	14.33	1.41	17.10
6	7.69	2.88	15.99	2.02	18.86
7	8.97	3.68	17.62	2.68	20.57
8	10.26	4.53	19.21	3.39	22.24
9	11.54	5.41	20.78	4.15	23.87
10	12.82	6.32	22.32	4.94	25.46
11	14.10	7.26	23.83	5.76	27.03
12	15.38	8.21	25.33	6.60	28.58
13	16.67	9.18	26.81	7.47	30.10
14	17.95	10.17	28.28	8.36	31.60
15	19.23	11.18	29.73	9.28	33.09

Confidence intervals (%) for the binomial distribution (N = 79-84)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 78 (continued)					
16	20.51	12.20	31.16	10.21	34.55
17	21.79	13.24	32.59	11.16	36.00
18	23.08	14.29	34.00	12.12	37.43
19	24.36	15.35	35.40	13.10	38.85
20	25.64	16.42	36.79	14.10	40.26
21	26.92	17.50	38.16	15.11	41.65
22	28.21	18.59	39.53	16.13	43.03
23	29.49	19.70	40.89	17.16	44.39
24	30.77	20.81	42.24	18.21	45.75
25	32.05	21.93	43.58	19.27	47.09
26	33.33	23.06	44.92	20.34	48.42
27	34.62	24.20	46.24	21.42	49.74
28	35.90	25.34	47.56	22.51	51.06
29	37.18	26.50	48.87	23.61	52.36
30	38.46	27.66	50.17	24.72	53.65
31	39.74	28.83	51.46	25.85	54.93
32	41.03	30.01	52.75	26.98	56.20
33	42.31	31.19	54.02	28.12	57.46
34	43.59	32.39	55.30	29.27	58.71
35	44.87	33.59	56.56	30.43	59.96
36	46.15	34.79	57.82	31.60	61.19
37	47.44	36.01	59.07	32.78	62.41
38	48.72	37.23	60.31	33.97	63.63
39	50.00	38.46	61.54	35.16	64.84
N = 79					
0	0.00	0.00	4.56	0.00	6.49
1	1.27	0.03	6.85	0.01	9.03
2	2.53	0.31	8.85	0.13	11.21
3	3.80	0.79	10.70	0.43	13.21
4	5.06	1.40	12.46	0.86	15.10
5	6.33	2.09	14.16	1.39	16.90
6	7.59	2.84	15.80	1.99	18.64
7	8.86	3.64	17.41	2.65	20.33
8	10.13	4.47	18.98	3.35	21.97
9	11.39	5.34	20.53	4.09	23.59
10	12.66	6.24	22.05	4.87	25.17
11	13.92	7.16	23.55	5.68	26.72
12	15.19	8.10	25.03	6.51	28.25
13	16.46	9.06	26.49	7.37	29.75
14	17.72	10.04	27.94	8.25	31.24
15	18.99	11.03	29.38	9.15	32.71
16	20.25	12.04	30.80	10.07	34.16
17	21.52	13.06	32.20	11.01	35.59
18	22.78	14.10	33.60	11.96	37.01
19	24.05	15.14	34.98	12.93	38.41
20	25.32	16.20	36.36	13.91	39.80
21	26.58	17.27	37.72	14.90	41.18
22	27.85	18.35	39.07	15.91	42.54
23	29.11	19.43	40.42	16.93	43.89
24	30.38	20.53	41.75	17.96	45.23
25	31.65	21.63	43.08	19.00	46.56
26	32.91	22.75	44.40	20.06	47.88
27	34.18	23.87	45.71	21.12	49.19
28	35.44	25.00	47.01	22.20	50.49
29	36.71	26.14	48.31	23.29	51.78
30	37.97	27.28	49.59	24.38	53.06
31	39.24	28.44	50.87	25.49	54.33
32	40.51	29.60	52.15	26.60	55.59
33	41.77	30.77	53.41	27.73	56.84
34	43.04	31.94	54.67	28.86	58.08
35	44.30	33.12	55.92	30.00	59.31
36	45.57	34.31	57.17	31.16	60.53
37	46.84	35.51	58.40	32.32	61.75
38	48.10	36.71	59.64	33.49	62.95
39	49.37	37.92	60.86	34.66	64.15
N = 80					
0	0.00	0.00	4.51	0.00	6.41
1	1.25	0.03	6.77	0.01	8.92
2	2.50	0.30	8.74	0.13	11.08
3	3.75	0.78	10.57	0.43	13.05
4	5.00	1.38	12.31	0.85	14.92
5	6.25	2.06	13.99	1.37	16.70
6	7.50	2.80	15.61	1.96	18.42
7	8.75	3.59	17.20	2.61	20.09
8	10.00	4.42	18.76	3.31	21.72
9	11.25	5.28	20.28	4.04	23.31
10	12.50	6.16	21.79	4.81	24.87
11	13.75	7.07	23.27	5.61	26.41
12	15.00	8.00	24.74	6.43	27.92
13	16.25	8.95	26.18	7.28	29.41
14	17.50	9.91	27.62	8.14	30.88
15	18.75	10.89	29.03	9.03	32.33
16	20.00	11.89	30.44	9.94	33.77
17	21.25	12.89	31.83	10.86	35.19

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 80 (continued)					
18	22.50	13.91	33.21	11.80	36.59
19	23.75	14.95	34.58	12.75	37.98
20	25.00	15.99	35.94	13.72	39.35
21	26.25	17.04	37.29	14.70	40.72
22	27.50	18.10	38.62	15.70	42.07
23	28.75	19.18	39.95	16.70	43.41
24	30.00	20.26	41.28	17.72	44.73
25	31.25	21.35	42.59	18.75	46.05
26	32.50	22.45	43.89	19.79	47.36
27	33.75	23.55	45.19	20.84	48.65
28	35.00	24.67	46.48	21.90	49.94
29	36.25	25.79	47.76	22.97	51.21
30	37.50	26.92	49.04	24.05	52.48
31	38.75	28.06	50.30	25.14	53.74
32	40.00	29.20	51.56	26.24	54.99
33	41.25	30.35	52.82	27.35	56.22
34	42.50	31.51	54.06	28.46	57.45
35	43.75	32.68	55.30	29.59	58.68
36	45.00	33.85	56.53	30.72	59.89
37	46.25	35.03	57.76	31.87	61.09
38	47.50	36.21	58.98	33.02	62.29
39	48.75	37.41	60.19	34.18	63.47
40	50.00	38.60	61.40	35.35	64.65
N = 81					
0	0.00	0.00	4.45	0.00	6.33
1	1.23	0.03	6.69	0.01	8.82
2	2.47	0.30	8.64	0.13	10.95
3	3.70	0.77	10.44	0.42	12.90
4	4.94	1.36	12.16	0.84	14.74
5	6.17	2.03	13.82	1.36	16.50
6	7.41	2.77	15.43	1.94	18.21
7	8.64	3.55	17.00	2.58	19.86
8	9.88	4.36	18.54	3.26	21.47
9	11.11	5.21	20.05	3.99	23.04
10	12.35	6.08	21.53	4.75	24.59
11	13.58	6.98	23.00	5.53	26.11
12	14.81	7.90	24.45	6.35	27.60
13	16.05	8.83	25.88	7.18	29.08
14	17.28	9.78	27.30	8.04	30.53
15	18.52	10.75	28.70	8.92	31.97
16	19.75	11.73	30.09	9.81	33.39
17	20.99	12.73	31.46	10.72	34.79
18	22.22	13.73	32.83	11.65	36.18
19	23.46	14.75	34.18	12.59	37.55
20	24.69	15.78	35.53	13.54	38.92
21	25.93	16.82	36.86	14.51	40.27
22	27.16	17.87	38.19	15.49	41.60
23	28.40	18.93	39.50	16.48	42.93
24	29.63	19.99	40.81	17.49	44.24
25	30.86	21.07	42.11	18.50	45.55
26	32.10	22.15	43.40	19.52	46.84
27	33.33	23.24	44.68	20.56	48.12
28	34.57	24.34	45.96	21.61	49.40
29	35.80	25.45	47.23	22.66	50.66
30	37.04	26.56	48.49	23.73	51.92
31	38.27	27.69	49.74	24.80	53.16
32	39.51	28.81	50.99	25.88	54.40
33	40.74	29.95	52.23	26.97	55.63
34	41.98	31.09	53.46	28.08	56.84
35	43.21	32.24	54.69	29.18	58.05
36	44.44	33.40	55.91	30.30	59.26
37	45.68	34.56	57.13	31.43	60.45
38	46.91	35.73	58.33	32.56	61.64
39	48.15	36.90	59.53	33.71	62.81
40	49.38	38.08	60.73	34.86	63.98
N = 82					
0	0.00	0.00	4.40	0.00	6.26
1	1.22	0.03	6.61	0.01	8.71
2	2.44	0.30	8.53	0.13	10.82
3	3.66	0.76	10.32	0.42	12.75
4	4.88	1.34	12.02	0.83	14.57
5	6.10	2.01	13.66	1.34	16.31
6	7.32	2.73	15.25	1.92	18.00
7	8.54	3.50	16.80	2.55	19.63
8	9.76	4.31	18.32	3.22	21.22
9	10.98	5.14	19.82	3.94	22.78
10	12.20	6.01	21.29	4.69	24.31
11	13.41	6.89	22.74	5.46	25.81
12	14.63	7.80	24.17	6.27	27.29
13	15.85	8.72	25.58	7.09	28.75
14	17.07	9.66	26.98	7.94	30.19
15	18.29	10.62	28.37	8.80	31.61
16	19.51	11.58	29.74	9.68	33.02
17	20.73	12.57	31.11	10.58	34.41

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 82 (continued)					
18	21.95	13.56	32.46	11.50	35.78
19	23.17	14.56	33.80	12.42	37.14
20	24.39	15.58	35.12	13.37	38.49
21	25.61	16.60	36.44	14.32	39.82
22	26.83	17.64	37.76	15.29	41.15
23	28.05	18.68	39.06	16.27	42.46
24	29.27	19.74	40.35	17.26	43.76
25	30.49	20.80	41.64	18.26	45.05
26	31.71	21.87	42.92	19.27	46.33
27	32.93	22.94	44.19	20.29	47.61
28	34.15	24.03	45.45	21.32	48.87
29	35.37	25.12	46.70	22.36	50.12
30	36.59	26.22	47.95	23.41	51.36
31	37.80	27.32	49.19	24.47	52.60
32	39.02	28.44	50.43	25.54	53.82
33	40.24	29.56	51.66	26.61	55.04
34	41.46	30.68	52.88	27.70	56.25
35	42.68	31.82	54.09	28.79	57.45
36	43.90	32.96	55.30	29.89	58.64
37	45.12	34.10	56.51	31.00	59.82
38	46.34	35.25	57.70	32.12	61.00
39	47.56	36.41	58.89	33.25	62.16
40	48.78	37.58	60.08	34.38	63.32
41	50.00	38.75	61.25	35.53	64.47
N = 83					
0	0.00	0.00	4.35	0.00	6.18
1	1.20	0.03	6.53	0.01	8.61
2	2.41	0.29	8.43	0.13	10.70
3	3.61	0.75	10.20	0.41	12.61
4	4.82	1.33	11.88	0.82	14.41
5	6.02	1.98	13.50	1.32	16.13
6	7.23	2.70	15.07	1.89	17.79
7	8.43	3.46	16.61	2.52	19.41
8	9.64	4.25	18.11	3.18	20.98
9	10.84	5.08	19.59	3.89	22.53
10	12.05	5.93	21.04	4.63	24.04
11	13.25	6.81	22.48	5.40	25.53
12	14.46	7.70	23.89	6.19	26.99
13	15.66	8.61	25.29	7.00	28.43
14	16.87	9.54	26.68	7.84	29.86
15	18.07	10.48	28.05	8.69	31.26
16	19.28	11.44	29.41	9.56	32.65
17	20.48	12.41	30.76	10.45	34.03
18	21.69	13.39	32.09	11.35	35.39
19	22.89	14.38	33.42	12.27	36.74
20	24.10	15.38	34.73	13.20	38.07
21	25.30	16.39	36.04	14.14	39.39
22	26.51	17.42	37.34	15.09	40.70
23	27.71	18.45	38.62	16.06	42.00
24	28.92	19.48	39.91	17.03	43.29
25	30.12	20.53	41.18	18.02	44.57
26	31.33	21.59	42.44	19.02	45.84
27	32.53	22.65	43.70	20.03	47.10
28	33.73	23.72	44.95	21.04	48.35
29	34.94	24.80	46.19	22.07	49.59
30	36.14	25.88	47.43	23.10	50.82
31	37.35	26.97	48.66	24.15	52.04
32	38.55	28.07	49.88	25.20	53.26
33	39.76	29.17	51.10	26.26	54.46
34	40.96	30.28	52.31	27.33	55.66
35	42.17	31.40	53.51	28.41	56.85
36	43.37	32.53	54.71	29.50	58.03
37	44.58	33.66	55.90	30.59	59.20
38	45.78	34.79	57.08	31.69	60.37
39	46.99	35.93	58.26	32.80	61.53
40	48.19	37.08	59.44	33.92	62.68
41	49.40	38.24	60.60	35.05	63.82
N = 84					
0	0.00	0.00	4.30	0.00	6.11
1	1.19	0.03	6.46	0.01	8.51
2	2.38	0.29	8.34	0.12	10.57
3	3.57	0.74	10.08	0.41	12.46
4	4.76	1.31	11.75	0.81	14.24
5	5.95	1.96	13.35	1.31	15.95
6	7.14	2.67	14.90	1.87	17.59
7	8.33	3.42	16.42	2.48	19.19
8	9.52	4.20	17.91	3.15	20.75
9	10.71	5.02	19.37	3.84	22.28
10	11.90	5.86	20.81	4.57	23.77
11	13.10	6.72	22.22	5.33	25.25
12	14.29	7.61	23.62	6.11	26.69
13	15.48	8.51	25.01	6.91	28.12
14	16.67	9.42	26.38	7.74	29.53
15	17.86	10.35	27.74	8.58	30.92

Confidence intervals (%) for the binomial distribution (N = 85-89)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p ($\frac{r}{N} \times 100$)	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 84 (continued)					
16	19.05	11.30	29.08	9.44	32.30
17	20.24	12.25	30.41	10.32	33.66
18	21.43	13.22	31.74	11.21	35.01
19	22.62	14.20	33.05	12.11	36.34
20	23.81	15.19	34.35	13.03	37.66
21	25.00	16.19	35.64	13.96	38.97
22	26.19	17.20	36.93	14.90	40.27
23	27.38	18.21	38.20	15.85	41.56
24	28.57	19.24	39.47	16.82	42.83
25	29.76	20.27	40.73	17.79	44.10
26	30.95	21.31	41.98	18.77	45.35
27	32.14	22.36	43.22	19.77	46.60
28	33.33	23.42	44.46	20.77	47.84
29	34.52	24.48	45.69	21.78	49.07
30	35.71	25.55	46.92	22.81	50.29
31	36.90	26.63	48.13	23.83	51.50
32	38.10	27.71	49.34	24.87	52.70
33	39.29	28.80	50.55	25.92	53.90
34	40.48	29.90	51.75	26.98	55.09
35	41.67	31.00	52.94	28.04	56.27
36	42.86	32.11	54.12	29.11	57.44
37	44.05	33.22	55.30	30.19	58.60
38	45.24	34.34	56.48	31.27	59.76
39	46.43	35.47	57.65	32.37	60.90
40	47.62	36.60	58.81	33.47	62.04
41	48.81	37.74	59.96	34.58	63.18
42	50.00	38.89	61.11	35.70	64.30
N = 85					
0	0.00	0.00	4.25	0.00	6.04
1	1.18	0.03	6.38	0.01	8.42
2	2.35	0.29	8.24	0.12	10.45
3	3.53	0.73	9.97	0.40	12.32
4	4.71	1.30	11.61	0.80	14.08
5	5.88	1.94	13.20	1.29	15.77
6	7.06	2.63	14.73	1.85	17.40
7	8.24	3.38	16.23	2.45	18.98
8	9.41	4.15	17.71	3.11	20.52
9	10.59	4.96	19.15	3.80	22.03
10	11.76	5.79	20.57	4.52	23.51
11	12.94	6.64	21.98	5.26	24.97
12	14.12	7.51	23.36	6.04	26.40
13	15.29	8.40	24.73	6.83	27.82
14	16.47	9.31	26.09	7.64	29.21
15	17.65	10.23	27.43	8.48	30.59
16	18.82	11.16	28.76	9.33	31.95
17	20.00	12.10	30.08	10.19	33.30
18	21.18	13.06	31.39	11.07	34.63
19	22.35	14.03	32.69	11.96	35.95
20	23.53	15.00	33.97	12.87	37.26
21	24.71	15.99	35.25	13.78	38.56
22	25.88	16.99	36.52	14.71	39.84
23	27.06	17.99	37.79	15.65	41.12
24	28.24	19.00	39.04	16.61	42.38
25	29.41	20.02	40.29	17.57	43.63
26	30.59	21.05	41.53	18.54	44.88
27	31.76	22.08	42.76	19.52	46.12
28	32.94	23.13	43.98	20.51	47.34
29	34.12	24.18	45.20	21.51	48.56
30	35.29	25.23	46.41	22.51	49.77
31	36.47	26.29	47.62	23.53	50.97
32	37.65	27.36	48.82	24.55	52.16
33	38.82	28.44	50.01	25.59	53.35
34	40.00	29.52	51.20	26.63	54.52
35	41.18	30.61	52.38	27.68	55.69
36	42.35	31.70	53.55	28.73	56.85
37	43.53	32.80	54.72	29.80	58.01
38	44.71	33.91	55.89	30.87	59.15
39	45.88	35.02	57.04	31.95	60.29
40	47.06	36.13	58.19	33.03	61.42
41	48.24	37.26	59.34	34.13	62.55
42	49.41	38.39	60.48	35.23	63.66
N = 86					
0	0.00	0.00	4.20	0.00	5.97
1	1.16	0.03	6.31	0.01	8.32
2	2.33	0.28	8.15	0.12	10.34
3	3.49	0.73	9.86	0.40	12.19
4	4.65	1.28	11.48	0.79	13.93
5	5.81	1.91	13.05	1.28	15.60
6	6.98	2.60	14.57	1.82	17.21
7	8.14	3.34	16.05	2.43	18.77
8	9.30	4.10	17.51	3.07	20.30
9	10.47	4.90	18.94	3.75	21.79
10	11.63	5.72	20.35	4.46	23.26
11	12.79	6.56	21.73	5.20	24.70
12	13.95	7.42	23.11	5.96	26.12
13	15.12	8.30	24.46	6.75	27.52
14	16.28	9.20	25.80	7.55	28.90
15	17.44	10.10	27.13	8.37	30.26
16	18.60	11.02	28.45	9.21	31.61
17	19.77	11.96	29.75	10.07	32.94
18	20.93	12.90	31.05	10.93	34.26
19	22.09	13.86	32.33	11.81	35.57
20	23.26	14.82	33.61	12.71	36.87
21	24.42	15.79	34.87	13.61	38.15
22	25.58	16.78	36.13	14.53	39.42
23	26.74	17.77	37.38	15.46	40.69
24	27.91	18.77	38.62	16.40	41.94
25	29.07	19.78	39.86	17.35	43.18
26	30.23	20.79	41.08	18.31	44.41
27	31.40	21.81	42.30	19.27	45.64
28	32.56	22.84	43.52	20.25	46.85
29	33.72	23.88	44.72	21.24	48.06
30	34.88	24.92	45.92	22.23	49.26
31	36.05	25.97	47.12	23.23	50.45
32	37.21	27.02	48.30	24.24	51.63
33	38.37	28.08	49.49	25.26	52.80
34	39.53	29.15	50.66	26.29	53.97
35	40.70	30.22	51.83	27.32	55.13
36	41.86	31.30	52.99	28.37	56.28
37	43.02	32.39	54.15	29.41	57.43
38	44.19	33.48	55.30	30.47	58.56
39	45.35	34.58	56.45	31.54	59.69
40	46.51	35.68	57.59	32.61	60.81
41	47.67	36.79	58.73	33.69	61.93
42	48.84	37.90	59.86	34.77	63.04
43	50.00	39.02	60.98	35.86	64.14
N = 87					
0	0.00	0.00	4.15	0.00	5.91
1	1.15	0.03	6.24	0.01	8.23
2	2.30	0.28	8.06	0.12	10.22
3	3.45	0.72	9.75	0.39	12.05
4	4.60	1.27	11.36	0.78	13.78
5	5.75	1.89	12.90	1.26	15.43
6	6.90	2.57	14.41	1.80	17.02
7	8.05	3.30	15.88	2.40	18.57
8	9.20	4.05	17.32	3.03	20.08
9	10.34	4.84	18.73	3.71	21.56
10	11.49	5.65	20.12	4.41	23.01
11	12.64	6.48	21.50	5.14	24.44
12	13.79	7.34	22.85	5.89	25.84
13	14.94	8.20	24.20	6.67	27.23
14	16.09	9.09	25.52	7.46	28.59
15	17.24	9.98	26.84	8.27	29.94
16	18.39	10.89	28.14	9.10	31.28
17	19.54	11.81	29.43	9.94	32.60
18	20.69	12.75	30.71	10.80	33.91
19	21.84	13.69	31.98	11.67	35.20
20	22.99	14.64	33.25	12.55	36.48
21	24.14	15.60	34.50	13.45	37.75
22	25.29	16.58	35.75	14.35	39.02
23	26.44	17.55	36.98	15.27	40.27
24	27.59	18.54	38.21	16.20	41.51
25	28.74	19.54	39.43	17.13	42.74
26	29.89	20.54	40.65	18.08	43.96
27	31.03	21.55	41.86	19.04	45.17
28	32.18	22.56	43.06	20.00	46.38
29	33.33	23.58	44.25	20.97	47.57
30	34.48	24.61	45.44	21.95	48.76
31	35.63	25.65	46.62	22.94	49.94
32	36.78	26.69	47.80	23.94	51.11
33	37.93	27.74	48.97	24.95	52.27
34	39.08	28.79	50.13	25.96	53.43
35	40.23	29.85	51.29	26.98	54.58
36	41.38	30.92	52.45	28.01	55.72
37	42.53	31.99	53.59	29.04	56.85
38	43.68	33.06	54.74	30.08	57.98
39	44.83	34.15	55.87	31.13	59.10
40	45.98	35.23	57.00	32.19	60.21
41	47.13	36.33	58.13	33.26	61.32
42	48.28	37.42	59.25	34.33	62.42
43	49.43	38.53	60.36	35.40	63.51
N = 88					
0	0.00	0.00	4.11	0.00	5.84
1	1.14	0.03	6.17	0.01	8.14
2	2.27	0.28	7.97	0.12	10.11
3	3.41	0.71	9.64	0.39	11.92
4	4.55	1.25	11.23	0.77	13.63
5	5.68	1.87	12.76	1.25	15.26
6	6.82	2.54	14.25	1.78	16.84
7	7.95	3.26	15.70	2.37	18.37
8	9.09	4.01	17.13	3.00	19.87
9	10.23	4.78	18.53	3.66	21.33
10	11.36	5.59	19.91	4.36	22.77
11	12.50	6.41	21.27	5.08	24.18
12	13.64	7.25	22.61	5.82	25.57
13	14.77	8.11	23.94	6.59	26.94
14	15.91	8.98	25.25	7.37	28.29
15	17.05	9.87	26.55	8.17	29.63
16	18.18	10.76	27.84	8.99	30.95
17	19.32	11.68	29.12	9.82	32.26
18	20.45	12.60	30.39	10.67	33.55
19	21.59	13.53	31.65	11.53	34.84
20	22.73	14.47	32.89	12.40	36.11
21	23.86	15.42	34.14	13.29	37.37
22	25.00	16.38	35.37	14.18	38.61
23	26.14	17.34	36.59	15.09	39.85
24	27.27	18.32	37.81	16.00	41.08
25	28.41	19.30	39.02	16.93	42.30
26	29.55	20.29	40.22	17.86	43.51
27	30.68	21.29	41.42	18.80	44.71
28	31.82	22.29	42.61	19.76	45.91
29	32.95	23.30	43.79	20.72	47.09
30	34.09	24.32	44.97	21.68	48.27
31	35.23	25.34	46.14	22.66	49.44
32	36.36	26.37	47.31	23.65	50.60
33	37.50	27.40	48.47	24.64	51.75
34	38.64	28.44	49.62	25.64	52.90
35	39.77	29.49	50.77	26.64	54.04
36	40.91	30.54	51.91	27.66	55.17
37	42.05	31.60	53.05	28.68	56.30
38	43.18	32.66	54.18	29.71	57.41
39	44.32	33.73	55.30	30.74	58.52
40	45.45	34.80	56.42	31.79	59.63
41	46.59	35.88	57.54	32.84	60.72
42	47.73	36.96	58.65	33.89	61.81
43	48.86	38.05	59.75	34.96	62.90
44	50.00	39.15	60.85	36.03	63.97
N = 89					
0	0.00	0.00	4.06	0.00	5.78
1	1.12	0.03	6.10	0.01	8.05
2	2.25	0.27	7.88	0.12	10.00
3	3.37	0.70	9.54	0.38	11.79
4	4.49	1.24	11.11	0.77	13.48
5	5.62	1.85	12.63	1.23	15.10
6	6.74	2.51	14.10	1.76	16.66
7	7.87	3.22	15.54	2.34	18.18
8	8.99	3.96	16.95	2.96	19.66
9	10.11	4.73	18.33	3.62	21.11
10	11.24	5.52	19.69	4.31	22.53
11	12.36	6.33	21.04	5.02	23.93
12	13.48	7.17	22.37	5.75	25.30
13	14.61	8.01	23.68	6.51	26.66
14	15.73	8.88	24.98	7.29	28.00
15	16.85	9.75	26.27	8.08	29.32
16	17.98	10.64	27.55	8.89	30.63
17	19.10	11.54	28.81	9.71	31.93
18	20.22	12.45	30.07	10.55	33.21
19	21.35	13.37	31.31	11.39	34.48
20	22.47	14.30	32.55	12.26	35.74
21	23.60	15.24	33.78	13.13	36.99
22	24.72	16.19	35.00	14.01	38.22
23	25.84	17.14	36.21	14.91	39.45
24	26.97	18.10	37.42	15.81	40.67
25	28.09	19.07	38.62	16.72	41.88
26	29.21	20.05	39.81	17.65	43.07
27	30.34	21.03	40.99	18.58	44.27
28	31.46	22.03	42.17	19.52	45.45
29	32.58	23.02	43.34	20.47	46.62
30	33.71	24.03	44.51	21.42	47.79
31	34.83	25.04	45.67	22.39	48.95
32	35.96	26.05	46.82	23.36	50.10
33	37.08	27.07	47.97	24.34	51.24
34	38.20	28.10	49.11	25.32	52.38
35	39.33	29.13	50.25	26.32	53.51
36	40.45	30.17	51.38	27.32	54.63
37	41.57	31.21	52.51	28.33	55.75
38	42.70	32.26	53.63	29.34	56.85
39	43.82	33.32	54.75	30.36	57.96
40	44.94	34.38	55.86	31.39	59.05
41	46.07	35.44	56.96	32.43	60.14
42	47.19	36.51	58.06	33.47	61.22
43	48.31	37.59	59.16	34.52	62.30

Confidence intervals (%) for the binomial distribution (N = 90-95)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 89 (continued)					
44	49.44	38.67	60.25	35.57	63.36
N = 90					
0	0.00	0.00	4.02	0.00	5.72
1	1.11	0.03	6.04	0.01	7.97
2	2.22	0.27	7.80	0.12	9.90
3	3.33	0.69	9.43	0.38	11.67
4	4.44	1.22	10.99	0.76	13.34
5	5.56	1.83	12.49	1.22	14.94
6	6.67	2.49	13.95	1.74	16.48
7	7.78	3.18	15.37	2.32	17.99
8	8.89	3.92	16.77	2.93	19.45
9	10.00	4.68	18.14	3.58	20.89
10	11.11	5.46	19.49	4.26	22.29
11	12.22	6.26	20.82	4.96	23.68
12	13.33	7.08	22.13	5.69	25.04
13	14.44	7.92	23.43	6.44	26.38
14	15.56	8.77	24.72	7.20	27.71
15	16.67	9.64	26.00	7.98	29.02
16	17.78	10.52	27.26	8.78	30.32
17	18.89	11.41	28.51	9.60	31.60
18	20.00	12.31	29.75	10.42	32.87
19	21.11	13.21	30.99	11.26	34.13
20	22.22	14.13	32.21	12.11	35.38
21	23.33	15.06	33.43	12.97	36.61
22	24.44	16.00	34.64	13.85	37.84
23	25.56	16.94	35.84	14.73	39.05
24	26.67	17.89	37.03	15.62	40.26
25	27.78	18.85	38.22	16.52	41.46
26	28.89	19.82	39.40	17.44	42.65
27	30.00	20.79	40.57	18.36	43.83
28	31.11	21.77	41.74	19.28	45.00
29	32.22	22.75	42.90	20.22	46.16
30	33.33	23.74	44.05	21.16	47.32
31	34.44	24.74	45.20	22.12	48.47
32	35.56	25.74	46.35	23.08	49.61
33	36.67	26.75	47.49	24.04	50.74
34	37.78	27.77	48.62	25.02	51.87
35	38.89	28.79	49.74	26.00	52.99
36	40.00	29.81	50.87	26.99	54.10
37	41.11	30.84	51.98	27.98	55.21
38	42.22	31.88	53.09	28.98	56.31
39	43.33	32.92	54.20	29.99	57.40
40	44.44	33.96	55.30	31.01	58.49
41	45.56	35.02	56.40	32.03	59.57
42	46.67	36.07	57.49	33.06	60.64
43	47.78	37.13	58.57	34.09	61.71
44	48.89	38.20	59.65	35.13	62.77
45	50.00	39.27	60.73	36.18	63.82
N = 91					
0	0.00	0.00	3.97	0.00	5.66
1	1.10	0.03	5.97	0.01	7.88
2	2.20	0.27	7.71	0.11	9.79
3	3.30	0.69	9.33	0.37	11.55
4	4.40	1.21	10.87	0.75	13.20
5	5.49	1.81	12.36	1.20	14.78
6	6.59	2.46	13.80	1.72	16.31
7	7.69	3.15	15.21	2.29	17.80
8	8.79	3.87	16.59	2.90	19.25
9	9.89	4.62	17.95	3.54	20.67
10	10.99	5.40	19.28	4.21	22.06
11	12.09	6.19	20.60	4.91	23.43
12	13.19	7.00	21.90	5.62	24.78
13	14.29	7.83	23.19	6.36	26.12
14	15.38	8.67	24.46	7.12	27.43
15	16.48	9.53	25.73	7.89	28.73
16	17.58	10.40	26.98	8.68	30.01
17	18.68	11.28	28.22	9.49	31.28
18	19.78	12.16	29.45	10.30	32.54
19	20.88	13.06	30.67	11.13	33.79
20	21.98	13.97	31.88	11.97	35.02
21	23.08	14.89	33.09	12.82	36.25
22	24.18	15.81	34.28	13.69	37.46
23	25.27	16.75	35.47	14.56	38.66
24	26.37	17.69	36.65	15.44	39.86
25	27.47	18.63	37.83	16.33	41.05
26	28.57	19.59	39.00	17.23	42.22
27	29.67	20.55	40.16	18.14	43.39
28	30.77	21.51	41.32	19.06	44.56
29	31.87	22.49	42.47	19.98	45.71
30	32.97	23.47	43.61	20.91	46.85
31	34.07	24.45	44.75	21.85	47.99
32	35.16	25.44	45.88	22.80	49.12
33	36.26	26.44	47.01	23.76	50.25
34	37.36	27.44	48.13	24.72	51.37
35	38.46	28.45	49.25	25.69	52.48
36	39.56	29.46	50.36	26.66	53.58
37	40.66	30.48	51.47	27.65	54.68
38	41.76	31.50	52.57	28.63	55.77
39	42.86	32.53	53.66	29.63	56.85
40	43.96	33.56	54.75	30.63	57.93
41	45.05	34.60	55.84	31.64	59.00
42	46.15	35.64	56.92	32.65	60.07
43	47.25	36.69	58.00	33.68	61.12
44	48.35	37.74	59.07	34.70	62.18
45	49.45	38.80	60.14	35.74	63.22
N = 92					
0	0.00	0.00	3.93	0.00	5.60
1	1.09	0.03	5.91	0.01	7.80
2	2.17	0.26	7.63	0.11	9.69
3	3.26	0.68	9.24	0.37	11.43
4	4.35	1.20	10.76	0.74	13.06
5	5.43	1.79	12.23	1.19	14.63
6	6.52	2.43	13.66	1.70	16.15
7	7.61	3.11	15.05	2.26	17.62
8	8.70	3.83	16.42	2.86	19.05
9	9.78	4.57	17.76	3.50	20.46
10	10.87	5.34	19.08	4.16	21.84
11	11.96	6.12	20.39	4.85	23.20
12	13.04	6.93	21.68	5.56	24.53
13	14.13	7.74	22.95	6.29	25.85
14	15.22	8.58	24.21	7.04	27.15
15	16.30	9.42	25.46	7.80	28.44
16	17.39	10.28	26.70	8.58	29.71
17	18.48	11.15	27.93	9.38	30.97
18	19.57	12.03	29.15	10.18	32.22
19	20.65	12.92	30.36	11.00	33.45
20	21.74	13.81	31.56	11.83	34.67
21	22.83	14.72	32.75	12.68	35.89
22	23.91	15.63	33.94	13.53	37.09
23	25.00	16.55	35.11	14.39	38.28
24	26.09	17.48	36.29	15.26	39.47
25	27.17	18.42	37.45	16.14	40.64
26	28.26	19.36	38.61	17.03	41.81
27	29.35	20.31	39.76	17.93	42.97
28	30.43	21.27	40.90	18.83	44.12
29	31.52	22.23	42.04	19.75	45.27
30	32.61	23.20	43.18	20.67	46.40
31	33.70	24.17	44.30	21.60	47.53
32	34.78	25.15	45.43	22.53	48.65
33	35.87	26.13	46.54	23.48	49.77
34	36.96	27.12	47.66	24.43	50.87
35	38.04	28.12	48.76	25.38	51.98
36	39.13	29.12	49.86	26.35	53.07
37	40.22	30.12	50.96	27.32	54.16
38	41.30	31.13	52.05	28.29	55.24
39	42.39	32.15	53.14	29.28	56.32
40	43.48	33.17	54.22	30.27	57.38
41	44.57	34.19	55.30	31.26	58.45
42	45.65	35.22	56.37	32.26	59.50
43	46.74	36.26	57.44	33.27	60.55
44	47.83	37.30	58.50	34.28	61.60
45	48.91	38.34	59.56	35.30	62.64
46	50.00	39.39	60.61	36.33	63.67
N = 93					
0	0.00	0.00	3.89	0.00	5.54
1	1.08	0.03	5.85	0.01	7.72
2	2.15	0.26	7.55	0.11	9.59
3	3.23	0.67	9.14	0.37	11.31
4	4.30	1.18	10.65	0.73	12.93
5	5.38	1.77	12.10	1.18	14.48
6	6.45	2.40	13.52	1.68	15.98
7	7.53	3.08	14.90	2.24	17.44
8	8.60	3.79	16.25	2.83	18.86
9	9.68	4.52	17.58	3.46	20.25
10	10.75	5.28	18.89	4.12	21.62
11	11.83	6.05	20.18	4.80	22.96
12	12.90	6.85	21.45	5.50	24.29
13	13.98	7.66	22.72	6.22	25.59
14	15.05	8.48	23.97	6.96	26.88
15	16.13	9.32	25.20	7.72	28.16
16	17.20	10.17	26.43	8.49	29.42
17	18.28	11.02	27.65	9.27	30.66
18	19.35	11.89	28.85	10.07	31.90
19	20.43	12.77	30.05	10.88	33.12
20	21.51	13.66	31.24	11.70	34.33
21	22.58	14.55	32.42	12.53	35.54
22	23.66	15.46	33.60	13.37	36.73
23	24.73	16.37	34.76	14.23	37.91
24	25.81	17.29	35.92	15.09	39.08
25	26.88	18.21	37.08	15.96	40.25
26	27.96	19.14	38.22	16.84	41.41
27	29.03	20.08	39.36	17.72	42.56
28	30.11	21.03	40.50	18.62	43.70
29	31.18	21.98	41.63	19.52	44.83
30	32.26	22.93	42.75	20.43	45.96
31	33.33	23.89	43.87	21.35	47.08
32	34.41	24.86	44.98	22.27	48.19
33	35.48	25.83	46.09	23.20	49.29
34	36.56	26.81	47.19	24.14	50.39
35	37.63	27.79	48.28	25.09	51.48
36	38.71	28.78	49.38	26.04	52.57
37	39.78	29.78	50.46	27.00	53.65
38	40.86	30.77	51.54	27.96	54.72
39	41.94	31.78	52.62	28.93	55.79
40	43.01	32.78	53.69	29.91	56.85
41	44.09	33.80	54.76	30.89	57.90
42	45.16	34.81	55.83	31.88	58.95
43	46.24	35.84	56.88	32.87	59.99
44	47.31	36.86	57.94	33.88	61.03
45	48.39	37.89	58.99	34.88	62.06
46	49.46	38.93	60.03	35.90	63.09
N = 94					
0	0.00	0.00	3.85	0.00	5.48
1	1.06	0.03	5.79	0.01	7.64
2	2.13	0.26	7.48	0.11	9.49
3	3.19	0.66	9.04	0.36	11.19
4	4.26	1.17	10.54	0.72	12.80
5	5.32	1.75	11.98	1.16	14.33
6	6.38	2.38	13.38	1.67	15.82
7	7.45	3.05	14.74	2.21	17.26
8	8.51	3.75	16.08	2.80	18.67
9	9.57	4.47	17.40	3.42	20.05
10	10.64	5.22	18.70	4.07	21.40
11	11.70	5.99	19.97	4.74	22.74
12	12.77	6.77	21.24	5.44	24.05
13	13.83	7.57	22.49	6.15	25.34
14	14.89	8.39	23.72	6.88	26.62
15	15.96	9.22	24.95	7.63	27.88
16	17.02	10.05	26.16	8.39	29.13
17	18.09	10.90	27.37	9.17	30.36
18	19.15	11.76	28.56	9.96	31.59
19	20.21	12.63	29.75	10.76	32.80
20	21.28	13.51	30.93	11.57	34.00
21	22.34	14.39	32.10	12.39	35.19
22	23.40	15.29	33.26	13.22	36.37
23	24.47	16.19	34.42	14.07	37.54
24	25.53	17.09	35.57	14.92	38.71
25	26.60	18.01	36.71	15.78	39.86
26	27.66	18.93	37.85	16.65	41.01
27	28.72	19.86	38.98	17.52	42.15
28	29.79	20.79	40.10	18.41	43.28
29	30.85	21.73	41.22	19.30	44.40
30	31.91	22.67	42.33	20.20	45.52
31	32.98	23.62	43.44	21.10	46.63
32	34.04	24.58	44.54	22.02	47.73
33	35.11	25.54	45.64	22.94	48.83
34	36.17	26.51	46.73	23.86	49.92
35	37.23	27.48	47.82	24.80	51.00
36	38.30	28.46	48.90	25.74	52.08
37	39.36	29.44	49.98	26.68	53.15
38	40.43	30.42	51.05	27.64	54.21
39	41.49	31.41	52.12	28.59	55.27
40	42.55	32.41	53.18	29.56	56.32
41	43.62	33.41	54.24	30.53	57.37
42	44.68	34.41	55.29	31.51	58.41
43	45.74	35.42	56.34	32.49	59.44
44	46.81	36.44	57.39	33.48	60.47
45	47.87	37.46	58.43	34.47	61.50
46	48.94	38.48	59.46	35.47	62.51
47	50.00	39.51	60.49	36.48	63.52
N = 95					
0	0.00	0.00	3.81	0.00	5.42
1	1.05	0.03	5.73	0.01	7.56
2	2.11	0.26	7.40	0.11	9.40
3	3.16	0.66	8.95	0.36	11.08
4	4.21	1.16	10.43	0.72	12.67
5	5.26	1.73	11.86	1.15	14.19
6	6.32	2.35	13.24	1.65	15.66
7	7.37	3.01	14.59	2.19	17.09
8	8.42	3.71	15.92	2.77	18.49
9	9.47	4.42	17.22	3.39	19.85
10	10.53	5.16	18.51	4.03	21.19

Confidence intervals (%) for the binomial distribution (N = 96-99)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p ($\frac{r}{N} \times 100$)	95% Lower	95% Upper	99% Lower	99% Upper
N = 95 (continued)					
11	11.58	5.92	19.77	4.69	22.51
12	12.63	6.70	21.03	5.38	23.81
13	13.68	7.49	22.26	6.08	25.09
14	14.74	8.30	23.49	6.81	26.36
15	15.79	9.12	24.70	7.55	27.61
16	16.84	9.94	25.90	8.30	28.84
17	17.89	10.78	27.10	9.07	30.07
18	18.95	11.63	28.28	9.85	31.28
19	20.00	12.49	29.46	10.64	32.48
20	21.05	13.36	30.62	11.44	33.67
21	22.11	14.23	31.78	12.25	34.85
22	23.16	15.12	32.94	13.08	36.02
23	24.21	16.01	34.08	13.91	37.18
24	25.26	16.91	35.22	14.75	38.34
25	26.32	17.81	36.35	15.60	39.48
26	27.37	18.72	37.48	16.46	40.62
27	28.42	19.64	38.60	17.32	41.75
28	29.47	20.56	39.71	18.20	42.87
29	30.53	21.49	40.82	19.08	43.98
30	31.58	22.42	41.92	19.97	45.09
31	32.63	23.36	43.02	20.86	46.19
32	33.68	24.31	44.11	21.77	47.28
33	34.74	25.26	45.20	22.68	48.37
34	35.79	26.21	46.28	23.59	49.45
35	36.84	27.17	47.36	24.51	50.53
36	37.89	28.14	48.43	25.44	51.59
37	38.95	29.11	49.50	26.38	52.66
38	40.00	30.08	50.56	27.32	53.71
39	41.05	31.06	51.62	28.26	54.76
40	42.11	32.04	52.67	29.22	55.81
41	43.16	33.03	53.72	30.18	56.84
42	44.21	34.02	54.77	31.14	57.88
43	45.26	35.02	55.81	32.11	58.90
44	46.32	36.02	56.85	33.09	59.92
45	47.37	37.03	57.88	34.07	60.94
46	48.42	38.04	58.90	35.06	61.95
47	49.47	39.05	59.93	36.05	62.95
N = 96					
0	0.00	0.00	3.77	0.00	5.37
1	1.04	0.03	5.67	0.01	7.49
2	2.08	0.25	7.32	0.11	9.30
3	3.13	0.65	8.86	0.36	10.97
4	4.17	1.15	10.33	0.71	12.54
5	5.21	1.71	11.74	1.14	14.05
6	6.25	2.33	13.11	1.63	15.51
7	7.29	2.98	14.45	2.17	16.92
8	8.33	3.67	15.76	2.74	18.30
9	9.38	4.38	17.05	3.35	19.66
10	10.42	5.11	18.32	3.98	20.99
11	11.46	5.86	19.58	4.64	22.29
12	12.50	6.63	20.82	5.32	23.58
13	13.54	7.41	22.04	6.02	24.85
14	14.58	8.21	23.26	6.73	26.10
15	15.63	9.02	24.46	7.47	27.34
16	16.67	9.84	25.65	8.21	28.57
17	17.71	10.67	26.83	8.97	29.78
18	18.75	11.51	28.00	9.74	30.98
19	19.79	12.36	29.17	10.52	32.17
20	20.83	13.21	30.33	11.32	33.35
21	21.88	14.08	31.47	12.12	34.52
22	22.92	14.95	32.61	12.93	35.68
23	23.96	15.83	33.75	13.76	36.83
24	25.00	16.72	34.88	14.59	37.97
25	26.04	17.62	36.00	15.43	39.11
26	27.08	18.52	37.11	16.28	40.24
27	28.13	19.42	38.22	17.13	41.36
28	29.17	20.33	39.33	18.00	42.47
29	30.21	21.25	40.43	18.87	43.57
30	31.25	22.18	41.52	19.75	44.67
31	32.29	23.10	42.61	20.63	45.76
32	33.33	24.04	43.69	21.52	46.85
33	34.38	24.98	44.77	22.42	47.92
34	35.42	25.92	45.84	23.33	48.99
35	36.46	26.87	46.91	24.24	50.06
36	37.50	27.82	47.97	25.16	51.12
37	38.54	28.78	49.03	26.08	52.17
38	39.58	29.75	50.08	27.01	53.22
39	40.63	30.71	51.13	27.94	54.26
40	41.67	31.68	52.18	28.88	55.30
41	42.71	32.66	53.22	29.83	56.33
42	43.75	33.64	54.25	30.78	57.35
43	44.79	34.63	55.29	31.74	58.37
44	45.83	35.62	56.31	32.71	59.38
45	46.88	36.61	57.34	33.68	60.39
46	47.92	37.61	58.36	34.65	61.39
47	48.96	38.61	59.37	35.63	62.39
48	50.00	39.62	60.38	36.62	63.38
N = 97					
0	0.00	0.00	3.73	0.00	5.32
1	1.03	0.03	5.61	0.01	7.41
2	2.06	0.25	7.25	0.11	9.21
3	3.09	0.64	8.77	0.35	10.86
4	4.12	1.13	10.22	0.70	12.42
5	5.15	1.69	11.62	1.13	13.91
6	6.19	2.30	12.98	1.61	15.35
7	7.22	2.95	14.30	2.14	16.76
8	8.25	3.63	15.61	2.71	18.13
9	9.28	4.33	16.88	3.31	19.47
10	10.31	5.06	18.14	3.94	20.78
11	11.34	5.80	19.39	4.59	22.08
12	12.37	6.56	20.61	5.26	23.35
13	13.40	7.33	21.83	5.95	24.61
14	14.43	8.12	23.03	6.66	25.85
15	15.46	8.92	24.22	7.39	27.08
16	16.49	9.73	25.40	8.12	28.29
17	17.53	10.55	26.57	8.87	29.50
18	18.56	11.38	27.73	9.63	30.69
19	19.59	12.22	28.89	10.41	31.87
20	20.62	13.07	30.03	11.19	33.03
21	21.65	13.93	31.17	11.99	34.19
22	22.68	14.79	32.30	12.79	35.34
23	23.71	15.66	33.42	13.61	36.49
24	24.74	16.54	34.54	14.43	37.62
25	25.77	17.42	35.65	15.26	38.74
26	26.80	18.32	36.76	16.10	39.86
27	27.84	19.21	37.86	16.94	40.97
28	28.87	20.11	38.95	17.80	42.07
29	29.90	21.02	40.04	18.66	43.17
30	30.93	21.93	41.12	19.53	44.26
31	31.96	22.85	42.20	20.40	45.34
32	32.99	23.78	43.27	21.28	46.41
33	34.02	24.70	44.34	22.17	47.48
34	35.05	25.64	45.41	23.07	48.55
35	36.08	26.58	46.46	23.97	49.60
36	37.11	27.52	47.52	24.87	50.65
37	38.14	28.47	48.57	25.79	51.70
38	39.18	29.42	49.61	26.70	52.74
39	40.21	30.37	50.65	27.63	53.77
40	41.24	31.33	51.69	28.56	54.80
41	42.27	32.30	52.72	29.49	55.82
42	43.30	33.27	53.75	30.44	56.84
43	44.33	34.24	54.77	31.38	57.85
44	45.36	35.22	55.79	32.33	58.85
45	46.39	36.20	56.81	33.29	59.86
46	47.42	37.19	57.82	34.26	60.85
47	48.45	38.18	58.82	35.22	61.84
48	49.48	39.17	59.83	36.20	62.82
N = 98					
0	0.00	0.00	3.69	0.00	5.26
1	1.02	0.03	5.55	0.01	7.34
2	2.04	0.25	7.18	0.11	9.12
3	3.06	0.64	8.69	0.35	10.75
4	4.08	1.12	10.12	0.69	12.30
5	5.10	1.68	11.51	1.12	13.78
6	6.12	2.28	12.85	1.60	15.21
7	7.14	2.92	14.16	2.12	16.60
8	8.16	3.59	15.45	2.69	17.95
9	9.18	4.29	16.72	3.28	19.28
10	10.20	5.00	17.97	3.90	20.58
11	11.22	5.74	19.20	4.54	21.87
12	12.24	6.49	20.41	5.21	23.13
13	13.27	7.26	21.62	5.89	24.38
14	14.29	8.04	22.81	6.59	25.61
15	15.31	8.83	23.99	7.31	26.82
16	16.33	9.63	25.16	8.04	28.03
17	17.35	10.44	26.31	8.78	29.22
18	18.37	11.26	27.47	9.53	30.40
19	19.39	12.10	28.61	10.30	31.57
20	20.41	12.93	29.74	11.07	32.72
21	21.43	13.78	30.87	11.86	33.87
22	22.45	14.64	31.99	12.65	35.01
23	23.47	15.50	33.11	13.46	36.15
24	24.49	16.36	34.21	14.27	37.27
25	25.51	17.24	35.31	15.09	38.38
26	26.53	18.12	36.41	15.92	39.49
27	27.55	19.01	37.50	16.76	40.59
28	28.57	19.90	38.58	17.60	41.69
29	29.59	20.79	39.66	18.46	42.77
30	30.61	21.70	40.74	19.31	43.85
31	31.63	22.61	41.80	20.18	44.92
32	32.65	23.52	42.87	21.05	45.99
33	33.67	24.44	43.93	21.93	47.05
34	34.69	25.36	44.98	22.81	48.11
35	35.71	26.29	46.03	23.70	49.15
36	36.73	27.22	47.07	24.60	50.20
37	37.76	28.16	48.12	25.50	51.23
38	38.78	29.10	49.15	26.41	52.26
39	39.80	30.04	50.18	27.32	53.29
40	40.82	30.99	51.21	28.24	54.31
41	41.84	31.95	52.23	29.17	55.32
42	42.86	32.90	53.25	30.10	56.33
43	43.88	33.87	54.27	31.03	57.34
44	44.90	34.83	55.28	31.97	58.33
45	45.92	35.80	56.29	32.92	59.33
46	46.94	36.78	57.29	33.87	60.31
47	47.96	37.76	58.29	34.82	61.30
48	48.98	38.74	59.28	35.79	62.27
49	50.00	39.73	60.27	36.75	63.25
N = 99					
0	0.00	0.00	3.66	0.00	5.21
1	1.01	0.03	5.50	0.01	7.27
2	2.02	0.25	7.11	0.11	9.03
3	3.03	0.63	8.60	0.34	10.65
4	4.04	1.11	10.02	0.69	12.18
5	5.05	1.66	11.39	1.11	13.64
6	6.06	2.26	12.73	1.58	15.06
7	7.07	2.89	14.03	2.10	16.44
8	8.08	3.55	15.30	2.66	17.78
9	9.09	4.24	16.56	3.25	19.10
10	10.10	4.95	17.79	3.86	20.39
11	11.11	5.68	19.01	4.50	21.66
12	12.12	6.42	20.22	5.15	22.91
13	13.13	7.18	21.41	5.83	24.15
14	14.14	7.95	22.59	6.52	25.37
15	15.15	8.74	23.76	7.23	26.57
16	16.16	9.53	24.91	7.95	27.76
17	17.17	10.33	26.06	8.69	28.94
18	18.18	11.15	27.20	9.43	30.11
19	19.19	11.97	28.34	10.19	31.27
20	20.20	12.80	29.46	10.96	32.42
21	21.21	13.64	30.58	11.73	33.56
22	22.22	14.48	31.69	12.52	34.69
23	23.23	15.33	32.79	13.32	35.81
24	24.24	16.19	33.89	14.12	36.93
25	25.25	17.06	34.98	14.93	38.03
26	26.26	17.93	36.07	15.75	39.13
27	27.27	18.80	37.15	16.58	40.22
28	28.28	19.69	38.22	17.42	41.31
29	29.29	20.57	39.29	18.26	42.38
30	30.30	21.47	40.36	19.11	43.45
31	31.31	22.36	41.41	19.96	44.52
32	32.32	23.27	42.47	20.82	45.57
33	33.33	24.18	43.52	21.69	46.63
34	34.34	25.09	44.56	22.56	47.67
35	35.35	26.01	45.60	23.44	48.71
36	36.36	26.93	46.64	24.33	49.75
37	37.37	27.85	47.67	25.22	50.77
38	38.38	28.78	48.70	26.12	51.80
39	39.39	29.72	49.72	27.02	52.81
40	40.40	30.66	50.74	27.93	53.83
41	41.41	31.60	51.76	28.84	54.83
42	42.42	32.55	52.77	29.76	55.83
43	43.43	33.50	53.77	30.69	56.83
44	44.44	34.45	54.78	31.62	57.82
45	45.45	35.41	55.77	32.55	58.81
46	46.46	36.38	56.77	33.49	59.79
47	47.47	37.34	57.76	34.44	60.76
48	48.48	38.32	58.75	35.39	61.73
49	49.49	39.29	59.73	36.34	62.70

Confidence intervals (%) for the binomial distribution (N = 100-150)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 100					
0	0.00	0.00	3.62	0.00	5.16
1	1.00	0.03	5.45	0.01	7.20
2	2.00	0.24	7.04	0.10	8.94
3	3.00	0.62	8.52	0.34	10.55
4	4.00	1.10	9.93	0.68	12.06
5	5.00	1.64	11.28	1.09	13.51
6	6.00	2.23	12.60	1.56	14.92
7	7.00	2.86	13.89	2.08	16.28
8	8.00	3.52	15.16	2.63	17.61
9	9.00	4.20	16.40	3.21	18.92
10	10.00	4.90	17.62	3.82	20.20
11	11.00	5.62	18.83	4.45	21.45
12	12.00	6.36	20.02	5.10	22.70
13	13.00	7.11	21.20	5.77	23.92
14	14.00	7.87	22.37	6.45	25.13
15	15.00	8.65	23.53	7.15	26.32
16	16.00	9.43	24.68	7.87	27.51
17	17.00	10.23	25.82	8.59	28.68
18	18.00	11.03	26.95	9.33	29.84
19	19.00	11.84	28.07	10.08	30.98
20	20.00	12.67	29.18	10.84	32.12
21	21.00	13.49	30.29	11.61	33.25
22	22.00	14.33	31.39	12.39	34.37
23	23.00	15.17	32.49	13.18	35.49
24	24.00	16.02	33.57	13.97	36.59
25	25.00	16.88	34.66	14.77	37.69
26	26.00	17.74	35.73	15.59	38.77
27	27.00	18.61	36.80	16.40	39.86
28	28.00	19.48	37.87	17.23	40.93
29	29.00	20.36	38.93	18.06	42.00
30	30.00	21.24	39.98	18.90	43.06
31	31.00	22.13	41.03	19.75	44.12
32	32.00	23.02	42.08	20.60	45.17
33	33.00	23.92	43.12	21.46	46.21
34	34.00	24.82	44.15	22.32	47.25
35	35.00	25.73	45.18	23.19	48.28
36	36.00	26.64	46.21	24.07	49.30
37	37.00	27.56	47.24	24.95	50.32
38	38.00	28.48	48.25	25.84	51.34
39	39.00	29.40	49.27	26.73	52.35
40	40.00	30.33	50.28	27.63	53.35
41	41.00	31.26	51.29	28.53	54.35
42	42.00	32.20	52.29	29.44	55.35
43	43.00	33.14	53.29	30.35	56.33
44	44.00	34.08	54.28	31.27	57.32
45	45.00	35.03	55.27	32.19	58.30
46	46.00	35.98	56.26	33.12	59.27
47	47.00	36.94	57.24	34.06	60.24
48	48.00	37.90	58.22	34.99	61.20
49	49.00	38.86	59.20	35.94	62.16
50	50.00	39.83	60.17	36.89	63.11

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 125					
0	0.00	0.00	2.91	0.00	4.15
1	0.80	0.02	4.38	0.00	5.79
2	1.60	0.19	5.66	0.08	7.21
3	2.40	0.50	6.85	0.27	8.51
4	3.20	0.88	7.99	0.54	9.73
5	4.00	1.31	9.09	0.87	10.91
6	4.80	1.78	10.15	1.25	12.05
7	5.60	2.28	11.20	1.66	13.16
8	6.40	2.80	12.22	2.09	14.24
9	7.20	3.35	13.23	2.56	15.30
10	8.00	3.90	14.22	3.04	16.35
11	8.80	4.48	15.20	3.54	17.37
12	9.60	5.06	16.17	4.05	18.39
13	10.40	5.65	17.13	4.58	19.39
14	11.20	6.26	18.08	5.13	20.37
15	12.00	6.87	19.02	5.68	21.35
16	12.80	7.50	19.95	6.24	22.32
17	13.60	8.13	20.88	6.82	23.28
18	14.40	8.76	21.80	7.40	24.23
19	15.20	9.41	22.71	7.99	25.17
20	16.00	10.06	23.62	8.59	26.11
21	16.80	10.71	24.53	9.19	27.03
22	17.60	11.37	25.43	9.81	27.96
23	18.40	12.04	26.32	10.43	28.87
24	19.20	12.71	27.21	11.05	29.78
25	20.00	13.38	28.09	11.68	30.68
26	20.80	14.06	28.97	12.32	31.58
27	21.60	14.74	29.85	12.96	32.48
28	22.40	15.43	30.72	13.61	33.36
29	23.20	16.12	31.59	14.26	34.25
30	24.00	16.82	32.46	14.92	35.13
31	24.80	17.51	33.32	15.58	36.00
32	25.60	18.22	34.18	16.25	36.87
33	26.40	18.92	35.03	16.92	37.74
34	27.20	19.63	35.88	17.59	38.60
35	28.00	20.34	36.73	18.27	39.46
36	28.80	21.05	37.58	18.95	40.31
37	29.60	21.77	38.42	19.64	41.16
38	30.40	22.49	39.26	20.33	42.01
39	31.20	23.22	40.10	21.03	42.85
40	32.00	23.94	40.93	21.72	43.69
41	32.80	24.67	41.77	22.42	44.52
42	33.60	25.40	42.60	23.13	45.36
43	34.40	26.14	43.42	23.84	46.18
44	35.20	26.87	44.25	24.55	47.01
45	36.00	27.61	45.07	25.26	47.83
46	36.80	28.35	45.89	25.98	48.65
47	37.60	29.10	46.70	26.70	49.47
48	38.40	29.84	47.52	27.43	50.28
49	39.20	30.59	48.33	28.16	51.09
50	40.00	31.34	49.14	28.89	51.90
51	40.80	32.10	49.95	29.62	52.70
52	41.60	32.85	50.75	30.36	53.50
53	42.40	33.61	51.56	31.10	54.30
54	43.20	34.37	52.36	31.84	55.09
55	44.00	35.14	53.16	32.58	55.88
56	44.80	35.90	53.95	33.33	56.67
57	45.60	36.67	54.75	34.08	57.46
58	46.40	37.44	55.54	34.84	58.24
59	47.20	38.21	56.33	35.60	59.02
60	48.00	38.98	57.11	36.36	59.80
61	48.80	39.76	57.90	37.12	60.57
62	49.60	40.54	58.68	37.89	61.35

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 150					
0	0.00	0.00	2.43	0.00	3.47
1	0.67	0.02	3.66	0.00	4.85
2	1.33	0.16	4.73	0.07	6.03
3	2.00	0.41	5.73	0.23	7.13
4	2.67	0.73	6.69	0.45	8.16
5	3.33	1.09	7.61	0.73	9.15
6	4.00	1.48	8.50	1.04	10.11
7	4.67	1.90	9.38	1.38	11.04
8	5.33	2.33	10.24	1.74	11.95
9	6.00	2.78	11.08	2.12	12.85
10	6.67	3.24	11.92	2.52	13.73
11	7.33	3.72	12.74	2.94	14.60
12	8.00	4.20	13.56	3.36	15.45
13	8.67	4.70	14.36	3.80	16.29
14	9.33	5.20	15.16	4.25	17.13
15	10.00	5.71	15.96	4.71	17.96
16	10.67	6.22	16.74	5.18	18.77
17	11.33	6.74	17.52	5.65	19.59
18	12.00	7.27	18.30	6.13	20.39
19	12.67	7.80	19.07	6.62	21.19
20	13.33	8.34	19.84	7.11	21.98
21	14.00	8.88	20.60	7.61	22.77
22	14.67	9.43	21.36	8.12	23.55
23	15.33	9.98	22.11	8.63	24.33
24	16.00	10.53	22.86	9.14	25.10
25	16.67	11.09	23.61	9.66	25.87
26	17.33	11.65	24.36	10.19	26.63
27	18.00	12.21	25.10	10.72	27.39
28	18.67	12.78	25.84	11.25	28.15
29	19.33	13.35	26.57	11.79	28.90
30	20.00	13.92	27.30	12.33	29.64
31	20.67	14.49	28.03	12.87	30.39
32	21.33	15.07	28.76	13.42	31.13
33	22.00	15.65	29.49	13.97	31.87
34	22.67	16.24	30.21	14.52	32.60
35	23.33	16.82	30.93	15.08	33.34
36	24.00	17.41	31.65	15.64	34.06
37	24.67	18.00	32.36	16.20	34.79
38	25.33	18.59	33.07	16.77	35.51
39	26.00	19.19	33.79	17.34	36.23
40	26.67	19.78	34.49	17.91	36.95
41	27.33	20.38	35.20	18.48	37.67
42	28.00	20.98	35.91	19.06	38.38
43	28.67	21.59	36.61	19.64	39.09
44	29.33	22.19	37.31	20.22	39.80
45	30.00	22.80	38.01	20.80	40.50
46	30.67	23.41	38.71	21.39	41.21
47	31.33	24.02	39.41	21.98	41.91
48	32.00	24.63	40.10	22.57	42.61
49	32.67	25.24	40.79	23.16	43.30
50	33.33	25.86	41.48	23.76	44.00
51	34.00	26.47	42.17	24.35	44.69
52	34.67	27.09	42.86	24.95	45.38
53	35.33	27.71	43.55	25.56	46.07
54	36.00	28.33	44.23	26.16	46.75
55	36.67	28.96	44.92	26.77	47.44
56	37.33	29.58	45.60	27.37	48.12
57	38.00	30.21	46.28	27.98	48.80
58	38.67	30.84	46.95	28.60	49.48
59	39.33	31.47	47.63	29.21	50.15
60	40.00	32.10	48.31	29.83	50.82
61	40.67	32.73	48.98	30.44	51.50
62	41.33	33.36	49.65	31.06	52.17
63	42.00	34.00	50.32	31.69	52.83
64	42.67	34.64	50.99	32.31	53.50
65	43.33	35.27	51.66	32.94	54.17
66	44.00	35.91	52.33	33.56	54.83
67	44.67	36.55	52.99	34.19	55.49
68	45.33	37.20	53.66	34.82	56.15
69	46.00	37.84	54.32	35.46	56.80
70	46.67	38.49	54.98	36.09	57.46
71	47.33	39.13	55.64	36.73	58.11
72	48.00	39.78	56.30	37.37	58.76
73	48.67	40.43	56.95	38.01	59.41
74	49.33	41.08	57.61	38.65	60.06
75	50.00	41.74	58.26	39.29	60.71

Confidence intervals (%) for the binomial distribution (N = 200-400)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 200					
0	0.00	0.00	1.83	0.00	2.61
1	0.50	0.01	2.75	0.00	3.66
2	1.00	0.12	3.57	0.05	4.55
3	1.50	0.31	4.32	0.17	5.38
4	2.00	0.55	5.04	0.34	6.16
5	2.50	0.82	5.74	0.54	6.91
6	3.00	1.11	6.42	0.78	7.64
7	3.50	1.42	7.08	1.03	8.35
8	4.00	1.74	7.73	1.30	9.05
9	4.50	2.08	8.37	1.59	9.73
10	5.00	2.42	9.00	1.88	10.40
11	5.50	2.78	9.63	2.19	11.06
12	6.00	3.14	10.25	2.51	11.71
13	6.50	3.51	10.86	2.84	12.35
14	7.00	3.88	11.47	3.17	12.99
15	7.50	4.26	12.07	3.51	13.62
16	8.00	4.64	12.67	3.86	14.25
17	8.50	5.03	13.26	4.21	14.87
18	9.00	5.42	13.85	4.57	15.48
19	9.50	5.82	14.44	4.93	16.09
20	10.00	6.22	15.02	5.29	16.70
21	10.50	6.62	15.60	5.66	17.30
22	11.00	7.02	16.18	6.04	17.90
23	11.50	7.43	16.75	6.42	18.50
24	12.00	7.84	17.33	6.80	19.09
25	12.50	8.26	17.90	7.18	19.68
26	13.00	8.67	18.47	7.57	20.26
27	13.50	9.09	19.03	7.96	20.85
28	14.00	9.51	19.59	8.35	21.43
29	14.50	9.93	20.16	8.75	22.00
30	15.00	10.35	20.72	9.15	22.58
31	15.50	10.78	21.27	9.55	23.15
32	16.00	11.21	21.83	9.95	23.72
33	16.50	11.64	22.38	10.36	24.29
34	17.00	12.07	22.94	10.77	24.86
35	17.50	12.50	23.49	11.18	25.42
36	18.00	12.94	24.04	11.59	25.99
37	18.50	13.37	24.59	12.00	26.55
38	19.00	13.81	25.13	12.42	27.11
39	19.50	14.25	25.68	12.84	27.66
40	20.00	14.69	26.22	13.26	28.22
41	20.50	15.13	26.77	13.68	28.77
42	21.00	15.57	27.31	14.10	29.32
43	21.50	16.02	27.85	14.53	29.87
44	22.00	16.46	28.39	14.95	30.42
45	22.50	16.91	28.92	15.38	30.97
46	23.00	17.36	29.46	15.81	31.51
47	23.50	17.81	30.00	16.24	32.06
48	24.00	18.26	30.53	16.67	32.60
49	24.50	18.71	31.06	17.11	33.14
50	25.00	19.16	31.60	17.54	33.68
60	30.00	23.74	36.86	21.97	39.01
70	35.00	28.41	42.05	26.51	44.22
80	40.00	33.15	47.15	31.16	49.33
90	45.00	37.98	52.18	35.90	54.34
100	50.00	42.87	57.13	40.74	59.26

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 300					
0	0.00	0.00	1.22	0.00	1.75
1	0.33	0.01	1.84	0.00	2.45
2	0.67	0.08	2.39	0.03	3.05
3	1.00	0.21	2.89	0.11	3.61
4	1.33	0.36	3.38	0.22	4.14
5	1.67	0.54	3.85	0.36	4.65
6	2.00	0.74	4.30	0.52	5.14
7	2.33	0.94	4.75	0.68	5.62
8	2.67	1.16	5.19	0.86	6.08
9	3.00	1.38	5.62	1.05	6.54
10	3.33	1.61	6.04	1.25	7.00
11	3.67	1.84	6.47	1.45	7.44
12	4.00	2.08	6.88	1.66	7.89
13	4.33	2.33	7.30	1.88	8.32
14	4.67	2.57	7.71	2.10	8.75
15	5.00	2.83	8.11	2.33	9.18
16	5.33	3.08	8.52	2.55	9.61
17	5.67	3.34	8.92	2.79	10.03
18	6.00	3.59	9.32	3.02	10.45
19	6.33	3.86	9.71	3.26	10.86
20	6.67	4.12	10.11	3.50	11.27
21	7.00	4.39	10.50	3.75	11.68
22	7.33	4.65	10.89	3.99	12.09
23	7.67	4.92	11.28	4.24	12.50
24	8.00	5.19	11.67	4.49	12.90
25	8.33	5.47	12.06	4.75	13.30
26	8.67	5.74	12.44	5.00	13.70
27	9.00	6.01	12.82	5.26	14.10
28	9.33	6.29	13.21	5.52	14.49
29	9.67	6.57	13.59	5.78	14.89
30	10.00	6.85	13.97	6.04	15.28
31	10.33	7.13	14.35	6.30	15.67
32	10.67	7.41	14.72	6.57	16.06
33	11.00	7.69	15.10	6.83	16.45
34	11.33	7.98	15.48	7.10	16.84
35	11.67	8.26	15.85	7.37	17.23
36	12.00	8.55	16.22	7.64	17.61
37	12.33	8.83	16.60	7.91	18.00
38	12.67	9.12	16.97	8.18	18.38
39	13.00	9.41	17.34	8.45	18.76
40	13.33	9.70	17.71	8.73	19.14
41	13.67	9.99	18.08	9.00	19.52
42	14.00	10.28	18.45	9.28	19.90
43	14.33	10.57	18.82	9.56	20.28
44	14.67	10.86	19.18	9.84	20.65
45	15.00	11.16	19.55	10.12	21.03
46	15.33	11.45	19.92	10.40	21.41
47	15.67	11.74	20.28	10.68	21.78
48	16.00	12.04	20.65	10.96	22.15
49	16.33	12.33	21.01	11.24	22.53
50	16.67	12.63	21.38	11.53	22.90
60	20.00	15.62	24.98	14.40	26.58
70	23.33	18.66	28.54	17.35	30.19
80	26.67	21.75	32.05	20.34	33.75
90	30.00	24.87	35.53	23.39	37.27
100	33.33	28.02	38.98	26.47	40.74
125	41.67	36.03	47.47	34.35	49.25
150	50.00	44.20	55.80	42.45	57.55

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
N = 400					
0	0.00	0.00	0.92	0.00	1.32
1	0.25	0.01	1.38	0.00	1.84
2	0.50	0.06	1.79	0.03	2.30
3	0.75	0.15	2.18	0.08	2.72
4	1.00	0.27	2.54	0.17	3.11
5	1.25	0.41	2.89	0.27	3.50
6	1.50	0.55	3.24	0.39	3.87
7	1.75	0.71	3.57	0.51	4.23
8	2.00	0.87	3.90	0.65	4.58
9	2.25	1.03	4.23	0.79	4.93
10	2.50	1.21	4.55	0.94	5.27
11	2.75	1.38	4.87	1.09	5.61
12	3.00	1.56	5.18	1.25	5.94
13	3.25	1.74	5.49	1.41	6.27
14	3.50	1.93	5.80	1.57	6.60
15	3.75	2.11	6.11	1.74	6.93
16	4.00	2.30	6.41	1.91	7.25
17	4.25	2.49	6.72	2.08	7.57
18	4.50	2.69	7.02	2.26	7.88
19	4.75	2.88	7.32	2.44	8.20
20	5.00	3.08	7.62	2.62	8.51
21	5.25	3.28	7.91	2.80	8.82
22	5.50	3.48	8.21	2.98	9.13
23	5.75	3.68	8.50	3.17	9.43
24	6.00	3.88	8.80	3.36	9.74
25	6.25	4.09	9.09	3.54	10.04
26	6.50	4.29	9.38	3.73	10.35
27	6.75	4.50	9.67	3.92	10.65
28	7.00	4.70	9.96	4.12	10.95
29	7.25	4.91	10.25	4.31	11.25
30	7.50	5.12	10.53	4.51	11.55
31	7.75	5.33	10.82	4.70	11.84
32	8.00	5.54	11.11	4.90	12.14
33	8.25	5.75	11.39	5.10	12.44
34	8.50	5.96	11.68	5.30	12.73
35	8.75	6.17	11.96	5.50	13.02
36	9.00	6.38	12.24	5.70	13.32
37	9.25	6.60	12.52	5.90	13.61
38	9.50	6.81	12.81	6.10	13.90
39	9.75	7.03	13.09	6.30	14.19
40	10.00	7.24	13.37	6.51	14.48
41	10.25	7.46	13.65	6.71	14.77
42	10.50	7.67	13.93	6.92	15.06
43	10.75	7.89	14.21	7.12	15.34
44	11.00	8.11	14.48	7.33	15.63
45	11.25	8.33	14.76	7.54	15.92
46	11.50	8.54	15.04	7.75	16.20
47	11.75	8.76	15.32	7.96	16.49
48	12.00	8.98	15.59	8.16	16.77
49	12.25	9.20	15.87	8.37	17.05
50	12.50	9.42	16.15	8.59	17.34
60	15.00	11.65	18.88	10.72	20.14
70	17.50	13.90	21.59	12.90	22.90
80	20.00	16.19	24.26	15.11	25.62
90	22.50	18.50	26.91	17.35	28.32
100	25.00	20.83	29.54	19.63	30.98
125	31.25	26.74	36.04	25.41	37.55
150	37.50	32.74	42.45	31.32	43.98
175	43.75	38.83	48.77	37.34	50.31
200	50.00	44.99	55.01	43.47	56.53

Confidence intervals (%) for the binomial distribution (N = 500-700)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
	N = 500				
0	0.00	0.00	0.74	0.00	1.05
1	0.20	0.01	1.11	0.00	1.48
2	0.40	0.05	1.44	0.02	1.84
3	0.60	0.12	1.74	0.07	2.18
4	0.80	0.22	2.04	0.13	2.50
5	1.00	0.33	2.32	0.22	2.80
6	1.20	0.44	2.59	0.31	3.10
7	1.40	0.56	2.86	0.41	3.39
8	1.60	0.69	3.13	0.52	3.68
9	1.80	0.83	3.39	0.63	3.96
10	2.00	0.96	3.65	0.75	4.23
11	2.20	1.10	3.90	0.87	4.50
12	2.40	1.25	4.15	0.99	4.77
13	2.60	1.39	4.41	1.12	5.04
14	2.80	1.54	4.65	1.25	5.30
15	3.00	1.69	4.90	1.39	5.56
16	3.20	1.84	5.14	1.52	5.82
17	3.40	1.99	5.39	1.66	6.07
18	3.60	2.15	5.63	1.80	6.33
19	3.80	2.30	5.87	1.95	6.58
20	4.00	2.46	6.11	2.09	6.83
21	4.20	2.62	6.35	2.23	7.08
22	4.40	2.78	6.59	2.38	7.33
23	4.60	2.94	6.82	2.53	7.58
24	4.80	3.10	7.06	2.68	7.82
25	5.00	3.26	7.29	2.83	8.07
26	5.20	3.42	7.53	2.98	8.31
27	5.40	3.59	7.76	3.13	8.56
28	5.60	3.75	7.99	3.28	8.80
29	5.80	3.92	8.22	3.44	9.04
30	6.00	4.08	8.45	3.59	9.28
31	6.20	4.25	8.69	3.75	9.52
32	6.40	4.42	8.92	3.91	9.76
33	6.60	4.59	9.14	4.06	9.99
34	6.80	4.75	9.37	4.22	10.23
35	7.00	4.92	9.60	4.38	10.47
36	7.20	5.09	9.83	4.54	10.70
37	7.40	5.26	10.06	4.70	10.94
38	7.60	5.43	10.28	4.86	11.17
39	7.80	5.61	10.51	5.03	11.41
40	8.00	5.78	10.73	5.19	11.64
41	8.20	5.95	10.96	5.35	11.87
42	8.40	6.12	11.18	5.51	12.11
43	8.60	6.29	11.41	5.68	12.34
44	8.80	6.47	11.63	5.84	12.57
45	9.00	6.64	11.86	6.01	12.80
46	9.20	6.81	12.08	6.17	13.03
47	9.40	6.99	12.30	6.34	13.26
48	9.60	7.16	12.53	6.51	13.49
49	9.80	7.34	12.75	6.67	13.72
50	10.00	7.51	12.97	6.84	13.95
60	12.00	9.28	15.18	8.53	16.21
70	14.00	11.08	17.35	10.26	18.44
80	16.00	12.90	19.51	12.02	20.65
90	18.00	14.73	21.65	13.80	22.83
100	20.00	16.58	23.78	15.60	24.99
125	25.00	21.26	29.04	20.17	30.31
150	30.00	26.01	34.23	24.83	35.55
175	35.00	30.82	39.36	29.57	40.72
200	40.00	35.68	44.44	34.38	45.82
225	45.00	40.58	49.48	39.24	50.86
250	50.00	45.53	54.47	44.16	55.84

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
	N = 600				
0	0.00	0.00	0.61	0.00	0.88
1	0.17	0.00	0.93	0.00	1.23
2	0.33	0.04	1.20	0.02	1.54
3	0.50	0.10	1.45	0.06	1.82
4	0.67	0.18	1.70	0.11	2.08
5	0.83	0.27	1.93	0.18	2.34
6	1.00	0.37	2.16	0.26	2.59
7	1.17	0.47	2.39	0.34	2.83
8	1.33	0.58	2.61	0.43	3.07
9	1.50	0.69	2.83	0.52	3.30
10	1.67	0.80	3.04	0.62	3.53
11	1.83	0.92	3.26	0.72	3.76
12	2.00	1.04	3.47	0.83	3.98
13	2.17	1.16	3.68	0.94	4.21
14	2.33	1.28	3.88	1.04	4.42
15	2.50	1.41	4.09	1.16	4.64
16	2.67	1.53	4.29	1.27	4.86
17	2.83	1.66	4.50	1.38	5.07
18	3.00	1.79	4.70	1.50	5.29
19	3.17	1.92	4.90	1.62	5.50
20	3.33	2.05	5.10	1.74	5.71
21	3.50	2.18	5.30	1.86	5.92
22	3.67	2.31	5.50	1.98	6.12
23	3.83	2.45	5.70	2.10	6.33
24	4.00	2.58	5.89	2.23	6.54
25	4.17	2.71	6.09	2.35	6.74
26	4.33	2.85	6.29	2.48	6.95
27	4.50	2.99	6.48	2.60	7.15
28	4.67	3.12	6.67	2.73	7.35
29	4.83	3.26	6.87	2.86	7.55
30	5.00	3.40	7.06	2.99	7.76
31	5.17	3.54	7.25	3.12	7.96
32	5.33	3.68	7.45	3.25	8.16
33	5.50	3.82	7.64	3.38	8.35
34	5.67	3.96	7.83	3.51	8.55
35	5.83	4.10	8.02	3.64	8.75
36	6.00	4.24	8.21	3.78	8.95
37	6.17	4.38	8.40	3.91	9.15
38	6.33	4.52	8.59	4.04	9.34
39	6.50	4.66	8.78	4.18	9.54
40	6.67	4.81	8.97	4.31	9.73
41	6.83	4.95	9.16	4.45	9.93
42	7.00	5.09	9.34	4.58	10.12
43	7.17	5.23	9.53	4.72	10.32
44	7.33	5.38	9.72	4.86	10.51
45	7.50	5.52	9.91	4.99	10.70
46	7.67	5.67	10.09	5.13	10.90
47	7.83	5.81	10.28	5.27	11.09
48	8.00	5.96	10.47	5.41	11.28
49	8.17	6.10	10.65	5.55	11.47
50	8.33	6.25	10.84	5.68	11.67
60	10.00	7.72	12.68	7.09	13.56
70	11.67	9.21	14.51	8.52	15.43
80	13.33	10.72	16.32	9.98	17.28
90	15.00	12.24	18.11	11.45	19.12
100	16.67	13.77	19.89	12.94	20.93
125	20.83	17.65	24.31	16.73	25.41
150	25.00	21.58	28.67	20.58	29.83
175	29.17	25.56	32.98	24.49	34.18
200	33.33	29.57	37.26	28.44	38.49
225	37.50	33.61	41.51	32.44	42.76
250	41.67	37.69	45.73	36.48	46.98
275	45.83	41.79	49.92	40.56	51.17
300	50.00	45.92	54.08	44.68	55.32

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
	N = 700				
0	0.00	0.00	0.53	0.00	0.75
1	0.14	0.00	0.79	0.00	1.06
2	0.29	0.03	1.03	0.01	1.32
3	0.43	0.09	1.25	0.05	1.56
4	0.57	0.16	1.46	0.10	1.79
5	0.71	0.23	1.66	0.15	2.01
6	0.86	0.32	1.86	0.22	2.22
7	1.00	0.40	2.05	0.29	2.43
8	1.14	0.49	2.24	0.37	2.63
9	1.29	0.59	2.43	0.45	2.83
10	1.43	0.69	2.61	0.53	3.03
11	1.57	0.79	2.79	0.62	3.23
12	1.71	0.89	2.98	0.71	3.42
13	1.86	0.99	3.15	0.80	3.61
14	2.00	1.10	3.33	0.89	3.80
15	2.14	1.20	3.51	0.99	3.99
16	2.29	1.31	3.69	1.09	4.17
17	2.43	1.42	3.86	1.19	4.36
18	2.57	1.53	4.03	1.29	4.54
19	2.71	1.64	4.21	1.39	4.72
20	2.86	1.75	4.38	1.49	4.90
21	3.00	1.87	4.55	1.59	5.08
22	3.14	1.98	4.72	1.70	5.26
23	3.29	2.09	4.89	1.80	5.44
24	3.43	2.21	5.06	1.91	5.61
25	3.57	2.32	5.23	2.01	5.79
26	3.71	2.44	5.40	2.12	5.97
27	3.86	2.56	5.56	2.23	6.14
28	4.00	2.67	5.73	2.34	6.31
29	4.14	2.79	5.90	2.45	6.49
30	4.29	2.91	6.06	2.56	6.66
31	4.43	3.03	6.23	2.67	6.83
32	4.57	3.15	6.39	2.78	7.01
33	4.71	3.27	6.56	2.89	7.18
34	4.86	3.39	6.72	3.01	7.35
35	5.00	3.51	6.89	3.12	7.52
36	5.14	3.63	7.05	3.23	7.69
37	5.29	3.75	7.21	3.35	7.86
38	5.43	3.87	7.38	3.46	8.03
39	5.57	3.99	7.54	3.58	8.19
40	5.71	4.11	7.70	3.69	8.36
41	5.86	4.24	7.86	3.81	8.53
42	6.00	4.36	8.02	3.92	8.70
43	6.14	4.48	8.19	4.04	8.86
44	6.29	4.60	8.35	4.16	9.03
45	6.43	4.73	8.51	4.27	9.20
46	6.57	4.85	8.67	4.39	9.36
47	6.71	4.97	8.83	4.51	9.53
48	6.86	5.10	8.99	4.63	9.70
49	7.00	5.22	9.15	4.74	9.86
50	7.14	5.35	9.31	4.86	10.03
60	8.57	6.60	10.90	6.06	11.66
70	10.00	7.88	12.47	7.29	13.27
80	11.43	9.17	14.02	8.53	14.86
90	12.86	10.47	15.57	9.79	16.44
100	14.29	11.78	17.10	11.06	18.01
125	17.86	15.09	20.90	14.29	21.87
150	21.43	18.44	24.66	17.57	25.68
175	25.00	21.83	28.38	20.90	29.45
200	28.57	25.25	32.07	24.26	33.18
225	32.14	28.69	35.74	27.66	36.87
250	35.71	32.16	39.39	31.09	40.54
275	39.29	35.65	43.01	34.55	44.17
300	42.86	39.16	46.62	38.03	47.78
325	46.43	42.68	50.20	41.54	51.37
350	50.00	46.23	53.77	45.07	54.93

Confidence intervals (%) for the binomial distribution (N = 600-1000)

See Appendixes B & C for the method of calculation and how to interpolate or extrapolate for limits not contained in the table

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
	N = 800				
0	0.00	0.00	0.46	0.00	0.66
1	0.13	0.00	0.69	0.00	0.93
2	0.25	0.03	0.90	0.01	1.15
3	0.38	0.08	1.09	0.04	1.37
4	0.50	0.14	1.28	0.08	1.57
5	0.63	0.20	1.45	0.13	1.76
6	0.75	0.28	1.63	0.19	1.95
7	0.88	0.35	1.79	0.26	2.13
8	1.00	0.43	1.96	0.32	2.31
9	1.13	0.52	2.12	0.39	2.48
10	1.25	0.60	2.29	0.47	2.66
11	1.38	0.69	2.45	0.54	2.83
12	1.50	0.78	2.61	0.62	3.00
13	1.63	0.87	2.76	0.70	3.16
14	1.75	0.96	2.92	0.78	3.33
15	1.88	1.05	3.07	0.87	3.49
16	2.00	1.15	3.23	0.95	3.65
17	2.13	1.24	3.38	1.04	3.82
18	2.25	1.34	3.53	1.12	3.98
19	2.38	1.44	3.68	1.21	4.14
20	2.50	1.53	3.83	1.30	4.29
21	2.63	1.63	3.98	1.39	4.45
22	2.75	1.73	4.13	1.48	4.61
23	2.88	1.83	4.28	1.57	4.76
24	3.00	1.93	4.43	1.67	4.92
25	3.13	2.03	4.58	1.76	5.07
26	3.25	2.13	4.73	1.85	5.23
27	3.38	2.24	4.87	1.95	5.38
28	3.50	2.34	5.02	2.04	5.53
29	3.63	2.44	5.16	2.14	5.69
30	3.75	2.54	5.31	2.24	5.84
31	3.88	2.65	5.46	2.33	5.99
32	4.00	2.75	5.60	2.43	6.14
33	4.13	2.86	5.74	2.53	6.29
34	4.25	2.96	5.89	2.63	6.44
35	4.38	3.07	6.03	2.73	6.59
36	4.50	3.17	6.18	2.83	6.74
37	4.63	3.28	6.32	2.92	6.89
38	4.75	3.38	6.46	3.02	7.03
39	4.88	3.49	6.60	3.12	7.18
40	5.00	3.60	6.75	3.23	7.33
41	5.13	3.70	6.89	3.33	7.48
42	5.25	3.81	7.03	3.43	7.62
43	5.38	3.92	7.17	3.53	7.77
44	5.50	4.02	7.31	3.63	7.92
45	5.63	4.13	7.45	3.73	8.06
46	5.75	4.24	7.60	3.84	8.21
47	5.88	4.35	7.74	3.94	8.35
48	6.00	4.46	7.88	4.04	8.50
49	6.13	4.57	8.02	4.15	8.64
50	6.25	4.67	8.16	4.25	8.79
60	7.50	5.77	9.55	5.30	10.22
70	8.75	6.88	10.93	6.37	11.64
80	10.00	8.01	12.29	7.45	13.04
90	11.25	9.14	13.65	8.55	14.42
100	12.50	10.29	14.99	9.66	15.80
125	15.63	13.18	18.33	12.47	19.20
150	18.75	16.10	21.63	15.33	22.55
175	21.88	19.06	24.90	18.23	25.87
200	25.00	22.03	28.15	21.15	29.15
225	28.13	25.03	31.38	24.11	32.41
250	31.25	28.05	34.59	27.09	35.64
275	34.38	31.08	37.78	30.09	38.85
300	37.50	34.13	40.96	33.11	42.04
325	40.63	37.20	44.12	36.16	45.21
350	43.75	40.28	47.27	39.22	48.36
375	46.88	43.37	50.40	42.30	51.49
400	50.00	46.48	53.52	45.40	54.60

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
	N = 900				
0	0.00	0.00	0.41	0.00	0.59
1	0.11	0.00	0.62	0.00	0.82
2	0.22	0.03	0.80	0.01	1.03
3	0.33	0.07	0.97	0.04	1.21
4	0.44	0.12	1.13	0.07	1.39
5	0.56	0.18	1.29	0.12	1.56
6	0.67	0.25	1.45	0.17	1.73
7	0.78	0.31	1.60	0.23	1.89
8	0.89	0.38	1.74	0.29	2.05
9	1.00	0.46	1.89	0.35	2.21
10	1.11	0.53	2.03	0.41	2.36
11	1.22	0.61	2.18	0.48	2.51
12	1.33	0.69	2.32	0.55	2.66
13	1.44	0.77	2.46	0.62	2.81
14	1.56	0.85	2.60	0.69	2.96
15	1.67	0.94	2.73	0.77	3.11
16	1.78	1.02	2.87	0.84	3.25
17	1.89	1.10	3.01	0.92	3.39
18	2.00	1.19	3.14	1.00	3.54
19	2.11	1.28	3.28	1.08	3.68
20	2.22	1.36	3.41	1.16	3.82
21	2.33	1.45	3.54	1.24	3.96
22	2.44	1.54	3.68	1.32	4.10
23	2.56	1.63	3.81	1.40	4.24
24	2.67	1.72	3.94	1.48	4.38
25	2.78	1.81	4.07	1.56	4.52
26	2.89	1.90	4.20	1.65	4.65
27	3.00	1.99	4.33	1.73	4.79
28	3.11	2.08	4.47	1.82	4.92
29	3.22	2.17	4.60	1.90	5.06
30	3.33	2.26	4.72	1.99	5.20
31	3.44	2.35	4.85	2.07	5.33
32	3.56	2.44	4.98	2.16	5.46
33	3.67	2.54	5.11	2.25	5.60
34	3.78	2.63	5.24	2.33	5.73
35	3.89	2.72	5.37	2.42	5.86
36	4.00	2.82	5.49	2.51	6.00
37	4.11	2.91	5.62	2.60	6.13
38	4.22	3.00	5.75	2.69	6.26
39	4.33	3.10	5.88	2.78	6.39
40	4.44	3.19	6.00	2.86	6.52
41	4.56	3.29	6.13	2.95	6.66
42	4.67	3.38	6.26	3.04	6.79
43	4.78	3.48	6.38	3.13	6.92
44	4.89	3.57	6.51	3.22	7.05
45	5.00	3.67	6.63	3.32	7.18
46	5.11	3.77	6.76	3.41	7.31
47	5.22	3.86	6.88	3.50	7.44
48	5.33	3.96	7.01	3.59	7.57
49	5.44	4.05	7.13	3.68	7.70
50	5.56	4.15	7.26	3.77	7.82
60	6.67	5.13	8.50	4.70	9.10
70	7.78	6.11	9.72	5.65	10.36
80	8.89	7.11	10.94	6.61	11.61
90	10.00	8.12	12.15	7.59	12.85
100	11.11	9.13	13.35	8.57	14.07
125	13.89	11.70	16.32	11.06	17.11
150	16.67	14.29	19.27	13.60	20.10
175	19.44	16.91	22.18	16.16	23.06
200	22.22	19.55	25.08	18.75	25.99
225	25.00	22.20	27.96	21.37	28.90
250	27.78	24.87	30.83	24.00	31.79
275	30.56	27.56	33.68	26.66	34.66
300	33.33	30.26	36.52	29.33	37.52
325	36.11	32.97	39.35	32.01	40.36
350	38.89	35.69	42.16	34.72	43.18
375	41.67	38.42	44.97	37.43	45.99
400	44.44	41.17	47.76	40.16	48.79
425	47.22	43.92	50.54	42.90	51.57
450	50.00	46.68	53.32	45.66	54.34

r	p $\left(\frac{r}{N}\times 100\right)$	Confidence intervals 95% Lower	95% Upper	99% Lower	99% Upper
	N = 1000				
0	0.00	0.00	0.37	0.00	0.53
1	0.10	0.00	0.56	0.00	0.74
2	0.20	0.02	0.72	0.01	0.92
3	0.30	0.06	0.87	0.03	1.09
4	0.40	0.11	1.02	0.07	1.25
5	0.50	0.16	1.16	0.11	1.41
6	0.60	0.22	1.30	0.15	1.56
7	0.70	0.28	1.44	0.20	1.70
8	0.80	0.35	1.57	0.26	1.85
9	0.90	0.41	1.70	0.31	1.99
10	1.00	0.48	1.83	0.37	2.13
11	1.10	0.55	1.96	0.43	2.26
12	1.20	0.62	2.09	0.50	2.40
13	1.30	0.69	2.21	0.56	2.53
14	1.40	0.77	2.34	0.63	2.67
15	1.50	0.84	2.46	0.69	2.80
16	1.60	0.92	2.59	0.76	2.93
17	1.70	0.99	2.71	0.83	3.06
18	1.80	1.07	2.83	0.90	3.19
19	1.90	1.15	2.95	0.97	3.31
20	2.00	1.23	3.07	1.04	3.44
21	2.10	1.30	3.19	1.11	3.57
22	2.20	1.38	3.31	1.18	3.69
23	2.30	1.46	3.43	1.26	3.82
24	2.40	1.54	3.55	1.33	3.94
25	2.50	1.62	3.67	1.41	4.07
26	2.60	1.71	3.79	1.48	4.19
27	2.70	1.79	3.90	1.56	4.31
28	2.80	1.87	4.02	1.63	4.44
29	2.90	1.95	4.14	1.71	4.56
30	3.00	2.03	4.26	1.79	4.68
31	3.10	2.12	4.37	1.86	4.80
32	3.20	2.20	4.49	1.94	4.92
33	3.30	2.28	4.60	2.02	5.04
34	3.40	2.37	4.72	2.10	5.16
35	3.50	2.45	4.83	2.18	5.28
36	3.60	2.53	4.95	2.26	5.40
37	3.70	2.62	5.06	2.34	5.52
38	3.80	2.70	5.18	2.42	5.64
39	3.90	2.79	5.29	2.50	5.76
40	4.00	2.87	5.41	2.58	5.88
41	4.10	2.96	5.52	2.66	6.00
42	4.20	3.04	5.64	2.74	6.11
43	4.30	3.13	5.75	2.82	6.23
44	4.40	3.21	5.86	2.90	6.35
45	4.50	3.30	5.98	2.98	6.47
46	4.60	3.39	6.09	3.06	6.58
47	4.70	3.47	6.20	3.15	6.70
48	4.80	3.56	6.31	3.23	6.82
49	4.90	3.65	6.43	3.31	6.93
50	5.00	3.73	6.54	3.39	7.05
60	6.00	4.61	7.66	4.23	8.20
70	7.00	5.50	8.76	5.08	9.34
80	8.00	6.39	9.86	5.94	10.47
90	9.00	7.30	10.95	6.82	11.58
100	10.00	8.21	12.03	7.70	12.69
125	12.50	10.51	14.71	9.94	15.42
150	15.00	12.84	17.37	12.21	18.13
175	17.50	15.19	20.00	14.52	20.80
200	20.00	17.56	22.62	16.84	23.45
225	22.50	19.95	25.22	19.19	26.08
250	25.00	22.34	27.80	21.55	28.69
275	27.50	24.75	30.38	23.93	31.29
300	30.00	27.17	32.95	26.32	33.87
325	32.50	29.60	35.50	28.73	36.44
350	35.00	32.04	38.05	31.14	39.00
375	37.50	34.49	40.58	33.57	41.55
400	40.00	36.95	43.11	36.02	44.08
425	42.50	39.41	45.63	38.47	46.61
450	45.00	41.89	48.14	40.93	49.12
475	47.50	44.37	50.65	43.40	51.62
500	50.00	46.85	53.15	45.89	54.11

Appendix 2
How to calculate exact binomial confidence limits

Exact binomial confidence limits are calculated as follows:[29]

$$if\ \ r=0\ \ LCL=0;\ \ UCL=1-10^{\left(\frac{\log_{10}\left(\frac{1-CL}{2}\right)}{N}\right)}$$

$$if\ \ r=N\ \ LCL=10^{\left(\frac{\log_{10}\left(\frac{1-CL}{2}\right)}{N}\right)};\ \ UCL=1$$

$$if\ \ 0<r<N\ LCL=1-BETAINV\left(\left(\frac{1+CL}{2}\right),N-r+1,r\right);\ \ UCL=BETAINV\left(\left(\frac{1+CL}{2}\right),r+1,N-r\right)$$

where CL = confidence level as a proportion (for a 95% confidence level CL should be expressed as 0.95)

LCL & UCL are the lower and upper confidence limits respectively

r = the number of successes in N trials

---

[29] Daly S. Simple SAS macros for the calculation of exact binomial and Poisson confidence limits. *Comput. Biol. Med.*, **22**, 351-361, 1992

Appendix 3
Calculating binomial confidence limits not contained in Appendix 1

The following methods enable exact binomial confidence limits to be interpolated or extrapolated from the tables in Appendix 1:

1. If r is greater than the tabulated values in a particular column of N

Look up the corresponding values of the confidence intervals for *N-r*. The required confidence intervals are calculated as follows:

$$LCI = 100\text{-}UCI_{(N-r)} \qquad \text{Equation 1}$$

$$UCI = 100\text{-}LCI_{(N-r)} \qquad \text{Equation 2}$$

where: LCI = lower confidence level and
UCI = upper confidence interval

For example, to interpolate the 95% confidence intervals where N = 21 and r = 15, look up the corresponding values of the confidence intervals for N-r = 21-15 = 6. LCI = 100-52.18 = 47.82, UCI = 100-11.28 = 88.72.

2. If r is lies between two values in a particular column of N ($r_1 < r < r_2$)

Calculate $p_r = \frac{r}{N} \times 100$, then look up the corresponding p values for r_1 and r_2. The required confidence intervals can then be calculated by applying the following formula:

$$CI_{r1} + \frac{p_r - p_{r1}}{p_{r2} - p_{r1}} \times \left(CI_{r2} - CI_{r1}\right) \qquad \text{Equation 3}$$

For example, to interpolate the 95% confidence intervals where N = 200 and r = 85, calculate $p_r = \frac{r}{N} \times 100 = \frac{85}{200} = 42.5$, look up the corresponding p values and confidence intervals for 80 and 90 and insert them into the formula:

$$LCI = 33.15 + \frac{42.50 - 40.00}{45.00 - 40.00} \times \left(37.98 - 33.15\right) = 35.56$$

$$UCI = 47.15 + \frac{42.50 - 40.00}{45.00 - 40.00} \times \left(52.18 - 47.15\right) = 49.67$$

3. If N lies between two columns of N ($N_1 < N < N_2$)

Calculate $p_r = \frac{r}{N} \times 100$, then look up the corresponding confidence limits for p_1 and p_2 where $p_1 < p_r < p_2$ for N_1 and N_2 respectively. Next, determine the interpolated confidence limits for p_r for both N_1 and N_2 as in Equation 3. Finally, the required confidence intervals can be calculated by applying the following formula:

$$CI_{pr\ lower1} + \frac{N - N_1}{N_2 - N_1} \times \left(CI_{pr\ upper} - CI_{pr\ lower}\right) \qquad \text{Equation 4}$$

For example, to interpolate the 95% confidence intervals for N = 270 and r = 22 where N lies between N_1 = 200 and N_2 = 300, firstly calculate $p_r = \frac{r}{N} \times 100 = 8.15$ then look up the corresponding p values for p_1 and p_2 where $p_1 < p_r < p_2$ for N_1 and N_2 respectively and apply Equation 3:

$N2$ = 300 interpolated

$$LCI = 5.19 + \frac{8.15 - 8.00}{8.33 - 8.00} \times (5.47 - 5.19) = 5.32 \qquad UC1 = 11.67 + \frac{8.15 - 8.00}{8.33 - 8.00} \times (12.06 - 11.67) = 11.85$$

N_1 = 200 interpolated

$$LCI = 4.64 + \frac{8.15 - 8.00}{8.50 - 8.00} \times (5.03 - 4.64) = 4.76 \qquad UC1 = 12.67 + \frac{8.15 - 8.00}{8.50 - 8.00} \times (13.26 - 12.67) = 12.85$$

Insert the interpolated confidence limits into Equation 4:

$$LCI = 4.76 + \frac{270 - 200}{300 - 200} \times (5.32 - 4.76) = 5.15 \qquad UCI = 12.85 + \frac{270 - 200}{300 - 200} \times (11.85 - 12.85) = 12.15$$

4. If *N* is greater than 1,000

Calculate $p_r = \frac{r}{N} \times 100$, then look up the corresponding confidence limits for p_r and apply the following formulae:

$$LCI = p_r - (p_r - LCI_{pr}) \times \sqrt{\frac{1000}{N}} \qquad \text{Equation 5}$$

$$UCI = p_r + (UCI_{pr} - p_r) \times \sqrt{\frac{1000}{N}} \qquad \text{Equation 6}$$

For example, to extrapolate the 95% confidence intervals for N = 3000 and r = 54 calculate $p_r = \frac{r}{N} \times 100 = 1.80$ then look up the lower and upper confidence limits for p_r = 1.80 and insert them into Equations 5 and 6 respectively:

$$LCI = 1.80 - (1.80 - 1.07) \times \sqrt{\frac{1000}{3000}} = 1.38 \qquad UCI = 1.80 + (2.84 - 1.80) \times \sqrt{\frac{1000}{3000}} = 2.4$$

Bibliography

Covello V.T. & Merkhofer M.W. (1993). – Risk assessment methods. Approaches for assessing health and environmental risks. Plenum Press, New York.

Cullen A.C. & Frey H.C. (1999). – Probabilistic techniques in exposure assessment. A handbook for dealing with uncertainty in models and inputs. Plenum Press, New York.

Daly S. (1992). – Simple SAS macros for the calculation of exact binomial and Poisson confidence limits. *Comput. Biol. Med.*, **22**, 351-361.

Martin S.W., Meek A.H. & Willeberg P. (1987). – Veterinary epidemiology. Principles and methods. Iowa State university Press, Ames.

Merkhofer M.W. (1987). – Quantifying judgmental uncertainty: methodology, experiences and insights. *IEEE Transactions on Systems, Man and Cyberkinetics*, **17**, 741-752.

Snedecor G.W. & Cochran W.G. (1967). – Statistical methods. Oxford & IBH Publishing Co., New Delhi.

Thrushfield M. (1997). – Veterinary epidemiology. Blackwell Science Ltd., United Kingdom.

Vose D. (1997). – Risk analysis in relation to the importation and exportation of animal products. *In* Contamination of animal products: prevention and risks for animal health (P. Sutmoller, ed.). *Rev. sci. tech. Off. int. Epiz.*, **16** (1), 17-29.

Vose D. (2000). – Risk analysis, a quantitative guide. John Wiley and Sons, Chichester.

Index