

## Shiny rrisk user guide



Shiny rrisk user guide



# Shiny rrisk user guide

Your quick start to perform probabilistic risk modelling

## Getting started

Read here how to manage the data file that contains the complete documentation and executable code of your model.

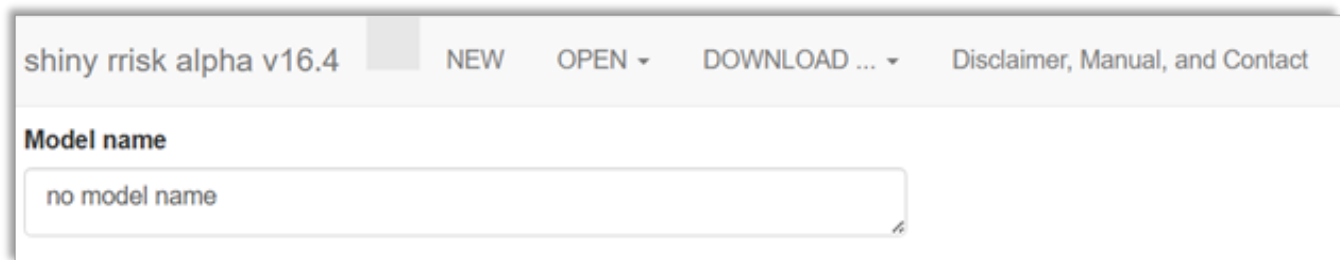


Figure 1: Main screen, top row of shiny rrisk.

The top row of the app displays the version of the app and provides basic functionality for managing models ([Figure 1](#)).

<b>NEW</b>	Check here if you wish to create a new model and enter a Model name.
<b>OPEN</b>	Open a locally saved model or one of the demo models implemented in the tool. The extension of model files is “risk”.
<b>SAVE</b>	[@Robert: consider change of button name] Save the model, results or the model report to your local system.
<b>ABOUT</b>	[@Robert: consider change of button name] Check out here for important information about the app .

### Avoid loss of your data

The app runs on a server provided by the German Federal institute for Risk Assessment (BfR). Your data is not accessible to BfR or any other party and irretrievably lost upon termination of the browser session or interrupt of the server connection. Save your model locally after each important modification.

### Good practice

- We recommend using “<agent>in<commodity>” as standard format for naming models.
- We recommend discussing the model concept with stakeholders (see tip below) in order to clarify the task and available data and resources.

[Read more](#)

## First things first

Before developing a mathematical or probabilistic model it is advisable to clarify some important points.

- What is the question you wish to answer?
- What is the starting point and final outcome of the scenario you wish to model?
- What are the relevant/influential events and pathways?
- What are the relevant populations and subgroups to be considered?
- If exposure factors and disease statuses are involved, are there accepted case definitions available?

## Building blocks of a model

A probabilistic risk model typically represents a series of events or situations that can be linked together in a logical or chronological order, whereby the final event represents the outcome of interest. For example, consider the hypothetical shipment of one race horse from Argentina to Germany. The task is to quantify the probability of the horse to be infected with equine infectious anemia (EIA, see [FLI, 2017](#), accessed 9.5.2025). Note that this example is for illustration only. Each building block of the model requires clarifications and definitions which are tailored to the specific problem at hand. Some of the clarifications can only be provided by the customer (risk manager) and others by domain experts risk assessors. For larger projects, stakeholders should be involved in this process.

- **Risk question:** defines the outcome of interest; here the quantification of the probability that the introduced horse is infected with EIA virus.
- **Scope:** defines what is included or excluded from the assessment; here economic and epidemiological consequences are beyond the scope.
- **Pathway and scenario:** define all relevant steps and events; here the selection of the horse, legal and sanitary transport requirements, transport, border health checks should be considered.
- **Model parameters (denoted “nodes” in shiny rrisk):** define all input and output quantities required in the model; here the minimum set of parameters include the prevalence of EIA in Argentina, the probability of infection at boarding and the probability of infection at introduction of the animal after de-boarding formalities. Additional parameters may apply.
- **Components of the model:** may structure the parameters into logical groups which makes it easier to discuss the model with individual domain experts and provides a useful structure of the report; typical components for a larger import risk models include the entry assessment, exposure assessment, consequence assessment and risk estimation (see [OIE, 2018](#), accessed 9 May 2025).

## Visualisation

It is recommended to sketch a visual representation of the conceptual model. The flowchart for the simple example above ([Figure 2](#)) may support the discussion about required data and additional nodes such as for diagnostic testing.

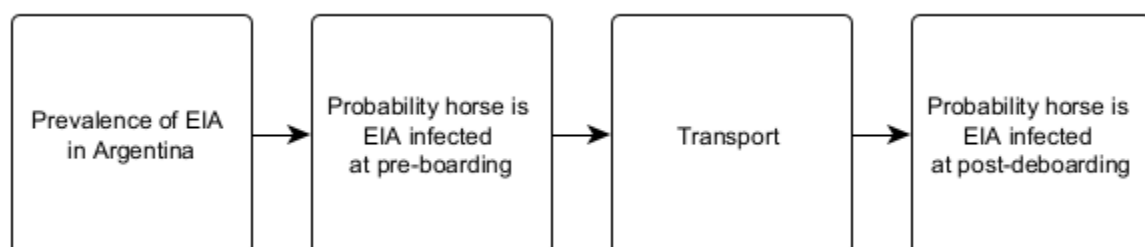


Figure 2: Example for a conceptual model (made with yEd).

## Model creation

At the end of this section, you will be able to build a model from scratch.

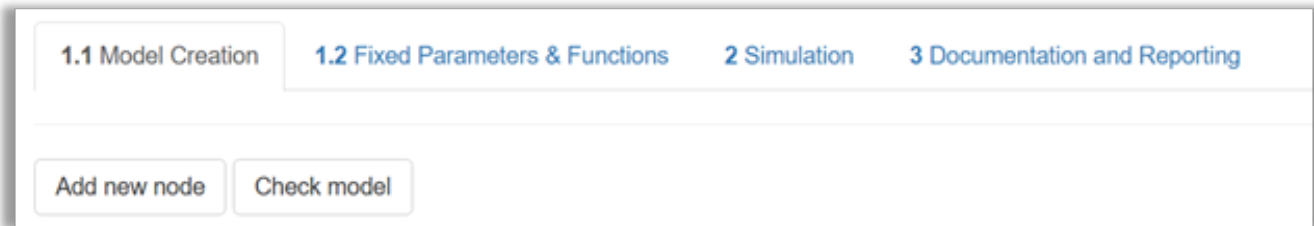


Figure 3: Main screen, second row of shiny rrisk.

A model in shiny rrisk consists of so-called nodes. The user may also define fixed parameters and functions available for the definition of nodes (see @FixedParametersFunctions).

**Add node**    [@Robert: consider change of button name] Create a new node and launch its verbal description and mathematical definition.

**Check model**    Verify that a simulation model can be built from the mathematical and numerical model input you have provided.

### Avoid loss of your data

Remember to save your model after adding each node.

### Good practice

- We recommend using “<agent>in<commodity>” as standard format for naming models.
- We recommend discussing the model concept with stakeholders (see tip below) in order to clarify the task and available data and resources.

### Read more

## First things first

Before developing a mathematical or probabilistic model it is advisable to clarify some important points.

- What is the question you wish to answer?
- What is the starting point and final outcome of the scenario you wish to model?
- What are the relevant/influential events and pathways?

- What are the relevant populations and subgroups to be considered?
- If exposure factors and disease statuses are involved, are there accepted case definitions available?

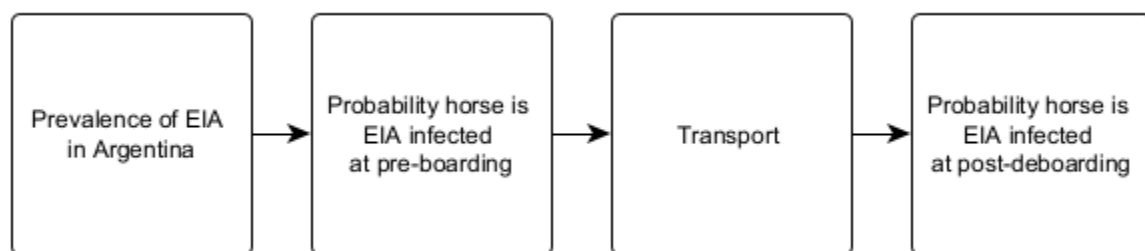
## Building blocks of a model

A probabilistic risk model typically represents a series of events or situations that can be linked together in a logical or chronological order, whereby the final event represents the outcome of interest. For example, consider the hypothetical shipment of one race horse from Argentina to Germany. The task is to quantify the probability of the horse to be infected with equine infectious anemia (EIA, see [FLI, 2017](#), accessed 9.5.2025). Note that this example is for illustration only. Each building block of the model requires clarifications and definitions which are tailored to the specific problem at hand. Some of the clarifications can only be provided by the customer (risk manager) and others by domain experts risk assessors. For larger projects, stakeholders should be involved in this process.

- **Risk question:** defines the outcome of interest; here the quantification of the probability that the introduced horse is infected with EIA virus.
- **Scope:** defines what is included or excluded from the assessment; here economic and epidemiological consequences are beyond the scope.
- **Pathway and scenario:** define all relevant steps and events; here the selection of the horse, legal and sanitary transport requirements, transport, border health checks should be considered.
- **Model parameters (denoted “nodes” in shiny rrisk):** define all input and output quantities required in the model; here the minimum set of parameters include the prevalence of EIA in Argentina, the probability of infection at boarding and the probability of infection at introduction of the animal after de-boarding formalities. Additional parameters may apply.
- **Components of the model:** may structure the parameters into logical groups which makes it easier to discuss the model with individual domain experts and provides a useful structure of the report; typical components for a larger import risk models include the entry assessment, exposure assessment, consequence assessment and risk estimation (see [OIE, 2018](#), accessed 9 May 2025).

## Visualisation

It is recommended to sketch a visual representation of the conceptual model. The flowchart for the simple example above ([Figure 2](#)) may support the discussion about required data and additional nodes such as for diagnostic testing.



Example for a conceptual model (made with yEd).

## Verbal definition of a node

Read here how easily you will generate an integrated documentation of each node.

Figure 4: Descriptive part of node definition.

Upon launching [Add node](#), a window pops-up with several free-text fields.

<b>Name*</b>	[@Robert: consider change of button] Each node has to have a unique name starting with a letter. Numbers and underscore are allowed. An error message occurs upon missing or invalid entry.
<b>Unit</b>	The physical measurement unit should be entered here if applicable.
<b>Grouping</b>	[@Robert: consider change of button name] A grouping name may refer to a particular module of the model or step in the event tree. It is possible to define membership of one node to more than one group in which case the group names are comma-separated. The groups are visualised in the graphical representation of the model and are used to organise the model documentation.
<b>Source</b>	The source of information about the node should be provided according to good scientific practice. If a bibliographic reference is available it should be entered using the format “Author(s), Year (doi code)”.
<b>Description</b>	A verbal account of all relevant information required for documentation and review should be provided. This may include any underlying assumptions and uncertainties. The text may be structured using sub-headers (see example for markdown syntax below).
<b>*</b>	Required

**Avoid loss of your data**

Remember to save your model after adding each node.

**Example for markdown syntax (Alizarin red S in Eel)**

Markdown syntax is available in all free text fields in shiny risk to generate formatted output in the report file. Some examples are given below ([Figure 5](#)).

Syntax in free-text field	Result in formatted model report
# Header level 1	Header level 1 formatted according to style template
## Header level 2	Header level 2 formatted according to style template
### Header level 3	Header level 3 formatted according to style template
<b>**Bold**</b>	<b>Bold</b>
<i>*italics*</i>	<i>Italics</i>
<b><i>*** Bold italics***</i></b>	<b><i>Bold italics</i></b>
...	...

Figure 5: Selection of common markdown commands for use in free-text fields.

You can find more information at [RStudio, 2016](#) Also check out the documentation of the Alizarin red S in Eel model, which makes use of markdown in the description of the model and model parameters (e.g., bootstrapping).

**Good practice**

- Be consistent with the style of node names within your project. Short names are preferred for reasons of readability and aesthetics in tables and figures (e.g., vowels can be omitted to shorten a name).
- Provide units for all nodes. This allows checking the unit of the outcome quantity.
- Use a consistent logic to define groups names. These groups may, for example, represent the sections of risk assessment (hazard identification, hazard characterisation, exposure, risk characterisation), steps of an exposure pathway or modules of a microbial risk assessment.

**Read more**

- The rules for naming nodes in shiny risk are inherited from and ensure compatibility with the [software R](#).
- Node names may be changed later on if needed (see section on editing model). In this case shiny risk automatically updates all references in the mathematical expressions (not in the free-text fields) to maintain consistency.

## Mathematical definition of a node

You will understand how to define a node in shiny risk, which is the essential tasks in model creation.

Figure 6: Mathematical part of node definition.[@Robert: consider change to “Choose node defintion\*”]

Several options are available for defining a node, including user defined, parametric distribution, bootstrapping and EFSA EKE. The first option allows you to define a node using a mathematical expression ([Figure 7](#)).

## User defined expression

Figure 7: Choose user defined to enter a mathematical expression.

**Expression** [Robert: consider change button name] A mathematical expression is entered here that can be evaluated similar to a line of code in the [software R](#). At least one term of the expression is another (distribution) node, which is referred to as a “parent node”. Parent nodes that do not yet exist in the model are referred to as “implicit nodes” and will be generated automatically by shiny risk.

### Example for an expression using a parent node (Alizarin Red S in eel model)

In the [Alizarin Red S in eel](#) model the outcome  $Y$  (dietary intake of ARS per kg body weight and year through consumption of eel marked with ARS) is defined in terms of  $I$  (concentration of ARS in edible tissues at time of catch) and  $M$  (mean consumption of European eel per year per body weight). Thus,  $I$  and  $M$  are both parents nodes of  $Z$ . This example also illustrates the common case that the final outcome is defined in terms of parent nodes ([Figure 8](#)).



Figure 8: Expression for the outcome  $Y$  in the Alizarin Red S in eel model.

### Example for an expression using implicit nodes

Consider the simple model

$$A = B * C.$$

The user can start creating the model by defining the node  $A$ . As a result, shiny risk generates not only the node  $A$ , but also two implicit nodes  $B$  and  $C$  based on the mathematical expression for  $A$ . However, the model cannot be executed until all implicit nodes are correctly defined. The same model could also be implemented in reverse order, starting with the independent nodes  $B$  and  $C$ .

### Good practice

- Use the model graph (see later) to identify nodes that were implicitly generated based on a spelling mistake in the node definition.
- Code expressions may contain any function that is part of the R base package.

### Read more

The tool's ability to create implicit nodes allows users to define a model by starting with the result function and working backwards to define all parent nodes until the complete model is created. This offers maximum flexibility when building a model.

## Parametric distributions

Upon completing this section, you will master the use of probability distributions in shiny risk, a prerequisite for reflecting variability and uncertainty in your model.

Figure 9: Choose a distribution family.

Once you have selected [parametric distributions](#) you can choose from a number of discrete and continuous distribution families ([Figure 9](#)).

Choose parametric  
distribution

Select here the distribution family for the new node.

## Discrete distributions

The following five discrete distributions are available to generate a probabilistic node in shiny rrisk: discrete distributions: binomial, hypergeometric, Poisson, negative binomial and discrete.

## Continuous distributions

The following thirteen continuous distributions are available to generate a probabilistic node in shiny rrisk: uniform, beta, modified PERT, exponential, Gaussian, lognormal, Weibull, gamma, inverse gamma, shifted log-logistic, triangular, general and cumulative.

## Special distributions

Shiny rrisk provides two special distributions. The [sigma](#) distribution is .... Please complete. The [Furry process \(Only birth\)](#) is a stochastic process where the probability of an event occurring at a given time depends on the number of events that have already occurred and the rate of new events is proportional to the current population size. The documentation of these distribution is found in ... Javier's model if we include it in the demo package..

## Variability and uncertainty

The user defines whether a distribution characterises variability or uncertainty. Variability is a reflection of a feature that varies among individuals in a population, e.g., body weight of exposed children of a defined age group (see [HEV in liver sausages](#) example below). This should be distinguished from uncertainty, which is a quantification of the limited knowledge about a population statistic, which is an unknown single value, e.g. the true mean body weight of individuals in the defined age group and region. Read more about this topic below.

## Context-sensitive distribution parameters

Upon launching any of the distributions, a context-sensitive interface appears for completing the definition of required model parameters. This is illustrated using the the modified PERT distribution in the [HEV in liver sausages](#) example below.

## Distribution fitting

A shiny rrisk module for fitting distributions is currently being developed.

## Known parameter values versus parameter uncertainty

Distribution parameters can be entered as single numbers (see [HEV in liver sausages](#) example below). A single number (scalar value) implies that the true parameter value is exactly known. In this case, the distribution reflects variability under the simplifying assumption that both the distribution family as well as the distribution

reflects variability under the simplifying assumption that both the distribution family as well as the distribution parameters are exactly known. In many cases, however, an uncertainty distribution is required to reflect uncertainty about the parameter value(s) or even a functional dependency of the parameter on other nodes. The latter is illustrated with the [L.monocytogenes in cheese](#) example model.

### Lower and/or upper limit for truncated distribution

In shiny risk, many distributions can be defined in a truncated version by defining a [lower](#) and/or [upper](#) limit.

#### Example for parametric distribution (HEV in liver sausages)

The objective of the example model [HEV in liver sausages](#) is to estimate the exposure with Hepatitis E virus (HEV) via consumption of sausage products made from pork (see shiny risk model report for further details). The model contains the variable [Liver\\_prop](#) to describe the amount of liver as ingredient in liver sausages in percent. This continuous quantity reflects variability (sausages contain variable amounts of liver depending on recipes and other factors). The modellers have chosen a modified PERT distribution with minimum 10%, mode 30% and maximum 50% for parameterisation. Note that shape parameter of 4 in the modified PERT is equivalent to the standard 3-parameter PERT model.

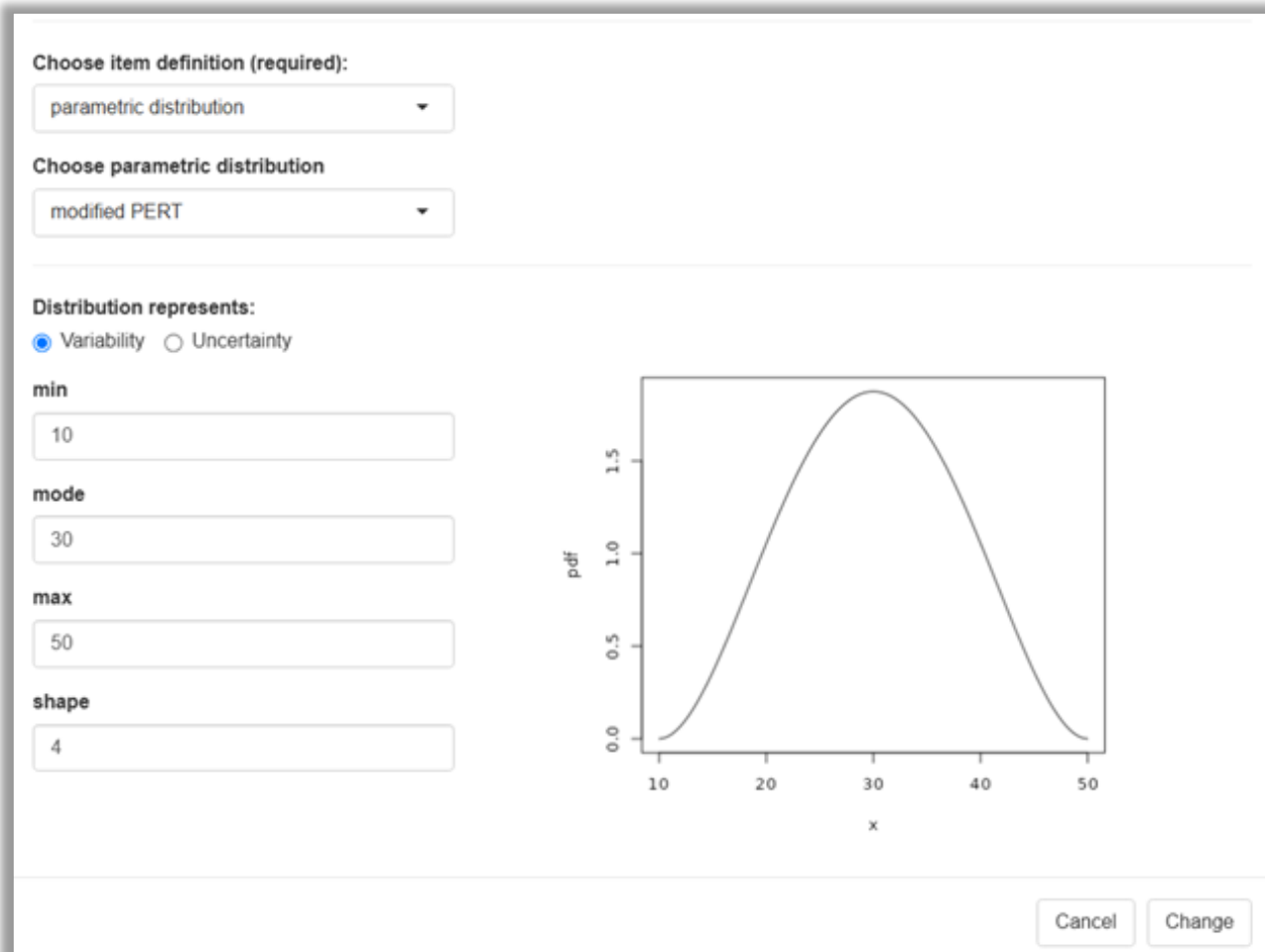
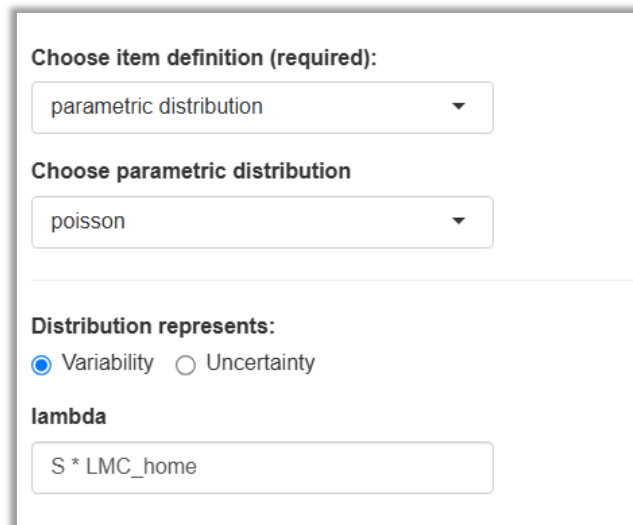


Figure 10: Modified Pert distribution model.

#### Example for a distribution parameter with functional dependency (L.monocytogenes in cheese)

The objective of the [L.monocytogenes in cheese](#) model is to estimate the risk of infection with *Listeria* (*L.*)

The objective of the **L.monocytogenes in cheese** model is to estimate the risk of infection with *Listeria* (L.) monocytogenes per meal. The model uses the variable dose of ingested L. monocytogenes per serving,  $D_{dist}$  modelled as Poisson distribution (Figure 11). The expression entered as distribution parameter  $\lambda$  uses as parent nodes the serving size  $S$  and the concentration of L. monocytogenes in the cheese at home  $LMC_{home}$ , both of which express variability as well. This results in a dispersion of dose  $D_{dist}$  much wider than what would be expected with a fixed value for the distribution parameter  $\lambda$ .



Choose item definition (required):

parametric distribution

Choose parametric distribution

poisson

Distribution represents:

☒ Variability ☐ Uncertainty

lambda

S \* LMC\_home

Figure 11: Poisson distribution model with functional dependency.

## Read more

### Quick guide for choosing appropriate distributions

Some text required here.

### Defining variability and uncertainty

“**Variability** reflects the fact that a variable is observed under different conditions. This generally refers to existing differences between individuals, and/or variation in time and space. Variability describes a property of the population. Variability in the population should be described but cannot be reduced. However, the variability in the data used for an assessment can be reduced by applying selection criteria (e. g. excluding individuals with specific traits). Stratification is another approach to reduce variability within the generated strata of the data. Changes over time occur on individual level (repeated observations, individual growth, changed behaviour or traits) as well as population level (population trends). The latter are often considered along with spatial factors. In case of changing exposure conditions of a population – e. g. due to (regional) changes in market supply over time – the variance of influential parameters might change as well. **Uncertainty** reflects the fact that the knowledge required for any step of the estimation process (problem formulation, scenario, model, parameters, calculations) is limited. Parameter uncertainty may be due to measurement errors at the individual level of observation and all sources of bias when selecting and aggregating observations into summary statistics for a given target population. The degree of uncertainty can be reduced on the basis of knowledge, at least in principle.” (Heinemeyer et al., 2022)

Note that uncertainty does not only apply to the level of a parameter (node) but can be related to problem formulation, scenario, model and calculations. In shiny risk, the user should describe those uncertainties in the free-text field of the

model description (see [Alizarin Red S in eel](#) model as an example).

## Using distributions to reflect variability and uncertainty

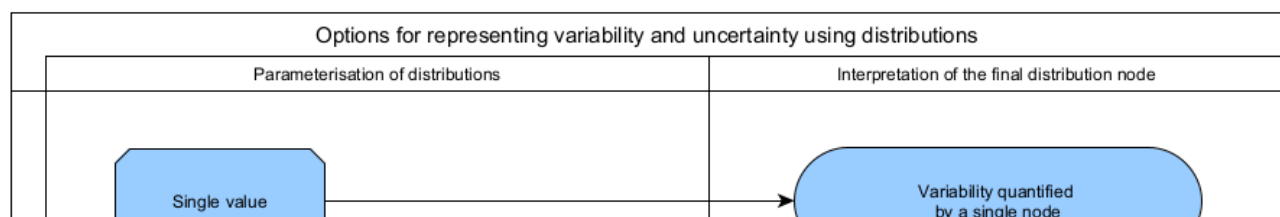
A distribution node can be defined by the user to represent the variability of a variable of interest, e.g. body weight. One typical situation is to parameterise the distribution using a single value for each distribution parameter, e.g., a single mean and a single standard deviation for the distribution of body weight obtained from observed data. The same principle applies for distributions with one parameter (e.g., Poisson) or distributions with three parameters as shown in the example above using the modPert ([Figure 10](#)). The variability of the quantity of interest is represented by one distribution node, which is parameterised with a single value for each distribution parameter. This is the standard situation for characterising variability (see top row shaded in blue in [Figure 12](#)).

More complex sources of variability may occur. For example, it may be known that a parameter of a distribution node for variability varies among sub-groups of the population. Typical examples include the litter effect in experimental animal studies or the cluster effect in observational studies. In this case the parent node for the distribution parameter represents variability at a higher aggregation level and “final” distribution node represents overdispersion, i.e. variability from various sources. Consider for example a Bernoulli distribution (for variability) with a binary outcome space (positive/negative), parameterised with a parameter for prevalence. It is a common finding in veterinary epidemiology that herd-level prevalences have a distribution. In such cases it may be possible to characterise the observed distribution of herd-level prevalences using a beta distribution and to use this beta as parent node to the Bernoulli.

Additional variability may also be due to structural sources. For example, the mean value of bacterial load may depend on temperature. In this case, the mean of the variability distribution can be modeled as a function of temperature. The factor of interest can be continuous (e.g., temperature) or discrete (e.g., age groups). In both cases, the independent factor is represented as parent node in the model and the final variability node represents variability due to structural (or functional) sources of variability.

The other motivation for using distributions in risk models is to represent parameter uncertainty. Assume you have a binomial distribution node in your model to represent variability of the random number of positive outcomes out of a fixed number of trials, given a defined outcome probability. A beta distribution for the latter would represent parameter uncertainty and act as parent node to the binomial distribution. The node `mean_log10_c0` from the [E.coli in beef](#) example is another illustration for an uncertainty parent node to a variability node. The bootstrap method provides a powerful solution for obtaining distributions for parameter uncertainty (see later). Independent of the technique used, the interpretation of the final distribution node is a quantification of variability accounting for parameter uncertainty.

Last but not least consider a node in your model that directly represents a parameter uncertainty. The classical example for this is prevalence. The prevalence is an unknown single value, i.e. only one value can be true at the same time. The statistical uncertainty in the estimation of prevalence is represented by a beta distribution with parameters  $k + 1$  and  $n - k + 1$ , where  $k$  and  $n$  denote the number of cases and sample size in estimation of prevalence. we don't have a beta in our example models... unbelievable;). Since prevalence is a common parameter in may risk models, its quantification of uncertainty using a single node is a typical situation in risk modelling (bottom row shaded blue in [Figure 12](#)).



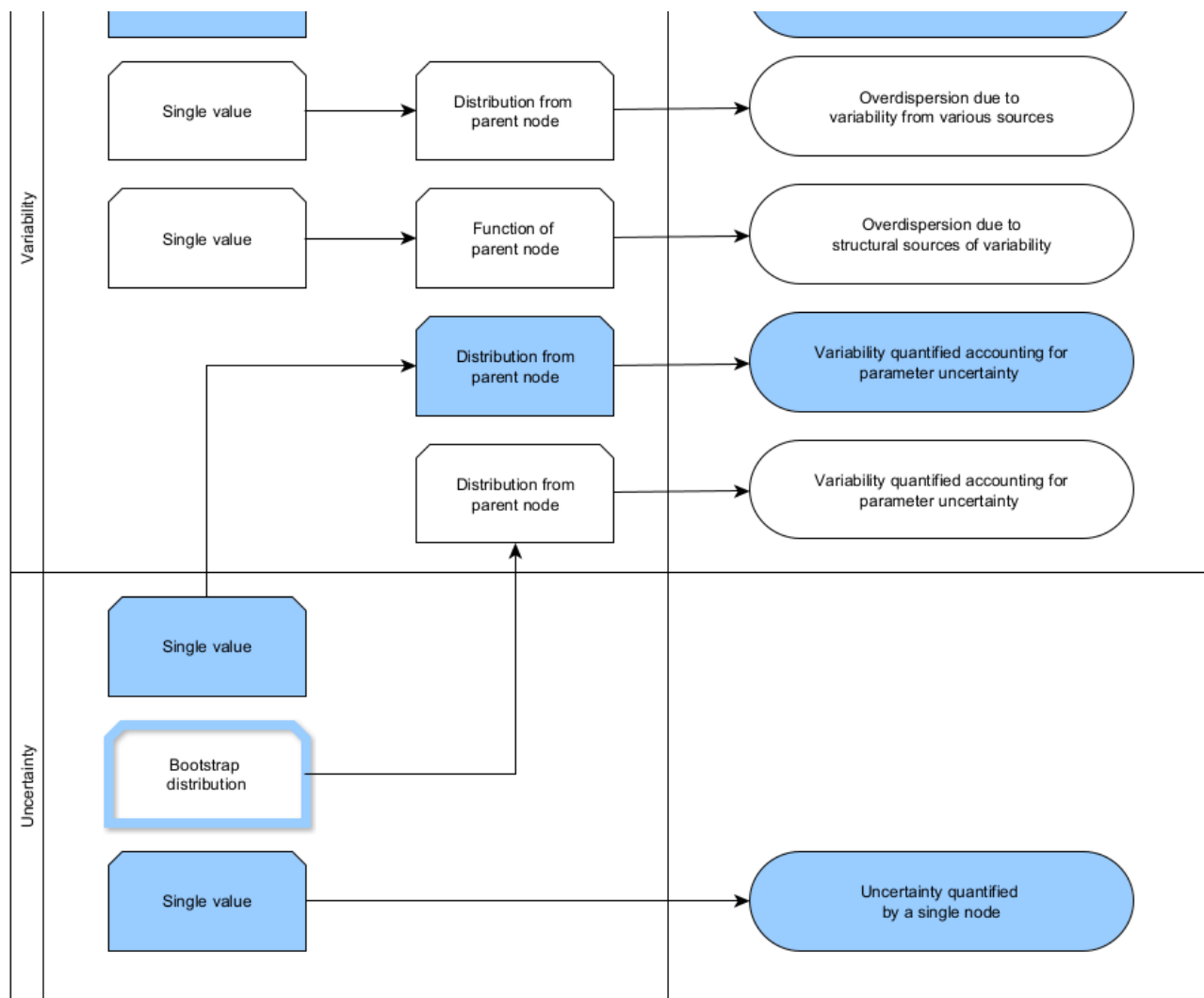


Figure 12: Parameterisations of distributions to reflect variability and uncertainty. Typical situations are shaded in blue. See text for more details.

Greiner M, Sanchez J, and who else  
contributes, 2025