

iDist Tutorial

This tutorial explains how the iDist online tool can be used to analyze individual patient data to estimate parameters and parametric distributions to be implemented in patient-level health economic models. A case study is used to do so and several methodological aspects to the tool are discussed.

Table of contents:

1. Explanation of the case study	1
2. Dataset structure and supported parameters	2
3. Visualization of the dataset	3
4. Defining the general variables	4
5. Defining the competing risk variables	4
6. Analyzing the dataset	6

iDist is powered by



website: www.personex.nl

e-mail: info@personex.nl

1. Explanation of the case study

iDist includes the option to use a test dataset to try out the tool without the need for uploading a custom dataset. This test dataset is generated based on an exemplary scenario, i.e. a hypothetical case study. The case study is about patients that might get a recurrence event after a surgical procedure. If a recurrence event occurs, the patient may experience major complications or end up in palliative care at one point. A graphical illustration of this clinical pathway is presented below (Figure 1).

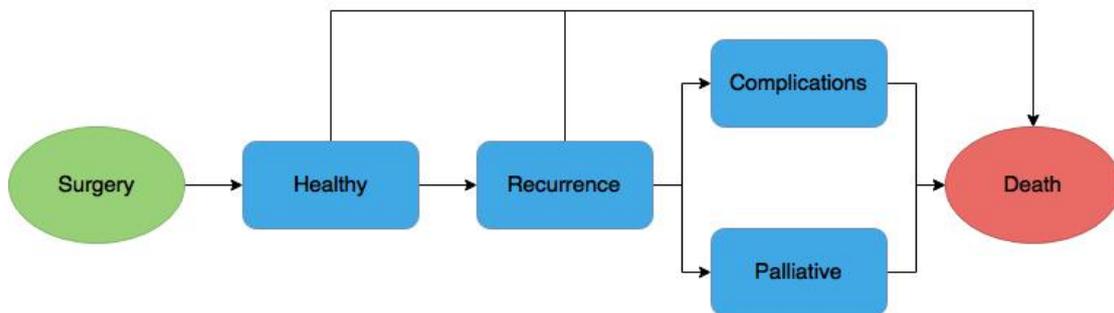


Figure 1. Graphical representation of the case study's clinical pathway.

The test dataset includes observations of 381 hypothetical patients that were followed throughout this clinical pathway. If an observation for a patient is not available, the corresponding cell is empty (i.e. NA). The following observations are included in the dataset:

patient.id	<i>unique patient identifier</i>
healthy.time	<i>time spent in the Healthy state</i>
healthy.cat	<i>competing risks identifier for the Healthy state: (1) death or (2) recurrence</i>
recurrence.time	<i>time spent in the Recurrence state</i>
recurrence.cat	<i>competing risks identifier for the Recurrence state: (1) death, (2) major complications, or (3) palliative</i>
recurrence.minorComp	<i>identifier for minor complications in the Recurrence state: (1) yes or (0) no</i>

complications.time	<i>time spent in the Major Complications state</i>
palliative.time	<i>time spent in the Palliative state</i>
costs.surgery	<i>costs of the surgery</i>
costs.recurrence	<i>costs per day in the Recurrence state</i>
costs.complications	<i>costs per day in the Major Complications state</i>
costs.palliative	<i>costs per day in the Palliative state</i>
costs.minorComp	<i>costs of minor complications</i>
util.healthy	<i>utility in the Healthy state</i>
util.recurrence	<i>utility in the Recurrence state</i>
util.complications	<i>utility in the Complications state</i>
util.palliative	<i>utility in the Major Complications state</i>
disutil.minorComp	<i>disutility of minor complications</i>

2. Dataset structure and supported parameters

In order for iDist to analyze datasets properly, the datasets need to be structured as defined in this tutorial. First it is important to point out the types of parameters that are supported. iDist will estimate probabilities of events, such as the occurrence of minor complications in the Recurrence state of the hypothetical case study. Furthermore, several parametric distributions are supported to describe, for example, costs, utilities, and survival data. The supported distributions are the Beta, Gamma, Normal, and Weibull distribution. The current version of iDist does not support the analysis of censored survival data, though a version that does support this is currently being developed. Survival data that is subject to competing risks can also be analyzed using iDist. For this, Weibull distributions are used. If no observation is available for a patient, the corresponding value can be left empty or defined as "NA". Obviously, each type of observation requires the data to be defined appropriately:

probability	<i>integer values with 1 indicating an event and 0 indicating no event</i>
Beta distribution	<i>numeric values between 0 and 1</i>
Gamma distribution	<i>numeric values larger than 0</i>

Normal distribution	<i>numeric values</i>
Weibull distribution	<i>numeric values larger than 0</i>
event column	<i>integer values</i>

3. Visualization of the dataset

A dataset can be loaded into iDist on the *Data Upload* tab by selecting a file from the user's computer and selecting the separator and decimal point used to define the dataset. Next the file can be read by pressing the *Read File* button. The supported file formats are text files (.txt) and comma delimited files (.csv). If no dataset is available to upload to iDist, the test dataset as described in this tutorial can be loaded by selecting this option rather than the option to upload a custom dataset.

After a dataset is loaded into iDist, the structure of the loaded dataset should be checked on the *Data Table* tab. This table displays the dataset that is currently loaded into the tool. Using the slider, the number of digits that is displayed can be adjusted. If the dataset is not loaded correctly, the setting on the *Data Upload* need to be adjusted and the file need to be read again by pressing the *Update* button. If the dataset is loaded correctly, the process can continue by further visualizing and analyzing the dataset.

The dataset can be visualized further on the *Plot Data* tab. Histograms that include a density curve and summary statistics can be generated on this tab. The first step in generating these plots is selecting the column that contains the data to be plotted. If the data to be plotted is survival data that includes competing risks, this needs to be indicated (*yes/no*). Otherwise (*no* selected), i.e. the data represents survival data without competing risks or utilities for example, the data can be plotted directly using the *Plot Data* button.

If the data represents survival data including competing risks (*yes* selected), a column needs to be selected that contains information on which event corresponds to the observed survival time, i.e. time-to-event. The events need to be defined by integers. For example, a one (1) indicates death and a two (2) indicates progression). After the column that contains the event data corresponding to the selected survival data column is selected, the event to be

plotted needs to be selected and the histogram can be generated using the *Plot Data* button.

After plotting the histogram using the *Plot Data* button, a parametric distribution can be fitted to the data and a simulated density curve will be added to the plot. If the selected distribution does not match the data, e.g. a Gamma distribution is selected for data that includes negatives, this will be indicated below the plot in the summary statistics area. Plotting the data and adding different distributions, helps to select the appropriate settings for the data analysis in the next steps.

4. Defining the general variables

If a dataset is loaded into the tool, variables for which parameters need to be estimated from the dataset can be easily defined on the *General Variables* tab. Please notice that only variables representing survival data that includes competing risks need to be defined on the *Competing Risks* tab (see Section 5), other types of variables are defined here at the *General Variables* tab. As pointed out in Section 2, these types of variables include probabilities and variables defined by Beta, Gamma, Normal, or Weibull distributions.

Variables can be added to the analysis using the *Add Variable* button. To define a variable, simply select the column that contains the data and select the type of the variable. The provided input can be checked, i.e. whether the selected variable type matches the selected data, using the *Check Input* button. The last variable can be deleted using the *Delete Variable* button. Use the *Clear Variables* button to clear all provided input. Once all relevant “general” variables are defined and the provided input is checked, variables that represent survival data including competing risks can be defined on the *Competing Risks* tab or the analysis can be started on the *Data Analysis* tab otherwise.

5. Defining the competing risk variables

Variables representing survival data including competing risks can be defined on the *Competing Risks* tab in almost the same way as all other variables are defined at the *General Variables* tab. Variables can be added or deleted and the provide

input can be cleared. However, for the competing risks variables not only the column that contains the data to be analyzed needs to be selected, also the column that contains the corresponding event data, as defined earlier, needs to be selected. No variable type needs to be selected, because Weibull distributions are fitted to all competing risk survival data in the current version of iDist. Please also be aware that this current version does not support censored survival data. An update of the tool, which allows for censored data and different types of survival models, e.g. Gamma, is currently being developed.

According to literature, different approaches can be taken to handle competing risks in patient level models. iDist provides input for discrete event simulation models, following the approach of selecting the event first and the time-to-event second. After analysis, event-specific probabilities are available, which can be used to select the event to occur to a patient in the simulation. Besides this probability of occurrence, each event also gets its own Shape and Scale parameter to define a Weibull distribution. After selecting the event using the probabilities, a time-to-event can be drawn from the corresponding Weibull distribution. Please see Figure 2 for a graphical representation of this process for the event to occur to a simulated patient in the *Recurrence* state of the case study.

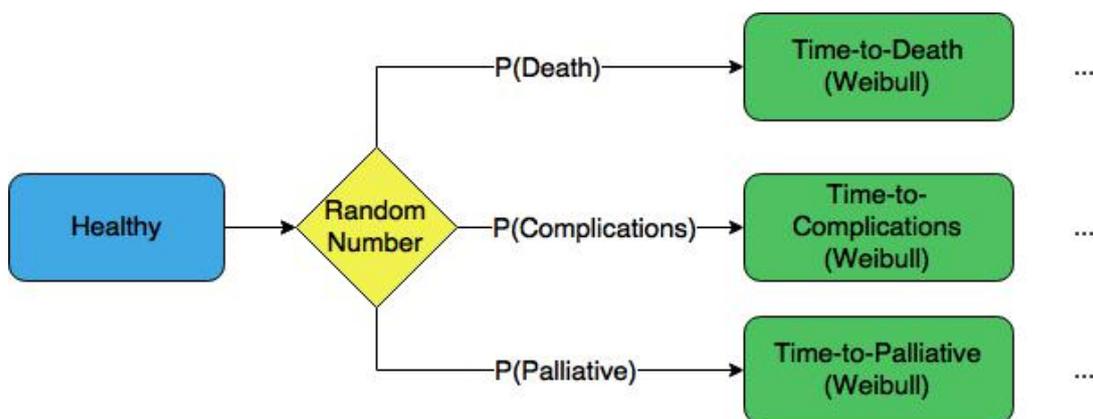


Figure 2. Graphical representation of how iDist analyzes competing risks and how these can be implemented in corresponding discrete event simulation models.

6. Analyzing the dataset

After the dataset is loaded, appropriate distributions are determined by plotting the data, and all variables are defined, the analysis can be performed on the *Data Analysis* tab. A *Base-case Analysis* can be performed, which will result in the estimated value for each parameter. Additionally, output for a *Probabilistic Sensitivity Analysis* can be generated, which will repeatedly bootstrap the loaded dataset and estimate all parameters to the bootstrap samples. If input for a probabilistic sensitivity analysis (PSA) is required, the required number of *simulation* runs (i.e. PSA runs/samples) needs to be defined. The random number seed used can also be defined to enable obtaining the same results at a later moment.

The analysis can be started using the *Start data analysis* button. The progress of the analysis is visible in the upper-right corner. After the dataset is analyzed, the results can be observed in the data table presented below the *Start data analysis* button and downloaded to the user's computer. The results are exported as a text-file (.txt) and can be loaded into a simulation software environment or office programs, such as Microsoft Excel.

Thank you for using iDist! We hope that the tool was helpful and that you will send us an e-mail (info@personex.nl) or a message (<https://personex.nl/contact/>) to provide us any feedback on your experiences, so we can improve the tool.