



# **Producing Plots to Assess Goodness-of-Fit for the Multistage Weibull (MSW) Time-To-Tumor Model *[gofplot\_msw()]***

## **USER MANUAL**

Chemical & Pollutant Assessment Division  
Center for Public Health and Environmental Assessment  
Office of Research and Development  
U.S. Environmental Protection Agency

## Table of Contents

|            |   |           |
|------------|---|-----------|
| <b>1.0</b> | <b>Introduction .....</b>   | <b>3</b>  |
| 1.1        | Acknowledgements .....  | 3         |
| 1.2        | Disclaimer .....  | 3         |
| <b>2.0</b> | <b>Installing and Using the <i>gofplot_msw()</i> Plotting Functions .....</b> | <b>4</b>  |
| 2.1        | Installing <i>gofplot_msw()</i> .....   | 4         |
| 2.2        | Using <i>gofplot_msw()</i> .....  | 4         |
| <b>3.0</b> | <b>Types of Plots Incorporated in MSW Module .....</b>                        | <b>6</b>  |
| 3.1        | Probability vs. Time Plot (type="pr") .....                                   | 6         |
| 3.2        | Dose-Response Plot (type="dr") .....  | 6         |
| 3.3        | Hazard Plot (type="hz") .....   | 6         |
| 3.4        | Quantile-Quantile Plot (type="qq") .....                                      | 7         |
| 3.5        | Probability-Probability Plot (type="pp") .....                                | 7         |
| <b>4.0</b> | <b>Generating the Plots .....</b>   | <b>8</b>  |
| 4.1        | Usage .....   | 8         |
| 4.2        | Arguments .....   | 8         |
| 4.3        | Example Data Sets .....   | 9         |
| <b>5.0</b> | <b>Probability Models Used by <i>gofplot_msw()</i> .....</b>                  | <b>10</b> |
| 5.1        | Fatal Risk .....  | 10        |
| 5.2        | Incidental Risk .....   | 10        |
| <b>6.0</b> | <b>Goodness-of-Fit and Diagnostic Plots .....</b>                             | <b>12</b> |
| <b>7.0</b> | <b>References .....</b>   | <b>14</b> |
| <b>8.0</b> | <b>Example Plots .....</b>  | <b>15</b> |

## Table of Figures

|  |    |
|--|----|
| Figure 1: Generating probability vs. time (type="pr") plots for the Non-Fatal tumor model .....            | 15 |
| Figure 2: Generating probability vs. time (type="pr") plots for the Non-Fatal and Fatal tumor models ..... | 16 |
| Figure 3: Generating a Dose-Response (type="dr") plot for the Non-Fatal tumor model .....                  | 17 |
| Figure 4: Generating a cumulative hazard (type="hz") plot for the Non-Fatal tumor model .....              | 18 |
| Figure 5: Generating a cumulative hazard (type="hz") plot for the Fatal tumor model .....                  | 19 |
| Figure 6: Generating a quantile-quantile (type="qq") plot for the Non-Fatal tumor model .....              | 20 |

---

## 1.0 Introduction

This user manual describes installation and use of plotting functions developed for use with output files produced by the “MSW” (multistage Weibull) program for time-to-tumor (time-to- event) analysis. These functions are described in aggregate as “*gofplot\_msw()*”, the name of the main function, the only one invoked by the user.

The functions were developed in the R language (R Core Team 2012). These R functions are intended for comparing the fitted, parametric MSW model to an empirical, nonparametric time-to-event model that is fitted to the same data. Comparisons between the parametric and nonparametric models provide a subjective assessment for goodness-of-fit of the multistage Weibull model to the data. This document describes the use and interpretation of various types of plots which *gofplot\_msw()* can generate.

A separate user manual describes the MSW program. Users of *gofplot\_msw()* must be familiar with the MSW user manual in order to understand, interpret, and report results from MSW and *gofplot\_msw()*. A user also should be familiar with the basics of R.

Users of these programs are expected to be familiar with basic methods of survival analysis including graphical diagnostic methods (Collett 1994; Klein and Moeschberger 2003; Lawless 2003) and with the R package ‘*survival*’.

## 1.1 Acknowledgements

The R programs in *gofplot\_msw* (and the MSW programs in C) were originally developed by Battelle Memorial Institute for the U.S.E.P.A. during 2006-2009 (contract EP-C-04-027). EPA staff subsequently modified the plotting programs in R and wrote new modules for dose- response plotting and hazard plotting.

The MSW program uses the C version (donlp2-intv) of the optimizer *donlp2*, copyright dated June 21, 2004 by Peter Spellucci. A condition of use specified by the author is: “The use of donlp2 must be acknowledged, in any publication which contains results obtained with it or parts of it. Citation of the author’s name and netlib-source is suitable.” Note that this may apply to plotting the results from MSW by using *gofplot\_msw*.

## 1.2 Disclaimer

The views expressed in this manual are those of the author(s) and do not necessarily represent the views or the policies of the U.S. Environmental Protection Agency.

## 2.0 Installing and Using the *gofplot\_msw()* Plotting Functions

### 2.1 Installing *gofplot\_msw()*

The following instruction assumes very little experience with R. An experienced user may also want to review this section before setting up *gofplot\_msw*. User is assumed to have installed MSW (*msw.exe*) and know how to use it.

To install the R plotting functions in *gofplot\_msw*:

1. Ensure that R is installed on the user's computer, and note the folder in which it is installed. R is available for download free-of-charge from the websites <http://www.r-project.org> or at <http://cran.r-project.org/>.
2. Create a Windows folder (subdirectory). Download and unzip(extract) contents of the zip file *gofplot\_msw.zip* into this folder.

For this example, we will assume that *gofplot\_msw* is installed in folder "C:\gofplot\_msw\". Make sure there is a copy of *msw.exe* in the same folder for convenience.

3. Make a Windows shortcut to R, place it in the folder, and install the R plotting function.
4. Right click on the R icon and click on "Create Shortcut". Move the shortcut to "C:\gofplot\_msw\". Double click to test the shortcut. An R console window should open.
5. Entering the R command `> getwd()`, the response should be:  

```
[1] "C:/gofplot_msw"
```
6. If the response is different, change the work directory by typing  

```
> setwd ("C:/gofplot_msw/")
```

**Note:** The forward slash is an R convention.
7. Install the R functions by entering the command  

```
> source("gofplot_msw.R")
```
8. Typing `> ls()` will display the names of the R functions.

### 2.2 Using *gofplot\_msw()*

This section describes a step-by-step procedure for using *gofplot\_msw()*.

1. Create a folder for new data (e.g., "C:\timetotumor\_1").
2. For this demonstration, copy the data files (with file extensions '.(d)') and output files (with file extensions '.out') from the folder where *gofplot\_msw* is installed to this new folder.
3. Return to the folder "C:\gofplot\_msw\" and open an R console.

4. Change the working directory by typing the command

```
> setwd ("C:/timetotumor_1")
```

Note the forward slash!

5. We will use the files EX1\_C+F+I\_Grouped\_Irisk.(d) and EX1\_C+F+I\_Grouped\_Irisk.out.

6. Type the command

```
> gofplot_msw(dname="EX1_C+F+I_Grouped_Irisk.(d)", oname =  
"EX1_C+F+I_Grouped_Irisk.out", type = "pr")
```

This should produce a plot in the R window. Right-clicking in the plot will list options for saving the plot or printing it. To save it as a PDF or PNG file in R, read the R documentation by typing

```
> ?pdf
```

or

```
> ?png
```

Details about plotting options and command arguments are described in Section 4.0 on page 8.

Here is an example of commands for writing the figure to a png file:

```
> png (filename="plot1.png", width=7, height=7, units="in", res=200)
```

Be sure to use file extension .png in the file name.

```
> gofplot_msw(dname="EX1_C+F+I_Grouped_Irisk.(d)", oname =  
"EX1_C+F+I_Grouped_Irisk.out", type = "pr")
```

```
> dev.off() # closes the png 'device' & writes the new file
```

## 3.0 Types of Plots Incorporated in MSW Module

The general objective of the MSW plotting tool is to assess the goodness-of-fit of the (parametric) MSW model by comparing it to a nonparametric model fitted to the same data. The nonparametric model imposes only the most necessary restrictions (esp. monotonicity) on the relationship between time, dose, and probability of tumor onset or death, with no assumption made on the specific distributional form of the data. By minimizing the restrictions on the structure of the model, the empirical nonparametric model fits the data as “closely” as possible.

The MSW plotting tool includes several plot types found useful for evaluating goodness-of-fit of survival functions (Collett 1994, 2003; Klein and Moeschberger 2003). It allows the user to plot the following types of graphs, by specifying the *type=""* argument in *gofplot\_msw()*:

### 3.1 Probability vs. Time Plot (type=“pr”)

In this plot (Figures 1 and 2), which is the default plot for *gofplot\_msw()*, the fitted distribution function is plotted against time, separately for each dose level specified in the input (\*.d) file. The MSW model is plotted as a smooth curve and the nonparametric model is displayed as a series of points (representing a right-continuous step function). If both Fatal and Incidental contexts occur in the data, two smooth curves and two series of points are plotted, a solid curve and filled points for the Fatal tumor response, and a dashed line and unfilled points for the Incidental tumor response.

### 3.2 Dose-Response Plot (type=“dr”)

This plot (Figure 3) shows response probability in relation to dose for a fixed time (usually 104 weeks for rodent bioassays) and also shows the BMD and BMDL values from the output file. In fitting the MSW model, the user will have specified either the Fatal or Incidental risk response at a specific time as a basis for calculating the benchmark dose (BMD). The predicted probability values from the distribution functions of both the MSW and nonparametric models for the specified risk response at the specified time are plotted against the range of dose values occurring within the data. The parametric MSW model is plotted as a smooth curve across the range of doses. The nonparametric probability estimates are plotted as points at the distinct dose values that occur within the data. The BMD and its lower confidence limit are also plotted and their values are printed in the plot legend.

### 3.3 Hazard Plot (type=“hz”)

This option (Figures 4 and 5) plots the log (base e) of the cumulative hazard function against event time for the parametric (MSW) model and for a nonparametric estimate. The log cumulative hazard plot is linear for a Weibull hazard function; thus, the parametric estimate is necessarily linear. If the nonparametric estimate is not approximately linear, the MSW model may not be suitable. A plot is also produced showing the nonparametric estimates for the dose groups in a single figure.

### **3.4 Quantile-Quantile Plot (type="qq")**

Using the fitted nonparametric model, predicted probability values from the distribution function are calculated at distinct observation times that represent quantiles of the predicted probabilities based on this model (Figure 6). Subsequently, the quantiles (in time units) for those predicted probabilities are calculated using the fitted MSW model, using the inverse of the estimated density function. The quantiles from the MSW model are then plotted against quantiles from the nonparametric model, with each plot point representing the quantiles at the given (predicted) probabilities. The graph also includes a straight line with unit gradient slope through the origin. The goodness-of-fit for the MSW model can be assessed by how well the plotted values follow this straight line. Quantile plots are generated for each dose level specified in the input (\*.d) file.

### **3.5 Probability-Probability Plot (type="pp")**

The predicted probability values from the distribution functions of both the MSW and nonparametric models are calculated for distinct observation time values that occur within the data. The probabilities from the MSW model are then plotted against the probabilities from the nonparametric model, with each plot point representing the two probabilities at a given observation time. The graph also includes a straight line with unit gradient (i.e., 1:1) slope through the origin. The goodness-of-fit for the MSW model can be assessed by how well the plotted values follow this straight line. Probability plots are generated for each dose level specified in the input (\*.d) file (this plot is not illustrated).

## 4.0 Generating the Plots

The R function *gofplot\_msw()* serves as the primary user interface to the MSW plotting routines. The *gofplot\_msw()* function is called by specifying *gofplot\_msw* on the R command line, followed by specifying the following five parameters (function arguments) in parentheses. It is recommended that the value of each parameter be specified by name (e.g., 'type="dr"') and in the order given below. The user must have first executed the MSW program and generated the output file before calling the plotting functions.

Figures 1-6 show how to call the function *gofplot\_msw* to generate certain types of plots, and how the plots appear upon executing these function calls.

### 4.1 Usage

```
gofplot_msw(dname="Filename.(d)", oname="Filename.out", type="dr",  
title = NULL, digit = 4, mult = TRUE, Prange01="TRUE")
```

### 4.2 Arguments

#### ***dname***

This required parameter specifies the name of the file containing the time-to-tumor data used by the MSW.exe module to fit the MSW model. The filename '*dname*' may be written with or without the extension '*.(d)*'. If the file is not in the R working directory, one may use the full path, e.g., C:\\mydata\\fname.(d). Windows path references should work, e.g., for data in a subdirectory data of the working directory, use *gofplot\_msw*("\\.\\data\\ fm\_alvbr\_1stage\_auto").

Recall that in R, one must use either the double backslash or a single forward slash in a pathname, but never the single backslash that is used by Windows. If the filename plus extension exceeds 4 characters, the data file need not have extension '*.(d)*' – whatever path and filename were typed will be used. However, we recommend using the standard MSW file extension '*.(d)*'.

#### ***oname***

This optional parameter specifies the name of the output file (\*.out) generated by the MSW module - this contains information on the MSW model fitted to the data specified by *dname*. The parameter must be specified when the file names of *dname* and *oname* differ (not a good practice!). If *oname* is not given, the output file (\*.out) must reside in the same directory as *dname*; the path and filename specified for *dname* will be used, with file extension '*.out*' rather than '*.(d)*'.

#### ***type***

This optional parameter specifies the type of plots to be generated for each dose level. If no value is specified by the user, the parameter value defaults to "pr". Possible values are:

- "pr" – plots the parametric (MSW) and nonparametric probability functions
- "dr" – generates a dose-response plot
- "hz" – generates a cumulative hazard plot
- "qq" – to generate Quantile-Quantile plots
- "pp" – to generate Probability-Probability plots.

***title***

This optional parameter is used to specify a title for figures where multiple plots appear on one page (when parameter `Mult=TRUE`). The title will appear at the top of the figure and thus will apply to all the plots on the page.

***digit***

This optional parameter is used to specify the number of significant digits used to display Dose values in the plots. If no value is specified by the user, the default is 4 significant digits.

***mult***

This optional parameter equals `TRUE` if multiple plots are to be placed on a single page (the default), and `FALSE` if a page is to hold a single plot. If this parameter equals `TRUE`, then up to four plots (2x2) will be displayed on a single page.

***save.Plot***

Setting this argument to `"TRUE"` will cause a plot to be exported to the file type specified by argument ***save.Plot.type***. To see the valid types, type `?savePlot` or `help(savePlot)` in the R console. The types best suited for importing graphics into your documents are likely to be `"pdf"`, `"png"`, `"tiff"` and `"jpeg"`.

***haz.npar.type***

Specifies the type of nonparametric estimate to use in the hazard plot (`type="hz"`). Valid choices are `"fh"` (the default), `"iso"`, and `"km"`. These correspond to Fleming-Harrington, isotonic regression, and Kaplan-Meier (Product-Limit) estimates.

***Prange01***

Logical argument. If `TRUE`, set range of probabilities in plot axis to `[0,1]`. Default is `FALSE` (use observed range of probabilities to determine plot axis range using style `"r"` as in `par(yaxs="r")` ).

A plot can also be copied from the R console by right-clicking inside the plot and selecting from the pop-up menu, then pasting the graphic into an open document.

Function *gofplot\_msw* invisibly returns a list containing the fitted models, named `par_resultI`, `par_resultF`, `npar_fh`, `npar_iso`, and `npar_km`. These represent the MSW models (`"par"`) and the nonparametric models (`"npar"`) for Incidental (I) and Fatal (F) risk. The parametric fits contain risk probabilities at equally spaced times ranging from the lowest to highest observed times. The nonparametric fits contain risk probabilities at the observed times. The object `npar_fh` is null unless a hazard plot was requested (`type="hz"`).

## 4.3 Example Data Sets

A number of example MSW input and output data sets are provided with the *gofplot\_msw()* download package. These serve to illustrate different types of data and risk responses, and will assist users in becoming familiar with use of *gofplot\_msw* and with MSW and time-to-tumor data.

## 5.0 Probability Models Used by *gofplot\_msw()*

Animal observations from a carcinogenicity bioassay may be classified (see the user manual for program MSW) as: Fatal (F), meaning that the animal death is attributed to the tumor; Incidental (I), meaning that the animal died from another cause (including a planned sacrifice) and the tumor was found to be present during necropsy; right Censored (C), meaning that no tumor (of the sort under consideration) was present, and Unknown (U), meaning that the animal was not necropsied (usually these are uncommon and occur because of decomposition or cannibalism).

### 5.1 Fatal Risk

**Multistage Weibull parametric model (plotted as a smooth curve):**

object "par\_resultF"

The distribution function for Fatal Risk is:

$$F(t, d) = F(t, d, t_0, c, \beta_0, \beta_1, \dots, \beta_k) = 1 - \exp \left\{ - (t - t_0)^c \sum_{i=0}^k \beta_i d^i \right\}$$

where the maximum likelihood estimates from MSW are substituted for the model parameters.

**Nonparametric model (plotted as points representing a step function):**

object "npar\_km"

The censoring indicator for Fatal Risk is 1 for "F" observations and is 0 if death from tumor has not occurred (context or class C, I, and U). Data are thus right-censored. The nonparametric maximum likelihood estimator in this case is the Kaplan-Meier (Product-Limit) estimator. The R function *survfit()* is used to compute the KM estimate of the survival function  $S(t)$  for death from tumor. The plotted function is  $1 - S(t) = F(t)$ .

### 5.2 Incidental Risk

**Multistage Weibull parametric model (plotted as a smooth curve):**

object "par\_resultI"

The distribution function for Incidental Risk (non-fatal tumors) is:

$$G(t, d) = G(t, d, c, \beta_0, \beta_1, \dots, \beta_k) = 1 - \exp \left\{ - t^c \sum_{i=0}^k \beta_i d^i \right\}$$

where the maximum likelihood estimates from MSW are substituted for the model parameters.

***Nonparametric model (plotted as points representing a step function):***

object "npar\_iso"

The censoring indicator  $\delta$  for Incidental Risk is 1 for "I" and "F" observations (i.e., if growth of a tumor has already begun) and is 0 if a tumor has not yet occurred (context "C"). These observations represent current-status data (e.g., Lawless 2003): if  $X_{(i)}$  is the time when a tumor just becomes detectable and  $T_{(i)}$  is the observation time, either  $X_{(i)} \leq T_{(i)}$  or  $X_{(i)} > T_{(i)}$  (the subscript indicates that the data are sorted in ascending order of the  $T_{(i)}$ ,  $i = 1, 2, \dots, n$ ). The nonparametric maximum likelihood estimator (NPMLE) for such data is the isotonic regression for the cumulative sum of indicators on the index  $i$  (Groeneboom and Wellner 1992). This estimator is equivalent to that of Ayer et al. (1955). The R function *isoreg()* is used for computing the NPMLE of the distribution function ( $F(t) = 1 - S(t)$ ) for the appearance of detectable tumors. The NPMLE based on *isoreg()* is used for all plot types except the cumulative hazard plot, for which the nonparametric estimate is based on Turnbull's EM approach.<sup>1</sup>

---

<sup>1</sup> See comments in function '*hazardplot\_msw()*'. Also see documentation for package *survival*, functions *Surv()* and *survfit()*, R Core Team 2012. The two nonparametric estimates appeared similar or indistinguishable for all data sets so far tested.

## 6.0 Goodness-of-Fit and Diagnostic Plots

The MSW software does not report a chi-square goodness-of-fit table. Methods similar to the chi-square goodness-of-fit test that is applied to BMDs quantal models do not apply to the MSW model with censored data. Development of a suitable goodness-of-fit test, especially for heavily censored and current-status data, is difficult.<sup>2</sup>

Recent work<sup>3</sup> explores new approaches for goodness-of-fit and confidence intervals on the nonparametric survivor function for current-status and interval-censored data, but we have not attempted implementation.

This leaves us with graphical comparison of the MSW parametric model and the nonparametric model for judging goodness-of-fit.<sup>4</sup> This software provides several types of diagnostic plots as described above. More detailed advice on use of these plots may be found in textbooks on survival analysis (Anderson et al. 1993; Collett 1994, 2003; Klein and Moeschberger 2003; Lawless 2003).

Carcinogenicity data are heavily censored and events (Incidental and Fatal observations) may be sparse in some dose groups. This limits ability to evaluate model fit and distinguish between alternative models. The best one should expect from these plots, applied to such data, is to occasionally identify cases of poor fit that do not strongly support reliance on the BMD/BMDL inference.

The cumulative probability plot (type="pr") and dose-response plot (type="dr") will show how well the model agrees with the data. Also, agreement of the model with data is indicated by a linear pattern for hazard (type="hz"), probability-probability (type="pp") and quantile-quantile (type="qq") plots.

We have found it useful to check the cumulative probability plot (type="pr"), dose-response (type="dr"), and hazard (type="hz") plots. The 'qq' and 'pp' plots have specific uses when considering models other than Weibull, for example loglogistic, and evaluating an accelerated failure time model (ibid.). The hazard plot may indicate when an alternative to the Weibull model needs to be considered.

When the multistage Weibull model is used to infer a benchmark dose (BMD), one should be concerned about the goodness-of-fit at the corresponding time and dose level. Usually the BMD is estimated for a nominal rodent lifetime, say 104 weeks (2 years). Often the BMD falls near or below the lowest dose. In such a case, we desire a good fit at longer times and lower doses. That does not imply that the fit at earlier times or higher doses should be ignored.

The hazard plot (type="hz") is helpful in evaluating reasonableness of the multistage Weibull model and alternative models (ibid.). The hazard plot option (type="hz") will produce plots of the parametric (MSW) and nonparametric estimates separately for each dose group (Figure 4), and a final plot showing the nonparametric estimates for all dose groups in one figure. The log cumulative hazard for the nonparametric estimate will be linear with time if the data agree with a Weibull model (for the estimated parametric model, it is necessarily linear), with a slope determined by the Weibull shape parameter. Increasing log cumulative hazard plots, parallel (with the same slope) and with different intercepts for different dose groups, are consistent with the Weibull model. If the slope of

---

<sup>2</sup> In principle, it should be possible to implement a bootstrap goodness-of-fit test (Lawless 2003, ibid.).

<sup>3</sup> See References Section: Ren 2003, 2008; Banerjee & Wellner 2005; Koul & Yi 2006; Sun et al. 2007; Koul & Aggarwal 2008; Maathuis & Hudgens 2011.

<sup>4</sup> In fact, Klein and Moeschberger (2003, Ch. 12.5) favor graphical examination over formal significance tests, and demonstrate how graphical diagnostics can inform alternative modeling choices and yield other insights.

the nonparametric hazard function increases or decreases systematically with dose, a different model may be appropriate or competing risks may need to be considered.

Even when interest lies in the Incidental risk, it can be useful to examine the cumulative hazard for the Fatal risk function, which is identical to the Incidental risk function shifted to the right by  $t_0$ . There are more points of support for the nonparametric estimate and so it is easier to judge agreement with the parametric estimate. One way to generate this plot is to run MSW to estimate the Fatal Risk model (Figure 5), and to use those data and output files as arguments to *gofplot\_msw()*. Note that the cumulative hazard function for Fatal risk is curved, not linear, in the plot (Figure 5). Alternatively, one can generate the same plot using the MSW data and output files for Incidental risk by using a special argument to *gofplot\_msw()*. This avoids re-running MSW just to obtain the fatal risk model. If the argument *haz.npar.type* is set to *haz.npar.type="km"* the hazard plot appropriate for Fatal risk is printed, albeit with a subtitle "Incidental Risk" - read from the output file - so it would be best to use the *title* argument of *gofplot\_msw()* to remind you that the Fatal risk function is plotted. Two cautions: (1) normally a user should not specify the argument *haz.npar.type*. (2) the BMD and BMDL in the output file for the Incidental risk model are still appropriate for Incidental risk, not Fatal risk, as is the dose-response plot (*type="dr"*).

One final caution: When occurrence of tumor contexts (*I*, *F*, and *C*) is highly confounded with dosage (*d*) or observation time (*t*), estimates can be unreliable<sup>6</sup>. Model diagnostic plots, parameter confidence limits, and the table of response categories by dose groups in the output file can help to identify such cases.

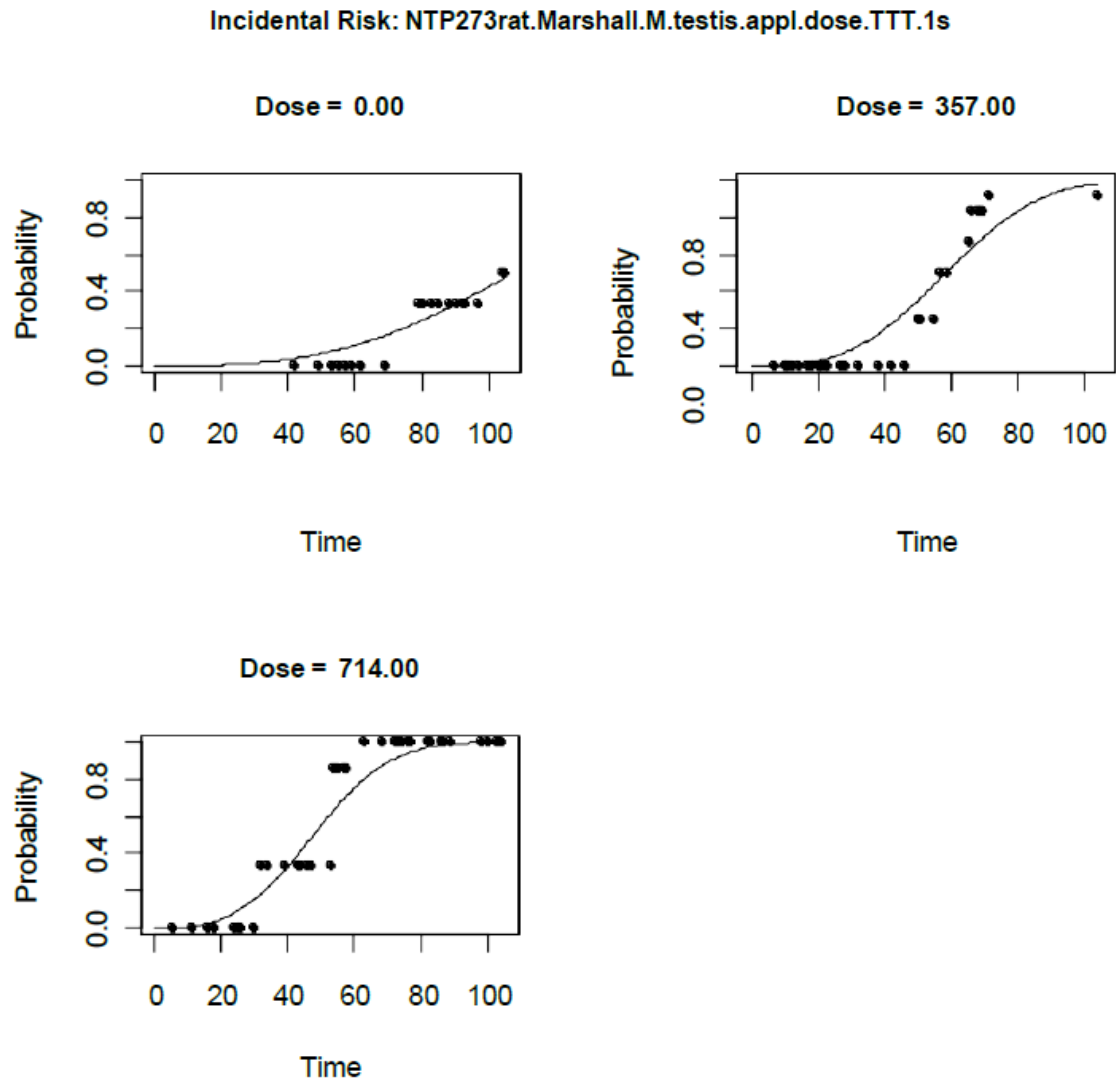
## 7.0 References

- Anderson, P.K., Ø.Borgan, R.D.Gill and N.Keiding. 1993. Statistical Models Based on Counting Processes, New York:Springer.
- Ayer, M., H.D. Brunk, G.M. Ewing, W.T. Reid and E. Silverman. 1955. An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics* **26**: 641-647.
- Banerjee, M. and J.A. Wellner. 2005. Confidence intervals for current status data, *Scandinavian J. Stat.* **32**: 405-424
- Collett, D. 1994. Modelling Survival Data in Medical Research. Chapman & Hall. (and 2<sup>nd</sup> ed. 2003)
- Groeneboom, P. and J.A. Wellner. 1992. Information Bounds and Nonparametric Maximum Likelihood Estimation. Basel: Birkhäuser.
- Klein, J.P., and M.L. Moeschberger. 2003. Survival Analysis. 2<sup>nd</sup> ed. Springer.
- Koul, K.L. and T. Yi. 2006. Goodness-of-fit testing in interval censoring case 1, *Statistics & Probability Letters* **76**: 709–718
- Koul, K.L. and D. Aggarwal. 2008. Minimum empirical distance goodness-of-fit tests for current status data, *J. Indian Statistical Association* **46**: 79-125
- Lawless, J.F. 2003. Statistical Models and Methods for Lifetime Data. 2<sup>nd</sup> ed. Hoboken: Wiley- Interscience.
- Maathuis, M. and Hudgens, M. 2011. Nonparametric inference for competing risks current status data with continuous, discrete or grouped observation times, *Biometrika* **98**: 325–340
- R Core Team. 2012. R: A Language and Environment for Statistical
- Ren, J.-J. 2003. "Goodness of fit tests with interval censored data. *Scandinavian Journal of Statistics* **30**: 211–226
- Ren, J.-J. 2008. Smoothed weighted empirical likelihood ratio confidence intervals for quantiles, *Bernoulli* **14**: 725-748
- Sun, J., L. Sun and C. Zhu. 2007. Testing the proportional odds model for interval-censored data, *Lifetime Data Analysis* **13**: 37-50

## 8.0 Example Plots

Figure 1: Generating probability vs. time (type="pr") plots for the Non-Fatal tumor model

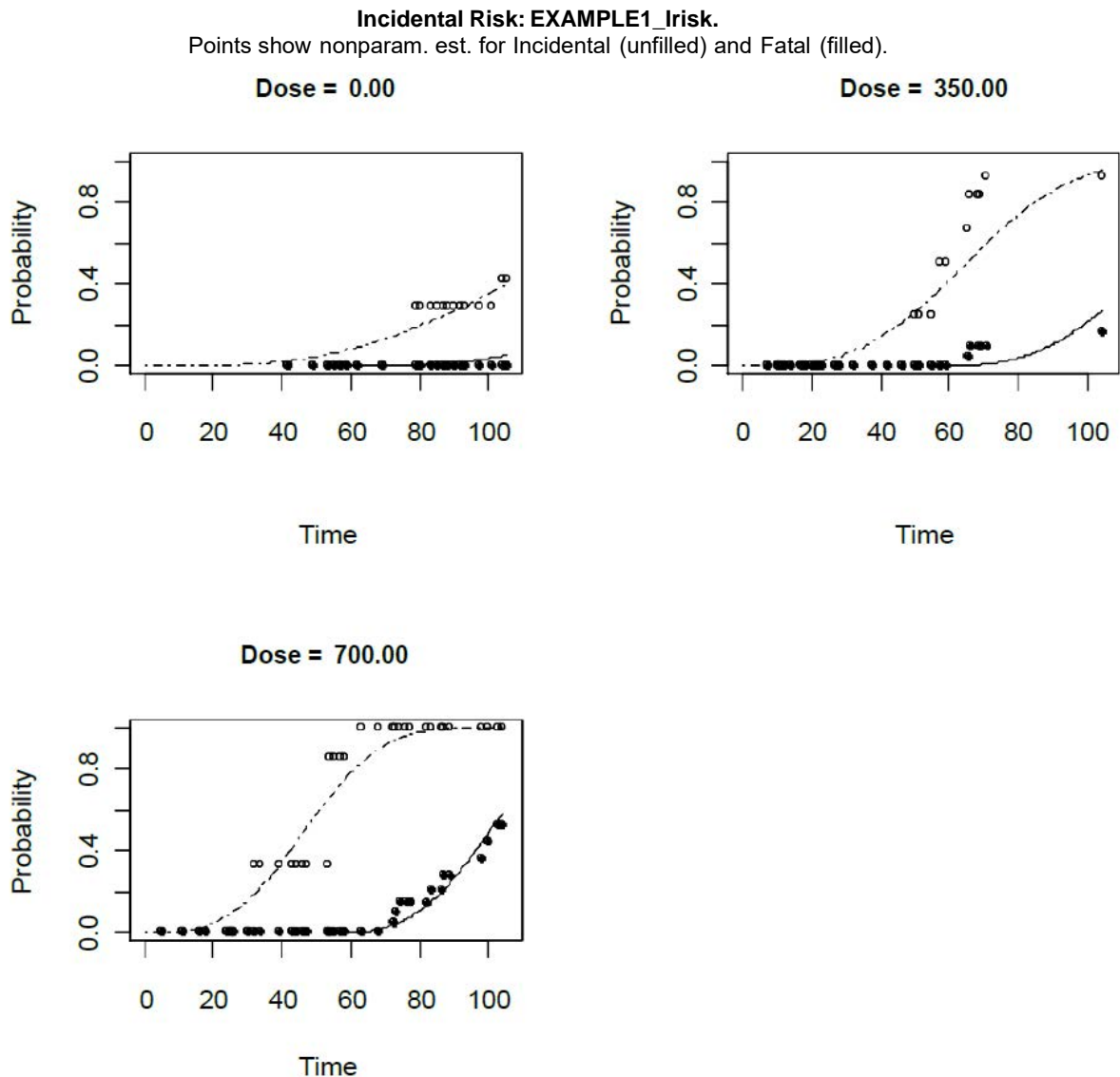
```
gofplot_msw("NTP273rat.Marshall.M.testis.appl.dose.TTT.1s")
```



Legend: The solid line represents probabilities as determined from the fitted MSW model, and the filled points represent probabilities as determined from the nonparametric model.

Figure 2: Generating probability vs. time (type="pr") plots for the Non-Fatal and Fatal tumor models

```
gofplot_msw("EXAMPLE1_Irisk", type="pr")
```

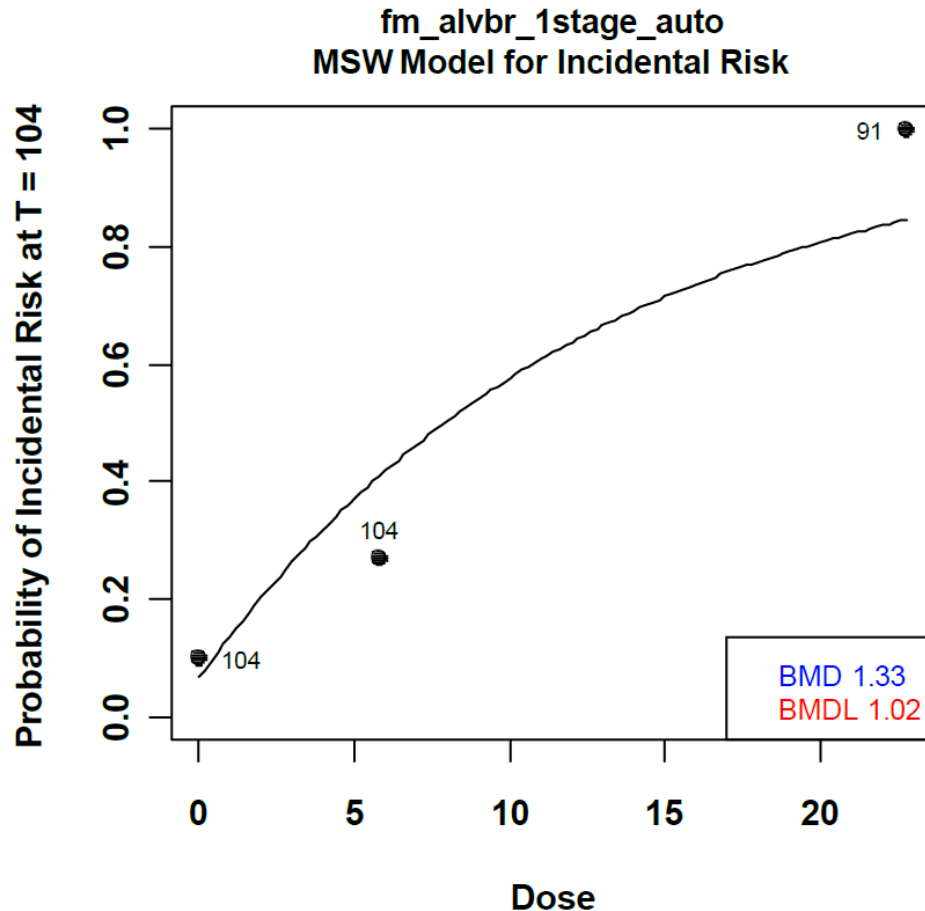


Legend: The solid line represents probabilities as determined from the fitted MSW model for fatal tumor risk, and the filled points represent probabilities as determined from the nonparametric model. The dashed line represents probabilities as determined from the fitted MSW model for incidental tumor risk, and the unfilled symbols represent probabilities as determined from the corresponding nonparametric model. The title "Incidental Risk" refers to the user-specified risk reported in the output file "EXAMPLE1\_Irisk.out", for which a BMD and BMDL were calculated (these are not shown in this type of plot).

Figure 3: Generating a Dose-Response (*type="dr"*) plot for the Non-Fatal tumor model

```
gofplot_msw("fm_alvbr_1stage_auto", type="dr")
```

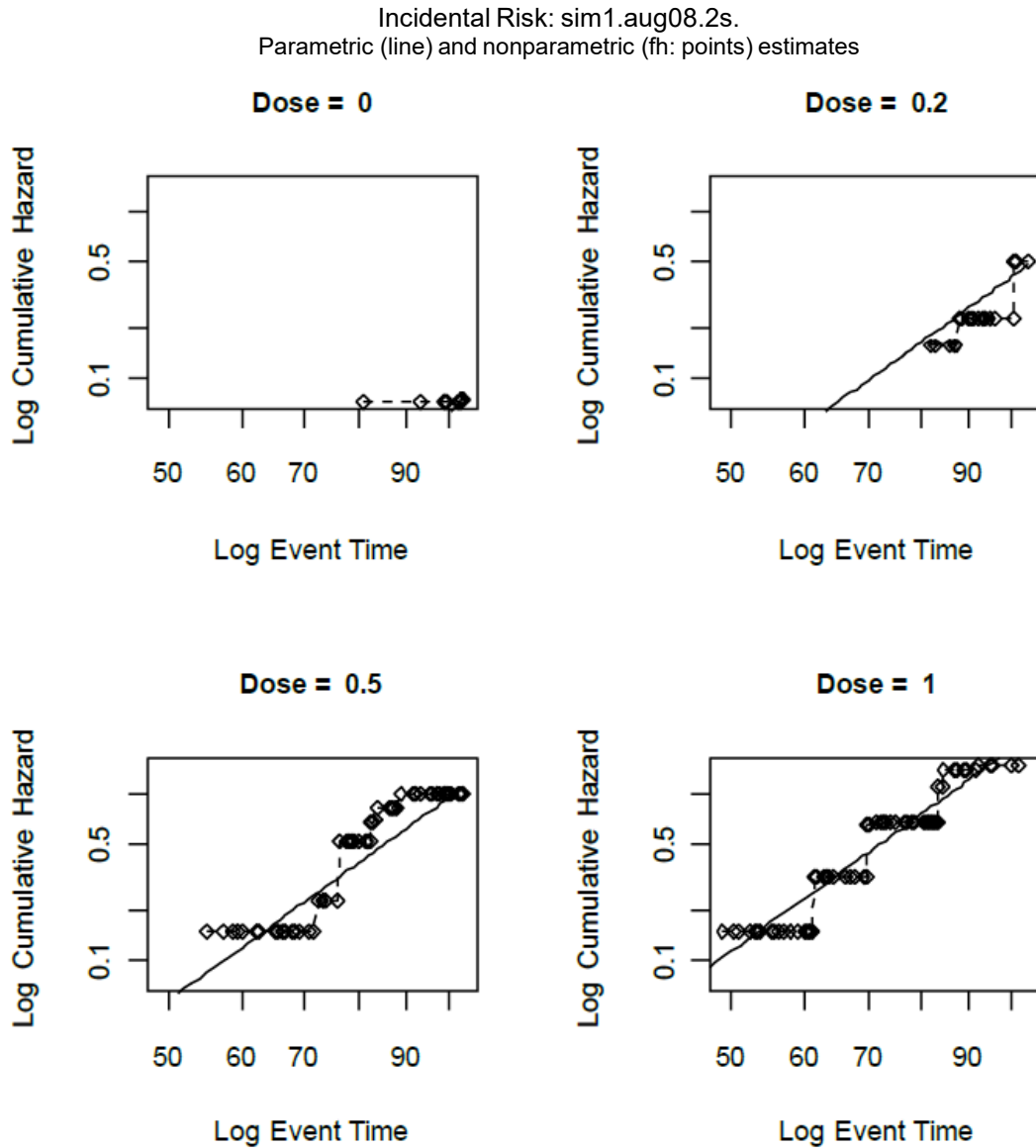
BMD for Incidental Risk at T = 104, Extra Risk level = 0.1, conf. level = 0.9 points  
show nonparametric estimate for nearest times at obsvd. doses



Legend: The solid line represents the fitted MSW model, and the filled points represent probabilities as determined from the nonparametric model. The latter must fall on observed times; the high dose group has no nonparametric estimate at 104 weeks, so the probability estimate at the time nearest 104 weeks (91 weeks) is plotted.

Figure 4: Generating a cumulative hazard (type="hz") plot for the Non-Fatal tumor model

```
gofplot_msw("sim1.aug08.2s", type="hz")
```

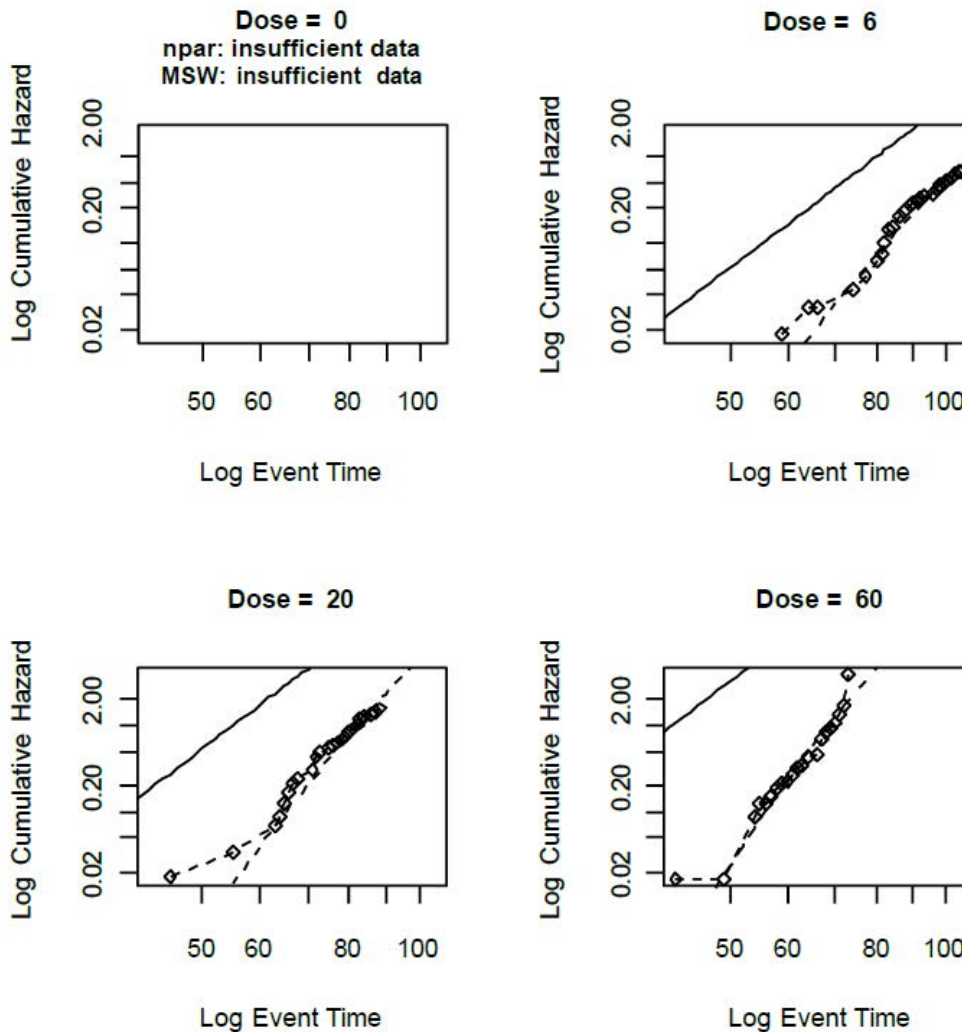


Legend: The solid line represents the MSW model (for which the log cumulative hazard is necessarily linear). The unfilled points represent log cumulative hazard values determined from the nonparametric model.

Figure 5: Generating a cumulative hazard (type="hz") plot for the Fatal tumor model

```
gofplot_msw("TCP_Fmouse_ASCT_grp_rnd_FI_2s", type="hz")
```

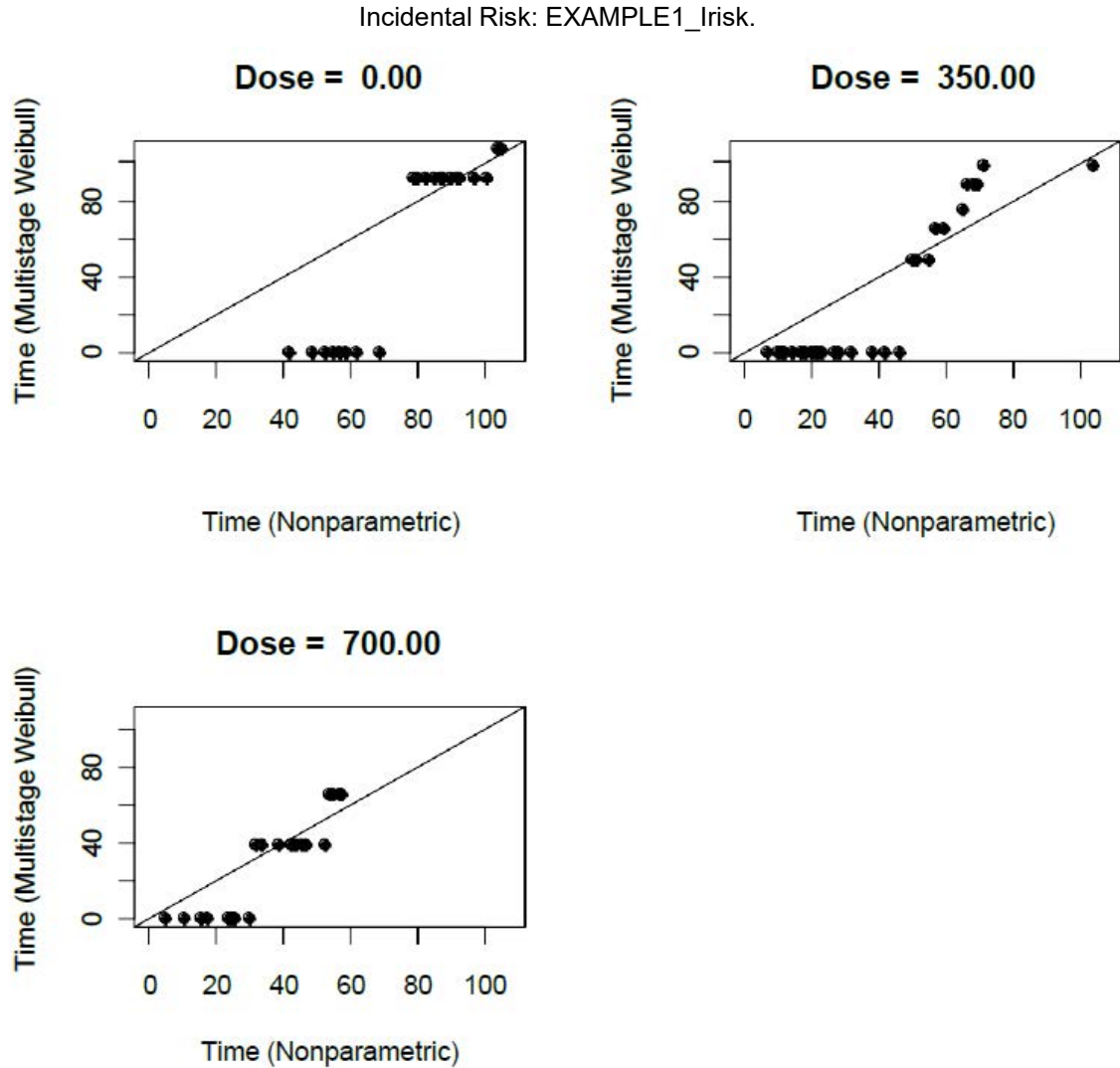
Fatal Risk: TCP\_Fmouse\_ASCT\_grp\_rnd\_FI\_2s  
Parametric (line) and nonparametric (km: points)  
estimates parametric shifted by  $t_0$  (dashed line).



Legend: Cumulative hazard plot for Fatal Risk. There were many more Fatal (F) observations and the estimate of  $t_0$  was 27 weeks. Because  $t_0 > 0$ , the MSW model for Fatal Risk was plotted at  $t+t_0$  as a dashed line, which is non-linear in this log plot. The solid line represents the MSW model without the time shift. On the log-log scale, the MSW model for fatal risk is not linear and it is not parallel to the model for Incidental Risk (solid line), which is necessarily linear in a (Weibull) cumulative hazard plot. The unfilled points represent log cumulative hazard values determined from the nonparametric Kaplan-Meier (Product-Limit) estimates. There is reasonably good agreement with the nonparametric estimates.

Figure 6: Generating a quantile-quantile (type="qq") plot for the Non-Fatal tumor model

```
gofplot_msw("EXAMPLE1_Irisk", type="qq")
```



**Legend:** The solid line represents an ideal 1:1 relation between parametric and nonparametric quantiles. The filled symbols represent the nonparametric estimates and the corresponding parametric estimates. The title "Incidental Risk" refers to the user-specified risk reported in the output file "EXAMPLE1\_Irisk.out."