iSQoL2 package
Version 6.1 (2022/12/22)
Tsuey-Hwa Hu

News for the version 6.1:

- Mean cost function before the extrapolation time point was estimated by the average of the monthly cost data for each month.
- New method for generating the survival function of the healthy reference group. (Lo WC, Hu TH, Hwang JS (2023)).
- The package: ffbase is unnecessary.
- Allow to convert the output file *.rds to *.csv.
- Update the abridged life tables of Taiwan from 1991-2021.

1. **Description**

The **iSQoL2** is a free **R** package under Microsoft Windows. It is written using the open source software **R** (version 4.2.2) and eight R packages (**data.table, ff**, **lokern**, **namespace**, **plotrix**, **rms**, **survival** and **tcltk**). To use **iSQoL2,** you need to install **R** and the seven **R** packages on your computer with Microsoft Windows. The next two sections describe how to install and start **iSQoL2** on **R**.

The **iSQoL2** provides the following functions:
(1) Major functions: The details are described in section 4.
- (A) To extrapolate survival function and estimate life expectancy (LE) for an index cohort with a specific disease or conditions.
- (B) To extrapolate mean cost function and estimate the expected lifetime cost (COST) for an index cohort when subjects' long-term records of medical expenditures are available.
- (C) To extrapolate mean QoL function and estimate the quality-adjusted life expectancy (QALE) for an index cohort. It needs a sample of subjects' repeated QoL measurements.

Note that the execution of the last two functions is based on the extrapolated survival function of (A). That is, the lifetime survival rate must be estimated before the execution of (B) and (C).
(2) Additional functions: The details are described in section 5.

The execution of all additional functions is based on the temporary output files of main function.
- ■ To compare life expectancy (LE), the expected lifetime cost (COST) or quality-adjusted life expectancy (QALE) between two groups.
- ■ To compare the expected loss-of-LE or loss-of-QALE between two groups.
- ■ To calculate cost-effectiveness ratio (CER).
- ■ To calculate incremental cost-effectiveness ratio (ICER).
- ■ Plot multiple curves as follows:
    - − Survival function, quality-adjusted survival function
    - − Loss function of LE or QALE
    - − Mean cost function

2. **How to install iSQoL2**

Please follow the steps below to install **iSQoL2** and some **R** packages.

Step1. Check whether **R** (version ≥4.0.x for Windows) is already installed on your computer. If not, please download **R** from https://cran.r-project.org/ and install it.

Step2. Download **iSQoL2_MyWork.zip** from http://sites.stat.sinica.edu.tw/isqol/ and unzip it to a folder (say, **C:/Project**) on your computer.

Step3. Open the file **C:/Project /iSQoL2_MyWork/Rpath.txt** to correct the location of **Rgui.exe** (the execution file for starting R). You can right-click **R** icon on your desktop and then click **Properties** to find the location displayed in the **Target** field.

Step4. Click **C:/Project /iSQoL2_MyWork/Ropen.bat** to start up **R**.

Step5. Execute the command **source('Start_iSQoL2.R')** in **R** console to make initial settings of **iSQoL2** and automatically install the required **R** packages (**data.table**, **ff**, **lokern**, **namespace**, **plotrix**, **rms**…) from http://cran.uk.r-project.org. In this step, you must have an internet connection.

Step6. Quit **R** to finish the installation, then see the next section for how to start **iSQoL2** on R.

NOTE:

After finishing the above steps, a folder named ".iSQoL2" will be created in the windows personal directory (typically "C:/Users/…/Documents/.iSQoL2"). It is also the default output folder of **iSQoL2**. The folder contains an initialization parameter file "InitialParameters.txt" and a folder "Example". The detailed descriptions of the file "InitialParameters.txt" are shown in Table 1; user can change the initialization parameters from **[iSQoL2/Update initial parameters]** menu. The folder "Example" contains 8 example data files and 3 input files. These files are listed below and the details of the file format are described in section 4.

⸺ 8 example data files:
- Data files of the index cohort I: AgeSurF120_1.txt, CostF120_1.txt, QolF120_1.txt.
- Data files of the index cohort II: AgeSurF120_2.txt, CostF120_2.txt, QolF120_2.txt.
- Average QoL in each age-sex group for the general reference population: Qol_ref1.txt and Qol_ref2.txt.

⸺ 3 input files: Each input file contains file names of data sets and parameters for running the major function of **iSQoL2**. The files inputS.csv, inputC.csv and inputQ.csv are the input file used to execute the estimation of LE, COST and QALE, respectively.

Table 1. The descriptions of the parameters file "InitialParameters.txt".

Parameter	Description
• ..OutDir = ""	• The output directory of the additional functions. The default is the windows "personal" directory (typically "C:/Users/…/Documents/.iSQoL2").
• ..GraphType = 8	• File format for saving the current figure. Default is Tagged Image File Format (tiff). ..GraphType = 1(wmf), 2(emf), 3(png), 4(jpg), 5(jpeg), 6(bmp), 7(tif), 8(tiff), 9(ps), 10(eps), 11(pdf)
• ..Xlab = "Time in months"	• Horizontal axis title of a plot.
• ..ConvertToCsv = FALSE	• Convert the output file *.rds to *.csv.
• ..OneWindow = 1	• Value is 0 for using multiple graphics devices to display different plots while you execute the major function. The default is equal to 1 for using single graphics device. Suggest that the value is 1 for running on multiple data sets at the same time.
• ..SMethod = 1	• The value is 1 or 2, respectively, for using Kaplan-Meier method or Weibull model to estimate survival function during follow-up time. Default is 1.
• ..n.knots = 5	• The number of knots for the rolling extrapolation algorithm based on the restricted cubic spline (RCS) model. Default is 5.
• ..FirstN.Sdel = 6	• The data from the first 6(=..FirstN.Sdel) time points are excluded for the first step of modeling logit[W(t)].

• ..TransformType=1	• The type of transformation for the ratio (W(t)=S_{ind}(t)/S_{ref}(t), $0 < W(t) < 1$) of the index's survival function (S_{ind}) to the reference's survival function (S_{ref}). There are two types listed below, and the default is the first type.
	  – ..TransformType is 1 for logit(.) transformation, i.e., use the relationship between logit[W(t)] and t to fit RCS model.
	  – ..TransformType is 2 for -log(-log(.)) transformation, i.e., use the relationship between -log[-log(W(t))] and log(t) to fit the RCS model.

• ..SepHRefSel2 = 36 ..SepSD = 0.5	• The two parameters are for estimating the hazard function ($h_H(t)$) of the healthy reference group, and available for the version ≥ 6.0. Where $h_H(t) = h_G(t) - (DM + ..\mathrm{SepSD} \times SD)$. The $h_H(t)$ is the hazard function of the general reference group. DM and SD are the average and standard deviation of the difference ($h_G(t) - h_I(t)$, $t=(F - ..\mathrm{SepRefSel2} + 1)$, …, F) of hazard between the

general reference and index group, respectively.

- ..SepHRefSel2: Number of time points used to estimate DM and SD. Default value is 36 months.
- ..SepSD: Initial value is 0.5.

- ..DiscountFromT = -1
- Time (t^{th} month) to start discount in estimating COST/QALE/LE. The value is a positive integer or -1. The default value is equal to -1 representing t = (1 + maximum follow-up time of the index cohort).

- ..LifetableType = 1
- Method for generating conditional survival function ($S_G(t|s, a)$) of the general references with sex=*s*, age=*a*. $S_G(t|s, a)$ can be expressed as

$$S_G(t|s, a) = S_G(t - 1|s, a) \times \left(1 - q_{a+t-1,s}^{y*}\right)$$

where $q_{a+t-1,s}^{y*}$ is the probability that someone with sex s and aged exactly $a + t - 1$ years old will die before reaching age $a + t$ based on the standard life table of year $y*$.

- When ..LifetableType = 0, set $y*$ equal to $q_{a+t-1,s}^{y}$.
- When ..LifetableType = 1, set $y*$ equal to $q_{a+t-1,s}^{y+t-1}$.

where y is the onset year. The default of ..LifetableType is 1.

3. How to start iSQoL2 on R

Suppose you have installed **iSQoL2** in the directory **C:/Project**. Follow the following steps to start **iSQoL2** on R.

Step1. Make sure the path in the file **C:/Project/iSQoL2_MyWork/Rpath.txt** is the same as the location of **Rgui.exe**. If not, **R** will not start up successfully.

Step2. Click **C:/Project/iSQoL2_MyWork/Ropen.bat** to start up **R**.

Step3. Execute the command **source("Start_iSQoL2.R")** in R console to create **iSQoL2** menu as shown in the figure below.
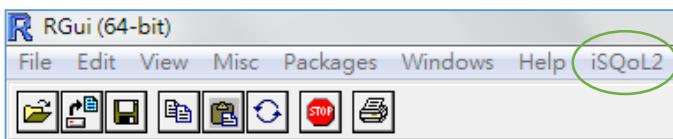
Then you can click **Run** within **iSQoL2** menu to start the estimate of life expectancy/lifetime cost/quality-adjusted life expectancy. The other commands within **iSQoL2** menu are described in Table 2.

Table 2. The command descriptions of **iSQoL2** menu.

Command	Description
Run:	*Major functions of **iSQoL2**. The details of the execution are described in Section 4.*
• (A) Survival extrapolation and LE	• To estimate the extrapolated survival function and LE. The temporary output file (*.rds) of this execution will be used for running (B), (C), and functions in **Additional functions** menu and **Discount for LE/QALE/COST** menu.
• (B) COST	• Use the extrapolated survival function to extrapolate the mean cost function and estimate COST. The temporary output file (*.rds) of this execution is a necessary file for running functions in **Additional functions** menu and **Discount for LE/QALE/COST** menu.
• (C) QALE	• Use the extrapolated survival function to extrapolate the mean QoL function and estimate QALE. The temporary output file (*.rds) of this execution is a necessary file for running functions in **Additional functions** menu and **Discount for LE/QALE/COST** menu.
Additional functions:	*The commands are applications based on the temporary output file (*.rds) of the major function ((A), (B) or (C)). The details are described in Section 5.*
• Comparison of LE/QALE/COST	• Compare LE, COST or QALE over a given time period between two groups. It is necessary to provide the temporary output files (*.rds) of (A), (B) or (C) for two groups.
• Comparison of the loss of LE/QALE	• Compare the loss of LE or QALE over a given time period between two

	groups. It is necessary to provide the temporary output files (*.rds) of (A) or (C) for two groups.
• CER	• Calculate the cost-effectiveness ratio (CER). It is necessary to provide the temporary output files (*.rds) of (B) and (A)/(C).
• ICER	• Calculate the incremental cost-effectiveness ratio (ICER). It is necessary to provide the temporary output files (*.rds) of (B) and (A)/(C) for two groups.
• Plot multiple curves	• Plot multiple curves for − survival function or quality-adjusted survival function − loss function of LE/QALE − mean cost function
Discount for LE/QALE/COST	• Use the given annual discount rate(s) to discount LE/QALE/COST estimated by major functions of **iSQoL2**. It is necessary to provide the temporary output files (*.rds) of (A), (C) or (B).
Close Graphics:	Close all graphics devices while too many R graphics devices in **R** console.
Update Initial Parameters:	Update the parameters described in Table 1. If you want to change any parameter value, click it to edit and update the settings of parameters.
Help:	Open iSQoL2Help.pdf.

4. How to execute the major function

4.1. How to execute the estimation of life expectancy (LE)

The following steps provide step by step on how to execute the extrapolation of survival function beyond the follow-up time and the estimation of LE.

Notation definition:
- F_s: The maximum follow-up time.
- F: The last month before starting extrapolation of survival function, $F \le F_s$.
- L: The time of ending extrapolation, $F \le F_s \le L$. When $L = F_s$, the system does not execute the rolling extrapolation algorithm.
- W(t): The ratio of the index's survival function ($S_{ind}(t)$) to the reference's survival function ($S_{ref}(t)$). The $S_{ref}(t)$ is often the general reference's survival function. If the index's survival function is larger than the general reference's survival, $S_{ref}(t)$ is the healthy reference's survival function generated by the system.
- H: The number of time points (months) in the last segment of RCS (restricted cubic spline) function for modeling the transformation of W(t).

Step1. **Prepare your data files with tab-delimited or space-delimited text format and save them in UTF-8 without BOM.**
   Suggest you use notepad++ to edit these files. The files are described below.
   - ■ **Survival, age and sex file**:
     — Five columns without specifying column names.
       - ➤ 1st column: The onset year (A.D.) of the index cohort.
       - ➤ 2nd column: The onset age (year) of the index cohort.
       - ➤ 3rd column: The gender code is 1 for male and 0 for female.
       - ➤ 4th column: The survival time (i.e. time from onset to death or last known alive, positive real value and in month) for the index cohort. The maximum follow-up time is denoted as F_s.
       - ➤ 5th column: The status indicator is 0 for censored case (alive at the end of follow-up) and 1 for completed case (dead during the F_s-month follow-up).
       - ➤ 6th column: The weights (>0) for each individual of the index cohort. When the biased estimator of survival function may be due to confounding factors, the weights will be used to estimate the adjusted survival function for the index cohort and the reference population. If the column is omitted, the system will assign an equal weight to all individuals of index cohort. Please see the reference: Xie J. and Liu C. (2005).

- The first three columns are used to generate the survival times of the reference population matched by the onset year, onset age and gender of the index cohort.
- The last three columns are used to estimate the survival function of the index cohort during the follow-up period.
- Two example files named "AgeSurF120_1.txt" and "AgeSurF120_2.txt" with 5000 people and 120 months follow-up are located in the directory "C:/Users/.../Documents/.iSQoL2/Example". The two data sets have 5 columns and an equal weight of each individual.

- **Create a directory containing annual life table files**: These files located in a directory are for creating the survival function of the reference population to provide assistance to survival extrapolation of the index cohort.
  - First you need to create a folder, then put the annual life table files in it.
  - The annual life table files need to be in the format of: &&****.txt. The && are two English letters. The **** is the 4-digit year (A.D.). e.g. TW1999.txt
  - Each file has 3 columns with no specified column names. The 1st column is age in years. The 2nd and 3rd columns are the probability of death within one year for male and female respectively.
  - When the death rate of a 110-year-old is missing, it is assumed to be 1.
  - The system will use these files to generate survival time for 100,000 individuals of the reference population which are matched by the onset year, onset age and gender of the index cohort.
  - For example, the abridged life tables of Taiwan from 1991-2021 are included in the folder "C:/Project/iSQoL2_MyWork/Rpackages/iSQoL2/LifetableTW".

Step2. **Create an input file in EXCEL and save it as a CSV file with comma-separated values**. The CSV file specifies the input files and parameters for executing the extrapolation of survival function. There are 7 columns and greater than or equal to 2 rows described below.
- Two or more rows:
  - 1st row: the column names.
  - 2nd row, 3rd row, etc.: Input parameters of the first index cohort, the second index cohort etc.
- 7 columns:
  - 1st column: Number (*NB*) of bootstrap samples for estimating standard error of LE. The value is integer and must be greater than and equal to 30. Set to 100 by default.

— 2nd column: The parameter, F ($F \leq F_s$), is the last month (since onset) before starting the rolling extrapolation algorithm. Set the default $F = F_s$ of survival data.

— 3rd column: The parameter, L, is the time of ending the extrapolation. The value is an integer and must be greater than or equal to F. That is, the rolling extrapolation algorithm will predict survival rate at times $F + 1, F + 2, ..., L$. When the survival rate at time L approaches 0.005, then L is suitable.

— 4th column: The parameter, H, is the number of time points in the last segment of the restricted cubic spline (RCS) function for modeling the transformation of the ratio ($W(t)=S_{ind}(t)/S_{ref}(t)$) of survival function. In other words, the transformation of W(t), $(F - H + 1) \leq t \leq F$, are restricted to a straight line. Set the default value to 12. The H is often between 12 and 24.

Note: The $(F - H + 1)$ is the last knot of RCS model.

— 5th column: The directory of the folder containing the annual life table files prepared in Step 1. The default is set as "C:/Project/iSQoL2_MyWork/Rpackages/iSQoL2/LifetableTW" (the abridged life tables of Taiwan).

— 6th column: The file with full path has onset year, onset age, gender and survival data of the index cohort prepared in Step 1.

— 7th column: The output directory for saving temporary output files and graph files. Note that the directories in each row must be different for different index cohorts.

■ If you want to make sure whether the values *L*, *F* and *H* are suitable, see Appendix for the details of the parameters selection.

■ An example file named "InputS.csv" is located at "C:/Users/…/Documents/.iSQoL2/Example/InputS.csv". It contains three rows, i.e., two datasets will be executed; the first row shows the column names. The settings for each column are shown below.

— *NB* (Number of bootstrap samples) = 100.

— *F* (the last month before starting extrapolation) = 120 months.

— *L* (Time of ending extrapolation) = 618 months.

— *H* (Number of months beyond the last knot) = 12.

— Use the abridged life tables of Taiwan area to generate the reference population.

— The file name of age-sex and survival data for the index cohort: "C:/Users/…/Documents/.iSQoL2/Example/AgeSurF120_1.txt" and "C:/Users/…/Documents/.iSQoL2/Example/AgeSurF120_2.txt".

— Output directory: "C:/Users/…/Documents/.iSQoL2/Example/Out1S" and "C:/Users/…/Documents/.iSQoL2/Example/Out2S".

Step3. **Begin extrapolating survival function and estimate life expectancy**.

(3.1) Click [**iSQoL2/Run/(A) survival extrapolation and LE**] menu to select an input file ("InputS.csv") to start the execution. Note that if the program unexpectedly stops executing, it may be caused by increasing the transformation of W(t) near the last few months of *F*. Thus, we suggest that you use a smaller value of *F* (say, *F* - 6) and re-run the step.

(3.2) After finishing the execution, a text output and four figures are displayed in **R** console (Depending on your setting, the four figures may replace one another as R draws them, so you may only be able to examine the last figure in **R** console). These, alongside a temporary file, are saved in the output directory that you set in the input file and described below.

— A text output: (*_sy.txt)

The results of the first data set shown as Table 4.1.1 contain the estimates, bootstrapped standard errors, bootstrapped mean, bootstrapped median and 95% bootstrap confidence intervals of LE from onset to the last month before starting extrapolation (Index_F) and to the time of ending extrapolation (Index_L) for the index cohort. The results also include the estimates for the general reference population from onset to the time of ending extrapolation (Ref_L).

Table 4.1.1 The text output of the estimated LE is shown in **R** console.

— Figures:
- SurvF.*: The estimated survival probability of the index cohort and the general reference population from onset to the given *F*.
- LogitWF.*: The transformation of W(t) from onset to the given *F* and the prediction curve of the restricted cubic spline (RCS) regression model. If the

index's survival function is larger than the general reference's survival function, the denominator of W(t) is the healthy reference's survival function.

- LogitWL.*: The transformation of W(t) from onset to the given *L*. If the index's survival function is larger than the general reference's survival function, the denominator of W(t) is the healthy reference's survival function.
- SurvL.*: The estimated survival probability of the index cohort and the general reference population from onset to the given *L*.

— A temporary file:

- The file named "SurvivalOnly.rds" is a serialized object that has a list of **R** objects contained (*NB*+1) estimated survival functions of the index cohort and the general reference population, etc. User can use the command "readRDS" to get it in **R**. Note that the file will be necessary for the estimation of COST (Section 4.2), QALE (Section 4.3), the execution of additional functions (Section 5) and the discount for LE/QALE/COST (Section 6).

4.2. **How to execute the estimation of lifetime cost (COST)**

The following steps provide step by step on how to execute the extrapolation of mean cost function and the estimation of COST for the index cohort. The method uses the estimated survival rate from onset ($t=0$) to time of the ending extrapolation ($t=L$) to extrapolate the mean cost function, and then estimate COST using the extrapolated survival function and mean cost function. Therefore, user needs to execute the estimation of LE before executing the estimation of COST.

Notation definition:
- F_c: The maximum observed time of the costs since onset.
- F: The last month before starting extrapolation of the mean cost function, $F \leq F_c$.
- L: The time of ending extrapolation, $F \leq F_c \leq L$. The parameter has been decided on the input parameter L of the estimated LE.
- K and B: The parameters for extrapolating the mean cost function.

Step1. **Prepare your data files with tab-delimited or space-delimited text format and save them in UTF-8 without BOM**.

Suggest using notepad++ to edit these files. The files are described below.

- **Cost file**:

The file has (F_c +2) columns and no column names.

— 1st column: The value is the survival time from the onset to the time when the subject died during the F_c-month period, otherwise input -1 for censored subjects.

— 2nd to the last column: Monthly accumulated cost per subject at each month from the onset to F_c. The costs at t=0, 1, 2, …, F_c are recorded in the second, third, fourth, …, (F_c +2)th column, respectively. Each row indicates a subject's monthly costs. Note three things:

(a) When a subject has no medical cost in any 1-month period during the follow-up time, the monthly cost is set to 0.

(b) The cost values beyond death or loss of follow-up are given -1 (or a negative integer).

(c) Total costs associated with onset event before the onset can be summed into the cost at onset (recorded in the second column). If a subject had no extra cost record before onset, input 0 in the second column.

— The example file named "costF120.txt" is saved in the directory "C:/Users/…/ Documents/.iSQoL2/Example". It is the collected monthly costs of 5000 individuals with F_c =120 months.

■ **A temporary output file of the estimated LE**:

The temporary file named "SurvivalOnly.rds" is an output file of the execution in the section 4.1.

⸺ The example is "C:/Users/.../Documents/.iSQoL2/Example/Out1S/Temporary/SurvivalOnly.rds".

Step2. **Create an input file in EXCEL and save it as a CSV file with comma-separated values**.

The CSV file contains data filenames and parameters for executing the estimation of lifetime cost. There are 7 columns and greater than or equal to 2 rows described below.

■ Two or more rows:

⸺ 1st row: the column names.

⸺ 2nd row, 3rd row, ... : File names of data files and the input parameters of the first index cohort, the second index cohort, etc.

■ 7 columns:

⸺ 1st column: Number (*NB*) of bootstrap samples for estimating the standard error of lifetime cost. The value is integer and greater than or equal to 30. Set 100 as the default value.

⸺ 2nd column: The parameter, *F*, is the last month before starting to extrapolate mean cost function. Set $F = F_c$ as the default value.

⸺ 3rd and 4th column: The parameters, K and B, are for extrapolating the mean cost function estimated by the average costs $\overline{C_k}$ at $k = 0, 1, ..., K, K^+$ months. The average costs $\overline{C_k}$ are calculated by the observed costs of subjects who died or alive during $(F - B + 1)^{th}$ and F^{th} month. The value of B is a positive integer, and K is -1 or 0 or a positive integer. When K=-1, the extrapolated mean cost function is assumed constant over time. Set K=12 and B=12 as the defaults.

⸺ 5th column: Cost file name with full path for the index cohort.

⸺ 6th column: An output directory of the estimated LE described in the section 4.1.

⸺ 7th column: Output directory for each index cohort is used to save the output files and graph files. Note that the directories must be different for each data set (i.e. each row of the input file "InputC.csv").

■ An example file "C:/Users/.../Documents/.iSQoL2/Example/InputC.csv" contains three rows. That is, two data sets will be executed. The settings are given below.

⸺ *NB* (Number of bootstrap samples) = 100 for two data sets.

⸺ $F = F_c$ =120 months for the two data sets.

⸺ The parameters *K*=12 and *B*=12 for the two data sets.

⸺ Annual discount rate = "0" for the two data sets

⸺ The file names of cost data are "C:/Users/.../Documents/.iSQoL2/Example

/costF120_1.txt" and "C:/Users/…/Documents/.iSQoL2/Example /costF120_2.txt".

— Two output directories of the estimated life expectancy are "C:/Users/…/Documents/.iSQoL2/Example/Out1S" and "C:/Users/…/Documents/.iSQoL2/Example/Out2S".

— Output directories for saving the output files are "C:/Users/…/ Documents/.iSQoL2/Example/Out1C" and "C:/Users/…/ Documents/.iSQoL2/Example/Out2C".

Step3. **Start to extrapolate mean cost function and estimate COST**.

(3.1.) Edit the input file "C:/Users/…/.iSQoL2/Example/InputC.csv" mentioned above to correct the values of F, B and K if necessary. For example, replace K with 24 for the second data set.

(3.2.) Click [**iSQoL2/Run/(B) COST**] to select the input file "InputC.csv" to start the execution.

(3.3.) After finishing the execution, a text output and figures are shown in **R** console and saved in the output directory specified in the file "InputC.csv". A temporary file name Cost*.rds is also saved in the specified output directory. The details are described below.

— A text output:

The results shown in Table 4.2.1 contain the estimate, bootstrapped standard errors, bootstrapped mean, bootstrapped median and 95% bootstrap confidence intervals of the expected cost from onset to the given F and from the onset to the end of extrapolation (L) for the index cohort. The results also show the estimate of LE.

Table 4.2.1 The text output of the estimated lifetime cost is shown in **R** console.

— Figures:

- CjB*K*.*: The average cost in K months prior to death and the average cost of subjects who won't die in K months for the given K and B.
  For example, the figure is shown as Figure 4.2.1(A) for (B, K) = (12, 12). It is saved to a file named "CjB12K12.tiff" and its values are saved in the file "CjB12K12.txt" for the raw data (Bootstrapped=0).
- MeanCostF.*: The mean cost function over time from onset to F (the previous month before starting extrapolation) is plotted as Figure 4.2.1(B).
- MeanCostE.*: The mean cost function over time from onset to the end of extrapolation (L) is plotted as Figure 4.2.1(C).

— A temporary file:

- A file named "CostRate*.rds" is saved in the folder named "Temporary". It is a serialized object that has a list of **R** object contained the mean cost function and survival-weighted mean cost function from onset to the end of the extrapolation for the index cohort, etc. This file will be used for the execution of the section 5 and 6.

Figure 4.2.1 The output figures after the execution of the COST.
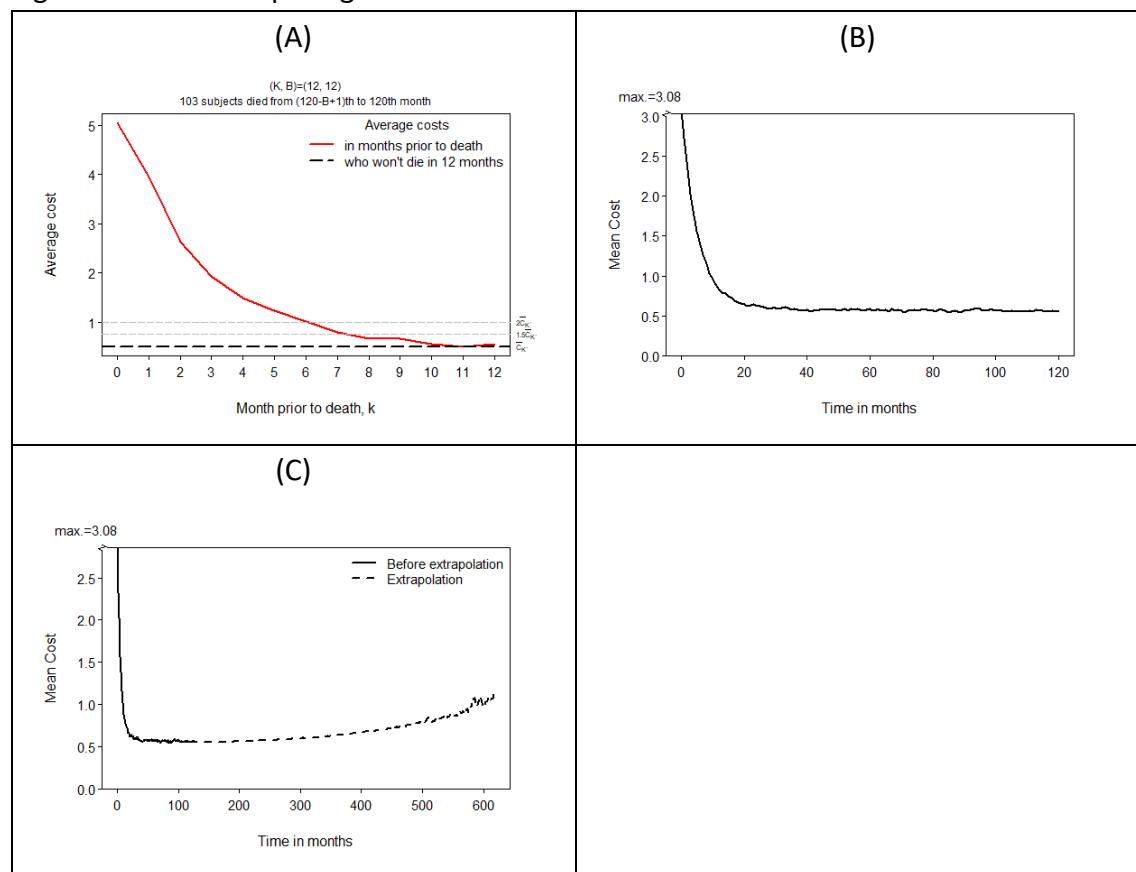
4.3. **How to execute the estimation of quality-adjusted life expectancy**

The following steps provide step by step on how to execute the extrapolation of mean QoL function and the estimation of QALE (quality-adjusted life expectancy, also called the expected QAS).

The method uses survival rate from onset (t=0) to time of ending extrapolation (t=L) to extrapolate the mean QoL function, and then estimate the quality-adjusted life expectancy with survival rate and mean QoL function. Therefore, user needs to execute the estimation of LE described in Section 4.1 before the execution of QALE.

Notation definition:
- F_q: The maximum assessment time of QoL.
- L: The time of ending extrapolation, $F_q \leq L$.
- b: The bandwidth of the kernel smoothing function in order to estimate the mean QoL function from the diagnosis time to the maximum assessment time.
- K and B: The parameters for extrapolating the mean QoL function.
- F_b (=F): The previous month before starting extrapolation, is equal to (100-b)th percentile of all assessment times of QoL.

Step1. **Prepare your data files with tab-delimited or space-delimited text format and save them in UTF-8 without BOM**.
- ■ **QoL data file of an index cohort**:
  The file has five columns and the column names recorded in the first row. The descriptions below are the definition of each column.
  — 1st column: ID number for each individual with QoL measurements of a sample from an index cohort. The values are positive and integer.
  — 2nd column: The survival time (i.e. month from onset to death or last known alive) for each individual of an index cohort.
  — 3rd column: The status indicator at the end of follow-up time is 0 for censored case (alive) and 1 for completed case (dead).
  — 4th column: Time ($\geq$ 0, month) from onset to the assessment time of the QoL. Each individual in the sample has one or more assessment time points from onset to the end of follow-up time. The maximum assessment time is denoted F_q.
  — 5th column: The QoL value at the assessment time. The value is in the closed interval [0, 1]. Each individual has one or more QoL scores from onset to the end of follow-up time.

Two example files are "QolF120_1.txt" and "QolF120_2.txt" saved in "C:/Users/…/ Documents/.iSQoL2/Example". Each individual has one or more QoL scores for the two data sets. The column name of QoL recorded at 5th column is "QoL1".

■ **QoL data file of the general reference population**:
The file has three columns and the column names have to be given in the first row. Each row consists of sex, age and the average QoL of the age-sex group for the general reference population. The descriptions below are the definition of each column.

— 1st column: The gender code of the general reference population is 1 for male and 0 for female.

— 2nd column: Age (year) for the general reference group.

— 3rd column: The average QoL utility (e.g., EQ-5D) of each age-sex group is given at this column. The column name must be the same as the column name of the 5th column in QoL data file of the index cohort.

Note: When each age-sex group of the general reference population is assumed having the same average QoL, set the value of the first and second column at the second row to -1 and the average QoL in the 3rd column.

Two example files are "qol_ref1.txt" and "qol_ref2.txt".

• The first file has one QoL measurements named "QoL1"; every general reference individual has the same average QoL score recorded 0.9 for "QoL1".

• The second file has also one QoL measurement named "QoL1", but each age-sex group has different average QoL score.

■ **Survival, age and sex data file of an index cohort:**
The file is the survival and age-sex data file described in Section 4.1. It is used to generate QoL scores of the general reference population matched by onset age and gender of an index cohort.

Note: When each age-sex group of the general reference population has the same average QoL, user does not need to provide it. For example, the QoL data file of the general reference population is set as the file "qol_ref1.txt", and then the age-sex and survival data file of the index cohort is not needed.

■ **A temporary output file of the estimated LE**:
The temporary file named "SurvivalOnly.rds" is an output file of the execution in the section 4.1.

— An example is "C:/Users/…/Documents/.iSQoL2/Example/Out1S/Temporary/SurvivalOnly.rds".

Step2. **Create an input file in EXCEL and save it as a CSV file with comma-separated values**.
The CSV file contains data filenames and parameters for preparing the estimation of QALE. There are 10 columns and greater than or equal to 2 rows.

- ■ Two or more rows:
  - ⎯ 1st row: the column names for each column.
  - ⎯ 2nd row, 3rd row, ...: File names of data files and the input parameters for the first index cohort, the second index cohort, etc.

- ■ 10 columns:
  - ⎯ 1st column: Number (*NB*) of bootstrap samples for estimating the standard error of QALE. The value is integer and greater than or equal to 30. Set 100 as the default value.
  - ⎯ 2nd column: The parameter is called "bandwidth" and denoted "*b*". It is specifying the width of the kernel smoothing function in order to estimate the mean QoL function from the diagnosis time to the maximum assessment time (F_q). It is a percentage value (%), often between 1 and 5.
  - ⎯ 3rd and 4th column: The parameters, *K* and *B*, are for extrapolating the mean QoL function. Set K =12 and B = 36 as the defaults. When the extrapolated mean QoL is assumed constant over time, the value of *K* is set to -1.
  - ⎯ 5th column: The mean QoL for dead individuals of the index cohort. Default is 0.
  - ⎯ 6th column: The filename with full path recorded the QoL scores of the index cohort.
  - ⎯ 7th column: The filename with full path recorded the average QoL score at each age-sex group of the general reference population.
  - ⎯ 8th column: The filename with full path recorded age-sex and survival data of the index cohort. When the filename in 7th column has the same average QoL at each age-sex group of the general reference population, set the value in this column to the blank value.
  - ⎯ 9th column: An output directory of the estimation of LE described in Section 4.1.
  - ⎯ 10th column: Output directory for saving the text outputs and graphs of the estimated QALE. When the input file (InputC.csv) has two or more rows, the values at this column must be specified the different directories.

- ■ An example file "C:/Users/…/Documents/.iSQoL2/Example/InputQ.csv" contains three rows, i.e., two data sets will be executed.
  - ⎯ The first dataset is assumed to have the same average QoL of the general reference individuals, and then the value of 8th column in the second row is set to blank value.
  - ⎯ The second data set has different average QoL at each age-sex group of the general reference population, and then the value of 8th column in the third row

must be provided the filename recorded age-sex and survival data of the index cohort.

Step3. **Start to extrapolate mean QoL function and estimate QALE**

(3.1.) Edit the input file "C:/Users/…/.iSQoL2/Example/InputQ.csv" to change b, B and K with suitable values if necessary. For example, replace K with 8 for the first data set.

(3.2.) Click [**iSQoL2/Run/(C) QALE**] menu to select the input file "InputQ.csv" to start execution.

(3.3.) After finishing the execution, a text output, figures and a temporary file named *.rds are shown in **R** console and saved in the output directory specified at the last column of the file "InputQ.csv".

— A text output:

The results shown in Table 4.3.1 contain the estimates, bootstrapped standard errors, mean, median and 95% confidence intervals of QALE from onset to F_q and L for the index cohort. The results also include the estimates followed to the extrapolation time for the general reference population. The summary table is saved to a file named "*_sy.txt".

Figure 4.3.1 The text output of the estimated QALE is shown in **R** console.

- Figures:
  - *B*DK*.tiff: The average QoL in *K* months prior to death and the average QoL of subjects who won't die in *K* months. In the example, the figure is shown as Figure 4.3.2(A) for (*B*, *K*) = (36, 8).
  - *_MeanF.tiff: The mean QoL function with *b*=5% over time from onset to F_q (maximum assessment time) is plotted as Figure 4.3.2(B).
  - *_MeanE1.tiff: The mean QoL function over time from onset to the end of extrapolation for the index cohort is plotted as Figure 4.3.2(C).
  - *_MeanE2.tiff: The mean QoL functions over time from onset to the end of extrapolation for the index cohort and the general reference population are plotted as Figure 4.3.2(D).
  - *_E1.tiff: The figure as Figure 4.3.2(E) includes survival function, mean QoL function and the quality-adjusted survival function over time from onset to the end of extrapolation for the index cohort.
  - *_E2.tiff: The figure as Figure 4.3.2(F) includes the quality-adjusted survival functions over time from onset to the end of extrapolation for the index cohort and the general reference population.
- A temporary file:
  - A file named "QAS_*.rds" is saved in a folder "Temporary" of the specified output directory. It is a serialized object that has a list of **R** object contained the extrapolated quality-adjusted survival function for the index cohort and the general reference population, etc. This file will be necessary and used for the execution of the section 5 and 6.

Figure 4.3.2 The output figures after the execution of the expected quality-adjusted life expectancy.
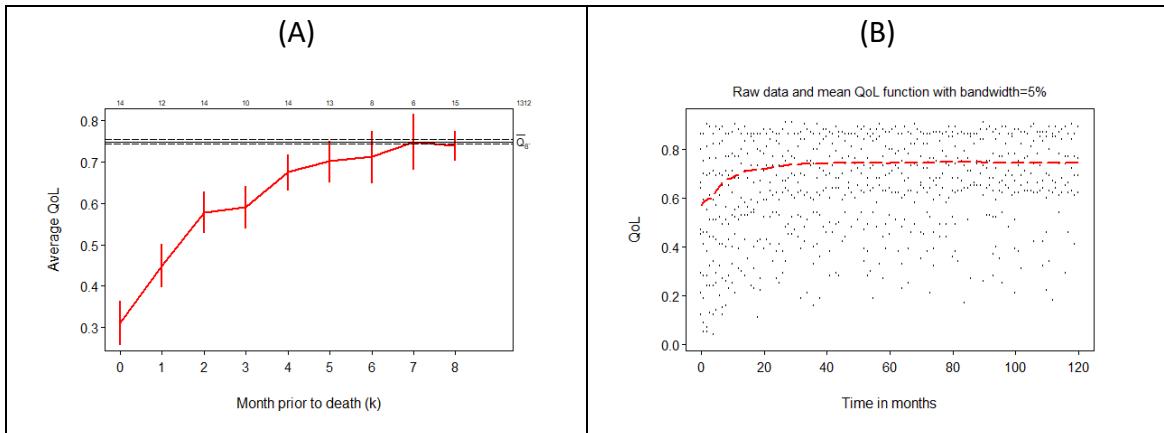
Figure 4.3.2 The output figures after the execution of the expected quality-adjusted life expectancy.
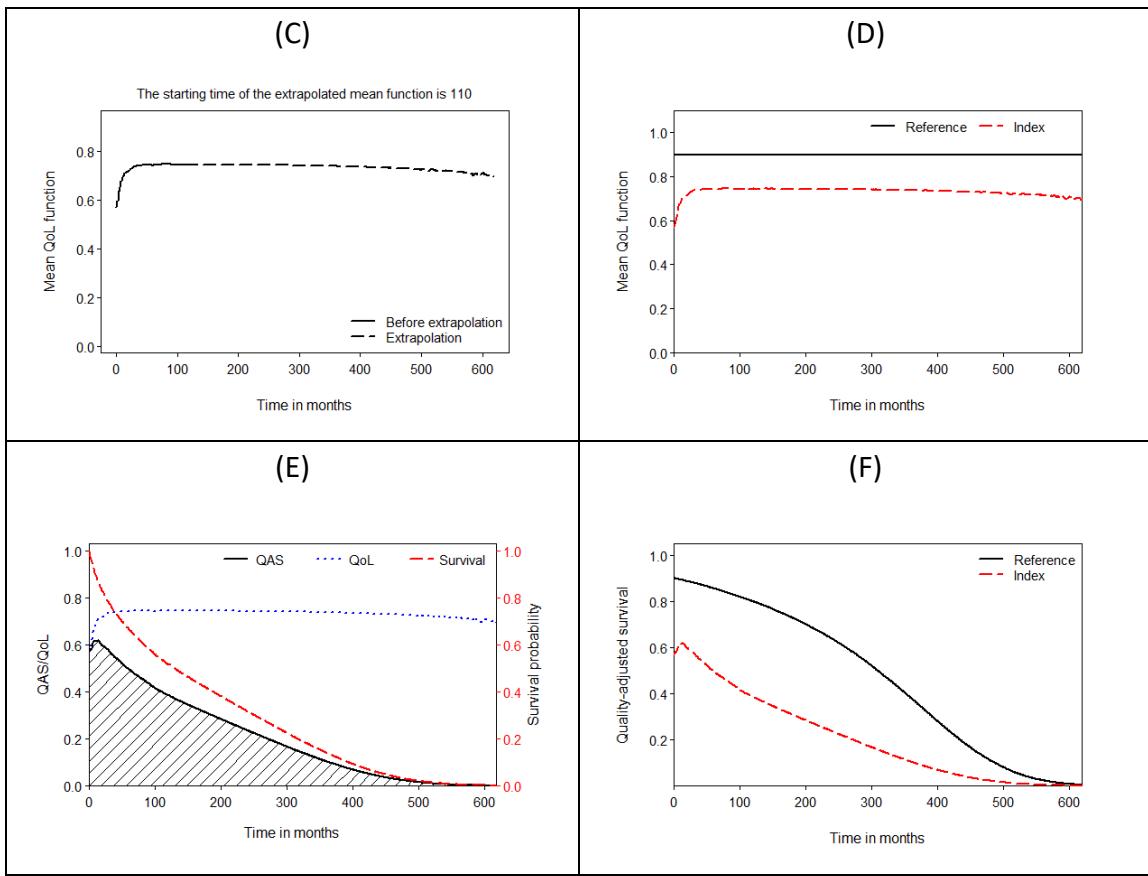
5. Additional functions

The following functions are executed based on a temporary output file (*.rds) of the major function. Before executing, you need to provide the temporary file *.rds as the input file.

5.1 The expected difference of LE/QALE/COST between two groups

The function is used to compare the expected survival (expected cost/expected quality-adjusted survival) over the given time period between two groups. It also computes the expected years of life lost (EYLL) and loss-of-QALE.

Figure 5.1 Input window for executing the expected difference between two groups.

(A)   (B)
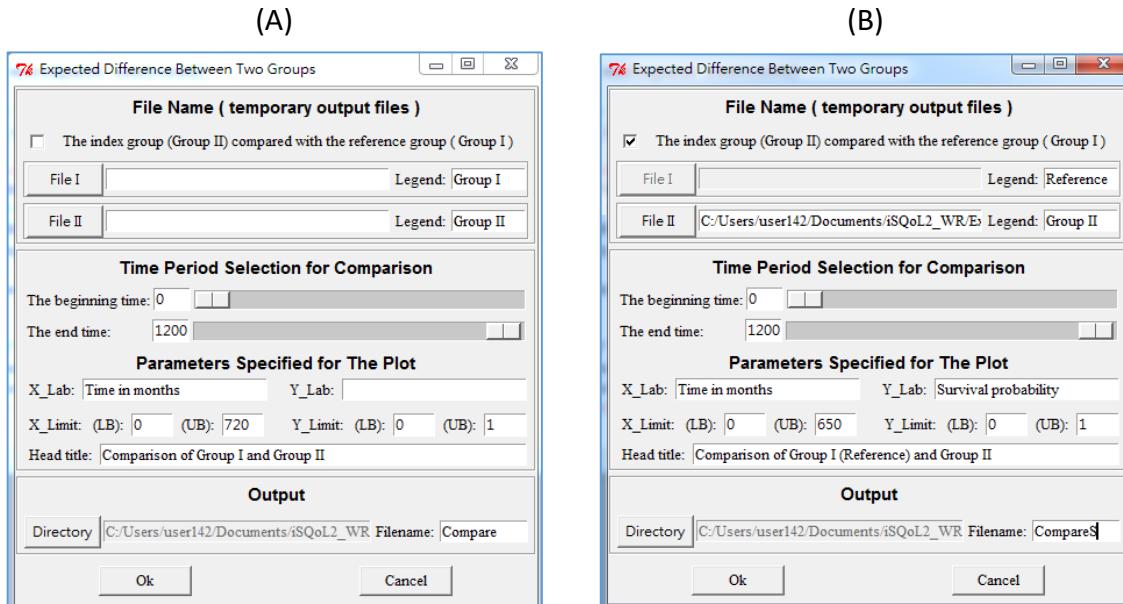
The steps below describe how to run the function.

Step1. Click [**iSQoL2/Additional Functions/comparison of LE/QALE/COST**] to open an input window as Figure 5.1(A).

— Click [**File I**] button and [**File II**] button to select two temporary files named like "Survival*.rds" , "CostRate*.rds" or "QAS*.rds".

— If tick the checkbox [**The index group (Group II) compared with the reference group (Group I)**] and select a temporary file (*.rds) for [**File II**], the result is the expected difference between the index cohort and the general reference population.

— **To compute loss-of-LE (EYLL):** Tick the checkbox [**The index group (Group II) compared with the reference group (Group I)**]; click [**File II**] button to select a temporary file named "Survival*.rds" as Figure 5.1(B); set the beginning time and the end time as 0 and 1200 respectively.

- **To compute loss-of-QALE:** Tick the checkbox [**The index group (Group II) compared with the reference group (Group I)**]; click [**File II**] button to select a temporary file named "QAS*.rds" as Figure 5.1(B); set the beginning time and the end time as 0 and 1200 respectively.

Step2. Fill in all inputs and then click **Ok** to execute the estimation. The results contain the summary of the estimates and a figure.

- The summary outputs in **R** console contain the expected difference, the bootstrapped standard error, p value of Z-test statistics, and the percentile of the bootstrap estimates (month-unit and year-unit). The outputs also provide the estimated LE/QALE/COST (month-unit and year-unit) which is the output of major function for two groups. The results as Table 5.1 are shown in **R** console and saved to a file specified in the input window.

Table 5.1 Summary output for the comparison of LE between the index cohort and the general reference population.

```
The difference of the expected survival(Group I - Group II)
over the time period [0, 618] between two groups:
-------------------------------------------------------------------------
$`Difference:`
   Time_Period Estimate     SE z-statistic p-value    50%   2.5%  97.5%
  month: 0-618   160.99   5.90       27.29    0.00 159.74 146.74 168.92
   year: 0-51.5   13.42   0.49       27.29    0.00  13.31  12.23  14.08


$`The expected survival during time period [0, 618]:`
          Group1(month) Group2(month) Group1(year) Group2(year)
Estimate         325.13         164.1       27.094        13.68
SE                 0.91           5.8        0.076         0.49
50%              325.00         165.5       27.083        13.79
2.5%             323.27         155.6       26.939        12.97
97.5%            326.83         178.4       27.235        14.87



$Inputs:
 ***              Group I: Reference (expected survival)
 ***             Group II: Group II (expected survival)
 ***               File I:
C:/Users/user1/Documents/.iSQoL2/Example/Out1S/Temporary/SurvivalOnly.
 ***              File II:
C:/Users/user1/Documents/.iSQoL2/Example/Out1S/Temporary/SurvivalOnly.
 ***        Extrapolation: 618 (Group I), 618 (Group II) months
 *** Number of bootstrap: 100 (Group I), 100 (Group II)
 ***              Outfile: C:/Users/user1/Documents/.iSQoL2/Compare.txt

Notes:
(1) Group I is the reference population. If the extrapolation time L is the end of life,
    the estimated EYLL (expected years of life lost) is 12.187 years (=146.25 months).
(2) Y-axis label: Survival function.
```

The example in Table 5.1 which sets Group II as the general reference group is for

calculating EYLL. Thus, the expected difference is called the expected years of life lost (EYLL) and equal to 12.17 years. The results also show that the estimates of LE are 26.214 and 14.047 years for the general reference population and index cohort, respectively.

— The figure includes two curves for the index cohort and the general reference population as shown in Figure 5.2. It is shown in **R** graphics device and saved to a user-specified file.

Figure 5.2 The survival function between the index cohort and the general reference population.
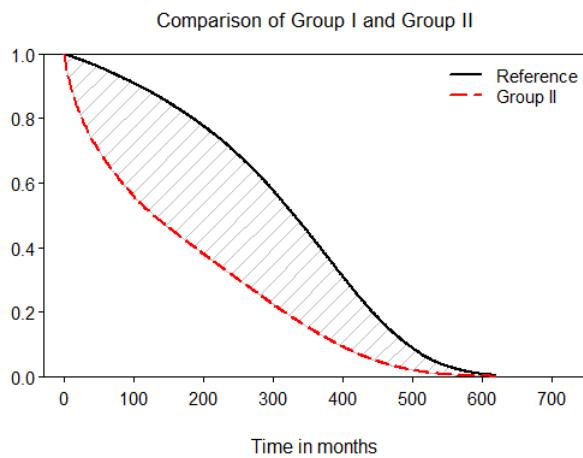
[Appendix 5.1] Estimate of the expected difference

Let X_0 and Y_0 be the estimates calculated with the raw data of Group I and II. $(X_1, \ldots, X_R)$ and $(Y_1, \ldots, Y_R)$ be the estimates calculated with R bootstrapped data sets for Group I and Group II, respectively. Then, the difference between two groups is equal to $D_i = X_i - Y_i$, where i =0, 1, …, R. Thus, the values in the table of the expected difference are below.

- The "Estimate" is equal to D_0.
- The "SE" is the standard deviation of the estimates D_0, D_1, …, D_R.
- The "z-statistic" ("Estimate" divide by "SE") and "p-value" are used to determine whether the difference is significant.
- 50%, 2.5% and 97.5% are 50th, 2.5 th and 97.5 th percentile of the estimates D_0, D_1, …, D_R, respectively. The 2.5 th and 97.5 th percentile is the lower bound and upper bound of the 95% confidence interval.

5.2 Comparison of the loss of LE/QALE between two groups

The function is used to compare the loss of the expected survival (or quality-adjusted survival) over the given time period between two groups. If the comparison from the diagnosis to the end of life, the estimated difference is the difference of EYLL (or loss-of-QALE) between two groups (i.e. Difference = $EYLL_1 - EYLL_2$ = [loss-of-LE_1] − [loss-of-LE_2] or Difference = [loss-of-$QALE_1$] − [loss-of-$QALE_2$]).

The execution steps are shown below.

Step1. Click [**iSQoL2/Additional Functions/comparison of the loss of LE/QALE**] to open an input window as the left-side of Figure 5.2.

Step2. Fill in all inputs in the inputs window.
- ⸺ Select two temporary files named "Survival*.rds" (or "QAS*.rds") by clicking [**File I**] button and [**File II**] button.
- ⸺ Set [**The beginning time**] and [**The end time**] to 0 and 1200, respectively. Then, the result is the difference of the loss from onset to the end of extrapolation.

Step3. Then click **Ok** to get a figure and the estimates. The figure as the right-side of Figure 5.2 shows two loss functions of Group I and II. The estimates (month-unit and year-unit) contain the difference, a bootstrapped standard error (S.E.) of the difference and 2.5th, 50th and 97.5th percentiles of the bootstrapped differences. The result also provides the Z-test statistic and p-value for testing whether the difference between two groups is significant. The additional results are loss-of-LE (or loss-of-QALE) with month-unit and year-unit for two groups.

Figure 5.2 Inputs window for the comparison of the loss-of-LE (or loss-of-QALE).
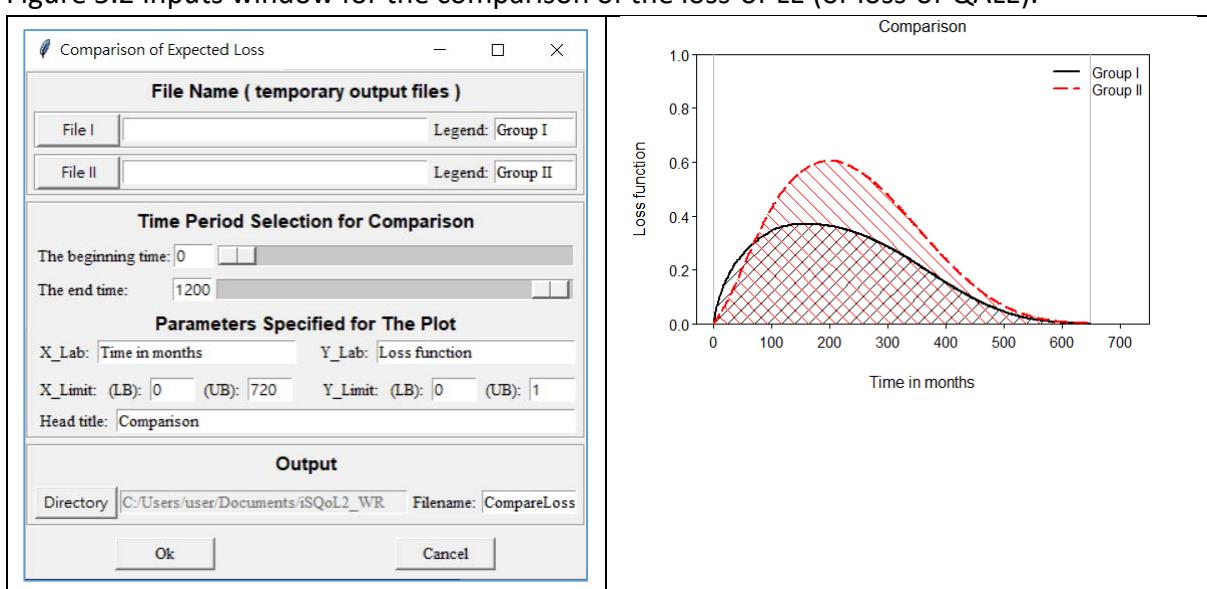
5.3 CER

The function is used to calculate the cost-effectiveness ratio (CER) which is the ratio of COST to LE (or QALE). The results are the estimates of CER, a bootstrapped standard error (S.E.) of CER and the 2.5th, 50th and 97.5th percentiles of the bootstrapped CERs.

How to execute the function?
⸺ Click [**iSQoL2/Additional Functions/CER**] to open an input window shown in Figure 5.3.
⸺ Fill all inputs and then click **Ok** button to start the execution. Two things need to be noted.
   (1) Select a temporary file named "CostRate*.rds" from clicking on [**COST**] button.
   (2) Select a temporary file named "Survival*.rds" or "QAS*.rds" from clicking on [**QALE/LE**] button
⸺ All results are shown in **R** console and saved to a file specified in the input window.
   • Text outputs include the estimated CER, a bootstrapped S.E. and the 2.5th, 50th and 97.5th bootstrap percentiles. (month-unit and year-unit)

Figure 5.3 Inputs window for the execution of CER.
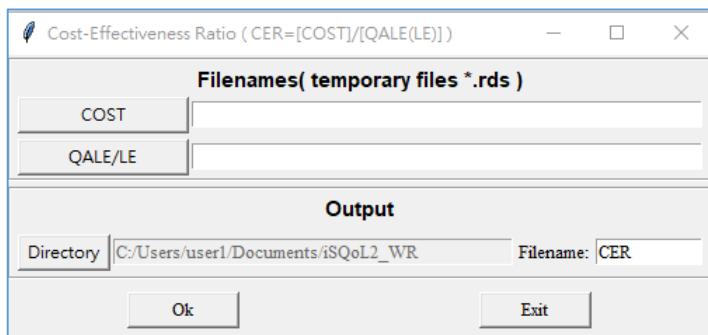
5.4 ICER

The function is used to calculate the incremental cost-effectiveness ratio (ICER). The CER is defined by $[Cost_2 - Cost_1]/[LE_2 - LE_1]$ or $[Cost_2 - Cost_1]/[QALE_2 - QALE_1]$ for two groups. The execution results contain the estimates of ICER, a bootstrapped standard error (S.E.) of ICER and the 2.5th, 50th and 97.5th percentiles of the bootstrapped CERs.

How to execute the function?
— Click [**iSQoL2/Additional Functions/ICER**] to open an input window as shown in Figure 5.4.
— Fill all inputs and then click **Ok** button to start the execution. Three things need to be noted.
  - [**COST I**] and [**QALE/LE I**] are for the estimated lifetime cost and life expectancy (or the expected quality-adjusted survival) of the first index cohort, respectively. [**COST II**] and [**QALE/LE II**] are for the second index cohort.
  - Select a temporary file named "CostRate*.rds" from clicking on [**COST ***] button.
  - Select a temporary file named "Survival*.rds" or "QAS*.rds" from clicking on [**QALE/LE ***] button.
— The text outputs of the estimation and two figures are shown in **R** console and saved in a file specified in the input window.
  - Text outputs include the estimated ICER, a bootstrapped S.E. and the 2.5th, 50th and 97.5th bootstrap percentiles. (month-unit and year-unit)
  - One figure shows two survival functions (or quality-adjusted survival function) of two groups. The other figure shows the survival-weighted mean cost functions (the product of the survival function and the mean cost function) for two groups.

Figure 5.4 Inputs window for the the execution of ICER.
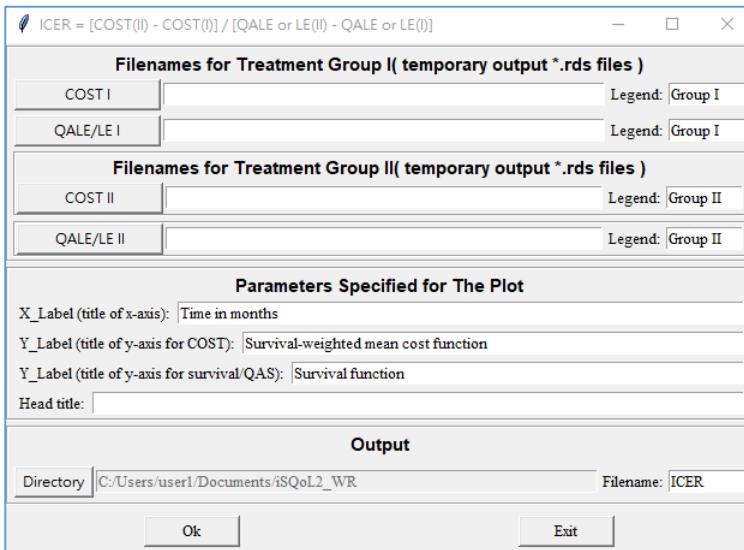
5.5 Plot multiple curves

The function is used to plot multiple curves in a graphics device. The multiple curves are one of the following estimated functions.

—   Survival functions.
—   Quality-adjusted survival functions.
—   The loss functions.
—   The mean cost functions.

Figure 5.6 Inputs window for plotting the multiple curves.
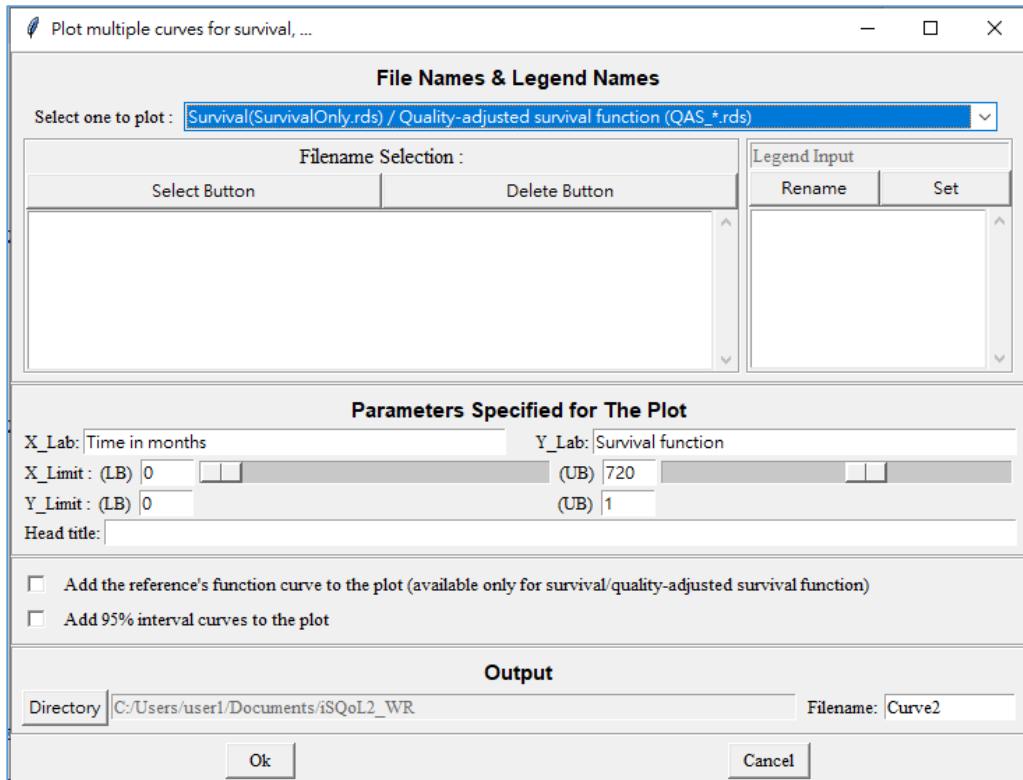

How to execute the function?

—   Click [**iSQoL2/Additional Functions/plot multiple curves**] to open an input window as Figure 5.6.

—   Choose a function to plot from [**Select one to plot**] menu.

—   Click **Select Button (Delete Button)** to select (delete) one or more target files.

   • The file in the left box is the temporary output files named *.rds.

   • The right box is the legend for each selected file. If you want to change the legend name, select one from the right box, then click **Rename** button to alter it and click **Set** button to change it.

   Note that:

- When the plot is shown in **R** console, please move the mouse cursor to the graphics coordinate for specifying the legend location.
- If legend names are empty, no legend will appear in this plot.
— Fill other inputs and click **Ok** button to start the execution. Then, the text outputs and a figure are shown in **R** console and saved to a file that you specify in the input window.

6. Discount for LE/QALE/Cost

The command is in order to discount LE, COST or QALE which has been estimated in Section 4.1 or 4.2 or 4.3, respectively. Before running the command, you need to prepare a file which is the input file (Input*.csv) of Section 4. Then, click the [**iSQoL2/Discount for LE/QALE/COST**] menu to open the input window shown below.
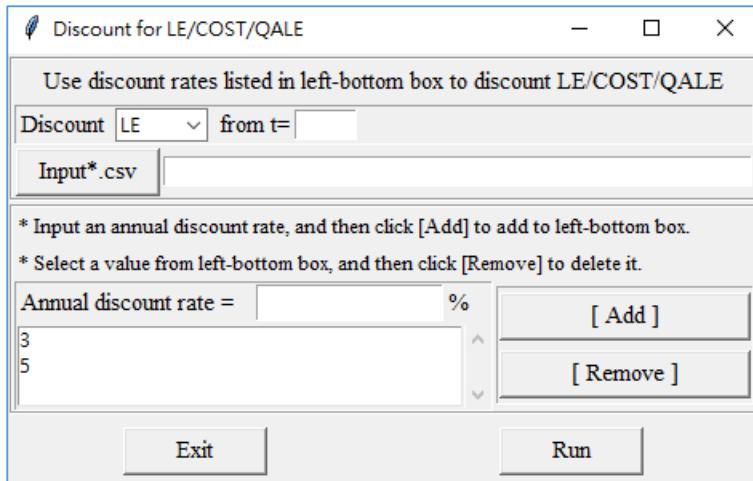
Next, follow the steps below to execute the command.
(1) Select one of three items: LE, QALE, COST for discounting.
(2) Input the starting time t (month after onset) for discounting.
(3) Select an input file (Input*.csv) described in Step 2 of Section 4. Note that the estimates will be discounted with the same discount rate(s) for all data sets of Input*.csv.
(4) Add or remove an annual discount rate in the left-bottom box.
   - How to add it?
     Input an annual discount rate and click [**Add**] to put in the left-bottom box.
   - How to remove it?
     Select an annual discount rate in the left-bottom box, then click [**Remove**] to delete.
(5) Click [**Run**] to discount with annual discount rate(s) listed in the left-bottom box.

The results contain the estimate of LE/QALE/COST calculated by the input annual discount rate(s). These will be shown on R console and saved in the specific directory (the last column of Input*.csv). For each data set, there are two output files, and the filenames format are &_Rate*_sy.txt and &_Rate*_#.rds.

7. Reference

1. Adler D, Gläser C, Nenadic O, Oehlschlägel J, Schuemie M, Zucchini W (2021). ff: Memory-Efficient Storage of Large Data on Disk and Fast Access Functions_. R package version 4.0.5, https://CRAN.R-project.org/package=ff.

2. Chung CH, Hu TH, Wang JD, Hwang JS. Estimation of Quality-Adjusted Life Expectancy of Patients With Oral Cancer: Integration of Lifetime Survival With Repeated Quality-of-Life Measurements. Value Health Reg Issues (2020) 21:59–65. doi: 10.1016/j.vhri.2019.07.005

3. Harrell Jr FE (2022). _rms: Regression Modeling Strategies_. R package version 6.3-0, https://CRAN.R-project.org/package=rms.

4. Hwang, J.S., Tsauo, J.Y. and Wang, J.D. (1996). Estimation of expected quality adjusted survival by cross-sectional survey. Statistics in Medicine, Vol. 15, 93-102.

5. Hwang, J.S. and Wang, J.D. (1999). Monte Carlo estimation of expected quality-adjusted survival for follow-up studies. Statistics in Medicine, 18, 1627-1640.

6. Hwang, J.S. and Wang, J.D. (2004). Integrating health profile with survival for quality of life assessment. Quality of Life Research, 13, 1-14.

7. Hwang, J.S., Hu TH, Lee L and Wang JD. (2017). Estimating lifetime medical costs from censored claims data. Health Economics, 26, e332-344.

8. Lemon, J. (2006) Plotrix: a package in the red light district of R. R-News, 6(4): 8-12.

9. Lo WC, Hu TH, Hwang JS (2023). Lifetime exposure to PM2.5 air pollution and disability-adjusted life years due to cardiopulmonary disease: A modeling study based on nationwide longitudinal data. Science of the Total Environment, 855. https://doi.org/10.1016/j.scitotenv.2022.158901.

10. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

11. Terry M. Therneau and Patricia M. Grambsch (2000). Modeling Survival Data: Extending the Cox Model. Springer, New York. ISBN 0-387-98784-3.

12. Therneau T (2015). A Package for Survival Analysis in S. Version 2.38, <URL: http://CRAN.R-project.org/package=survival>.

13. Xie J. and Liu C. (2005) Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. Statistics in medicine.