である．実際に計算してみると，

$$\pi(p_1, p_2 | x_1, x_2) = \pi(p_1, p_2;\ \alpha_1 + x_1, \alpha_2 + x_2, \alpha_3 + x_3),$$
$$\text{ただし，} x_3 = n - x_1 - x_2.$$

となる．記号で書くと，$(p_1, p_2) \sim \mathrm{Dir}(\alpha_1 + x_1, \alpha_2 + x_2, \alpha_3 + x_3)$ である．

以上をまとめると次のようになる．

事前分布		データ		事後分布
$\mathrm{Dir}(\alpha_1, \alpha_2, \alpha_3)$	and	$\mathrm{Mn}(n, p_1, p_2)$	$\rightarrow$	$\mathrm{Dir}(\alpha_1 + x_1, \alpha_2 + x_2, \alpha_3 + x_3)$

$(\alpha_1, \alpha_2, \alpha_3) = (2, 3, 4)$ および $(x_1, x_2, x_3) = (5m, 4m, 3m)$ とし，$m = 0, 1, 2, 3$ と変化させて事前分布と事後分布のグラフを描いてみる．

4. ディリクレ分布の性質

ディリクレ分布は，その周辺分布もディリクレ分布になるという性質を持っている．特に，ある1つの変数に着目すると，周辺分布はベータ分布になる．$(p_1, p_2) \sim \mathrm{Dir}(\alpha_1, \alpha_2, \alpha_3)$ のとき，例えば p_1 はベータ分布 $\mathrm{Beta}(\alpha_1, \alpha_2 + \alpha_3)$ に従う．したがって，平均および分散は次のように計算できる．

$$\mathrm{E}[p_1] = \frac{\alpha_1}{\alpha_1 + \alpha_2 + \alpha_3}, \quad \mathrm{Var}[p_1] = \frac{\alpha_1(\alpha_2 + \alpha_3)}{(\alpha_1 + \alpha_2 + \alpha_3)^2(\alpha_1 + \alpha_2 + \alpha_3 + 1)}.$$

- 期待値はパラメータ $(\alpha_1, \alpha_2, \alpha_3)$ の比率で決まり，
- （期待値が同じ＝パラメータの比が同じときには,）
  パラメータの総和 $\alpha_1 + \alpha_2 + \alpha_3$ が大きいほど分散は小さくなる．

5. 専門家の意見をディリクレ分布で表現する

Bayes の定理は平たく言えば，情報の更新ルールである．専門家の意見は次のように Bayes の定理の実例と見立てることができる．

1. 専門家は勉強などを通じて当初の意見（ディリクレ分布）をもっており，
2. 臨床などの経験（多項分布に従うデータ）と統合して，
3. 意見が修正される（ディリクレ分布）．

次のスキームで専門家1人1人の意見をディリクレ分布で表現する．

- 各カテゴリの best estimate が期待値に一致するようにする．
  ⇒ パラメータの比率が決まる．
- Best estimate の最大値を与えるカテゴリに着目し，区間推定の幅とディリクレ分布（ベータ分布）に基づく幅が一致するようにする．
  ⇒ パラメータの総和が決まる．

ディリクレ分布ではパラメータの個数が少ないため，全てのカテゴリで区間推定の幅とディリクレ分布に基づく幅を一致させることはできない．Best estimate の最大値を与えるカテゴリのみに着目したのはこのためである．

6. 専門家の意見を統合する

まず，2 人の専門家の意見の統合の仕方を説明する．専門家 A が $(p_1, p_2) \sim \mathrm{Dir}(\alpha_1, \alpha_2, \alpha_3)$ という意見をもち，専門家 B が $(p_1, p_2) \sim \mathrm{Dir}(\beta_1, \beta_2, \beta_3)$ を持っているとする．このとき，2 人の意見を次のように統合する．

$$\frac{\pi(p_1, p_2;\ \alpha_1, \alpha_2, \alpha_3)\,\pi(p_1, p_2;\ \beta_1, \beta_2, \beta_3)}{\displaystyle\int_0^1 dp_1 \int_0^{1-p_1} dp_2\ \pi(p_1, p_2;\ \alpha_1, \alpha_2, \alpha_3)\,\pi(p_1, p_2;\ \beta_1, \beta_2, \beta_3)} \tag{1}$$

これは Bayes の定理と同様の計算である．実際に計算すると，$(p_1, p_2) \sim \mathrm{Dir}(\alpha_1+\beta_1, \alpha_2+\beta_2, \alpha_3+\beta_3)$ となる．このスキームを記号で次のように表しておく．

A の意見		B の意見		統合された意見
$\mathrm{Dir}(\alpha_1, \alpha_2, \alpha_3)$	and	$\mathrm{Dir}(\beta_1, \beta_2, \beta_3)$	$\rightarrow$	$\mathrm{Dir}(\alpha_1+\beta_1, \alpha_2+\beta_2, \alpha_3+\beta_3)$

$(\alpha_1, \alpha_2, \alpha_3) = (2, 3, 4)$ および $(\beta_1, \beta_2, \beta_3) = (15, 12, 9)$ の場合のグラフを描いてみる．

3 人以上の専門家の意見を統合するときは，上記の方法を順次適用すればよい．

以上

(別添3-1)

The DALY Calculator

A graphical user interface for stochastic DALY calculation in ***R***

— Manual version 1.1.0 —

Brecht Devleesschauwer[1,2] **Arie Havelaar[3,4]** **Juanita Haagsma[5]**

Nicolas Praet[6] **Niko Speybroeck[2]**

1 Department of Virology, Parasitology and Immunology, Faculty of Veterinary Medicine, Ghent University, Ghent, Belgium
2 Institute of Health and Society, Faculty of Public Health, Université catholique de Louvain, Brussels, Belgium
3 Laboratory for Zoonoses and Environmental Microbiology, National Institute for Public Health and the Environment, Bilthoven, The Netherlands
4 Institute for Risk Assessment Sciences, University Utrecht, Utrecht, the Netherlands
5 Department of Public Health, Erasmus Medical Center, Erasmus University, Rotterdam, The Netherlands
6 Department of Biomedical Sciences, Institute of Tropical Medicine, Antwerp, Belgium

Corresponding Author:
Brecht Devleesschauwer
Department of Virology, Parasitology and Immunology
Faculty of Veterinary Medicine
Ghent University
Salisburylaan 133
9820 Merelbeke
Belgium
e-mail: Brecht.Devleesschauwer@UGent.be

ABSTRACT

The disability-adjusted life year or DALY is an increasingly used measure of population health. The **DALY** package, accessible in the **R** statistical programming environment, provides a Graphical User Interface (GUI) for calculating DALYs and performing uncertainty analysis. The latter is done through Monte Carlo simulations, which are used to compute DALY credibility intervals based on the uncertainty in the various input parameters. The GUI of the *DALY Calculator* allows for ease and flexibility, while the **R** environment enables more advanced computations and graphical functions. The underlying calculation methods are designed to promote consistency in the uncertainty analysis of DALYs. These methods are demonstrated through the inclusion of two examples from the foodborne disease burden literature (i.e., the burden of neurocysticercosis in West-Cameroon and the burden of congenital toxoplasmosis in the Netherlands). We believe that the *DALY Calculator* can become a useful tool for students, scientists and health officials involved in the quantification of public health problems.

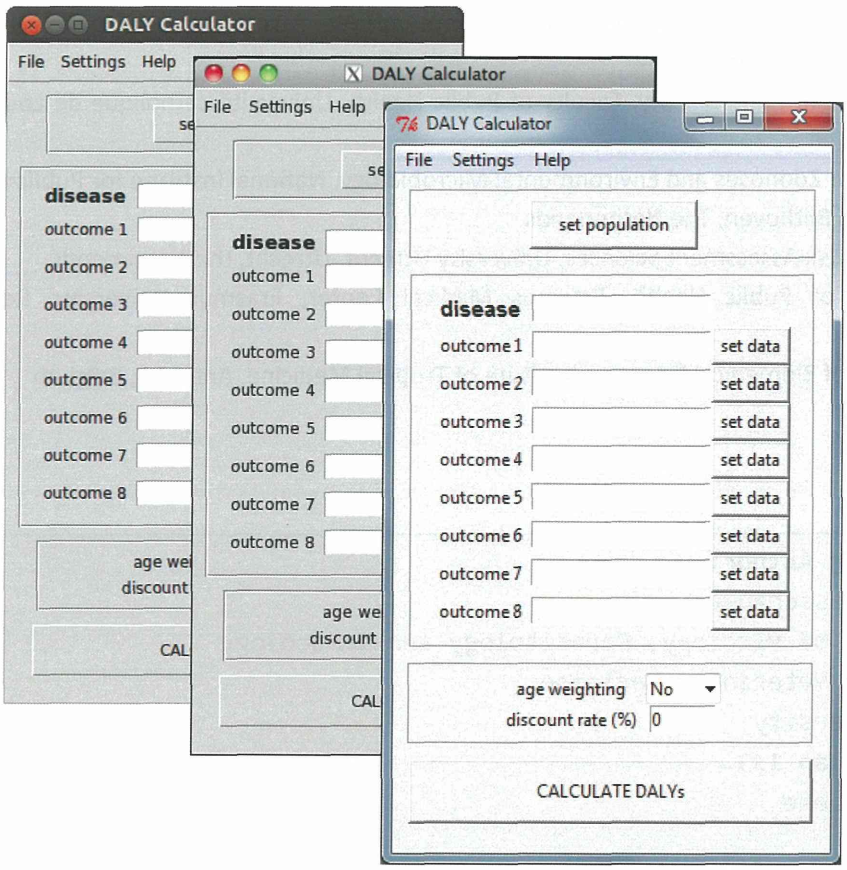


Figure 1. DALY Calculator main window in Ubuntu Linux, Mac OS X 10.6, and Windows 7

1. Introduction

The ***disability-adjusted life year*** **or DALY** is a summary measure of population health widely used in disease burden assessment studies and cost-utility analyses (Murray and Lopez, 1996; Lopez et al., 2006). DALYs represent the number of healthy life years lost due to a disease or disability, and do so by incorporating non-fatal and fatal health outcomes, calculated as the years of life lived with disability (YLD) and the years of life lost due to premature death (YLL), respectively. In order to make a direct comparison of YLDs and YLLs possible, disability weights are assigned to various degrees of disability, ranging from zero (full health) to one (worst possible health state). In addition to these disability weights, the DALY calculation requires the choice of two other social values, i.e., age weighting and time discounting. The former gives a higher weight to the healthy life years lived between the age of 9 and 54, as this period of life is considered to be socially more important than the younger and older life spans (Murray, 1994). The latter discounts the years of healthy life lived in the future, at a rate of (usually) 3%. The incorporation of a time discounting rate reflects similar practices in economic assessments, and would prevent policy makers from saving resources for a possible future eradication program, instead of investing in currently available, but less effective, intervention measures (the so-called "disease eradication and research paradox"; Murray, 1994).

The formulas for calculating the YLLs and YLDs are presented in Murray (1994) and Murray and Acharya (1997), and can easily be incorporated in a spreadsheet, such as the "**DALY calculation template**"[1] prepared by the World Health Organization for its Global Burden of Disease (GBD) project (Mathers et al., 2001). This template allows the computation of YLLs and YLDs for both sexes and various age groups, which are then summarized into a single deterministic DALY measure.

The reliability of the final DALY result depends heavily on the quality of the epidemiological data, which are commonly derived from routine data collection systems, scientific literature, and expert elicitation. The estimates provided by these data sources include an inherent level of uncertainty, mainly due to sampling error, diagnostic uncertainty, and population heterogeneity. To reflect this stochastic nature, epidemiological parameters are often accompanied by a confidence or credibility interval, or represented as a probability distribution, rather than being represented by a single point estimate. **Monte Carlo simulations** have been suggested as the appropriate technique to incorporate this uncertainty in the final DALY result (de Vocht et al., 2010), which can then be presented as a point estimate with a credibility

[1] http://www.who.int/entity/healthinfo/bodreferencedalycalculationtemplate.xls

interval. This interval allows the assessment of the level of uncertainty in the total DALYs, and facilitates a more reliable comparison of the health impact of different diseases.

To our knowledge, however, there are **no standardized tools** available for stochastic DALY calculation. Therefore, we designed a Graphical User Interface (GUI) for calculating DALYs that allows for the incorporation of input uncertainty and the computation of a DALY credibility interval through Monte Carlo simulations. This program, the *DALY Calculator*, is designed to be used by a variety of users, with different levels of statistical skills, to allow maximum flexibility, and to promote consistency in the DALY uncertainty analysis.

The remainder of this manual is structured as follows. A description of how to obtain and install the *DALY Calculator* is given in Chapter 2. The five basic steps to calculate DALYs with the *DALY Calculator* are presented in Chapter 3. Using the table widgets of the *DALY Calculator* is described in Chapter 4. Manipulating the *DALY Calculator* output in the **R** environment is presented in Chapter 5. The two built-in examples are introduced in Chapter 6. The features to save and load input data are described in Chapter 7. Current limitations and envisioned future improvements are listed in Chapter 8. Finally, an overview of the different versions of the DALY Calculator is given in Chapter 9.

The computational details of the uncertainty analysis performed by the *DALY Calculator* are outlined in Annex 1.

2. Installing and running the *DALY Calculator*

The *DALY Calculator* is developed in **R**, an open-source environment for statistical programming and graphics (R Development Core Team, 2012). The use of the *DALY Calculator* requires the prior installation of **R**, which can be freely downloaded from the Comprehensive R Archive Network (CRAN): http://cran.r-project.org/

The installation procedure under Mac OS X is more complicated. Two additional tools need to be installed:

- The **X Window System (X11)**: if this is properly installed, the file `X11.app` should appear in the Utilities folder under Applications in the Finder. If not, you can install it from your Mac OS X installation disc.
- **Tcl/Tk for X Windows**: this can be installed by downloading and installing `tcltk-8.5.5-x11.dmg` from http://cran.r-project.org/bin/macosx/tools/

The **DALY** package is available on the CRAN repository, and can be installed by accessing the following menu in **R**:

Packages > Install package(s)...

Next, the required CRAN mirror has to be selected, and the **DALY** package can be selected and installed from the list of available packages.

Alternatively, the download and installation of the **DALY** package may be invoked directly from the **R** console, by calling:

```
install.packages("DALY")
```

Finally, every time the *DALY Calculator* has to be used, the **DALY** package has to be loaded in the **R** environment, by typing the following command in the **R** console (Figure 2):

```
library(DALY)
```

This call will load and attach the **DALY** package, and initiate the main window of the *DALY Calculator* (Figure 1, 3-5). To re-initiate this window, call the following function in the **R** console:

```
DALYcalculator()
```

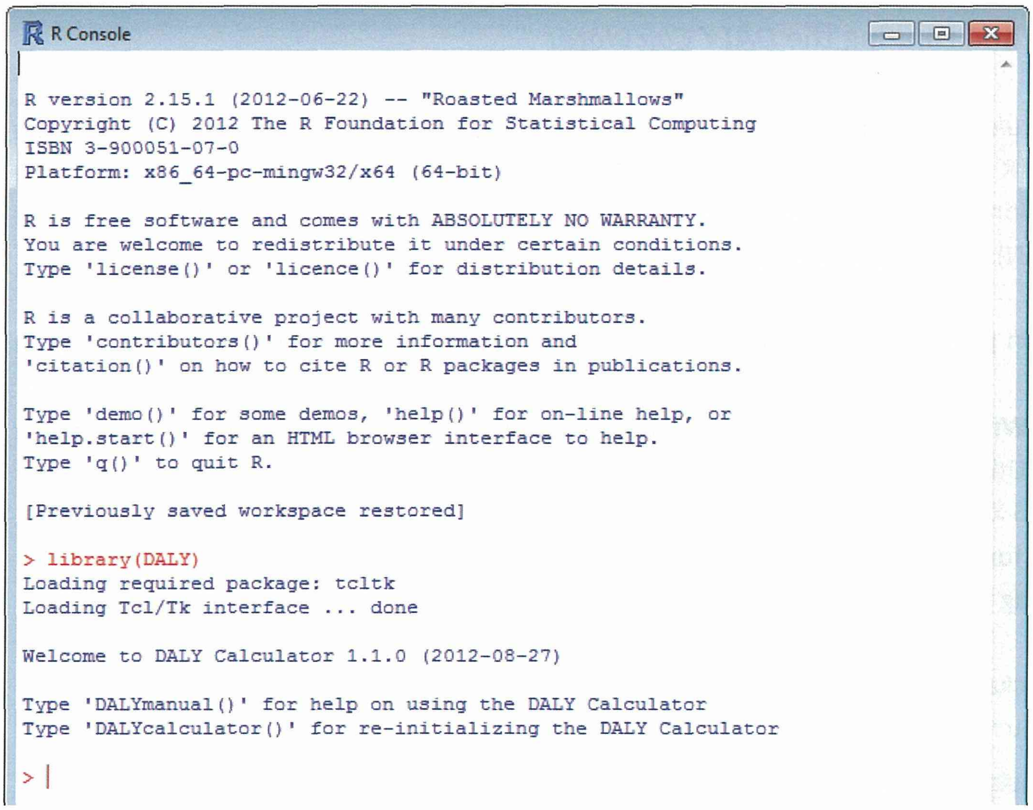


Figure 2. Initializing the DALY Calculator

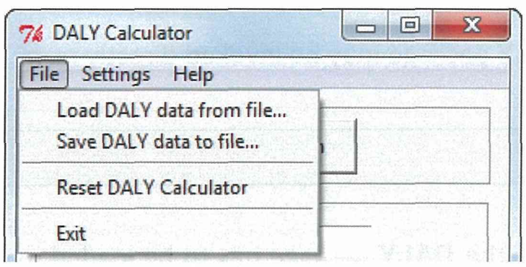


Figure 3. DALY Calculator 'File' menu

Figure 4. DALY Calculator 'Settings' menu

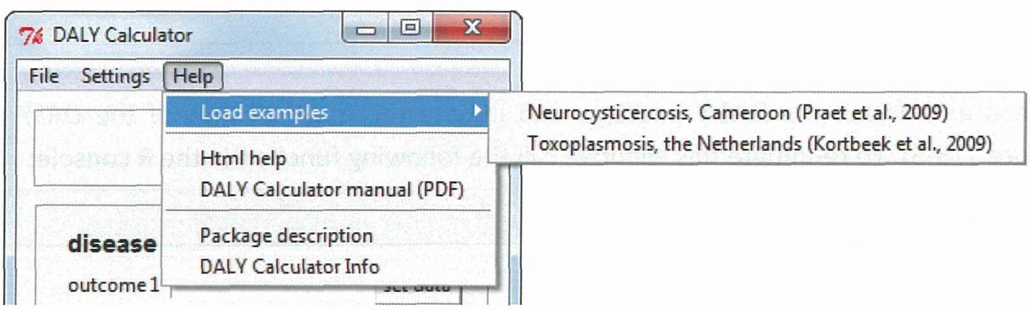


Figure 5. DALY Calculator 'Help' menu

3. Getting started with the *DALY Calculator*

The *DALY Calculator* is designed to be used in an **outcome-based** as well as an **agent-based** approach. Depending on the approach, YLDs, YLLs, DALYs and incident cases and deaths can be computed for a combination of up to eight different disease categories of one outcome, or of up to eight different outcomes of one agent.

According to the GBD studies, the **incidence-based** approach was favored over the prevalence-based approach. The current version of the *DALY Calculator* is able to calculate DALYs for a basic incidence-based disease model with well-defined incidence and/or mortality rates per disease category or outcome.

The default age groups used by the *DALY Calculator* are the **five age groups** used by the GBD 1990 study: 0-4; 5-14; 15-44; 45-59; 60+ (Lopez and Murray, 1996). At least one combination of sex and age group has to be set in order to proceed with the DALY calculation.

To compute YLDs, YLLs, DALYs, incident cases and deaths, **five steps** have to be followed (as illustrated in Figure 6). These steps will be explained using one of the two built-in examples (The burden of *Taenia solium* cysticercosis in Cameroon; Praet et al., 2009). Users can load this example by selecting:

Help > Load examples > Neurocysticercosis, Cameroon (Praet et al., 2009)

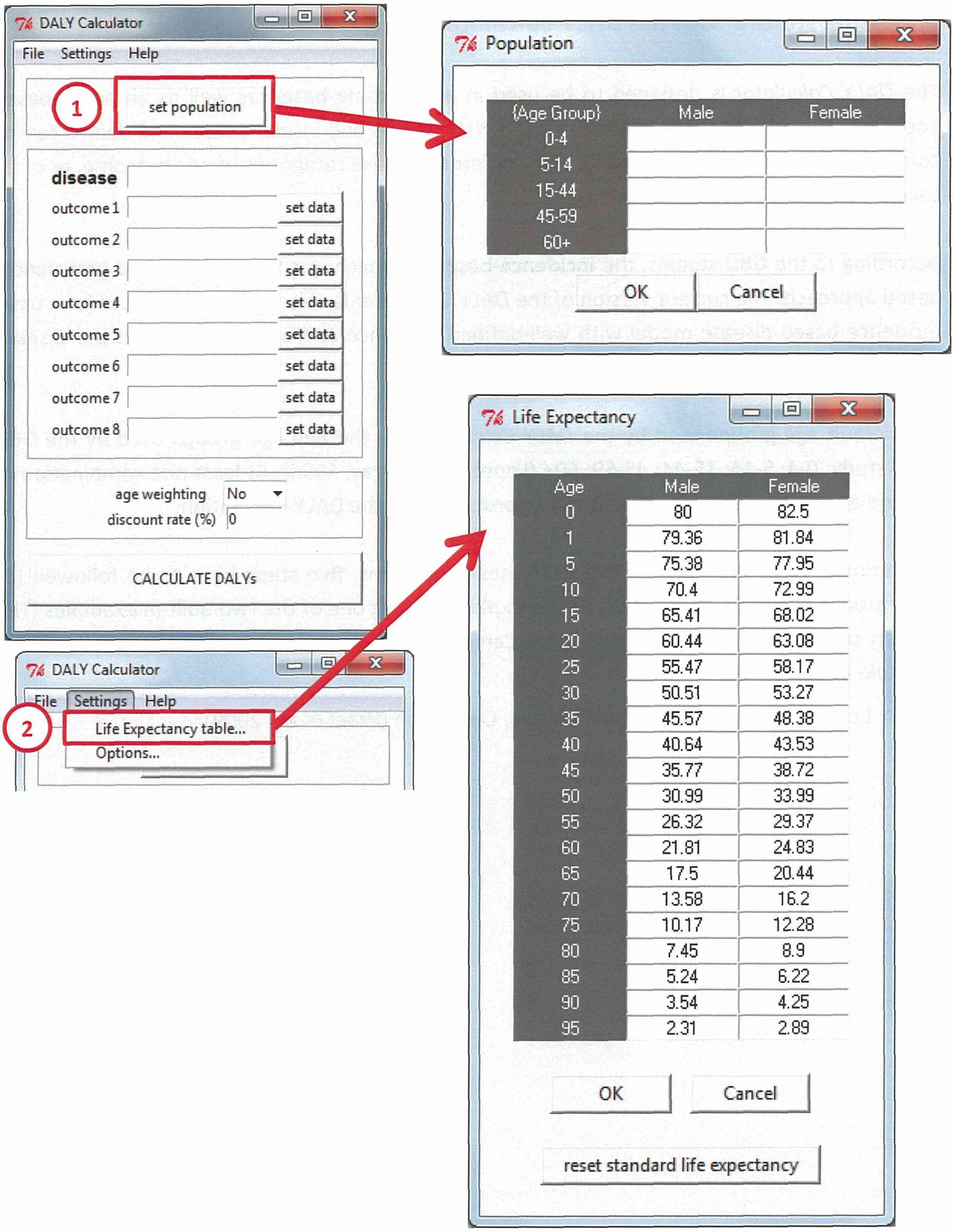


Figure 6. DALY calculation with the DALY Calculator

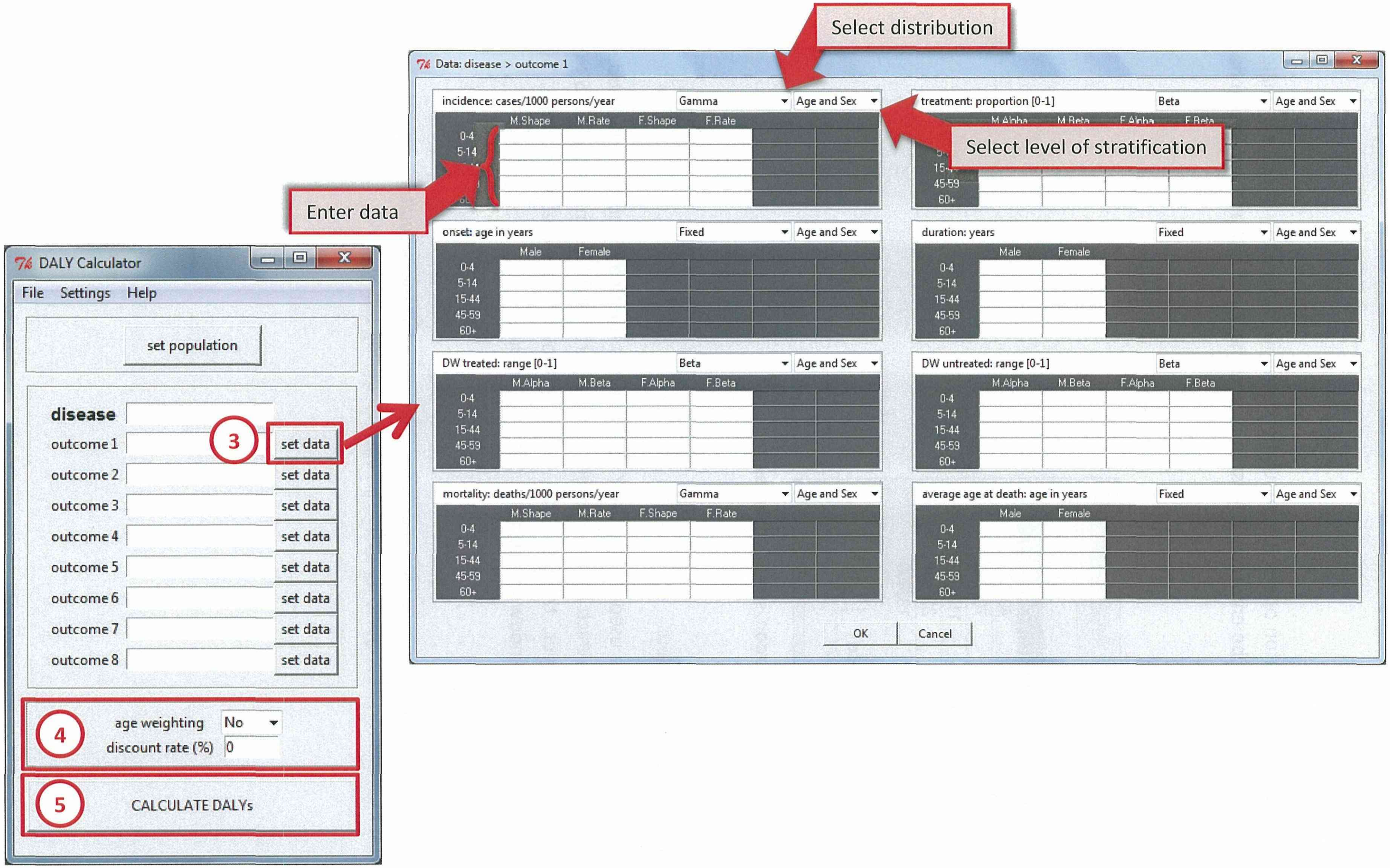
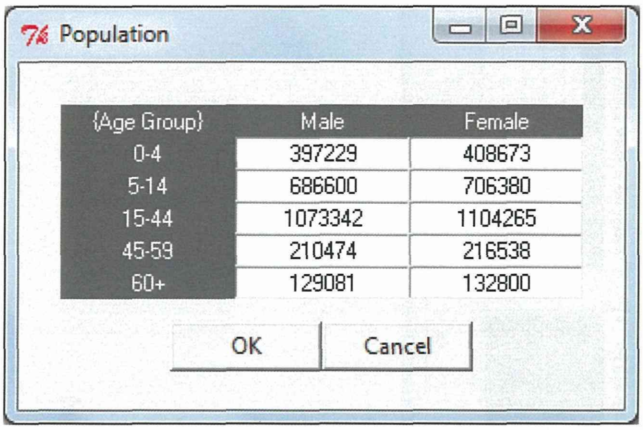


Figure 6. DALY calculation with the DALY Calculator (continued; M = Male, F = Female)

① Set the population table

Clicking the "set population" button opens the population window, where the number of males and females, per age group, can be entered. At least one combination of sex and age group has to be set.



(Age Group)	Male	Female
0-4	397229	408673
5-14	686600	706380
15-44	1073342	1104265
45-59	210474	216538
60+	129081	132800

Figure 7. Population of West-Cameroon, stratified by sex and age

② Set the life expectancy table

The default life expectancy table used by the *DALY Calculator* is the Coale and Demeny model life-table West, level 26 and 25, which has a life expectancy at birth of 80 for males and 82.5 for females (Murray, 1994). However, the user can define his own life expectancy table, by accessing the 'Life Expectancy' window through the 'Settings' menu:

Settings > Life Expectancy

Praet et al. (2009) applied the standard life expectancy table, which is the default life expectancy table of the DALY Calculator. Therefore, no action is required during this step.

③ Set the input parameters, per disease category or outcome

After entering the disease and outcome names, the epidemiological data and disability weights can be entered by clicking the "set data" button. For every parameter, the user can specify the **distribution** by selecting one of following:

- **Beta-Pert** (mode; min; max)
- **Beta** (alpha; beta)
- **Gamma** (shape; rate)
- **Normal** (mu; sigma)
- **Lognormal-geometric** (logmean; logsigma)
- **Lognormal-arithmetic** (mean; sigma)
- **Uniform** (min; max)
- **Fixed**

Next, the user has to select the specific **level of stratification** for every parameter. The following four stratification levels are available:

- **Age and Sex** (i.e., full stratification)
- **Age** (i.e., data is stratified by age group, but not by sex)
- **Sex** (i.e., data is stratified by sex, but not by age group)
- **None** (i.e., no stratification, data applies to total population)

For calculating the **YLDs and incident cases**, the following tables have to be completed:

- **Incidence**: number of new cases per 1,000 persons per year
- **Treatment**: proportion of patients receiving proper treatment; range [0-1]
- **Onset of the disease**: age of onset in years
- **Duration of the disease**: duration in years
- **Disability Weight for treated cases**: range [0-1]
- **Disability Weight for non-treated cases**: range [0-1]

For calculating the **YLLs and deaths**, the following tables have to be completed:

- **Mortality**: number of deaths per 1,000 persons per year
- **Average age at death**: age at death in years; based on these values, the *DALY Calculator* will compute the corresponding life expectancies according to the Life Expectancy table.

By default, the distributions of incidence and mortality will be set to "Gamma", those of the proportion treated and the disability weights to "Beta", and those of onset, duration and average age at death to "Fixed"; the level of stratification for all parameters is set to full stratification (i.e., "Age and Sex").

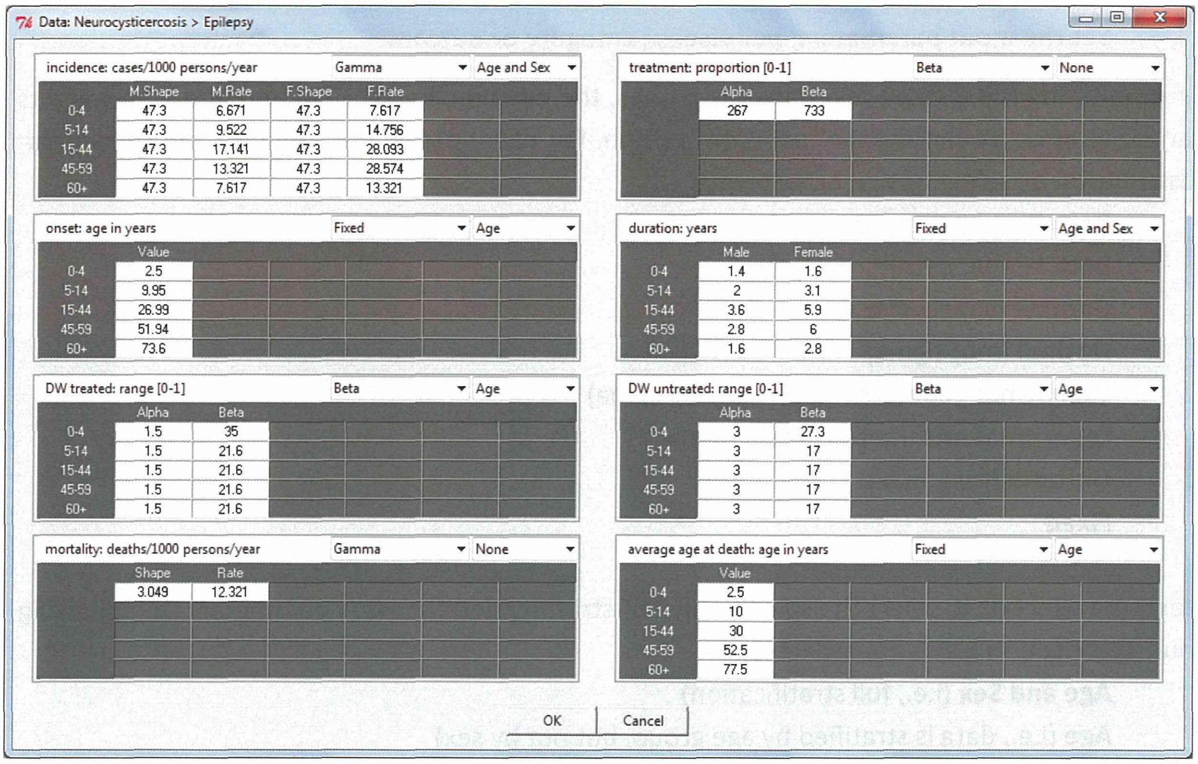


Figure 8. Epidemiological parameters for the DALY calculation of Neurocysticercosis in Cameroon, based on Praet et al. (2009)

④ Set the social values:

Age weighting yes or no
Discount rate (%) ...

The default social values applied by the *DALY calculator* are uniform age weights (i.e., no age weighting) and a zero discount rate. However, the user is given the possibility to define the required set of social values, and to alter these values to assess their influence on the final result.

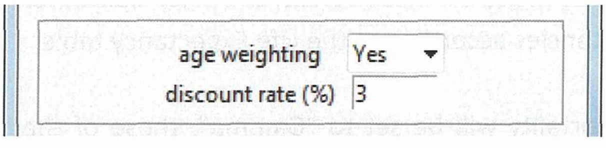


Figure 9. DALY Calculator main window, with social values set to full age-weighting and a 3% discount rate, as applied by Praet et al. (2009)

⑤ CALCULATE DALYs

Clicking the "CALCULATE DALYs" button will read the data and compute the YLDs, YLLs, DALYs and incident cases and deaths per disease category/outcome. Standard, the overall mean, median and a 95% credibility interval of these results will be printed to the **R** Console, and a histogram of total DALYs, with accompanying density distribution and credibility interval, will be displayed (Figure 10).

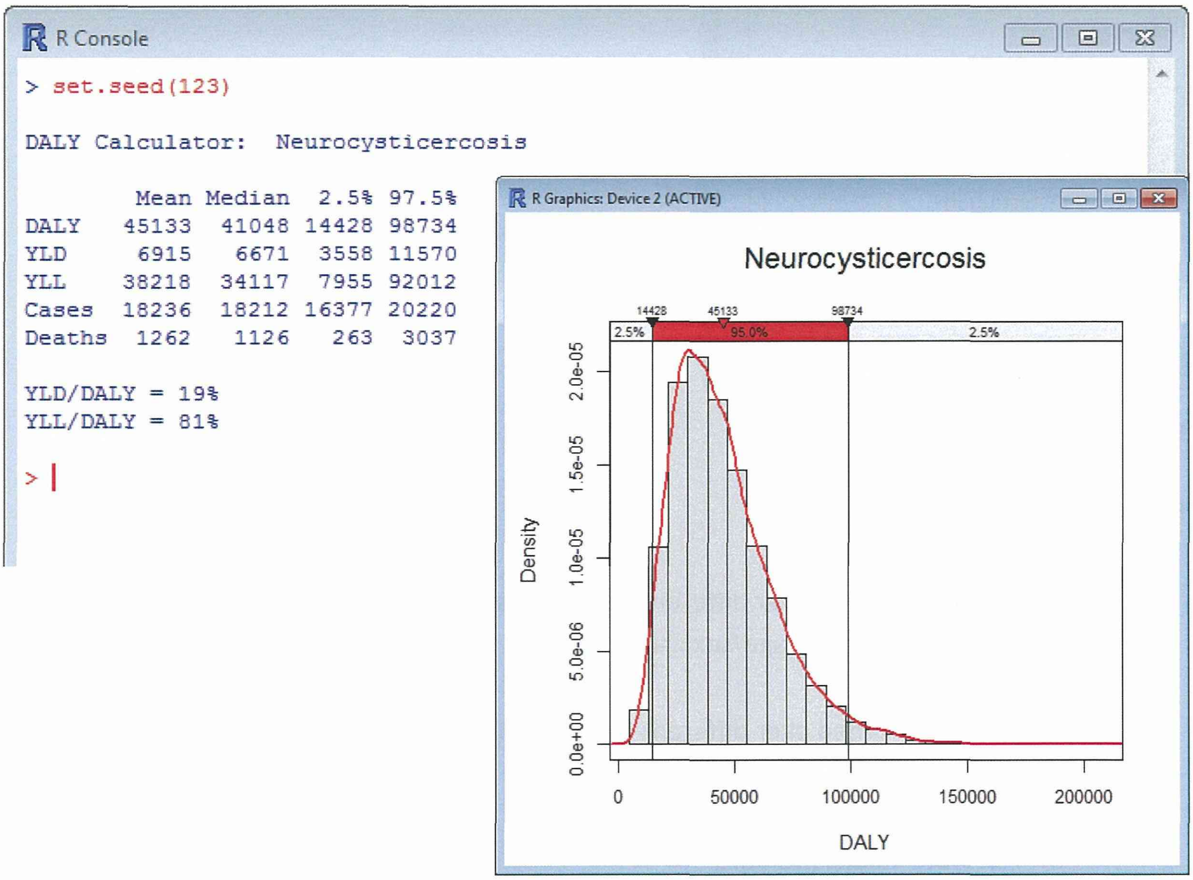


Figure 10. Standard output of the DALY calculation for *Taenia solium* Neurocysticercosis in Cameroon (based on Praet et al., 2009)

A more detailed output of the DALY calculation can be obtained by changing the output settings in the 'Options' window (Figure 11):

Settings > Options

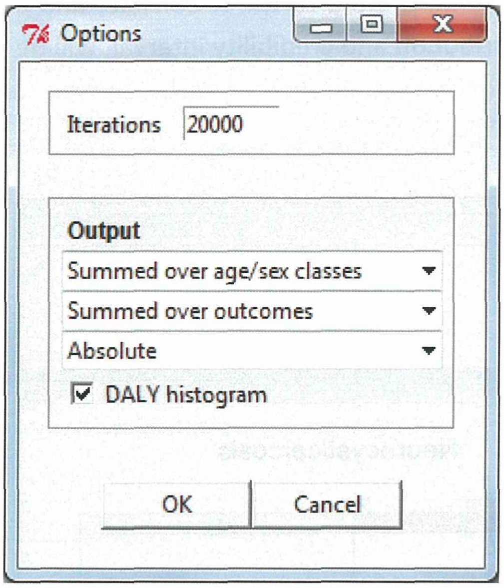


Figure 11. Options window

The following settings are possible:

- Age/Sex classes
    - Results summed over different age/sex classes (default)
    - Results shown per age/sex class
- Outcomes
    - Results summed over different outcomes (default)
    - Results shown per outcome
- Absolute/Relative
    - Absolute values (default)
    - Relative values, i.e., per 1,000 population

More advanced output, both numerically as graphically, can be obtained by saving the *DALY Calculator* output to an **R** object, and manipulating this object in the **R** environment (see Chapter 5 for more details and examples).

4. Using the *DALY Calculator* table widgets

All tables of the *DALY Calculator* are built using the **`Tktable`** toolkit. The navigation and editing properties of this tabulator toolkit differ from those of common spreadsheet documents, which may cause confusion. Some clues are therefore useful to get started:

- Keyboard navigation is only possible through the arrow keys; pressing the RETURN (ENTER) key will not change focus to the underlying cell, but will append blank space to the currently active cell.
- To select all values of a column (row), you can click the column (row) header. Selecting all values in the table is possible by clicking the header in the top-left corner of the table. Figure 12 gives some examples.
- Removing the value of one or more cells can be done by selecting the corresponding cells and pressing CTRL+X (also on Mac).
- Pasting values into a table (eg, after copying them from a spreadsheet document), can be done by selecting the top-left cell of the desired range, and pressing CTRL+V (also on Mac).

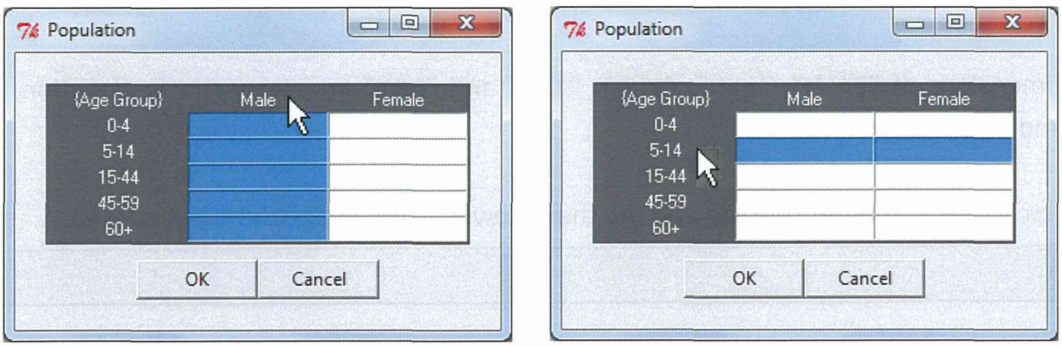


Figure 12. Selecting columns/rows in *DALY Calculator* table widgets.

5. Handling *DALY Calculator* output in the R environment

Instead of clicking the "CALCULATE DALYs" button in the *DALY Calculator* main window, the DALY calculation process may also be invoked by calling the `getDALY()` function from the **R** console. This function initiates the Monte Carlo simulation process, and returns the simulated DALYs, YLDs, YLLs and incident cases and deaths. Of course, this will only take place if the data have been entered, as outlined before.

The results of the DALY calculation process can be stored in an object, say `'x'`, as follows:

```
x <- getDALY()
```

5.1. Structure of the object returned by `getDALY()`

The `getDALY()` function returns an object of class `'DALY'`, which inherits from class `'list'`. The returned object is a list containing the following elements:

`i`	For each outcome *i*, an unnamed list containing simulated DALYs, YLDs, YLLs and incident cases and deaths, per age/sex class;
`"pop"`	A matrix containing the population data;
`"name"`	The name of the disease.

Each list of simulated results for a certain outcome *i* contains the following six elements:

`"DALY"`	3-dimensional vector of simulated DALYs;
`"YLD"`	3-dimensional vector of simulated YLDs;
`"YLL"`	3-dimensional vector of simulated YLLs;
`"INC"`	3-dimensional vector of simulated incident cases;
`"MRT"`	3-dimensional vector of simulated deaths;
`"name"`	The name of the outcome.

The three dimensions of `"DALY"`, `"YLD"`, `"YLL"`, `"INC"` and `"MRT"` are, respectively, iteration, age group, and sex.

To view the structure of the returned object, call the following command from the **R** console:

```
str(x)
```

5.2. Methods for objects of class `'DALY'`

Four methods have been made available for the objects of class `'DALY'`:

`print()`	print results, summed over age/sex classes;
`summary()`	print results, per age/sex class;
`hist()`	histogram of total DALYs, YLDs, YLLs, incident cases or deaths;
`aggregate()`	aggregate simulations by outcome, age/sex, or both.

In both the `print()` and `summary()` methods, the user can specify whether or not to print the results per outcome (argument `outcome`), and whether or not to print the results relative to the population (i.e., per 1000 population; argument `relative`). By default, `outcome` and `relative` are set to FALSE.

The argument `digits` of both the `print()` and `summary()` methods controls the number of decimal digits to be printed. By default, `digits` equals zero. The argument `prob` of the `print()` method sets the range of the printed credibility interval. By default, `prob` equals 0.95.

For example, if you wanted to see the results of the Neurocysticercosis example expressed relative to the population size, and with 90% credibility intervals, you would call:

```
> print(x, relative = TRUE, prob = 0.90, digits = 2)

DALY Calculator:  Neurocysticercosis

Total population:  5065382

       Mean Median   5%   95%
DALY   8.82   8.00 3.37 17.00
YLD    1.36   1.31 0.78  2.10
YLL    7.47   6.63 2.08 15.65
Cases  3.60   3.59 3.29  3.92
Deaths 0.25   0.22 0.07  0.52

YLD/DALY = 19%
YLL/DALY = 81%
```

The `hist()` method plots a standardized histogram of simulated results. By default, argument `xval` is set to "DALY", argument `prob` to 0.95, and argument `central` to "mean". The histogram shown in Figure 10 is a result of these default settings.

If you wanted to see a histogram of, say, the number of deaths in the Neurocysticercosis example, with an indication of the median and a 90% credibility interval, you would call:

```
hist(x, xval = "Deaths", prob = 0.90, central = "median")
```

*For more info on the available methods for class `'DALY'`, please visit the **R** help pages for the corresponding methods. To access the **R** help page for, say, `print.DALY()`, call `?print.DALY`.*

5.3. Handling objects of class 'DALY'

The strength of the implementation of the *DALY Calculator* is that its results can be directly manipulated from within the **R** environment. This allows users to obtain virtually any possible numerical or graphical output they desire. Of course, this also requires a more advanced knowledge of the **R** programming language. Some hints and examples might, however, get you started.

Each of the elements from `'x'` may be extracted using the `'[['` operator. Named elements may also be extracted using the `'$'` operator. For example, we can view the population matrix of the Neurocysticercosis example as follows:

```
> x[["pop"]]  # equivalent to 'x$pop'
        [,1]    [,2]
[1,]  397229  408673
[2,]  686600  706380
[3,] 1073342 1104265
[4,]  210474  216538
[5,]  129081  132800
```

Likewise, the simulated DALYs due to epilepsy (i.e., outcome 1 of the Neurocysticercosis example), may be extracted as follows (note the double indexing):

```
x[[1]][["DALY"]]
```

The `aggregate()` method is a utility function used to sum up simulated results by outcome, age/sex, or both (controlled by argument `by`). This method returns a list of aggregated DALYs, YLDs, YLLs, incident cases and deaths, as well as the population matrix and disease name.

```
aggregate(x, by = "outcome")  # aggregate by outcome, sum over age/sex class
aggregate(x, by = "class")    # aggregate by age/sex class, sum over outcomes
aggregate(x, by = "total")    # sum over outcomes and age/sex classes
```

For more information on the `aggregate()` method, please visit its **R** help page:

```
?aggregate.DALY
```