

The logo for CanReg 5 is centered at the top of the page. It consists of a large, light-brown circle with a subtle gradient. Overlaid on the circle is the text "CanReg" in a bold, dark brown sans-serif font, and a large, bright orange number "5" below it.

# CanReg 5

**Open Source**

By: Morten Johannes Ervik, IARC

Version: October 18, 2016



# Contents

<b>I. Introduction</b>	<b>9</b>
<b>1. What is CanReg?</b>	<b>11</b>
1.1. A short introduction to the database structure in CanReg5 . . . . .	11
1.2. Forum, Issue tracker, community site, twitter . . . . .	12
1.3. Getting hold of the latest version of CanReg . . . . .	12
1.4. Logfile . . . . .	12
<b>II. Installing and running CanReg5</b>	<b>15</b>
<b>2. Software installation</b>	<b>17</b>
2.1. Install a recent Java Runtime Environment . . . . .	17
2.2. Install CanReg5 . . . . .	17
2.2.1. From CanReg5.zip . . . . .	17
2.2.2. From CanReg5-Setup.zip (Windows only) . . . . .	17
2.3. Recommended Third Party tools . . . . .	17
2.3.1. Any post script viewer . . . . .	17
2.3.2. R . . . . .	18
2.4. Un-install CanReg5 . . . . .	18
<b>3. First launch</b>	<b>19</b>
3.1. Run CanReg5 . . . . .	19
3.2. Demo system . . . . .	20
3.3. Install a new CanReg5 system . . . . .	20
3.4. Convert the CanReg4 system definitions . . . . .	20
3.5. Setting up or modifying a CanReg system using the built in editor . . . .	22
3.5.1. Modifying a dictionary . . . . .	25
3.5.2. Modifying a group . . . . .	25
3.5.3. Modifying a variable . . . . .	26
3.5.4. Set up person search variables . . . . .	27
3.5.5. Coding . . . . .	27
3.5.6. Settings . . . . .	28
3.5.7. Saving the system . . . . .	28

3.6. Launching the CanReg server . . . . .	28
3.6.1. Launching server from command line . . . . .	29
3.7. Login . . . . .	29
3.7.1. Locally . . . . .	29
3.7.2. In a network . . . . .	30
3.8. Import the dictionaries from CanReg4 . . . . .	31
3.9. Import the data from CanReg4 . . . . .	33
3.10. Import data from other programs . . . . .	35
 <b>III. Working with CanReg5</b>	 <b>37</b>
 <b>4. Start</b>	 <b>39</b>
4.1. Welcome screen . . . . .	39
4.2. Login to CanReg . . . . .	39
4.2.1. Advanced login settings . . . . .	39
4.2.1.1. Test connection . . . . .	40
4.2.1.2. Get IP Address . . . . .	40
4.2.1.3. Port . . . . .	40
4.2.1.4. Automatically launch local server on startup . . . . .	40
4.2.1.5. Single user mode . . . . .	41
4.3. Install New System . . . . .	41
 <b>5. Data entry</b>	 <b>43</b>
5.1. Browse/edit . . . . .	43
5.1.1. Table . . . . .	43
5.1.2. Sorty by . . . . .	43
5.1.3. Range . . . . .	44
5.1.4. Filter . . . . .	44
5.1.4.1. Examples of how to use . . . . .	44
5.1.5. Filter wizard . . . . .	45
5.1.5.1. For example . . . . .	45
5.1.6. Display variables . . . . .	47
5.1.7. Navigate buttons . . . . .	47
5.2. Edit record . . . . .	47
5.2.1. System variables . . . . .	49
5.2.2. Obsolete button. . . . .	50
5.2.3. Check . . . . .	50
5.2.4. Person search . . . . .	51
5.3. Dictionary editor . . . . .	52
5.3.1. Export Dictionary to file . . . . .	52
5.3.2. Display/edit ... Select . . . . .	53
5.3.3. Update . . . . .	53
5.3.4. Import Dictionary from file . . . . .	53

5.4. Population Dataset Editor . . . . .	54
5.4.1. Manual entry/validation . . . . .	54
5.4.2. Import population data set from CanReg4 . . . . .	57
5.5. Import . . . . .	59
5.5.1. Discrepancies . . . . .	62
5.5.2. Max. Lines / Test only . . . . .	63
5.5.3. CanReg DATA . . . . .	63
5.5.3.1. Perform Checks . . . . .	63
5.5.3.2. Perform Person Search . . . . .	63
5.5.3.3. Query New First Name . . . . .	64
<b>6. Analysis</b>	<b>65</b>
6.1. Export data . . . . .	65
6.1.1. Options . . . . .	66
6.1.2. Export file format . . . . .	66
6.1.3. Titles, variable names . . . . .	66
6.1.3.1. Titles . . . . .	66
6.1.3.2. Short variable name . . . . .	67
6.1.3.3. Long Variable name . . . . .	67
6.1.4. Variables . . . . .	67
6.1.5. Date Format . . . . .	67
6.1.6. Export setup . . . . .	68
6.2. Table builder . . . . .	68
6.2.1. Example . . . . .	68
6.2.2. File formats . . . . .	71
6.2.2.1. Portable Document Format (PDF) . . . . .	74
6.2.2.2. Post Script (PS) . . . . .	74
6.2.2.3. Scalable Vector Graphics (SVG) . . . . .	74
6.2.2.4. Portable Network Graphics (PNG) . . . . .	74
6.2.2.5. Windows Meta File (WMF) . . . . .	74
6.2.2.6. Tabulated data (HTML) . . . . .	74
6.2.3. CanReg5 Chart Editor . . . . .	74
6.2.4. CanReg and SEER*Stat . . . . .	74
6.2.4.1. In CanReg . . . . .	75
6.2.4.2. In SEER*Prep . . . . .	75
6.2.4.3. In SEER*Stat . . . . .	77
6.3. Frequency distributions . . . . .	77
6.3.1. Pivot tables . . . . .	79
<b>7. Management</b>	<b>81</b>
7.1. Back up and restore . . . . .	81
7.1.1. Perform backup . . . . .	81
7.1.2. Restore from backup . . . . .	81
7.1.2.1. Restore with “Install new system” . . . . .	82

7.1.2.2. Restore with “Restore from backup”-tool . . . . .	82
7.2. Manage users . . . . .	82
7.2.1. Change your own password . . . . .	83
7.2.2. Database password . . . . .	83
7.2.3. User manager . . . . .	83
7.3. Quality control . . . . .	84
7.3.1. Name and sex . . . . .	84
7.3.2. Person search . . . . .	86
7.3.2.1. Advanced . . . . .	86
7.3.2.2. Results . . . . .	87
7.4. Options . . . . .	89
7.4.1. [UPDATED]Language . . . . .	89
7.4.2. Look & Feel . . . . .	89
7.4.2.1. Change font . . . . .	89
7.4.2.2. Change font style . . . . .	89
7.4.2.3. Show only outline . . . . .	90
7.4.3. Periodic backup . . . . .	90
7.4.4. CanReg version . . . . .	91
7.4.5. Paths . . . . .	92

## IV. Appendix 93

<b>A. Frequently asked questions (FAQ)</b>	<b>95</b>
A.1. Server . . . . .	95
A.2. Conversion CanReg4 to CanReg5 . . . . .	95
A.3. Dictionary . . . . .	96
A.4. Analysis/tables . . . . .	96
A.5. Import . . . . .	97
A.6. Database design . . . . .	97
A.7. Install/update CanReg5 . . . . .	99
<b>B. Known issues</b>	<b>101</b>
B.1. Known bugs (errors) . . . . .	101
B.2. Known limitations . . . . .	101
<b>C. Migrating from CanReg4 - Step by Step</b>	<b>103</b>
C.1. Step 1 - Import the variable definitions of CanReg4 to CanReg5 . . . . .	103
C.2. Step 2 - Import the dictionary from your CanReg4 installation . . . . .	105
C.3. Step 3 - Import the data from your CanReg4 installation . . . . .	106

<b>D. Advanced database management</b>	<b>109</b>
D.1. The System definition file (the XML file)	109
D.1.1. Advanced options in the CanReg5 person search	109
D.1.1.1. Weight	109
D.1.1.2. Discriminatory power	109
D.1.1.3. Reliability	109
D.1.1.4. Prescence	109
D.1.1.5. Threshold	109
<b>E. Technical details on the Table Builder</b>	<b>111</b>
E.1. Configuration files	111
E.1.1. The file structure and parameters	111
E.1.1.1. table_label	111
E.1.1.2. table_description	111
E.1.1.3. table_engine	111
E.1.1.4. engine_parameters	113
E.1.1.5. preview_image	113
E.1.1.6. ICD_groups_labels	113
E.1.1.7. ICD10_groups (or ICD_groups)	113
E.2. Engines	114
E.2.1. Incidence Rates Table(“incidencerates”)	114
E.2.1.1. line_breaks	114
E.2.2. Number of Cases ( “number of cases”)	114
E.2.2.1. line_breaks	115
E.2.3. Population Pyramids ( “populationpyramids”)	115
E.2.4. Top 10 pie chart ( “top10piechart”)	115
E.2.5. The R engines ( “r-engine” and “r-engine-grouped”)	115
E.3. CanReg and R	115
E.3.1. Configuration file parameters	115
E.3.1.1. r_scripts	116
E.3.1.2. file_types_generated	116
E.3.2. Arguments	117
E.3.3. Population file	118
E.3.4. Incidence file	118
E.3.4.1. r-engine	118
E.3.4.2. r-engine-grouped	119
E.3.5. SVG support	119
<b>F. Third party tools</b>	<b>121</b>
F.1. Java Runtime Environment	121
F.1.1. Issues with memory	121
F.1.1.1. On Windows:	121
F.1.1.2. On Solaris/Linux:	121

F.2. Analysis	122
F.2.1. R	122
F.2.2. Epi Info	122
F.2.3. SEER*Stat and SEER*Prep	122
F.2.4. PSPP	122
F.2.5. SOFA	123
F.3. Data cleaning and record linkage	123
F.3.1. Google Refine	123
F.3.2. Talend Open Studio	123
F.3.3. FRIL: Fine-Grained Records Integration and Linkage Tool	123
F.3.4. LinkPlus	123
F.4. Data security	124
F.4.1. GnuPG	124
F.4.2. TrueCrypt	124
F.5. Graphics display/editing	124
F.5.1. PostScript Viewer	124
F.5.2. Inkscape	125
F.6. General	125
F.6.1. Notepad++	125
<b>G. Scripts</b>	<b>127</b>
G.1. Ruby scripts	127
G.1.1. Script to generate icd10 codes and cancer groups	127
G.1.1.1. Prerequisites	127
G.1.1.2. Usage	127
<b>H. About</b>	<b>129</b>
H.1. Copyright	129
H.2. Credits	129
H.2.1. Lead developement and design:	129
H.2.2. Additional developement	129
H.2.3. Additional design	129
H.2.4. Consultants:	130
H.2.5. Technical consultants	130
H.2.6. Thanks	130
H.2.7. Alpha-testers	131
H.2.8. Beta-testers	131
H.2.9. Translators	131
H.3. References	132
H.4. Acknowledgments	132
H.5. Licenses	132
<b>I. Changelog</b>	<b>133</b>



# Part I.

## Introduction



# 1. What is CanReg?

CanReg5 is an open source tool to input, store, check and analyse cancer registry data. It has modules to do data entry, quality control, consistency checks and basic analysis of the data. The main improvements from the previous version are the new database engine, the improved multi user capacities and that the development is managed like an open source project. Also included is a tool to facilitate the set up of a new or modification of an existing database by adding new variables, tailoring the data entry forms etc.

## 1.1. A short introduction to the database structure in CanReg5

New in CanReg5 is the three level table structure. Where CanReg4 stored everything in one big table of tumours, CanReg5 splits this information in three tables: Patient, Tumour and Source. For each patient, you can store as many tumour records as you need, and for each tumour you can store as many source records as you need. This allows us to do more with our databases, for example related to completeness by counting number of sources, but it poses some problem that might require manual intervention during the conversion process of a system from CanReg4 to CanReg5.

For example, some of you might store multiple sources in each tumour record in CanReg4. This should be split into several source records for this tumour record in CanReg5, but this is not an easy task to automate since all registries that have opted for this have solved it in a different way in CanReg4.

One way around this is to put all the fields related to the source table in CanReg5 so that you are sure not to lose any data, and then start from the date you start using CanReg5 to store multiple source records per tumour.

The best way, but a more time consuming one, is to set up the source table (by editing the system definition XML) to only contain the data you want to store per source and then work on the exported file from CanReg4 and import additional source information at a later stage. (General import of data is not yet functioning adequately in CanReg5 - only import & migration of old data.)

You can of course choose not to use the source table, as well - just record the source information per tumour like you would in CanReg4. This can be set up while migrating your system definition files.

## 1. What is CanReg?

### 1.2. Forum, Issue tracker, community site, twitter

We have created a project page at GitHub to help us keep track of issues with CanReg5: <https://github.com/IARC-CSU/CanReg5>. Using tools there allows you to see what error reports other CanReg users have already filed and if solutions have already been proposed and also discuss potential improvements for CanReg5, before posting your own queries.

You can of course still send us emails at [canreg@iarc.fr](mailto:canreg@iarc.fr).

If you encounter any problems please provide a description of it along with the specifications of your computer. (Operating System (Windows XP, Vista, 7, OSX, Linux?), memory, processor speed etc.) Also it would be very useful if you can precise the **version** and the **build code** of your CanReg5. This can be found on the bottom left of the welcome screen and on the “About” screen. (For example Version: “4.99.0b586”)

Please be aware that some problems can be avoided by installing the latest version of the Java Runtime Environment (Version 6 or newer) before you start. (Available from <http://java.com/en/download/manual.jsp>.)

Videos documenting certain operations described below can be downloaded from: <http://www.iacr.com.fr/CanReg5/videos.zip>.

Last, but not, least: CanReg has its own stream on twitter. Please follow: <http://twitter.com/canreg> for updates.

### 1.3. Getting hold of the latest version of CanReg

If you are running version 4.99.1 (or newer) of CanReg, you can launch CanReg and click “Options...”, go to the “Advanced” tab. (See There you see your current version, i.e. 4.99.1. If you click “Check” CanReg will look for an updated version. Afterwards you can click “Download latest version” to get the zip-file containing the most recent version of CanReg5.

(If you have version 4.99.0 or no CanReg5 at all you can download the newest version from here: <http://www.iacr.com.fr/CanReg5/CanReg5.zip>)

### 1.4. Logfile

CanReg generates a logfile when you run it. This file is called canreg5client.log and is located in your home folder. (On windows it is most probably under C:\Documents and Settings\<your username> - you can access this by browsing to %HOMEPATH%

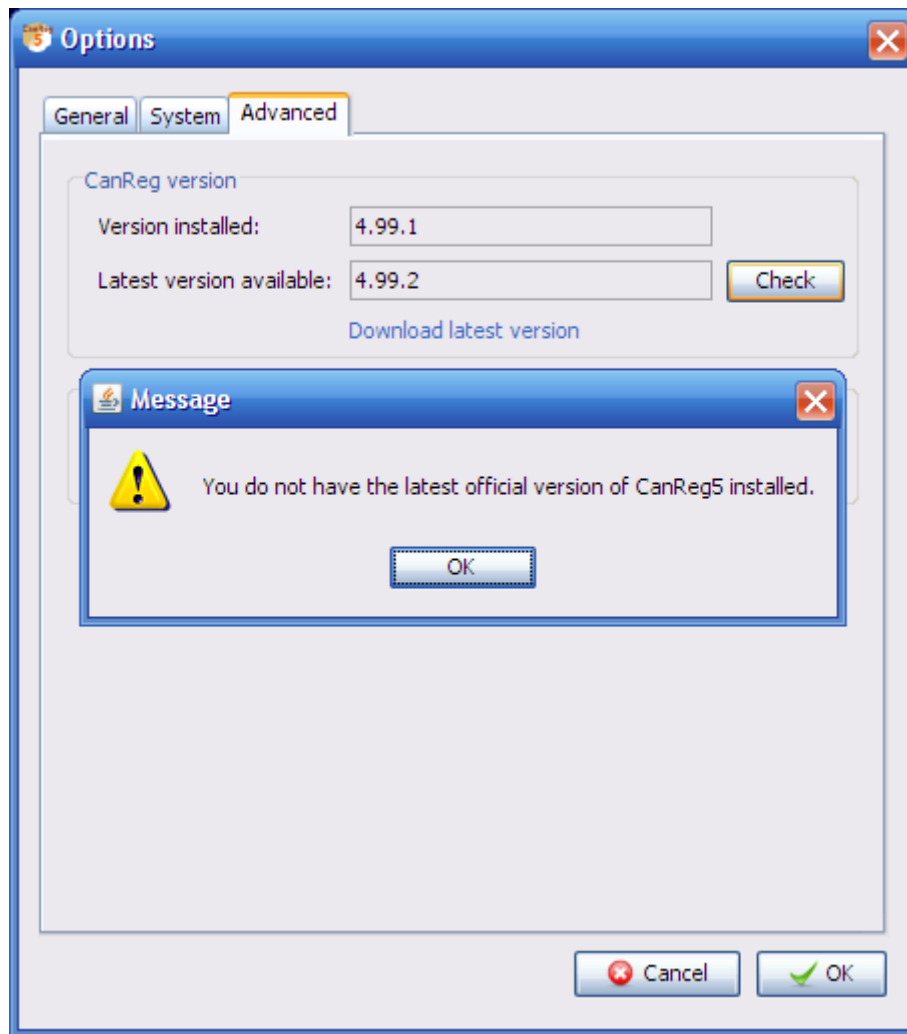


Figure 1.1.: Checking if you have the latest version of CanReg installed.

Depending on how you have configured your computer it might show up as canreg5client.) If you can attach this to emails with feedback/queries it would be very useful. Please note that this file is overwritten each time CanReg is started, so you need to “take it” just after, for example, your error occurs.

## 1. What is CanReg?

Example content of a log-file:

```
<?xml version="1.0" encoding="windows-1252" standalone="no"?>
<!DOCTYPE log SYSTEM "logger.dtd">
<log>
  <record>
    <date>2009-06-25T16:09:27</date>
    <millis>1245938967921</millis>
    <sequence>0</sequence>
    <logger>canreg.client.CanRegClientApp</logger>
    <level>INFO</level>
    <class>canreg.client.CanRegClientApp</class>
    <method>startup</method>
    <thread>10</thread>
    <message>CanReg version: 4.99.9b668 (20090625160546)</message>
  </record>
  ...
</log>
```

## Part II.

### Installing and running CanReg5





## 2. Software installation

### 2.1. Install a recent Java Runtime Environment

Before you install and run CanReg5 for the first time it is recommended that you install the latest Java 7 Runtime Environment (March 2016: Version 8 Update 77 )  
You can get that from here: <http://java.com/en/download/manual.jsp>

### 2.2. Install CanReg5

CanReg5 is compatible with most major operating systems and the default distribution of CanReg is simply a zip-archive.

#### 2.2.1. From CanReg5.zip

To install it simply extract the content to a new folder, for example on your desktop.  
(It is important to keep the same directory structure as inside the zip-file.)

If you have tools like 7-zip or WinRAR installed it is just a matter of copying/moving the zip-file to the folder you want CanReg to reside, right-clicking it and choose “Extract here...”

#### 2.2.2. From CanReg5-Setup.zip (Windows only)

An alternative way to install CanReg is by running the CanReg5-Setup.exe from within CanReg5-Setup.zip. This is a standard windows installer that will install CanReg5 in your “Program Files” folder (by default). It is just a matter of clicking “next” ,”next”, “next” etc.

### 2.3. Recommended Third Party tools

It is recommended that you also install the following third party software:

#### 2.3.1. Any post script viewer

CanReg can produce tables in the post script format. To view these you need a post script viewer like GhostView. (See [F.5.1 on page 124.](#))

## 2. Software installation

### 2.3.2. R

Some of the more advanced analytical features in CanReg relies on the free software called R. Please install this if you need to use that functionality. (See [F.2.1 on page 122](#).) As of version 5.00.31, CanReg5 has a built in offline R package installer that can be installed in parallel with CanReg5 and called from CanReg5 using the menu option “Tools”->“Install R packages”. (This is only visible if you have installed the R package installer.)

## 2.4. Un-install CanReg5

If you wish to un-install CanReg5, delete the following folders (or rename if you have anything valuable in them):

.CanRegClient and .CanRegServer in your user folder.

(On Windows you can get to your user folder by pressing Windows Key + R and then entering %UserProfile% and click OK. See [http://en.wikipedia.org/wiki/Home\\_directory](http://en.wikipedia.org/wiki/Home_directory) for more information.)

Afterwards you can delete the folder from [2.2](#).

## 3. First launch

### 3.1. Run CanReg5

Go to the folder you installed CanReg5 in and double click on the coffee cup icon (CanReg.jar). (If, at this point, CanReg5 does not start you might have to update your Java Runtime Environment and retry. (See 2.1 on page 17.)

You should now be presented with the CanReg5 welcome window (See figure 3.1).

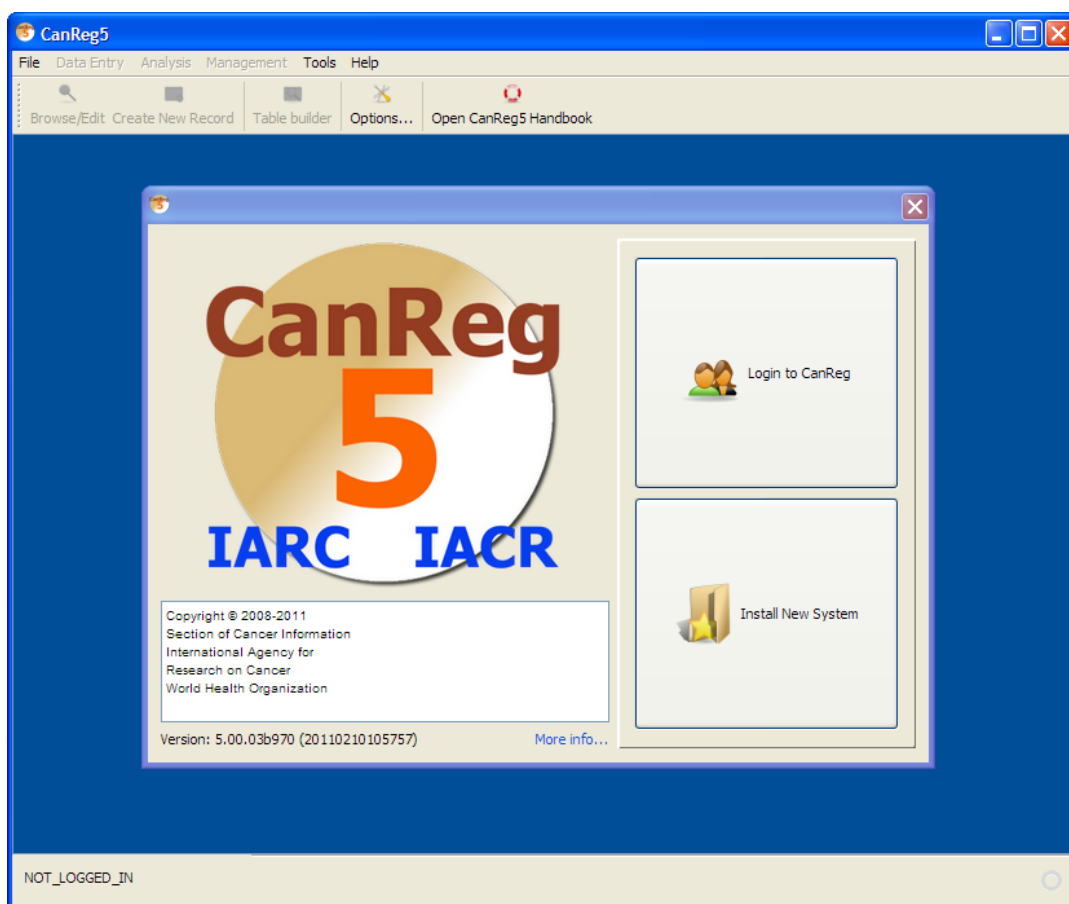


Figure 3.1.: The welcome window

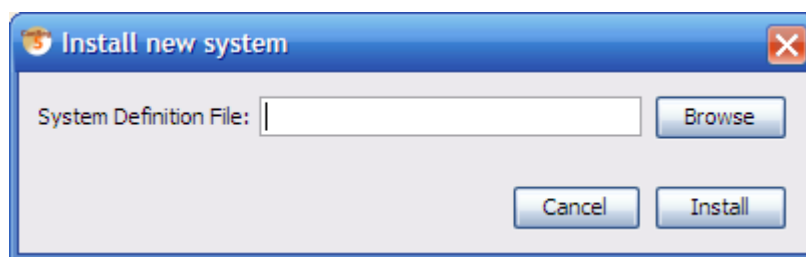
### 3. First launch

## 3.2. Demo system

Included is the dictionary for the training system located in the demo-folder in the zip-file. With this you can get a demo-system up and running to test some data entry. To run this demo system, install the TRN-file to set up the system. Afterwards you should import the dictionary using “Data Entry” - ”Edit dictionary” and “Import complete dictionary from a file”, before you start to enter data.

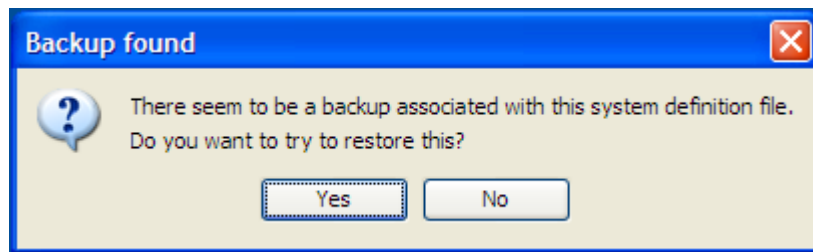
## 3.3. Install a new CanReg5 system

If you want to install an already provided system definition (for example the demo system TRN) please click “Install New System”. CanReg5 will present you with the following message:



Click browse to find the system definition file. (If you want to use the TRN demo system look in demo/database and select TRN.xml and click open and then Install.)

If the XML file is associated with a backup of CanReg5 the program will ask you if you want to restore from this backup.



If you click yes CanReg will launch the server with the newly installed system definitions and restore this backup.

## 3.4. Convert the CanReg4 system definitions

If you have a CanReg4 system you can use tools built into CanReg to help you migrate this to CanReg5.

First import the variables of CanReg4 to CanReg5 - the system definition of CanReg4.

### 3.4. Convert the CanReg4 system definitions

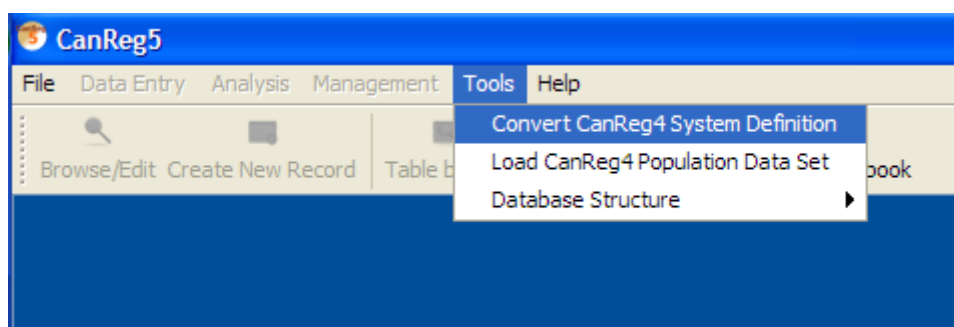


Figure 3.2.: CanReg systems definition converter menu

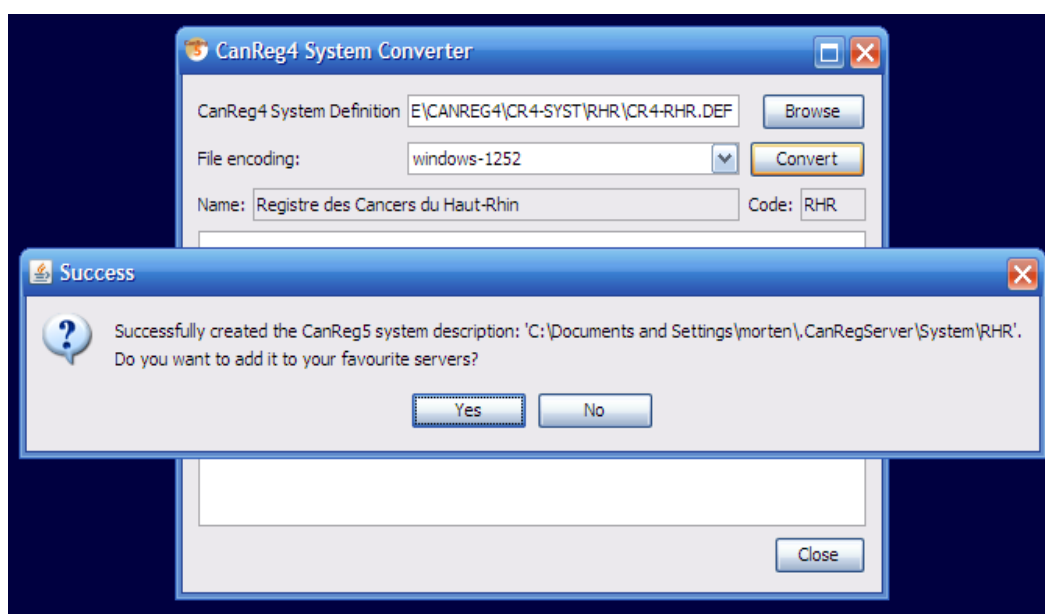
Go to “Tool” in CanReg5 menu and click “Convert system definition” (See figure 3.2.)

Do “Browse” to find your CanReg4 system definition file. (This is a file located in the folder `\\CR4SHARE\CANREG4\CR4-SYST\` followed by your 3 letter registry code i.e. TRN whose name is ending in .DEF (i.e. CR4-TRN.DEF).)

Select your CanReg4 file and double click it or click “Open”.

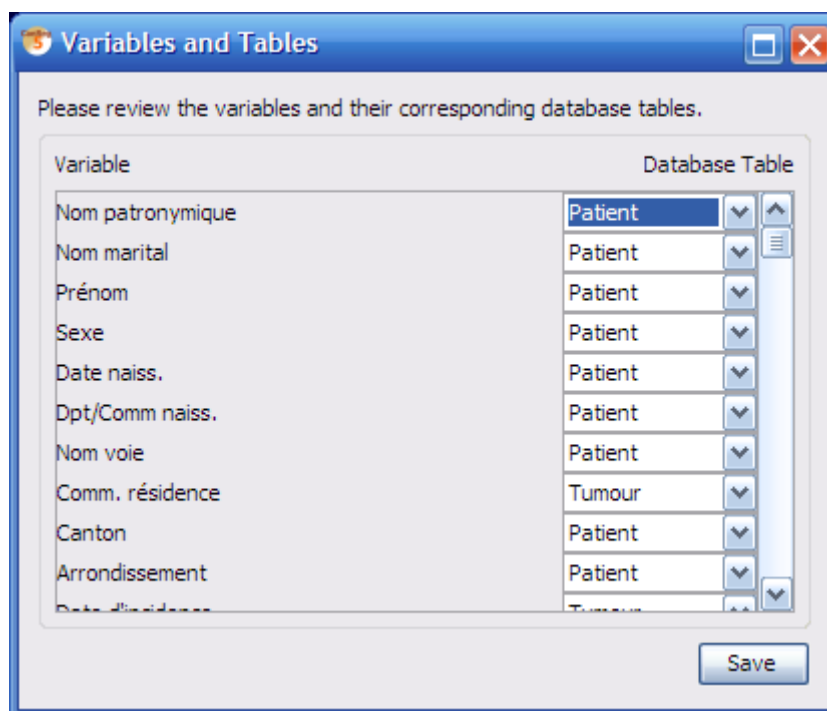
Click “Convert”.

The program will then ask you if you want to add this server to your favourites. Click “Yes” here.



### 3. First launch

The next step is the trickiest one during the conversion. Since we go from a tumour based database structure with only one big table with all the tumour and patient related information to a structure with both a table for tumour related information and patient related information we need to specify what variable goes in what table of CanReg5. We recommend putting the unique patient related information (name, date of birth and follow-up variables) in the patient table, source information in the source table and pretty much the rest (tumour information, age, address etc) in the tumour table.



The program presents an initial proposal that you might agree with, but please go through one by one the variables and decide.

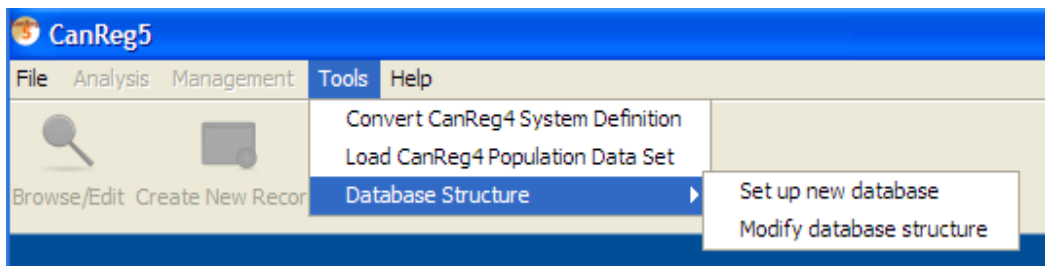
Click “Save”. You have now created an XML file that describes your CanReg5 system.

Optional: Before you proceed to the next step and launch the server you can, if you want(!), manually edit this XML file you have created by opening it in a text editor or a dedicated XML editor. The file is located in your user folder under .CanRegServer. (On my machine, for example, running Windows XP it is under: C:\Documents and Settings\morten\.CanRegServer\System.)

## 3.5. Setting up or modifying a CanReg system using the built in editor

To modify an existing CanReg system or set up a new one you can use tools built into CanReg5.

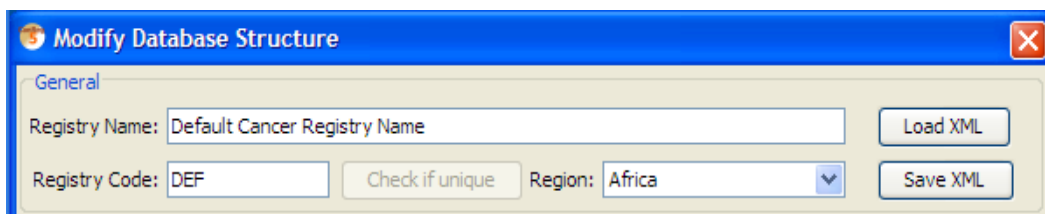
### 3.5. Setting up or modifying a CanReg system using the built in editor



They can be found under Tools -> Database Structure.

**Note: Before using this tool it is highly recommended to perform a backup of your CanReg5 database!**

For now please use “Set up new database”. This will give you the Modify Database Structure window.



Please click Load XML and either pick your own XML or, if you start from scratch, the TRN.xml or DEF.xml found in the CanReg installation. (You will need to load an existing XML to be able to create a working XML for your CanReg system.)

Please note that certain modifications done using this tool will impact the structure of the CanReg database to such an extent that it will have to be rebuilt afterwards. Others like renaming groups, changing the displayed name of a variable or reordering the variables are purely cosmetic and do not impact the database structure as such. If you wish to do changes to the structure of the database you'll need to export your data prior to those changes, delete the database files of the CanReg system, do required modifications using this tool or directly in the XML, relaunching the CanReg server and then import the data (this again will potentially have to be adapted to the structural changes). CanReg will warn you if you have done changes that require a rebuild of the database. (Fields with pink background in this editor - as well as adding or removing variables, leads to this.)

When this has loaded you'll see all the info specifying this CanReg system. On the top you can specify the registry name, registry code and region of the registry. Below you have a list of the Dictionaries, then the Groups and then the Variables.

### 3. First launch

**Modify Database Structure**

Registry Name:

Registry Code:   Region:

Dictionaryes Groups **Variables** Search Variables Coding Settings Indexes

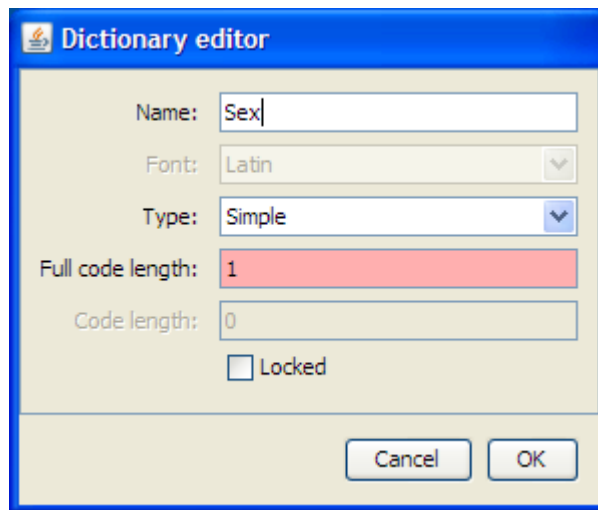
* Hospital (StdVar: Source1) (Dict: Hospital Group: Hospital)	<input type="button" value="Add"/>	<input type="button" value="Remove"/>	<input type="button" value="Edit"/>
Unit (Dict: Unit Group: Hospital)	<input type="button" value="Add"/>	<input type="button" value="Remove"/>	<input type="button" value="Edit"/>
Path lab n° (Text, Length: 8, Group: Hospital)	<input type="button" value="Add"/>	<input type="button" value="Remove"/>	<input type="button" value="Edit"/>
Case n° (Text, Length: 6, Group: Hospital)	<input type="button" value="Add"/>	<input type="button" value="Remove"/>	<input type="button" value="Edit"/>
* Date last contact (StdVar: Lastcontact) (Date, Group: Follow	<input type="button" value="Add"/>	<input type="button" value="Remove"/>	<input type="button" value="Edit"/>
Status (Dict: Status Group: Follow up)	<input type="button" value="Add"/>	<input type="button" value="Remove"/>	<input type="button" value="Edit"/>
* ICC code (StdVar: ICC) (Text, Length: 3, Group: Tumour)	<input type="button" value="Add"/>	<input type="button" value="Remove"/>	<input type="button" value="Edit"/>
MiddleName (Text, Length: 15, Group: Patient)	<input type="button" value="Add"/>	<input type="button" value="Remove"/>	<input type="button" value="Edit"/>

To add a dictionary, group or variable, click add in the proper pane. This will then appear as the last item in the corresponding list for you to edit.

Clicking edit on any button related to a dictionary, group or variable brings up the respective editor.



### 3.5.1. Modifying a dictionary

The image shows a 'Dictionary editor' dialog box with a blue title bar. It contains several input fields: 'Name' with the text 'Sex', 'Font' with a dropdown menu showing 'Latin', 'Type' with a dropdown menu showing 'Simple', 'Full code length' with a red background and the value '1', and 'Code length' with the value '0'. There is also a 'Locked' checkbox which is unchecked. At the bottom right are 'Cancel' and 'OK' buttons.

Using the dictionary editor you can modify any dictionary in CanReg5. The fields are as follows:

**Name:** The name of the dictionary

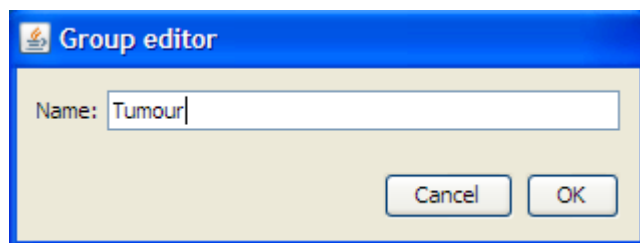
**Type:** This can either be “Simple” or “Compound”. A “Simple” dictionary is a plain list of codes and corresponding labels, whereas a “compound” dictionary has two levels of refinement. For example the user can pick the two first digits and then the last digit, as in the above example.

**Full code length:** The number of characters the codes for this dictionary takes up in the database.

**Code length:** The number of characters in the first level of refinement in the case of a compound dictionary.

**Locked:** Will you allow the super user to modify this dictionary using the tools in CanReg, or should it be locked?

### 3.5.2. Modifying a group

The image shows a 'Group editor' dialog box with a blue title bar. It contains a single input field 'Name' with the text 'Tumour'. At the bottom right are 'Cancel' and 'OK' buttons.

Using the group editor you can modify any group in CanReg5.

**Name:** The name of the group

### 3.5.3. Modifying a variable

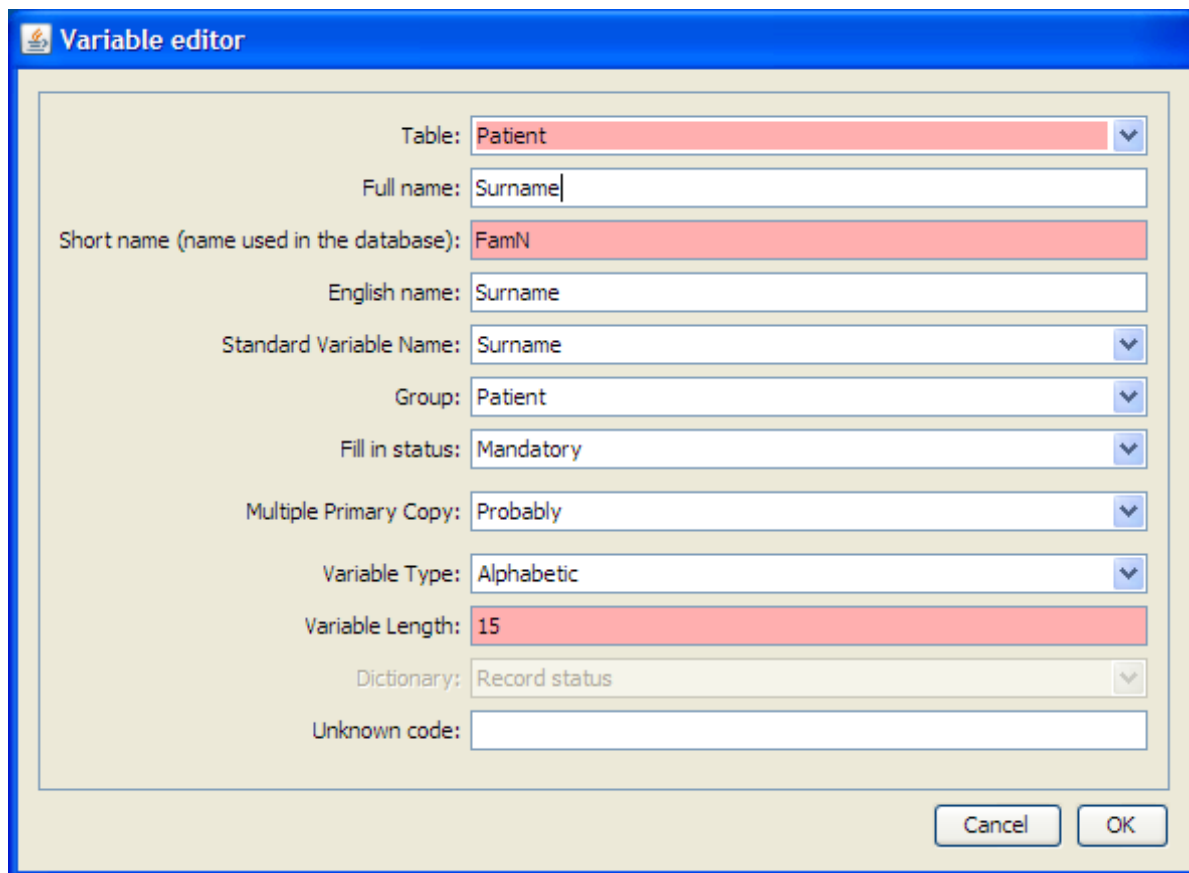
A screenshot of the 'Variable editor' dialog box. The dialog has a blue title bar with the text 'Variable editor' and a small icon. The main area is light beige and contains several fields and dropdown menus. The fields are: 'Table:' with a dropdown showing 'Patient'; 'Full name:' with a text box containing 'Surname'; 'Short name (name used in the database):' with a text box containing 'FamN'; 'English name:' with a text box containing 'Surname'; 'Standard Variable Name:' with a text box containing 'Surname'; 'Group:' with a dropdown showing 'Patient'; 'Fill in status:' with a dropdown showing 'Mandatory'; 'Multiple Primary Copy:' with a dropdown showing 'Probably'; 'Variable Type:' with a dropdown showing 'Alphabetic'; 'Variable Length:' with a text box containing '15'; 'Dictionary:' with a dropdown showing 'Record status'; and 'Unknown code:' with an empty text box. At the bottom right are 'Cancel' and 'OK' buttons.

Table: Patient

Full name: Surname

Short name (name used in the database): FamN

English name: Surname

Standard Variable Name: Surname

Group: Patient

Fill in status: Mandatory

Multiple Primary Copy: Probably

Variable Type: Alphabetic

Variable Length: 15

Dictionary: Record status

Unknown code:

Cancel OK

Using the variable editor you can modify any variable stored in the CanReg5 database. The fields are as follows:

**Full name:** The name of the variable as displayed in data entry forms etc.

**Short name:** The name of the variable in the database. (This should be without any blanks and other special characters and reasonably short.)

**English name:** It is useful to provide an English name for the variable in case you want to collaborate with people in other countries.

**Standard Variable Name:** This maps the variable to a standard CanReg5 variable for the purpose of edit checks and analysis.

**Group:** The choice of group only affects the display during data entry.

**Fill in status:** Can be set to “Mandatory”, “Optional”, “Automatic” or “System”, depending on if you want to force the registrar to provide this information before confirming the record.

**Multiple Primary Copy:** Legacy information. Leave as “Other”.

**Variable Type:** Can be “Alphabetic” (for plain text), “Asian text” (legacy field, same as “Alphabetic”), “Date”, “Dictionary”, “Number” and “Text Area”.

**Variable Length:** The length of the variable in characters.

**Dictionary:** If you chose “Dictionary” as type of variable you’ll have to choose a dictionary here.

**Unknown code:** Here you can specify the unknown code of this variable.

**Table:** Choose the table where this variable should be stored.

### 3.5.4. Set up person search variables

The 'Search Variables' editor interface shows a list of variables with their respective weights and a 'Remove' button for each. The variables are: Surname (Weight: 25.0), First names (Weight: 20.0), Birth date (Weight: 16.0), Sex (Weight: 5.0), and Tribe (Weight: 8.0). At the bottom, there is a 'Minimum match' field set to 70.0% and an 'Add variable' button.

Variable	Weight	Action
Surname	25.0	Remove
First names	20.0	Remove
Birth date	16.0	Remove
Sex	5.0	Remove
Tribe	8.0	Remove

Minimum match: 70.0 %

Add variable

Using this editor you can change the variables that come into play during person search in CanReg5, and their respective weights and minimum match criteria.

### 3.5.5. Coding

The 'Coding' editor interface shows fields for Male code (1), Female code (2), and Unknown sex code (9). It also has a Date Format dropdown (dd/mm/yyyy), a Date Separator field (/), a Morphology Length field (4), and a checkbox for IARC/IACR Basis of Diagnosis codes (checked).

Male code: 1 Female code: 2 Unknown sex code: 9

Date Format: dd/mm/yyyy Date Separator: /

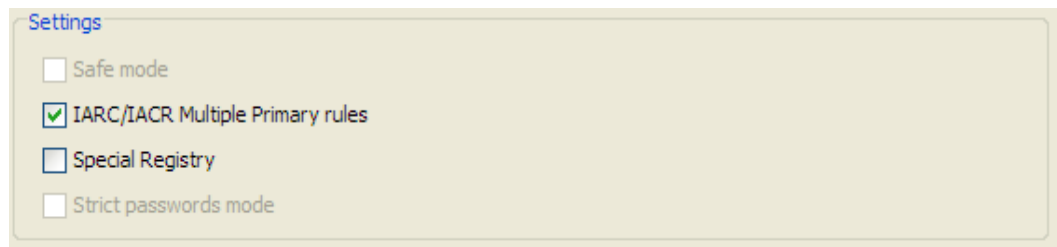
Morphology Length: 4

☒ IARC/IACR Basis of Diagnosis codes

### 3. First launch

Here you can change some coding settings of your CanReg system. (Not yet implemented.)

#### 3.5.6. Settings



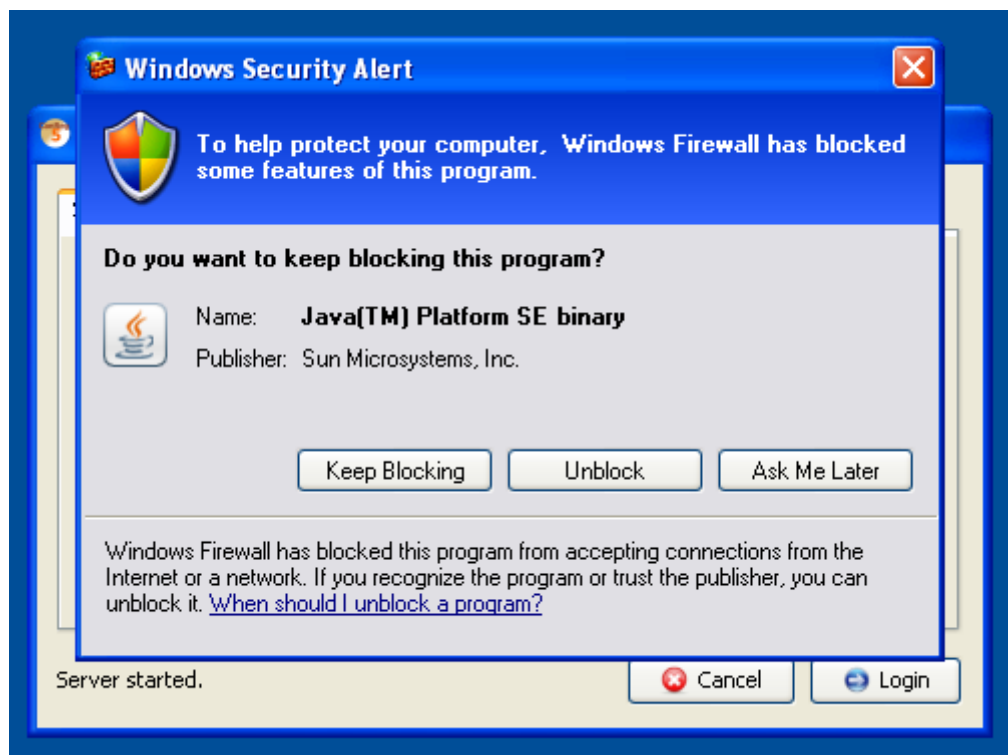
Here you can change some settings of your CanReg system. (Not yet implemented.)

#### 3.5.7. Saving the system

By clicking Save XML the system XML will be saved to the system folder of CanReg under the name <your system code>.xml (for example TRN.xml), ready for use.

### 3.6. Launching the CanReg server

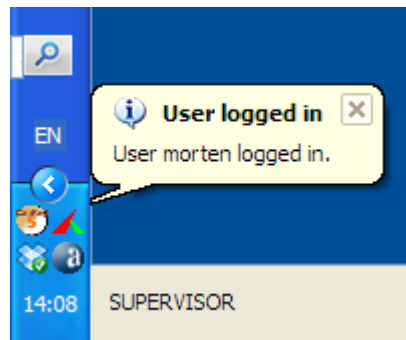
After clicking “Login” on the welcome screen of CanReg you get the login screen. To launch the CanReg server click “Settings”. Click “Launch Server”.



If you get a java firewall query, please confirm that it is OK that java can communicate through your firewall by clicking “Unblock”, “OK” or “Yes”. If this is the first time you launch the server on this machine it will automatically create the database needed for CanReg5.

If your database is encrypted you will get a message regarding this and you need to provide the database password to open it. Please note that this can be different from your user’s password.

When the CanReg5 server is running on your machine you should see an icon in your notification area.



#### 3.6.1. Launching server from command line

To launch CanReg5 from command line, you’ll have to open command line on your computer (Windows: Win+R, “cmd”) change into the folder you have *installed* CanReg5 (Windows: “cd %programfiles(x86)%\CanReg5”) and type the following to launch the server:

“java -jar CanReg5.jar <registry-code>” (where <registry-code> should be replaced by the code of your registry, ie “java -jar CanReg5.jar TRN-E”)

## 3.7. Login

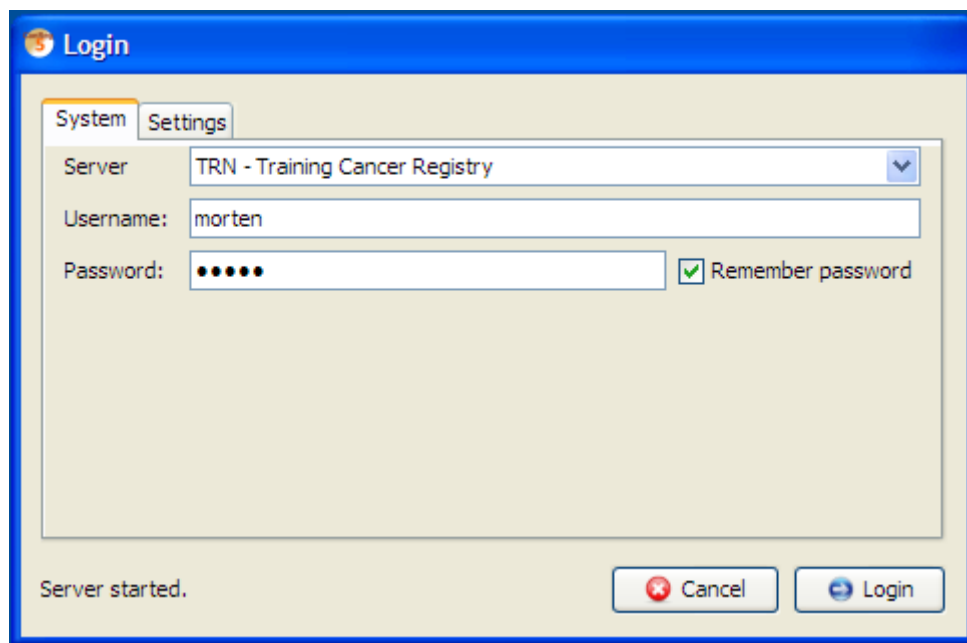
After launching the server you can log on to your CanReg system.

#### 3.7.1. Locally

If you want to log in to a CanReg server running on your local machine, after either installing a CanReg5 system XML or converted your CanReg4 system definition files you go to the “System” tab of the “Login”-window and choose your server from the drop down list. (Most probably already selected.) (The default username is “morten” and password is “ervik”. (All in small letters with no double quotes.) Click “Login” and you’ll be logged on.)

### 3. First launch

If you get an error message saying “Could not log in to the CanReg server on localhost with the given credentials.”, please make sure that you have entered the correct username and password and that the server is indeed running. (See “Launching the CanReg server” above.)



#### 3.7.2. In a network

If you want to log on to a CanReg server running on another machine in your network you need to know the address of that machine. (Either it's IP address or name on the network.)

To find the IP address of a CanReg server you can go to the Settings tab on the “Login”-window and tick “Advanced” to get access to some more advanced tools, like the “Get IP Address” tool. Click this and you will get a message saying “The IP address of (your machine) is www.xxx.yyy.zzz. (Most probably something like 10.0.0.x or 192.168.0.x.) Take a note of those numbers.

Launch CanReg on the machine you want to run CanReg on. Click “Login” to get to the “Login” screen. There you can click “Settings” and type the IP address, www.xxx.yyy.zzz, you found above in the “Server URL” field along with the system code for your registry. (For example TRN.) If you click “Add server to list” the program will test the connection to the server and if this is OK this network server will be added to the list of servers you can log in to from this CanReg installation.

Click the “System” tab and choose this networked server from the drop down list of servers, enter username and password. (The default username is “morten” and password is “ervik”. (All in small letters with no double quotes.) Click “Login” and you'll be logged on.) Click “Login” and you'll be logged on.)

### 3.8. Import the dictionaries from CanReg4

If you get an error message saying “Could not log in to the CanReg server on localhost with the given credentials.”, please make sure that you have entered the correct username and password and that the server is indeed running. (See “Launching the CanReg server” above.)

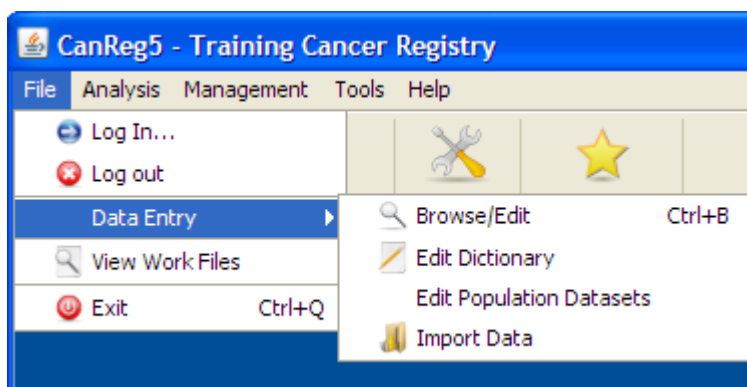
The next time you want to log on to this server all you have to do is launch CanReg, select this server, enter username and password and click “Login”.

Please note that you do *not* need to install the system definition or convert from CanReg4. You only need the code (for example TRN) of the CanReg system you want to connect to.

## 3.8. Import the dictionaries from CanReg4

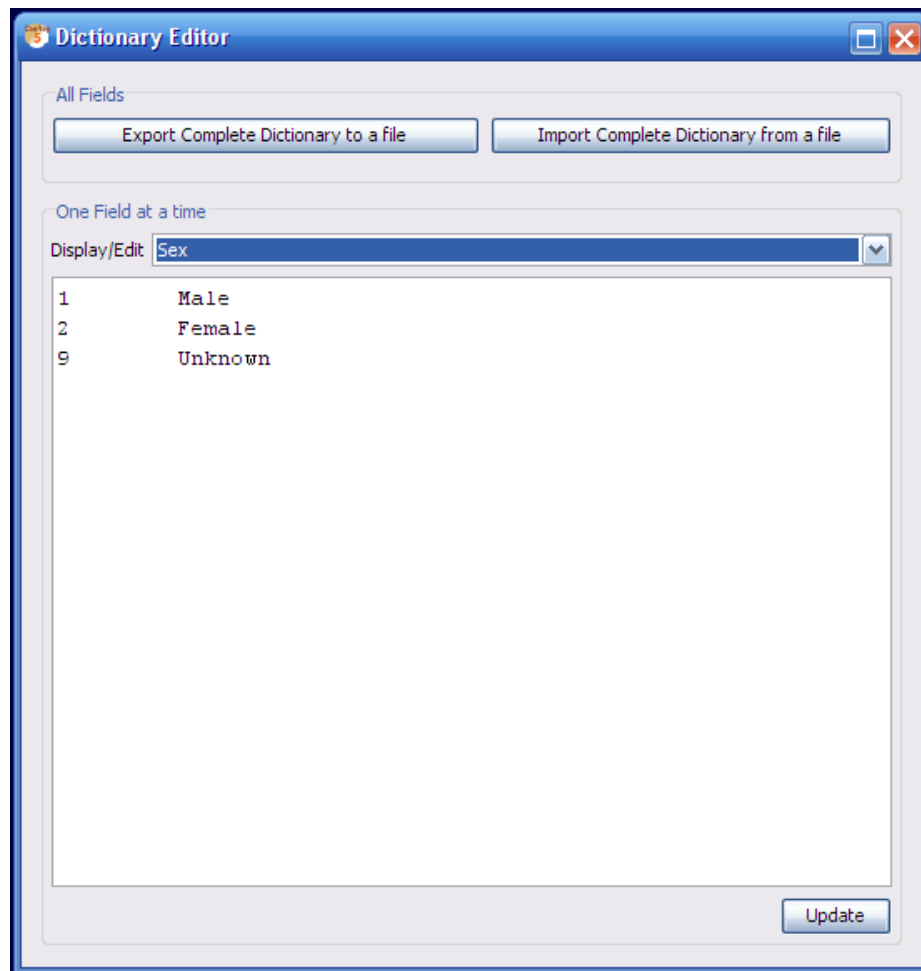
If you are migrating from CanReg4 make sure to export the most updated dictionary from your CanReg4 system. (In CanReg4: “Data Entry”, “Dictionary”, “Export dictionary to text file”) If you want to use the demo system, the dictionary is located in: demo/dictionary.

- Go to “File”, “Data Entry”, “Edit dictionary” in CanReg5



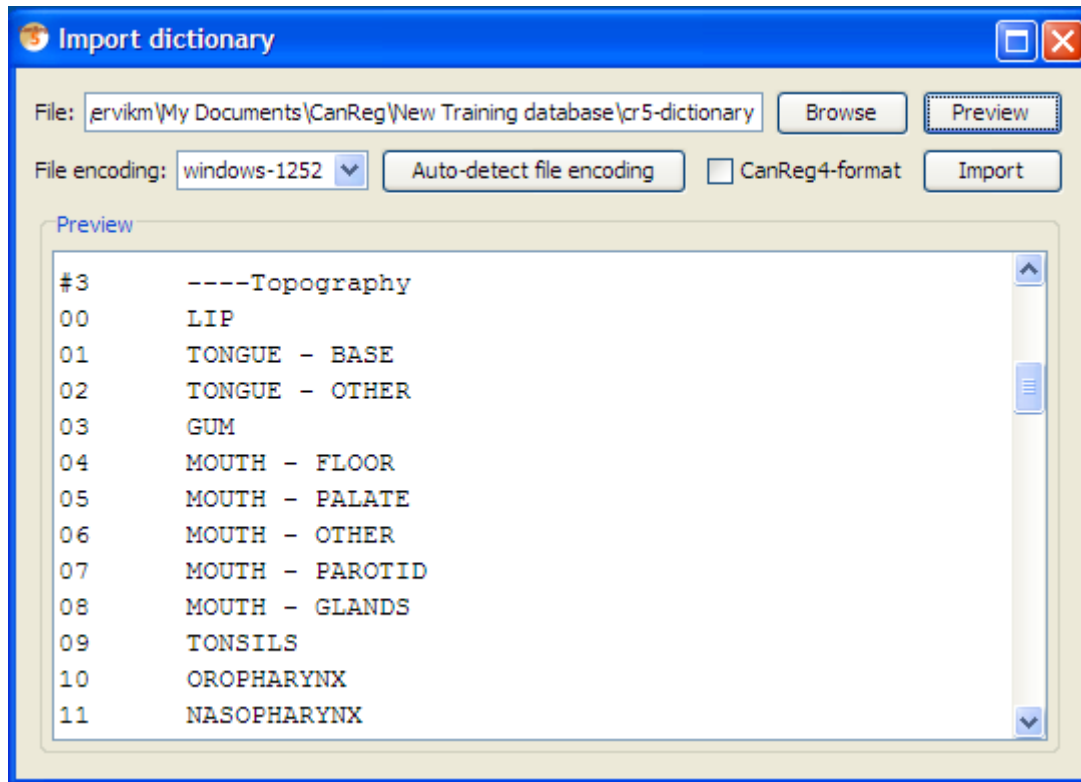
- Click on “Import complete dictionary from file”.

### 3. First launch



- Browse and select the dictionary from you CanReg4 work folder or elsewhere.
- Do Preview
- Tick “CanReg4 Format” if you are migrating from CanReg4, leave unticked if you are using the demo system or otherwise are importing a CanReg5 formatted dictionary.





- Click Import. This might take some time. Please note the bar in the lower right indicating that the program is busy.
- Afterwards you will receive a message of success imported.
- Click OK.
- Go back to “File”, “Data Entry”, “Edit dictionary” and verify that the dictionaries have been imported.

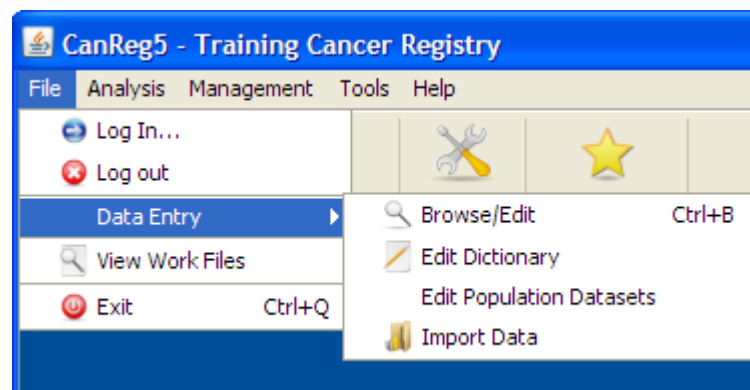
### 3.9. Import the data from CanReg4

Make sure to export the most updated data from your CanReg4 system.

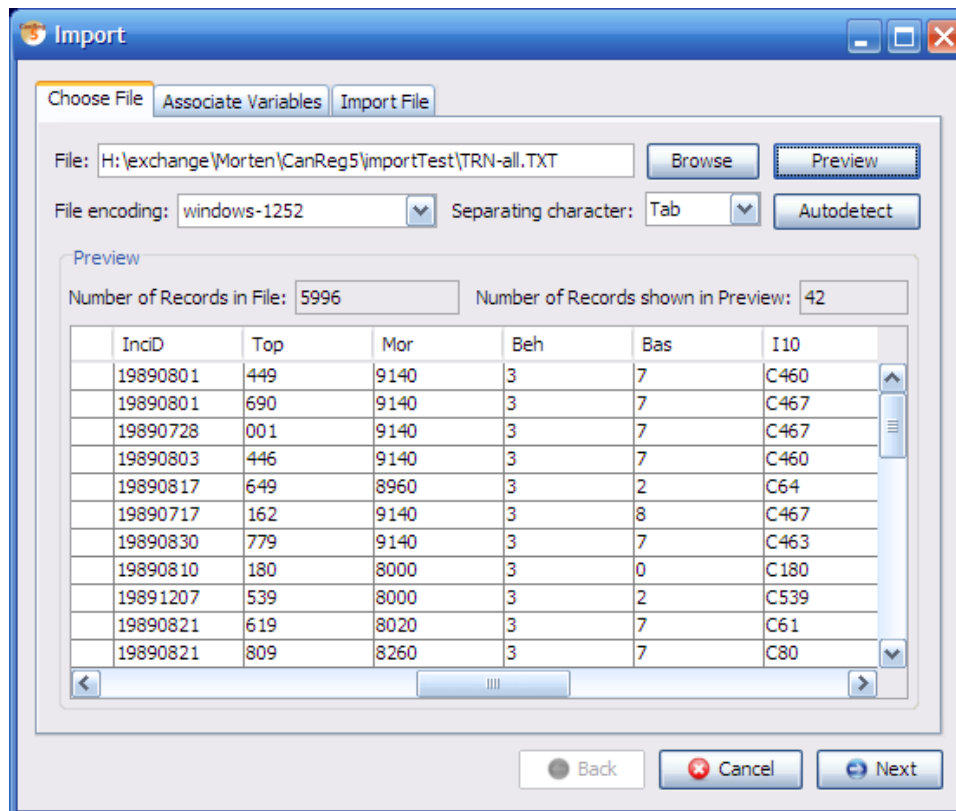
- In CanReg4: “Analysis”, “Export data”
- Tick “Export all variables”.
- Choose variables names short
- Under “Export File options” choose “Comma separated variables”
- Untick “Format date”
- Untick “Correct Unknown”

### 3. First launch

- Click “write data to file” and pick a file name that you can find back easily in CanReg5. For example on the desktop. Click “save”.
- Take a look at the data you have now exported and close CanReg4.
- Back in CanReg5 do “File”, “Data Entry” and “Import Data”.



- The program will ask you if you have all your data in one file. Answer “yes” as this is the case when migrating from CanReg4.
- Click “Browse” and locate the file from step A. Select it and click “Open”. You can if you want preview the file to see that you picked the right one and that the file looks OK. If for example Arabic names are garbled you should try to choose another “File encoding” (Default for Arabic text is ISO-8859-6).
- **Set “Separating character” to Comma.** (Or whatever separating characters your file has.)



- Click Preview to see that the data looks OK.
- Click “Next” (or select the tab “Associate Variables”)
- This lets you associate the variables in the file to import with the variables in the database. CanReg5 will find most of these associations by itself, but you should revise them to see if they look OK. Look for variable names in bold, as they are the one that are not assigned at all.
- Click “Next” (or select the tab “Import File”)
- Click “Import” (leave everything as by default – the import function only works on empty CanReg databases as per now...)
- Let CanReg5 import the data (this might take a while) and click “OK”.
- Click “Browse/Edit” and “Refresh Table” to see that the data has arrived well.

### 3.10. Import data from other programs

You can import data from other programs than CanReg4 by using the import tool in CanReg5. The only thing to pay attention to is that the data has to be coded in exactly the same way as in the CanReg5 database.

### 3. *First launch*

- Dates should be coded as year month day (yyyyMMdd) – ISO 8601
- Topography in 3 digits ICD-O-3 with no leading C.
- Morphology in 4 (or 5) digits ICD-O-3.

Other fields with dictionaries, like for example addresses should follow the dictionary defined for them in CanReg5.

The data can either be in a single file as the example for CanReg4, or in one separate file for patient-information, tumour information and source information (with pointers to link sources to tumours and tumours to patients).

## Part III.

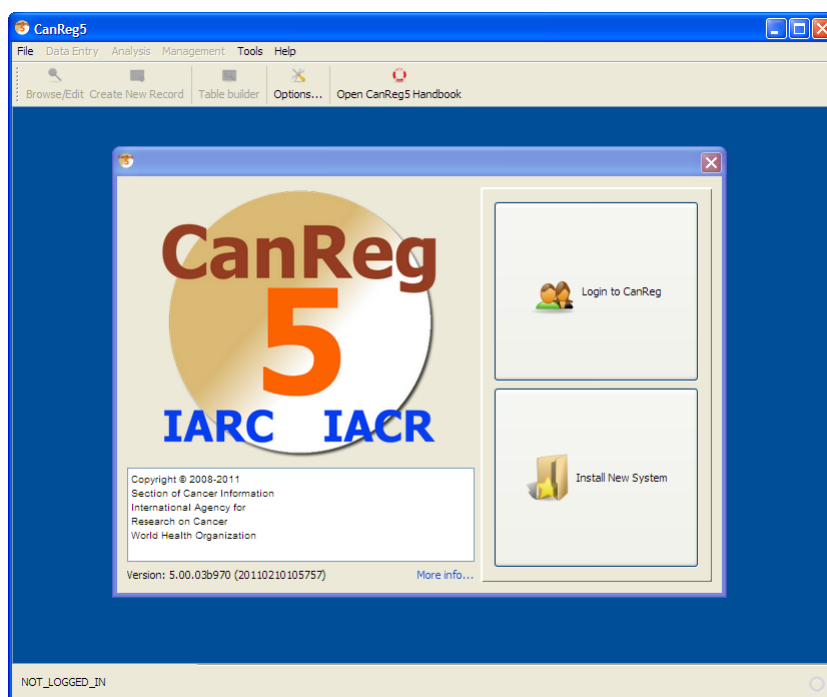
# Working with CanReg5



## 4. Start

### 4.1. Welcome screen

There are two options in this CanReg5 welcome screen: "Login to CanReg" and "Install New System".



### 4.2. Login to CanReg

If CanReg5 has already been setup on this computer for your cancer registry, then click on this option to pass to the "Login/Password" panel. (See 3.7 on page 29 for more information.)

#### 4.2.1. Advanced login settings

Under Settings you can tick Advanced to get access to some more advanced settings for CanReg. (See figure 4.1 on the next page.)

#### 4. Start

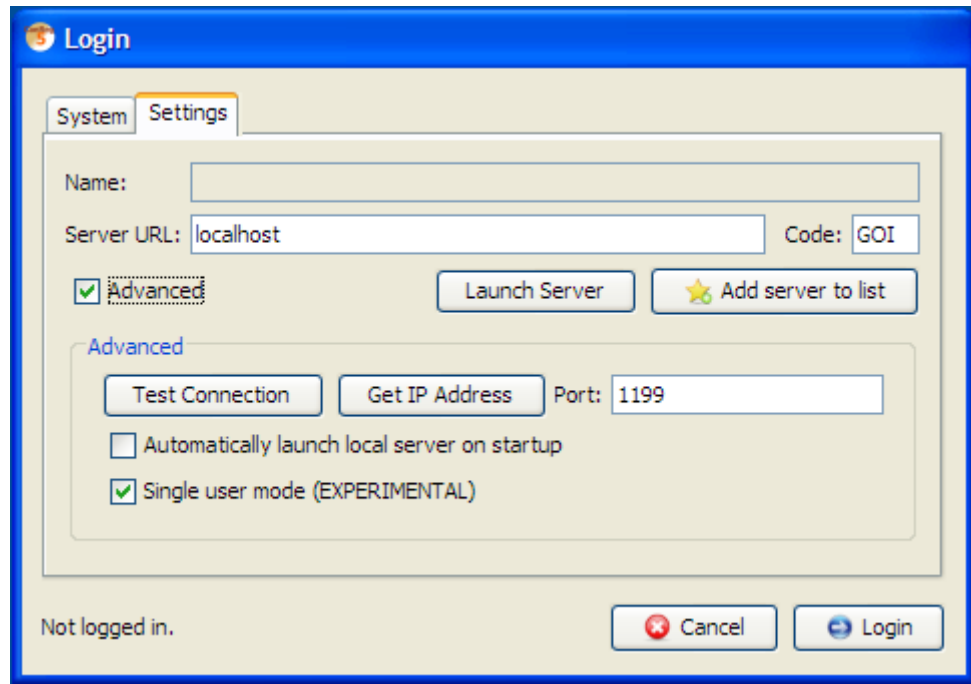


Figure 4.1.: Login settings

##### 4.2.1.1. Test connection

will test the connection to the server specified - without connecting to it.

##### 4.2.1.2. Get IP Address

will try to get the ip address of the URL specified in the Server URL field. If this is set to localhost it will give you the IP address of the machine you are on. This can be usefull when installing CanReg in a network.

##### 4.2.1.3. Port

lets you change the port the server is running on. This can be useful if another program is running on the default port and you want to change that or if you need to tunnel CanReg traffic through a VPN-network or similar.

##### 4.2.1.4. Automatically launch local server on startup

is by default ticked after installing a new XML. This does what it says - it will try to launch the last used CanReg server on this machine when starting CanReg. Basically it clicks "Launch Server" for you when opening the login screen.



#### **4.2.1.5. Single user mode**

lets you log on to your CanReg system without starting the network components. This can be useful if you are the only user on CanReg this session. Like this you avoid network related errors due to for example laptops falling asleep. Also, it is slightly safer, since you can be sure that nobody can log on to your system but you (even if they have gotten hold of your username and password and server address).

### **4.3. Install New System**

The CanReg5 program has been installed but you wish to add the Registry definition for a particular Cancer Registry. To install a new system you can use this module. Click browse and choose the XML-file that corresponds to your system. (See [3.3 on page 20](#) for more information.)



## 5. Data entry

From the data entry menu you can open the Browse/edit data view, edit dictionary view, edit population datasets view or import data view.

### 5.1. Browse/edit

This part of CanReg5 allows you to view and edit the database records.

For Data Entry purposes, you can use this Browse part to look for a particular record to Edit, or to see if a particular person has a cancer notification already stored.

The table below shows the data - move (with the "Scroll Bars") horizontally to see other variables, or vertically to view other records.

You can use the Filter, Index and Ranges to select which records to show, and the Variables radio buttons to select the variables columns.

Use these buttons to go to the Edit Form:

- Create next record: If you have checked that the patient has no record already, use this option to create a new blank edit form. The next available registration number will be assigned when you save the record.

- Edit Table record: to edit the record highlighted by the blue bar in the data table.

- Edit/create Patient ID: Before clicking this button, fill in the Registration number of the record you wish to edit. If the record exists already, you can edit it; if not, this number will be assigned to a new blank record. USE THIS OPTION TO SET THE REGISTRATION NUMBER FOR YOUR FIRST RECORD.

- Re-draw table:

If you have made changes to the database, use this button to update the table displayed.

#### 5.1.1. Table

This lets you select the table to look at.

#### 5.1.2. Sorty by

The records will be ordered (or sorted) by the variable chosen.

### 5.1.3. Range

You can specify the Range start and end values for that sequenced variable.  
Example of how to use:

- Sort by Date of Incidence,
- show records of years 2000 and 2001 only:
- Range = Incidence Date
- Range Start = 2000
- Range End = 20019999

### 5.1.4. Filter

To select records. (Use "Range" as primary selection - it is quicker)

Operator	Description
=	Equal
<>	Not equal
>	Greater than
<	Less than
>=	Greater than or equal
<=	Less than or equal
BETWEEN	Between an inclusive range
LIKE	Search for a pattern (use % as wildcard)
IN	If you know the exact value you want to return for at least one of the columns.

Logical Operator	Description
AND	match both criterias
OR	match one of the criterias

#### 5.1.4.1. Examples of how to use

- sex = '1' (all male cases.)

- age  $\geq 60$  (cases aged 60 and more.)
- sex = '2' and age  $< 60$  (female cases aged more than 60.)
- age BETWEEN 45 AND 60 (cases from patients aged from 45 to 60 (inclusive))
- age  $< 15$  OR age  $> 60$  (patients aged less than 15 and more than 60)
- name = 'Smith' (Name is "Smith")
- name LIKE 'Sm%' (Name begins with "Sm".)
- basis = '7' or Basis = '5' (Basis is 7 or 5.)
- topog LIKE '50%' (for all Breast cases.)

### 5.1.5. Filter wizard

The filter wizard is here to help you build filters. (See figure 5.1 on the next page.) It is a fast method to specify filter, or selection, criteria.

#### 5.1.5.1. For example

, to select

Females over 60 years old.... (make sure you have selected Tumour+Patient table)

Launch Filter Wizard and click on ...

1. Variable - "Sex"
2. Operator - "="
3. Value - "Female" (from Dictionary)
4. Logical Operator - "And"
5. "Add"
6. Variable - "Age"
7. Operator - ">"
8. Value - type "60"
9. "Add"
10. "OK"

For some combinations using "AND" and "OR" you may need to add brackets after.

e.g.

Topog = '220' AND (Basis='1' OR Basis='2')

## 5. Data entry

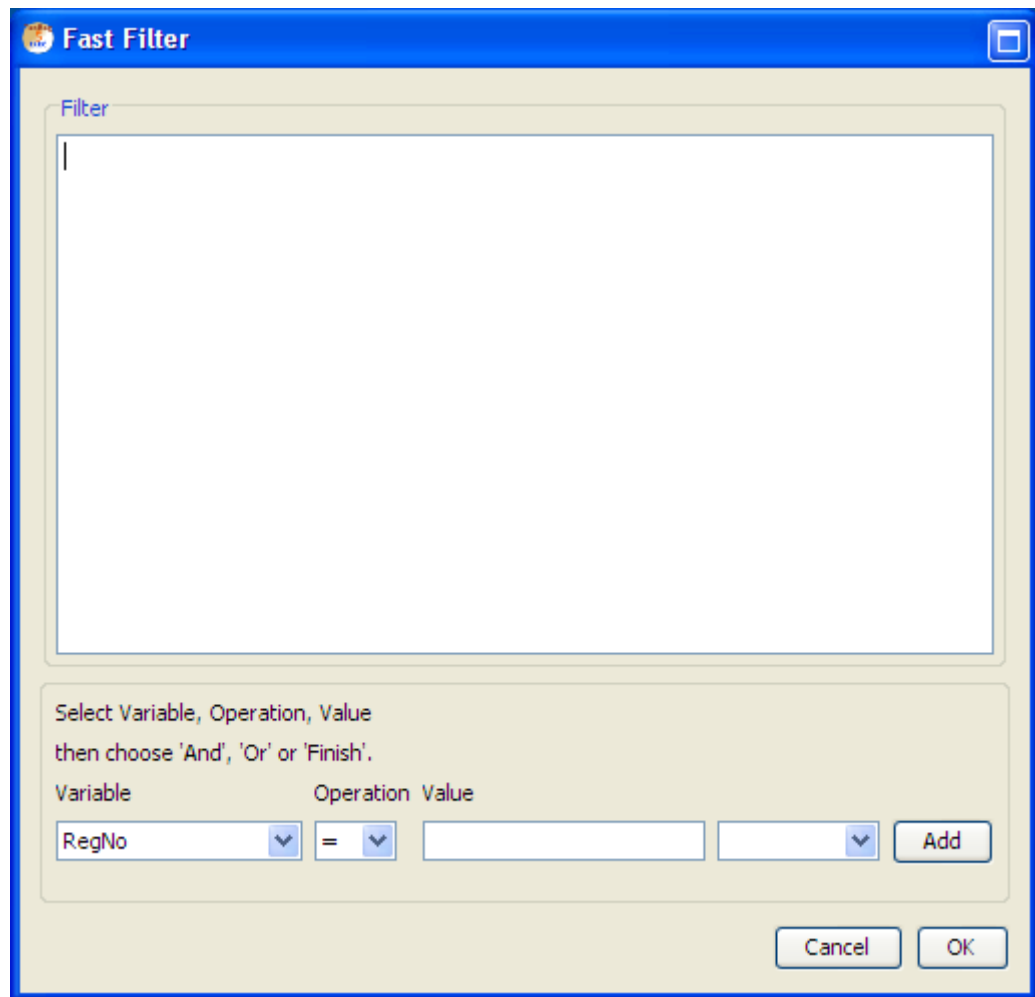
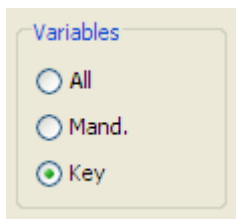


Figure 5.1.: Filter Wizard

### 5.1.6. Display variables



Click on a "radio button" above to display either:

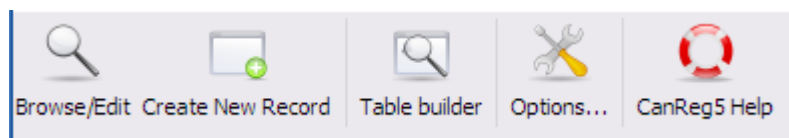
- All variables
- Mandatory variables (those that MUST be filled in the Edit form)
- Key variables (Names, Age, Date Incidence, Topog etc)

### 5.1.7. Navigate buttons

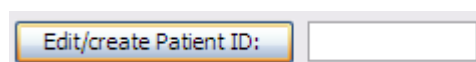
Click on the Navigation buttons below to move record: Top, Bottom, Up, Down.

## 5.2. Edit record

To get to a data entry form either press Create New Record from the menu bar



or enter a new record number in the browser and click "Edit Patient ID:"



or double click on any record in the browser.

## 5. Data entry

The screenshot displays a software interface for data entry, divided into two main sections: 'Patient' and 'Tumour'.

**Patient Section:**

- Buttons: 'Add tumour record', 'Merge with other Patient ID', 'Enable obsolete records', 'Patient/Tumour', 'Print'.
- Search: 'Person Search' with 'Run' and 'Menu' buttons.
- Fields: 'First names', 'MiddleName', 'Maiden name', 'Surname', 'Sex', 'Birth date', 'Race', 'Occupation'.

**Tumour Section:**

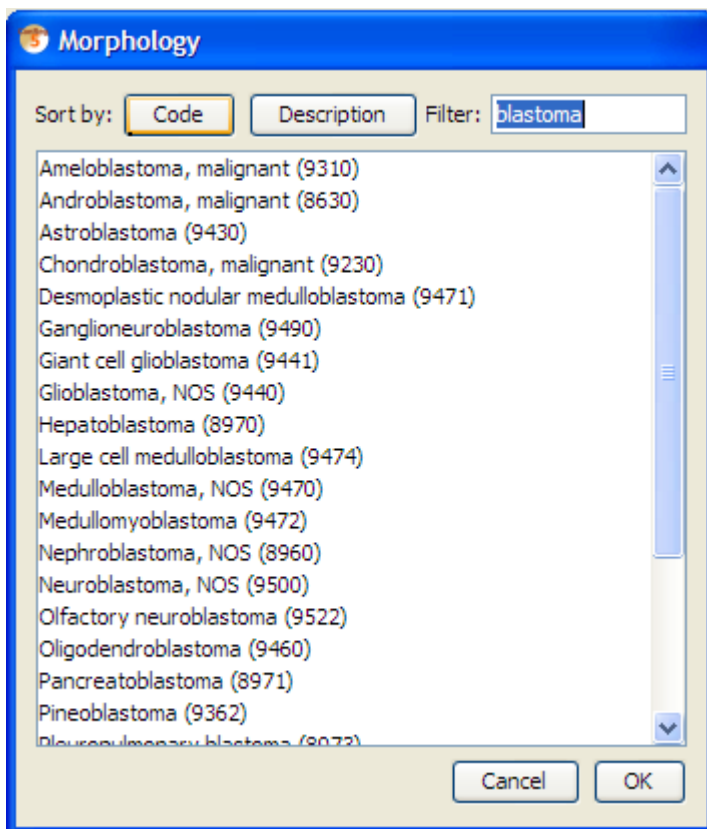
- Buttons: 'Checks', 'MP Search', 'Record Status', 'Sequence', 'Updated', 'Menu', 'Save'.
- Fields: 'Incidence date', 'Topography', 'Morphology' (with value '9310' and 'Ameloblastoma, malignant'), 'Behaviour', 'Basis diagnosis'.

You can move around the Data Entry form in the following ways:

- Tab and shift-tab to move to the next or previous variable;

With the mouse, simply click on the variable you wish to edit, or on the dictionary box to select from the popup of the valid dictionary codes.





The pink fields are the mandatory ones. (Date format is still set to yyyyMMdd, but this will be improved later.)

When you have finished entering the data, you must perform the checks as described in the system variables part below.

### 5.2.1. System variables

After data entry, you must perform the following steps:

**Person Search** Searches for any records that might belong to the same person.

**Check** Performs various consistency checks ( [5.2.3 on the following page](#)) on the data you have entered.

**Record Status** All new records are set to "Pending" and cannot be "Confirmed" until the "Check" and the "Person search" have been successfully performed. Only confirmed cases are used for analysis. Only a user with "Supervisor" permission level can confirm rare or multiple primary cases, or delete records.

**Save** Save record to the database

The "Updated" box shows the date that the record was last edited and by whom.

## 5. Data entry

### 5.2.2. Obsolete button.

This will flag the record as obsolete, so that it will not show up in analysis. It is a way to keep duplicate records that might contain valuable information.

### 5.2.3. Check

The screenshot shows the 'Edit Checks' dialog box. It has a blue title bar with the text 'Edit Checks' and standard window controls (minimize, maximize, close). The dialog is divided into several sections. The first section is 'Mandatory variables' with a list box containing 'All Present'. The second section is 'First Name and Sex' with a list box containing 'Sex/First Name: OK'. The third section is 'Mandatory variables' with a list box containing 'Cross-check conclusion: Valid'. The fourth section is 'Local checks' with an empty list box. At the bottom, there are two groups. The first group is 'Generated variables' with sub-sections for 'ICD-10' (showing 'C46.0') and 'ICCC' (showing '09c'). The second group is 'Results' with an 'OK' button. A large green checkmark and 'OK' button are at the bottom right.

Any variables found in error or query will be marked in red in the data entry form. There are three sections to the checks:

**Mandatory variables** Indicates any variables, defined as mandatory for your Registry, that have not been filled in. If the value is really not available, then fill in "9" or "99" etc. - the code for "Unknown".

**First Name and Sex** Checks the combination of First Name and Sex. e.g. "Mary", "male" would probably be an error! A name that is really used by both sexes can be defined as "Unisex".

**Cross checks** These are the same consistency checks as in the IARC Tools "Check" program. Some combinations would be marked as errors:

e.g. Sex = Female and Topography = Prostate, while others could be marked as "Rare". Only a Supervisor can confirm a Rare case.

As well as performing these checks, this function also determines the ICD-10 code derived from the ICD-O Topography and Morphology.

For more information on these checks please see: <http://www.iacr.com.fr/TechRep42-MPrules.pdf>

(The following checks are implemented in CanReg5: "Age and Morphology", "Age and Topography", "Age, IncidenceDate, and BirthDate", "Age, Topography, and Morphology", "Basis", "DateOfLastContact", "Grade", "Morphology", "Sex and Morphology", "Sex and Topography", "Topography and Behaviour", "Topography and Morphology", and "Topography".)

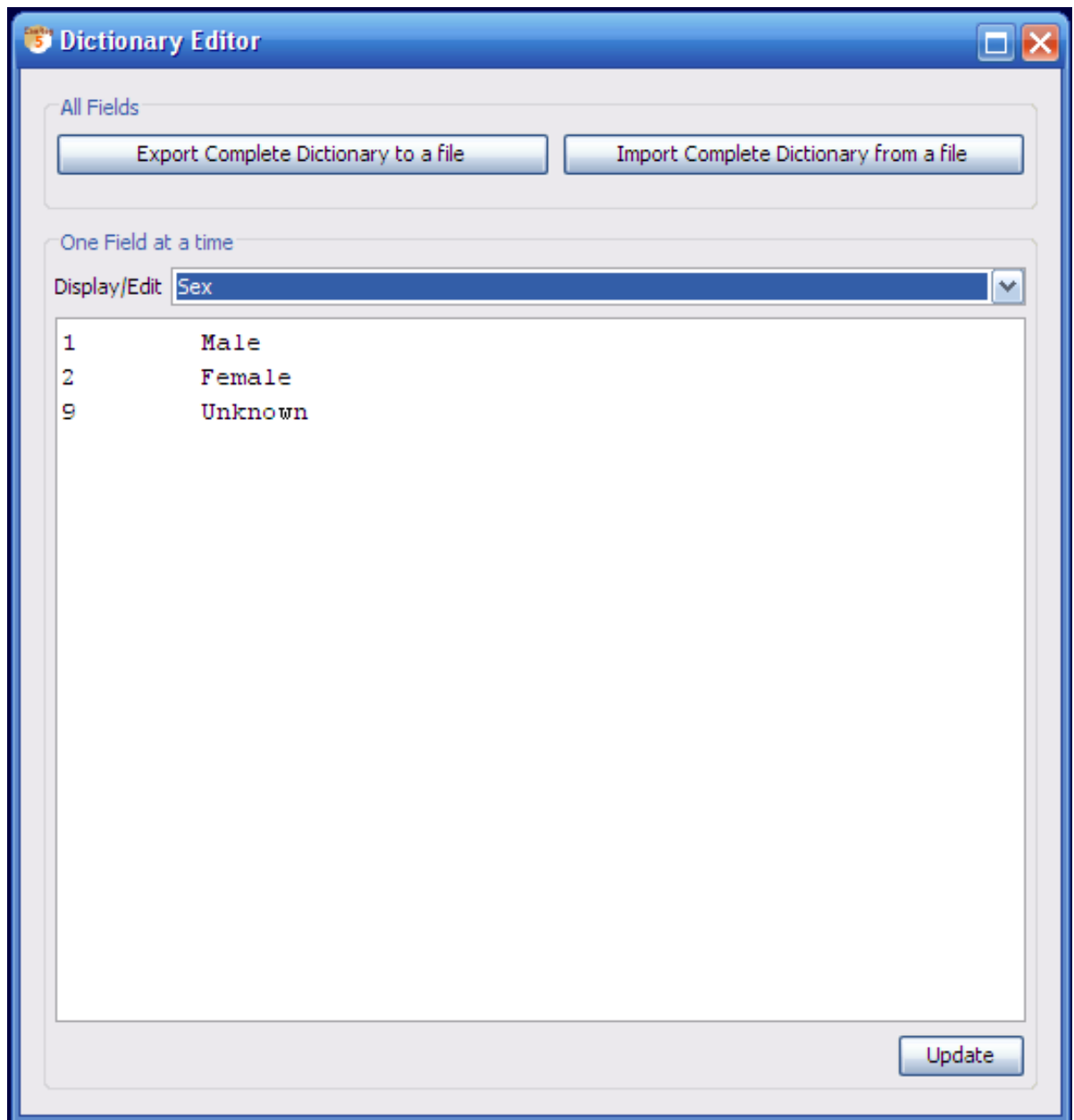
#### 5.2.4. **Person search**

The whole database is searched using probability matching for any records that might belong to the same person. All personal data such as Date of Birth, Place of Residence, Id Number, plus a phonetically simplified form of the name are used in this search. (This can be tailored by the administrator of the CanReg5 system.) Any other record with a percentage match higher than the "Minimum Match" is displayed.

If no match is found, a message will be displayed to that effect.

Otherwise, the computer only displays possibilities - YOU must decide if it really is the same person.

## 5.3. Dictionary editor



The dictionary editor lets the supervisor users edit the coding schemes of the various variables in CanReg5.

### 5.3.1. Export Dictionary to file

Export the current set of CanReg5 dictionaries to a tab-separated file for editing in for example Excel.

### 5.3.2. Display/edit . . . Select

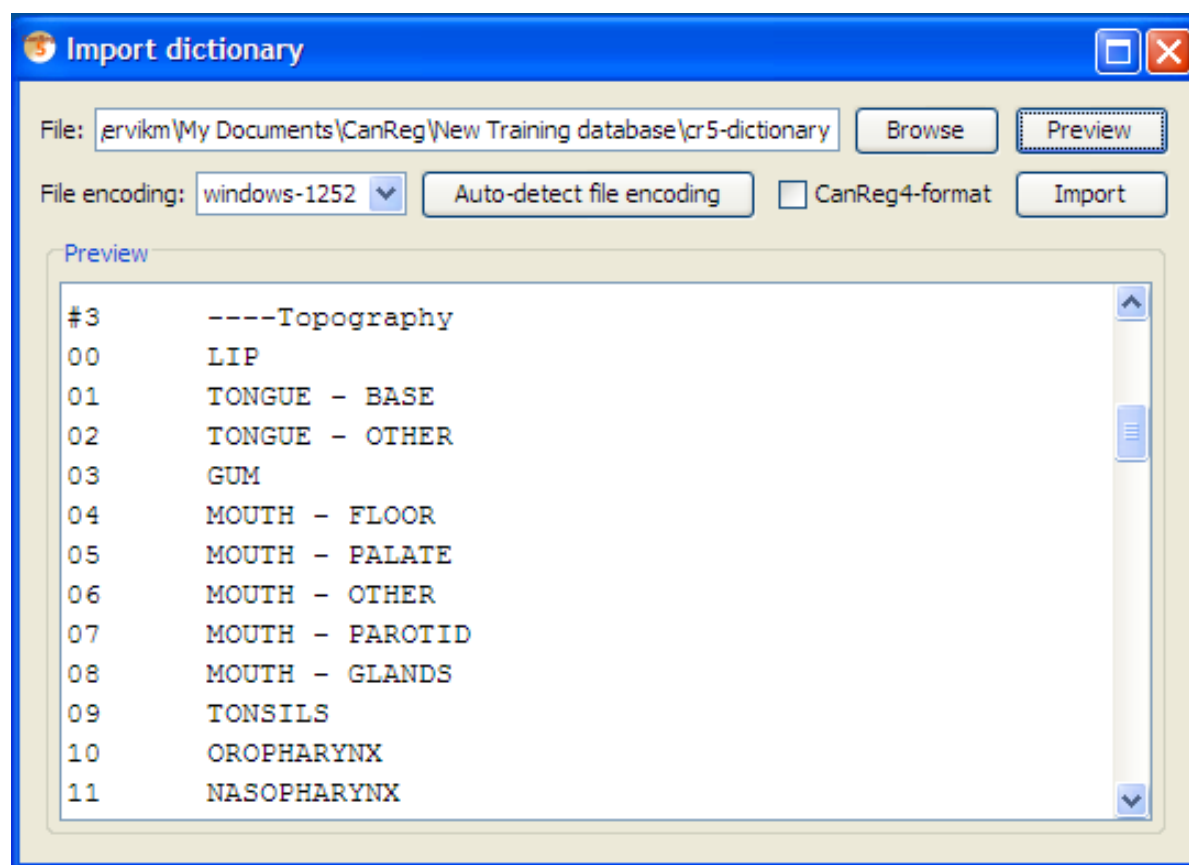
This will display any dictionary picked by the user for editing directly in CanReg. The format will be standard tab-separated values so that the user can also copy and paste this into general spreadsheet applications.

### 5.3.3. Update

This will import the dictionary picked by the user from the text area. The format must be tab-separated values. This means that the users can copy and paste directly from general spreadsheet applications (i.e. Excel).

### 5.3.4. Import Dictionary from file

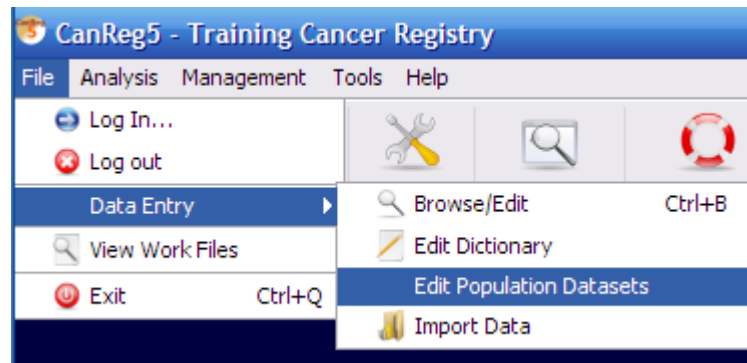
Import a complete set of CanReg5 dictionaries from a tab-separated file. (Or a two-space-separated CanReg4 dictionary.)



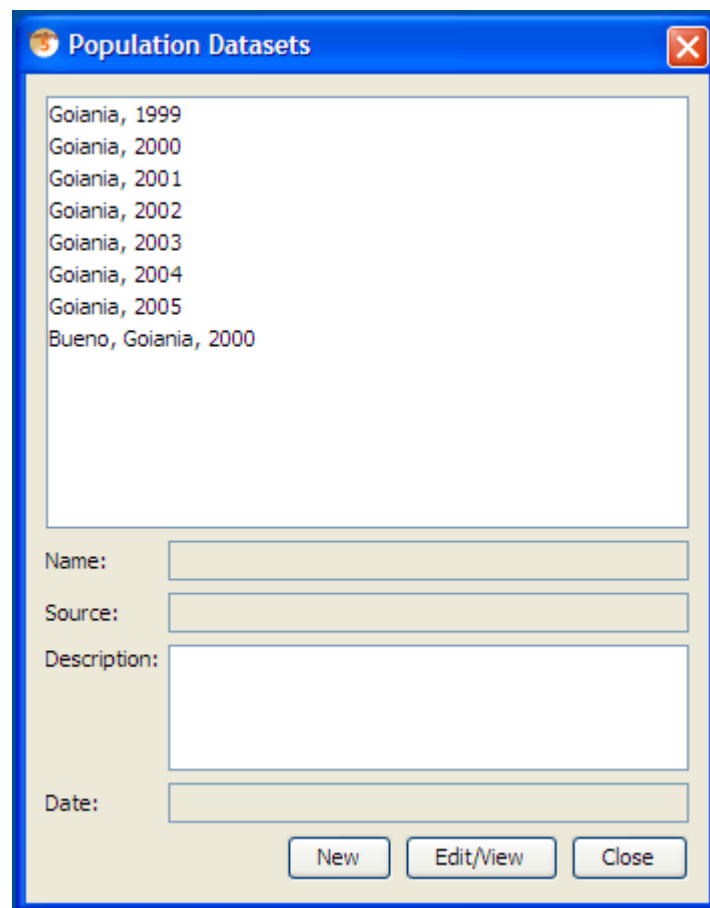
## 5.4. Population Dataset Editor

### 5.4.1. Manual entry/validation

The Population Dataset editor lets you edit population data set to be used in the table builder. This is located under File – Data Entry Edit Population Dataset:



When you start it you get to the list of all your population datasets.



Add one by clicking New. This opens the Population Dataset editor:

Population Data Set Editor

Details Population Data Set Pyramid

Name: Bueno, Goiania, 2000

Filter: Addr = '74210' Filter Wizard

Source: Government Census

Description:

Age group structure: 0-4, 5-9, 10-14, 15-19, ..., 80+ Other

Date: 20000701

Standard Population: World Standard Population Edit

Delete Locked Cancel Save

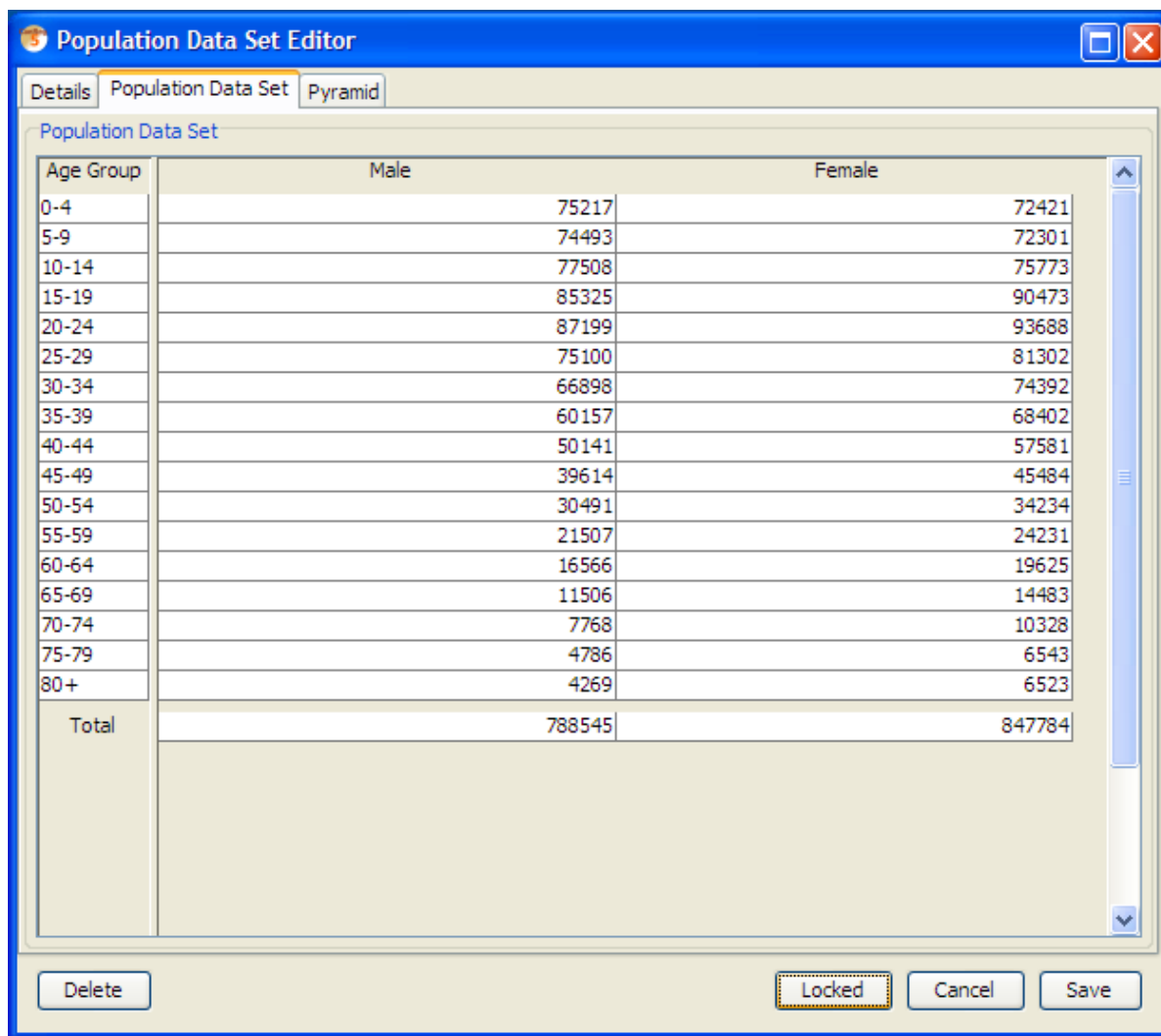
Fill in the details:

- A name for the dataset
- "Filter" or selection criteria, so the program only selects records corresponding to the population (e.g. Address code  $\geq 10$  and Address code  $\leq 19$ ) (Basically if it does not cover your entire area of your database.)
- A source of this data (e.g. whether Government Census, or Estimation).
- Some description (less than 255 characters)
- Choose the age group structure.
- Set the date when the population was at this amount. (In the example above it is mid 1992)
- The Standard population used for ASRs when building tables with this set.

## 5. Data entry

"Age Standardised" rates are calculated in order to compare rates from different countries that have different age profiles. Normally the "Standard" population is the World standard included here. (If you wish to change this, choose another standard population (or click on the standard pop."Edit" button.)

Then fill the population dataset itself:



The screenshot shows a software window titled "Population Data Set Editor". It has three tabs: "Details", "Population Data Set" (which is selected), and "Pyramid". The "Population Data Set" tab displays a table with three columns: "Age Group", "Male", and "Female". The table lists age groups from "0-4" to "80+", with a "Total" row at the bottom. The population counts for males and females are provided for each age group. At the bottom of the window, there are four buttons: "Delete", "Locked" (which is highlighted with a yellow border), "Cancel", and "Save".

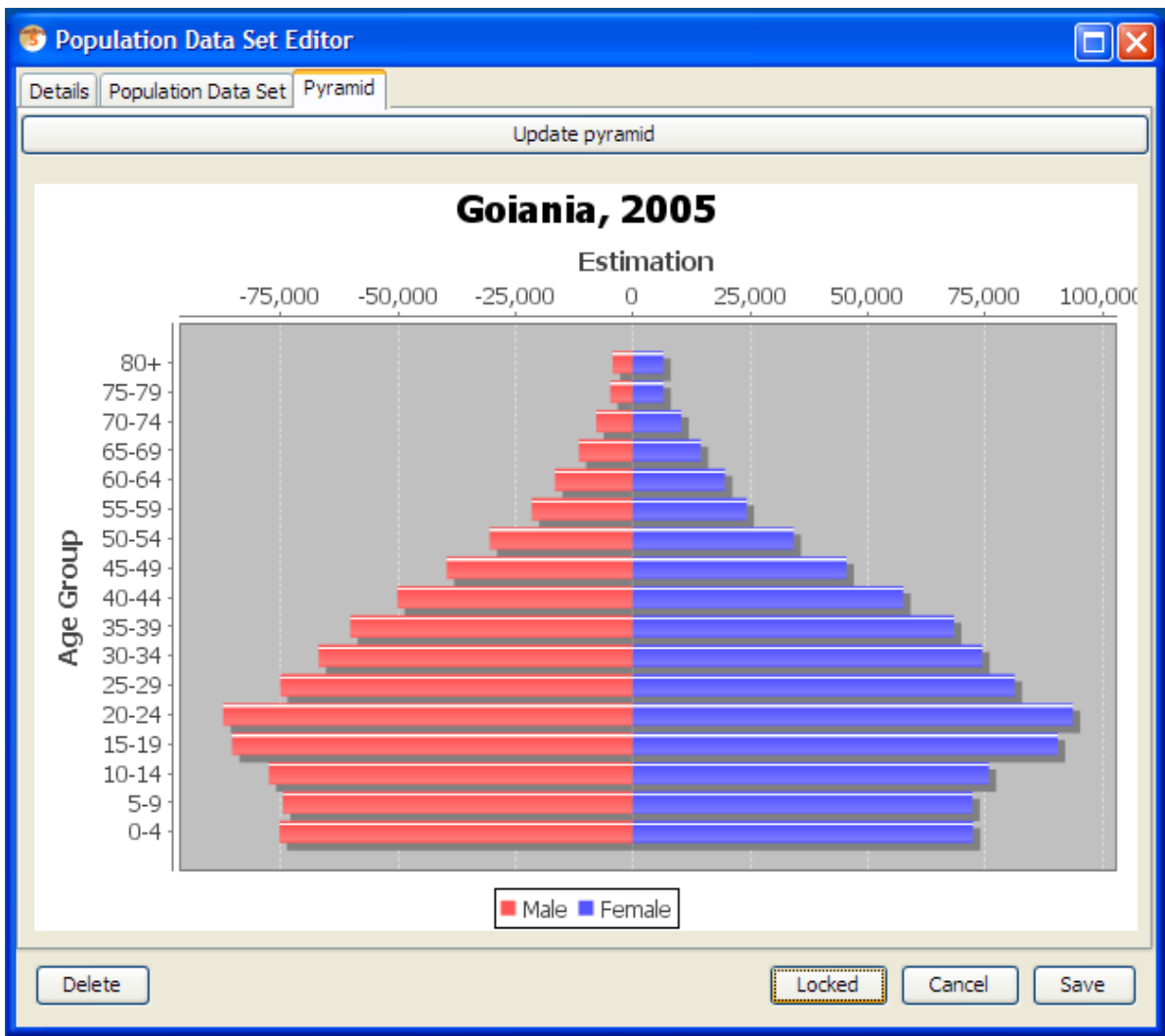
Age Group	Male	Female
0-4	75217	72421
5-9	74493	72301
10-14	77508	75773
15-19	85325	90473
20-24	87199	93688
25-29	75100	81302
30-34	66898	74392
35-39	60157	68402
40-44	50141	57581
45-49	39614	45484
50-54	30491	34234
55-59	21507	24231
60-64	16566	19625
65-69	11506	14483
70-74	7768	10328
75-79	4786	6543
80+	4269	6523
Total	788545	847784

(Please note that you can copy and paste population datasets back and forth from general spreadsheets like Excel.)

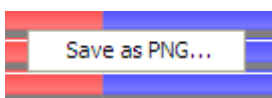
Click save to save your population dataset to the database.

You can also take a look at the population pyramid of the current population data set by going to the Pyramid tab.





This can be saved as an image to disk by right-clicking on the image and choosing “Save as PNG...”

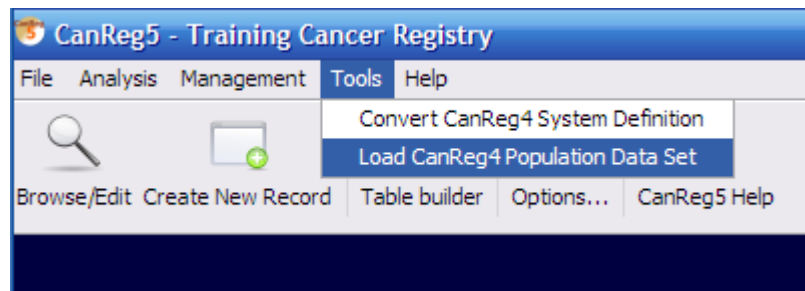


And choosing a proper file name.

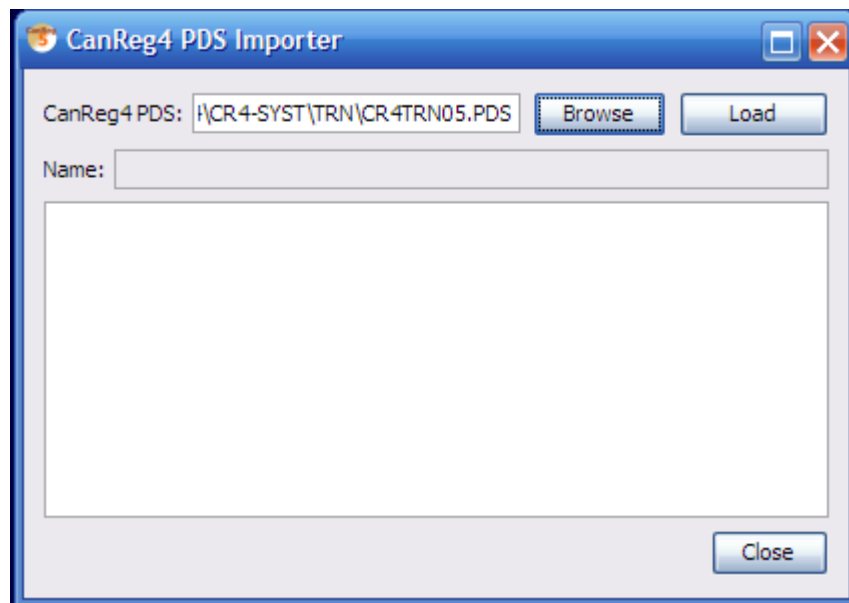
### 5.4.2. Import population data set from CanReg4

Alternatively you can import population data sets from CanReg4. To do this you go to the Tools menu and choose “Load CanReg4 Population Dataset”.

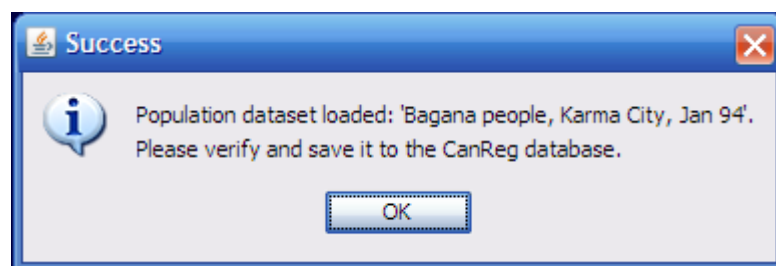
## 5. Data entry



Click Browse to find the population dataset:



Then click Load and a confirmation message that the population dataset has been loaded will appear.



Click "OK".

Next step is to revise the population dataset and see to that it has been imported correctly.

One important thing to do is to see to that the filter is correct. That for example the search variables are enclosed by 's. If you need to change anything in the dataset you need to unlock it by toggling the “Locked”-toggle.

## 5.5. Import

The import function lets you import data from other CanReg systems or other programs.

(For a detailed walk through of how to get the data from CanReg4 to 5 see [3.9 on page 33](#).)

Data in an external file may be added to the CanReg database by importing. It should be of the following format:

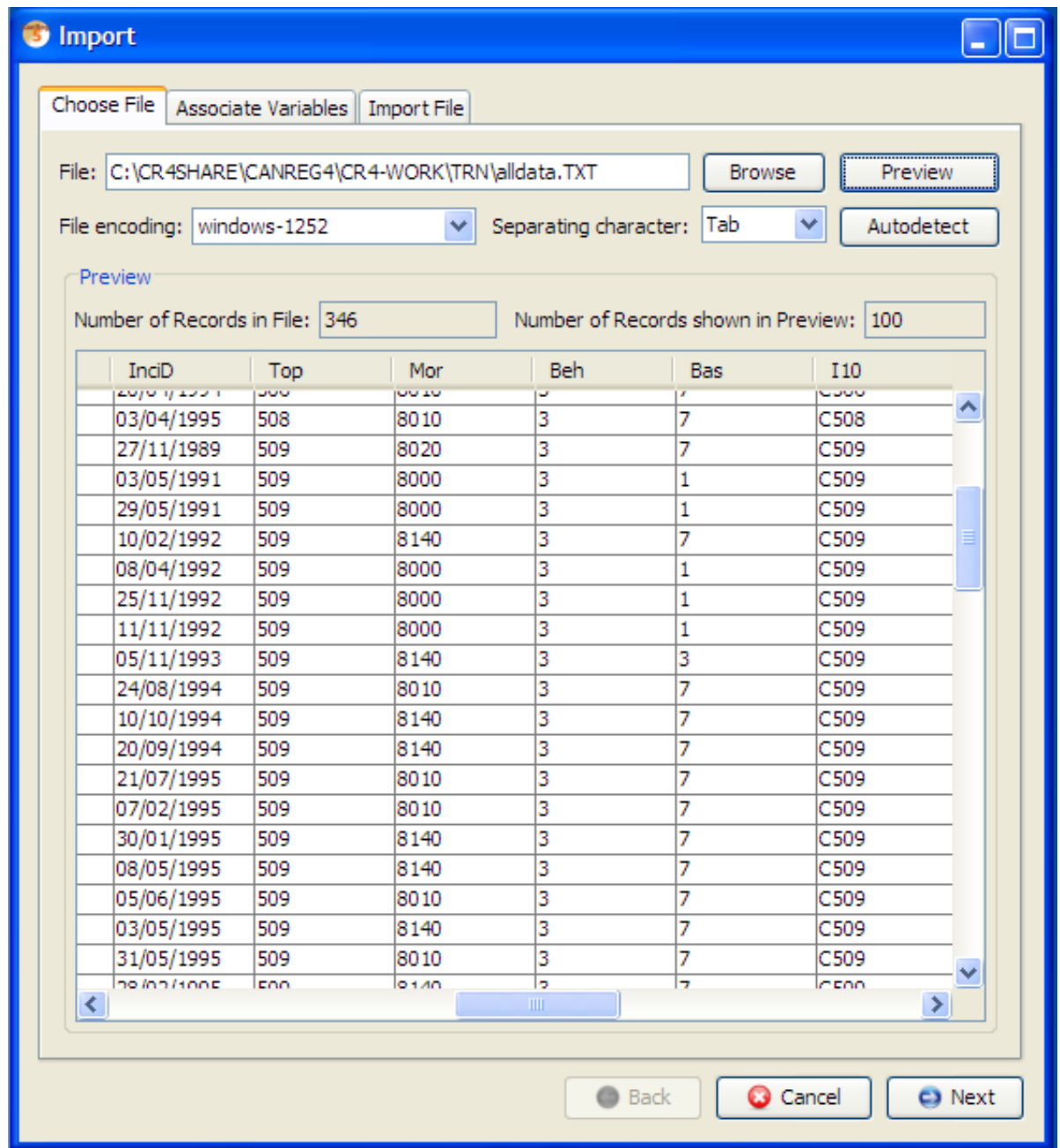
- Tab delimited
- Comma Separated Variables
- Character delimited

## 5. Data entry

The easiest way is to have the variable names at the top of each column.

There are four main steps to importing a data file. (If you don't have all your data in one file, but rather one file per table you need to repeat the first two steps for each file.)

- 1) Choose the file(s).



Here you also might need to specify the character encoding of your file as well as the separating character used, if CanReg does not detect it automatically. Please verify using preview.

- 2) Identify the variables.

The 'Import' dialog box has three tabs: 'Choose File', 'Associate Variables' (selected), and 'Import File'. The 'Associate Variables' tab displays a table with two columns: 'Variable in file to be imported' and 'Variable in CanReg database'. The table contains the following rows:

Variable in file to be imported	Variable in CanReg database
RegNo	RegNo
<b>RecS</b>	<b>RecS</b>
Chec	Chec
PerS	PerS
FamN	FamN
FirstN	FirstN
MaidN	MaidN
Sex	Sex
Age	Age
BirthD	BirthD
Trib	Trib
Addr	Addr
Occu	Occu
InciD	InciD
Top	Top
Mor	Mor
Beh	Beh
Bas	Bas
I10	I10
MPcode	MPcode

At the bottom of the dialog are three buttons: 'Back', 'Cancel', and 'Next'.

3) Choose the various options - see specific helps.

## 5. Data entry

The screenshot shows the 'Import' dialog box with the following settings:

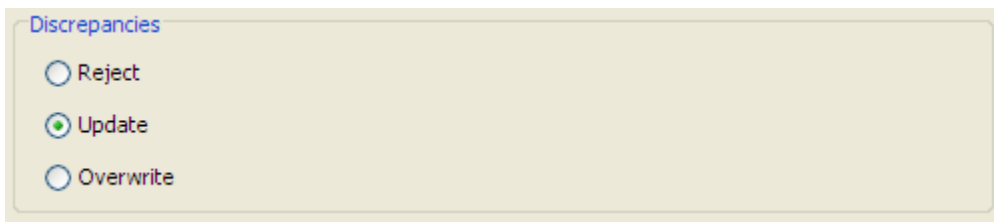
- Discrepancies:** ☐ Reject, ☒ Update, ☐ Overwrite
- Max Lines:** [Empty text field], ☒ Test Only
- CanReg data:** ☐ Do Checks, ☒ Person Search, ☒ Query New name

Buttons: Import (green checkmark), Back, Cancel, Next.

4) Import the file (maybe in "test only" mode first)

### 5.5.1. Discrepancies

A discrepancy is when a record is found with the same registration number as one in the database, but there are differences in some of the data.



Discrepancies

☐ Reject

☒ Update

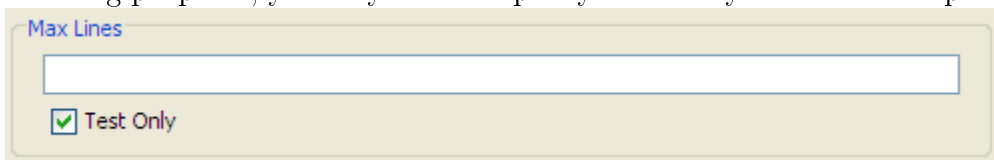
☐ Overwrite

Click on a "radio button" to either:

- Reject these discrepancies (they will not be imported)
- Update them (any new data will be copied over to the database record)
- Overwrite (ALL variables will be copied over, even empty ones)

### 5.5.2. Max. Lines / Test only

For testing purposes, you may wish to specify how many lines of the import file to read.

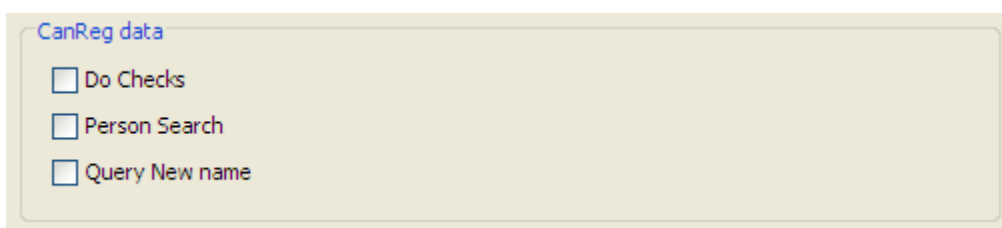


Max Lines

☒ Test Only

With "Test only" ticked, NO data is actually added to your database; only a report is generated showing what WOULD happen. It lists discrepancies, possible matches, rare or error cases etc.

### 5.5.3. CanReg DATA



CanReg data

☐ Do Checks

☐ Person Search

☐ Query New name

#### 5.5.3.1. Perform Checks

If the data to import was not created using CanReg, or if it is a Pending case, then the Checks must and will be performed.

If however, the case has already passed the Checks, the Checks will NOT be performed again unless specified by ticking this option.

#### 5.5.3.2. Perform Person Search

Normally, when importing CanReg data, the Person Search will still be performed even if already done. If this option is NOT ticked, then the Person Search will NOT be repeated in this case. This is only advisable in the case of having no original data.

#### **5.5.3.3. Query New First Name**

For data that has not already passed the checks, the First Name and Sex combination will be checked and updated. Tick this option if you wish all NEW names to be set as pending.

If you are starting a new registry then you probably don't want all new names to be set as pending, however, if you have several years of data already, then it would be advisable to query names not already known.



## 6. Analysis

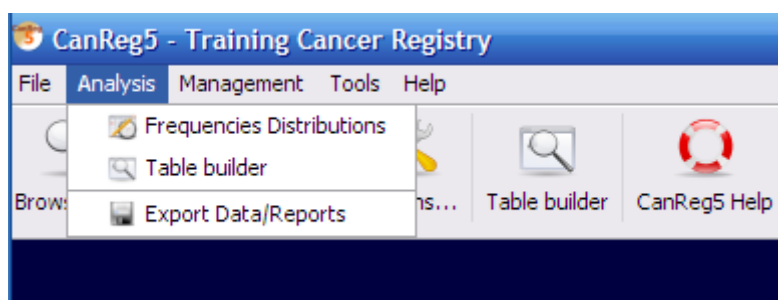
### 6.1. Export data

To export (write out) all, or part of, your CanReg5 data to an external text file.

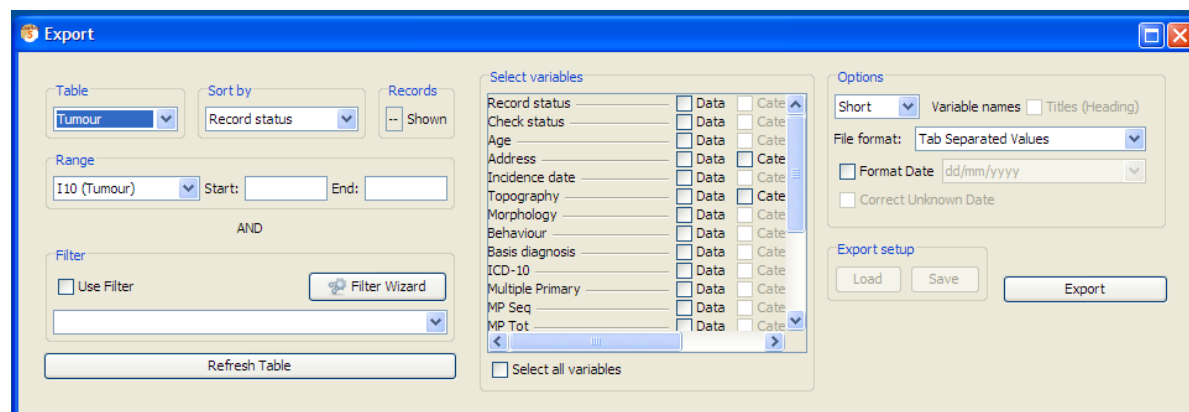
There are two main reasons for doing this:

- To be able to *import* the file into another program (e.g. *Stata*, *SPSS*, a general spreadsheet like *Excel* or *Calc* or *Access*) for further analysis.
- To produce a report, or case listing, that could be read into *Word* and printed out.

To export your data go to Analysis -> Export Data/Reports:



You will be presented with a screen that resembles the browse-screen:



The following steps need to be taken:

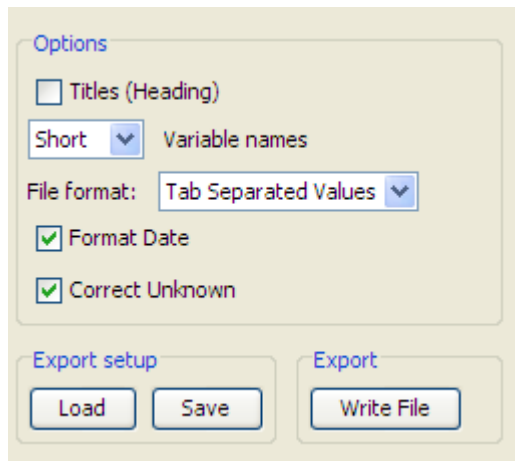
- Specify the records you want to be selected by using the Filter( 5.1.4 on page 44) and Range( 5.1.3 on page 44) options, and the order in which they will be written using the Sort by( 5.1.2 on page 43).

## 6. Analysis

- Select the Variables( [6.1.4 on the next page](#)) to display.
- Choose the Variable headers to include.
- Choose the File Format ( [6.1.2](#)) suitable for your needs.

Please note that some of the functionality, like the ability to store export-setups is not yet implemented.

### 6.1.1. Options

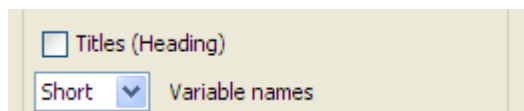


### 6.1.2. Export file format

All file formats produce text files, with one case per line (a new line at the end of each case). They all have default extension .TXT except for "Comma Separated Variable", which has .CSV. "Tab Delimited" writes each variable separated by the TAB character. "Comma Separated Values" encloses each variable in quotes, and separates by a comma.

If you export data to a “Tab Separated” or a “Comma Separated” file you can open this in general spreadsheets (like *Excel*).

### 6.1.3. Titles, variable names



#### 6.1.3.1. Titles

Will write at the top of your export file:

- the filter criteria, index and ranges used,

- today's date.

This option is useful is you are writing a report, or case listing. (Not yet implemented.)

### 6.1.3.2. Short variable name

Puts the abbreviated names of the variables at the top of each column. (If this file were imported into, for example *Access* or *R* then these would automatically become the names of the variables.)

### 6.1.3.3. Long Variable name

Writes the full name of the variable at the top of each column.

### 6.1.4. Variables

Select variables

Record status	<input type="checkbox"/>	Data	<input type="checkbox"/>	Cate
Check status	<input type="checkbox"/>	Data	<input type="checkbox"/>	Cate
Age	<input type="checkbox"/>	Data	<input type="checkbox"/>	Cate
Address	<input type="checkbox"/>	Data	<input type="checkbox"/>	Cate
Incidence date	<input type="checkbox"/>	Data	<input type="checkbox"/>	Cate
Topography	<input type="checkbox"/>	Data	<input type="checkbox"/>	Cate
Morphology	<input type="checkbox"/>	Data	<input type="checkbox"/>	Cate
Behaviour	<input type="checkbox"/>	Data	<input type="checkbox"/>	Cate
Basis diagnosis	<input type="checkbox"/>	Data	<input type="checkbox"/>	Cate
ICD-10	<input type="checkbox"/>	Data	<input type="checkbox"/>	Cate
Multiple Primary	<input type="checkbox"/>	Data	<input type="checkbox"/>	Cate
MP Seq	<input type="checkbox"/>	Data	<input type="checkbox"/>	Cate
MP Tot	<input type="checkbox"/>	Data	<input type="checkbox"/>	Cate

☐ Select all variables

Select the variables to export.

Tick the boxes next to the variable name to select (or deselect). They will appear in the data grid after you have clicked refresh table. (Note that even though you select to export the category names or descriptions they only show up as codes in the preview window.)

You can drag the grid columns to change the order of the variables.

Click on "All variables" to select them all.

### 6.1.5. Date Format

☒ Format Date

☒ Correct Unknown

Tick "Format Date", for example, to export "21/04/2001" instead of numeric form "20010421". (The date format can be any format specified by a text string using dd

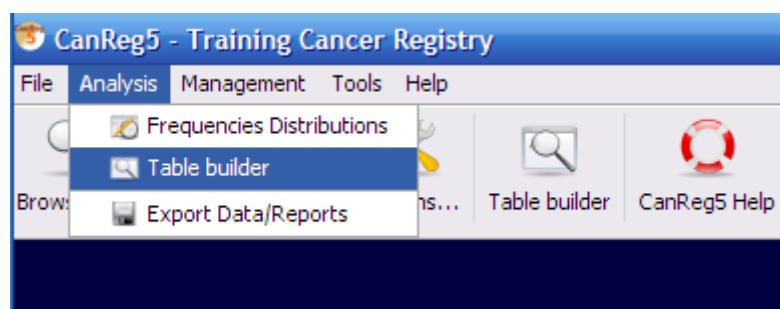
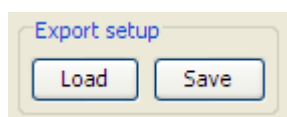


Figure 6.1.: Table builder in the menu.

for days, MM for months, and yyyy for year, for example: “yyyyMMdd” (the default), “dd/MM/yyyy”, or “MM.dd.yyyy”).

Tick "Correct Unknown" so that unknown day will be written as "01" and unknown month as "07". This is necessary if you wish to import the data into *Excel*, or any other software that will reject invalid dates. (Please note that you lose information on what dates are unknown and what dates are really mid-year.)

### 6.1.6. Export setup



This module is used to load previously made export settings or save the current settings. This includes the filter, the sequence, the variables to export and the various options. (Not yet implemented.)

## 6.2. Table builder

The Table builder (See figure 6.2 on the next page) lets you build incidence tables etc in CanReg. You find it under analysis → table builder.

### 6.2.1. Example

When you start it you first choose the type of table you want to produce. (Please note that not all the tables in the list has been implemented.)

Pick Incidence per 100,000 by age group (Period):

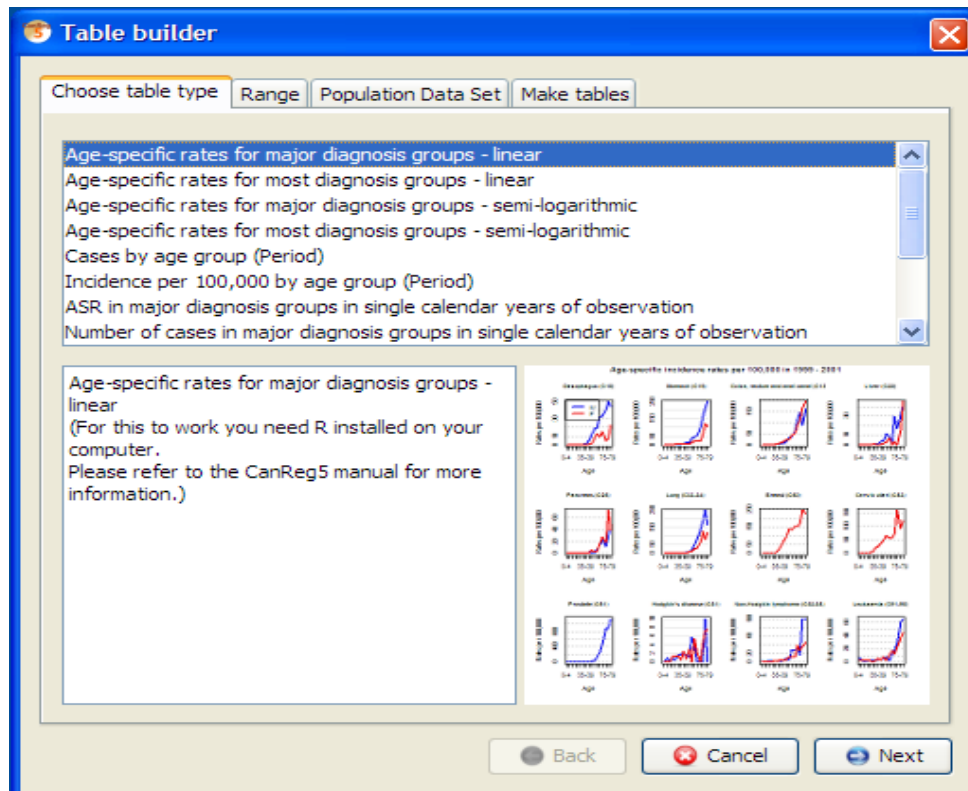


Figure 6.2.: Table builder

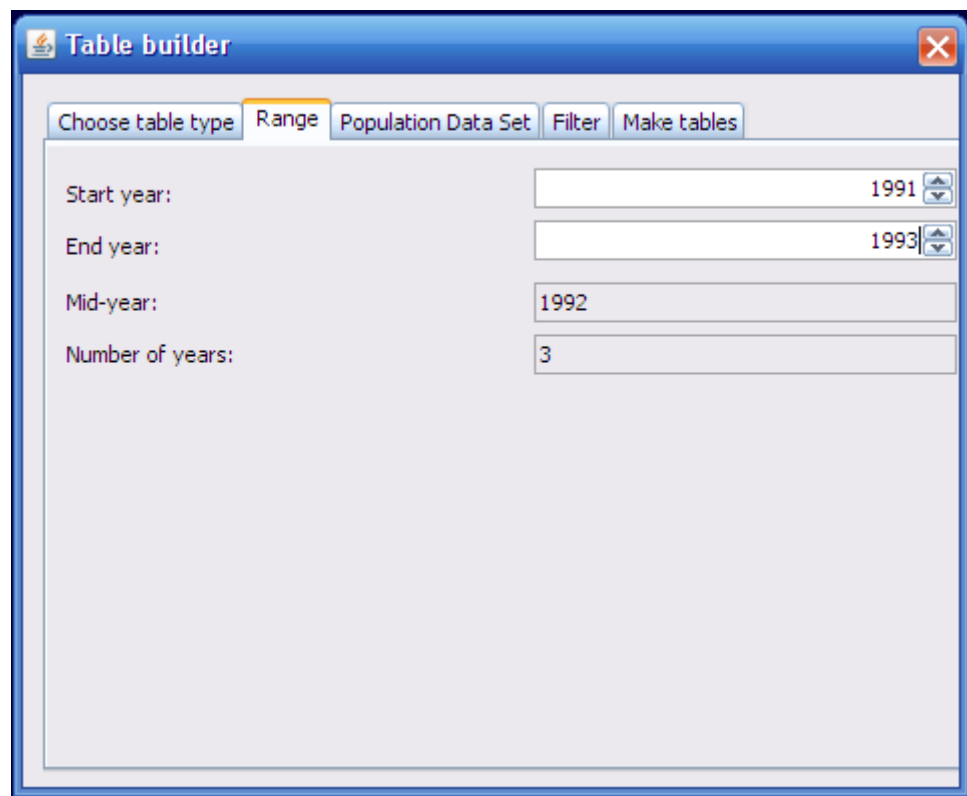
## 6. Analysis

The incidence rate is defined as:

$$\frac{\text{incidence cases per year}}{\text{population at risk}} \times 10000$$

This gives an idea of the risk of getting each type of cancer - the tables consist of Incidence Rates by Sex, Age group and ICD10 cancer type.

Click on range to set the range of your analysis and set it to match the analysis you want to do, for example here we want to look at Karma City, 1991 to 1993:



The screenshot shows a software window titled "Table builder" with a blue border and a close button in the top right corner. Inside the window, there are five tabs: "Choose table type", "Range" (which is highlighted with an orange border), "Population Data Set", "Filter", and "Make tables". Below the tabs, there are four input fields with labels to their left: "Start year:" with a value of 1991, "End year:" with a value of 1993, "Mid-year:" with a value of 1992, and "Number of years:" with a value of 3. Each input field has a small up/down arrow icon on its right side.

Figure 6.3.: The range chooser.

In order to create tables of incidence rates, we need to know the size of the population at risk. Therefore, a "Population Data Set" is needed.

Click "Population Data Set". Pick one population data set per year. (Please note that this can be the same for all three if that year is representative of the period. See [figure 6.4 on the facing page](#)).

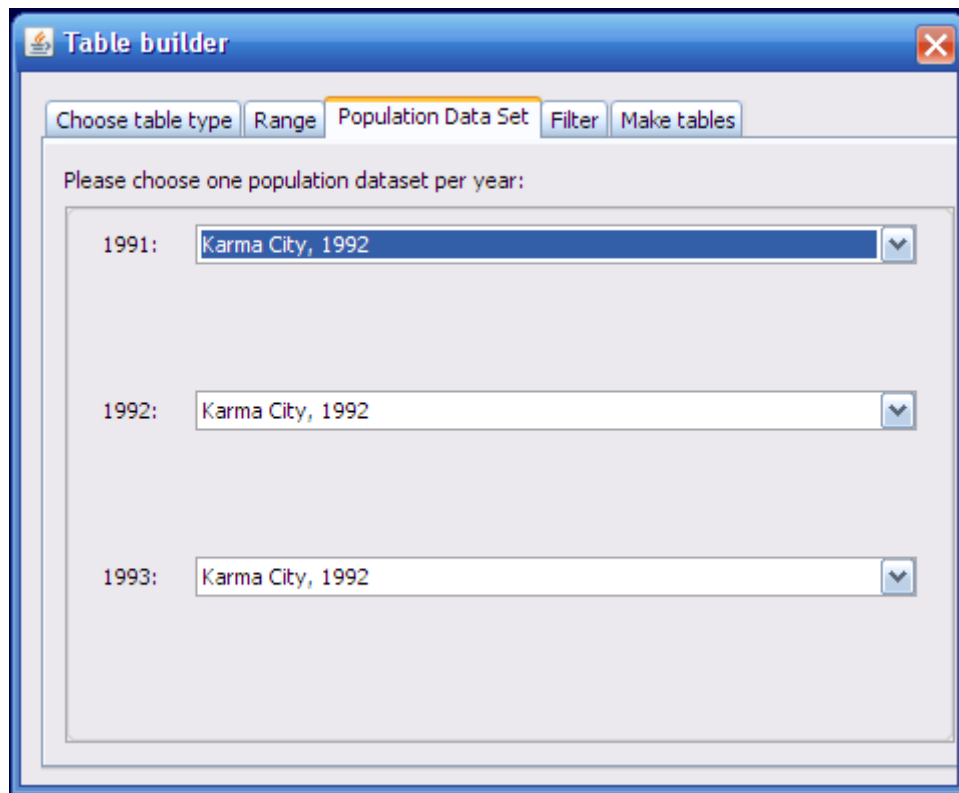


Figure 6.4.: The population dataset chooser.

Then you can go to the Make tables tab to generate the actual tables. Click “Generate post script (PS) files” (See figure 6.5 on the next page.) and choose a file name. (If the table generates more than one file (like it is the case for incidence per 100,000 some number or text will be added to the name you give for each file.)

You get a message saying, “Tables built.”

Click OK and if you have a program that can read PostScript (See page 96.) files the tables will be displayed after you press OK. (You might have to go to the Orientation menu in *GSView* and choose Portrait to get the table in the proper orientation.)

### 6.2.2. File formats

The various tables can be generated in various file formats, depending on what they support.

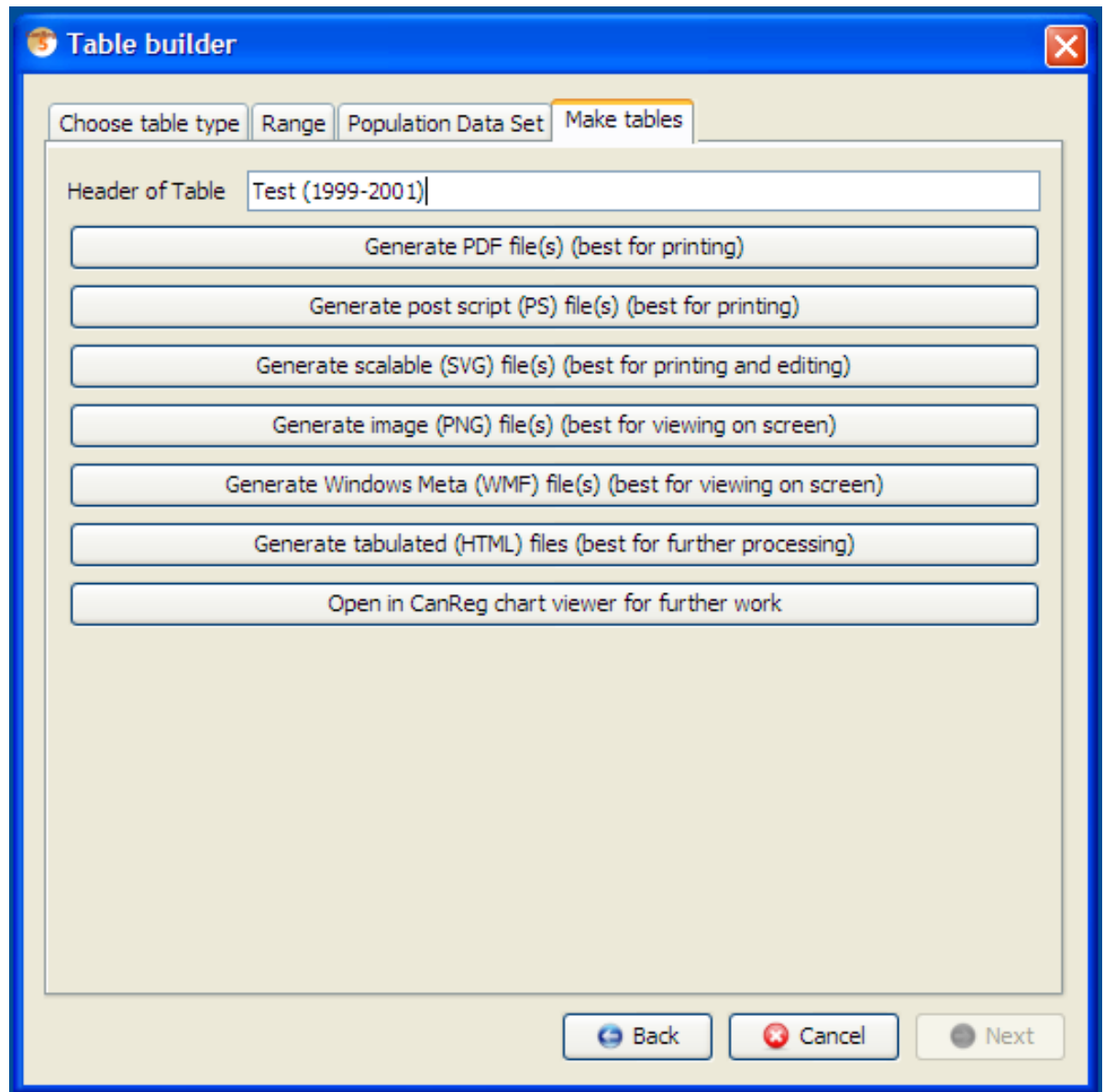


Figure 6.5.: File formats selector



## Training Cancer Registry (1991-1993)

Incidence per 100,000 by age group (Period) - Male

SITE	ALL AGE AGES UNK	0.	5.	10.	15.	20.	25.	30.	35.	40.	45.	50.	55.	60.	65.	70.	75+	CRUDE RATE	(%)	CUM 0-64	CUM 0-74	ASR (W)	ICD
Lip	1 0	-	-	-	-	-	0.5	-	-	-	-	-	-	-	14.3	-	-	0.1	0.1	0.00	0.07	0.4	C00
Tongue	3 0	-	-	-	-	-	0.5	-	-	1.9	-	-	-	-	-	-	-	0.2	0.3	0.01	0.01	0.2	C01-02
Mouth	4 0	-	-	-	-	-	0.5	-	1.2	-	2.8	-	-	-	14.3	-	-	0.3	0.4	0.02	0.09	0.7	C03-06
Salivary glands	3 0	-	-	-	-	-	0.5	-	-	1.9	2.8	-	-	-	-	-	-	0.2	0.3	0.03	0.03	0.3	C07-08
Tonsil	5 0	-	-	-	-	-	0.5	0.8	-	1.9	7.1	3.6	7.1	-	-	-	-	0.3	0.5	0.07	0.07	0.7	C09
Other oropharynx	0 0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	0.0	0.00	0.00	0.0	C10
Nasopharynx	6 0	-	0.5	-	-	-	0.5	0.8	-	1.9	2.8	3.6	-	-	-	-	-	0.4	0.6	0.05	0.05	0.6	C11
Hypopharynx	0 0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	0.0	0.00	0.00	0.0	C12-13
Pharynx unspecified	3 0	-	-	-	-	-	-	-	-	1.9	5.7	-	-	-	-	-	-	0.2	0.3	0.04	0.04	0.5	C14
Esophagus	59 0	-	-	-	0.6	0.5	0.5	0.8	-	1.9	17.0	50.7	35.5	103.0	85.9	36.0	128.9	3.8	6.2	1.06	1.67	15.3	C15
Stomach	22 0	-	-	-	-	1.0	0.5	0.8	-	3.9	5.7	18.1	-	25.7	43.0	18.0	25.8	1.4	2.3	0.27	0.58	4.8	C16
Small intestine	1 0	-	-	-	-	-	-	-	-	-	2.8	-	-	-	-	-	-	0.1	0.1	0.01	0.01	0.2	C17
Colon	11 0	-	-	-	-	-	-	1.6	1.2	1.9	5.7	-	-	8.6	14.3	18.0	25.8	0.7	1.2	0.10	0.26	2.3	C18
Rectum	12 0	-	-	-	-	0.5	-	-	3.7	-	2.8	3.6	-	8.6	28.6	18.0	25.8	0.8	1.3	0.10	0.33	2.7	C19-20
Anus	2 0	-	-	-	-	-	-	-	-	1.9	-	-	-	-	14.3	-	-	0.1	0.2	0.01	0.08	0.5	C21
Liver	47 0	-	-	0.6	-	1.0	2.2	0.8	5.0	15.6	8.5	21.7	49.7	25.7	43.0	36.0	12.9	3.0	4.9	0.66	1.05	8.6	C22
Gallbladder etc.	0 0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	0.0	0.00	0.00	0.0	C23-24
Pancreas	6 0	-	-	-	-	-	-	-	1.2	1.9	2.8	3.6	-	8.6	14.3	-	-	0.4	0.6	0.09	0.16	1.3	C25
Nose, sinuses etc.	6 0	-	-	-	-	-	-	0.8	-	-	5.7	-	7.1	-	14.3	-	12.9	0.4	0.6	0.07	0.14	1.4	C30-31
Larynx	4 0	-	-	-	-	-	0.5	0.8	-	2.8	3.6	-	-	-	-	-	-	0.3	0.4	0.04	0.04	0.4	C32
Trachea, bronchus and lung	17 0	-	-	-	-	0.5	-	0.8	5.0	1.9	2.8	10.9	7.1	25.7	14.3	36.0	-	1.1	1.8	0.24	0.49	3.4	C33-34
Other thoracic organs	1 0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.1	0.1	0.04	0.04	0.3	C37-38
Bone	12 0	-	1.1	0.6	0.6	0.5	-	0.8	-	-	-	3.6	7.1	8.6	14.3	18.0	12.9	0.8	1.3	0.11	0.28	2.2	C40-41
Melanoma of skin	2 0	-	-	-	-	-	-	-	-	-	11.4	10.9	-	-	28.6	36.0	-	0.1	0.2	0.05	0.05	0.5	C43
Other skin	13 0	-	0.4	-	-	-	0.5	-	-	-	-	-	-	-	-	-	-	0.8	1.4	0.12	0.44	2.9	C44
Mesothelioma	0 0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C45
Kaposi sarcoma	473 0	-	7.8	9.7	4.8	3.0	21.3	51.4	82.7	78.6	48.3	47.1	85.2	25.7	71.6	36.0	206.2	30.3	49.6	2.77	3.30	40.0	C46
Connective and soft tissue	12 0	-	0.5	1.2	-	-	-	1.6	-	3.9	2.8	3.6	7.1	-	-	-	25.8	0.8	1.3	0.10	0.10	1.6	C47-C49
Breast	6 0	-	-	-	-	-	-	0.8	-	-	2.8	-	-	14.2	8.6	-	12.9	0.4	0.6	0.13	0.13	1.4	C50
Penis	12 0	-	-	-	0.6	-	-	0.8	-	5.8	-	-	7.1	-	14.3	71.9	12.9	0.8	1.3	0.07	0.50	2.9	C60
Prostate	63 0	-	-	-	-	-	-	-	-	1.9	5.7	25.4	14.2	77.2	157.6	251.7	219.1	4.0	6.6	0.62	2.67	19.5	C61
Testis	3 0	-	-	-	-	-	-	-	1.2	-	-	-	-	8.6	-	18.0	-	0.2	0.3	0.05	0.14	0.8	C62
Other male genital organs	3 0	-	-	0.6	-	-	-	-	-	-	-	-	7.1	-	-	18.0	-	0.2	0.3	0.04	0.13	0.7	C63
Kidney	9 0	1.8	-	-	-	-	-	1.6	-	1.9	-	-	-	-	14.3	-	-	0.6	0.9	0.03	0.10	0.9	C64
Renal pelvis	0 0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	0.0	0.00	0.00	0.0	C65
Ureter	0 0	-	-	-	-	-	-	-	-	-	2.8	3.6	-	8.6	-	-	-	0.0	0.0	0.00	0.00	0.0	C66
Bladder	9 0	0.4	-	-	-	-	-	0.8	-	-	-	-	-	-	-	-	25.8	0.6	0.9	0.08	0.26	2.0	C67
Other urinary organs	0 0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	0.0	0.00	0.00	0.0	C68
Eye	29 0	3.5	1.1	0.6	0.6	1.5	0.5	2.4	1.2	5.8	2.8	7.2	7.1	-	-	-	-	1.9	3.0	0.17	0.17	2.2	C69
Brain, nervous system	2 0	0.4	-	-	-	-	-	-	-	-	-	-	-	8.6	-	-	-	0.1	0.2	0.04	0.04	0.4	C70-72
Thyroid	3 0	-	-	-	0.6	-	0.5	-	-	-	-	-	-	8.6	-	-	-	0.2	0.3	0.05	0.05	0.4	C73
Adrenal gland	1 0	0.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.1	0.1	0.00	0.00	0.0	C74
Other endocrine	1 0	-	-	-	-	0.5	-	-	-	-	-	-	-	-	-	-	-	0.1	0.1	0.00	0.00	0.0	C75
Hodgkin disease	3 0	-	0.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.2	0.3	0.05	0.12	0.8	C81
Non-Hodgkin lymphoma	40 0	3.5	6.4	3.6	1.8	1.0	1.6	1.6	2.5	1.9	5.7	3.6	7.1	8.6	14.3	-	-	2.9	4.8	0.24	0.31	3.3	C82-85,C96
Immunoproliferative diseases	0 0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	0.0	0.00	0.00	0.0	C86
Multiple myeloma	3 0	-	-	-	-	-	-	0.8	-	-	-	-	-	8.6	-	18.0	-	0.2	0.3	0.05	0.14	0.8	C80
Lymphoid leukaemia	2 0	0.4	-	-	-	-	-	-	-	1.9	-	-	-	-	-	-	-	0.1	0.2	0.01	0.01	0.2	C91
Myeloid leukaemia	1 0	-	-	-	-	-	-	-	-	-	-	3.6	-	-	-	-	-	0.1	0.1	0.02	0.02	0.2	C92-94
Leukaemia unspecified	4 0	-	-	0.6	-	0.5	-	0.8	1.2	-	-	-	-	-	-	-	-	0.3	0.4	0.02	0.02	0.2	C95
Myeloproliferative disorders	0 0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	0.0	0.00	0.00	0.0	MFPD
Myelodysplastic syndromes	0 0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.0	0.0	0.00	0.00	0.0	MDS
Other and unspecified	41 0	-	1.1	-	1.2	-	1.0	1.1	2.4	5.0	9.7	2.8	10.9	28.4	25.7	14.3	103.1	2.6	4.3	0.45	0.52	6.8	C64U
All sites	966 0	20.2	19.9	13.3	8.3	30.0	61.7	104.7	107.3	159.5	156.1	242.8	298.1	411.9	630.3	665.2	850.7	61.9	8.17	14.65	135.1	ALL	
All sites but C44	953 0	19.8	19.9	13.3	8.3	30.0	61.1	104.7	107.3	159.5	144.8	232.0	298.1	411.9	601.6	629.3	850.7	61.0	100.0	8.05	14.21	132.2	ALL/C44

Figure 6.6.: Incidence Rates per 100.000

### 6.2.2.1. Portable Document Format (PDF)

PDF is a well known standard document format that can be read in for example Adobe Reader, installed on most computers

### 6.2.2.2. Post Script (PS)

Post Script is the predecessor of PDF and is an open format that can be displayed using for example GSview.

### 6.2.2.3. Scalable Vector Graphics (SVG)

Scalable Vector Graphics, or SVG, is a scalable editable file format that lets you further tailor the graphics. This can be viewed in many programs, including most modern internet browsers (as it is part of the HTML5 standard), and further edited in programs like InkScape or Adobe Illustrator. They can also be directly embedded in many documents using Google Docs or Open Office.

### 6.2.2.4. Portable Network Graphics (PNG)

PNGs are useful for presentation on screen or online, but less adapted to print, as they are not scalable.

### 6.2.2.5. Windows Meta File (WMF)

WMFs are useful for presentation on screen, but less adapted to print, as they are not scalable. (If possible it is better to use PNG as that is an open format.

### 6.2.2.6. Tabulated data (HTML)

This is the data from the tables in HTML format for further processing.

## 6.2.3. CanReg5 Chart Editor

CanReg5 has a built in chart editor for charts that support it. Using this you can preview and modify the chart before saving it as PNG, PDF or SVG or printing it. (See figure 6.7 on the facing page.)

## 6.2.4. CanReg and SEER\*Stat

CanReg provides a facility to export data to be used in the free SEER\*stat. (Windows only.) To do so you'll have to first get hold of and install SEER\*Stat and SEER\*Prep from the SEER website. (<http://seer.cancer.gov/seerprep/> and <http://seer.cancer.gov/seerstat>)

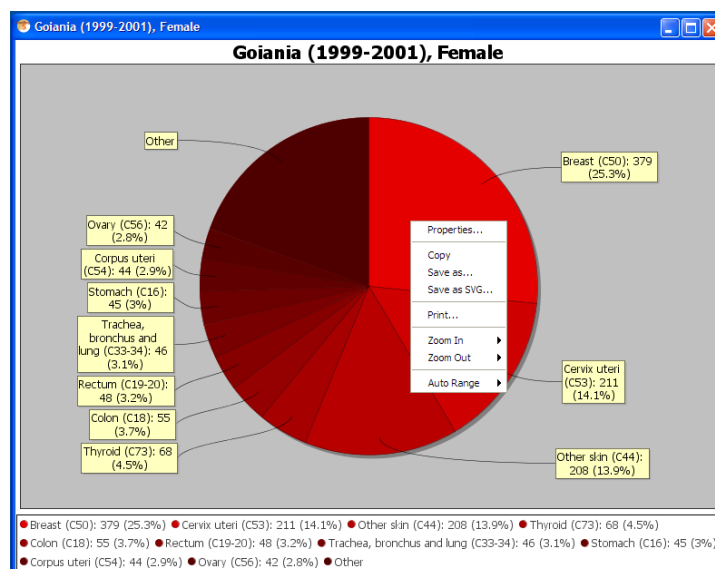


Figure 6.7.: CanReg5 chart editor

Please note that the following limits apply (for now, as SEER\*Stat wants things like this, but in the future we might be able to work around them):

- PatientIDs have to be 8 digits or less.
- Population datasets has to have 5 year age groups.

#### 6.2.4.1. In CanReg

Afterwards you can use the table builder in CanReg and choose the SEER\*Prep as table type, specify the range of years you want to look at, the population datasets matching each year in the period and click “Generate files for SEER\*Prep”. This will ask you to specify the base name of the files for SEER\*Prep, i.e. “Goiania1999-2001”. (Please note that you can only use ASCII (English) characters in the paths of the datafiles.) CanReg will add “cases-” and “pop-” for respectively the incidence file and the population file, as well as the suffix “.txd” so that SEER\*Prep recognises the files. CanReg will also create a file describing the data for SEER\*Prep - a so-called .dd-file.

Start SEER\*Prep.

#### 6.2.4.2. In SEER\*Prep

Load the .dd file you created with CanReg using the little folder icon.

## 6. Analysis

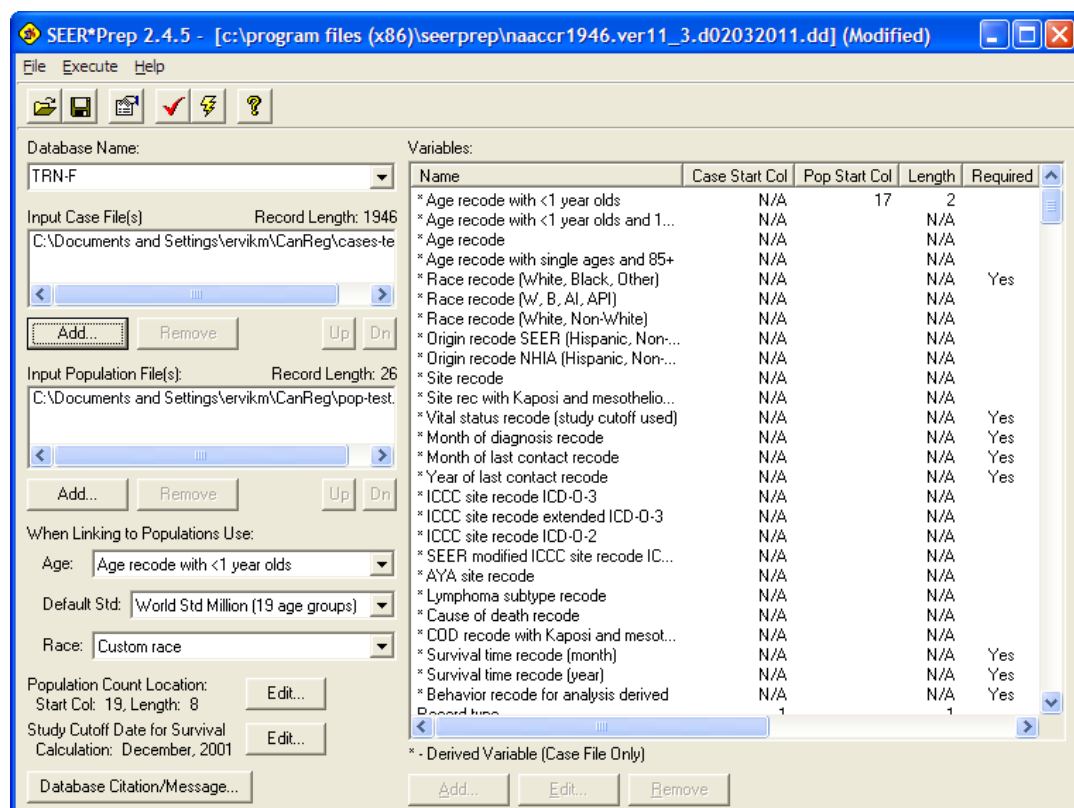


Figure 6.8.: SEER\*Prep in action

You can also do this manually. You then have to first specify that your files follow the NAACCR 1946 v11.3 format, by opening the corresponding .dd file from the folder you installed SEER\*prep to. (Most probably “naaccr1946.ver11\_3.d02032011.dd”) Then under “Database name:” you can specify something more instructive than the default. Then you add your case file and population files from above using the “Add...” buttons. Do the following choices under “When Linking to Populations Use:”

- Age: “Age recode with <1 years old”
- Default Std: “World Std. Million (19 age groups) (recommended)”
- Race: “Custom race”
- Leave Population Count Location as it is (Start col 19, length 8)

Click “Edit...” on the study cutoff date for survival and set it to something after the last entry in the database. (The date of the export?) (This is not important for now as the survival features in SEER\*Stat is not valid for data from CanReg yet.)

Click the “Lightning” button. (Or go to the “Execute” menu and choose “Create file...” SEER\*Prep will then ask you where you want it to write it’s report file. Choose somewhere you’ll be able to find it back, i.e. the Desktop. (If you have already created a database with this name, SEER\*Prep will ask you if you want to overwrite it. Choose “yes”, if you want.)

Next you get some choices regarding invalid values. (The first time you might want to choose “Do not exclude any records”.)

That’s it. SEER\*Prep will generate the files that SEER\*Stat needs. Exit SEER\*Prep and start SEER\*Stat.

(A more detailed walk through on this process can be found here:

<http://seer.cancer.gov/seerprep/example.html>)

#### 6.2.4.3. In SEER\*Stat

Your data from above should now appear in SEER\*Stat under the name you chose for the database above ready to be exploited in SEER\*Stat.

You might have to create a new profile, by going to the “Profile” menu and choosing “Add New...”

Click “Add Local...” for Data Locations and choose the folder that SEER\*Prep writes data to.

(You can also untick the Client-Server ssp-connection if you want.)

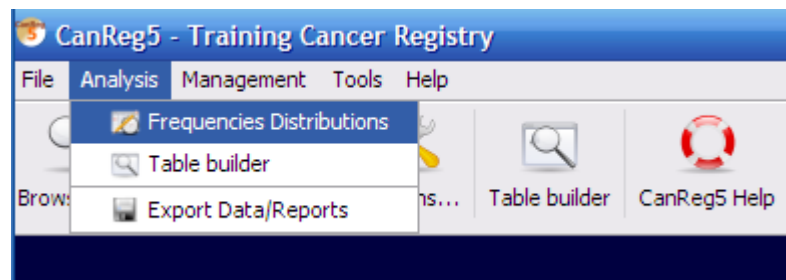
Under “Other variables” you have to specify valid paths for the “User Variables”, “Open or Save” and “Temporary Files”.

## 6.3. Frequency distributions

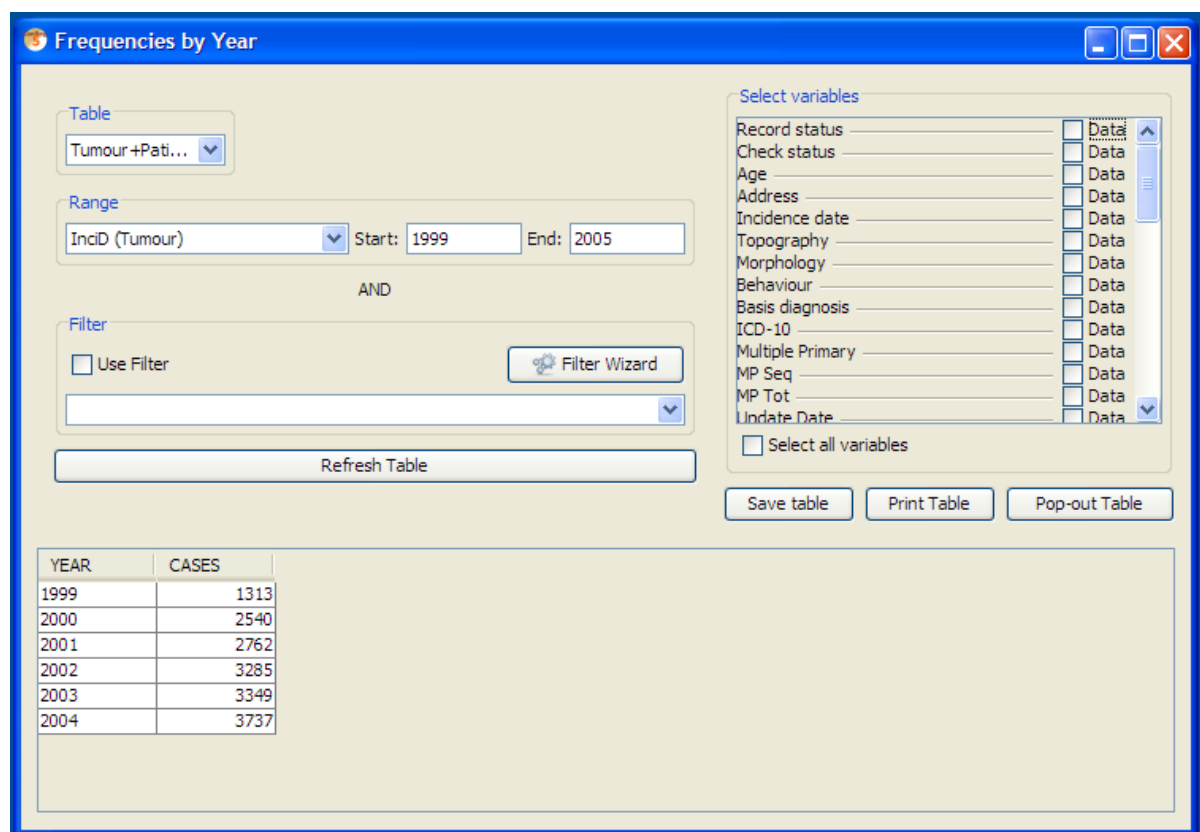
Frequencies distributions let you look at the data in your database as frequencies by year. You can cross-tabulate several variables. To start this module go to Analysis –

## 6. Analysis

Frequencies Distributions:



If you click Refresh table with no filter and no selected variables you get a table of cases per year.



YEAR	CASES
1999	1313
2000	2540
2001	2762
2002	3285
2003	3349
2004	3737

You can sort by any field by clicking its header. For example by number of cases:

YEAR	CASES ▾
1995	1643
1994	996
1993	837
1992	707
1991	665
1996	568
1990	396
1989	177
1901	1
1939	1
1977	1
1980	1
1984	1
1986	1
1987	1
1999	1

You can filter the result by adding a filter like for example on incidence date. You can also add as many variables as you want.

With the save table button you can write the table to a comma separated file (.CSV) that can be opened in most programs (*Excel*, *Stata*, *R*...) for further analysis.

The table can also be selected and copied and pasted into *Excel*, for example. (No right-click shortcut for that is implemented yet, but you can select the lines you want and press Ctrl-C (on Windows and Linux) or Apple-C (on Mac) and paste it into other programs.

### 6.3.1. Pivot tables

If you save a table with only one stratifier (selected variable) CanReg5 also saves a pivot table based on the data in the table. Here's an example with the sex variable:

	A	B	C	D	E	F	G	H	I	J	K	L
1	Sex	SEX	1999	2000	2001	2002	2003	2004	2005	2006	2010	MISSING
2	Male	1	595	1097	1269	1397	1428	1697	1947	2113	0	0
3	Female	2	718	1443	1493	1888	1921	2040	2082	2241	0	0
4		MISSING	0	0	0	0	0	0	0	0	1	13





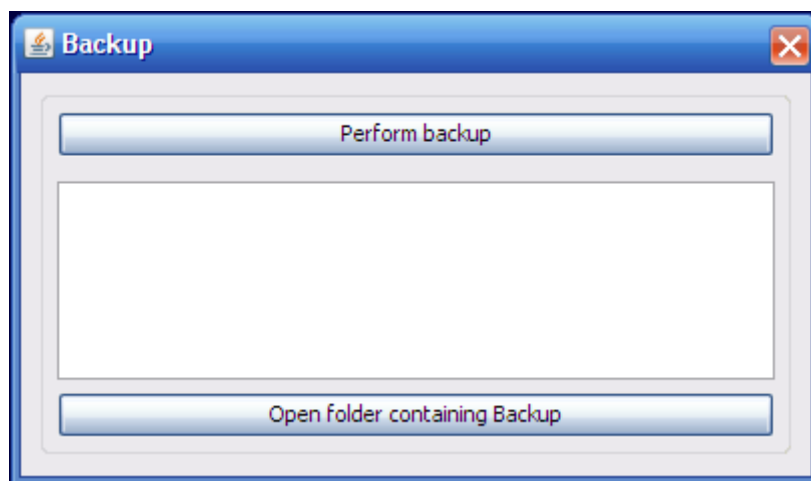
## 7. Management

### 7.1. Back up and restore

Backup-functionality can be found under the Management menu.

#### 7.1.1. Perform backup

Under “Management” click “Backup”



Then click “Perform backup”. This creates the backup of the CanReg5 database on the server machine. If you are on the server machine, you can see the files you created by clicking “Open folder containing Backup”. It is stored in the **CanReg server** folder under Backup and 3 digit code of the registry and then the date of the backup. On my machine, for example, it is “C:\Documents and Settings\morten\.CanRegServer\Backup\TRN\2009-01-14”.

#### 7.1.2. Restore from backup

If you are on the server machine you can restore the backup you created above by clicking “Restore” in the “Management”-menu restore your Registry definition, population dataset, dictionary codes and any data from a CanReg5 backup.

You will probably perform this either:

- at first installation, or

## 7. Management

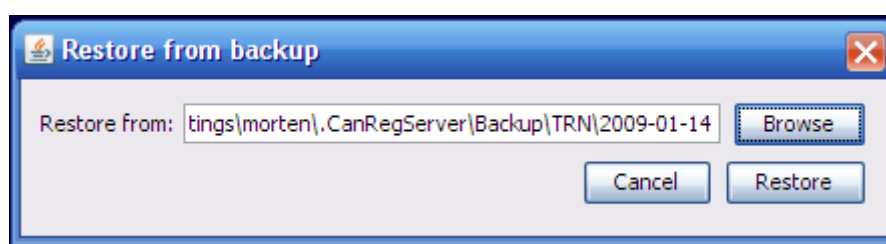
- if re-installing on a new computer, or
- recovering from a lost or damaged computer situation.

There are two ways to do this. The easiest one is to use the tool “Install new system” that we have access to when CanReg launches.

### 7.1.2.1. Restore with “Install new system”

Click “Install new system” and browse to the XML file describing this system and click Open. If this is part of a backup CanReg will tell you that a backup has been detected and asks you if you want to restore this. Click yes and it will restore this.

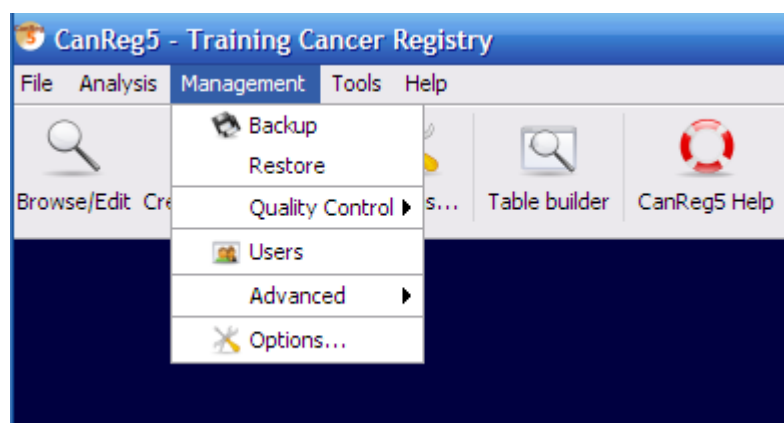
### 7.1.2.2. Restore with “Restore from backup”-tool



Simply choose the folder corresponding to the backup you want to restore using the “Browse” button or enter the path manually and then click “Restore”. You need to specify the folder containing a folder with a folder named your registry code. That is most of the time the date of backup. In the example above we restore a backup from January 2009. If you select the folder within this “date” folder you will get an error message and the restore will not be performed.

## 7.2. Manage users

The user manager is located under management – users:



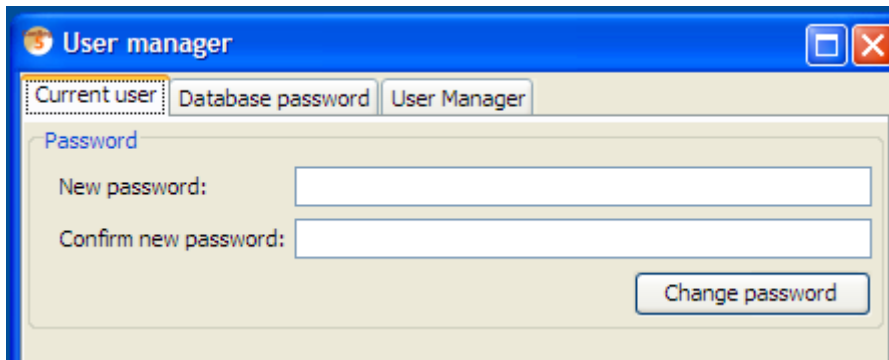


Figure 7.1.: Change your own password

### 7.2.1. Change your own password

To change your own password, go to the Current user tab in the user manager and enter your new password twice. (See figure 7.1.)

### 7.2.2. Database password

The CanReg5 database can be encrypted using 56-bit DES encryption. To enable this you need to provide a minimum 8 characters long password on the “Database password” tab. (See figure 7.2 on the following page.) This protects your database in case somebody gets hold of it. The user will now be asked for this password everytime she launches the server. (Once it is running others can connect to the database as usual.) This is a one way process - the only way to remove the encryption of your database is to export all the data, build an empty database and import everything again. The password can, however, be changed at any point in time.

Please note that if you loose this password you loose access to all the data in CanReg! (And it is nothing that can be done about it.)

This functionality is only available to users with supervisor access.

### 7.2.3. User manager

If you are logged in with Supervisor rights you have access to the User manager part of the user manager. (See figure 7.3 on page 85.) This allows you to add and delete users. Each user has their own login name, password and permission level.

- A Supervisor can use all options;
- A Registrar can perform most of data entry except to confirm rare cases or possible duplicates. Changing the Dictionary and User Administration (adding users, changing user levels) are also prohibited.

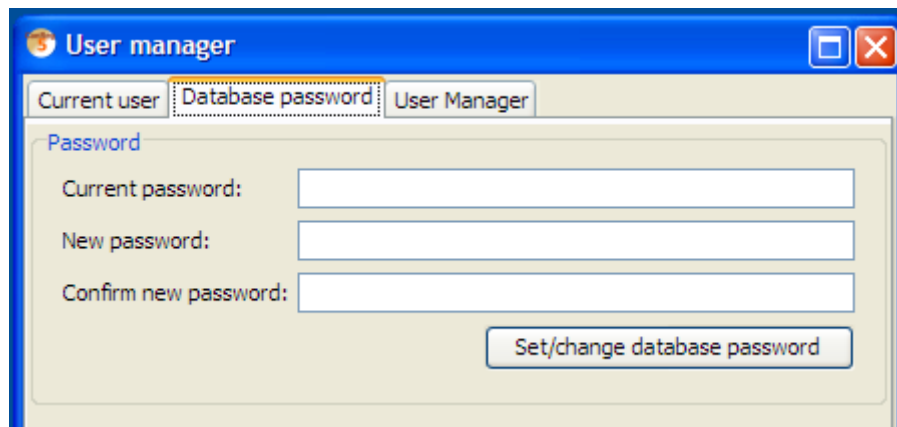


Figure 7.2.: Set database password

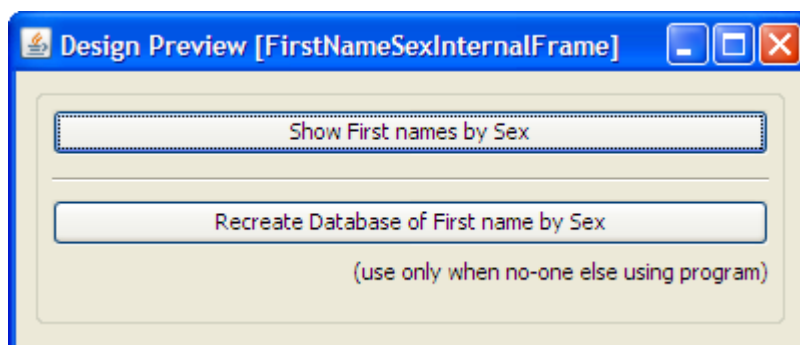
- An Analyst cannot make any changes to the database. Only analysis options are available.

To add a new user, click “Add user” fill in the user’s login name. This will create a default user with registrar permissions. To change this select the user and pick a suitable user right level and click “Update user”.

You can change a permission level for a user and hit "Update user", or delete a user by selecting the user and then clicking the "Delete" button. The default password of any user is its user name. This should of course be changed at first login by the user himself.

## 7.3. Quality control

### 7.3.1. Name and sex



Click the button "Show First names by Sex" to view all the names used in your registry:

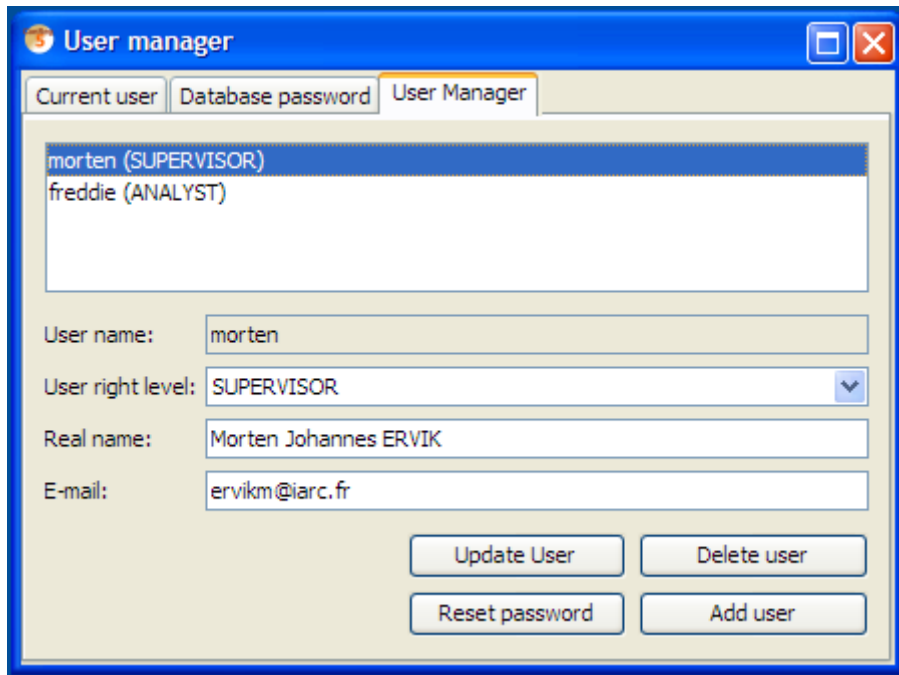


Figure 7.3.: User manager

- Male names
- Female names
- Unisex names (used by both sexes)

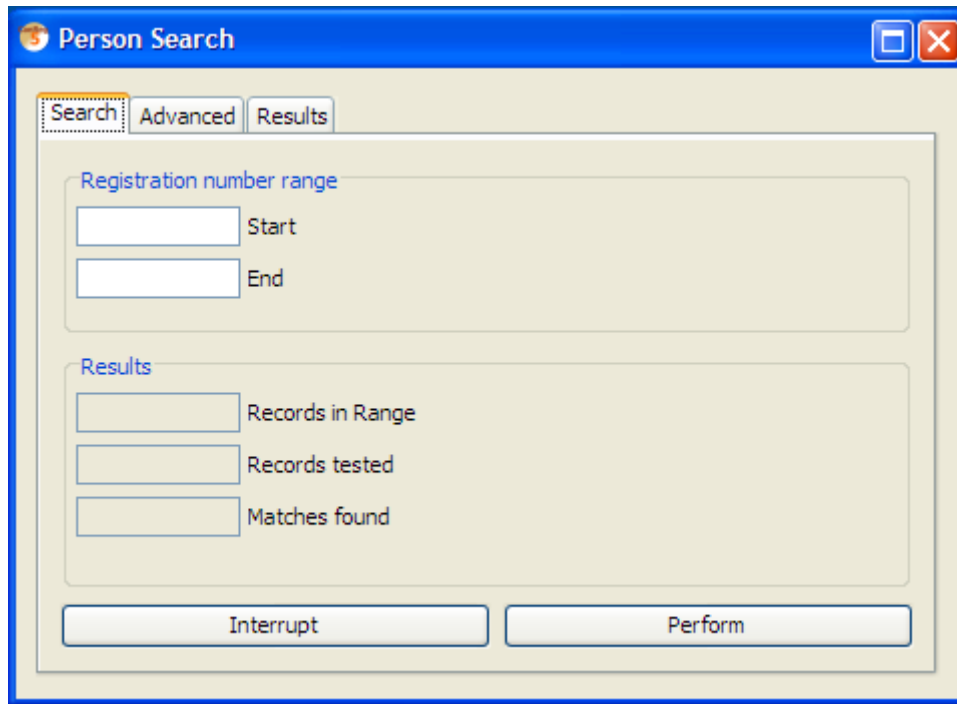
You should periodically review these lists to check for obvious errors.

New names are automatically added to the lists.

However, if you have made corrections and some names are in the wrong category, the supervisor can recreate the lists by clicking the second button. In a network environment, only do this when nobody else is using the program.

## 7. Management

### 7.3.2. Person search

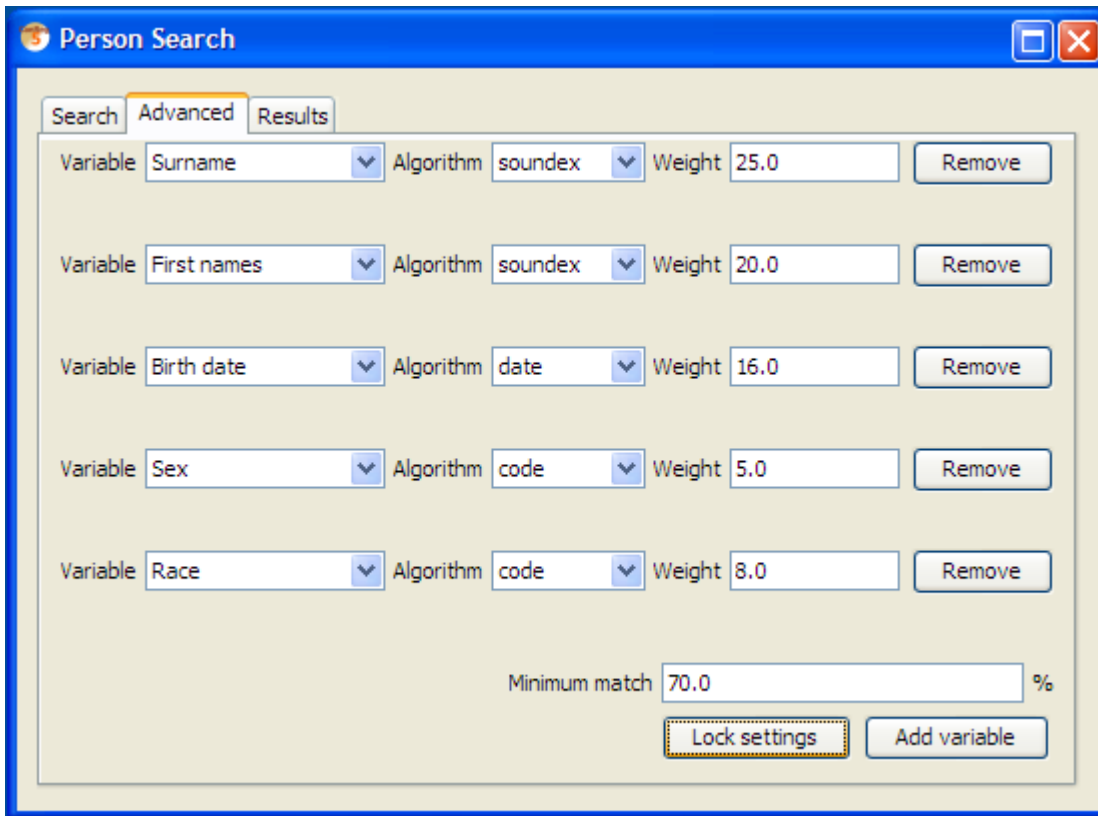


The image shows a software window titled "Person Search" with a blue title bar and standard window controls. Inside, there are three tabs: "Search" (selected), "Advanced", and "Results". The "Search" tab contains two sections. The first section, "Registration number range", has two input fields labeled "Start" and "End". The second section, "Results", has three input fields labeled "Records in Range", "Records tested", and "Matches found". At the bottom of the window are two buttons: "Interrupt" and "Perform".

Registration number - range start and end.

#### 7.3.2.1. Advanced

Change weights and search variables.



The image shows a software window titled "Person Search" with a blue title bar and standard window controls. It contains three tabs: "Search", "Advanced" (which is selected), and "Results". The "Advanced" tab displays five search criteria, each with a "Variable" dropdown, an "Algorithm" dropdown, a "Weight" text box, and a "Remove" button. The criteria are: Surname (soundex, 25.0), First names (soundex, 20.0), Birth date (date, 16.0), Sex (code, 5.0), and Race (code, 8.0). At the bottom, there is a "Minimum match" text box set to 70.0%, a "Lock settings" button, and an "Add variable" button.

Variable	Algorithm	Weight	Action
Surname	soundex	25.0	Remove
First names	soundex	20.0	Remove
Birth date	date	16.0	Remove
Sex	code	5.0	Remove
Race	code	8.0	Remove

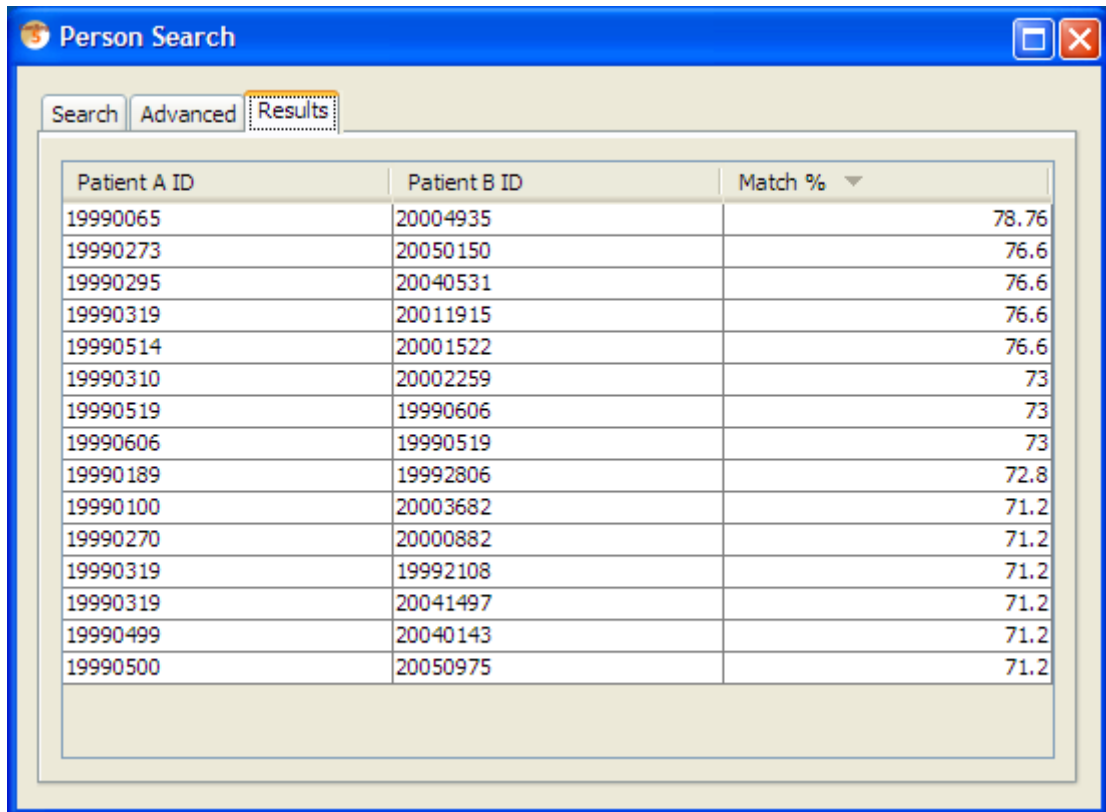
Minimum match: 70.0 %

Buttons: Lock settings, Add variable

#### 7.3.2.2. Results

Results from the search appears here as potential duplicates are found.

## 7. Management

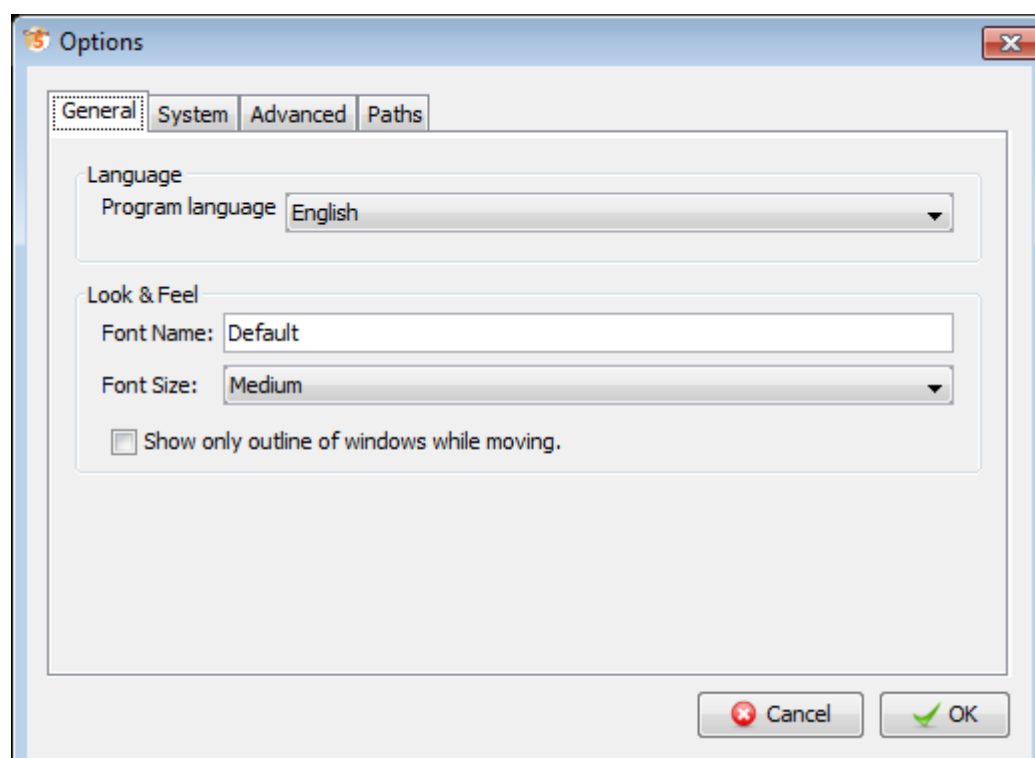


The image shows a software window titled "Person Search". It has three tabs: "Search", "Advanced", and "Results". The "Results" tab is selected, displaying a table with three columns: "Patient A ID", "Patient B ID", and "Match %". The table contains 18 rows of data, showing various patient IDs and their corresponding match percentages. The window has a blue title bar and standard window controls (minimize, maximize, close) in the top right corner.

Patient A ID	Patient B ID	Match %
19990065	20004935	78.76
19990273	20050150	76.6
19990295	20040531	76.6
19990319	20011915	76.6
19990514	20001522	76.6
19990310	20002259	73
19990519	19990606	73
19990606	19990519	73
19990189	19992806	72.8
19990100	20003682	71.2
19990270	20000882	71.2
19990319	19992108	71.2
19990319	20041497	71.2
19990499	20040143	71.2
19990500	20050975	71.2



## 7.4. Options



### 7.4.1. [UPDATED]Language

Change the program language and date format (for data entry) of CanReg5. (You need to restart your CanReg5 system if you change program language.) The date format can be any format specified by a text string using dd for days, MM for months, and yyyy for year, for example: “yyyyMMdd” (the default), “dd/MM/yyyy”, or “MM.dd/yyyy”.

### 7.4.2. Look & Feel

You need to restart your CanReg5 system if you change font style or font size.

#### 7.4.2.1. Change font

Set the font size to sizes from Small to Big.

#### 7.4.2.2. Change font style

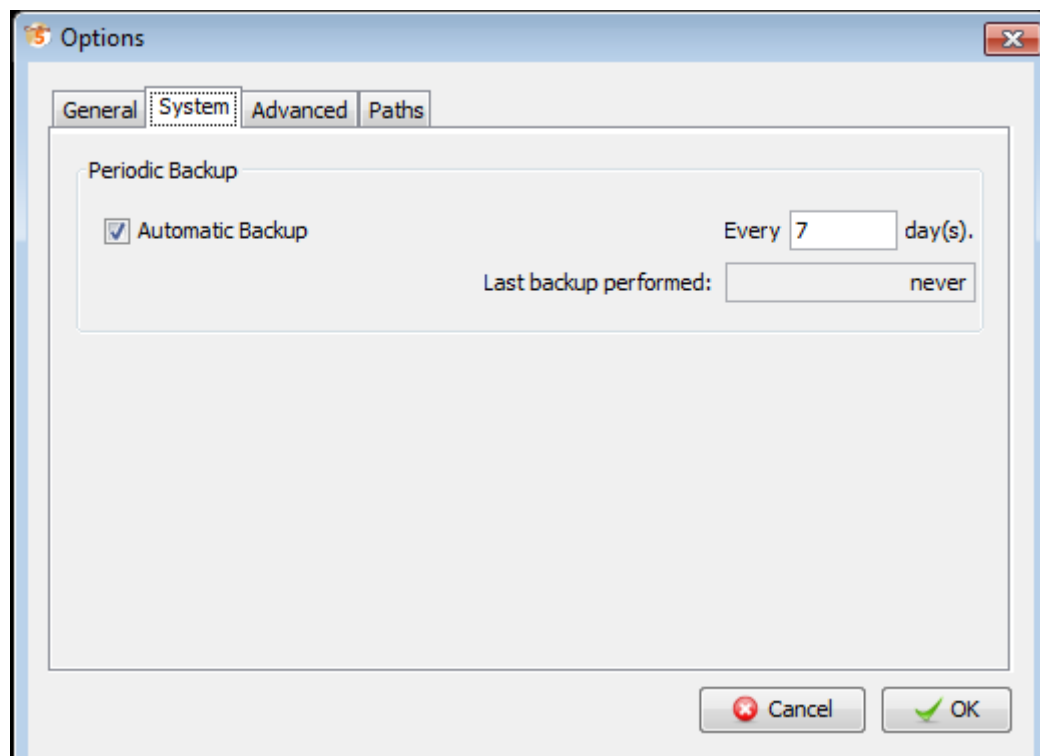
You can set the font of CanReg by typing the name of the font you want to use in the Font name text box. (To find out what fonts you have installed on your Windows system you can take a look at the Fonts section on your Windows Control Panel.) You can use this if the default font doesn't contain all the characters you would like to use or if you just prefer the look of another font.

## 7. Management

### 7.4.2.3. Show only outline

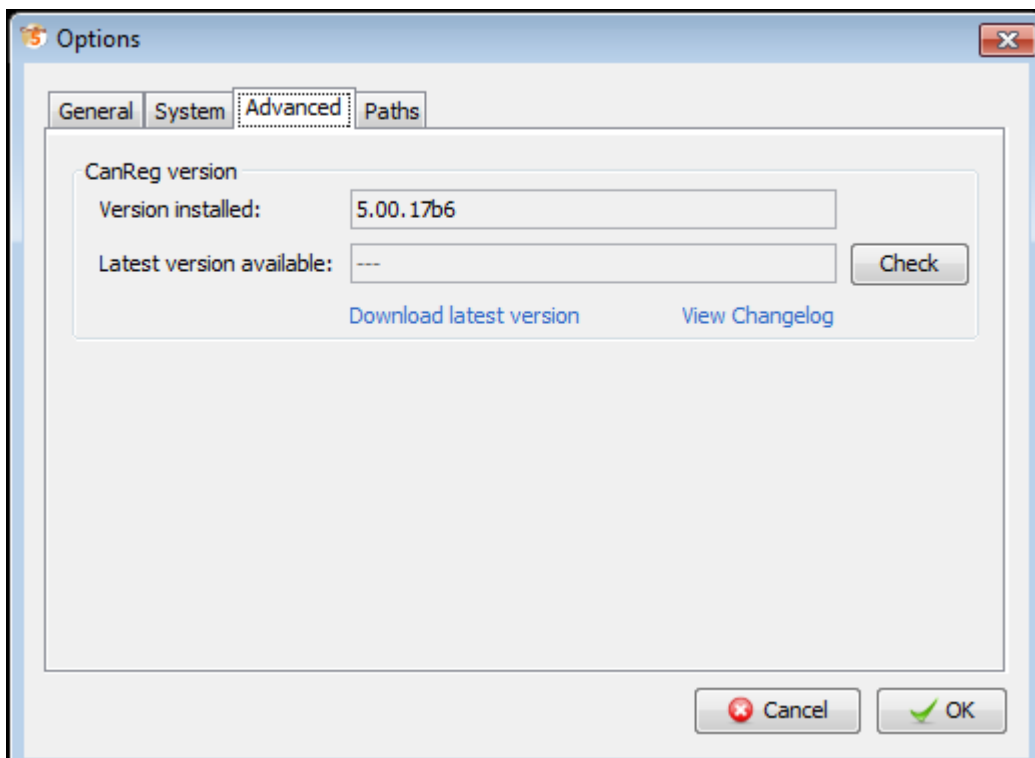
The “Show only outline of windows while moving” will increase the performance of the user interface of CanReg. Tick this if moving windows around the screen is slow.

### 7.4.3. Periodic backup



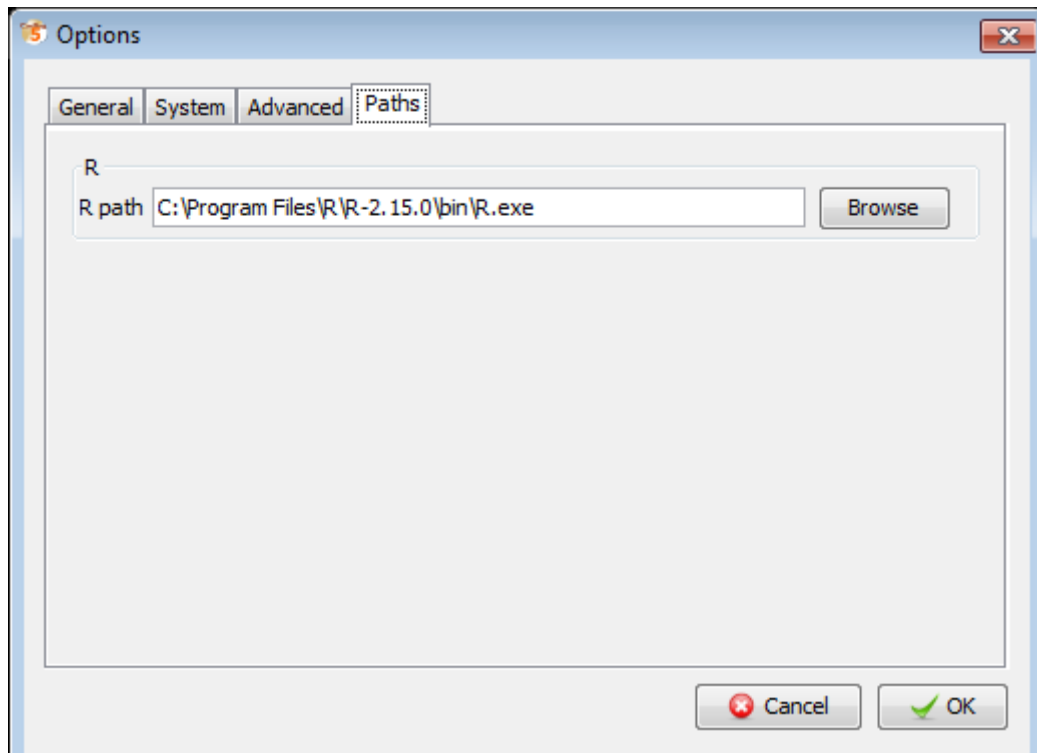
Reminder backup options

#### 7.4.4. CanReg version



Let the program check to see if the latest version of CanReg is installed. This will access the internet to find the most recent version. You can click “Download latest version” to get the latest universal version of CanReg (The zip-file.) or click “View Changelog” to see what is new in CanReg.

### 7.4.5. Paths



Here you can set the paths to the various helper applications of CanReg. (So far only “R”.)

# Part IV.

## Appendix



# A. Frequently asked questions (FAQ)

## A.1. Server

**Q:** When I click the “Launch Server”/“Test Connection” button, it takes **more than 3 minutes** to launch the server and I get the message that the “Server [is] already running”. Afterwards I cannot log in.

**A:** Xenios found the following solution: On our PCs, we use Microsoft Internet Explorer 7. In the “Tools / Internet Options / Connections / LAN Settings” we have a tick in the checkbox for “Use a proxy server for your LAN” and the address and port of the proxy server filled in accordingly. I put a tick in the checkbox for “Bypass proxy server for local addresses”, clicked the “Advanced” button and typed the IP address of the PC (localhost) in the “Do not use proxy server for addresses beginning with:” box.

**Q:** When I try to log on to the CanReg server I get the message: “Exception creating connection to: aaa.bbb.ccc.ddd; nested exception is: java.net.SocketException: Malformed reply from SOCKS server”

**A:** This can be solved using the same procedure as the previous answer.

**Q:** Is CanReg5 only designed for local networks or does it allow remote registration?

**A:** CanReg5 is designed for local networks, yes, but traffic is over standard TCP/IP so you can tunnel this through secure channels (SSH/VPN etc) to do remote registration. (This is not "officially" supported, as we haven't been able to test it enough (yet).)

## A.2. Conversion CanReg4 to CanReg5

**Q:** In what table should the variable age be stored?

**A:** The tumour table. Like that, if the same patient has a new tumour you can (probably) keep the patient record and just add a new tumour record. Birth date is stored in the patient table. (Incidence date with the tumour.)

### *A. Frequently asked questions (FAQ)*

**Q:** Do I need to “install” the CanReg5 system definition file after converting from CanReg4 using the built in tool?

**A:** After converting the system you don’t need to “install” it afterwards as the XML file is automatically copied to your system folder during conversion.

**Q:** I get errors during import of the data from CanReg4. The process stops after a certain percentage every time.

**A:** Try exporting your data with a comma separated variables instead of the default tab-separated ones (or vice versa) and see if that helps.

**Q:** In CanReg5, how do you clear previous data and import fresh ones from canreg4.

**A:** To clear the previous data from CanReg5 you need to delete the database files. These can be found in your user folder under .CanRegServer\Database (On my machine it is under C:\Documents and Settings\ervikm\.CanRegServer\Database ). Please note that this deletes the dictionaries and population data as well, so you might want to export those prior to deleting this. Then you can relaunch CanReg and the server and the empty database will be rebuilt. It is a good idea to keep a backup of an empty database - just containing dictionaries (and population data) to get back to this state of the database if needed.

## **A.3. Dictionary**

**Q:** Can I import dictionaries from other CanReg systems to my own?

**A:** Since most CanReg systems have different dictionary structure (length of codes, order of dictionaries etc.) you need to import the dictionary corresponding to your system or do necessary modifications.

## **A.4. Analysis/tables**

**Q:** What program can I use to view the post script files with?

**A:** PostScript is an open standard, so you can use many different tools to view them. (You can in many cases even send them directly to a printer.) Apple’s OSX and most Linux-distributions (Ubuntu, RedHat, SuSE etc) come with a tool to view them by default. On Windows the tool I recommend is the open sourced and free GSview. (Available from: <http://pages.cs.wisc.edu/~ghost/gsview/> ) To run GS View you need to install Ghostscript first. This can be downloaded from here: <http://pages.cs.wisc.edu/~ghost/doc/GPL/gpl864.htm> (Scroll all the way down, under the heading Microsoft windows and download the “GPL Ghostscript 8.64 for 32-bit Windows (the common variety)”



(<http://mirror.cs.wisc.edu/pub/mirrors/ghost/GPL/gs864/gs864w32.exe>) Run this file to install Ghostscript. Then you can get GSView from here: <http://pages.cs.wisc.edu/~ghost/gsview/get49.htm> (Most probably, you should pick the Win32 self extracting archive - the first download option. <http://mirror.cs.wisc.edu/pub/mirrors/ghost/ghostgum/gsv49w32.exe>) Run this file to install GSView.

## A.5. Import

**Q:** Some letters are distorted/missing in records after an import. (For example when importing Arabic names.) Why is that?

**A:** If the data is from a program that does not code the data using Unicode (for example previous versions of CanReg) you need to specify the coding scheme/“codepage” during import of that file to your database. If you pick the wrong one your data might get distorted. To solve this problem you need to re-import the data. Please use the preview button during import to see to that you have the right coding scheme.

**Q:** While importing my data the process brakes down before the end and I get an error message saying that there is something wrong with the file.

**A:** First of all make sure that you have the latest version of CanReg installed, as some of the earlier versions had a bug that manifested itself while importing files. Then look at the line where the import brakes down and see if you have some unclosed quotes in any of the fields. This will brake down the import as CanReg allows qoutes around strings that contain the separating character, but not unclosed quotes.  
Another possibility is that your Java Runtime Environment has been set up with not enough memory. To address this please see [F.1.1 on page 121](#).

## A.6. Database design

**Q:** What is the minimum data set that one would collect?

**A:** The minimum data set required is up to your registry to define. Please refer to [this publication](#) for more information: ([Chapter 6](#), for example. ) The bare minimum are detailed in [Figure A.1 on the next page](#).

**Q:** In what database table should the variable X be stored?

**A:** In the **patient** table you store the information about the patient that never changes (or at least very seldom) in the patient table. (That is birth date, sex, names etc.) You also store follow up information in the patient table. (Vital

## A. Frequently asked questions (FAQ)

- Name(s):
  - String, According to local usage, but at least one, eg First name, Middle name, Last Name, Father's name, Grandfather's name, Maiden Name.
- Gender/Sex:
  - Coded, 1 Male; 2 Female; 9 Unknown; (Some registries use 3 for Other.)
- Birth date:
  - String, Stored in the database as yyyyMMdd (but can, in principle, be entered in any format), Allows 99 for unknown day and 99 for unknown month.
- Incidence date:
  - String, Stored in the database as yyyyMMdd (but can, in principle, be entered in any format), Allows 99 for unknown day and 99 for unknown month.
- Address:
  - Coded, According to local usage/needs.
- Ethnic group (optional, but in some cases encouraged):
  - Coded, According to local usage/needs.
- Age:
  - Integer (For legacy reasons this is input as 3 digits, with 999 as unknown. Some registries collect only two digits so unknown age is coded as 99 and 98 and greater is coded 98.)
- (Most valid) Basis of diagnosis:
  - Coded, (1,2,3,4) Non-microscopic, (5,6,7,8) Microscopic, 9 Unknown, 0 Death certificate only.
- Source of information:
  - Coded, According to local usage, eg hospital number.
- Source of information - additional information:
  - String, According to local usage, eg hospital record number, name of physician.

Note: Registries can choose to collect Date of Birth OR Age, but most choose to collect both.

Figure A.1.: Minimum variables.

status, current address, etc.) In the **tumour** table you store all data that can potentially differ between two tumours of the same patient. (Topography, morphology, age, address at the time of the tumour (coded) etc.) Finally, in the **source** table you store all information about the sources of information of any given tumour. (Source name (preferably coded), source type, date etc.)

**Q:** After adding a new/removing an old variable to collect in the database what do I need to do?

**A:** First of all, before adding or removing variables using the built in tool or editing the underlying XML file you should take a backup of your old database. Then you need to export all the data - one table at a time from CanReg5 - so that you have 3 files. One for the patient table, one for the tumour table and one for the source table. Then you need to export the dictionaries. After editing the database structure you will have to delete the old database files, relaunch CanReg5 and import these files to the empty database. Basically what CanReg does when you launch the server is first to load the XML describing the database, then checks to see if the database files exist in the Database folder. If this does exist it checks to see if it corresponds with the XML. If it does not it will not be able to launch the server. (I'll look into creating a better set of error messages.) If it does not find a database folder with the registry code found in the XML it will generate an empty one.

## A.7. Install/update CanReg5

**Q:** Parts of CanReg5 are no longer working after upgrading to the latest version. What should I do?

**A:** First, uninstall CanReg5 and reinstall the latest version. See if that helps. If not write a bug report to [canreg@iarc.fr](mailto:canreg@iarc.fr).



## B. Known issues

### B.1. Known bugs (errors)

- Using a population dataset with a cut off age at less than 85 while building some R-tables leads to errors in the last age group in the tables.
  - Severity: Shouldn't cause loss of data
  - Priority: Medium
  - Category: Tables

### B.2. Known limitations

- You need a population dataset with 5 years age groups for many of the tables to work properly...
- Age can not yet be calculated automatically.
- The result set in the browser is sometimes very slow to scroll around in.
  - Solution: Use filters to minimize the number of records shown at any time or only browse the "Patient" or the "Tumour" table.
  - Severity: Shouldn't cause loss of data
  - Priority: Low
  - Category: Database/Browse
-



## C. Migrating from CanReg4 - Step by Step

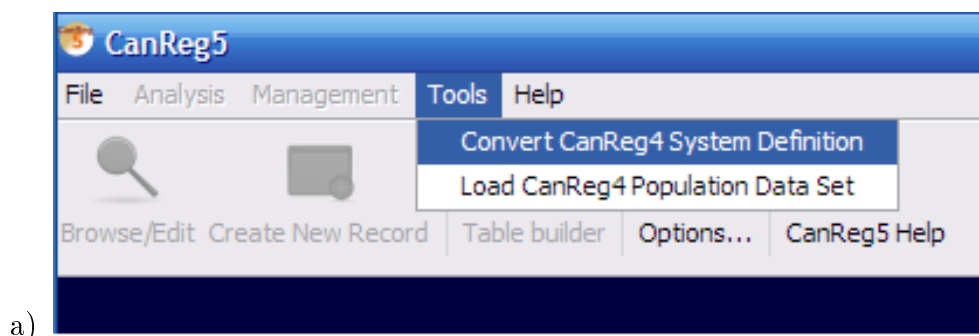
(For a video demonstration of this process, please see webinar 3 - available from the GICR website under “Resources”. (<http://gicr.iarc.fr/en/resources.php#TRAINING>))

Install CanReg5. (See **II on page 15**) Start CanReg5 and it presents you the Welcome window. Do NOT click anything here just yet.

### C.1. Step 1 - Import the variable definitions of CanReg4 to CanReg5

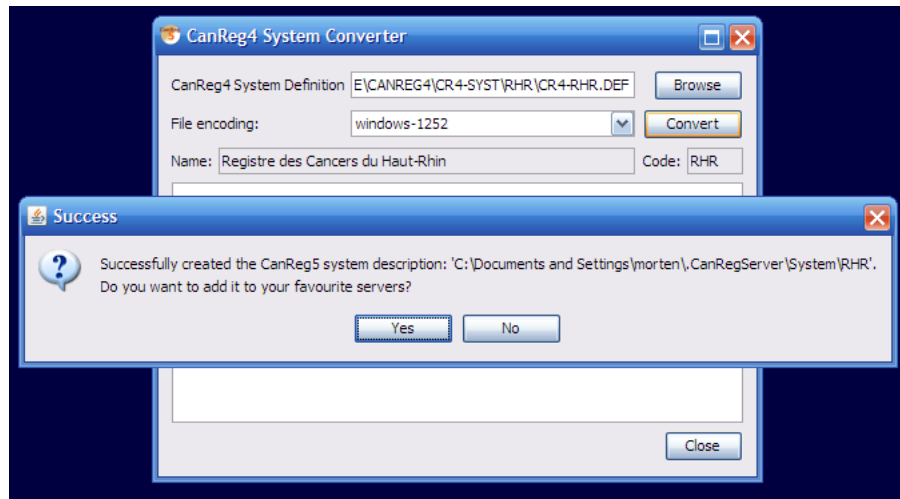
The first step is to import the variables of CanReg4 to CanReg5 - the system definition of CanReg4.)

1. Go to “Tool” in CanReg5 menu and click “Convert system definition”.



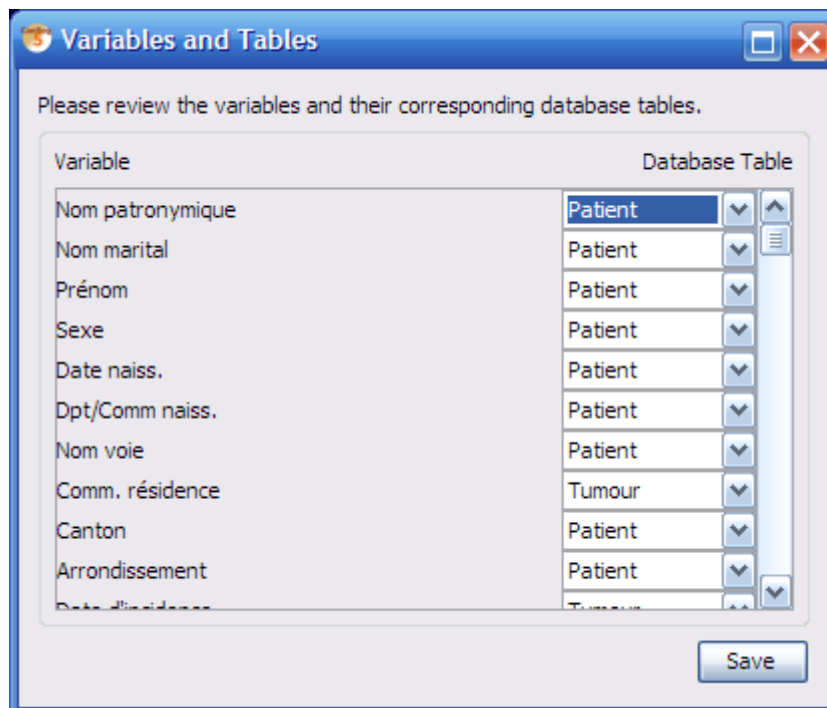
2. Do “Browse” to find your CanReg4 system definition file. (This is a file located in the folder \CR4SHARE\CANREG4\CR4-SYST\ followed by your 3 letter registry code i.e. TRN whose name is ending in .DEF (i.e. CR4-TRN.DEF).)
3. Select your CanReg4 file and double click it or click “Open”.
4. Click “Convert”.
5. The program will then ask you if you want to add this server to your favourites. Click “Yes” here.

### C. Migrating from CanReg4 - Step by Step



a)

6. The next step is the trickiest one during the conversion. Since we go from a tumour based database structure with only one big table with all the tumour and patient related information to a structure with both a table for tumour related information, one for patient related information and yet another one for source information(1.1) we need to specify what variable goes in what table of CanReg5. We recommend putting the unique patient related information (name, date of birth and follow-up variables) in the patient table, source information in the Source table and pretty much the rest (tumour information, age, address etc) in the tumour table.
7. The program presents an initial proposal that you might agree with, but please go through one by one the variables and decide.



a)



## C.2. Step 2 - Import the dictionary from your CanReg4 installation

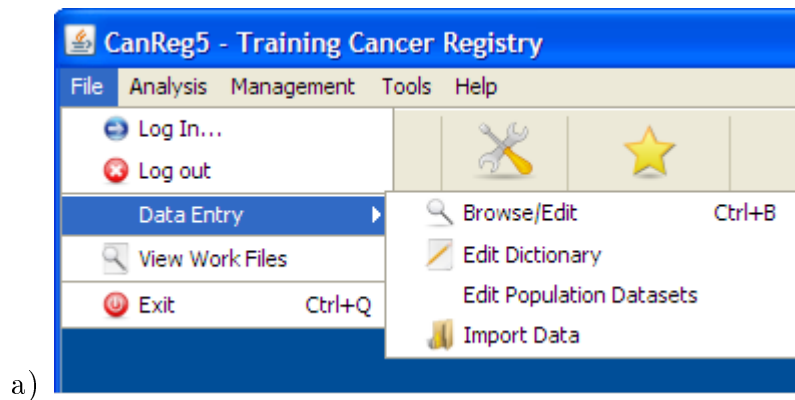
8. Click “Save”. You have now created an XML file that describes your CanReg5 system.
9. Optional: Before you proceed to the next step and launch the server you can, if you want(!), edit this XML file you have created. Either using the built in graphical tool in CanReg5 (3.5) or manually by opening it in a text editor or a dedicated XML editor. The file is located in your user folder under .CanRegServer. (On my machine, for example, running Windows XP it is under: C:\Documents and Settings\morten\.CanRegServer\System.)
10. Click “Login”.
11. Launch the CanReg server
  - a) Click “Settings”
  - b) Click “Launch Server”. (If you get a java firewall query, please confirm that it is OK that java can communicate through you firewall by clicking “OK” or “Yes”.) If this is the first time you launch the server on this machine it will automatically create the database needed for CanReg5.
12. Now it is time to log into the system.
  - a) Click “System”
  - b) Enter username: “morten”
  - c) Enter password: “ervik”
  - d) Click “Login”. The system will now log you onto your CanReg server.

## C.2. Step 2 - Import the dictionary from your CanReg4 installation

The thing we want to do now is to import the *dictionary* from your CanReg4 installation or demo system. (Earlier we only imported the description of what variables exist in you canreg4 database.) This is demonstrated in the video called 06-import-dictionary.avi.

1. If you are migrating from CanReg4 make sure to export the most updated dictionary from your CanReg4 system. (In CanReg4: “Data Entry”, “Dictionary”, “Export dictionary to text file”)
2. Go to “File”, “Data Entry”, “Edit dictionary” in CanReg5

### C. Migrating from CanReg4 - Step by Step



3. Click on “Import complete dictionary from file”.
4. Browse and select the dictionary from you CanReg4 work folder or elsewhere.
5. If CanReg does not detect the encoding of the file automatically, please select it from the drop down list.
6. Click “Preview” to take a look at the dictionary. (Please note that only the first hundred lines or so are shown.)
7. Tick “CanReg4 Format” if you are migrating, leave unticked if you are using the demo system or otherwise are importing a CanReg5 formatted dictionary.
8. Click “Import”. This might take some time. Please note the bar in the lower right indicating that the program is busy.
9. Afterwards you will receive a message of success imported. Click OK.
10. Go back to “File”, “Data Entry”, “Edit dictionary” and verify that the dictionaries have been imported.

### C.3. Step 3 - Import the data from your CanReg4 installation

The next step is to import the *data* from CanReg4 to CanReg5. This is demoed in the video called 07-import-data.avi.

1. Make sure you export the most updated data from your CanReg4 system.
  - a) In CanReg4: “Analysis”, “Export data”
  - b) Tick “Export all variables”.
  - c) Choose variables names short
  - d) Under “Export File options” choose “Comma separated variables”
  - e) Untick “Format date”

### C.3. Step 3 - Import the data from your CanReg4 installation

- f) Untick “Correct Unknown”
  - g) Click “write data to file” and pick a file name that you can find back easily in CanReg5. For example on the desktop. Click “save”.
  - h) Take a look at the data you have now exported and close CanReg4.
  - i) Take a note of the number of records. (This should later match the number of *Tumour records* in your CanReg5 system.)
2. Back in CanReg5 do “File”, “Data Entry” and “Import Data”.
  3. Click “Browse” and locate the file from step A. Select it and click “Open”. You can if you want preview the file to see that you picked the right one and that the file looks OK. If for example Arabic names are garbled you should try to choose another “File encoding” (Default for Arabic text is ISO-8859-6).
  4. Set “Separating character” to Comma. (Or whatever separating character your file has.)
  5. Click Preview to see that the data looks OK.
  6. Click “Next” (or select the tab “Associate Variables”)
  7. This lets you associate the variables in the file to import with the variables in the database. CanReg5 will find most of these associations by itself, but you should revise them to see if they look OK. Look for variable names in bold, as they are the one that are not assigned at all.
  8. Click “Next” (or select the tab “Import File”)
  9. Click “Import” (leave everything as by default – the import function only works on empty CanReg databases as per now...)
  10. Let CanReg5 import the data (this might take a while) and click “OK”.
  11. Click “Browse/Edit” and “Refresh Table” to see that the data has arrived well. *Verify that you have the same number of tumours in the tumour table as in CanReg4.* Double click a record to take a look at it.



# D. Advanced database management

This appendix covers some advanced database management topics

## D.1. The System definition file (the XML file)

### D.1.1. Advanced options in the CanReg5 person search

You have access to a couple of extra settings in the CanReg5 system definition file to improve the detection of duplicate patients (in addition to *weight*), namely *discriminatory power*, *reliability*, and *presence*.

#### D.1.1.1. Weight

The weight of the link. These can add up to more/less than 100 (but will then be normalized to 100). The weight is the score that the link will get if there is a 100% match on that variable. Some algorithms also return numbers between 0 and a 100, and if so, the score is adjusted.

#### D.1.1.2. Discriminatory power

Does nothing at the moment!

(Discriminatory Power calculator: [http://insilico.ehu.es/mini\\_tools/discriminatory\\_power/index.php](http://insilico.ehu.es/mini_tools/discriminatory_power/index.php))

#### D.1.1.3. Reliability

Does nothing at the moment!

How reliable is the data. Scale from 0 to 1.

#### D.1.1.4. Presence

Does nothing at the moment!

How often is the variable filled. Scale from 0 to 1.

#### D.1.1.5. Threshold

Set between 0 and 100. Cases with higher match percentage than this shows up in person search.



## E. Technical details on the Table Builder

The graphical tool to build tables in CanReg5, the Table Builder, can be modified to suit particular needs of users via configuration files. CanReg comes with some default ones, but these can be added to as the user sees fit.

### E.1. Configuration files

The default configuration files of these tables can be found in the “conf/tables” folder in your CanReg5 installation. The user provided ones are found in the “conf/tables” folder in the “.CanRegClient” folder. For each line in the Table Builder you will find a corresponding .conf-file. (See figure [E.1](#))

The user can change or add new configuration files and they will appear in this list.

#### E.1.1. The file structure and parameters

The easiest way to familiarize oneself with the structure of these configuration files is to open one and take a look. The first thing we notice is that you have several parameters and labels grouped by curly brackets. The entry “*table\_label*” for example contains one entry which is the label, or name, of the table. This is what will be displayed in the list in the table builder. Let’s go through the standard groups one by one:

##### E.1.1.1. *table\_label*

The name or label of the table.

##### E.1.1.2. *table\_description*

The description of the table - more information that the user will see when he selects the table in the list.

##### E.1.1.3. *table\_engine*

This is the name of the engine that will produce the table when the user click *Generate* in CanReg5. More on this in the next sub section.

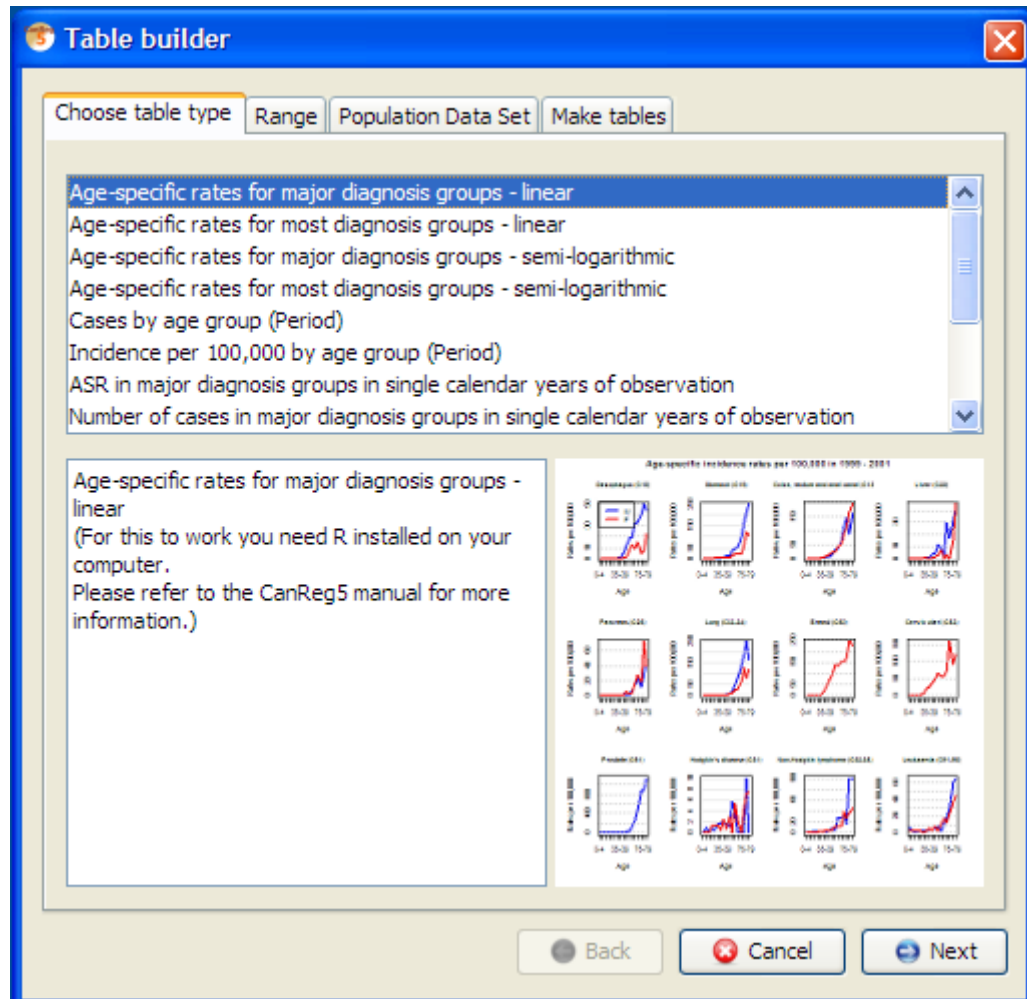


Figure E.1.: The Table Builder



**E.1.1.4. engine\_parameters**

These are parameters to the different engines, and can be anything the engine understand. For example a code saying “noC44” to tell the engine that this should be counts excluding skin cancer - if the engine supports that parameter.

**E.1.1.5. preview\_image**

This is the name of a file found in one of the previews folders (within the “conf/tables” folders mentioned above) that will displayed in CanReg when the user has selected this table. The user provided previews override the system provided ones.

**E.1.1.6. ICD\_groups\_labels**

Many tables also need a list of ICD group labels to display. This is just a list of labels with the 3 first characters detailing if this should be displayed for male/female cases. For example the line:

```
"11 Lip"
```

Means that the name of this site/group of sites is Lip and it should be displayed for both male and female.

```
"01 Ovary"
```

on the other hand means that this should only be displayed for female cases.

**E.1.1.7. ICD10\_groups (or ICD\_groups)**

Each of the labels from ICD\_groups\_labels must have a corresponding definition detailing the actual ICD10 codes that goes into this group. This can be found in ICD10\_groups. Line number 1 in ICD\_groups\_labels corresponds to line 1 in ICD10\_groups, line 2 to line 2 etc. These groups are parsed and interpreted run-time by CanReg so you can define your own groups by following the following rules:

1. For a single site enter the ICD10 code for that site, i.e. “C09”.
2. For a single subsite enter the ICD10 code for that site, i.e. “C000”.
3. For a range enter it as <site>-<site without the leading 'C'>, i.e. "C30-31".
4. For a series of sites enter them like <site or range>, <site or range> etc., i.e. "C47,C49" or "C82-85,C96"
5. There are some special sites that can be defined by using their standard names. These can not be part of any range or series:
  - a) "MPD" - Myeloproliferative disorders

- b) "MDS" - Myelodysplastic syndromes
- c) "O&U" - Other and unspecified
- d) "ALL" - All sites
- e) "ALLbC44" - All sites but C44

## **E.2. Engines**

CanReg5 has an ever increasing number of built in engines to produce tables. Their names:

- “incidencerates”
- “numberofcases”
- “populationpyramids”
- “top10piechart”
- “r-engine”
- "r-engine-grouped"

### **E.2.1. Incidence Rates Table(“incidencerates”)**

This is the engine that produces a A4 ready to print overview of number of cases per 100.000 per site in the database. (See figure [6.6 on page 73](#).)

#### **E.2.1.1. line\_breaks**

One special parameter in the configuration file of this table is the line\_breaks. It is a list of pairs of numbers detailing where to put grey boxes behind the numbers to make them easier to read. For example

```
"8,9"  
"17,-1"  
"21,1"
```

This means that at line 8 we start a 9 line big box. Stopping at line 17. Picks up again on line 21 for just one line.

### **E.2.2. Number of Cases (“number of cases”)**

This is the engine that produces a A4 ready to print overview of number of cases per site in the database.

#### E.2.2.1. line\_breaks

This has the same special parameter `line_breaks` as the “incidencerates” engine.

### E.2.3. Population Pyramids (“populationpyramids”)

This is the simplest configuration file so far as it looks like this:

```

table_label {
    "Population Pyramids"
}

table_description {
    "Population Pyramids"
}

table_engine {
    "populationpyramids"
}

preview_image {
    "populationpyramid.png"
}

```

### E.2.4. Top 10 pie chart (“top10piechart”)

Here you place your “candidate” groups in the `ICD10_groups` and `ICD_groups_labels` in the .conf file and the engine produces a pie chart of the 10 most common cancers in your registry for the period you have provided in either PNGs or SVGs. (See [E.2 on page 120](#))

If you pass “noC44” as one of the engine parameters you get tables excluding skin cancer.

### E.2.5. The R engines (“r-engine” and “r-engine-grouped”)

You can call R programs directly from within the TableBuilder of CanReg5. This is detailed in [E.3](#).

## E.3. CanReg and R

### E.3.1. Configuration file parameters

Like with the other Table Builder engines of CanReg you can configure the R engines as well using configuration files. Specific to the R engines are the following parameters:

### E.3.1.1. `r_scripts`

This can be one or many R scripts that gets called by CanReg to produce tables. (See the listing on page 116 for a skeleton script.)

### E.3.1.2. `file_types_generated`

This details what filetypes this R script can generate. For example:

```
file_types_generated {  
  "pdf"  
  "ps"  
  "svg"  
  "html"  
}
```

Listing E.1: A sample main R script (See the folder `conf/tables/r` for more info)

```
Args <- commandArgs(TRUE)  
## Find directory of script  
initial.options <- commandArgs(trailingOnly = FALSE)  
file.arg.name <- "--file="  
script.name <- sub(file.arg.name, "",  
initial.options[grepl(file.arg.name, initial.options)])  
script.basename <- dirname(script.name)  
  
source(paste(sep="/", script.basename, "makeTable.R"))  
source(paste(sep="/", script.basename, "checkArgs.R"))  
source(paste(sep="/", script.basename, "makeFile.R"))  
  
## outFile out <- checkArgs(Args, "--out")  
fileType <- checkArgs(Args, "--ft")  
  
# Load dependencies  
source(paste(sep="/", script.basename, "myLibrary1.R"))  
...  
source(paste(sep="/", script.basename, "myLibraryN.r"))  
  
# The incidence file  
fileInc <- checkArgs(Args, "--inc")  
dataInc <- read.table(fileInc, header=TRUE)  
  
# The population file  
filePop <- checkArgs(Args, "--pop")  
dataPop <- read.table(filePop, header=TRUE)
```

```
#do your plot
generatedFileNames = createMyNiceFigure(dataInc , dataPop)
dev.off()

# write the name of any file created by R to out
cat(generatedFileNames)
```

### E.3.2. Arguments

When calling the R scripts CanReg provides the following arguments:

- *-ft=<filetype>* The file type - either pdf, png, wmf, ps, html or svg.
  - This is how CanReg gives information to the R program on what file(s) should be created.
- *-out=<filename>* The base of the report file name - or output file name if you want.
  - The R program generates either a file with that name - with added standard file type suffix, or a series of file names with this as the base. One important point is that the way CanReg detects what files that actually has been created is by listening to the standard output of R. All the R program has to do is to write (print or cat) the files generated to standard out and CanReg will try to do system calls to display them when they are done.
- *-pop=<filename>* Population file name.
- *-inc=<filename>* Incidence file name.

Some scripts can also take additional arguments like:

- *-logr* Turn logarithmic plots on.
- *-onePage=axb* Generate summary page(s) with all the plots laid out in a columns and b rows.

Example call:

```
"C:\Program Files\R\R-2.12.2\bin\R.exe"
--slave --file=
"C:\Users\ervikm\Desktop\CanReg\.\conf\tables\r\test.r"
--args -ft=svg -out="C:\Documents and Settings\ervikm\Desktop\test"
-pop="C:\Users\ervikm\Temp\pop4834689295162083243.tsv"
-inc="C:\Users\ervikm\Temp\inc4297000362541616146.tsv"
-onePage=3x4
```

(The user never sees this, but it can be useful to keep this in mind when writing your own R scripts.)

### **E.3.3. Population file**

CanReg writes a population file to a temporary file with the following columns:

1. YEAR
2. AGE\_GROUP\_LABEL
3. AGE\_GROUP
4. SEX
5. COUNT

### **E.3.4. Incidence file**

CanReg writes an incidence file to a temporary file with a structure depending on the script and the engine.

#### **E.3.4.1. r-engine**

If you use the r-engine you can specify the columns you need from CanReg with the `variables_needed` configuration file parameter. For example:

```
variables_needed {  
    "Age"  
    "ICD10"  
    "Sex"  
}
```

This will create an incidence file with the following columns:

1. YEAR (added automatically)
2. AGE
3. ICD10
4. SEX
5. CASES (added automatically)

#### **E.3.4.2. r-engine-grouped**

If you use the grouped version of the r-engine CanReg will group the cancer cases according to the groups defined in the configuration file and create an incidence file with the following columns:

1. YEAR
2. ICD10GROUP
3. ICD10GROUPLABEL
4. SEX
5. AGEGROUP
6. MORPHOLOGY
7. BEHAVIOUR
8. BASIS
9. CASES

#### **E.3.5. SVG support**

To be able to generate SVG files in the R programs you need to install a package called “RSvgDevice” from within R. To install this package you need to launch R and type in the following command.

```
install.packages("RSvgDevice")
```

Then select a country (mirror) close to you and click OK. This will install the package and it is ready for R to use. You can now close R and if all went well start producing SVG figures calling R functions from CanReg.

(In the latest version of the R programs distributed with CanReg there's built in functionality to automatically install this package if you are connected to the internet.)

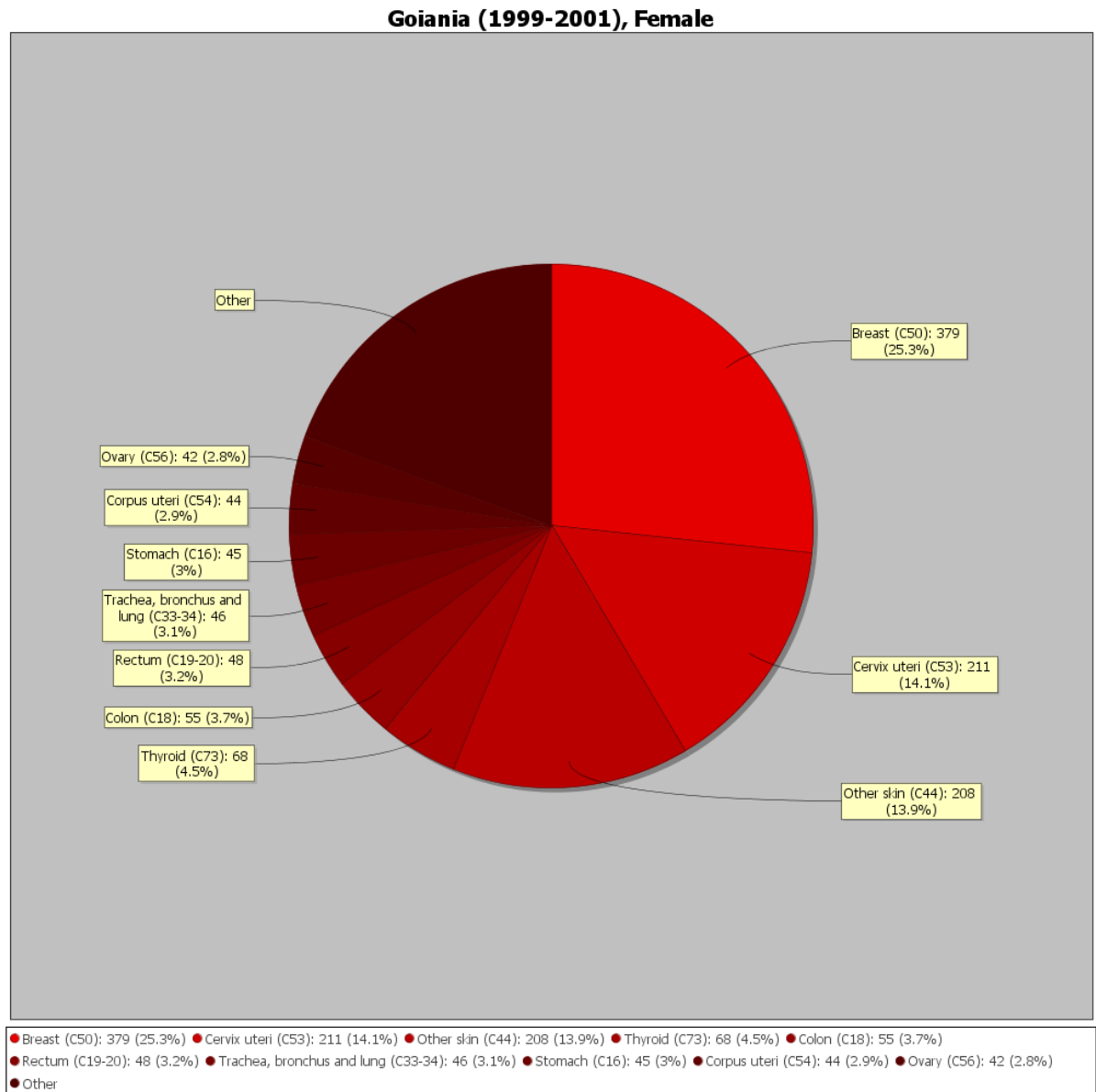


Figure E.2.: Pie Chart



## F. Third party tools

Here's a list of useful free third-party tools to use with CanReg5.

### F.1. Java Runtime Environment

To run CanReg at all you need a Java Runtime Environment. You can get it from <http://java.com/en/download/manual.jsp>.

#### F.1.1. Issues with memory

By default most JVM implementations limit themselves to using a small amount of RAM (eg: 64MB), even if your computer has a lot of system RAM (eg: 1GB). This can lead to out of memory conditions, even if your computer has lots of free memory. To configure your JVM to use more memory, complete the following steps:

##### F.1.1.1. On Windows:

Go to Control Panel | Java , in 'Java' panel, click view Java applet running settings. Set Java Running Parameters to:

```
-Xmx256M -Xms256M -Xss128k
```

Save and apply these changes. You must close and restart CanReg (and potential web browsers open) for these changes to take effect.

These settings require 256MB of contiguous memory be available. You may need to restart your computer before these settings will work.

##### F.1.1.2. On Solaris/Linux:

```
cd /bin
./ControlPanel
```

Change Java Applet running Parameters to: -Xmx256M -Xms256M -Xss128k

Press OK, restart CanReg (and potential web browsers open) for these changes to take effect.

This will configure your JVM to use a fixed 256MB of RAM.

If you have multiple versions of java on same machine, please perform the above operations for all versions.

## **F.2. Analysis**

### **F.2.1. R**

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.

More information about it on: <http://www.r-project.org/>

To download R go to <http://cran.r-project.org/mirrors.html> and choose a mirror near you, for example <http://cran.univ-lyon1.fr/>. Choose your platform, i.e. Windows, click the link that says “base”, i.e. <http://cran.univ-lyon1.fr/bin/windows/base/> and then click “Download R...”. The latest as of the 18th of October 2016 is version 3.3.1 and you can get more information on the installation procedure here: <http://cran.r-project.org/src/base/NEWS.html>.

### **F.2.2. Epi Info**

From the Epi Info website:

“Physicians, nurses, epidemiologists, and other public health workers lacking a background in information technology often have a need for simple tools that allow the rapid creation of data collection instruments and data analysis, visualization, and reporting using epidemiologic methods. Epi Info™, a suite of lightweight software tools, delivers core ad-hoc epidemiologic functionality without the complexity or expense of large, enterprise applications.”

More information about it on: <http://www.cdc.gov/epiinfo/>

The newest version of EpiInfo is windows based, but CDC also provides a DOS based version, Epi Info 6. (Available here: [http://www.cdc.gov/epiinfo/html/ei6\\_downloads.htm](http://www.cdc.gov/epiinfo/html/ei6_downloads.htm).)

### **F.2.3. SEER\*Stat and SEER\*Prep**

From the SEER website:

“The SEER\*Stat statistical software provides a convenient, intuitive mechanism for the analysis of SEER and other cancer-related databases. It is a powerful PC tool to view individual cancer records and to produce statistics for studying the impact of cancer on a population.”

To get your data from CanReg to SEER\*Stat you need to go via SEER\*Prep. (See [6.2.4 on page 74](#) for more information.)

### **F.2.4. PSPP**

“PSPP is a program for statistical analysis of sampled data. It is a Free replacement for the proprietary program SPSS, and appears very similar to it with a few exceptions.” Available from <http://www.gnu.org/software/pspp/>.

### **F.2.5. SOFA**

SOFA - Statistics Open For All The user-friendly, open-source statistics, analysis, and reporting package. Available from <http://www.sofastatistics.com/home.php>.

## **F.3. Data cleaning and record linkage**

### **F.3.1. Google Refine**

Google Refine is a free and open source tool that allows you to standardise and clean your data prior to importing it into CanReg5. Please note that even though you access and manipulate your data using a web browser your data is kept locally on your machine.

More information about it on: <http://code.google.com/p/google-refine/>

### **F.3.2. Talend Open Studio**

Talend Open Studio is a more heavy duty package to clean and standardise your data before importing it into CanReg. It is free and open source.

More information about it on: <http://www.talend.com/>

### **F.3.3. FRIL: Fine-Grained Records Integration and Linkage Tool**

From their website:

“FRIL is FREE open source tool that enables fast and easy record linkage. The tool extends traditional record linkage tools with a richer set of parameters. Users may systematically and iteratively explore the optimal combination of parameter values to enhance linking performance and accuracy. “

More information about it on: <http://fril.sourceforge.net/index.html>

### **F.3.4. LinkPlus**

From the CDC website:

“Link Plus is a probabilistic record linkage program developed at CDC’s Division of Cancer Prevention and Control in support of CDC’s National Program of Cancer Registries (NPCR). Link Plus is a record linkage tool for cancer registries. It is an easy-to-use, standalone application for Microsoft® Windows® that can run in two modes—

To detect duplicates in a cancer registry database. To link a cancer registry file with external files. Although originally designed to be used by cancer registries, the program can be used with any type of data in fixed width or delimited format. Used extensively across a diversity of research disciplines,

## *F. Third party tools*

Link Plus is rapidly becoming an essential linkage tool for researchers and organizations that maintain public health data.”

More information about it on:<http://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm>

## **F.4. Data security**

### **F.4.1. GnuPG**

- Complete and free implementation of the OpenPGP standard. (<http://gnupg.org/>)
- Allows to encrypt and sign your data and communication.
- Easy to use and set up.
- OS independant
  - Gpg4win on Windows <http://www.gpg4win.org>
  - GPGTools on Mac OS X
- Integrates with email clients
  - Enigmail

### **F.4.2. TrueCrypt**

- Free open-source disk encryption software.
- On-the-fly-encrypted volume (data storage device).
- For Windows 7/Vista/XP, Mac OS X, and Linux.
- <http://www.truecrypt.org>

## **F.5. Graphics display/editing**

### **F.5.1. PostScript Viewer**

Some of the tables generated by CanReg5 are PostScript files. PostScript is an open standard, so you can use many different tools to view them. (You can in many cases even send them directly to a printer.) Mac OS X and most Linux-distributions come with tools to view them by default.

On Windows, the tool I recommend is the open sourced and free GSview.

To run GS View you need to install Ghostscript first. This can be for example be downloaded from here: <http://pages.cs.wisc.edu/~ghost/doc/GPL/gpl900.htm> (Scroll all the way down, under the heading Microsoft windows and download the “GPL Ghostscript 9.00 for 32-bit Windows (the common variety)” (<http://mirror.cs.wisc.edu/pub/mirrors/ghost/GPL/gs900/gs900w32.exe>))

Run this file to install Ghostscript.

Then you can get GS View from here:

<http://pages.cs.wisc.edu/~ghost/gsview/get49.htm> (Most probably, you should pick the Win32 self extracting archive - the first download option.

<http://mirror.cs.wisc.edu/pub/mirrors/ghost/ghostgum/gsv49w32.exe>)

Run this file to install GS View.

## F.5.2. Inkscape

Inkscape is a free open source program to display and edit SVGs and other vector based files.

From their website:

“An Open Source vector graphics editor, with capabilities similar to Illustrator, CorelDraw, or Xara X, using the W3C standard Scalable Vector Graphics (SVG) file format.

Inkscape supports many advanced SVG features (markers, clones, alpha blending, etc.) and great care is taken in designing a streamlined interface. It is very easy to edit nodes, perform complex path operations, trace bitmaps and much more.”

Available from: <http://inkscape.org/>

## F.6. General

### F.6.1. Notepad++

Notepad++ is an excellent open source text editor that is perfect for writing code, cleaning a dataset using, for example regular expressions, or just simply looking at data from CanReg. Available from: <http://notepad-plus-plus.org/> (It has nothing to do with the Notepad that is distributed with Windows - it merely shares the first part of it’s name with it.)



# G. Scripts

## G.1. Ruby scripts

### G.1.1. Script to generate icd10 codes and cancer groups

This tool lets you add icd10 codes and user defined groups to data in a csv file.

#### G.1.1.1. Prerequisites

##### Install JRuby

1. Download the latest JRuby from here: <http://jruby.org/download> fitting your machine
2. Run the installer and follow on screen instructions. (Make sure you add jruby to the path when asked.)
3. Windows: Run the “install.bat” in the script folder. Linux/Mac: run the following commands:

```
jgem install bundler
jruby --1.9 -S bundle install
```

#### G.1.1.2. Usage

```
jruby --1.9 icd10grouper.rb [options]
```

##### Example:

```
jruby --1.9 icd10grouper.rb -t Topography -m Morphology -b Behaviour
-s Sex -c , -g Groups.conf -f LotsOfData.csv
```

##### Arguments:

```
-t TOPOGRAPHY_COLUMN_NAME, Name of the column containing topography.
Default: TOP --topography-column-name
-m MORPHOLOGY_COLUMN_NAME, Name of the column containing morphology.
Default: MOR --morphology-column-name
```

## *G. Scripts*

```
-b BEHAVIOUR_COLUMN_NAME, Name of the column containing behaviour.  
Default: BEH --behaviour-column-name  
-s SEX_COLUMN_NAME, Name of the column containing sex.  
(Coded: 1-Male, 2-Female) Default: SEX --sex-column-name  
-c SEPARATING_CHARACTER, Separating character in the files.  
Default: tab --separating-character  
-g GROUP_CONF_FILE, File containing the definitions of the groups.  
Default: Groups.conf --group-conf-file  
-p PATH_TO_CANREG, Path to the CanReg.jar-file.  
Default: ../../CanReg.jar --path-to-canreg  
-f, --in-file IN_FILE File to process.  
Default: "Here be Data.txt"  
-h, --help Display help functions.
```



# H. About

## H.1. Copyright

© 2008-2016 International Agency for Research on Cancer World Health Organization  
All rights reserved.

## H.2. Credits

### H.2.1. Lead developement and design:

- Morten Johannes Ervik, IARC/CSU

### H.2.2. Additional developement

- Andy Cooke (Original quality control module)
- Jacques Ferlay (Original PostScript code, conversion tables, and original quality control module)
- Anahita Rahimi (R code for analysis)
- Sebastien Antoni (R code for analysis)
- Hemant Dhivar (Migration Tool)

### H.2.3. Additional design

- Andy Cooke
- Maria Paula Curado
- Lydia Voti
- Freddie Bray
- David Forman
- Max Parkin

#### **H.2.4. Consultants:**

- Philippe Autier, IARC/BIO
- Hoda Anton-Culver, University of California Irvine, USA
- Joe Harford, NCI, USA
- Hai-Rim Shin, IARC/DEA
- John Young, Emory University, USA

#### **H.2.5. Technical consultants**

- Michel Smans, IARC/ITS
- Lucile Alteyrac, IARC/BIO

#### **H.2.6. Thanks**

- Kai Toedter (JCalendar)
- Jeremy Dickson (CachingTableAPI)
- Alexandre Moore (Icons)
- Brian Cole (PagingTableModel)
- Stephen Kelvin (XTableColumnModel)
- Ashok Banerjee and Jignesh Mehta (ExcelAdapter)
- Glen Smith et al (opencsv)
- Dem Pilafian (Bare Bones Browser Launch)
- Jesse Wilson (Glazed Lists)
- Frank Tang (juniversalchardet)
- Object Refinery Limited (JFreeChart)
- Chris Evans (jcommons)
- The Buzz Media (imgscalr – Java Image Scaling Library)
- The Apache Software Foundation (Batik)
- iText Software Corp (iText)

### **H.2.7. Alpha-testers**

- Mathieu Mazuir, IARC/DEP
- Antoine Buemi, Registre des Cancers du Haut-Rhin

### **H.2.8. Beta-testers**

- Xenios Anastassiades, Ministry of Health, Cyprus
- Deborah Bringman, University of California Irvine, USA
- Amr Ebeid, Gharbiah Cancer Registry, Egypt
- Ibrahim Abd-Elbar Seif Eldin, Gharbiah Cancer Registry, Egypt
- Sarah Marshall, University of California Irvine, USA
- Omar Nimri, Jordan Cancer Registry, Jordan
- M. Ramadan, Gharbiah Cancer Registry, Egypt
- Kevin Ward, Emory University, USA
- Cankut Yakut, Izmir Cancer Registry, Turkey
- Antoine Buemi, Registre des Cancers du Haut-Rhin, France
- Gonalo Forjaz, Azores Cancer Registry, Portugal

### **H.2.9. Translators**

- **French:** Joannie Lortet-Tieulent, IARC, France
- **Portuguese:** Edesio Martins, Goiania Cancer Registry, Brazil
- **Russian:** Evgeniya Ostroumova, IARC, France and Anton Ryzhov, Ukraine
- **Spanish:** Graciela Cristina Nicolas, Cordoba, Argentina
- **Chinese:** Yayun Dai, IARC, France and Shixuan Liu, China
- **Portuguese, Portugal:** Gonalo Lacerda, Azores, Portugal
- **Turkish:** Cankut Yakut, Izmir Cancer Registry, Turkey

## H.3. References

- J Ferlay et al, Check and Conversion Programs for Cancer Registries (IARC/IACR Tools for Cancer Registries), IARC Technical Report No. 42, Lyon, 2005 (Available at: <http://www.iacr.com.fr/TR42.htm>)
- O.M. Jensen et al, Cancer Registration: Principles and Methods, IARC Scientific Publication No. 95, Lyon, 1991 (Available online at <http://www.iarc.fr/en/publications/pdfs-online/epi/sp95/index.php>)
- E Steliarova-Foucher et al, Childhood Classification, Third Edition, CANCER Volume 103, 1457-1467
- K Fogel, Producing Open Source Software, First Edition, O'Reilly, 2005
- J Bloch, Effective Java, Second Edition, Sun Microsystems, 2008

## H.4. Acknowledgments

- CanReg1: Allen Bieber
- CanReg2: Stéphane Olivier
- CanReg3 and 4: Andy Cooke

## H.5. Licenses

- The RSSutils library is copyright 2003 Sun Microsystems, Inc. ALL RIGHTS RESERVED
- The Soundex library is copyright 1996-2002 Ian F. Darwin, <http://www.darwinsys.com/>, ALL RIGHTS RESERVED

# I. Changelog

## Changelog

### 5.00.40 BETA

- Date format for data entry can now be chosen at user level. (Under "Options".)
- Added detection of uninstallation/reinstallation/updates of R/GS software.
- Improved year selection in table builder.
- Updated R libraries and R script for installation of those.
- Character separated file extension now defaults to CSV even for tab separated files.
- Added another common population format for children ([0,5>,[5,10>,[10,15>).
- Fixed a bug in that hindered user setting number of cancers in top N cancer trends. (R tables configuration files.)
- Updated Spanish translation.
- Various fixes.

### 5.00.39

- Users can now perform "Exact search" while entering a new case.
- System definition files are now always saved in UTF-8.
- [ModifuDatabaseStructureInternalFrame] Save button now disabled while saving a file.
- Slight restructure of the main menus to make them more consistent.
  - Options are now always visible – even when not logged in.
  - Migrate from CR4 is a separate sub-menu under Tools.
  - Import under data entry always wants data in CR5 format, the CR4 import function is under Tools->Migrate from CR4.
- Enter now skips to next field during data entry.
- Fixed a bug in the filter wizard where it would not pick up the change of table when changing to a join of all three tables.
- Fixed a bug where the date was sometimes not saved if it was unparsable.
- Frequencies by year now does not allow selection of patients or sources alone.
- Range and filters now work again when selecting only Tumour table in frequencies by year.
- Fixed a bug where pie-charts and bar-charts would not be generated if any cases were recorded with unknown sex codes.

## *I. Changelog*

- Versioning system changed from Mercurial to Git.
- Better handling of Unicode characters in output files.
- Other improvements.

### 5.00.38

- Improved handling of non-standard database filenames.
- Improved error messages.
- Escaped character in export file header
- Other bugfixes.

### 5.00.37

- Fixed a bug in the Date of Last contact check introduced in the previous version.
- Fixed a bug where updates to Source records were sometimes not imported.
- Other minor fixes.

### 5.00.36

- Fixed a bug occurring in some tables if the decimal point was not '.' (ie. Russian locale.)
- Saving frequency tables now also produces pivot tables.
- Hid "Other" age group structure button.
- Unknown age code is now dynamic.
- DLC check now with improved handling of unknown dates.
- Allow a lenience for a date being prior to another date in the case of unknown dates.
- DateHelper function "years between" two dates now takes better into account unknown dates.
- Other minor fixes.

### 5.00.35

- Fixed a bug in the import of datasets split in three tables where sometimes not all cases were imported.

### 5.00.34

- Fixed (Temporarily) the Age-Specific Time Trends script. (It now computes rates/trends for age groups  $\geq 40$  and for all sites together only)
- Fixed a bug where some non-Latin characters were not exported correctly, but showed up as question-marks. Now properly implemented as unicode (UTF-8).
- Fixed a bug related to import of patients with multiple tumours.
- Turkish translation updated.
- Other minor fixes.

### 5.00.33

- Corrected the Age Specific Rates script (was generating fake values, used for graphs, when exporting as CSV)
- Removes the double extension from output files if HTML is selected (e.g. test.html.html)
- Added SVG saving option to the Top Cancers Barchart script
- Improved Population Pyramid script
- Added some more default age group structures. (Cut off at 70 and 60.)
- Added Beta Version of Time Trends (Age Specific) + Other functions
- Temporarily removed values in barchart (bug)
- Other minor fixes

#### 5.00.32

- Time trends added.
- CI5editorial tables added as an analysis option.
- AgeSpecificRatesTopX table now sorts cancers by ASR.
- Population pyramid improved.
- ASR Top Cancers bar chart added.
- Fixed a bug where ASR was by 1.000.000 instead of 100.000 in the charts of top 10 cancers.
- Fixed a bug where R would not be called in the background on Linux/OSX.
- Improved the handling of obsolete cases in analysis.
- Improved the handling of pending/deleted/duplicate cases in analysis.
- All sites but skin added to quality indicators.
- Added truncation by age groups.
- Changed the way external programs are called. (Now calling with an array of strings instead of calling with just a string.) This solves problems on Mac and Linux, and improves security on Windows.
- Other updates to the R scripts/analytical functions.
- Updated the code for barcharts and piecharts to take into account upcoming changes to ggplot.
- Russian translation updated.
- Other minor fixes.

#### 5.00.31

- Age specific rates for top X cancer sites figure added (SA).
- Data Quality table added (SA).
- Updated other R scripts.
- Added functionality to distribute R packages for CanReg5.
- Table Header and Table Label now passed as arguments to the R scripts.
- Improved detection of lines in output from R.
- R now only installs needed dependencies – not recommended ones.

## *I. Changelog*

- Exportizer path added to options.
- Other minor fixes.

### 5.00.30

- Fixed a bug where some times a record would be locked and inaccessible after creating a new record specifying the ID.
- Cleaned some code.
- Other minor fixes.

### 5.00.29

- Improved handling of patients missing tumour records.
- Improved functionality for user created table definitions and R scripts.
- Improved readability of some text on screen.
- Other minor fixes.

### 5.00.28

- Improved handling of separating characters within strings while exporting data.
- Reference population now displayed on Age Specific Rates per 100.000 table.
- Added a new age group structure.
- Changed an inconsistency in the way that age group structures were displayed.
- Other minor fixes.

### 5.00.27

- The latest Java update (7u21 and 6u45) broke CanReg5 integration of R and GS due to a new restriction in spaces in file names. This is a quick work around that solves the problem on Windows machines, in most languages (including Spanish, Portuguese, and English).
- Fixed a bug in the testScript.r where generate SVG or PNG did not work.

### 5.00.26

- CanRegDAO should now be more thread safe.
- Improved the connection to SEER\*stat.
- Updated some R scripts.
- When restoring from backup, CanReg5 now always renames existing folders and files instead of overwriting them.
- Added Lastcontact and Vital Status to the SEERprep.conf file.
- Handbook: Updated to a more compact book layout.
- Handbook: Fixed some lines/graphics that went into the margins.
- "Latest news" now working again. (The down time was due to a change in Twitters API.)



- Other minor fixes.

#### 5.00.25

- R script generated by CanReg5 now functions like the R scripts in the script folder by cat'ing out the file names generated. CanReg then reads from a report file instead of STOUT to decide on files generated by R.
- Improved the "makeSureGgplot2IsInstalled.R" script by changing the repo to a dynamic one and the installation of packages to local folder more explicit.
- Made the export window more dynamic.
- Fixed a bug in the range filter panel occuring when the list of previous filters was missing.
- Started migrating away from the Swing Application Framework.
- Updated many libraries.
- Other fixes.

#### 5.00.24

- Fixed an error in ggplot pie charts occuring if any category had a 0 count.
- Updated: The latest ggplot2 no longer allows options(title=""), but needs labs(title="").
- Improved error handling in the population dataset editor.
- Modify Database Structure: Short name of new variables no longer default to "Default name" and there is now a check to see if they are empty before saving.
- Fixed a bug occuring when selecting no standard variable.
- Turkish translation partially done.
- Other fixes

#### 5.00.23

- Fixed bug occuring when top or morph did not have the appropriate length in the ICD03 to 10 module.
- Fixed a bug in the import function where the character set of the imported files were sometimes not taken into consideration.
- Fixed a bug in the person search where weights of the links were not taken into account.
- Fixed a bug occurring when a short name of a variable was a registered word.
- Other bug fixes.

#### 5.00.22

- Fixed a bug where the browser didn't work without a filter/range. (Added too many parentheses in last update.)
- New tables added. Previews updated.

## I. Changelog

- "Top10" and "Cases by age group" charts can now be generated using R (with ggplot).
- Generators for R and JChart implemented plus helper functions.
- A builder for R files added.
- "makeSureGgplot2IsInstalled" script added.
- Added some static paths to the globals.
- Added chart-type, file-type, count-type, and sex-labels to table builder interface.
- Other bug fixes.

### 5.00.21

- Cases by age group Pie and Bar charts added.
- Top 10 cancer sites can now also be represented as bar a chart.
- Added a write to PDF function to the built in canreg5 chart viewer.
- PDSEditor: "Save as new" button added as suggested by Max Parkin.
- If user enter name of PDS in standard format (Name, Year), the date is automatically filled/suggested.
- Removed the rates from the Cases table.
- Legends are now optional in charts – and definable in the .conf file associated with the table.
- The user can now copy the data behind the charts in the chartviewer to clipboard, ready to be pasted into general spread sheets, ie excel.
- Charts can now directly be written as csv files.
- Added some parentheses to account for ORs in filters.
- Fixed a bug in the top 10 cancers tables where sites with the same numbers of cases led to dropped data.
- AgeSpecificCases tables no longer counts all D's into Others & Unspecified.
- Upgraded JFreeChart to 1.00.14 and JCommons to 1.00.17.
- Upgraded OpenCSV to version 2.3.
- Improved the colours of the top N Bar Chart.
- Updated the top10bar preview image.
- Updated the counting and grouping algorithm of ICD10 codes.
- Internationalized some strings.

### 5.00.20

- The display now refreshes after removing a variable in the tool to edit the database structure.
- Export now takes into account UTF-8 characters.
- Some tables, previously only available as PS, now also available as PDFs (if GhostScript is installed).
- Data entry forms can now be written as a PDF.
- Improved detection of R-path and added GhostScript-path.
- Fixed a problem with blancs in the path of PS tables.

- Other fixes.

#### 5.00.19

- Fixed a potential null-pointer in the table builder if people wanted to use the filter wizard.
- Added JRuby helper scripts to use the CanReg5 conversions and grouping functionality on data not from CanReg. (See handbook.)
- Table builder now shows a warning if trying to build a table with no data.
- Fixed a name error in the changelog. Gonçalo Lacerda translated CanReg5 to Portuguese from Portugal.
- ICD10 codes will be generated even when checks are not performed, potential null pointer error if no name field was defined fixed.
- Updated the handbook.

#### 5.00.18

- Fixed a bug in the pie chart builder where cancer groups with the same count of cases only counted once.
- Fixed a potential null-pointer error while accessing the lock-file.
- Updated the docs.
- Various bug fixes.

#### 5.00.17

- Implemented a system to detect records not properly released from the server after an unclean break.
- Tablebuilder can now build tables without denominators.
- Users can now change the font and font size used in CanReg5.
- Tablebuilder automatically guesses population datasets to use in analysis based on the date of the stored sets.
- Database generator: Added user role column to the users table.
- Migrator: Added a migrator to add user roles.
- A Grade Field variable can now be automatically filled from the 6th digit in the morphology part of the ICD-O-3 codes, like Behaviour.
- Improved handling of null pointers in the system description editor.
- Fixed a bug where CanReg didn't launch if the working dir entry in the settings file was missing.
- Fixed a bug related to country specific locales.
- Fixed bugs in the population dataset system.
- Various bug fixes

#### 5.00.16

- Portuguese from Portugal added as language. (Thanks Gonçalo Lacerda)

## *I. Changelog*

- Language chooser in Option Frame now only displays translated languages.
- Fixed a potential null pointer error in the Population Datasets.
- DerbyDB engine updated to version 10.8.
- Tidied some old properties files.

### 5.00.15

- Fixed a nullpointer error in the rangefilter panel if the actionlistener had not been set.

### 5.00.14

- Fixed an import related bug where sometimes cases were missed during import!
- The browser sorts by default by Incidence Date/PatientID/SourceRecordID depending on tables shown.
- Updated the doc.
- Other minor fixes.

### 5.00.13

- Added functionality in the system definition editor to detect running databases and changes that impacts on the database structure and warns the user about that.
- Reverted back to the old way of accessing the twitter rss feed.
- Logo updated.
- Updated documentation and roadmap.
- Other bug fixes.

### 5.00.12

- Tables for "age specific cases per 100.000" and "age specific cases" can now be written to a character separated file and opened in general spreadsheets for further work.
- Duplicate entries no longer show up in the list of favourite servers.
- Added automatic verification of short name of variables (database variable name) in the system editor.
- Better feedback while logging in. A progress bar shows up if things take long.
- Implemented a changelog-viewer to see what's new in the current version of CanReg.
- Updated handbook.
- Updated translations.
- Various fixes.

### 5.00.11

- Fixed a bug in the topography/morphology check where too many rare cases were detected.

- Fixed a bug in the table builder where text instead of numbers in age, sex, morphology, topography etc would break the table instead of give a warning.
- Added a warning message when using the "Overwrite" option during import.
- Fixed a bug where csvreader was attempted closed even when null.
- Fixed a bug where windows programs could't detect the line endings of exported files (case listings/dictionaries).
- Other minor fixes.

#### 5.00.10

- Implemented export facilities to get data in a fixed width file following the NAACR 1946 v11.3 format from CanReg to SEER\*prep and then on to SEER\*stat.
- System editor GUI is now using tabs to split up variables, dictionaries, groups, etc...
- Improved scroll speed in system editor.
- Fixed bug in system editor where you needed to click several times on arrows if there were hidden variables involved.
- Fixed a bug where one couldn't add new variables after having removed some.
- After adding elements to the database structure editor the relevant editor opens.
- Replaced the old Twitter RSS URL with a more generic one using the Twitter API.
- Updated the French translation.
- Optimized some code.
- Other bugfixes.

#### 5.00.09

- Fixed a bug where the last character of some lines in the export file went missing.
- Fixed a bug where user couldn't save dictionary when a it only contained one entry.
- Fixed a bug where no error message was displayed when records were locked and tried to open.
- Removed warning message if no encoding is detected for dictionary during import.
- Removed most generic exceptions and replaced them with specific ones.
- Other bug fixes

#### 5.00.08

- Updated topography and morphology check.
- Patients comparator implemented.
- Internationalized pop up menus.

## *I. Changelog*

- Added functionality to convert CanReg4 system definitions in batch mode using `—convert <.def-file> [encoding]` as arguments.
- Updated French translation.
- Updated Chinese translation.
- Updated Spanish translation.
- Handbook is now using the book-layout (instead of article).
- Updated about.html
- Other small fixes.

### 5.00.07

- The database can now be encrypted with 56-bit DES using a minimum 8 character long boot password. The user must then provide this password during every server launch.
- All server calls now handles server disconnects and requests users to log in again. This should end problems on laptops falling asleep.
- The CanReg server can now be launched in single user mode without RMI (network calls).
- Launch server no longer hangs if XML contains an invalid standard variable name.
- More info shown about the database elements during migration/tailoring.
- Person Search and duplicate search renamed to Duplicate Patient Search.
- New source added to tumour by default on creation of new tumour.
- PatientID shows up as a title of the recorder when the patient has been assigned an ID.
- Improved the preview while importing data. Now properly supports other separating characters. No longer editable.
- Fixed a bug in the logout mechanism that would not redraw the desktop after logout.
- Fixed a bug in the dictionary editor where a series of only codes – no labels were accepted, but not added to the database.
- Improved handling of already running servers while launching a new one.
- More consistent layout in the record editor.
- R table builder allows for null as pops or incs. Better handling of nullpointers.
- Updated the R test script.
- Better exception handling.
- Chinese translation started.
- Other bug fixes.

### 5.00.06

- Implemented a table builder that calls R with any user specified standard variables.

- AgeSpecific incidence curves (linear and semi log) functionality implemented using R. (Thanks to Anahita Rahimi.)
- TableBuilder: user can now write many different file formats depending on what the various table builders support, PDF, PS, SVG, PNG, WMF, HTML etc.
- CanReg chart viewer implemented. Tables supporting this can be previewed directly in CanReg.
- You can now join all 3 tables in the browse/export/frequency tools
- .
- Improved error messages when filters are incorrect.
- Range can now be formed by any variables that is included in an index.
- Added code to migrate the database to 5.00.06 - add foreign keys etc. to speed up 3 way join
- Fixed some bugs in the population pyramid where totals showed up as 0s and the population name contained the name of one year of the population data set.
- Fixed a bug where a resultset was not closed properly
- Added functionality to create indexes and keys in a database.
- Added more variables to the import options.
- Implemented a simple pie chart of 10 most common cancers.
- Implemented a system to copy graphics from (and to) CanReg.
- PopPyramid now allows editing of the chart and printing using the ChartPanel from JFreeChart with my added SVG writer using Batik.
- PDS editor now displays male blue and women red.
- FastFilter now clears the filter if user changes dictionary.
- TableBuilder: fixed a bug where pending cases would show up in some of the tables, improved the performance of the filter.
- TableBuilderInternalFrame can now call HTML writers.
- Check Topo/morpho no longer breaks down if Morpho don't have a 4 digit code, but rather returns an error message.
- Password now kept as char array through the entire login process for security purposes.
- Using stringbuilders in CanRegDAO.
- Special characters no longer show up in ICD10 codes.
- Fixed a bug where comments in the ICD03to10 lookup file caused problems.
- Path to R installation added to the Options Pane.
- Implemented automatic detection of (one of the) user's R installations.
- View work files now uses platform independent system calls to open the folder view.
- CheckResult.Missing not displayed.
- Improved the handling of deletion of source records.
- Open backup folder now uses cross platform system calls.

## *I. Changelog*

- Filter is now cleared in the dictionary element chooser after a selection has been made.
- Mouse pointer also returns to normal if you view the charts in the built in chartviewer.
- `common.Tools`: better handling of null pointers.
- Created a `TableBuilderFactory` to encapsulate the definitions of the various table builders.
- Refactored and tidied some code.
- Other bug fixes

### 5.00.05

- Translated to Spanish by Graciela Cristina Nicolas.
- Implemented Topography/Morphology check.
- Fixed a memory leak during export.
- Install new system definition frame now detects backups in the same folder as the XML to streamline the initial installation process.
- Standard dictionaries are now filled with standard codes when the database is created.
- If you start CanReg with the registry code as argument it launches only this server – not the client.
- Updated the Age/Morphology, Age/Topography, Grade checks.
- Fixed a bug where dates would not be reported as missing although flagged as mandatory variables.
- Implemented system to request focus after pop up menu.
- The user can now press '?' to get the dictionary chooser.
- Browse and openFile updated. Now using `java.awt.Desktop` if possible – falling back on `BareBonesBrowserLaunch` if necessary.
- Updated the `BareBonesBrowserLaunch` class.
- The panels are now using the interfaces instead of implementations.
- Added a tray icon to show that the CanReg server is running.
- A system for shutting down the server properly put in place.
- `LoginInternalFrame`: the Launch server button gets reactivated if you modify the server code.
- System Tray notifications and popups implemented.
- Shows login frame after successfully installed system definition.
- Internationalized the splash screen messages.
- Updated the demo system, `TRN.xml`
- Javadoc expanded.
- Added some protection from nullpointers.
- Added some tooltips.
- Various fixes.

### 5.00.04

- Fixed the "dropped result set while browsing" bug



- Population data set editor improved.
- Added pyramids directly in the editor for immediate feedback.
- Population Pyramids in the PDS editor can now be saved as PNGs.
- Copy and paste menu for the population data set implemented.
- Improved layout of Export/report frame.
- Improved the layout of the import screen. (Added a scrollbar.)
- Registrar can no longer import files.
- Copy and paste menu for (most) text fields implemented.
- Fixed bug in system description affecting text areas.
- CanReg launch4j project created to facilitate launch on Windows machines.
- Started refactoring and updating tables and table builders.
- Refactored the cachingtableapi out of the main canreg-tree.
- Made sure old resultsets are properly dropped.
- Import complete dictionary no longer shows message as error but warning when no encoding is detected.
- The list of Population Data Sets are now updated in real time if entries are added/updated or deleted.
- Export of sources attached to a tumour table is now (properly) implemented.
- Sources' variable names are now numbered if more than 2.
- Integrated postscript-viewer test.
- TextArea of backupframe no longer editable.
- Tidied some exception handling.
- Tweaked the build.xml.
- Implemented a calculate age conversion.
- Converter and checker now only depends on the standardvariablenames.
- Added code to select a specific data element from the variableschooserpanel.
- Comments added.
- Varions fixes.

#### 5.00.03

- Turkish bug fixed. Changed all calls to toUpperCase() to a standardized static toUpperCaseStandardize() located in the Tool class. Default upper case and lower case locale set to ENGLISH.
- Merged the handbook and the manual into one PDF that can be updated independent of the CanReg releases.
- Frequencies by Year table can now be written to CSV file.
- Improved the layout of the ExportFrame.
- Export/report and Frequencies by year and now appends the .csv/.txt if the user does not specify this.
- DictionaryEntry can now be added to a tree to be sorted by either code or description.

## *I. Changelog*

- The dictionary chooser put in place. Users can now sort dictionary codes by either description or code.
- Implemented a filter for the dictionary element chooser using the Glazed Lists library <http://sites.google.com/site/glazedlists/>
- DictionaryImporter: Fixed a bug that added a space to the label of dictioanries imported from CR4.
- GUI for the Index-editor implemented. Fixed an update-bug in the database structure editor.
- Fixed a bug where the range sometimes did not work when a join of two tables were accessed.
- Group name now shows up in group editor.
- Import: performance fixes and tidied some code.
- Fixed some potential null-pointer errors.
- Fixed some localization issue.
- Auto detection of file encoding now works.
- FastFilter now uses the new dictionary element chooser.
- Removed the cancel option from "do you want to close"...
- Logging more info if something goes wrong during login.
- Added an easy access list of tables.
- Added links to news items in the "latest news" browser.
- Fixed a bug in the conversion from ICD O 3 to ICD10 where no ICD10 would be generated for some rare morphologies.
- No longer displays patient record numbers but patent ids as results of the person search.
- Implemented the GUI to let the user select types of algorithms for each variable in the person search, like alpha, number and date as well as soundex.
- Improved the database structure editor.
- Implemented user selectable types of algorithms for each variable in the person search, like alpha, number and date as well as soundex. This can be stored in the system definition XML file.
- Implemented a better way to store the person searcher in an XML.
- Updated the about.html.
- Table builder and export/report now launches faster.
- More info button added to the welcome frame.
- Latest News menu option: Added functionality to read the CanReg Twitter/RSS feed directly from the program.
- Check to see if a standard variable is already mappe to a variable in the database during system setup/tailoring.
- DatabaseStructure editor now displays a warning message if minimum required variables are not in place.
- Improved the GUI of the database variable editor screen.
- Code: Added override annotations, replaced some printstacktraces with proper logging of errors, replaced vectors with lists
- Fixed a bug where the compound dictionaries did not detect faulty (truncated) codes.

- Variable names are sorted in the rangefilter and the fastfilter.
- Updated the welcome frame.
- Performance improvements.
- Updated the about box.

#### 5.00.02

- Fixed a bug when the standard variable is a string of 0 lenght.
- Tidied some code.
- Added a menu option to file bug/issue reports.
- Dictionary Editor: Now uses StringBuilder to improve performance and allow for editing of bigger dictionaries.
- Handbook: Updated FAQ

#### 5.00.01

- Database: fixed a bug where some filters didn't work when joining two (or more) tables.
- Import: handles better errors when one line does not have enough elements, the apache licenced csvreader now used to parse the infile.
- Database: fixed a memory leak issue, improved efficiency of import function, improved error handling



# Index

- .CanRegClient, 18
- .CanRegServer, 18
- 3 digit code, 81
- add user, 84
- age, 95
- age group structure, 55
- age standardised rates, 56
- analysis, 50, 65, 96
- analyst, 84
- ASR, 55
- backup, 81, 90
- bugs, 101
- build code, 12
- CanReg version, 91
- CanReg4, 11, 20, 33, 57, 95, 103, 105, 106
- canreg5client.log, 12
- changelog, 133
- character delimited, 59
- check, 49, 50
- checks, 51, 63
- clear the previous data, 96
- codepage, 97
- coding, 28
- coding scheme, 97
- comma separated file, 79
- comma separated variables, 59, 96
- command line, 29
- community site, 12
- conversion CanReg4 to CanReg5, 95
- correct unknown, 68
- create new record, 47
- cross checks, 51
- cross-tabulate, 77
- CSV, 79
- data cleaning, 123
- data entry, 43
- data security, 124
- database design, 97, 99
- database table, 97
- default password, 84
- delete user, 84
- demo system, 20
- dictionary, 23, 25, 27, 96, 105
- dictionary editor, 25, 52
- discrepancies, 62
- disk encryption, 124
- display variables, 47
- download, 12
- duplicate records, 50
- edit record, 47
- Epi Info, 122
- errors, 101
- errors during import, 96
- Excel, 79
- export data, 65
- export dictionary, 52
- export file format, 66
- FAQ, 95
- file encoding, 34
- filter, 44, 78
- filter wizard, 45
- firewall, 29
- folder containing backup, 81
- format date, 67
- forum, 12

## Index

- frequency distributions, 77
- frequently asked questions, 95
- Ghostscript, 96, 125
- GnuPG, 124
- Google Refine, 123
- group, 23, 25
- group editor, 25
- GSview, 96, 124
- ICD10, 70
- import, 59, 97
- import data, 106
- import data from CanReg4, 33, 106
- import data from other programs, 35
- import dictionaries, 31, 96
- import dictionary, 53, 105
- import population data set, 57
- import variable definitions, 103
- incidence tables, 68
- installation, 17
- IP address, 30
- issue tracker, 12
- Java Runtime Environment, 121
- Java Runtime Environment, 17
- JRE, 17
- language, 89
- latest version, 12, 91
- launch server, 95
- launch the CanReg server, 28
- limitations, 101
- local machine, 29
- logfile, 12
- logical operator, 44
- login**, 29
- login name, 83
- look & feel, 89
- manage users, 82
- management, 81
- mandatory variable, 50
- mandatory variables, 49
- migrate, 20, 103
- minimum data set, 97
- minimum match, 27
- modifying a CanReg system, 22
- name, 50, 64, 84, 98
- name on the network, 30
- navigate buttons, 47
- network, 30
- notification area, 29
- obsolete, 50
- only outline, 90
- operator, 44
- options, 89
- password, 83
- patient table, 97
- PDS, 54, 70
- permission level, 83
- permissions, 84
- person search, 27, 49, 51, 63, 86
- pink fields, 49
- population data set, 70
- population dataset, 54
- population dataset editor, 54
- population pyramid, 56
- postscript files, 124
- postscript files, 71, 96
- postscript viewer, 124
- PS files, 124
- quality control, 84
- R, 79, 115, 122
- range, 44
- rare cases, 83
- record Status, 49
- registrar, 83
- registry code, 23
- remote registration, 95
- restore from backup, 81
- run**, 19
- save record, 49
- save table, 79
- search variables, 86

SEER\*Prep, 74, 122  
SEER\*Stat, 74, 122  
separating character, 34  
server, 28, 95  
setting up a CanReg system, 22  
settings, 28  
sex, 50, 64, 84, 98  
SOCKS server, 95  
software installation, 17  
sorty by, 43  
SSH, 95  
standard population, 55  
Stata, 79  
supervisor, 83  
system folder, 96  
system variables, 49  
  
tab delimited, 59  
table, 43  
table builder, 68  
tables, 96  
Talend Open Studio, 123  
TCP/IP, 95  
titles, 66  
TrueCrypt, 124  
tumour table, 99  
tunnel, 95  
twitter, 12  
  
un-install, 18  
update user, 84  
user folder, 18, 96  
user manager, 82  
user manager, 83  
  
variable, 23, 26  
variable definitions, 103  
variable editor, 26  
variable names, 66  
variables to export, 67  
version, 12  
VPN, 95  
  
weights, 27, 86  
welcome screen, 39  
  
XML, 11, 22, 23, 28, 96, 105