European Commission

JRC TECHNICAL REPORTS

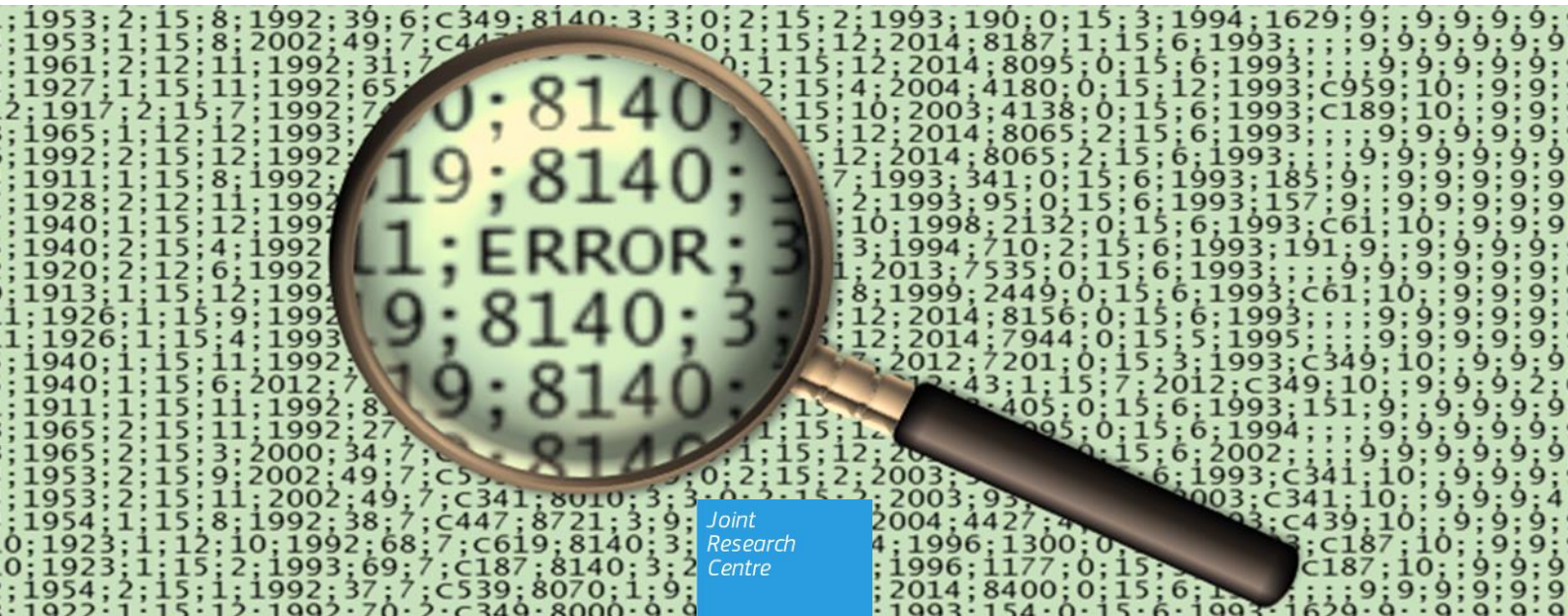
The JRC-ENCR Quality Check Software (QCS) for the validation of cancer registry data: user compendium

JRC-ENCR QCS Version 1.8

1.8.1 update

Francesco Giusti, Carmen Martos, Stefano Adriani, Tadeusz Dyba, Lena Voith von Voithenberg, Luciana Neamtiu, Raquel N. Carvalho, Giorgia Randi, Nadya Dimitrova, Nicholas Nicholson, Revveka Trigka, Enrico Ben, Emanuele Crocetti, Manola Bettio

2019



This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policy-making process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

Contact information
Name: Francesco Giusti
Address: Directorate F – Health, Consumers and Reference Materials, Health in Society Unit.
Via Enrico Fermi 2749, TP 127, I-21027 ISPRA (VA), Italy
E-mail: francesco.giusti@ec.europa.eu or JRC-ENCR@ec.europa.eu
Tel.: +39.0332.78.9420 or +39.0332.78.9926

JRC Science Hub
https://ec.europa.eu/jrc

JRC113014

Ispra: European Commission, 2019

© European Union, 2019

Reproduction is authorised provided the source is acknowledged.

How to cite: Francesco Giusti, Carmen Martos, Stefano Adriani, Tadeusz Dyba, Lena Voith von Voithenberg, Luciana Neamtiu, Raquel N. Carvalho, Giorgia Randi, Nadya Dimitrova, Nicholas Nicholson, Revveka N Trigka, Enrico Ben, Emanuele Crocetti, Manola Bettio, *The JRC-ENCR Quality Check Software (QCS) for the validation of cancer registry data: user compendium – version 1.8.1*, European Commission, Ispra 2019, JRC113014

All images © European Union 2019, except: page 4. Source: www.java.com

Printed in Italy

Table of contents

Introduction 1
1 Software overview and changes from the previous versions 2
2 System requirements and installation 3
2.1 Minimum system requirements 3
2.2 In case Java software is not installed on your computer 3
2.3 Further information and troubleshooting related to Java 5
2.4 How to install the QCS 7
2.5 Running the QCS on macOS 7
2.6 Running the QCS on Linux operating systems 7
2.7 Verify the correct installation 8
3 How to prepare an input file for the QCS 9
3.1 Incidence File 9
3.2 Mortality file 9
3.3 Population file 10
3.4 Life Table file 10
4 How does the software work? 11
5 Using the software 12
5.1 Running the software 12
5.2 Checking the files 13
5.2.1 Options 16
5.2.2 Help menu 17
5.3 Output files 18
6 How to interpret the output of incidence files provided by the QCS 23
Annex 1 – Known JRC-ENCR QCS issues and bugs 31
Annex 2 – List of error and warning codes 32
Annex 3 – Running the JRC-ENCR QCS in background 35

Introduction

In June 2015 the European Network of Cancer Registries (ENCR) and the European Commission Joint Research Centre (JRC) launched a **call for data** (https://encr.eu/call-for-data) for compilation of cancer indicators at the European level.

The call for data was organized with a view to minimizing the overheads on cancer registries, which previously were required to submit data in different formats to different sets of data users. The ENCR-JRC call centralized the data-collection process for the first time, thereby providing a type of data-brokering service on behalf of the European cancer registries.

One of the major outputs of the data call concerned the updating of the European monitoring system on cancer incidence and mortality (https://encr.eu/encr-jrc-project).

The description and specific data formats of the call are detailed in the submission guidelines, available both in the ENCR website (https://encr.eu/sites/default/files/pdf/2015_ENCR_JRC_Call_for_Data_Version_1_1.pdf) and in the JRC-ENCR portal (https://portal-encr.jrc.ec.europa.eu/).

The reliability and utility of the information provided by cancer registries is critically dependent on the quality of the data collected. The JRC, in cooperation with the ENCR, set up a working group involving the main stakeholders in the cancer-data domain in order to establish a comprehensive and standardised list of data quality checks to be adopted by European cancer registries and European projects for verification of data quality.

The outcome of the Working Group was a proposal on one common procedure for data quality checks among the European cancer-registry data stakeholders, reaching final agreements on case definition; the variables to be collected (including their format); and internal-consistency rules. The result of the initiative was an ENCR-endorsed set of recommendations to serve as guidelines for the data-acquisition procedure (https://encr.eu/sites/default/files/inline-files/Cancer_Data_Quality_Checks_Procedure_Report_online_0.pdf).

In order to enable cancer registries to perform data-quality checks independently and to test the adherence of their data to the required format of the recommendations, the JRC has developed a software toolkit (currently in version 1.8). The toolkit represents an important step in the process of facilitating the procedures to be followed by European cancer registries when submitting data in order to improve their quality and comparability.

Version 1.8 of the *JRC – ENCR Cancer Registries Data Quality Check Software*, which replaces version 1.7 distributed in 2016, includes the following features:

- checks on the data files format (for incidence, mortality, lifetables and population) and on variables names and order according to the data call protocol
- verification of variables' values
- cross checks among variables (internal consistency)
- check of multiple primary malignant tumours

The present report provides technical guidance to the software, and serves to help understand and interpret its output.

1 Software overview and changes from the previous versions

The JRC-ENCR Quality Check Software (QCS) version 1.8[1] is a stand-alone tool created for validating cancer-registries' data against the requirements of the protocol for the ENCR-JRC project on "Incidence and Mortality in Europe", detailed in version 1.1 (2018 update) of the ENCR-JRC report "A proposal on cancer data quality checks: one common procedure for European cancer registries".

QCS input files are incidence, mortality, population or life tables; the QCS output consists in a set of files containing warnings or errors found in the checked files.

In comparison to version 1.7, the 2018 version 1.8 release of the software includes the following enhancements:

- ability to be run also on macOS and Linux operating systems (see sections 2.5 and 2.6 below);
- the creation of a separate file reporting on Multiple Primary Malignant Tumours (*QCS-PrimaryDuplicates-Output.csv*);
- updated morphology families used by the Multiple Primary Malignant Tumour checks according to the 2011 update of ICD-O-3;
- corrections to the algorithms checking Multiple Primary Malignant Tumours;
- ability to check Multiple Primary Malignant Tumours also when there are errors in variables not related to morphology and topography, allowing for the three options: (**1**) skip primary duplicates, (**2**) perform primary duplicates checks for all records, (**3**) perform primary duplicates checks only for valid records (records with no errors, but with possible warnings);
- correction of the bug regarding age-class limits (W-AGMT check);
- correction of bugs in the implementation of Table 2 of the ENCR-JRC report "A proposal on cancer data quality checks: one common procedure for European cancer registries";
- completion of the list of values of the cause of death variable (27_Cause_death);
- improvement in flexibility of the software allowing values with different formats in checks related to variables, 30_pT, 32_pM, 33_cT, 35_cM;
- change to the format and univariate checks for variable 45_Breslow;

In comparison to version 1.8, the 2019 version 1.8.1 release of the software includes the following enhancements:

- ability to be run also in background. The GUI (standard execution) and Silent (background process) modes are available;
- optimized initialization at startup.

For the list of remaining known bugs and issues that will be addressed in a later release, please refer to ***Annex 1 – Known JRC-ENCR QCS issues and bugs***.

[1] https://encr.eu/tools-for-registries

2 System requirements and installation

This software has been developed for Windows operating systems that support Java (Windows 7 and above).

Version 1.8 of the QCS can also run on macOS and Linux operating systems (see sections 2.5 and 2.6) below. Sections 2.1-2.4 refer to Windows operating systems.

2.1 Minimum system requirements

- Windows Vista SP2 *OR* Windows 7 *OR* Windows 8 Desktop
- Support for DirectX 9 graphics and 32 MB of graphics memory

A Pentium 2 266 MHz or faster processor with at least 512 MB of physical RAM is recommended, with a minimum of 220 MB of free disk space to efficiently run the software.

2.2 In case Java software is not installed on your computer

Java software is needed to run the JRC-ENCR-QCS. In case Java is not installed on your computer, please follow the following steps, otherwise go to **section 2.3.**

- Go to Java.com and click on the **Free Java Download** button;
- On the browser download page click on the **Agree and Start Free Download** button;
- The File Download dialog box appears, click on the **Save File** button;
- Double click on the downloaded file in the Download Manager window or where you normally save downloaded files;
- Depending on your security settings, you may be presented with dialog boxes asking for permission to continue. Confirm that you want to proceed with the installation;
- The installation process starts. Click the **Install** button to accept the license terms and to continue.

Please refer to the following screenshots, referring to Java Version 8 Update 181:





After having completed all the steps of the installation process going through several consecutive dialog boxes, click **Close** on the last one and the Java installation process is finally completed.



Once Java software is correctly installed, you can install the JRC-ENCR-QCS.

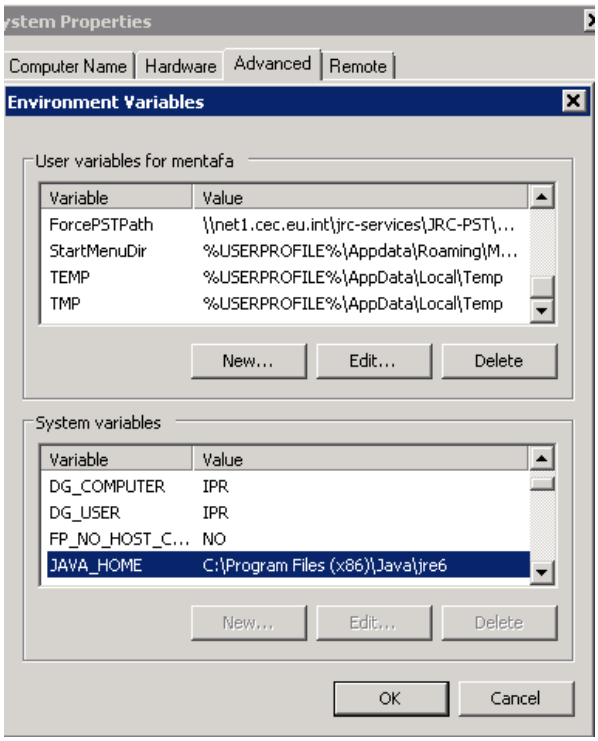
2.3 Further information and troubleshooting related to Java

If you need help in installing Java Runtime Environment installed on your machine, kindly ask to your System Administrator or local IT support to install it for you.

You will also need the JAVA_HOME environment variable correctly configured. Usually, this is done automatically. Please check with your System Administrator.

In Windows 7 (for other systems the procedure may vary) please refer to window *Start → Control Panel → System → Advanced System Settings → Environment variables* to configure the Java environment as follows:

Please refer to the next screenshot:



The official requirements for Java can be found here:
https://www.java.com/en/download/win_sysreq-sm.jsp

The required Java runtime environment can be downloaded from Oracle at https://www.java.com
Remember to choose the correct version for your operating system (Windows 32 bit or Windows 64 bit).

Please note: there are two versions of Java environments, Java Developer Kit (JDK) and Java Runtime Environment (JRE). **Please install JRE**.

Detect older versions (8u20 and later versions).

Starting with Java 8 Update 20 (8u20), on Windows systems, the Java Uninstall Tool is integrated with the installer to provide an option to remove older versions of Java from the system. The change is applicable to 32 bit and 64 bit Windows platforms.

Notifications about disabled Java and restoring prompts

The installer notifies you if Java content is disabled in web browsers, and provides instructions for enabling it. If you previously chose to hide some of the security prompts for applets and Java Web Start applications, the installer provides an option for restoring the prompts. The installer may ask you to reboot your computer if you chose not to restart an internet browser when it prompted you to do so.

Test Installation

To test that Java is installed and working properly on your computer, run this test applet (https://www.java.com/en/download/help/testvm.xml).

NOTE: You may need to restart (close and re-open) your browser to enable the Java installation in your browser.

Further information on how to install Java without third party sponsor offers: (https://www.java.com/en/download/faq/disable_offers.xml)

2.4 How to install the QCS

Once you download the latest version of the software from the ENCR website (https://encr.eu/tools-for-registries) please extract file ***JRC-ENCR-QCS-V1.8.1.zip*** on your computer.
You will be able to access folder "*JRC-ENCR-QCS-v1.8.1*" with all the related subfolders.

2.5 Running the QCS on macOS

1. Double click the ZIP file: the package will be unzipped in a new folder, having the same name of the ZIP package (but without any extension)

2. Press the combination *Command-Shift-U* (Command is the key with the Mac symbol) to open the Utility window

3. Double click the Terminal icon (or label, depending by your view settings) to open a Terminal window

4. Enter the Terminal window and move into the folder created at **step 1**. For example, if the target QCS file was named "JRC-ENCR-QCS-V1.8.1.zip", then you should execute the command:

   cd Desktop/JRC-ENCR-QCS-V1.8.1

5. Execute the file having the extension ".sh". For example if the file is named "start-jrc-encr-qcs.sh", then type the command:

   ./start-jrc-encr-qcs.sh

2.6 Running the QCS on Linux operating systems

1. Unzip the ZIP file into the directory where to wish to install the application. For example, if the target QCS file was named "JRC-ENCR-QCS-V1.8.1.zip" you should execute the command

   unzip JRC-ENCR-QCS-V1.8.1.zip

2. Move to the folder created at step 1. For example:

   cd JRC-ENCR-QCS-V1.8.1/

3. Make sure the ".sh" file has permissions for execution. If not, assign it executable permissions by typing the command:

   chmod +x start-jrc-encr-qcs.sh

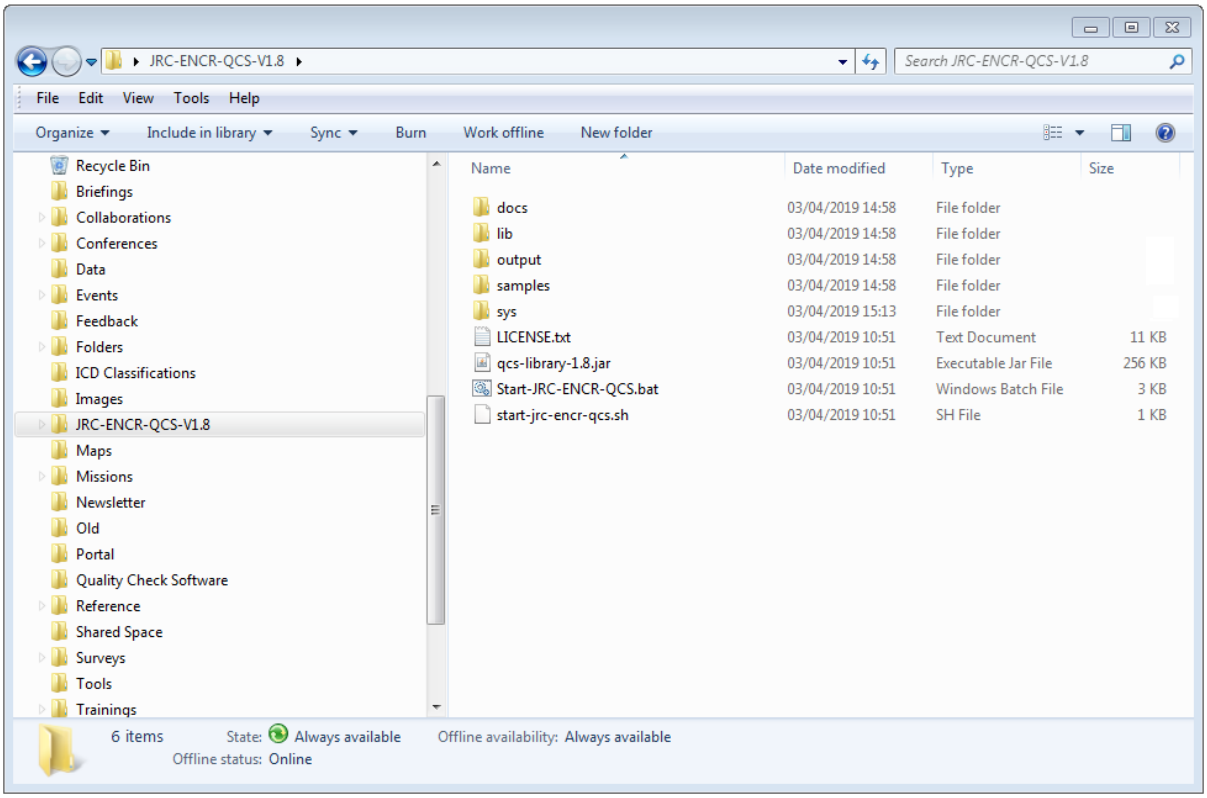
4. Execute the QCS by running the ".sh" file:

   ./start-jrc-encr-qcs.sh

2.7 Verify the correct installation

Navigate to the folder where you extracted the software and run it as specified in the next section of the manual, "*Running the Software*".

The expected directory structure is the following:



The folder *JRC-ENCR-QCS-V1.8.1* includes the following:

- The executable file *Start-JRC-ENCR-QCS.bat;*
- Files *start-jrc-encr-qcs.sh* and *LICENSE.txt*;
- The library *qcs-library-1.8.1.jar* file;
- Folders *docs, lib, output, samples* and *sys.*

docs: this folder contains all the documentation files of the software, such as the present report, the 2015 ENCR-JRC call for data protocol and the 2018 update (version 1.1) of the 2014 JRC Technical Report "A proposal on cancer data quality checks: one common procedure for European cancer registries" in pdf format.

lib: this folder includes library files used by the software at run-time.

output: this folder includes four subfolders, one for each of the different error reports that the QCS produces for the four type of files: *Incidence, Mortality, Population, LifeTables.*

samples: this folder contains two "empty" examples on the running of the application as a background process.

sys: this folder includes the following subfolders: *Config, Images, Log, Temp*. These folders are used by the operational system of the QCS. These folders contain resources used by the software at run-time.

3 How to prepare an input file for the QCS

In this section an example for each type of file accepted by the software is given.

The input files should be formatted as follows:

- Should be semicolon-separated files only;
- The first line should be the header.

3.1 Incidence File

First of all, you need to create the header of the file. For the incidence file the number of accepted variables for each record is 56 by default. By selecting a specific option, an additional variable is accepted (see section 5.2.1).

The file has a fixed structure (names, order and separation of variables by semicolon (**;**).

The header line is mandatory as such (please copy/paste the following, adding the line at the top of your incidence file).

1_Flag;2_Patient_ID;3_Tumour_ID;4_Day_DoB;5_Month_DoB;6_Year_DoB;7_Sex;8_Day_DoI;9_Month_DoI;10_Year_DoI;11_Age;12_BoD;13_Topo;14_Morpho;15_Beh;16_Grade;17_Autopsy;18_Vital_status;19_Day_FU;20_Month_FU;21_Year_FU;22_Survival;23_Laterality;24_Day_DoR;25_Month_DoR;26_Year_DoR;27_Cause_death;28_ICD_edition;29_TNM_prefix;30_pT;31_pN;32_pM;33_cT;34_cN;35_cM;36_Stage;37_TNM_edition;38_Cond_T;39_Cond_N;40_Cond_M;41_Dukes;42_FIGO;43_AArbor;44_Gleason;45_Breslow;46_EoD;47_Tsize;48_N_exam_nodes;49_N_met_nodes;50_Sent_nodes;51_Met_sent_nodes;52_Cfactor;53_Surgery;54_Systemic_th;55_Radiotherapy;56_BMtransp

Please note: do NOT put a semicolon at the end of the line. The line ends in "56_BMtransp" and not in "56_BMtransp**;**"

After the creation of the header, please proceed by creating the lines/records with the values of those variables. Only the core variables according to the *2015 ENCR-JRC Call for Data Protocol* are compulsory. The values for the optional variables can be missing.

When you finish inserting the records of your file, save it in csv or txt format.

You are now ready to load the incidence file into the JRC-ENCR QCS.

3.2 Mortality file

Similarly as above, you need first to create the header of the file. For mortality files the number of accepted variables is 5 by default. It is possible to have an additional variable (See section 5.2.1)

Please use the following line as header, copy/pasting it at the top of your file:

1_Calendar_Year;2_Sex;3_Age_Unit;4_Cause_Death;5_Number_Deaths

Please note: Please make sure that the variables are in the correct order, in the correct number and are separated by semicolons. The header line is mandatory. Do NOT put a semicolon at the end of each line.

After having created the header, please proceed by creating the lines/records with the values of those variables. When you finish creating the records of your file, save it in csv or txt format**.**

You are now ready to load the mortality file into the JRC-ENCR QCS.

3.3 Population file

Please create first the header of the file. For population files the number of accepted variables is 4 by default. It is possible to have an additional variable (See section 5.2.1).

Please use the following line as header, copy/pasting it at the top of your file:

1_Calendar_Year;2_Sex;3_Age_Unit;4_Number_of_residents

Please note: Please make sure that the variables are in the correct order, in the correct number and are separated by semicolons. The header line is mandatory. Do NOT put a semicolon at the end of the header.

After having created the header, please proceed by creating the lines/records with the values of those variables. When you finish creating the records of your file, save it in csv or txt format.

You are now ready to load the population file into the JRC-ENCR QCS.

3.4 Life Table file

Please create first the file header. For life table files the number of accepted variables is 4 by default. It is possible to have an additional variable (See section 5.2.1).

Please use the following line as header, copy/pasting it at the top of your file:

1_Calendar_Year;2_Sex;3_Age_Unit;4_ALL_causes_death

Please note: Please make sure that the variables are in the correct order, in the correct number and are separated by semicolons. The header line is mandatory. Do NOT put a semicolon at the end of each line.

After having created the header, please proceed by creating the lines/records with the values of those variables. When you finish creating the records of your file, save it in csv or txt format.

You are now ready to load the life table file into the JRC-ENCR QCS.

4 How does the software work?

The analysis process of an input *incidence* file is described below. Similar processes are performed for the other allowed input data files: *mortality*, *population* and *life table* files.

The software assumes that input files have *csv* or *txt* extensions. Files with *csv* and *txt* extension are shown first by default. Selecting the option "*All files*", files with extensions other than *csv* and *txt* are displayed. The incidence file should include 56 or 57 variables, semicolon-separated, and in the correct format (see below) as reported in in the *2015 ENCR-JRC Call for Data Protocol*.

The software checks that variable names are correct, and every single record is compliant with the valid format and value for each variable according to the *2015 ENCR-JRC Call for Data Protocol* as for:

- the number of variables;
- the presence of non-missing and non-blank values in the mandatory fields according to the *2015 ENCR-JRC Call for Data Protocol*;
- when applicable, the field content against a list of valid values. **Example**: patient's sex numeric value (variable 7_Sex) can be 1=male, 2=female, 3=other or 9=unknown. Every other value will produce an error;
- the field length, which must be within the allowed range. ***Example****: maximum length for patient ID (*variable 2_Patient_ID*) is 50 characters;*
- the validity of dates (also checking that dates are not set in the future);
- records failing the edits described in the 2018 update (version 1.1) to the 2014 JRC Technical Report "one common procedure for European cancer registries" (see also the *2015 ENCR-JRC Call for Data*).

Output messages from the JRC-ENCR QCS are saved in specific output. Five output files are generated (names below are relative to the *incidence* file):

1) *QCS-Incidence-FormatErrors.txt* – file listing all the records with format errors (wrong number of fields, etc.);
2) *QCS-Incidence-Output.pdf* – file with error and warning messages in pdf format including multiple primary malignant tumour warnings;
3) *QCS-Incidence-Output.txt* – file with error and warning messages in *txt* format including multiple primary malignant tumour warnings;
4) *QCS-output-tabular-forExcel.csv* – file with error and warning messages in *csv* format. This file can be imported by most software packages to allow for advanced data manipulation, such as linkage with the original file using the unique id patient+tumour id. If warnings for multiple primary malignant tumours are present they are not included in this file;
5) QCS-PrimaryDuplicates-Output.csv – the "QCS-output-tabular-forExcel.csv" file above does not contain the lines related to the multiple primary malignant tumour warnings (if any). Such lines are reported in the "QCS-PrimaryDuplicates-Output.csv" file instead.

5 Using the software

5.1 Running the software

- Please navigate to the folder in which you installed the software;
- Double click on the *Start-JRC-ENCR-QCS.bat* file;
- The user interface appears;



Note: The software runs <u>only</u> double clicking on the file ending in *.bat.* In case you click on the *.jar* file by mistake the following alert message is displayed: "*Please run the QCS from the .bat file*".

In this case, click "*OK"* and then run the software from the *.bat* file.

To quit the JRC-ECNR QCS just close the window, or select the "*Exit*" item in menu *File*.



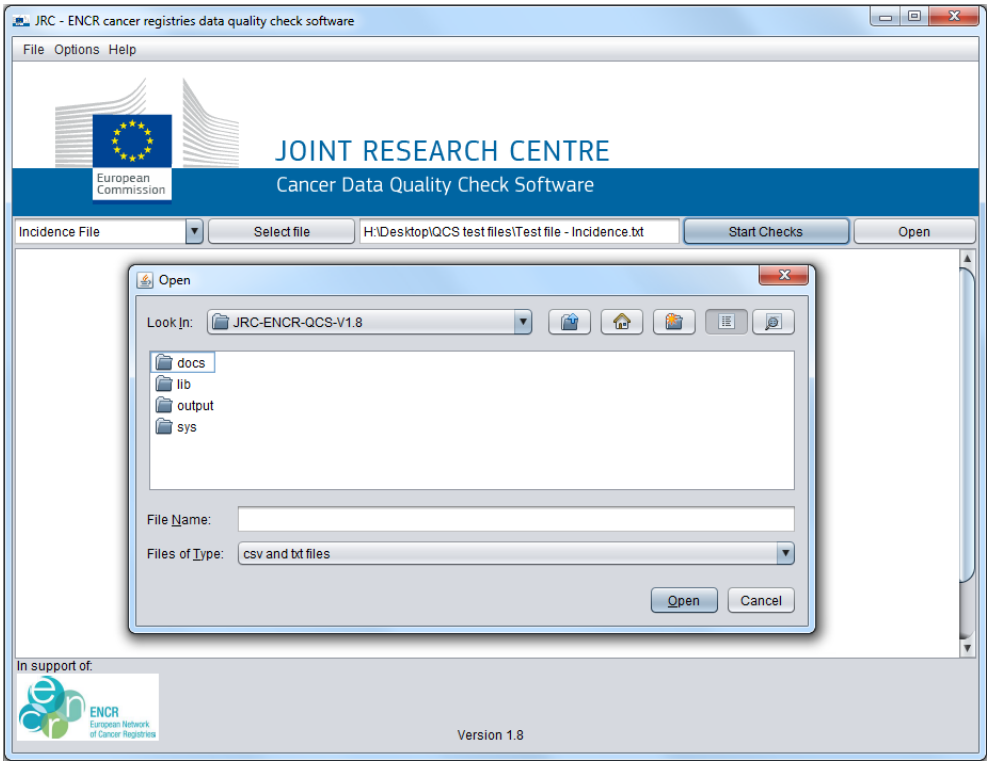
5.2 Checking the files

Select the type of file you want to check from the drop down menu.

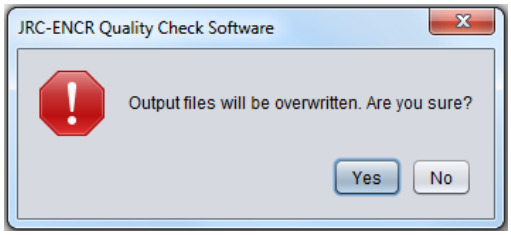
For instance, for checking an incidence file:

- Select the "*Incidence File*" option from the drop down menu;
- Press the "*Select File*" button;
- A file browsing window appears;
- Select the file to be checked.

The software accepts only files with semicolon (**;**) separated values (with extension such as *csv* or *txt*).



- Navigate to the folder where the incidence file to be checked is located, select it and press "*Open*";
- The "*Start Checks*" button will then become active and the full path of the file you have chosen will be displayed in the text box on the left of the button;
- Press the "*Start Checks*" button;
- A dialog box with the following message will appear: "*Output files will be overwritten. Are you sure?*"



If you had previously already checked the incidence file, please note that the output files will be overwritten. Please save them in a different folder or with a different name in case you want to keep them. After selecting "*Yes*", the software will run.

If you run the software for the first time, or you close it and open it again, a message for the initialisation of the system will be displayed on the screen (see screenshot below). It may take up to a few minutes for the software to initialise. The initialisation sets files and space on the disk for efficient running the software. This is necessary only once per session; in case you close the software, the initialisation will be performed again the next time you use the JRC-ECNR QCS.



Once launched, the quality check procedure will be running. The label of button "Start Checks" changes to "CANCEL", giving you the possibility to stop the process. Should this be the case, the following message is displayed in the text area: "*Processing interrupted by the user. Number of records read: …*"

If the process is not interrupted, a sequence of messages appears in the display text box. Those messages inform the user about the different steps of the procedure (see screenshot below).

Once the procedure is completed, the button label will revert to the "*Start Checks*" status and the software will be ready to check another file.

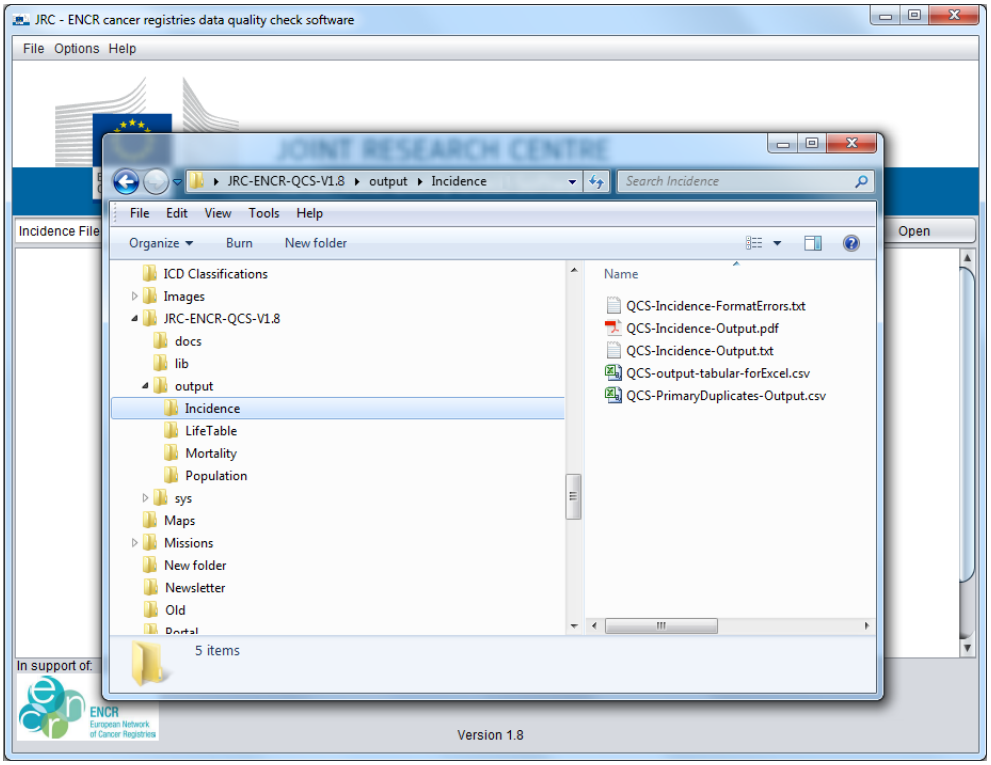


The output window of the software reports on the completed process, displaying:

- Number of records processed;
- Total number of errors;
- Total number of warnings with Flag=0;
- Total number of warnings with Flag=1;
- Total number of records rejected (this happens when the headers are correct but some of the variables are not present, not even left blank or with missing value);
- Total number of duplicate records (same patient id and tumour id for more than one record);
- Total number warnings for multiple primary malignant tumours;

Please note that *mortality*, *population* and *life table* files are not checked for duplicate records, as they contain aggregated data. Therefore for these files **the number of duplicated records** above is not displayed.

You can finally access the outputs, by clicking on "*Open*", and accessing the "*output*" folder, containing all the report files.

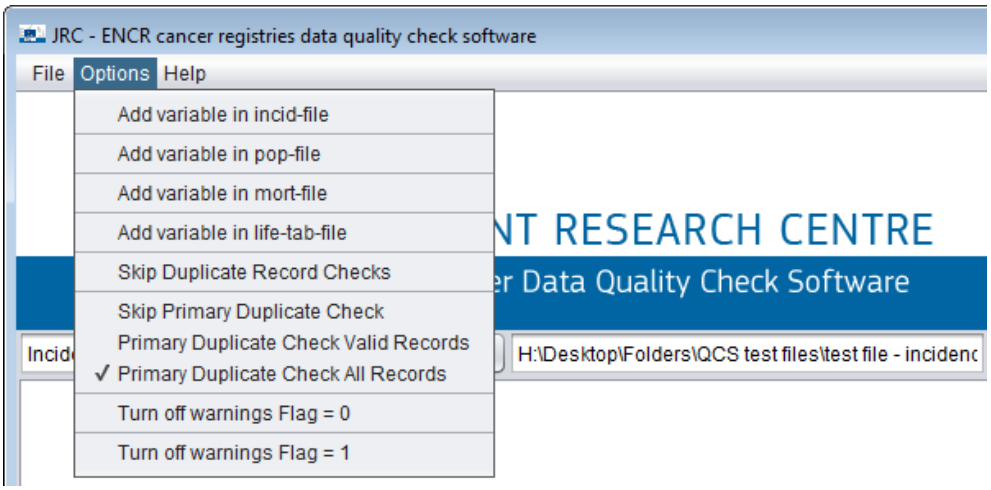


Similarly, *mortality*, *population* and *life table* files can be checked by selecting the type of the file from the drop down menu.

The procedure for checking such files is the same as described above for Incidence files.

5.2.1 Options

The Options menu enables to select additional choices before running the checks.



The following options are available:

- *Add variable in incid-file*. This option provides the possibility to add an extra variable to the incidence file. The extra variable **must be named** *57_Registry_Code*, and have an alphanumeric format. With this option the number of variables required by the JRC-ENCR QCS in the incidence file is increased by one;

- *Add variable in pop-file*. By selecting this option it is possible to add one extra variable to the population file. There is no compulsory name for this variable. With this option the number of variables required by the JRC-ENCR QCS in the population file is increased by one;
- *Add variable in mort-file*. By selecting this option it is possible to add one extra variable to the mortality file. There is no compulsory name for this variable. With this option the number of variables required by the JRC-ENCR QCS in the mortality file is increased by one;
- *Add variable in life-tab-file*. By selecting this option it is possible to add one extra variable to the life table file. There is no compulsory name for this variable. With this option the number of variables required by the JRC-ENCR QCS in the life table file is increased by one;
- *Skip Duplicate Record Check*. If this option is selected the search for duplicate records, i.e. more than one record with the same patient id and tumour id, is not performed by the software;
- *Skip Primary Duplicate Check.* This options allows to skip the search for Multiple Primary Malignant Tumours;
- *Primary Duplicate Check Valid Records*. If this option is selected the search for Multiple Primary Malignant Tumours is performed on records without errors (but is performed on records with warnings only). See Annex 2 for the list of errors and warnings;
- *Primary Duplicate Check All Records*. If this option is selected the search for Multiple Primary Malignant Tumours is performed on all records, except those with errors regarding topography and morphology (e.g. E-MOBE, see chapter 6 below);
- *Turn off warnings Flag = 0*. By selecting this option warnings are not returned for records with variable *1_Flag* equal to 0. Error codes are still reported by the software;
- *Turn off warnings Flag = 1*. By selecting this option warnings are not returned for records with variable *1_Flag* equal to 1. Error codes are still reported by the software;

5.2.2 Help menu

The "*Help*" menu includes option "*Manual and documentation*". This opens a folder with the present JRC-ENCR QCS user guidelines and the following files:

- *Cancer Data Quality Checks Procedure Report.pdf*: the *2018 update (version 1.1) of the 2014 JRC Technical Report "one common procedure for European cancer registries"*)
- *2015_ENCR_JRC_Call_for_Data_Version_1_1.pdf* - the submission guidelines for the ENCR-JRC project on "Incidence and Mortality in Europe".

In the "*Help*" menu you can also find the "*About*" item, with the credits and copyright statement.



5.3 Output files

The output files are located in the subfolders inside the "*output*" folder, depending on the type of the file. For example, output files for an Incidence file are located in the "**\JRC-ENCR-QCS-V1.8.1\output\Incidence"** folder.

The following four screenshots refer to the *QCS-Incidence-Output.pdf* file:

First page of the *QCS-Incidence-Output.pdf* file

```
**********************************************************************************************************************
QUALITY CHECK SOFTWARE REPORT - INCIDENCE
**********************************************************************************************************************
**********************************************************************************************************************
PROCESSING PARAMETERS
**********************************************************************************************************************
File process start: 2018-08-22 13:36:14.181
File process end: 2018-08-22 13:36:21.731
Skip duplicate record check: false
Extra variable (registry code): false
Skip Primary Duplicates check: false
Skip warnings with Flag = 0: false
Skip warnings with Flag = 1: false
Validate by: QCS Version 1.8 (s)
File Processed:
H:\Desktop\QCS test files\test file - incidence 2018-07-17.txt
**********************************************************************************************************************
PROCESSING STATISTICS
**********************************************************************************************************************
Number of records read:               11986
Total number of errors:                  31
Number of warnings Flag=0:                0
Number of warnings Flag=1:             2721
Total number of records rejected:         0
Total number of duplicate records:        0
Warnings for Multiple primary cancers:   14
**********************************************************************************************************************
KEY TO ERROR AND WARNING CODES
**********************************************************************************************************************
E-AGED: DoI - DoB different from Age
E-CoDV: Date of last known vital status not valid
E-CoDA: DoB + DoI not coherent (p.16)
E-DATE: DD/MM/YYYY Non valid
E-ECOD: ICD edition + Cause of death not valid
E-FORM: Format error
E-MISS: Value missing
E-MOBE: Morphology + Behaviour not valid
E-OUTR: Value out of range
E-SETO: Topography + Sex not valid (tab.4)
E-SURV: Survival can not be calculated
E-VBDA: VitalStatus + BoD + Autopsy not valid (p.40)
WARNING CODES:
W-AGMT: Unlikely Age + tumour type (tab.3)
W-BDMO: Morphology too specific (p.30)
W-BDMS: Morphology not specific enough (p.30)
W-BDMU: BoD + Morpho/Beh (p.30)
W-BDpM: BoD + pM not valid (p.40)
W-BDpN: BoD + pN not valid (p.40)
W-BDpT: BoD + pT not valid (p.40)
W-BEGR: Beh + grade not valid (tab.7)
W-BTNM: Beh + TNM not valid (p.41)
W-CdMN: Met_Nodes + condN not valid (p.41)
W-CdpM: TNM + condM not valid (p.41)
W-CdpN: TNM + condN not valid (p.41)
W-ECdN: EoD + condN not valid (p.41)
W-EDCD: EoD + cond T,N or M not valid (p.41)
W-EDNM: EoD + TNM not valid (p.40)
W-EoDN: EoD + met_Nodes not valid (p.41)
W-EoDT: EoD + Topography not valid (p.41)
W-ETNC: EoD + TNM/condN not valid (p.41)
W-MOGR: Morphology + Behaviour + grade not valid (tab.6)
W-MOTO: Morphology + Topography not valid (tab.8)
W-MPMT: Multiple primary malignant tumour (p. 42)
W-SEMO: Sex + Morphology not valid (tab.5)
W-TMED: EoD + M not valid (p.41)
W-TNED: EoD + N not valid (p.41)
W-TNMA: Age + TNM not valid (tab.2)
W-TNMC: Topo + TNM edition + T,N,M + Stage clinical (tab.2)
W-TNMM: Topo + TNM edition + T,N,M + Morpho (tab.2)
W-TNMn: TNM + met_Nodes not valid (p. 41)
W-TNMP: Topo + TNM edition + T,N,M + Stage pathological (tab.2)
W-TOLA: Topography + Laterality not valid (p.33)
W-TOPN: Topography + Nodes not valid (p.40)
W-TOTS: Topography + Tsize not valid (p.40)
W-TTED: EoD + T not valid (p.41)
```

Detail: upper section

```
**********************************************************************************************************************
QUALITY CHECK SOFTWARE REPORT - INCIDENCE
**********************************************************************************************************************
**********************************************************************************************************************
PROCESSING PARAMETERS
**********************************************************************************************************************
File process start: 2018-08-22 13:36:14.181
File process end: 2018-08-22 13:36:21.731
Skip duplicate record check: false
Extra variable (registry code): false
Skip Primary Duplicates check: false
Skip warnings with Flag = 0: false
Skip warnings with Flag = 1: false
Validate by: QCS Version 1.8 (s)
File Processed:
H:\Desktop\QCS test files\test file - incidence 2018-07-17.txt
**********************************************************************************************************************
PROCESSING STATISTICS
**********************************************************************************************************************
Number of records read:               11986
Total number of errors:                  31
Number of warnings Flag=0:                0
Number of warnings Flag=1:             2721
Total number of records rejected:         0
Total number of duplicate records:        0
Warnings for Multiple primary cancers:   14
**********************************************************************************************************************
KEY TO ERROR AND WARNING CODES
**********************************************************************************************************************
E-AGED: DoI - DoB different from Age
E-CoDV: Date of last known vital status not valid
E-CoDA: DoB + DoI not coherent (p.16)
E-DATE: DD/MM/YYYY Non valid
E-ECOD: ICD edition + Cause of death not valid
```

Processing parameters:

- *File process start, File process end*;
- *Skip duplicate record check*. "*False*" means that the user did not ask to skip the search for duplicate records, i.e. more than one record with the same patient id and tumour id. "*True*" means that the search for duplicate records will not be performed;
- *Extra variable (registry code)*. "*False*" means that the user did not add variable "*57_Registry_Code*" to the file, whereas "*true*" means that the extra variable has been added;
- *Skip Primary Duplicates check, Skip warnings with Flag = 0, Skip warnings with Flag = 1*. If "false" the user decided not to skip the respective check, whereas "*true*" means that the check was skipped;
- *Validate by*. The JRC-ENCR QCS version used to produce the report is added, e.g. "*QCS Version 1.8.1 (s)*". If value "*c*" appears in brackets after the version number, at least one configuration file was *customized* (i.e. modified) by the user before running the software. If the default value "*s*" appears, the *standard* version of the JRC-ENCR QCS was used;
- *File processed*. The name and the path of the file checked by the software is reported.

Processing statistics:

- *Number of records read, Total numbers of errors, Number of warnings Flag=0, Number of warnings Flag=1*;
- *Total number of records rejected*. Records are rejected whenever the headers are correct but some of the variables are not present, not even left blank or with missing value;
- *Total number with duplicate records*. Records with repeated patient id/tumour id values;
- *Warnings for Multiple primary cancers*. Number of records with warning for Multiple Primary Malignant Tumour.

Key to error and warning codes:

- Errors and warnings are reference by codes, described by short labels and accompanied by the reference to the relevant table or page from the 2018 update of the JRC Technical Report "*A proposal on cancer data quality checks: one common procedure for European cancer registries*". See also *Annex 2 – List of error and warning codes* below.

Detail: second page (summary of errors and warnings, multiple primary malignant tumours)

```
**********************************************************************************************************************
SUMMARY OF ERRORS BY CODE
**********************************************************************************************************************
----------------------------------------------
 E-FORM                                 1
----------------------------------------------
 E-MOBE                                 30
----------------------------------------------
**********************************************************************************************************************
SUMMARY OF WARNINGS BY CODE
**********************************************************************************************************************
----------------------------------------------
 W-BTNM                                 5
----------------------------------------------
 W-BDpT                                 12
----------------------------------------------
 W-BDMU                                 6
----------------------------------------------
 W-MOGR                                 72
----------------------------------------------
 W-MOTO                                 51
----------------------------------------------
 W-AGMT                                 6
----------------------------------------------
**********************************************************************************************************************
        MULTIPLE PRIMARY MALIGNANT TUMOUR CHECK
**********************************************************************************************************************
----------------------------------------------------------------------------------------------------------------------
 2_Patient_ID  14391                                  3_Tumour_ID   1
----------------------------------------------------------------------------------------------------------------------
 12_BoD  13_Topo  14_Morpho  15_Beh  7_Sex  DoI        DoB          18_Vital_Status 23_Laterality
----------------------------------------------------------------------------------------------------------------------
 7       C509     8500       3       2      15/5/2011  11/8/1946    1               2
----------------------------------------------------------------------------------------------------------------------
 2_Patient_ID  14391                                  3_Tumour_ID   2
----------------------------------------------------------------------------------------------------------------------
 12_BoD  13_Topo  14_Morpho  15_Beh  7_Sex  DoI        DoB          18_Vital_Status 23_Laterality
----------------------------------------------------------------------------------------------------------------------
 7       C509     8500       3       2      28/7/2001  11/8/1946    1               1
----------------------------------------------------------------------------------------------------------------------
```

Summary of errors by code: see *Annex 2 – List of error and warning codes*

Summary of warnings by code: see *Annex 2 – List of error and warning codes*

Multiple primary malignant tumour check: for each warning the following variables are reported: *2_Patient_ID*, *3_Tumour_ID*, *12_Bod* (basis of diagnosis), *13_Topo* (topography), *14_Morpho* (morphology), *15_Beh* (behaviour), 7_Sex, *DoI* (date of incidence), *DoB* (date of birth), *18_Vital_Status*, *23_Laterality*

Detail: page(s) with errors and warnings

```
**********************************************************************************************************************
ERRORS AND WARNINGS
**********************************************************************************************************************
----------------------------------------------------------------------------------------------------------------------
 2_Patient_ID                                          3_Tumour_ID
----------------------------------------------------------------------------------------------------------------------
 1_Flag  13_Topo  14_Morpho  15_Beh  7_Sex  DoI         DoB          Var_Name        Var_Value  Error_Code
----------------------------------------------------------------------------------------------------------------------
 1       C421     9731       3       2      4/11/2014  2/12/1958    14_Morpho       9731       W-MOTO
                                                                    13_Topo         C421       W-MOTO
----------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------------------------------
 2_Patient_ID                                          3_Tumour_ID   '
----------------------------------------------------------------------------------------------------------------------
 1_Flag  13_Topo  14_Morpho  15_Beh  7_Sex  DoI         DoB          Var_Name        Var_Value  Error_Code
----------------------------------------------------------------------------------------------------------------------
 1       C539     8000       3       2      30/6/2014  26/6/1970    14_Morpho       8000       W-BDMS
                                                                    12_BoD          7          W-BDMS
----------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------------------------------
 2_Patient_ID                                          3_Tumour_ID   '
----------------------------------------------------------------------------------------------------------------------
 1_Flag  13_Topo  14_Morpho  15_Beh  7_Sex  DoI         DoB          Var_Name        Var_Value  Error_Code
----------------------------------------------------------------------------------------------------------------------
 1       C445     8090       3       1      29/3/2014  31/5/1967    13_Topo         C445       W-TOLA
                                                                    23_Laterality   3          W-TOLA
```

Errors and warnings: for each warning or error the following variables are reported: *2_Patient_ID*, *3_Tumour_ID*, 1_Flag, *13_Topo* (topography), *14_Morpho* (morphology), *15_Beh* (behaviour), 7_Sex, *DoI* (date of incidence), *DoB* (date of birth), *Var_Name* and *Var_Value* (list of variables which caused the warning or error to be returned by the JRC-ENCR QCS, and their values), *Error_Code* (code according to list in *Annex 2 – List of error and warning codes*)

The following screenshots refer to the *QCS-output-tabular-forExcel.csv* file:

Line_nr	2_Patient_ID	3_Tumour_ID	1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Error_code	Error_Description	Var1_Name	Var1_Value	Var2_Name	Var2_Va
209	13198	1	1	C421	9731	3	2	04/11/2014	02/12/1958	W-MOTO	Morphology + Topography not valid	13_Topo	C421	14_Morpho	9731
213	13490	1	1	C539	8000	3	2	30/06/2014	26/06/1970	W-BDMS	Morphology not specific enough (p.30)	14_Morpho	8000	12_BoD	7
217	13498	1	1	C445	8090	3	1	29/03/2014	31/05/1967	W-TOLA	Topography + Laterality not valid	13_Topo	C445	23_Laterality	3
251	13555	2	1	C445	8092	3	1	17/08/2014	10/10/1972	W-TOLA	Topography + Laterality not valid	13_Topo	C445	23_Laterality	3
444	13787	1	1	C445	8743	2	1	09/10/2014	21/10/1953	E-MOBE	Morphology + Behavior not valid	14_Morpho	8743	15_Beh	2
874	14002	1	1	C445	8743	2	1	10/11/2014	10/10/1952	W-TOLA	Topography + Laterality not valid	13_Topo	C445	23_Laterality	2
1903	15011	1	1	C421	9761	3	1	15/09/2015	23/11/1969	W-MOTO	Morphology + Topography not valid	13_Topo	C421	14_Morpho	9761
1951	15077	1	1	C445	8743	2	2	19/09/2015	02/03/1947	E-MOBE	Morphology + Behavior not valid	14_Morpho	8743	15_Beh	2
2566	15701	1	1	C421	9960	3	2	01/11/2015	14/03/1948	W-BDMS	Morphology not specific enough (p.30)	14_Morpho	9960	12_BoD	5
2571	15709	1	1	C445	8090	3	2	10/10/2015	27/03/1943	W-TOLA	Topography + Laterality not valid	13_Topo	C445	23_Laterality	2
2575	15722	1	1	C421	9962	3	1	23/09/2015	18/01/1934	W-BDMU	BoD + Morpho/Beh (p.30)	14_Morpho	9962	12_BoD	6
2756	15929	1	1	C421	9731	3	1	12/08/2015	15/08/1933	W-MOTO	Morphology + Topography not valid	13_Topo	C421	14_Morpho	9731

Detail: left part

Line_nr	2_Patient_ID	3_Tumour_ID	1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB
209	13198	1	1	C421	9731	3	2	04/11/2014	02/12/1958
213	13490	1	1	C539	8000	3	2	30/06/2014	26/06/1970
217	13498	1	1	C445	8090	3	1	29/03/2014	31/05/1967
251	13555	2	1	C445	8092	3	1	17/08/2014	10/10/1972
444	13787	1	1	C445	8743	2	1	09/10/2014	21/10/1953
874	14002	1	1	C445	8743	2	1	10/11/2014	10/10/1952
1903	15011	1	1	C421	9761	3	1	15/09/2015	23/11/1969
1951	15077	1	1	C445	8743	2	2	19/09/2015	02/03/1947
2566	15701	1	1	C421	9960	3	2	01/11/2015	14/03/1948
2571	15709	1	1	C445	8090	3	2	10/10/2015	27/03/1943
2575	15722	1	1	C421	9962	3	1	23/09/2015	18/01/1934
2756	15929	1	1	C421	9731	3	1	12/08/2015	15/08/1933

Detail: right part

Error_code	Error_Description	Var1_Name	Var1_Value	Var2_Name	Var2_Value	Var3_Name
W-MOTO	Morphology + Topography not valid	13_Topo	C421	14_Morpho	9731	
W-BDMS	Morphology not specific enough (p.30)	14_Morpho	8000	12_BoD	7	
W-TOLA	Topography + Laterality not valid	13_Topo	C445	23_Laterality	3	
W-TOLA	Topography + Laterality not valid	13_Topo	C445	23_Laterality	3	
E-MOBE	Morphology + Behavior not valid	14_Morpho	8743	15_Beh	2	
W-TOLA	Topography + Laterality not valid	13_Topo	C445	23_Laterality	2	
W-MOTO	Morphology + Topography not valid	13_Topo	C421	14_Morpho	9761	
E-MOBE	Morphology + Behavior not valid	14_Morpho	8743	15_Beh	2	
W-BDMS	Morphology not specific enough (p.30)	14_Morpho	9960	12_BoD	5	

6 How to interpret the output of incidence files provided by the QCS

This section describes how to interpret the outcomes of the JRC-ENCR QCS for the variables and their combinations defined as "core" in the *2015 Call for Data Protocol* (https://encr.eu/sites/default/files/pdf/2015_ENCR_JRC_Call_for_Data_Version_1_1.pdf), and having an impact on the incidence estimation. Some examples related to warnings of multiple primary malignant tumours (MPMT) are also reported.

The code of the errors stars by **E**(**-**XXXX) and the code of the warnings by **W**(**-**XXXX).

1) Errors due to variable values and their format

- **E-OUTR**: out of range.

  When the variables have values different from the ones allowed by the *2015 Call for Data Protocol* or the 2018 update of the JRC Technical Report (https://encr.eu/sites/default/files/inline-files/Cancer_Data_Quality_Checks_Procedure_Report_online_0.pdf) the QCS returns error E-OUTR.

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C427	9800	3	2	15/9/1999	15/10/1910	13_Topo	C427	E-OUTR

  In this example the QCS gives the error E-OUTR because the topography C427 does not exist in the International Classification of Diseases for Oncology, third edition[2] (ICD-O-3).

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C620	9999	9	1	9/4/2000	31/12/1978	14_Morpho	9999	E-OUTR
							15_Beh	9	E-OUTR

[2] International Classification of Diseases for Oncology, Third Edition, First Revision. Geneva: World Health Organization, 2013.

In this example the QCS returns error E-OUTR because morphology 9999 does not exist in the ICD-O-3, and value 9 is not allowed according to the 2015 call for data protocol.

- **E-MISS**: value missing.

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C169	8140	3	1	99/10/1978	99/99/9999	11_Age	999	E-MISS

In this example the QCS gives error E-MISS because the core variable *age* is unknown and cannot be calculated.

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C187		0	1	99/9/2000	99/11/1927	14_Morpho		E-MISS

In this example the QCS returns error E-MISS because the core variable morphology has a missing value.

- **E-FORM**: format error.

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C443	80984	3	1	99/99/2012	99/99/9999	14_Morpho	80984	E-FORM

In this example the QCS gives error E-FORM because morphology should have four digits instead of five according to the ICD-O-3.

2) Errors due to inconsistency of the dates.

- **E-CoDA**: date of birth and date of incidence are not consistent.

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C159	8140	3	2	15/12/1992	30/8/2016	6_Year_DoB	2016	E-CoDA

In this example the QCS is gives error E-CoDA because the year of birth is later than the year of incidence.

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C741	9490	3	1	22/2/1992	15/3/1992	5_Month_DoB	3	E-CoDA

In this example the QCS gives error E-CoDA because the month of birth occurs after the year of incidence.

- **E-CoDV**: date of the incidence and date of the last known vital status are not consistent.

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C549	8000	3	2	5/8/2013	27/4/1933	8_Day_DoI	5	E-CoDV
							9_Month_DoI	8	E-CoDV
							10_Year_DoI	2013	E-CoDV
							19_Day_FU	9	E-CoDV
							20_Month_FU	8	E-CoDV
							21_Year_FU	2012	E-CoDV

In this example the QCS gives error E-CoDV because the date (year) of incidence occurs later than the date (year) of last known vital status.

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C160	8000	3	1	26/06/2009	23/10/1924	8_Day_DoI	26	E-CoDV
							9_Month_DoI	06	E-CoDV
							10_Year_DoI	2009	E-CoDV
							19_Day_FU	24	E-CoDV
							20_Month_FU	04	E-CoDV
							21_Year_FU	2009	E-CoDV

In this example the QCS gives error E-CoDV because the date (month) of incidence occurs later than the date (month) of last known vital status.

3) Errors and warnings due to tumour and demographic variables combinations.

- **E-SETO**: sex and topography combinations are not valid.

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
							7_Sex	1	E-SETO
0	C569	8000	3	1	99/99/2009	99/99/9999	13_Topo	C569	E-SETO

In this example the QCS returns error E-SETO because the combination topography=C569 (ovary) and sex=1 (men) is not valid.

- **W-AGMT**: age and morphology/topography combinations are unlikely.

2_Patient_ID					3_Tumour_ID				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C424	9652	3	1	99/12/2003	29/10/2003	11_Age	0	W-AGMT
							14_Morpho	9652	W-AGMT

In this example the QCS gives warning W-AGMT because the morphology 9652 (Hodgkin lymphoma, mixed cellularity, NOS) is unlikely between ages 0-2.

2_Patient_ID					3_Tumour_ID				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C619	8140	3	1	99/3/2007	29/5/1996	11_Age	10	W-AGMT
							13_Topo	C619	W-AGMT
							14_Morpho	8140	W-AGMT

In this example the QCS gives warning W-AGMT because the topography= C619 (prostate) in combination with morphology 8140/3 (adenocarcinoma, NOS) is unlikely under age of 40.

4) Errors and warnings due to tumour variables combinations.

- **E-MOBE**: morphology and behaviour combinations are not included in the ICD-O-3

Although according to Rule F of the ICD-O-3 it is exceptionally possible to have morphology and behaviour combination not listed in the ICD-O-3, the current version of the QCS (v.1.8) reports as errors the morphology and behaviour combinations not listed in the ICD-O-3. They will be marked as warnings in the next version of the software.

2_Patient_ID					3_Tumour_ID				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C569	8621	3	2	7/4/2005	10/6/1982	14_Morpho	8621	E-MOBE
							15_Beh	3	E-MOBE

In this example the QCS gives error E-MOBE because morphology=8621 (granulosa cell-theca cell tumour) and behaviour=3 (malignant tumour) combination are not included in the ICD-O-3.

The combination morphology and behaviour presented in the example above is possible but unlikely.

2_Patient_ID					3_Tumour_ID				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C421	9950	1	1	20/12/1989	7/3/1921	14_Morpho	9950	E-MOBE
							15_Beh	1	E-MOBE

In this example the QCS gives error E-MOBE because morphology 9950 (polycythaemia vera) in the ICD-O-3 has behaviour=3 (malignant tumour).

This term (polycythaemia vera) changed from borderline tumour (behaviour=1) in ICD-O-2[3], to malignant tumour (behaviour=3) in ICD-O-3.

- **W-BDMU**: basis of diagnosis and morphology/behaviour combinations are unlikely

2_Patient_ID					3_Tumour_ID				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C187	8210	2	2	99/11/1996	99/11/1922	12_BoD	6	W-BDMU
							15_Beh	2	W-BDMU

In the example above the QCS returns warning W-BDMU because the combination behaviour=2 (in situ tumour) and base of diagnosis=6 (histology of a metastasis) is not valid.

2_Patient_ID					3_Tumour_ID				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C421	9823	3	2	99/5/2013	99/7/1927	14_Morpho	9823	W-BDMU
							12_BoD	6	W-BDMU

In the example below the QCS gives warning W-BDMU because the combination base of diagnosis=6 (histology of a metastasis) and morphology (9823) coded as haematological malignancy is very unlikely. Usually haematological malignancies are diagnosed by cytology (base of diagnosis=5) or histology (base of diagnosis=7).

- **W-BDMO**: morphology too specific according to the basis of diagnosis

2_Patient_ID					3_Tumour_ID				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C209	8140	2	2	23/10/2013	28/11/1927	12_BoD	1	W-BDMO
							15_Beh	2	W-BDMO

In the example below the QCS returns warning W-BDMO because it is very unlikely to identify behaviour=2 (in situ tumour) if basis of diagnosis=1 (clinical).

[3] International Classification of Diseases for Oncology, Second Edition. Geneva: World Health Organization, 1990.

2_Patient_ID					3_Tumour_ID				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C199	8010	2	2	99/11/1999	99/9/1907	12_BoD 15_Beh	2 2	W-BDMO W-BDMO

As in the previous example, the QCS gives warning W-BDMO because it is very unlikely to identify behaviour=2 (in situ tumour) being the basis of diagnosis=2 (clinical investigation).

- **W-BDMS**: morphology not specific enough according to the basis of diagnosis

2_Patient_ID					3_Tumour_ID				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C341	8000	3	2	99/1/1991	99/1/1907	14_Morpho 12_BoD	8000 7	W-BDMS W-BDMS

In this example the QCS gives warning W-BDMS because the morphology= 8000 (neoplasm, malignant) is not specific enough taking into account the basis of diagnosis=7 (histology of a primary tumour).

2_Patient_ID					3_Tumour_ID				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C809	8001	3	2	99/5/1990	99/8/1914	14_Morpho 12_BoD	8001 7	W-BDMS W-BDMS

Regarding the morphology and basis of diagnosis, this example is similar to the previous one. In addition, basis of diagnosis=7 (histology of a primary tumour) is not coherent with topography=C809 (unknown primary site).

- **W-BTNM**: behaviour and TNM combination not valid

2_Patient_ID					3_Tumour_ID				
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C629	9061	3	1	99/2/2011	99/9/1991	15_Beh 30_pT	3 is	W-BTNM W-BTNM

In this example the QCS gives warning W-BTNM because behaviour=3 (malignant tumour) is not coherent with pathological T (pT)=is (carcinoma in situ).

- **W-MOGR**: morphology, behaviour and grade combinations are unlikely

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C569	8620	3	2	27/5/2011	10/7/1954	16_Grade	5	W-MOGR
							14_Morpho	8620	W-MOGR
							15_Beh	3	W-MOGR

The QCS gives warning W-MOGR because grade=5 (T-cell) is used to denote cell lineage for haematological malignancies (leukaemia and lymphoma). Morphology=8620 (granulosa cell tumour, malignant) is not a haematological malignancy.

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C445	9709	3	1	99/5/1998	99/11/1913	16_Grade	6	W-MOGR
							14_Morpho	9709	W-MOGR
							15_Beh	3	W-MOGR

In this example, the QCS gives warning W-MOGR because the morphology= 9709 (Cutaneous T-cell lymphoma, NOS) should have grade=5 (T-cell) instead of 6.

- **W-MOTO**: morphology and topography combinations are unlikely

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C779	8070	3	2	30/12/2008	31/10/1946	14_Morpho	8070	W-MOTO
							13_Topo	C779	W-MOTO

The QCS gives warning W-MOTO because topography=C779 (Lymph node, NOS) and morphology=8070 (squamous cell carcinoma, NOS); this combination is probably a metastasis and topography should be coded as C809.

2_Patient_ID						3_Tumour_ID			
1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C539	8120	2	2	9/11/2007	15/9/1959	14_Morpho	8120	W-MOTO
							13_Topo	C539	W-MOTO

The QCS gives warning W-MOTO because topography=C539 (cervix uteri) and morphology=8120 (transitional cell carcinoma, NOS); the combination is very rare.

5) Warnings for multiple primary malignant tumours (MPMT).

2_Patient_ID						3_Tumour_ID		
12_BoD	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	18_Vital_Status	23_Laterality
2	C717	8000	3	2	23/12/1996	31/12/1934	2	9
2_Patient_ID						3_Tumour_ID		
12_BoD	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	18_Vital_Status	23_Laterality
5	C717	9590	3	2	22/11/1996	31/12/1934	2	9

In this example, the QCS gives warning for MPMT because probably the two records are the same tumour.

2_Patient_ID						3_Tumour_ID		
12_BoD	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	18_Vital_Status	23_Laterality
7	C679	8130	3	1	20/5/1993	9/12/1920	2	9
2_Patient_ID	19931					3_Tumour_ID	2000	
12_BoD	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	18_Vital_Status	23_Laterality
6	C809	8000	3	1	6/11/2000	9/12/1920	2	9

The QCS gives warning for MPMT because probably the two records are the same tumour.

2_Patient_ID						3_Tumour_ID		
12_BoD	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	18_Vital_Status	23_Laterality
7	C501	8500	3	2	11/01/2001	30/07/1960	1	9
2_Patient_ID						3_Tumour_ID		
12_BoD	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	18_Vital_Status	23_Laterality
7	C508	8520	3	2	08/10/2002	30/07/1960	1	9

In this example, the QCS gives warning for MPMT because according to the 2004 International Rules for Multiple Primary cancer the two topographies are the same (C50), since the three first digits should be considered and the two morphologies are included in the same morphology group. In this case, only one tumour should be considered for incidence analysis.

Annex 1 – Known JRC-ENCR QCS issues and bugs

The following bugs and issues of the JRC-ENCR QCS still remain unsolved, and will be fixed at a later stage.

- The information on pT, pN, pN, cT, cN and cM provided by the Cancer Registry is not yet used to calculate TNM stage grouping. Thus, the check between stage provided by the Cancer Registry (36_Stage) and the one calculated by the software is not yet implemented.
- In case behaviour code is less than 3 (i.e. non malignant tumour) and grade code is less than 9 (i.e. grade is not unknown) the software does not report any error, while it should – with code *W-BEGR* – (see page 31 of the 2018 report "A proposal on cancer data quality checks: one common procedure for European cancer registries – version 1.1"). Please ask for technical support to the JRC (JRC-ENCR@ec.europa.eu) should you be interested in running this check.
- Code *W-ECdN* (invalid combination of extent of disease and condensed N) is present in the list of codes in the output, but is never returned by the JRC-ENCR QCS; when such combinations are present, code *W-EDCD* (invalid combination of extent of disease and condensed T, N or M) is given instead.
- Value "is" (in situ) is allowed for variable cT, whereas it should not (see page 40 of the 2018 report "A proposal on cancer data quality checks: one common procedure for European cancer registries – version 1.1").

Annex 2 – List of error and warning codes

The following is the list of error and warning codes reported in the two output files "*QCS-Incidence-Output.pdf*" and "*QCS-Incidence-Output.txt*". The page or table numbers referenced in the list are those of the 2018 update (version 1.1) of the 2014 ENCR-JRC report *"A proposal on cancer data quality checks: one common procedure for European cancer registries"*.

Error codes

E-AGED: Calculated (Date of incidence – Date of birth) in years differs from variable *11_Age* by more than one year.

E-CoDV: Date of last known vital status is not valid, e.g. when date of the incidence and date of the last known vital status are not consistent.

E-CoDA: Date of birth and date of incidence are not consistent, i.e. date of incidence occurs before date of birth.

E-DATE: Invalid date, e.g. February 29th in a non-leap year

E-ECOD: ICD[4] edition and cause of death combination are not valid, e.g. cause of death=157 (pancreatic cancer) and ICD edition=10 (the correct value for pancreatic cancer is C25 for ICD-10, and 157 in ICD-7, ICD-8 and ICD-9). The check is performed for ICD editions from 7 to 10.

E-FORM: Format error, e.g. when a character value is used when a numeric one is required.

E-MISS: Value missing, e.g. when core variable *11_Age* is unknown and cannot be calculated.

E-MOBE: Morphology and behaviour combinations are not included in the ICD-O-3.

E-OUTR: Value out of range; value is not in agreement with the ones allowed by the 2015 call for data protocol or the 2018 update (for instance, behaviour=6).

E-SETO: Sex and topography combinations are not valid (please refer to table 4 for the combinations between sex and topography considered to be unlikely).

E-VBDA: Combination of vital status, basis of diagnosis and autopsy not valid. See paragraph 3.3 *Specific checks for survival analysis* on page 40 for the allowed combinations.

[4] International Classification of Diseases (http://www.who.int/classifications/en/)

Warning codes

W-AGMT: Unlikely age and morphology/topography combination. See table 3 for the list of unlikely and rare combinations of age and tumour type.

W-BDMO: Morphology too specific according to the basis of diagnosis. See page 30 for valid combinations of basis of diagnosis and morphology.

W-BDMS: Morphology not specific enough according to the basis of diagnosis. See page 30 for valid combinations of basis of diagnosis and morphology.

W-BDMU: Basis of diagnosis and morphology/behaviour combination is unlikely. See page 30 for valid combinations of basis of diagnosis and morphology.

W-BDpM: Basis of diagnosis and pM combination is not valid. If pM is not MX and is not missing then basis of diagnosis should be 5, 7 or 6 (see page 40).

W-BDpN: Basis of diagnosis and pN combination is not valid. If pN is not NX and is not missing then basis of diagnosis should be 5 or 7 (see page 40).

W-BDpT: Basis of diagnosis and pT combination is not valid. If pT is not TX and is not missing then basis of diagnosis should be 7 (see page 40).

W-BEGR: Behaviour and grade combination is not valid. Only malignant tumours (behaviour=3) should be graded. However, the present version of the software is not returning this code (see Annex 1).

W-BTNM: Invalid behaviour and TNM combination, e.g. Behaviour=2 and pT=Tis (see page 41).

W-CdMN: Combination of Number of metastatic nodes and condensed N is not valid, i.e. Number of metastatic nodes>0 and condensed N=N0 (see page 41).

W-CdpM: Combination of TNM M and condensed M is not valid, e.g. M=0 and condensed M=M1 (see page 41).

W-CdpN: Combination of TNM N and condensed N is not valid, e.g. N=1 and condensed N=N0 (see page 41).

W-ECdN: Invalid combination of summary extent of disease and condensed N. Code W-ECdN is never returned by the JRC-ENCR QCS; when such combinations are present, code W-EDCD (see below) is given.

W-EDCD: Invalid combination of summary extent of disease and condensed T, N, or M, e.g. extent of disease=1 and condensed M=1 (see page 41).

W-EDNM: Inconsistency between Topography C809 (unknown primary site) and TNM, e.g. Topography=C809 and TNM N=0 and TNM M=0. See page 40.

W-EoDN: Inconsistency between summary extent of disease and number of metastatic nodes, i.e. summary extent of disease=1 (confined) and number of metastatic nodes>0 (see page 41).

W-EoDT: Inconsistency between topography and summary extent of disease, e.g. if topography is C809 (unknown primary site) and summary extent of disease is 1 (confined). See page 41 for the list of combinations that are returning this code.

W-ETNC: Invalid combination of summary extent of disease, cN, pN and condensed N (see page 41).

W-MOGR: Morphology and grade combination is unlikely (warning is given according to tables 6 and 7).

W-MOTO: Morphology and topography combination is unlikely (see table 8)

W-MPMT: Multiple primary malignant tumour (p. 42) The quality checklist of warnings for Multiple Primary Malignant Tumours (MPMTs) was developed by the JRC according to the current International Rules for Multiple Primary Cancers published in 2004 (http://www.encr.eu/sites/default/files/pdf/MPrules_july2004.pdf).

W-SEMO: Sex and morphology combination is unlikely, e.g. female with seminoma. See table 5 for the list of unlikely combinations.

W-TMED: Invalid combination of summary extent of disease, pM and cM, e.g. if summary extent of disease=3 (distant organs), pM=0 and cM=0 (see page 41).

W-TNED: Invalid combination of summary extent of disease, pN and cN, e.g. if summary extent of disease=1 (confined), pN=1 and cN=1 (see page 41).

W-TNMC: Combination of topography, TNM edition, cT, cN, cM and TNM stage grouping (provided by the cancer registry) is not valid, e.g. if topography=C34, TNM edition=7, cT=1a, cN=0, cM=X, stage=IA (see table 2).

W-TNMn: Inconsistency between TNM N and number of metastatic nodes (see page 41).

W-TNMP: Combination of topography, TNM edition, pT, pN, pM and TNM stage grouping is not valid, e.g. if topography=C34, TNM edition=7, pT=1a, pN=0, pM=X, stage=IA (see table 2).

W-TOLA: Topography and laterality combination is not valid according to the list of paired organs considered at page 33 of the 2018 report.

W-TOPN: Invalid combination of topography and number of examined and metastatic nodes. For certain topography codes the number of examined and metastatic nodes must be missing (either blank or 99), see page 40.

W-TOTS: Invalid combination of topography and tumour size, i.e. if topography is C80 (Unknown primary site), tumour size must be missing (either blank or 999), see page 40.

W-TTED: Invalid combination of summary extent of disease, pT and cT, e.g. if summary extent of disease=2, pT=1 and cT=1 (see page 41).

Annex 3 – Running the JRC-ENCR QCS in background

Overview

The JRC-ENCR QCS application can be run in two different modes or "moods". For the time being, the following "moods" are available:

- **GUI** (standard execution): open the main window and wait for user's actions
- **Silent** (background process): run in background and validate the file passed as argument

When executed in *silent* mode, the application accepts the following arguments:
-m=<mode> -f=<path_to_data_file> -s=<validation_schema>

Supported values are:
- **-m**: gui | silent
- **-f**: path to the file to be validated
- **-s**: incidence | lifetable | mortality | population

Warning

Some options are reserved for developing the application and MUST NOT be used by the final user:
- **-t**: index of the test to be executed
- **-c**: create the "checksum" files used to verify the integrity of the configuration

To acknowledge all options available from the command line, run the application with the **-h** option.

Sample scripts

The *samples* directory of the application contains two sample files showing examples of usage as a **background** process:
- **Run-qcs.bat**: example of executing the application in Windows OS
- **run-qcs.sh**: example of executing the application in Linux OS

Remark: the sample files listed above DO NOT provide complete management of possible execution errors, and DO NOT access (nor read, nor parse) the output reports produced at the end of the validation process. The actual management of the execution outcome MUST BE handled by the caller, with respect of his/her specific client's *execution context* (e.g. type of operative system, execution from webapp, execution as system service, etc.) and of the specific client's *needs and business* (e.g. validation of a single line, validation of big files, synchronous validation, asynchronous validation, etc.).
These sample files are provided only to show an example of executing the application as a background process and how to intercept the possible process outcomes.

Output reports

At the end of the validation process, the application should produce all output reports in path:
<application base path>/output

Guidelines

Some of the reports produced in the *output* directory are intended to be accessed directly by the final user, therefore are formatted in a human-friendly style (PDF or TXT). If the client application needs to read, parse, analyse or process the results of the validation process, usage of the following reports is recommended:

- **QCS-Incidence-FormatErrors.txt**: read this file in order to check that the input file was correctly formatted
- **QCS-output-tabular-forExcel.csv**: read this file in order to acknowledge the detailed result of the validation process, line by line. This should be the core report when the application is run as a background process
- **QCS-PrimaryDuplicates-Output.csv**: read this file in order to acknowledge the list of records which triggered a Primary Duplicate warning. This report should be ignored when the application is used to validate a single line

For details about the format of the three reports listed above, please refer to page 11 of the present user compendium.

Europe Direct is a service to help you find answers to your questions about the European Union
Free phone number (*): 00 800 6 7 8 9 10 11
(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server http://europa.eu

How to obtain EU publications

Our publications are available from EU Bookshop (http://bookshop.europa.eu),
where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents.
You can obtain their contact details by sending a fax to (352) 29 29-42758.

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub