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

JRC-ENCR QCS 2.3.1
JRC CSV Data layout converter

Changes introduced in the current document:

Manuela Flego, Stefano Adriani

Other contributors

Francesco Giusti, Carmen Martos, Antonino Brunetto, Tadeusz Dyba, Raquel N. Carvalho, Giorgia Randi, Joanna Bartnicka, Manola Bettio, Enrico Ben

JRC Cancer Information Group

Manuela Flego, Joanna Bartnicka., Cristina Bosetti, Giorgia Randi, Tadek Dyba, David Pettersson Raquel N. Carvalho, Manola Bettio

Update November 2025, QCS version 2.3

Based on ECIS Technical report (November 2025) and ECIS data call protocol (November 2025)

Contents

Acknowledgments

The authors acknowledge the following institutions and entities for the invaluable feedback provided running the previous versions of the JRC-ENCR Quality Check Software.

- East Switzerland Cancer Registry
- OmniSoftory Engineering S.A., Switzerland
- National Institute for Cancer Epidemiology and Registration (NICER), Switzerland
- National Agency for Cancer Registration (NACR), Switzerland
- Swiss Childhood Cancer Registry
- Basel Cancer Registry, Switzerland
- Bern Cancer Registry, Switzerland
- Samara Cancer Registry, Russian Federation
- Trapani Cancer Registry, Italy
- Cancer Registry of Puglia – Section of Health Local Unit Barletta – Andria – Trani, Italy
- Veneto Cancer Registry, Italy
- Castellon Cancer Registry, Spain
- Basque Country Cancer Registry, Spain
- Girona Cancer Registry, Spain
- Tarragona Cancer Registry, Spain
- Greater Poland Cancer Registry
- Polish National Cancer Registry
- Cancer Registry of Republic of Slovenia
- Morphological Tumour Registry, Luxembourg
- Cancer Registry of Norway
- Isère Cancer Registry, France
- Association of the French Cancer Registries (FRANCIM)
- Cancer Registry of Baden-Württemberg, Germany
- National Cancer Registry Ireland

Introduction

The European Network of Cancer Registries (ENCR) and the European Commission's Joint Research Centre (JRC) have launched in 2022 a data call for updating cancer burden indicators available in the European Cancer Information System (ECIS) (https://ecis.jrc.ec.europa.eu/), following an updated ECIS data call protocol.

In order to enable cancer registries to perform data quality checks and to test the adherence of their data to the protocol required format, the JRC starting from 2015 has been developing the JRC–ENCR Data Quality Check Software (QCS). Complementary to the QCS, the CSV Data Layout Converter has been released in 2021, as an additional tool to facilitate the preparation of the incidence file.

This version of the QCS is updated to the requirements of the latest ECIS data call protocol and the ECIS Technical report "Data quality checks for validation of the ECIS database". This new version benefits from the experience gathered in validation of submitted data from European registries, and from registries' feedback.

The development of the QCS is an evolving process, with regular updates that ensures alignment with European and International recommendations and classifications.

The present document provides technical guidance to the software and helps to understand and interpret its output.

The main changes and improvements included in this version compared to previous one, are detailed in ***Annex 1 – Changes and improvements from the previous versions***. A list of known bugs and issues to be addressed in a later release is available in ***Annex 2 – Known issues and future improvements***.

1 Software overview

The JRC-ENCR Quality Check Software (QCS) is an open-access software developed to facilitate standardisation and validation of population-based cancer registries (PBCRs) data.

The QCS has been designed as a stand-alone desktop application that can run locally by cancer registries, without the need of internet connection.

The majority of the checks performed by this version are based on latest version of the ECIS Technical report, which includes recommendations for checking internal consistency of the cancer incidence data and provides a comprehensive and standardised list of data quality checks to be adopted by European PBCRs.

The software checks the four different data files (incidence, mortality, population and life tables) required by the ECIS protocol. The QCS output consists of four files (in PDF, TXT and CSV format) containing the detailed list of warnings and errors detected in the validation run for each record of the input data file.

Population, life tables and mortality files have a simple structure, which require univariate checks validation (checks on single variables). Incidence files are more complex, and in addition to univariate validation the QCS performs checks between variables of the same record (multivariate checks) and between variables of different records (duplicates and multiple primary tumours checks).

2 System requirements and installation

The QCS has been initially developed for Windows operating systems that support Java (Windows 10 and above).

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

2.1 In case Java software is not installed on your computer

Java software is needed to run the QCS. In case Java is not installed on your computer, please install Java JDK version 17 or later, otherwise go to **section 2.3**.

2.2 Further information and troubleshooting related to Java

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

Remember to choose the correct version for your operating system (ex: Windows 32 bit or Windows 64 bit).

2.3 How to install the QCS

Download the latest version of the QCS software from the JRC-ENCR website and extract the files on your computer in a dedicated folder. Please follow the instructions below according to your PCs operating system.

2.4 Running on Windows

1. Open the folder containing the QCS
2. Double click on the "JRC-ENCR-QCS-2GB.bat" file (suggested option). If the QCS does not start or does not work properly, double click on the "JRC-ENCR-QCS.bat"

2.5 Running on macOS

1. Double click the Terminal icon (or label, depending by your view settings) to open a Terminal window
2. 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-version zip" (where *version* is the current software version, e.g. *version* = 2.2), then you should execute the command like below:

   cd Desktop/JRC-ENCR-QCS-version

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

   source jrc-encr-qcs.sh

2.6 Running the QCS on Linux operating systems

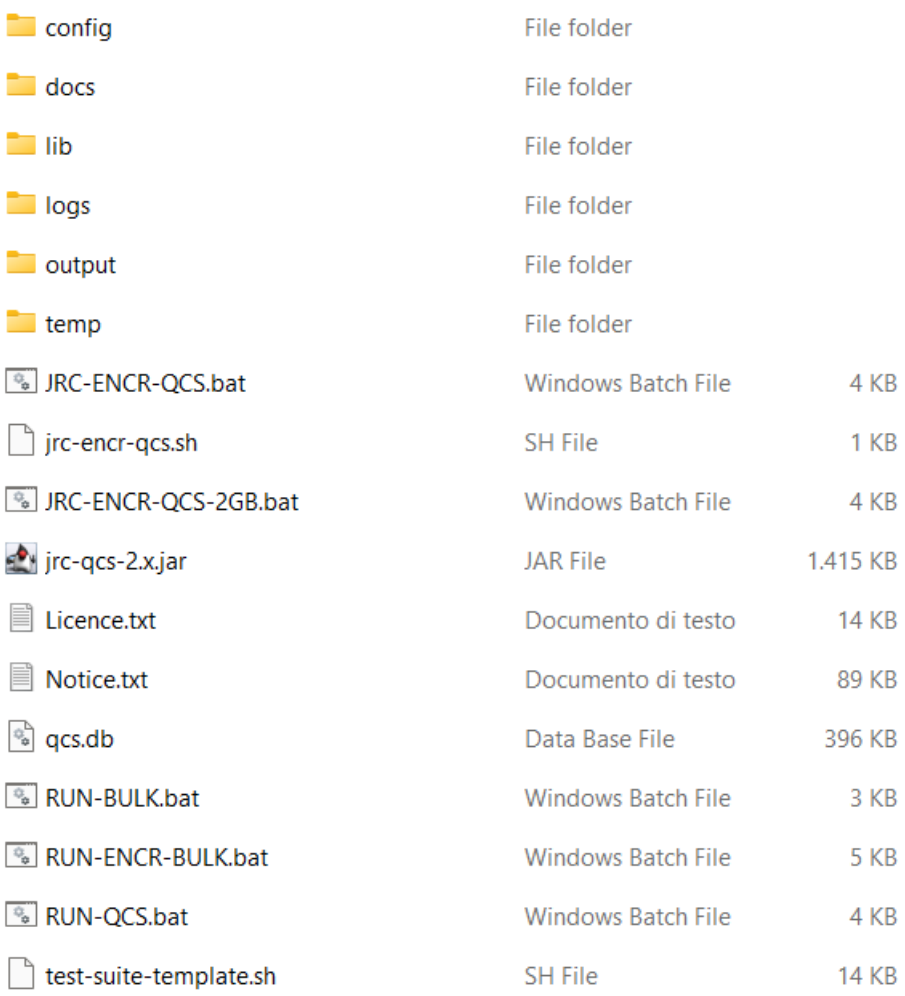
1. Move into the directory containing the QCS
2. Execute the QCS by running the ".sh" file:

   source start-jrc-encr-qcs.sh

2.7 Verify the correct installation

Navigate to the folder where you extracted the software and before running it (as explained in section 5.1), check if it is correctly installed.

The expected directory structure is shown below:



The folder *JRC-ENCR-QCS-V2.x* (where 2.x is the current version) includes the following files and folders:

- The License.txt file containing the usage license
- The executable files *JRC-ENCR-QCS.bat*, *JRC-ENCR-QCS-2GB.bat* for all JRC external users;
- The executable file *RUN-QCS.bat* for JCR internal users only;
- The executable files *RUN-BULK.bat* (for JCR internal users only) and *RUN-ENCR-QCS.bat*;
- Files *jrc-encr-qcs.sh* and *test-suite-template.sh*;
- The application's file *jrc-qcs-2.x.jar*;
- The QCS database *qcs.db*;
- Installation folders: *config*, *docs, lib* (installed with the QCS)*;*
- Runtime folders: *logs, output, temp* (created at first usage);

The file *JRC-ENCR-QCS.bat* will run the QCS with 1GB of RAM memory allocated by Java, whereas *JRC-ENCR-QCS-2GB.bat* will use 2GB of RAM memory.

The file *RUN-BULK.bat* will run the QCS as a background service (see detailed instructions in the manual included in the doc folder);

The file *jrc-encr-qcs.sh* is the standard ".sh" script for running the application on the macOS and Linux systems (see sections 2.5 and 2.6 above).

config: this folder contains configuration files, such as those for multivariate intra records checks, and general application settings (with the possibility to disable some functionalities) etc.

docs: this folder contains relevant documentation files in PDF format, i.e. the QCS user compendium (this document) and the manual "Run QCS as background service.pdf". The file *change_log.txt* is also included with the list of the most recent QCS versions and updates. The subfolder "ex*amples"* includes some sample scripts for advanced users (see *Annex 4 – Running the JRC-ENCR QCS in background*).

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

logs: this folder stores the log file of all the QCS activities.

output: this folder is created after the QCS is run for the first time. It includes four different subfolders with the output reports produced by the QCS for *Incidence, Mortality, Population* and *LifeTable* files respectively.

temp: this folder is created after the QCS is run for the first time. It contains all the raw files (working files) used to generate the final output reports.

3 How to prepare an input file for the QCS

In this section an example of input files accepted by the software is given. Input files should be formatted as follows:

- as text files (.csv or .txt) with semicolon (;) separators only;
- with first line as header.

3.1 Incidence File

The input incidence file must follow the format of the ECIS data call protocol. It can be either created following the instructions below (section 3.1.2) or using the JRC CSV Data layout converter (section 3.1.3).

3.1.1 The ECIS data call protocol

Please refer to the ECIS data call protocol, section 3.1.7 and Table 1 in which name, description and format of requested variables, with corresponding missing/unknown values and coding schema are summarized.

3.1.2 Incidence file creation

As a first step, the file header needs to be created with 39 variables, as per ECIS data call protocol. The incidence file has a fixed structure as for names, variables' order and semicolon (**;**) as column separators. As a consequence, the header line should be the following:

PAT; MoB; YoB; Age; Sex; Geo_Code; Geo_Label; TUM; MoI; YoI; BoD; Topo; Morpho; Beh; Grade; Autopsy; Vit_stat; MoF; YoF; Surv_time; ICD; CoD; TNM_ed; cT; cN; cM; pT; pN; pM; ToS; Stage; Surgery; Rt; Cht ; Tt; It; Ht; Ot; SCT

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

After the creation of the header line, you can populate your incidence file with records (rows) corresponding to incidence cases each. When you finish populating, please save it in csv or txt format and this will be your incidence file ready to be run by the QCS.

The paragraph below explains how to use the CSV Data layout converter tool if you need assistance in preparing or checking the incidence file format before running the QCS.

3.1.3 JRC CSV Data layout converter

The JRC CSV Data layout converter tool assists users (with Windows operating systems) in the creation of incidence files to be run by the QCS.

To use the tool, please select a data file to import and convert into the proper format. The file can be in any text format (CSV, TXT) with allowed supported columns separators as follows:

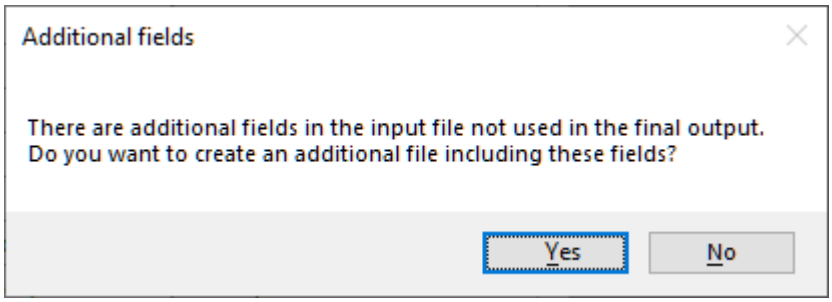
- **TAB** (tabulation)
- **|** (pipe)
- **,** (comma)
- **;** (semicolon).If the incidence file is formatted correctly, the converter tool maps the variables as defined in the protocol (Required protocol fields) with the corresponding variables of the input data file (Available fields).To facilitate the data import, the converter tool tries to automatically map fields with the same name: mapped variables are displayed in GREEN, while unmapped variables are marked in RED.

When automatic mapping is not possible, the user can:

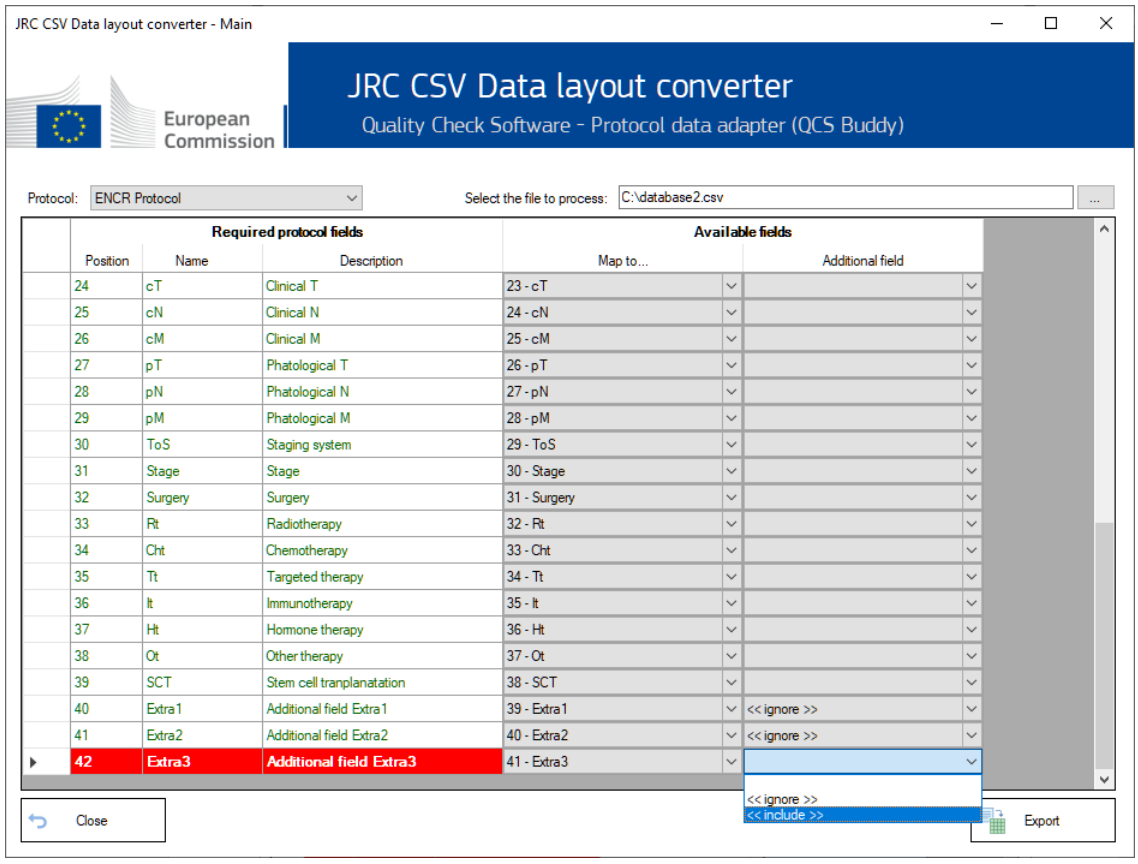
- either map manually the variables;
- or leave the field blank (for example, if no mapping is possible, data are not available, etc.).

Only when all the fields defined in the protocol are mapped (or blank), it is possible to Export the content of the original file in the format required by the ECIS protocol.

For variables in the data file not relevant to the ECIS protocol, the tool enables exporting an additional file including one or more of these variables in addition to those required by the ECIS protocol.



In the example above, variables "Extra1", "Extra2" and "Extra3" are not relevant for the ECIS protocol. Among these, "Extra3" (marked as to be included) will be kept in the export file.



3.2 Mortality file

Similarly to incidence files, the creation of the file header is the first step. For mortality files, the number of required variables is 5.

Depending on the reporting of deaths by single years of age (Age unit) or by age range (Age range), the line header for the mortality file is the following:

Calendar_Year;Sex;Age unit;Cause of death;Number of deaths

Calendar_Year;Sex;Age range;Cause of death;Number of deaths

Please note: make sure that the 5 variables are in the correct order, and separated by semicolons (;). The header line is mandatory. Do NOT put a semicolon at the end of header.

After having created the header, you can populate your mortality file. Once done, please save it in csv or txt format. You are now ready to run the mortality file by the QCS.

3.3 Population file

Please create first the header of the file. For population files, the number of required variables is 6.

Depending on the reporting of deaths by single years of age (Age unit) or by age range (Age range), the line header for the population file is the following:

Calendar Year;Sex;Age unit;Geo_code;Geo_label;Number of residents

Calendar Year;Sex;Age range;Geo_code; Geo_label;Number of residents

Please note: Please make sure that the 6 variables are in the correct order and separated by semicolons (;). The header line is mandatory. Do NOT put a semicolon at the end of the header.

After having created the header, you can populate your population file. Once done, please save it in csv or txt format. You are now ready to run the population file by the QCS.

3.4 Life Table file

Please create first the header of the file. For life table files, the number of required variables is 6.

The line header for the life table file is the following:

Calendar Year;Sex;Annual age;Geo_code; Geo_label;All causes death probability

Please note: Please make sure that the 6 variables are in the correct order and separated by semicolons. The header line is mandatory. Do NOT put a semicolon (;) at the end of each line.

After having created the header, you can populate your life table file. Once done, please save it in csv or txt format. You are now ready to run the population file by the 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 input data files: *mortality*, *population* and *life table* files.

The incidence file should include 39 variables, semicolon-separated (;), and in the correct format as described in sections 3.1.1 and 3.1.2 above.

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 ECIS protocol as for:

- the number of variables;
- the presence of non-missing and non-blank values in the fields affecting incidence calculation (if year of incidence (variable *YoI*) is missing, the software rises the error E-MISS etc.);
- the field content against a list of valid values, when applicable. **Example**: for patient's sex (variable "Sex") only four numeric values are allowed: 1=male, 2=female, 3=other or 9=unknown. Other values will produce an error (sex=4 will produce E-OUTR, value out of range; sex=X or 11 will produce E-FORM, format error, etc.);
- the field length, which must be within the allowed range. **Example**: the maximum length for Patient identification code (variable PAT) is 50 characters; the maximum length for Morphology (variable *Morpho*) is 4 digits (*Morpho*=82633 will produce E-FORM, format error, etc.);
- the validity of dates, including checking that some dates are not set in the future. **Example**: YoI=2040 will produce E-OUTR, value out of range, being the current year the maximum allowed value;
- records failing the edits described in the ECIS Technical Report.

All warning and error messages produced by the QCS in each validation run are saved in three specific output files (as an example, the names below are relative to the *incidence* file):

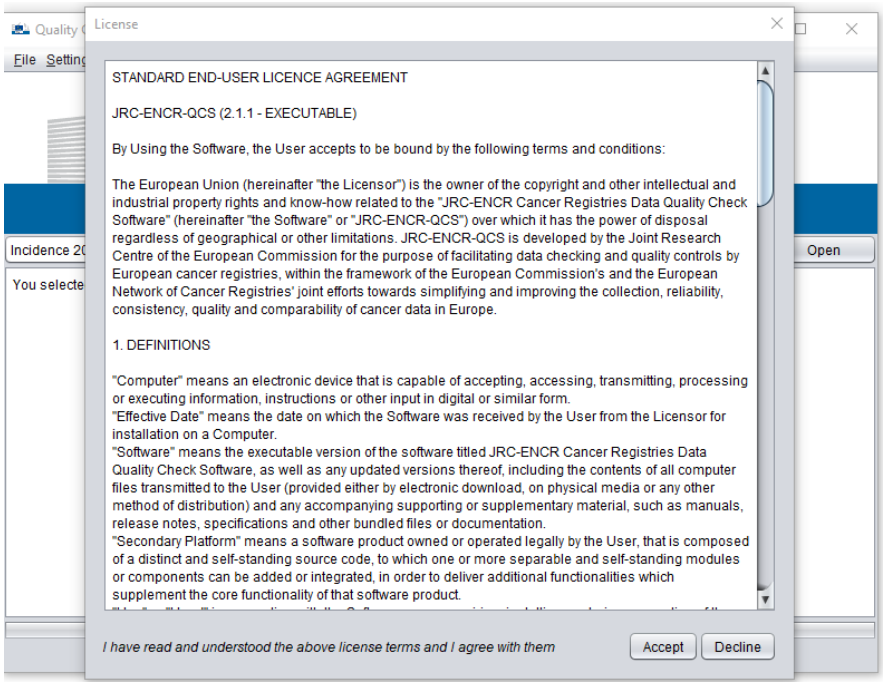
1) *QCS-Incidence-Output.pdf* – contains a summary of the execution process (date and time, name of the processed file, number of rows, total number of errors and warnings) and the detailed list of all errors and warnings detected for each record of the input data file. Please note that if the QCS output file "QCS-Incidence-Output.txt" is too large, the corresponding pdf file will not be produced; in this case, if in the output folder there is already an old PDF file produced in a previous run, it will be temporarily renamed "*_old", until a new PDF is produced that overwrites it.
2) *QCS-Incidence-Output.txt* – has the same content of the PDF file, but in text format;
3) *QCS-Incidence-Output.csv* – contains all errors and warnings in a format easily readable by automated procedures. This file can be used to load the results of the validation process in a database, or to perform detailed statistics and analyses. It can be imported by most software packages to allow for advanced data manipulation, such as linkage with the original file using the unique combination of patient/tumour ID. Warnings for multiple primary tumours are also included in this file. This format can be particularly useful if the input file generates a large number of errors.

5 Using the software

5.1 Running the software

- Please navigate to the folder in which you installed the software;
- Double click on the *JRC-ENCR-QCS.bat* file to launch the software (in case of any issue, try to run the QCS with 2GB of RAM memory by launching the file *JRC-ENCR-QCS-2GB.bat*); please note that the software runs only double clicking on the file ending in .bat;

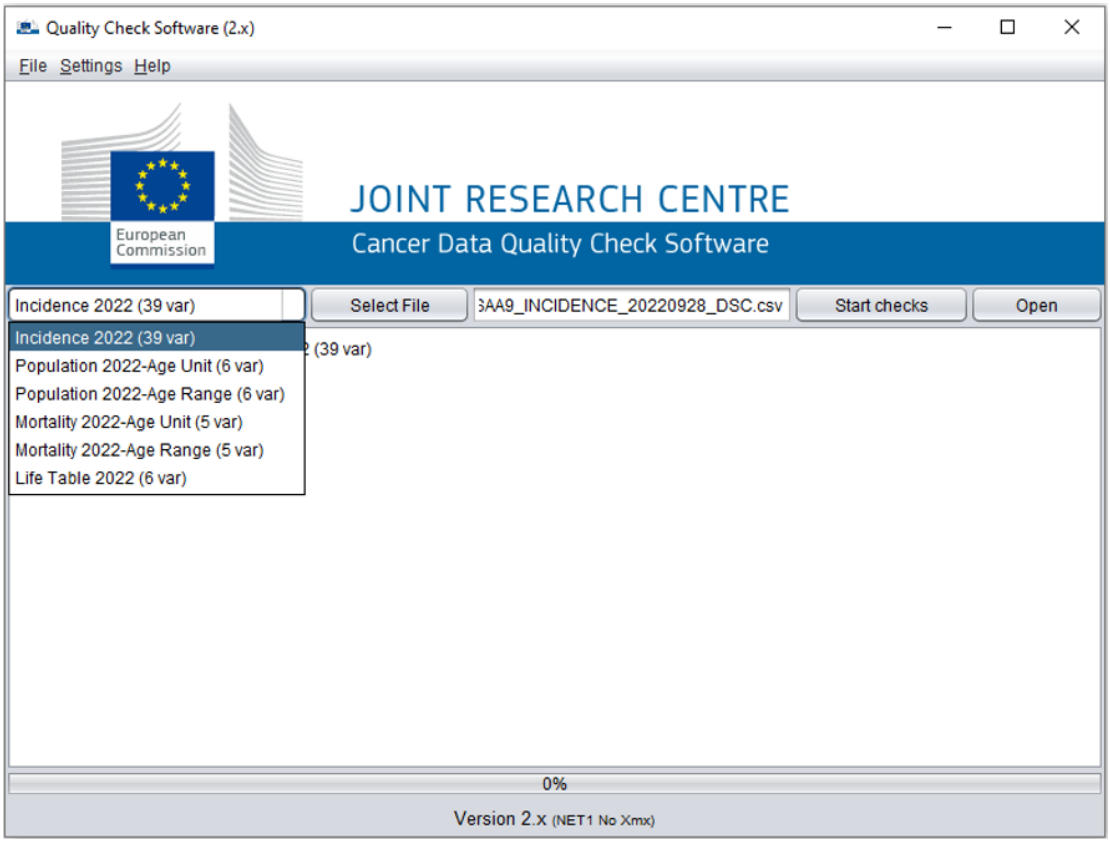
  The first time the application is launched a warning window opens for acceptance of the software license. If the license is not accepted, the application will close. Launch the application again and accept the license to run the software.



- To quit the QCS just close the window or select "Exit" in the File menu.

5.2 Checking the files

From the drop-down menu select the type of file you want to check (Incidence, Population etc.), as shown below.

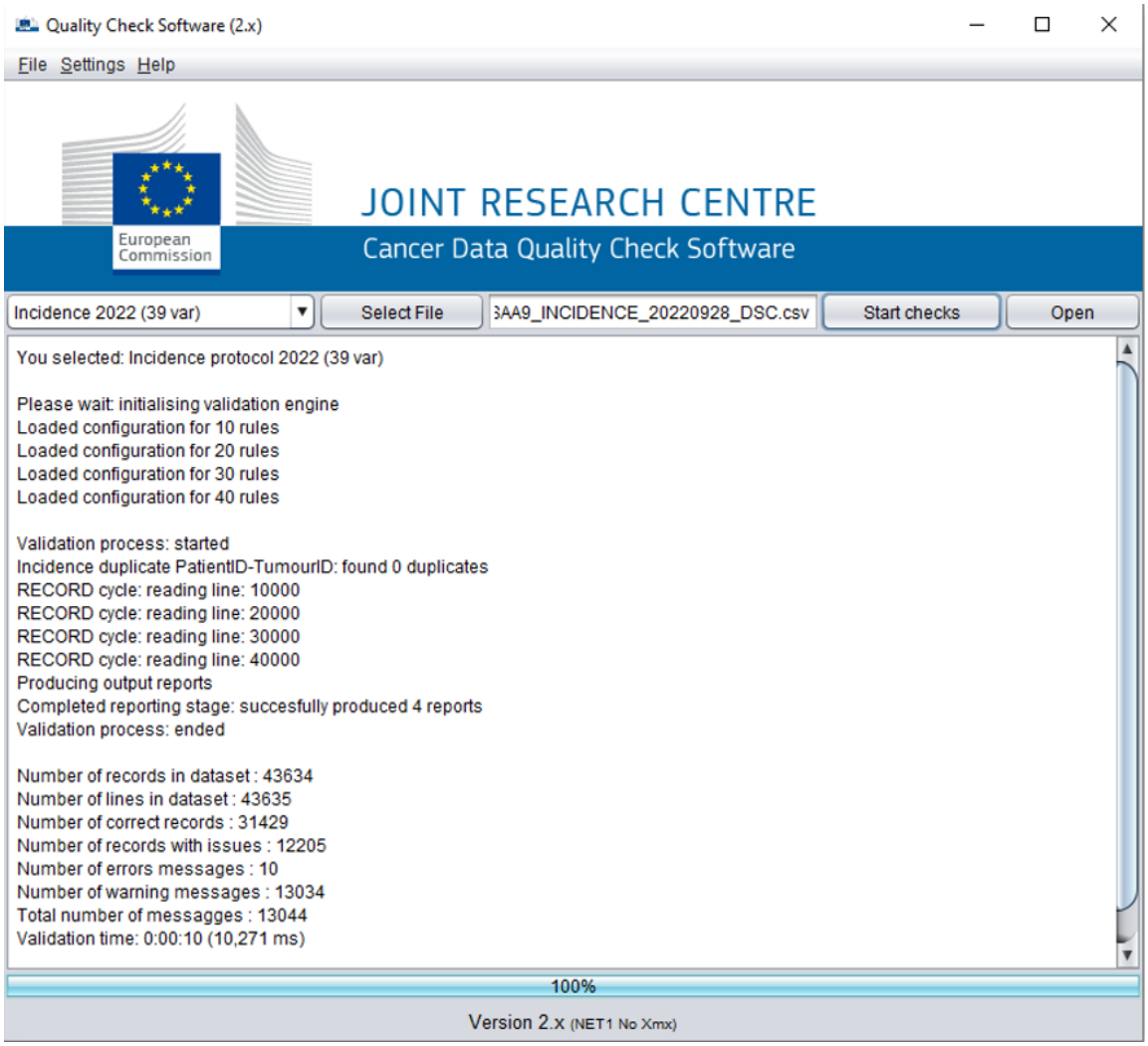


For checking the incidence file according to the ECIS data call protocol:

- Select the "*Incidence 2022 (39 var)*" option from the drop-down menu on the top left corner;
- Press the "*Select File*" button;
- A file browsing window will appear;
- Navigate to the folder where the file is located, select it (only files with extension csv or txt are accepted) and choose "*Open*" in the browsing window. The full path of the selected file will be displayed in the text box next to the "*Select file*" button;
- Press the "*Start Checks*" button to start the validation process.

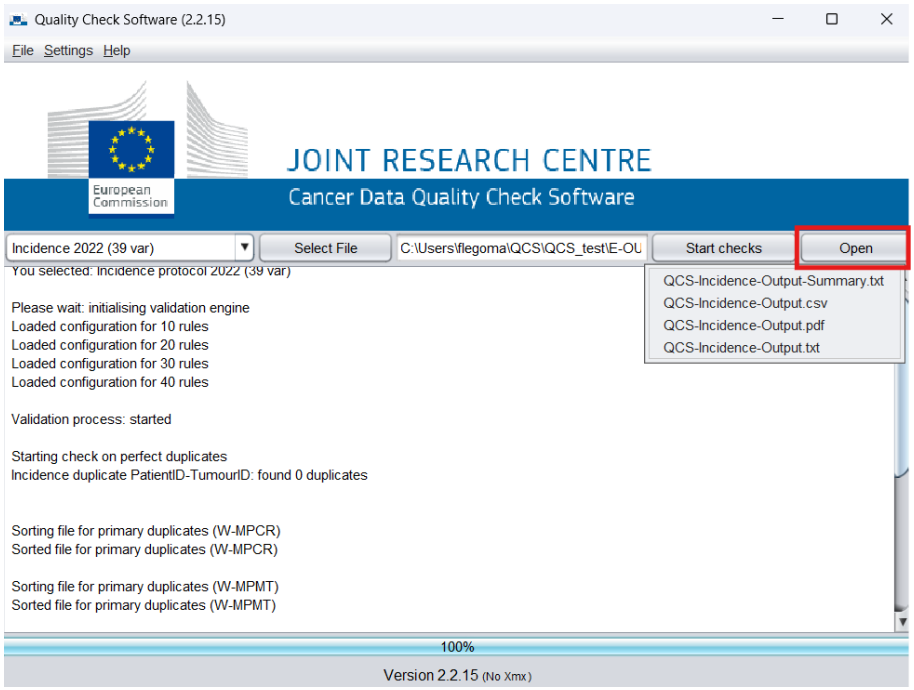
Please note that in each validation run the output files **will be overwritten**. To keep different QCS output files, please save in different folders or with different names.

While the validation process is running, the number of checked records keeps updating in the output window:



Once the validation process is completed, the output window displays a short summary report which includes the total number of checked records and the number of errors and warning messages.

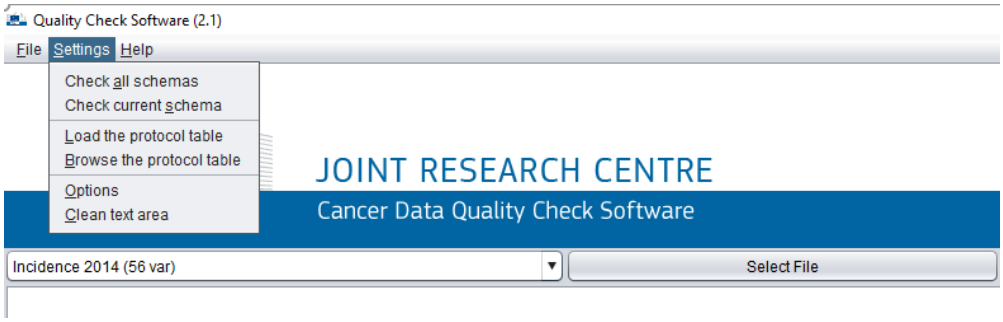
You can directly access the output files (CSV, PDF, TXT) by selecting the "*Open*" tab in the user window (see screenshot below). Click on a specific file to open it. These files are saved in the "*output/incidence*" folder created at the end of the validation process.



The procedure for checking *mortality*, *population* and *life table* files is the same as described above for *incidence* files.

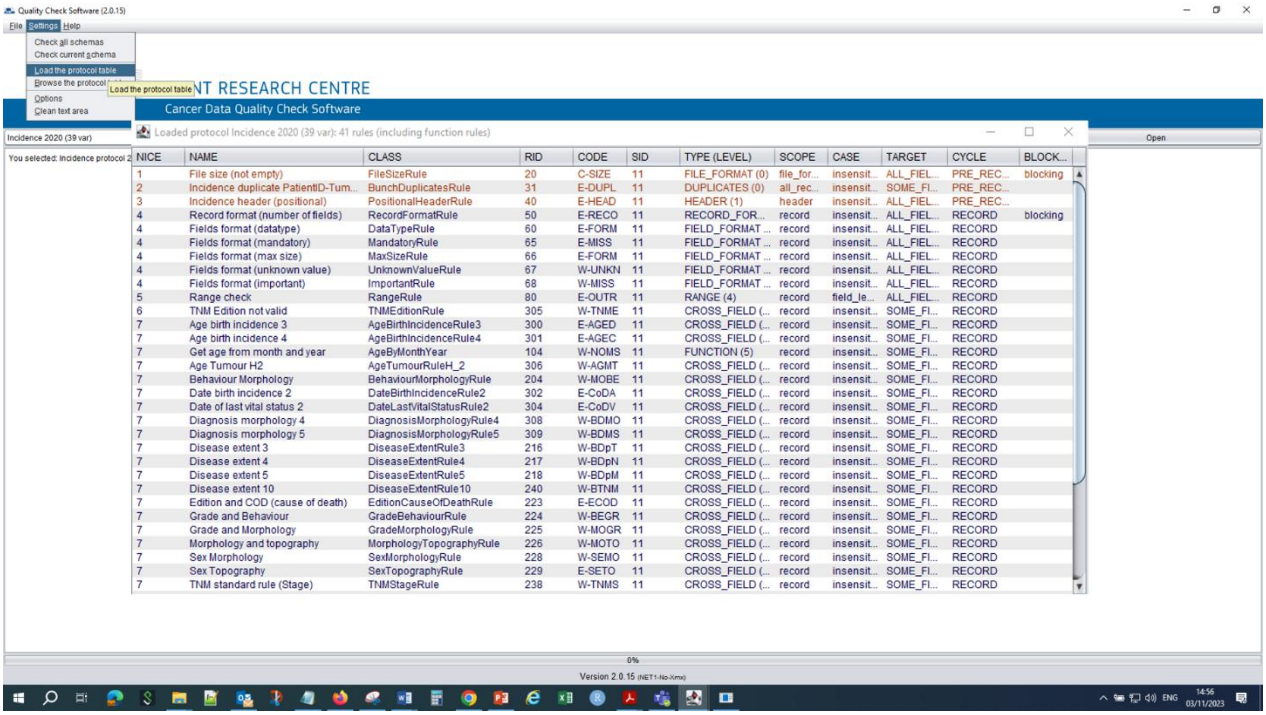
5.2.1 Settings and options

The "Settings" menu enables to select additional QCS functionalities.

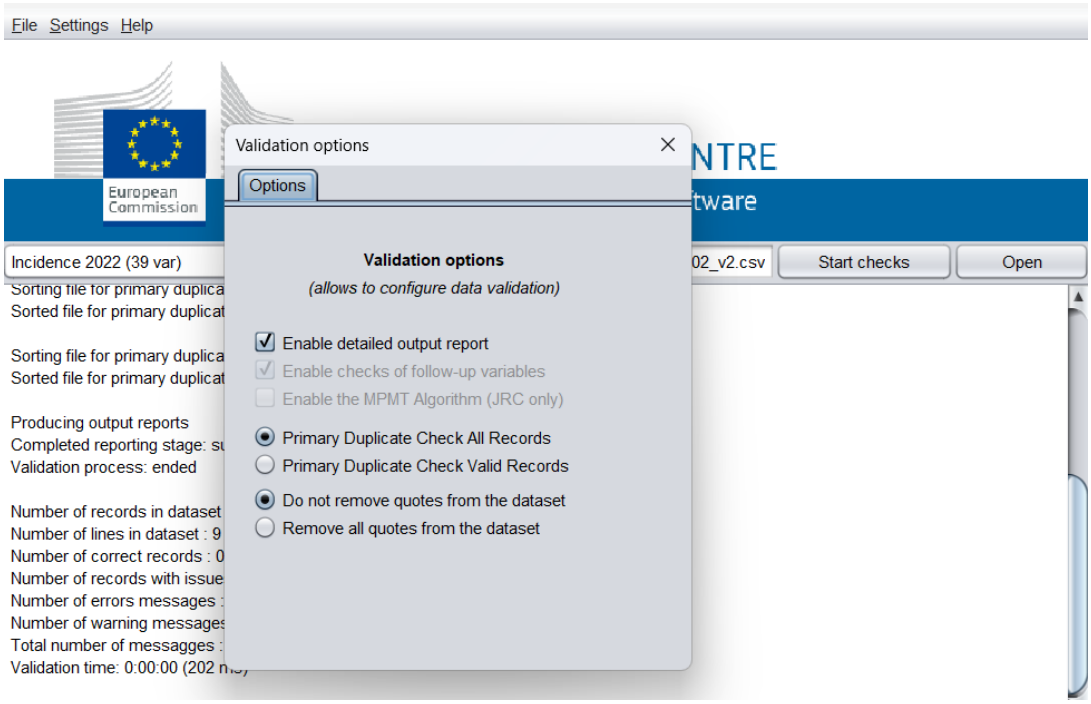


The following settings are available:

- *Check all schemas/Check current schema*. This functionality checks the existence of configuration files, the integrity of single files, the integrity of configuration files and returns the integrity status of either all schemas or the current schema;
- *Load the protocol tables/Browse the protocol table.* This functionality allows to load or browse the protocol table, listing all the protocol rules applied according to the selected protocol (see screenshot below);



- *Options*. When selected, all validation options are shown. The options active for general users (outside the JRC) are listed below:
  - *Enable detailed output report* allows the creation of either a detailed or aggregated report. A detailed report is created as the default option.



  - *Primary Duplicate Check All Records/Valid Records*. These options allow to have different conditions for checking multiple primary tumours:

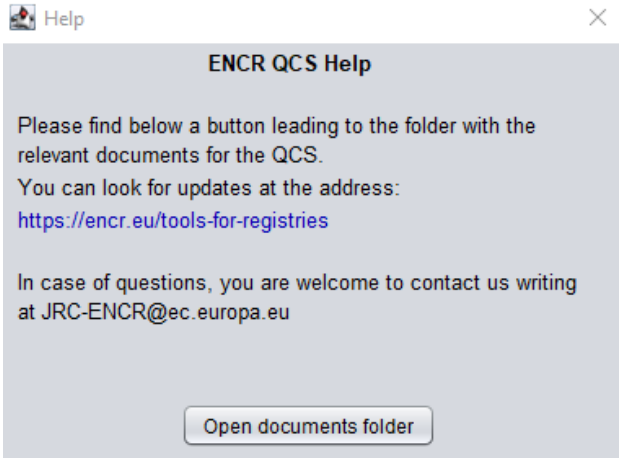
    With *Primary Duplicate Check All Records* this check is performed on all records except those with errors E-MISS (Value missing), E-FORM (Format error) and E-OUTR (Value out of range) on the morphology variable (Morpho).

    The option *Primary Duplicate Check Valid Records* is more selective because multiple primary tumours checks are performed only on records without the following errors/warnings: E-SETO (Topography + Sex not valid), E-AGED (DoI - DoB different from Age), E-AGEC (Age is invalid + impossible to calculate age from DoI – DoB), E-CoDA (DoB + DoI not coherent), W-AGMT (Unlikely Age + tumour type), W-MOTO (Morphology + Topography not valid) and errors as E-MISS, E-FORM and E-OUTR involving the variables topography (Topo) and morphology (Morpho) (see Annex 3 for the definition of error and warning codes)
  - *Remove all quotes from the dataset*. These options allow to remove double quotes characters ("") accidentally introduced into the dataset that could cause problems with the QCS. The default configuration is *Do not remove quotes from the dataset.*

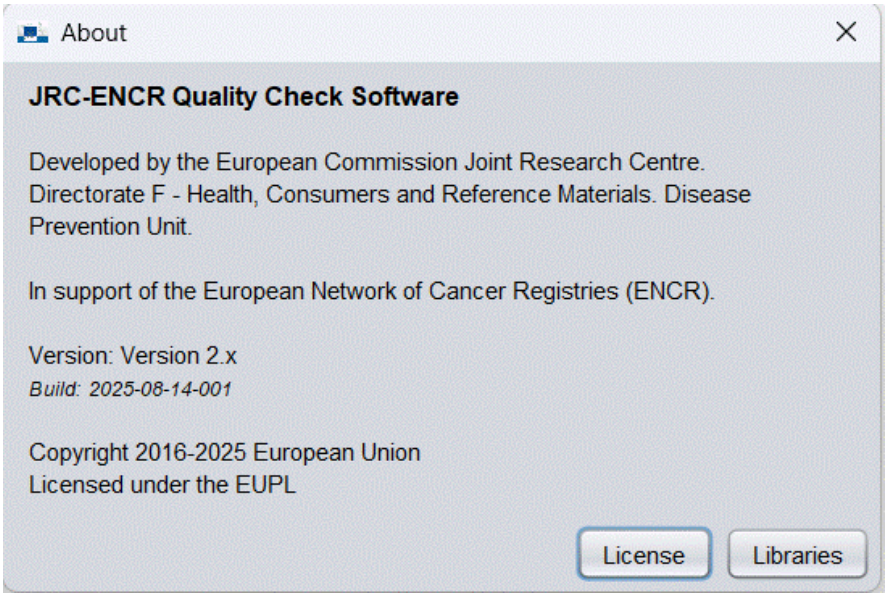
- *Clear text area* in the *Settings* menu deletes all text from the dialog box.

5.2.2 Help menu

This functionality includes a link to the folder with information on the QCS, a contact e-mail and the link to the QCS page on the ENCR website.



In the "*Help*" menu you can also find the "*About*" item, with copyright statement, the license and the list of *jar* libraries.



5.3 Output files

Each validation run generates four output files in PDF, TXT and CSV format. The output files are located in the related subfolders for *incidence*, *population*, *mortality* or *life table*, inside the "*output*" folder. For example, output files for incidence are located in the subfolder "**\JRC- ENCR-QCS-V2.x\output\incidence"**. Please note that if the QCS output file "QCS-Incidence-Output.txt" is too large, the corresponding pdf file will not be produced; in this case, if in the output folder there is already an old PDF file produced in a previous run, it will be temporarily renamed "*_old", until a new PDF is produced that overwrites it.

5.3.1 Details of pdf/txt output file

The file "QCS-Incidence-Output-Summary.txt" contains only a summary of the execution process (date and time, name of the processed file, number of rows, total number of errors and warnings). The other pdf/txt files include also the detailed list of all errors and warnings detected for each record of the input data file.

The following screenshots refer to the different sections of the *QCS-Incidence-Output.txt* (or *pdf*) file.

Detail of upper section

```
**********************************************************************************************************

QUALITY CHECK SOFTWARE REPORT - INCIDENCE

**********************************************************************************************************

**********************************************************************************************************

PROCESSING PARAMETERS

**********************************************************************************************************

File process start : 2023-10-30 11:0:47.748
File process end   : 2023-10-30 11:0:54.749

Validated by       : QCS Version 2.1 (run 1698660055579)

File Processed:
U:\1b-Cancer Information\QCS\jrc-qcs-2.1\Test.csv (CC89315B5D)

Validation options:

Validation protocol ID                 : 11
Name of validation protocol            : Incidence 2020 (39 var)
Print detailed TXT report              : true
Use reference data                     : true
Strip all quotes from dataset          : false
Print cross-record matching info       : true

ENCR Incidence options:

Acceptance MP criterion                 : 2
Age-Tumour rule (W-AGMT)                : hybrid 2023 (obj 558)
Apply MPMT merging algorithm            : false

**********************************************************************************************************

PROCESSING STATISTICS

**********************************************************************************************************

Number of records read                 : 25573
Total number of errors                 : 246
Total number of warnings               : 7273
Total number of records with issues    : 6552
Total number of rejected records       : 0
Preliminary rules                      : nothing critical

**********************************************************************************************************

KEY TO ERROR AND WARNING CODES

**********************************************************************************************************

E-AGEC: Age is invalid + impossible to calculate age from DoI - DoB
E-AGED: DoI - DoB different from Age
E-AUVS: Autopsy+Vital status+Surv time+Date of Incidence/Date of Followup not valid (Tab. 12)
E-BDVS: BoD+Vital status+Surv time+Date of Inidence/Date of Followup not valid (Tab. 11)
E-CoDA: DoB + DoI not coherent (p.16)
E-CoDV: Date of last known vital status not valid
```

Processing parameters:

- *File process start, File process end*;
- *Validated by*: The QCS version used to produce the report;
- *File processed*. The name and the path of the file checked by the software.

Processing statistics:

- *Number of records read, Total numbers of errors*, *Total number of warnings*;
- *Total number of records with issues*: each record can have more errors and/or warnings;
- *Total number of records rejected*: records are rejected whenever the header is correct but some of the variables are not present, not even left blank or with missing value.

Key to error and warning codes:

- Errors and warnings are referenced by codes, described by short labels and accompanied by the reference to the relevant table or page from the ECIS Technical Report "Data quality checks for validation of the ECIS database". See also Annex 3 – List of error and warning codes.

Detail of second part (summary of errors and warnings)

SUMMARY OF ERRORS BY CODE

E-AUVS	7
E-BDVS	239

SUMMARY OF WARNINGS BY CODE

W-AGMT	20
W-BDMO	236
W-BDMS	43
W-BDpM	1415
W-BDpN	123
W-BDpT	140
W-BEGR	282
W-BTNM	45
W-MOGR	718
W-MOTO	102
W-MPCR	18
W-MPMT	20
W-TBGR	45
W-TNMM	511
W-TNMS	3555

Summary of errors warnings by code: see *Annex3 – List of error and warning codes*.

Detail of Multiple Primary Tumours checks: W-MPCR and W-MPMT sections

The QCS 2.x performs MPMTs quality checks for computing incidence according to the 2004 International Rules for Multiple Primary Cancers, also referenced in the ECIS Technical Report (Chapter 4).

In addition to checks on multiple primary tumours with behavior 3 (malignant), the QCS 2.x performs also checks on tumours with behavior <3 using new and more precise filtering criteria. For example, multiple primary tumours with topography codes C50 or C65-C67 and behaviour 2 (in situ), or tumours with codes C70-C72, C751-C753 and behaviour 0 (benign) or 1 (uncertain) and specific morphologies, are also checked.

These filtering criteria are detailed in the configuration file located in the QCS 2.x subfolder "…\jrc-qcs-2.x\config\rules\encr2014\PrimaryMultipleFilter.csv".

```
**************************************************************************************************************
MULTIPLE PRIMARY MALIGNANT TUMOUR CHECK (W-MPCR)
**************************************************************************************************************

------------------------------------------------------
 PAT  234281                                 Tum  572373
 Duplicated record:                          Tum  572375
------------------------------------------------------
 BoD  Topo  Morpho        Beh  Sex  DoI         DoB
------------------------------------------------------
 2    C619  8000          3    1    12/2008     9/1924
------------------------------------------------------
------------------------------------------------------
 PAT  234281                                 Tum  572375
 Duplicated record:                          Tum  572373
------------------------------------------------------
 BoD  Topo  Morpho        Beh  Sex  DoI         DoB
------------------------------------------------------
 2    C343  8000          3    1    1/2012      9/1924
------------------------------------------------------
------------------------------------------------------
**************************************************************************************************************
MULTIPLE PRIMARY MALIGNANT TUMOUR CHECK (W-MPMT)
**************************************************************************************************************

------------------------------------------------------
 PAT  135639                                 Tum  165025
 Duplicated record:                          Tum  426753
------------------------------------------------------
 BoD  Topo  Morpho        Beh  Sex  DoI         DoB
------------------------------------------------------
 7    C509  8503          2    2    4/2007      8/1936
------------------------------------------------------
------------------------------------------------------
 PAT  135639                                 Tum  426753
 Duplicated record:                          Tum  165025
------------------------------------------------------
 BoD  Topo  Morpho        Beh  Sex  DoI         DoB
------------------------------------------------------
 7    C509  8480          3    2    9/2008      8/1936
------------------------------------------------------
------------------------------------------------------
```

For each MPMT warning the following variables are reported: PAT, Tum, BoD (basis of diagnosis), Topo (topography), Morpho (morphology), Beh (behaviour), Sex, DoI (date of incidence), DoB (date of birth).

The two MPMT output sections list all multiple primary tumours detected by the QCS 2.x and reports two types of warnings, **W-MPCR** or **W-MPMT**, according to some special conditions defined for subgroups of topography/morphology codes.

- The warning **W-MPMT** applies, under certain conditions, only to subgroups of multiple primary solid tumours with specific topography codes:
  - tumours diagnosed in paired organs (e.g. C50, C62, C69 etc.);
  - tumours which occur in some 4th character subcategory of colon (C18.0-C18.7) and skin (C44.0-C44.7);
  - tumours with different first 3 digits in the site code, belonging to groups of topography codes considered as a single site according to Table 14 of the ECIS Technical report (e.g. C01, C02; C00, C03-C06; C65-C68 etc.).

All equivalent tumours (same topography and morphology) with **W-MPMT** warnings in the

same patient should be counted only once for incidence analysis, according to the 2004 International Rules for Multiple Primary Cancers and the ENCR recommendations as referenced in the ECIS Technical report. See some examples at point 6) of Chapter 6.

- The warning **W-MPCR** applies to all tumours not belonging to the subgroups above, e.g. solid tumours with different topography codes such as C61 etc. and all hematological malignancies regardless of topography code.

  Equivalent tumours with **W-MPCR** warnings in the same patient should be further checked by the cancer registry to decide if only one should be considered for incidence analysis. See some examples at point 6) of Chapter 6.

Detail on section with errors and warnings

```
**********************************************************************************************************
ERRORS AND WARNINGS
**********************************************************************************************************

--------------------------------------------------------------------------------------------------------
PAT  208387                              Tum  248485
--------------------------------------------------------------------------------------------------------
BoD      Topo     Morpho   Beh   Sex   DoI       DoB       Var_Name        Var_Value  Error_Code
--------------------------------------------------------------------------------------------------------
7        C710     9441     3     1     1/2008    2/1937    Surgery         9          W-SUpT
                                                           pT              x          W-SUpT
--------------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------------------
PAT  521016                              Tum  655821
--------------------------------------------------------------------------------------------------------
BoD      Topo     Morpho   Beh   Sex   DoI       DoB       Var_Name        Var_Value  Error_Code
--------------------------------------------------------------------------------------------------------
7        C492     8920     3     2     11/2012   4/2008    Surgery         9          W-SUpT
                                                           pT              x          W-SUpT
--------------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------------------
PAT  1388134                             Tum  1514686
--------------------------------------------------------------------------------------------------------
BoD      Topo     Morpho   Beh   Sex   DoI       DoB       Var_Name        Var_Value  Error_Code
--------------------------------------------------------------------------------------------------------
7        C221     8000     3     2     11/2018   2/1949    Topo            C221       W-TNMM
                                                           Morpho          8000       W-TNMM
                                                           TNM_ed          8          W-TNMM
                                                           Stage           3B         W-TNMM
                                                           pT              x          W-TNMM
                                                           pN              x          W-TNMM
                                                           pM              0          W-TNMM
                                                           cT              2          W-TNMM
                                                           cN              1          W-TNMM
                                                           cM              0          W-TNMM
                                                           ---------------------------------
                                                           Surgery         9          W-SUpT
                                                           pT              x          W-SUpT
--------------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------------------
```

Errors and warnings: key variables are reported for each error and warning, namely: *PAT*, *Tum*, *Topo* (topography), *Morpho* (morphology), *Beh* (behaviour), Sex, *DoI* (date of incidence), *DoB* (date of birth), *Var_Name* and *Var_Value* (list of relevant variables which caused the warning/error with the corresponding name and value), *Error_Code* (code according to list in *Annex 3 – List of error and warning codes*).

5.3.2 Details of csv output files

The csv output file contains all errors and warnings in a format easily readable by automated procedures, which can be used to perform detailed statistics and analyses.

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

Detail of the left part of the csv file

Line	Patient	Tumour	BoD	Topo	Morpho	Beh	Sex	DoI	DoB
4	1388134	1514686	7	C221	8000	3	2	Nov-18	Feb-49
6	2709	2726	5	C424	9836	3	2	Aug-04	Dec-93
9	2778	2793	5	C424	9836	3	1	Apr-05	Nov-99
14	2972	2975	5	C424	9872	3	1	Sep-05	Sep-02
23	3222	3208	7	C480	9500	3	1	Jun-06	Sep-03
24	3223	3209	5	C424	9874	3	1	Jan-06	Jul-92

Detail of the right part of the csv file

Error_Cod	Error_Description	Var1_Nam	Var1_Valu	Var2_Nam	Var2_Valu	Var3_Nam	Var3_Valu
W-TNMM	Morphology not addressed by the Topograp	Topo	C221	Morpho	8000	TNM_ed	8
W-MOTO	Morphology + Topography not valid (tab.8)	Morpho	9836	Topo	C424		
W-MOTO	Morphology + Topography not valid (tab.8)	Morpho	9836	Topo	C424		
W-MOTO	Morphology + Topography not valid (tab.8)	Morpho	9872	Topo	C424		
W-TNMM	Morphology not addressed by the Topograp	Topo	C480	Morpho	9500	TNM_ed	6
W-MOTO	Morphology + Topography not valid (tab.8)	Morpho	9874	Topo	C424		

6 How to interpret the incidence files' output generated by the QCS

Two types of messages are included in the output reports: W-YYYY (warning code) and E-YYYY (error code), where the code YYYY identifies the specific type of message.
This section describes how to interpret the outcomes of the QCS for some of the variables having an impact on the incidence estimation. Some examples of errors on key variables, of warnings on TNM classification variables and on multiple primary tumours are also reported.

1) Errors due to single variable format and values

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

  When the variables have values different from those allowed by the ECIS *Data Call Protocol*, the QCS returns the error E-OUTR.

PAT 000001					Tum 02				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
1	C427	9800	3	2	9/2010	2/1924	Topo	C427	E-OUTR

  In this example the QCS gives the error E-OUTR on the variable *Topo* because topography C427 does not exist in the International Classification of Diseases for Oncology, third edition[15] (ICD-O-3).

PAT 000002					Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C620	9999	9	1	10/2012	4/1935	Morpho	9999	E-OUTR
							Beh	9	E-OUTR

  In this example the QCS returns the error E-OUTR on the variables *Morpho* and *Beh* because morphology 9999 does not exist in the ICD-O-3, and behaviour 9 is not allowed according to the ECIS data call protocol.

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

PAT 000003					Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C187		3	1	8/2011	3/1945	Morpho		E-MISS

  In this example the QCS returns the error E-MISS because the variable *morphology* (impacting on incidence calculations) is missing.

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

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

PAT 000002					Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C620	9999	9	1	10/2012	4/1935	Morpho	9999	E-OUTR
							Beh	9	E-OUTR

In this example the QCS returns the error E-OUTR on the variables *Morpho* and *Beh* because morphology 9999 does not exist in the ICD-O-3, and behaviour 9 is not allowed according to the ECIS data call protocol.

- **E-AGEC**: Age is invalid or missing, and it is impossible to calculate.

PAT 000004					Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C169	8140	3	2	11/2013	99/9999	Age	999	E-AGEC
							YoB	9999	E-AGEC
							YoI	2013	E-AGEC

In this example the QCS gives the error E-AGEC because the variable *Age* (impacting on incidence calculations) is unknown (999) and cannot be calculated from DoI (2013) – DoB (9999). All the related variables *Age*, *YoB* and *YoI* are reported with the same error.

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

PAT 000005					Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C443	80984	3	1	9/2011	2/1933	Morpho	80984	E-FORM

In this example the QCS gives the error E-FORM on variable *Morpho* (80984) because morphology should have four digits instead of five, according to the ICD-O-3.

2) Errors due to inconsistency of dates.

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

PAT 000006					Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C159	8140	3	2	12/1992	8/2016	YoB	2016	E-CoDA

In this example the QCS gives the error E-CoDA on the variable *YoB* because the year of birth (2016) is set after the year of incidence (1992).

PAT 000007					Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C741	9490	3	1	2/1992	3/1992	MoB	3	E-CoDA

In this example the QCS gives the error E-CoDA because the birth (month 3) occurs after the incidence month (2) of the same year.

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

PAT 000008					Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
2	C549	8000	3	2	8/2013	4/1933	MoI	8	E-CoDV
							YoI	2013	E-CoDV
							MoF	8	E-CoDV
							YoF	2012	E-CoDV

In this example the QCS gives the error E-CoDV because the year of incidence (*YoI* 2013) occurs later than the date of last known vital status (*YoF* 2012).

PAT 000009					Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
2	C160	8000	3	1	6/2009	10/1924	MoI	6	E-CoDV
							YoI	2009	E-CoDV
							MoF	4	E-CoDV
							YoF	2009	E-CoDV

In this example the QCS gives the error E-CoDV because the date of incidence (6/2009) occurs later than the date of last known vital status (4/2009).

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

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

PAT 000010					Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
2	C569	8000	3	1	10/2013	3/1935	Sex	1	E-SETO
							Topo	C569	E-SETO

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

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

PAT 000011					Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C424	9652	3	1	12/2003	10/2003	Age	0	W-AGMT
							Morpho	9652	W-AGMT

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

PAT 000012					Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C619	8140	3	1	3/2007	5/1996	Age	10	W-AGMT
							Topo	C619	W-AGMT
							Morpho	8140	W-AGMT

In this example the QCS gives the warning W-AGMT because the topography C619 (prostate) in combination with morphology 8140/3 (adenocarcinoma, NOS) is unlikely under the age of 35 (see Table 2 of the ECIS technical report).

4) Errors and warnings due to combinations of follow-up variables and basis of diagnosis.

The following rules have been implemented to check consistency between important variables related to survival.

- **E- AUVS:** autopsy, vital status, survival time and date of incidence/date of follow-up combination not valid.

This rule checks the consistency between the variables autopsy, vital status, survival time and date of incidence/date of follow-up, according to Table 12 of the ECIS technical report.

PAT 10855 Tum 13439

BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C779	9663	3	2	1/2004	7/1929	Autopsy	1	E-AUVS
							Vit_stat	2	E-AUVS
							Surv_time	221	E-AUVS
							YoI	2004	E-AUVS
							MoI	1	E-AUVS
							YoF	2004	E-AUVS
							MoF	9	E-AUVS

In this example the QCS gives the error E-AUVS because the vital status is set to 2 (dead) but the survival time is not zero and DoI is not equal to DoF.

- **E-BDVS:** basis of diagnosis, vital status, survival time and date of incidence/date of follow-up combination not valid.

  This rule checks consistency between the variables basis of diagnosis, vital status, survival and date of incidence/date of follow-up according to Table 11 of the ECIS technical report.

PAT 3 Tum 3

BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C039	8010	3	2	2/2021	10/1954			
							BoD	0	E-BDVS
							Vit_stat	1	E-BDVS
							Surv_time	984	E-BDVS
							YoI	2021	E-BDVS
							MoI	2	E-BDVS
							YoF	2023	E-BDVS
							MoF	10	E-BDVS

In this example the QCS gives the error E-BDVS because basis of diagnosis is 0 (DCO) but vital status is 1 (alive), survival time is not 0 and date of incidence is not equal to

date of follow-up.

- **E-VSBD:** vital status, basis of diagnosis and autopsy combination not valid.

  This new rule checks consistency between the variables vital status, basis of diagnosis and autopsy, according to Table 10 of the ECIS technical report.

PAT 3 Tum 3741

BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C509	8500	3	2	12/2004	11/1948	Vit_stat	1	E-VSBD
							BoD	0	E-VSBD
							Autopsy	1	E-VSBD

In this example the QCS gives the error E-VSBD because vital status set to 1 (alive) is not coherent with basis of diagnosis 0 (DCO) and autopsy 1 (incidental finding of cancer at autopsy). Autopsy should not be 1 and Basis of Diagnosis should not be 0.

- **W-SUMU:** combination of missing/unknown survival time and date of follow-up.

  This rule checks consistency between the variables survival time and date of follow-up if survival time is missing/unknown, according to Table 13 of the JRC technical report.

PAT 3					Tum 3				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C039	8010	3	2	2/2021	10/1954			
							Surv_time	99999	W-SUMU
							YoF	9999	W-SUMU
							MoF	10	W-SUMU

In this example the QCS gives the warning W-SUMU because survival time is unknown and the year of follow-up is unknown.

- **W-SUDA:** combination of survival time, date of follow-up and date of incidence.

  This rule checks consistency between the variables survival time, date of follow-up and date of incidence, according to Table 14 of the ECIS technical report.

PAT 3					Tum 3				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
0	C039	8010	3	2	2/2021	10/1954			
							Surv_time	45	W-SUDA
							YoI	2021	W-SUDA
							MoI	6	W-SUDA
							YoF	2021	W-SUDA
							MoF	10	W-SUDA

In this example the QCS gives the warning W-SUDA because survival time is 45 days but the difference between year of follow-up and year of incidence is 4 months.

5) Errors and warnings due to combinations of other tumour variables.

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

  According to Rule F of the ICD-O-3 it is exceptionally possible to have a morphology and behaviour combination not listed in the ICD-O-3. Starting from version 2.0 the QCS reports as warnings such combinations, while previous versions reported as errors (E-MOBE). The list of accepted combinations of morphology and behaviour is reported in Table 9 of the ECIS technical report; these combinations are now allowed and do not produce W-MOBE warnings.

PAT 000013 Tum 01

BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C569	8621	3	2	4/2005	6/1982	Morpho	8621	W-MOBE
							Beh	3	W-MOBE

In this example the QCS gives error the W-MOBE because morphology 8621 (granulosa cell-theca cell tumour) with behaviour 3 (malignant tumour) is not listed in the ICD-O-3. This combination is possible, but unlikely.

PAT 000014 Tum 01

BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C421	9950	1	1	12/1989	3/1921	Morpho	9950	W-MOBE
							Beh	1	W-MOBE

In this example the QCS gives the W-MOBE warning because morphology 9950 (polycythaemia vera) has behaviour 3 (malignant tumour) in ICD-O-3. This morphology (polycythaemia vera) changed from borderline tumour (behaviour 1) in ICD-O-2[16] to malignant tumour (behaviour 3) in ICD-O-3.

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

  This rule has been updated according to Table 4 of the ECIS technical report.

PAT 311683 Tum 729324

BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
2	C710	9530	0	2	6/2013	7/1936	BoD	2	W-BDMO
							Morpho	9530	W-BDMO
							Topo	C710	W-BDMO
							Beh	0	W-BDMO

In the example above the QCS returns the warning W-BDMO because the combination of morpho/beh 9530/0 (Meningioma, NOS) with topography C71 (Brain) and basis of diagnosis 2 (clinical) is not allowed.

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

- **W-BDMS**: morphology not specific enough according to the basis of diagnosis.
  This rule has been updated according to Table 4 of the ECIS technical report.

PAT 10542 Tum 13064

BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
5	C421	9866	3	1	9/2004	3/1936	BoD	5	W-BDMS
							Morpho	9866	W-BDMS
							Beh	3	W-BDMS

In this example the QCS gives the warning W-BDMS because the combination of *morpho/beh* 9866/3 (Acute promyelocytic leukemia,) with basis of diagnosis 5 (cytology) is not allowed.

PAT 30407 Tum 33085

BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C421	9896	3	2	4/2004	9/1953	BoD	7	W-BDMS
							Morpho	9896	W-BDMS
							Beh	3	W-BDMS

In this example the QCS gives the warning W-BDMS because the combination of *morpho/beh* 9896/3 (Acute myeloid leukemia) with basis of diagnosis 7 (histology) is not allowed.

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

  This rule has been updated according to the additional checks suggested in section 3.2.5 of the ECIS technical report.

PAT 000021 Tum 01

BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C629	9061	3	1	2/2011	9/1991	Beh	3	W-BTNM
							pT	is	W-BTNM
							cT	9	W-BTNM

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

PAT 4164 Tum 4209

BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C445	8720	2	1	10/2004	12/1962	Beh	2	W-BTNM
							pT	is	W-BTNM
							cT	x	W-BTNM
							Stage	9	W-BTNM

  In the example above the QCS gives the warning W-BTNM because behavior 2 (in situ) should have stage set to 0.

- **W-TBGR (NEW)**: topography, morphology, behaviour and grade combinations are unlikely.

  This new rule follows Table 5 of the ECIS technical report, checking the grade of some tumours with behavior <3 for specific morphologies. In addition to the combinations reported in Table 5, the grade of CNS tumours with behaviours 0 and 1 is also checked according to the new ENCR Recommendations for coding tumours of the CNS (to be published).

PAT 5104 Tum 5517

BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C679	8130	2	1	2/2004	7/1953	Topo	C679	W-TBGR
							Morpho	8130	W-TBGR
							Beh	2	W-TBGR
							Grade	2	W-TBGR

  In this example the QCS gives the warning W-TBGR because topography C679 (Bladder, NOS) with behavior 2 and morphology 8130 should have grade set to 1 or 3.

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

  This rule has been updated according to Tables 6 and 7 of the ECIS technical report.

PAT	000022				Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C569	8620	3	2	5/2012	7/1954	Grade	5	W-MOGR
							Morpho	8620	W-MOGR
							Beh	3	W-MOGR

In the example above the QCS gives the warning W-MOGR because grade 5 (T-cell) is used to denote cell lineage for haematological malignancies (leukaemia and lymphoma), while morphology 8620 (granulosa cell tumour, malignant) is not an hematological malignancy. In general, the combination between grades 5-8 and morphology out of the range 9590-9993 is impossible, and produces the warning W-MOGR, except for 9801/34.

PAT	000023				Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C445	9709	3	1	11/2013	4/1935	Grade	6	W-MOGR
							Morpho	9709	W-MOGR
							Beh	3	W-MOGR

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

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

This rule has been updated according to Tables 8 of the ECIS technical report.

PAT	000024				Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C779	8070	3	1	12/2008	10/1946	Morpho	8070	W-MOTO
							Topo	C779	W-MOTO

The QCS gives the warning W-MOTO because the combination of topography C779 (Lymph node, NOS) and morphology 8070 (squamous cell carcinoma, NOS) is not allowed; this combination is probably a metastasis and the topography should be coded as C809.

PAT	000025				Tum 01				
BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C539	8120	3	2	11/2007	9/1959	Morpho	8120	W-MOTO
							Topo	C539	W-MOTO

In the example above the QCS gives the warning W-MOTO because the combination of topography C539 (cervix uteri) and morphology 8120 (transitional cell carcinoma, NOS) is very rare.

- **W-TNMM**: TNM and stage are present, but morphology is not included in the TNM.

The groups of morphology codes used in the TNM classification rules have been updated to TNM edition 6, 7 and 8, according to Table 15 of the ECIS technical report.

PAT 000026 Tum 01

BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C505	9120	3	2	11/2007	3/1971	Topo	C505	W-TNMM
							Morpho	9120	W-TNMM
							TNM_ed	6	W-TNMM
							Stage	IIB	W-TNMM
							pT	3	W-TNMM
							pN	0	W-TNMM
							pM	0	W-TNMM
							cT	9	W-TNMM
							cN	9	W-TNMM
							cM	9	W-TNMM

In the example above the QCS returns the warning W-TNMM because the tumour is a breast angiosarcoma (morphology 9120) with stage IIB. When topography is set to C50 (breast) only carcinomas should be staged.

- **W-TNMS**: TNM and stage are not consistent.

PAT 000027 Tum 01

BoD	Topo	Morpho	Beh	Sex	DoI	DoB	Var_Name	Var_Value	Error_Code
7	C502	8140	3	2	8/2013	6/1965	Topo	C502	W-TNMS
							Morpho	8140	W-TNMS
							TNM_ed	7	W-TNMS
							Stage	IIIA	W-TNMS
							pT	3	W-TNMS
							pN	1	W-TNMS
							pM	1	W-TNMS
							cT	9	W-TNMS
							cN	9	W-TNMS
							cM	9	W-TNMS
							Grade	3	W-TNMS
							Age	48	W-TNMS
							Beh	3	W-TNMS

In the example above the QCS returns the warning W-TNMS because the tumour is a breast carcinoma with *pT*=3, *pN*=1, *pM*=1 and *Stage*=IIIA. This combination is not consistent, according to TNM Ed.7 *Stage* should be IV.

6) W-MPMT and W-MPCR: Warnings for multiple primary tumours.

The MPMTs checks performed by the QCS produce two types of warnings, W-MPMT and W-MPCR, according to some special conditions for topography/morphology codes as explained at § 5.3.1 above.

If a patient has two or more multiple primary tumours that produce a warning **W-MPMT**, only one tumour should be considered for incidence analysis according to 2004 International Rules for Multiple Primary cancer.

If a patient has two or more multiple primary tumours that produce a warning **W-MPCR**, probably the multiple records refer to the same tumour, therefore these records should be reviewed by CRs.

Below some examples:

PAT	000028				Tum	01
BoD	Topo	Morpho	Beh	Sex	DoI	DoB
7	C717	8000	3	2	12/2016	12/1954

PAT	000028				Tum	02
BoD	Topo	Morpho	Beh	Sex	DoI	DoB
7	C717	9590	3	2	11/2016	12/1954

In example above, the QCS gives the warning W-MPCR because probably the two records refer to the same tumour (unspecific morphology 8000, topography code C71).

PAT	000029				Tum	01
BoD	Topo	Morpho	Beh	Sex	DoI	DoB
7	C679	8130	3	2	5/2003	1/1930

PAT	000029				Tum	02
BoD	Topo	Morpho	Beh	Sex	DoI	DoB
7	C809	8000	3	2	11/2010	1/1930

In this example, the QCS gives the warning W-MPCR because the two records could refer to the same tumour (unspecific morphology 8000, unknown primary site C809).

PAT	000030				Tum	01
BoD	Topo	Morpho	Beh	Sex	DoI	DoB
7	C501	8500	3	2	1/2001	7/1960

PAT	000030				Tum	02
BoD	Topo	Morpho	Beh	Sex	DoI	DoB
7	C508	8520	3	2	10/2002	7/1960

In example above, the QCS gives the warning W-MPMT because according to the 2004 International Rules, the two records could refer to the same tumour of different laterality with same morphology, diagnosed in paired organs (Table 15 of the ECIS Technical report). In this case only one tumour should be considered for incidence analysis (the first to be diagnosed in 2001).

```
PAT  1141980                                  Tum  1230495
Duplicated record:                            Tum  1567154
-----------------------------------------------------------
BoD  Topo  Morpho        Beh  Sex  DoI         DoB
-----------------------------------------------------------
7    C182  8140          3    1    12/2017     4/1960
-----------------------------------------------------------
-----------------------------------------------------------
PAT  1141980                                  Tum  1567154
Duplicated record:                            Tum  1230495
-----------------------------------------------------------
BoD  Topo  Morpho        Beh  Sex  DoI         DoB
-----------------------------------------------------------
7    C184  8140          3    1    12/2018     4/1960
-----------------------------------------------------------
```

In this example, the QCS gives the warning W-MPMT because according to the 2004 International Rules, the two topographies are the same (single site C18) and the two morphologies are in the same group (single entity).
In this case, only one tumour should be considered for incidence analysis (the first to be diagnosed in 2017).

Annex 1 – Major changes and improvements

Major updates and new features available from previous versions are listed below. Please refer to the change log file "change_log.txt" in the docs folder for more details.

Changes from 2.2 to 2.3.1

GUI:

- A new option "Enable checks of follow-up variables" has been added in the Settings/Options menu (currently active for JRC internal users only)

New features:

- Implemented support for running the QCS in background as a service. This allows to validate dataset with just one or few lines in real time;

Univariate/multivariate checks have been updated according to the latest version of the ECIS Technical report:

- Update of univariate checks, compliant to Table1;
- Update of consistency checks between morphology and topography (rule W-MOTO) according to the updated Table 8;
- Update of consistency checks between basis of diagnosis,morphology and behaviour by topography (rule W-BDMO) according to the updated Table 4;
- Update of consistency checks between morphology and grade for haematological malignancies, according to the updated Table 6;
- Update of configuration files used to define multiple primaries malignant tumours (MPMT) groups (rules W-MPMT and W-MPCR);
- Update of checks between morphology and grade (rule W-MOGR) for haematological and solid tumours, and update of morphology and grade checks for CNS tumours according to the ENCR recommendations;
- Update of configuration file for checks between Stage and TNM variables (rule W-TNMS) (TNM tables with alphanumeric TNM values)
- Update of checks between basis of diagnosis, morphology and behaviour by topography (rule W-BDMS) according to the updated Table 4;
- Updated logic of univariate checks regarding E-MISS, W-MISS and W-UNKN for follow-up and stage variables;
- Updated conditions for multivariate checks of follow-up variables (rules E-VSBD, E-BDVS and E-AUVS) according to tables 10, 11 and 12;
- NEW multivariate checks for follow-up varaiables (rules W-SUMU and W-SUDA) according to new Tables 13 and 14.
- Updated logic for multivariate checks of morphology/behaviour and grade for some combinations with behaviour 3 (rule W-BEGR) according to Table 7 and with behaviour 2 (W-TBGR) according to Table 5;

Previous changes from 2.1 to 2.2

GUI:

- A reminder to accept the user license has been added the first time the software is launched;

- Additional messages on the status of the validation process have been added;
- A short summary on the results of the validation process has been added at the end of each run;
- A new option "Remove/do not remove quotes from the dataset" has been added in the Settings/Options menu to help users in cleaning the dataset.

Output Reports:

- Details on the set of "Validation options" used in the last run have been added.
- A new TXT output file with a summary of the validation output (from version 2.1.1)

Code and configuration files:

- Upgrade of Log4j to version 2.17.1 to mitigate vulnerability risk;
- Better usage of the computer's memory during validation;

Univariate/multivariate checks have been updated according to the latest version of the ECIS Technical report:

- Update of univariate checks, compliant to Table1;
- Update of consistency checks between morphology and topography (rule W-MOTO) according to the updated Table 8;
- Update of consistency checks between morphology and grade of CNS tumours with Beh=3 (rule W-MOGR), according to the new ENCR Recommendations for coding tumours of the CNS (to be published);
- Update of consistency checks between morphology and grade of CNS tumours with Beh<3 (rule W-TBGR), according to the new ENCR Recommendations for coding tumours of the CNS;
- Update of consistency checks between Surgery, Basis of Diagnosis (BoD), Autopsy and the variable pT, according to the updated rules on p.33 of the ECIS Technical report;
- Update of consistency checks between age and tumour type (rule W-AGMT) according to the updated Table 2;
- Update of consistency checks between basis of diagnosis, morphology and behaviour by topography (rules W-BDMO, W-BDMS), according to Table 4 and to the 2022 ENCR Recommendations on Basis of Diagnosis;
- Update of consistency checks between morphology and grade compliant to Table 5 and to the WHO grading system, table 27 of the ICD-O-3.1 (rule **W-TBGR**) for CNS tumours;
- Update of consistency checks between morphology and grade compliant to Table 6 and 7 of the ECIS Technical report for rule W-MOGR;
- Update of consistency checks between morphology and topography codes (rule W-MOTO) according to Table 8;
- Selected combinations of morphology/behaviour not included in any version of the ICD-O-3 are now accepted, according to Table 9;
- Update of consistency checks for follow-up variables according to Tables 10, 11, 12 (rules **E-VSBD**, **E-BDVS** and **E-AUVS** replacing the rule E-VBDA in previous versions of the QCS;

- Multivariate checks for treatment variables, basis of diagnosis and TNM stage implemented according to rules reported at p.33 of the latest version of the JRC report (new rules **W-SUBD** and **W-SUpT**);
- Update of consistency checks between behaviour and TNM/stage (rule **W-BTNM** now checking also the stage, see p.33 of the latest version of the JRC report);
- Update of consistency checks between TNM/stage and basis of diagnosis (rules **W-BDpT**, **W-BDpN** and **W-BDpM**, see p.33 of the latest version of the JRC report);
- Update of groups of morphology codes used to check consistency between TNM variables and stage according to TNM editions 6,7 and 8 (rules **W-TNMM** and **W-TNMS**);
- Update of NUTS codes for the geographical variables Geo_code according to the EUROSTAT Nuts 2021 classification (https://ec.europa.eu/eurostat/web/nuts/background);
- Update of MPMT checks according to the 2004 International Rules and Chapter 4 of the latest version of the ECIS Technical report:
  - Groups of morphology codes considered as a single entity updated according to Table 15, following the adoption of the new ICD-O-3.2 codes;
  - Groups of topography codes considered as a single site updated according to Table 14;
  - More precise filtering criteria for MPMTs for records with behavior less than 3, implemented in the new configuration file located in the QCS 2.1.x *config* folder: "…\jrc-qcs-2.1.x\config\rules\encr2014\PrimaryMultipleFilter.csv";
- More precise identification and classification of MPMTs (old rule W-MPMT splitted into the new rules **W-MPMT** and **W-MPCR**);
- Update of morphology groups for checking multiple primaries according to the latest version of the ECIS Technical report;
- Update of consistency checks between TNM values and basis of diagnosis (new rule **W-BDpM**);
- Update of topography codes according to ICD-O-3 (wrong code C74.4 has been removed).

Annex 2 – Known issues and bugs, and future improvements

Issues arising from cancer registries data validation through the QCS are continuously collected, to be fixed in next versions. The following lists includes already known issues:

- The QCS could not run properly on computers with very low RAM memory (less than 4GB). If the application does not start, or if it blocks during execution, make sure that your local machine supports at least 4GB of memory and that your internal policies allow to use all the available memory (e.g. you use a corporate PC);
- To avoid memory issues, the QCS should be run clicking on a BAT file containing a "2GB" label in the file's name. This solution could be blocked by strict system policies, such as those applied by universities, corporate environments, governments etc. If you experience memory issues and can't run the BAT file with the "2GB" label, please ask support to your local system administrator and ask to be granted the necessary permissions for running the "2GB" file;
- If the TXT files located in the Output directory cannot be opened with your text editor, please set Notepad++ as the default application to open them;
- The MPMT warning W-MPCR could be wrongly assigned to some rare combinations of solid or haematological MPMTs: in case, please notify;
- If the QCS is run while the PDF output report is still opened by the user, the PDF will **not** be updated. This happens on purpose, as a "security block" that avoids to delete the results of the previous run while the user is still using the software;
- If the input dataset has a critically wrong header (the first line), with errors or missing variables, the QCS might not correctly produce output reports;
- The QCS still checks for the previous version's warning W-SEMO, which is not relevant anymore for the ECIS data call protocol. Should your dataset trigger any W-SEMO message, please ignore related warnings.

Annex 3 – List of error and warning codes

This annex reports the list of error and warning codes included in the output files "*QCS- Incidence-Output.pdf*" and "*QCS-Incidence-Output.txt*". The table or page numbers referenced in the list are from the latest version of the ECIS Technical report.

Error codes

E-AGEC: Age is invalid or missing and cannot be calculated age by subtracting date of incidence from date of birth, since one or both dates are invalid or missing.

E-AGED: The age calculated as difference in years from (Date of incidence – Date of birth) differs from the value of variable *Age* by more than one year.

E-AUVS (NEW): the combination of Autopsy 1 with vital status and/or survival (Date of incidence – Date of follow-up in days) is not valid. Vital status should be 2 and Survival should be 0 (Table 12).

E-BDVS: the combination of Basis of diagnosis 0 (DCO) with vital status and/or survival is not valid. Vital status should be 2 and survival should be 0 (Table 11).

E-CoDA: The date of birth is not consistent with the date of incidence, i.e. the latter occurs before the date of birth.

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

E-DUPL: The same patient ID/tumour ID combination is repeated in two or more records.

E-ECOD: ICD[17] edition and cause of death combination are not valid, e.g. cause of death 157 (pancreatic cancer) with 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-HEAD: Errors in the file header, e.g invalid number/order of columns, invalid header's separator etc.

E-MISS: Value missing, e.g. when variable *morphology* is unknown. Thiserrors applies to variables whose invalid/missing/unknown values have an impact on incidence statistics.

E-OUTR: Value out of range; value is not in agreement with the ones allowed by the ECIS data call protocol (for instance, behavior equal to 6 or sex set to 4).

E-RECO: The record has the wrong number of fields.

E-SETO: Sex and topography combinations are not valid (Table 3).

E-VSBD: the combination of vital status 1 with autopsy and/or basis of diagnosis is not valid. Autopsy should be not 1 and Basis of diagnosis should be not 0 (Table 10).

[17] International Classification of Diseases (https://www.who.int/standards/classifications/other-classifications/international-classification-of-diseases-for-oncology)

Warning codes

W-AGMT: Unlikely age and morphology/topography combinations (Table 2).

W-BDMO: Morphology too specific according to the basis of diagnosis. See pages 17-18 for valid combinations of basis of diagnosis and morphology.

W-BDMS: Morphology not specific enough according to the basis of diagnosis. See page 19 for valid combinations of basis of diagnosis and morphology.

W-BDMU: Basis of diagnosis and morphology/behavior 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 1 then basis of diagnosis should be 7 or 6 (see page 33).

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 or 7 (see page 33).

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 33).

W-BEGR: Behavior and grade combination is not valid (Table 7).

W-BTNM: Invalid behavior and TNM combination, e.g. *Behavior*=3 and *pT*=*Tis* (see page 33).

W-EDIM: TNM edition and *pM* are not consistent. The warning is returned when TNM edition is 7 or 8, and *pM* or *cM* are "*X*", since this value should be "0".

W-MISS: Value missing, e.g. when variable *Autopsy* is blank. This check applies to variables whose invalid/missing/unknown values is not impacting on incidence statistics.

W-MOBE: Morphology and behavior combinations not included in the ICD-O-3 (some combinations are allowed, according to Table 9).

W-MOGR: Morphology and grade combination is unlikely (Tables 6 and 7).

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

W-MPCR: Multiple primary tumour (see § 5.3 of this document)

W-MPMT: Multiple primary tumour (see § 5.3 of this document)

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

W-SUDA: combination of survival time (in days) not consistent with date of incidence and date of follow-up (Tab. 14).

W-SUMU: combination of survival time missing/unknown and date of follow-up missing/unknown is not valid (Tab. 13).

W-TBGR (**NEW**): combination of morphology/behavior and grade not valid for tumours with behavior <3 (Table 5).

W-TNME: TNM and stage are present, but TNM edition is not valid or missing. The warning is returned since it is not possible to make a consistency check between TNM and stage.

W-TNMM: TNM and stage are present, but the morphology is not included in the TNM, e.g. when only carcinomas can be staged in a given topography, but stage is filled in for sarcomas.

W-TNMS: TNM and stage are not consistent, e.g. *pT* is 1, *pN* is 0, *pM* is 0 and stage is *IV*. In case both pathological (*pT*, *pN* and *pM*) and clinical (*cT*, *cN* and *cM*) TNM are provided for a tumour, the

QCS will check the consistency between the pathological TNM and stage.

W-UNKN: missing/unknown value for variables with no impact on incidence calculations, which however could be important for quality evaluations or survival analysis (e.g. basis of diagnosis=9, or year of follow up=9999).

Annex 4 – Running the JRC-ENCR QCS in background

Overview

The QCS can be run in different modes as listed below:

- **GUI** (standard execution): opens the main window and wait for user's actions;
- **Silent** (background process): run in background and validate the file passed as argument;
- **Introspection** (consolle): prints information about protocols and rules.

To view all supported arguments, please run the application with the **-h** option. The actual syntax may depend by your Java installation, and should be something like:

java -jar jrc-qcs-2.1.jar -h

The main supported runtime arguments are the following (see *Command line examples* below):

- **-gui**: opens the application in GUI mode (standard execution);
- **-v**: validates target dataset using the target validation protocol;
- **-p**: prints the list of rules defining target protocol;
- -**s**: prints the list of variables addressed by target protocol;
- -**fi**: prints all the information about the target variable (aka *field*);
- **-rf**: prints the list of rules addressing the target variable (aka *field*).

Command line examples

Validation of a dataset using the ECIS data call protocol (i.e. protocol with ID = 11):

*java -jar jrc-qcs-2.1.jar **-v** 11 C:/Home/incidence_1.csv*

List of rules for the ECIS data call protocol (i.e. protocol with ID = 11):

*java -jar jrc-qcs-2.1.jar **-p** 11*

Description of the topography field in the ECIS data call protocol (field with ID = 80):

*java -jar jrc-qcs-2.1.jar **-fi** 80*

List of rules addressing the topography field in the ECIS data call protocol (field ID = 80):

*java -jar jrc-qcs-2.1.jar **-rf** 80*

Sample scripts

The *Examples* 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 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 they are in a easy readable format (PDF or TXT). If the client application needs to read, parse, analyse or process the results of the validation process, usage of the following report is recommended:

- **QCS-Incidence-Output.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.