European
Commission

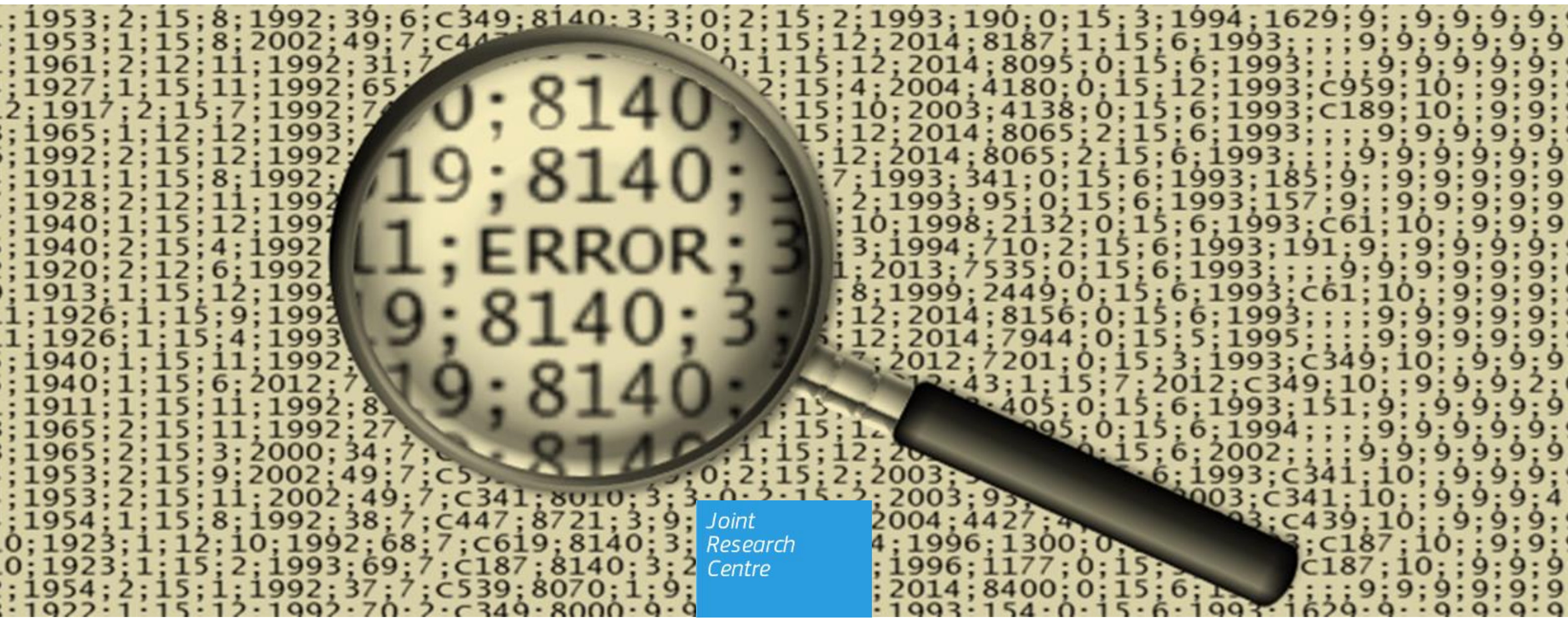
JRC SCIENTIFIC INFORMATION SYSTEMS
AND DATABASES REPORT

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

JRC-ENCR QCS 2.0
JRC CSV Data layout converter

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

2022



Joint
Research
Centre

This publication is a Scientific Information Systems and Databases 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 policymaking 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 that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information
E-mail: JRC-ENCR@ec.europa.eu

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

JRC127031

Ispra: European Commission, 2022

How to cite: Francesco Giusti, Carmen Martos, Stefano Adriani, Manuela Flego, Antonino Brunetto, Tadeusz Dyba, Lena Voith von Voithenberg, Luciana Neamtiu, Raquel N. Carvalho, Giorgia Randi, Nadya Dimitrova, Nicholas Nicholson, Revveka Trigka, Emanuele Crocetti, Manola Bettio, Enrico Ben, *The JRC-ENCR Quality Check Software (QCS) for the validation of cancer registry data: user compendium – version 2.0*, European Commission, Ispra 2022, JRC127031

Contents

Acknowledgments

The authors acknowledge the following institutions and entities for the invaluable feedback provided running the present and 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

After the 2015 **call for data** the European Network of Cancer Registries (ENCR) and the European Commission Joint Research Centre (JRC) have launched a new data call for updating cancer indicators available in the European Cancer Information System (ECIS) web application (https://ecis.jrc.ec.europa.eu/).

Unlike the previous one, the 2022 call is a rolling process; cancer registries will have the possibility to update their data once per year. The new data call protocol is available in the ENCR website here.

In order to enable cancer registries to perform data quality checks and to test the adherence of their data to the required format of the ENCR-JRC data calls, starting from 2015 the JRC has been developing the JRC–ENCR Cancer Registries Data Quality Check Software (QCS) (https://encr.eu/tools-for-registries).

The present version of the JRC-ENCR QCS is based on the 2022 data call protocol, and the experience gathered after validating over 30 million cases from around 150 population-based cancer registries in 35 European Countries with previous JRC-ENCR QCS versions. In addition, feedback from European cancer registries and institutions was taken into account for the improvement of the JRC-ENCR QCS.

Version 2.0 of the JRC-ENCR QCS, which replaces version 1.8.1 distributed in 2019, 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 (see section 3.1.1 below)
- verification of variables' formats and values
- cross checks among variables (internal consistency)
- check of multiple primary 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 2.0[1] is a stand-alone tool created for validating cancer-registries' data against the requirements of the latest ENCR-JRC call for data protocol for European population-based cancer registries. The majority of the checks are based on version 1.1 (2018 update) of the ENCR-JRC report "A proposal on cancer data quality checks: one common procedure for European cancer registries" (https://encr.eu/sites/default/files/inline-files/Cancer_Data_Quality_Checks_Procedure_Report_online_0.pdf). This report will be updated by the end of 2022.

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 1.8.1, the 2021 version 2.0 release of the software includes the following changes and enhancements:

- ability to be run, in addition to the 2022 ENCR-JRC call for data protocol, also on the previous (2015) call for data protocol;
- creation of a separate software, the *JRC CSV Data layout converter* (QCS Buddy) that will assist users in the preparation of the file to be run by the QCS;
- new consistency check between topography, TNM Edition, TNM, and stage introduced;
- new consistency check between topography, TNM Edition, TNM, stage and morphology introduced;
- new consistency check between TNM edition and pM introduced;
- new check on TNM edition value introduced;
- All TNM Checks: update to 8th edition (6th and 7th editions were already included in QCS version 1.8.1);
- updated morphology families used by checks involving TNM and morphology according to the ICD-O-3.2 update;
- updated morphology families used by the multiple primary tumour checks according to the ICD-O-3.2 update;
- inclusion of behaviour 2 (in situ) and behaviour 1 (uncertain and unknown behaviour) urological tumours (C65-C68 ICD-O-3 codes) as well as behaviour 1 and behaviour 0 (benign tumours) central nervous systems tumours (C70-C72 and C751-C753) in the multiple primary tumour checks.

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 future improvements***.

[1] Information on the QCS updates will be published on the following webpage: 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).

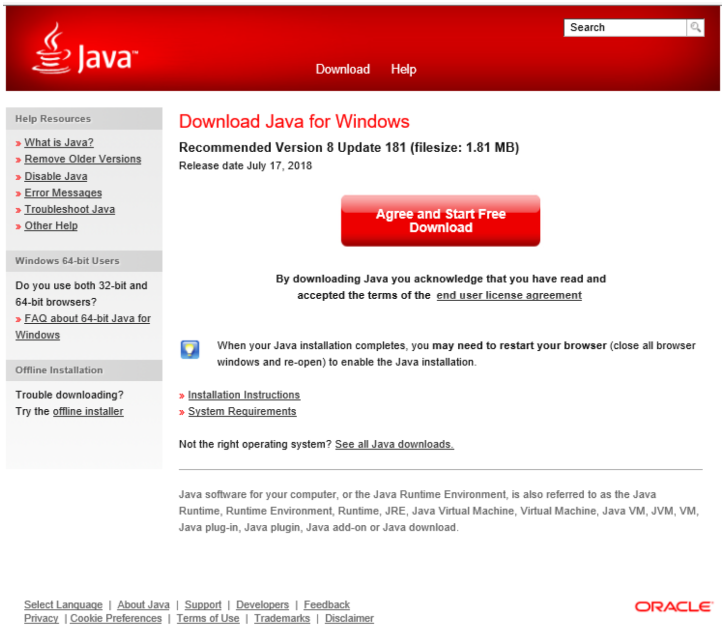
Version 2.0 of the QCS can also run on macOS and Linux operating systems (see sections 2.5 and 2.6) below. Sections 2.1-2.4 refer to Windows operating systems.

2.1 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:



Do you use both 32-bit and 64-bit browsers?
» FAQ about 64-bit Java for Windows

Trouble downloading?
Try the offline installer

Download Java for Windows

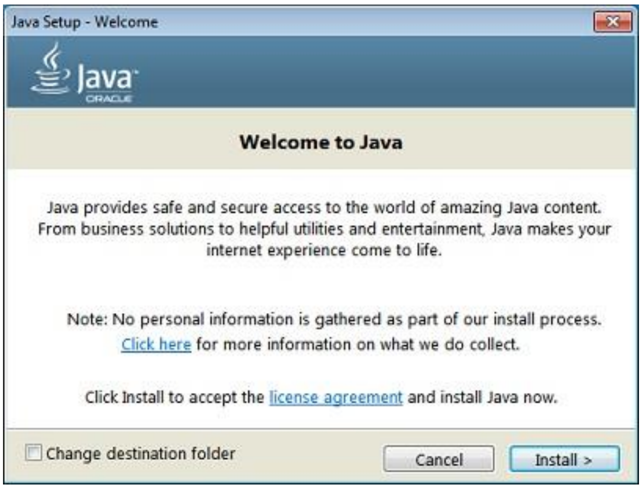
Recommended Version 8 Update 181 (filesize: 1.81 MB)
Release date July 17, 2018

By downloading Java you acknowledge that you have read and accepted the terms of the end user license agreement

When your Java installation completes, you may need to restart your browser (close all browser windows and re-open) to enable the Java installation.

Not the right operating system? See all Java downloads.

Java software for your computer, or the Java Runtime Environment, is also referred to as the Java Runtime, Runtime Environment, Runtime, JRE, Java Virtual Machine, Virtual Machine, Java VM, JVM, VM, Java plug-in, Java plugin, Java add-on or Java download.


Java Setup - Welcome

Welcome to Java

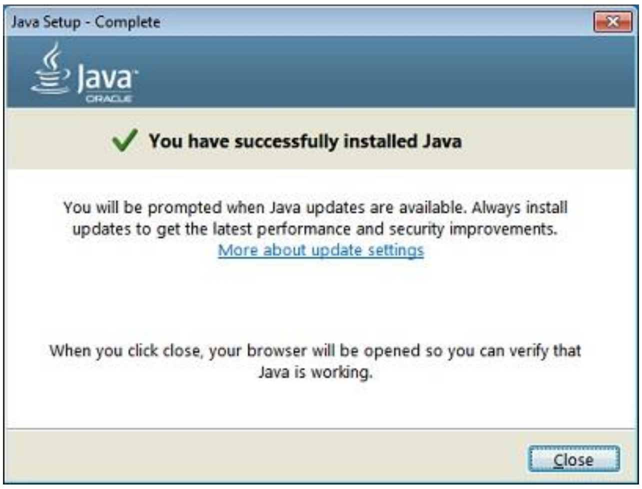
Java provides safe and secure access to the world of amazing Java content. From business solutions to helpful utilities and entertainment, Java makes your internet experience come to life.

Note: No personal information is gathered as part of our install process. Click here for more information on what we do collect.

Click Install to accept the license agreement and install Java now.

□ Change destination folder

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.


Java Setup - Complete

✓ You have successfully installed Java

You will be prompted when Java updates are available. Always install updates to get the latest performance and security improvements. More about update settings

When you click close, your browser will be opened so you can verify that Java is working.

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

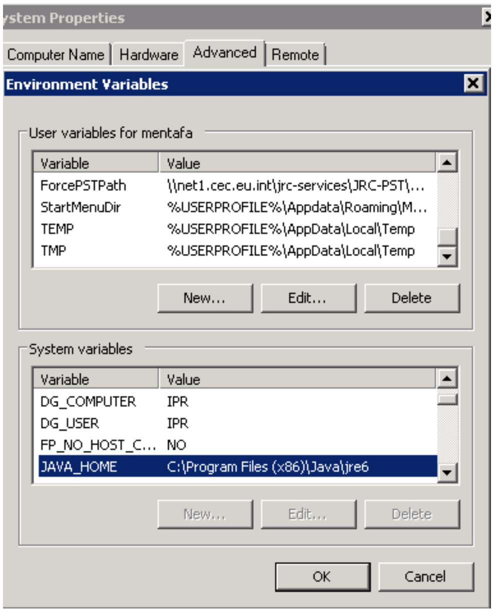
2.2 Further information and troubleshooting related to Java

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:



User variables for mentafa

Variable	Value
ForcePSTPath	\\net1.cec.eu.int\jrc-services\JRC-PST\...
StartMenuDir	%USERPROFILE%\Appdata\Roaming\M...
TEMP	%USERPROFILE%\AppData\Local\Temp
TMP	%USERPROFILE%\AppData\Local\Temp

System variables

Variable	Value
DG_COMPUTER	IPR
DG_USER	IPR
FP_NO_HOST_C...	NO
JAVA_HOME	C:\Program Files (x86)\Java\jre6

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)

2.3 How to install the QCS

Once you download the latest version of the software please extract file ***JRC-ENCR-QCS-V2.0.zip*** on your computer.
You will be able to access folder "*JRC-ENCR-QCS-V2.0*" with all the related subfolders.

2.4 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-V2.0.zip", then you should execute the command:

   ```
   cd Desktop/JRC-ENCR-QCS-V2.0
   ```

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
   ```

2.5 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-V2.0.zip" you should execute the command

   ```
   unzip JRC-ENCR-QCS-V2.0.zip
   ```

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

   ```
   cd JRC-ENCR-QCS-V2.0/
   ```

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
   ```

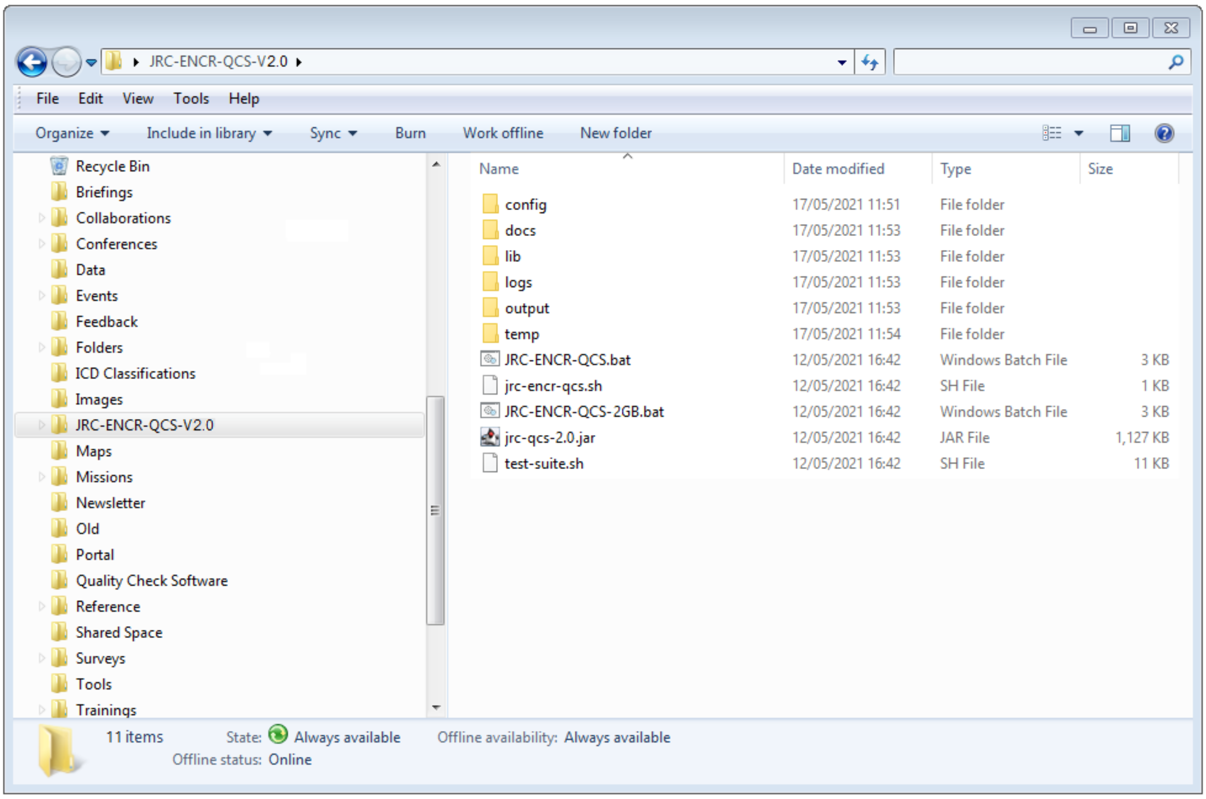
4. Execute the QCS by running the ".sh" file:

   ```
   ./start-jrc-encr-qcs.sh
   ```

2.6 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:



Name	Date modified	Type	Size
config	17/05/2021 11:51	File folder	
docs	17/05/2021 11:53	File folder	
lib	17/05/2021 11:53	File folder	
logs	17/05/2021 11:53	File folder	
output	17/05/2021 11:53	File folder	
temp	17/05/2021 11:54	File folder	
JRC-ENCR-QCS.bat	12/05/2021 16:42	Windows Batch File	3 KB
jrc-encr-qcs.sh	12/05/2021 16:42	SH File	1 KB
JRC-ENCR-QCS-2GB.bat	12/05/2021 16:42	Windows Batch File	3 KB
jrc-qcs-2.0.jar	12/05/2021 16:42	JAR File	1,127 KB
test-suite.sh	12/05/2021 16:42	SH File	11 KB

The folder *JRC-ENCR-QCS-V2.0* includes the following:

- The executable files *JRC-ENCR-QCS.bat* and *JRC-ENCR-QCS-2GB.bat;*
- Files *jrc-encr-qcs.sh* and *test-suite.sh*;
- The library *qcs-library-2.0.jar* file;
- Folders config, *docs, lib, logs, output, temp.*

File *JRC-ENCR-QCS.bat* will run the QCS with 1GB of RAM memory, whereas *JRC-ENCR-QCS-2GB.bat* will use 2GB of RAM memory.

File *jrc-encr-qcs.sh* is the standard SH script for running the application on the Linux system (see section 2.5 above)

config: this folder contains configurations files, such as those for values ranges, general application settings (with the possibility to disable some functionalities) and for the log file.

docs: this folder contains relevant documentation files of the software, i.e. the present report 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. Subfolder *samples* includes some sample scripts for advanced users (see *Annex 3 – Running the JRC-ENCR QCS in background*).

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

logs: this folder stores the logs of all the QCS activity.

output: it is created after the QCS is run for the first time. It 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.*

temp: this folder contains all the raw files (working file), which form the basis of the reports.

A separate folder for the JRC CSV Data layout converter is also created.

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

The file must follow the format of the call for data protocol (section 3.1.1). 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 new call for data protocol

The following are the variables of the new ENCR-JRC call for data protocol for European population-based cancer registries, with the required format.

Patient variables					
Variable name	**Variable description**	**Format**	**Maximum length**	**Missing/ unknown**	**Coding**
PAT[2]	Patient identification code	A	50	Not allowed	According to registry coding
MoB	Month of birth	F	2	99	Range of allowed values: 1 - 12
YoB	Year of birth	F	4	9999	Range of allowed values: > 1842 and ≤ the current year
Age	Age at diagnosis (incidence date) in years	F	3	999	Range of allowed values: ≥ 0 and < 121
Sex	Sex at birth	F	1	9	1 →Male 2 →Female 3 →Other

[2] *PAT* should be a code assigned by the registry that is not to be used elsewhere (e.g. in a hospital). So, it cannot be an official personal number. It may be an encrypted personal number as long as this specific encryption is not used by any other organisation. The JRC will provide the tool to the CRs to do it.

Geo_code	Code for the geographical area of residence at Diagnosis	A	10	XX99	NUTS 2 when available or the highest level of administrative sub-divison that can be provided[3]. Blank → not applicable
Geo_label	Name of the geographical area of residence at Diagnosis	A	50	9	Blank → not applicable
Tumour variables					
TUM	Tumour identification	A	50	Not allowed	According to registry coding
MoI	Month of incidence	F	2	99	Range of allowed values: 1 - 12
YoI	Year of incidence	F	4	Not allowed	Range of allowed values: From 1941 to present
BoD	Basis of diagnosis	F	1	9	0→Death certificate only 1→Clinical 2→Clinical investigation 4→Specific tumour markers 5→Cytology 6→Histology of a metastasis 7→Histology of a primary tumour
Topo	ICD-O-3 topography code	A	4	Not allowed	Valid code in ICD-O-3
Morpho	ICD-O-3 morphology code	F	4	Not allowed	Valid code in any ICD-O-3 version
Beh	ICD-O-3 behaviour	F	1	Not allowed	0→ Benign neoplasm 1→ Neoplasm of uncertain and unknown behaviour 2→ In situ neoplasm 3→ Malignant neoplasm
Grade[4]	ICD-O-3 grade of differentiation / immunophenotype	F	1	9	1→Grade I, Well differentiated 2→ Grade II, Moderately differentiated 3→ Grade III, Poorly differentiated 4→Grade IV, Undifferentiated, anaplastic 5→ T-cell; T-precursor 6→ B-Cell; Pre-B; B-precursor 7→ Null cell; Non T-non B 8→ NK cell (natural killer cell) 9→ Not applicable
Variables related to follow-up					

Variable name	Variable description	Format	Maximum length	Missing/ unknown	Coding
Autopsy[5]	Incidental finding of cancer at autopsy	F	1	9	0→No 1→Yes
Vit_stat	The last known vital status	F	1	9	1→ Alive 2→ Dead
MoF	Month of last known vital status	F	2	99	Range of allowed values: From 1 to 12
YoF	Year of last known vital status	F	4	9999	Range of allowed values: > 1941 and ≤ the current year
Surv_time	Duration of survival in days	F	5	99999	≥ 0

[3] NUTS 3 codes should be provided for regional registries covering NUTS 3 areas such as French *Départements*, Italian *Province* and Spanish *Provincias*.

[4] The *grade* of tumours of the central nervous system should be coded according to table 27 of ICD-O-3.

[5] In autopsy cases, incidentally found at autopsy, the *vital status* is always 2 (dead) and the *survival* time is 0 days.

ICD[6,7]	ICD edition for coding cause of death	F	2	99	Range of allowed values: <12 Blank → Not applicable
CoD[6,7]	Official underlying cause of death	A	4	R99 (ICD-10) 7999 (ICD-9)	According to ICD Blank → Not applicable
Stage variables					
TNM_ed	TNM edition	F	2	99	Allowed values: ≤ 8
cT[8]	Clinical T-category	A	12	9	According to the TNM Classification of Malignant Tumours Blank → not applicable
cN[8]	Clinical N-category	A	12	9	
cM[8]	Clinical M-category	A	12	9	
pT[8,9]	Pathological T-category	A	12	9	
pN[8,9]	Pathological N-category	A	12	9	
pM[8,9]	Pathological M-category	A	12	9	
ToS	Staging system	A	3	9	A → Ann Arbor/ Lugano stage D → Dukes' stage E → Extent of disease F → FIGO stage S → TNM stage, unknown whether clinical or pathological clS → clinical TNM stage paS → pathological TNM stage cpS → combination of clinical & pathological TNM stage coS → condensed TNM stage esS → essential TNM stage T1 → Tier 1 stage for paediatric tumours T2 → Tier 2 stage for paediatric tumours 8 → Other staging system
Stage variables					
Variable name	**Variable description**	**Format**	**Maximum length**	**Missing/ unknown**	**Coding**
Stage	Stage	F	1	9	0 → Stage 0, stage 0a, stage 0is, carcinoma in situ, non-invasive 1 → Stage I, FIGO I, localized, localized limited (L), limited, Dukes A 2 → Stage II, FIGO II, localized advanced (A), locally advanced, advanced, direct extension, Dukes B 3 → Stage III, FIGO III, regional (with or without direct extension), R+, N+, Dukes C 4 → Stage IV, FIGO IV, metastatic, distant, M+, Dukes D
Treatment variables					

[6] If the vital status is 1 (alive) the *CoD* and *ICD* should be left blank.

[7] if the vital status is 2 (dead) and the cause of death is unknown, CoD should be coded as R99 (ICD-10)/7999 (ICD-9) or 9999 and ICD should be coded as 99.

[8] If TNM is not available or not applicable, cTNM (*cT, cN, cM*) and pTNM (*cT, cN, cM*) should be coded as 9 and be left blank respectively and (if possible) *Staging system (ToS)* and *stage* should be coded.

[9] If cTNM is available and the primary tumour was not resected the pTNM (*pT, pN, pM*) should be left blank.

Surgery[10],[11]	Resection of the primary tumour	F	1	9	0 → No 1 → Yes, without specification 2 → Yes, local surgery only[12] 3 → Yes, ‘operative’ surgery[13]
Rt	Radiotherapy	F	1	9	0 → No 1 → Yes, without specification 2 → Yes, neoadjuvant (pre-operative) radiotherapy 3 → Yes, adjuvant (post-operative) radiotherapy
Cht	Chemotherapy	F	1	9	0 → No 1 → Yes, without other specification 2 → Yes, neoadjuvant (pre-operative) 3 → Yes, adjuvant (post-operative) 4 → Yes, both neoadjuvant and adjuvant
Tt[14]	Targeted therapy (including monoclonal antibodies)	F	1	9	0 → No 1 → Yes
It	Immunotherapy (excl. monoclonal antibodies)	F	1	9	0 → No 1 → Yes
Ht	Hormone therapy	F	1	9	0 → No 1 → Yes
Ot	Other or unspecified systemic therapy	F	1	9	0 → No 1 → Yes, without other specification 2 → Yes, neoadjuvant (pre-operative) 3 → Yes, adjuvant (post-operative)
SCT	Stem cell transplantation	F	1	9	0 → No 1 → Yes

[10] If available, type of surgery (*local surgery* [12] versus *operative surgery* [13]) should be coded for solid cancers of the following cancer sites: C01-C06, C16-C20, C30-C34, C53-C55, C61 and C65-C68. For other cancers, code 1 (surgery without specification) suffices.

[11] If both *local surgery* and *operative surgery* were performed for the same tumour, *operative surgery* should be coded.

[12] The following procedures should be coded as local surgery: polypectomy (mainly gastro-intestinal tract), transurethral resection (TUR; bladder & other urinary tract), cone biopsy/loop excision (cervix), as well as all other procedures which leave the organ in situ, such as cryosurgery, laser coagulation, thermoablation, radiofrequency ablation (RFA), etc.

[13] This includes all resections of the tumor which require the removal of an organ or a major part of that organ, such as a lobectomy, hemicolectomy, hysterectomy, cystectomy, prostatectomy, etc.

[14] Targeted therapy comprises all drugs that block the growth of cancer cells by inhibition of certain pathways in the cancer cell. Traditional chemotherapy also affects other cells in the body that divide quickly. The main categories of targeted therapy are small molecules (mostly tyrosine kinase inhibitors such as imatinib and many other -nibs) and monoclonal antibodies (such as rituximab and many other -mabs). Monoclonal antibodies are considered a form of immunotherapy but should be coded as targeted therapy.

3.1.2 Incidence file creation

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 39 by default.

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).

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**;**"

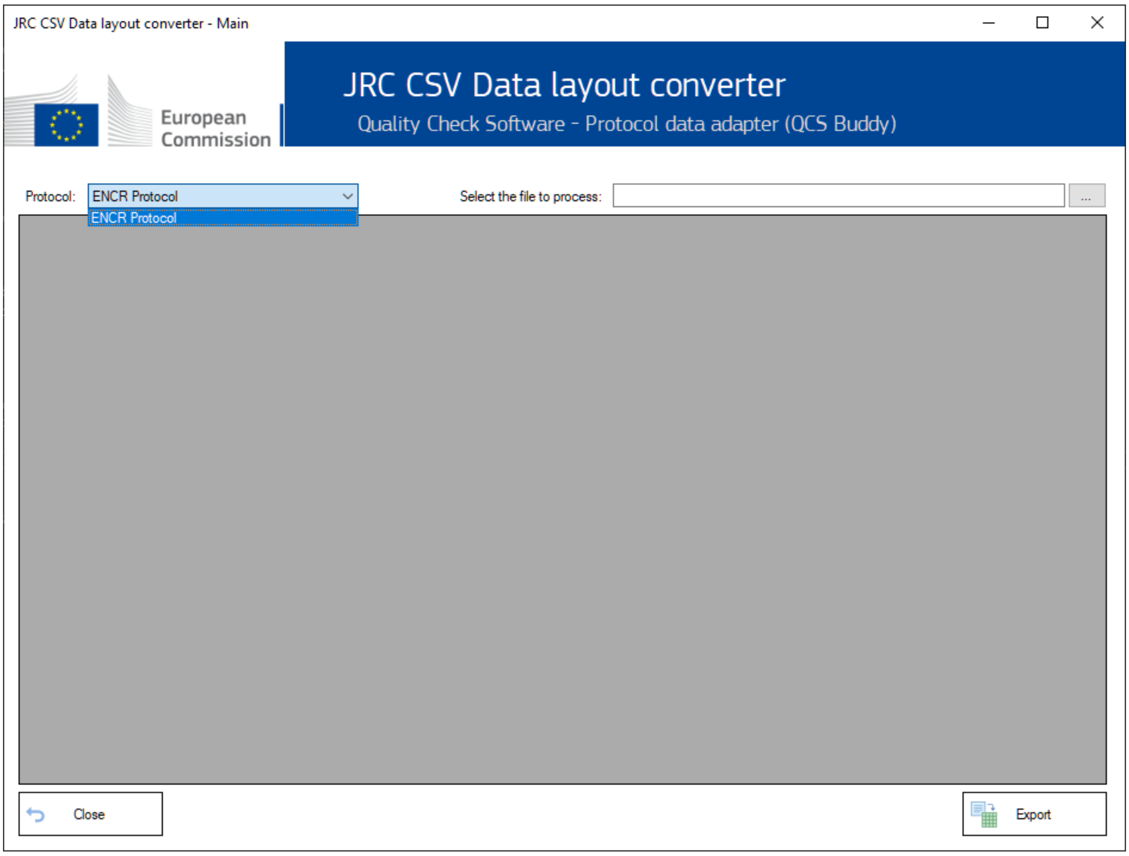
After the creation of the header, please proceed by creating the lines/records with the values of those variables.

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.

3.1.3 JRC CSV Data layout converter (QCS Buddy)

The JRC CSV Data layout converter (QCS Buddy) was created in order to assist users (with Windows operating systems) in the creation of incidence files to be checked with the QCS.

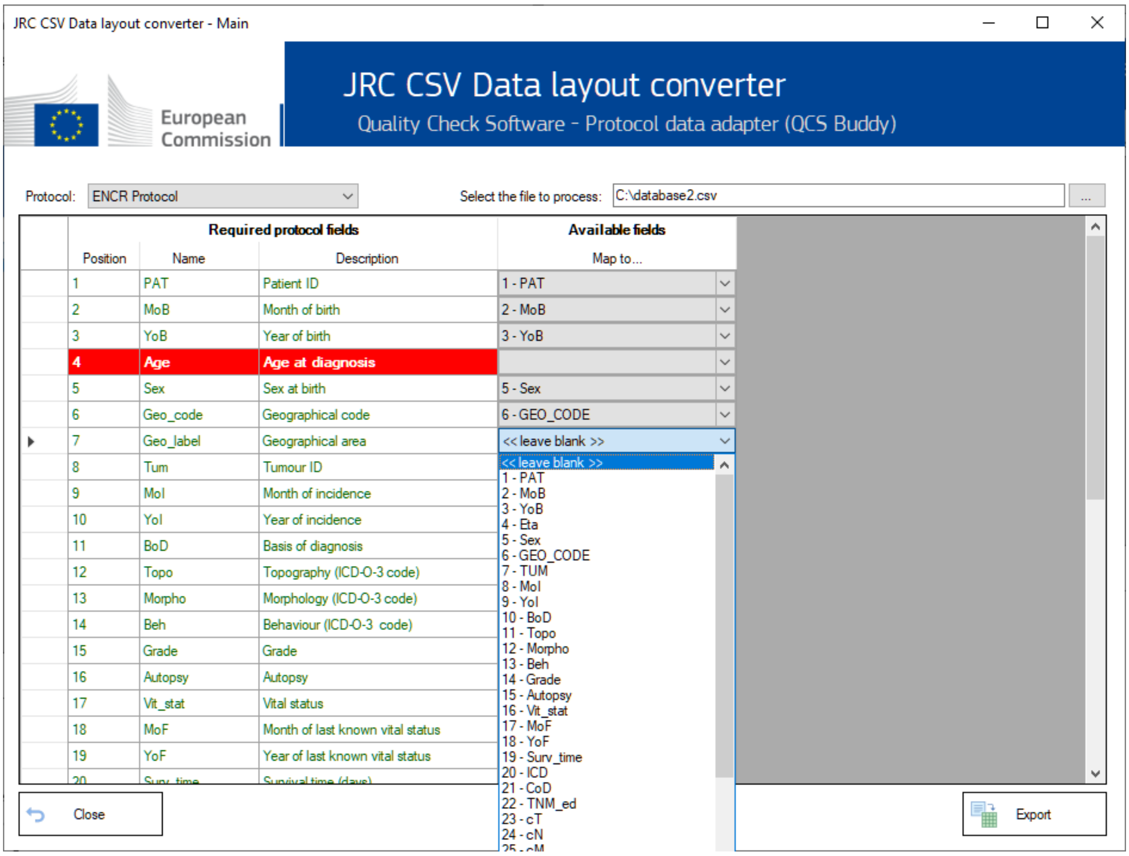

JRC CSV Data layout converter - Main
European Commission
JRC CSV Data layout converter
Quality Check Software - Protocol data adapter (QCS Buddy)
Protocol:
ENCR Protocol
ENCR Protocol
Select the file to process:
...
Close
Export

Option “ENCR Protocol” is the default one for the tool.

Select a data file to import and convert. The file can be in any text format (CSV) with columns separator. The following column separators are supported:

- **TAB** (tabulation)
- **|** (pipe)
- **,** (comma)
- **;** (semicolon)

If there are no errors, program shows the list of the fields defined in the protocol and the corresponding fields found in the data file.


JRC CSV Data layout converter - Main
European Commission
JRC CSV Data layout converter
Quality Check Software - Protocol data adapter (QCS Buddy)
Close
Export

Protocol: ENCR Protocol

Select the file to process: C:\database2.csv

Required protocol fields			Available fields
Position	Name	Description	Map to...
1	PAT	Patient ID	1 - PAT
2	MoB	Month of birth	2 - MoB
3	YoB	Year of birth	3 - YoB
4	Age	Age at diagnosis	
5	Sex	Sex at birth	5 - Sex
6	Geo_code	Geographical code	6 - GEO_CODE
7	Geo_label	Geographical area	<< leave blank >>
8	Tum	Tumour ID	
9	MoI	Month of incidence	
10	YoI	Year of incidence	
11	BoD	Basis of diagnosis	
12	Topo	Topography (ICD-O-3 code)	
13	Morpho	Morphology (ICD-O-3 code)	
14	Beh	Behaviour (ICD-O-3 code)	
15	Grade	Grade	
16	Autopsy	Autopsy	
17	Vit_stat	Vital status	
18	MoF	Month of last known vital status	
19	YoF	Year of last known vital status	
20	Surv_time	Survival time (days)	

<< leave blank >>
1 - PAT
2 - MoB
3 - YoB
4 - Eta
5 - Sex
6 - GEO_CODE
7 - TUM
8 - MoI
9 - YoI
10 - BoD
11 - Topo
12 - Morpho
13 - Beh
14 - Grade
15 - Autopsy
16 - Vit_stat
17 - MoF
18 - YoF
19 - Surv_time
20 - ICD
21 - CoD
22 - TNM_ed
23 - cT
24 - cN
25 - cM

To facilitate the data import process, the QCS Buddy tries to automatically map fields with the same name.

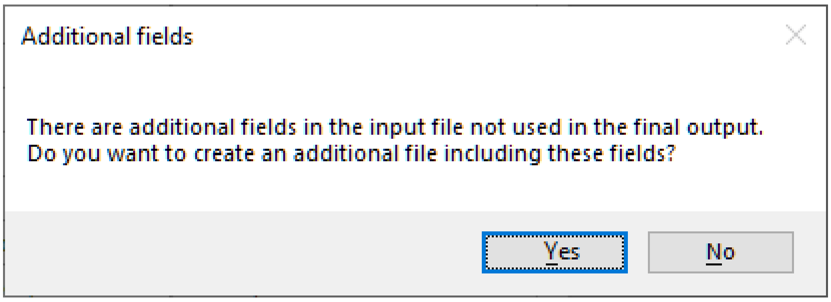
Mapped fields are displayed in GREEN, unmapped fields are displayed in RED.

For those fields where an automatic mapping was not possible (but in general for all fields) the user can:

1) Map the protocol field with one of the fields found in the data file
2) Leave the field blank (for example, if no mapping is possible, data are not available, etc..)

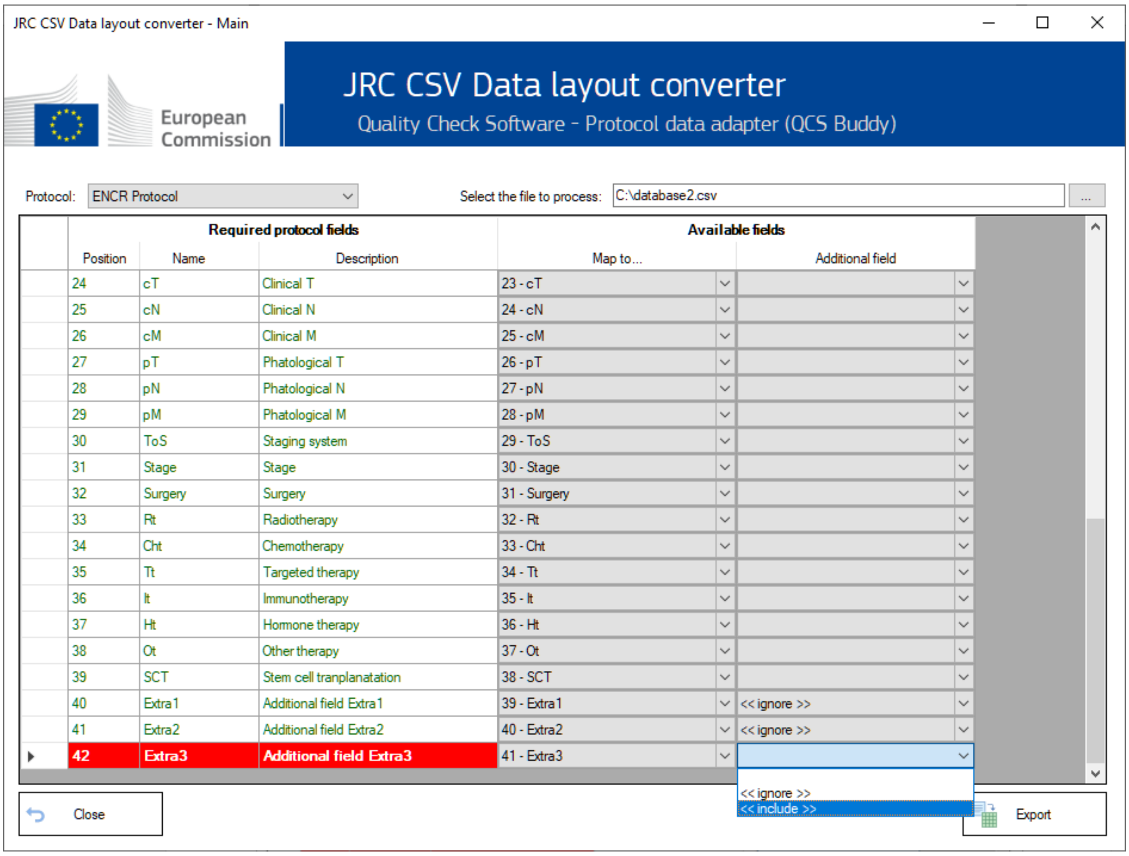
Only when all the fields defined in the protocol are mapped (or blank), it is possible to Export the content of the original file and “convert” it in the format defined in the ENCR-JRC protocol.

Additionally, if there are fields in the data file that are not used in the protocol, the tool asks if the user wants to export an additional file including one or more of these fields in addition to those expected in the ENCR-JRC protocol.



Additional fields

There are additional fields in the input file not used in the final output.
Do you want to create an additional file including these fields?



JRC CSV Data layout converter - Main

JRC CSV Data layout converter

Quality Check Software - Protocol data adapter (QCS Buddy)

Protocol: ENCR Protocol

Select the file to process: C:\database2.csv

Required protocol fields			Available fields	
Position	Name	Description	Map to...	Additional field
24	cT	Clinical T	23 - cT	
25	cN	Clinical N	24 - cN	
26	cM	Clinical M	25 - cM	
27	pT	Phatological T	26 - pT	
28	pN	Phatological N	27 - pN	
29	pM	Phatological M	28 - pM	
30	ToS	Staging system	29 - ToS	
31	Stage	Stage	30 - Stage	
32	Surgery	Surgery	31 - Surgery	
33	Rt	Radiotherapy	32 - Rt	
34	Cht	Chemotherapy	33 - Cht	
35	Tt	Targeted therapy	34 - Tt	
36	It	Immunotherapy	35 - It	
37	Ht	Hormone therapy	36 - Ht	
38	Ot	Other therapy	37 - Ot	
39	SCT	Stem cell tranplanatation	38 - SCT	
40	Extra1	Additional field Extra1	39 - Extra1	<< ignore >>
41	Extra2	Additional field Extra2	40 - Extra2	<< ignore >>
42	Extra3	Additional field Extra3	41 - Extra3	<< ignore >> << include >>

In this example there are 3 fields in the data file (Extra1, Extra2 and Extra3), and the user has chosen to export a file including only the Extra3 field. A new file will be exported, like the previous one but including the new Extra3 field (new fields are appended at the end of the record).

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.

Please use the following lines as header, copy/pasting the relevant one at the top of your file:

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

Calendar_Year;Sex;Age range;Cause of death;Number of 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 5.

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

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

Calendar Year;Sex;Age range;Geo_code;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 5.

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

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

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 39 variables, semicolon-separated, and in the correct format as reported in section 3.1.1 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 new *ENCR-JRC Call for Data Protocol* as for:

- the number of variables;
- the presence of non-missing and non-blank values in the fields affecting incidence calculation;
- when applicable, the field content against a list of valid values. **Example**: patient's sex numeric value (variable 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 identification code (*variable PAT*) 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) of the 2014 JRC Technical Report "one common procedure for European cancer registries" (see also the *2022 data call protocol*).

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

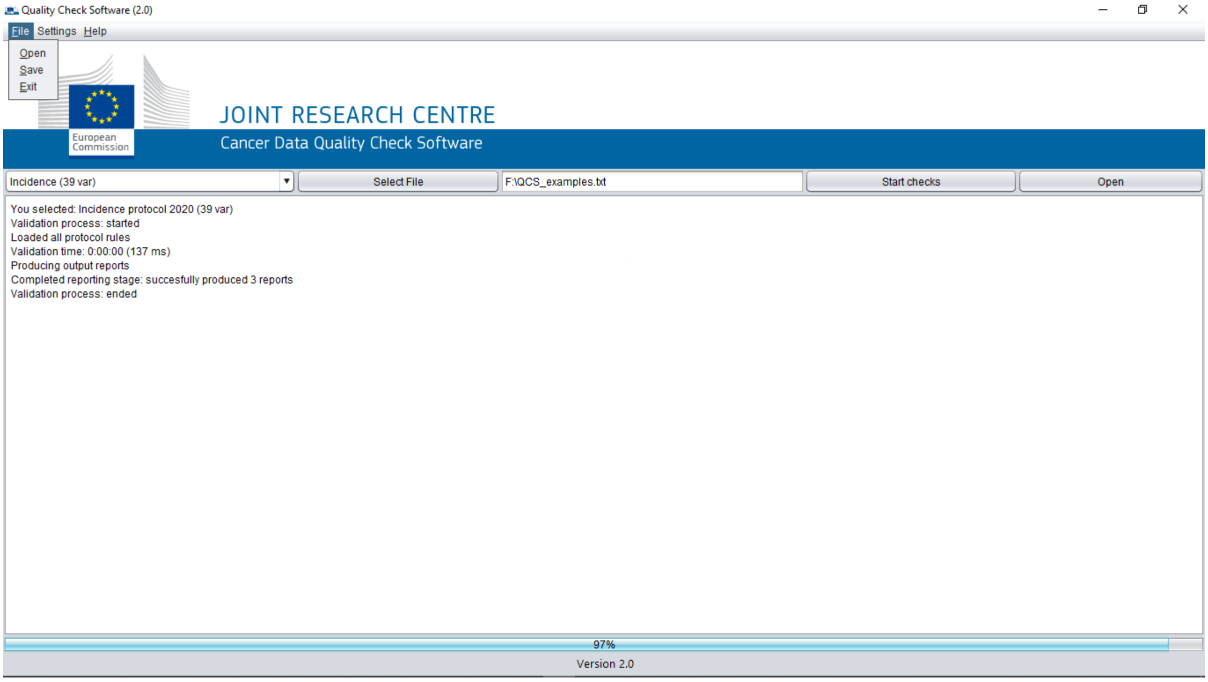
1) *QCS-Incidence-Output.pdf* – file with error and warning messages in pdf format including multiple primary tumour warnings;
2) *QCS-Incidence-Output.txt* – file with error and warning messages in *txt* format including multiple primary tumour warnings;

3) *QCS-Incidence-Output.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. Warnings for multiple primary tumours are also included in this file.

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 (In case of any issue, it is possible to try running the QCS with 2GB of RAM memory by launching file *JRC-ENCR-QCS-2GB.bat*);
- The user interface appears;



Incidence (39 var)

F:\QCS_examples.txt

```
You selected: Incidence protocol 2020 (39 var)
Validation process: started
Loaded all protocol rules
Validation time: 0:00:00 (137 ms)
Producing output reports
Completed reporting stage: succesfully produced 3 reports
Validation process: ended
```

Note: The software runs <u>only</u> double clicking on the file ending in *.bat.*

It is possible to save the current configuration of the JRC-ENCR QCS on a file, by selecting "*Save*" in menu *File*.

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

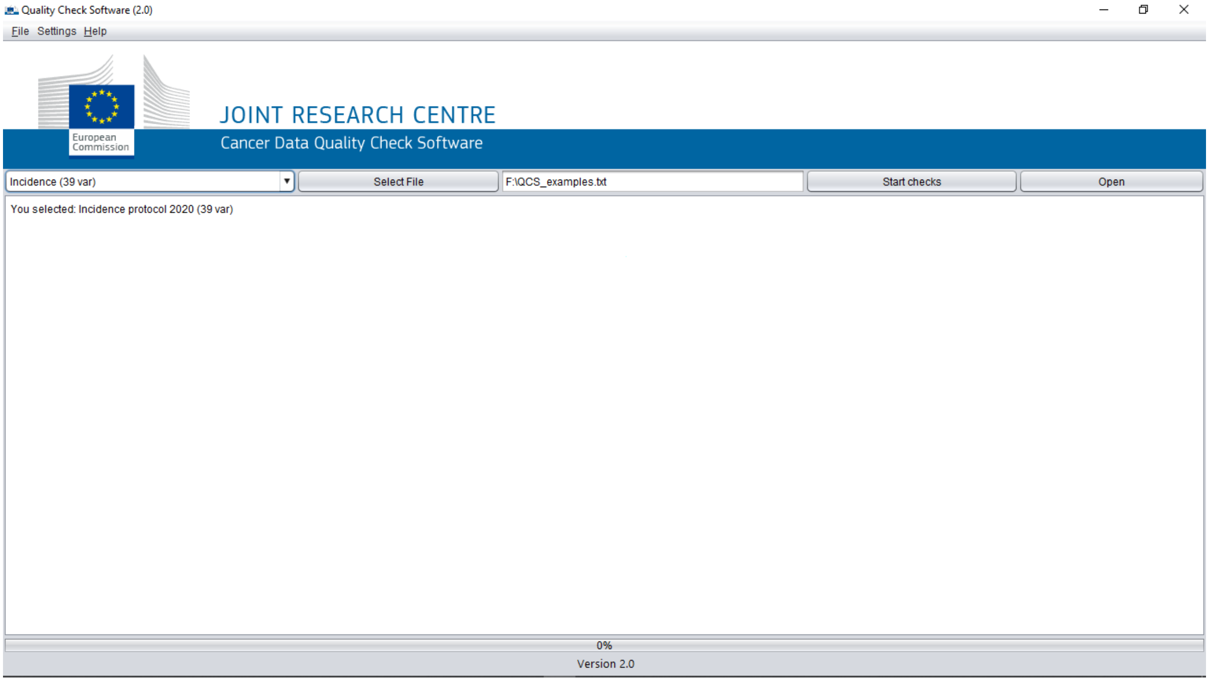
5.2 Checking the files

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

For instance, for checking an incidence file according to the 2022 data call protocol:

- Select the "*Incidence (39 var)*" 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*).



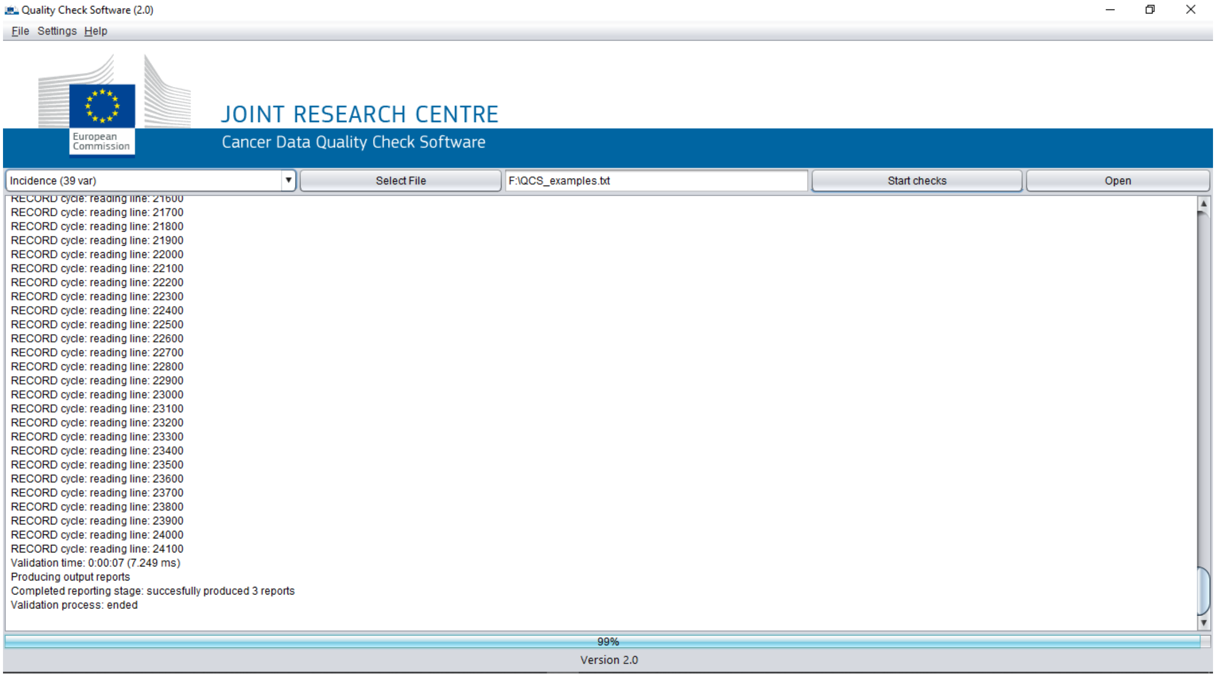
F:\QCS_examples.txt

You selected: Incidence protocol 2020 (39 var)

- Navigate to the folder where the incidence file to be checked is located, select it and press "*Open*";
- The full path of the file you have chosen will be displayed in the text box on the left of the "*Start Checks*" button;
- Press the "*Start Checks*" button;

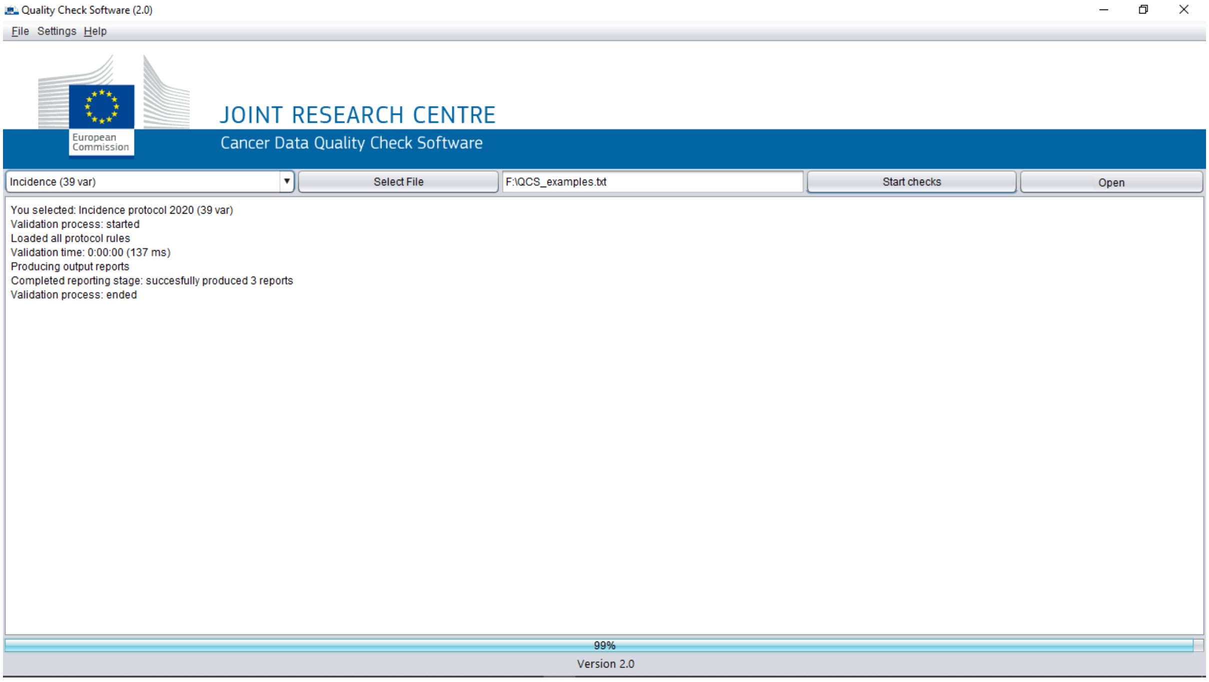
If you had previously already checked the incidence file, please note that the output files **will be overwritten**. Please save them in a different folder or with a different name in case you want to keep them.

While the software is running, the number of the checked record will appear in the display text box:



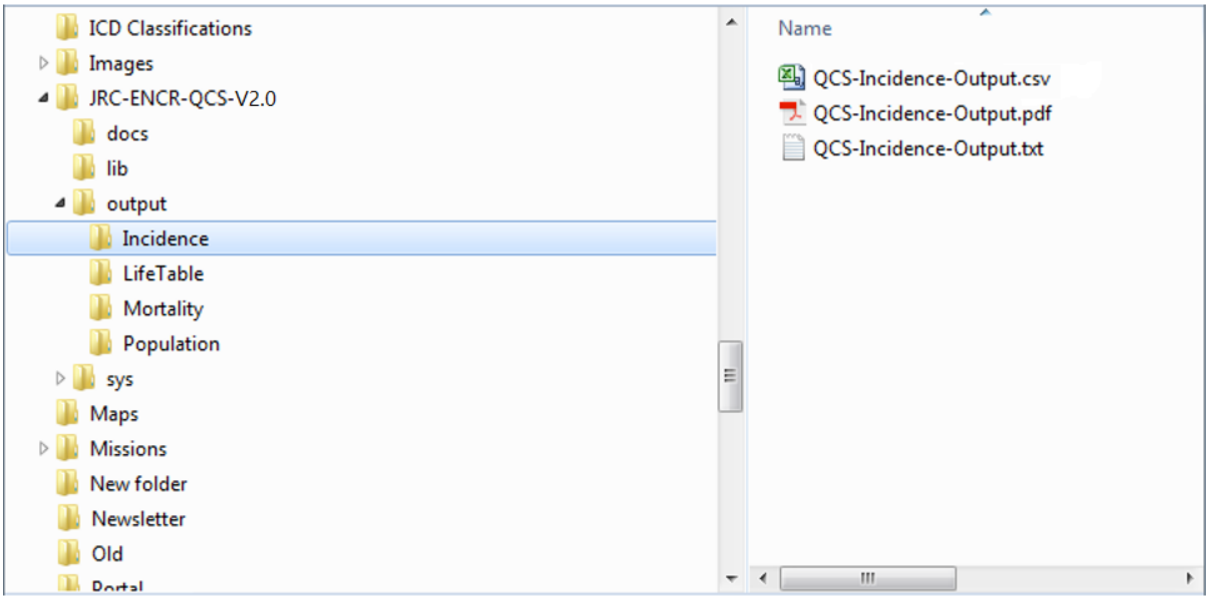
```
RECORD cycle: reading line: 21600
RECORD cycle: reading line: 21700
RECORD cycle: reading line: 21800
RECORD cycle: reading line: 21900
RECORD cycle: reading line: 22000
RECORD cycle: reading line: 22100
RECORD cycle: reading line: 22200
RECORD cycle: reading line: 22300
RECORD cycle: reading line: 22400
RECORD cycle: reading line: 22500
RECORD cycle: reading line: 22600
RECORD cycle: reading line: 22700
RECORD cycle: reading line: 22800
RECORD cycle: reading line: 22900
RECORD cycle: reading line: 23000
RECORD cycle: reading line: 23100
RECORD cycle: reading line: 23200
RECORD cycle: reading line: 23300
RECORD cycle: reading line: 23400
RECORD cycle: reading line: 23500
RECORD cycle: reading line: 23600
RECORD cycle: reading line: 23700
RECORD cycle: reading line: 23800
RECORD cycle: reading line: 23900
RECORD cycle: reading line: 24000
RECORD cycle: reading line: 24100
Validation time: 0:00:07 (7.249 ms)
Producing output reports
Completed reporting stage: succesfully produced 3 reports
Validation process: ended
```

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



```
You selected: Incidence protocol 2020 (39 var)
Validation process: started
Loaded all protocol rules
Validation time: 0:00:00 (137 ms)
Producing output reports
Completed reporting stage: succesfully produced 3 reports
Validation process: ended
```

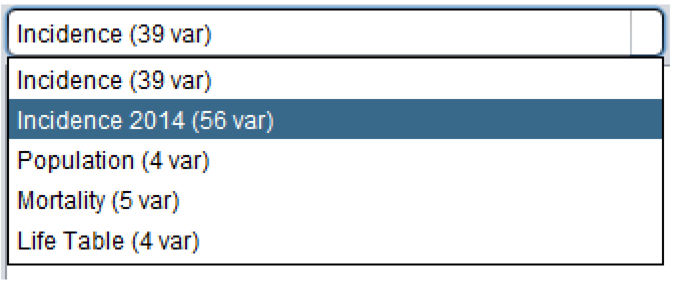
You can finally access the outputs, by clicking on "*Open*", and accessing the "*output*" folder, containing all the report files.



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.

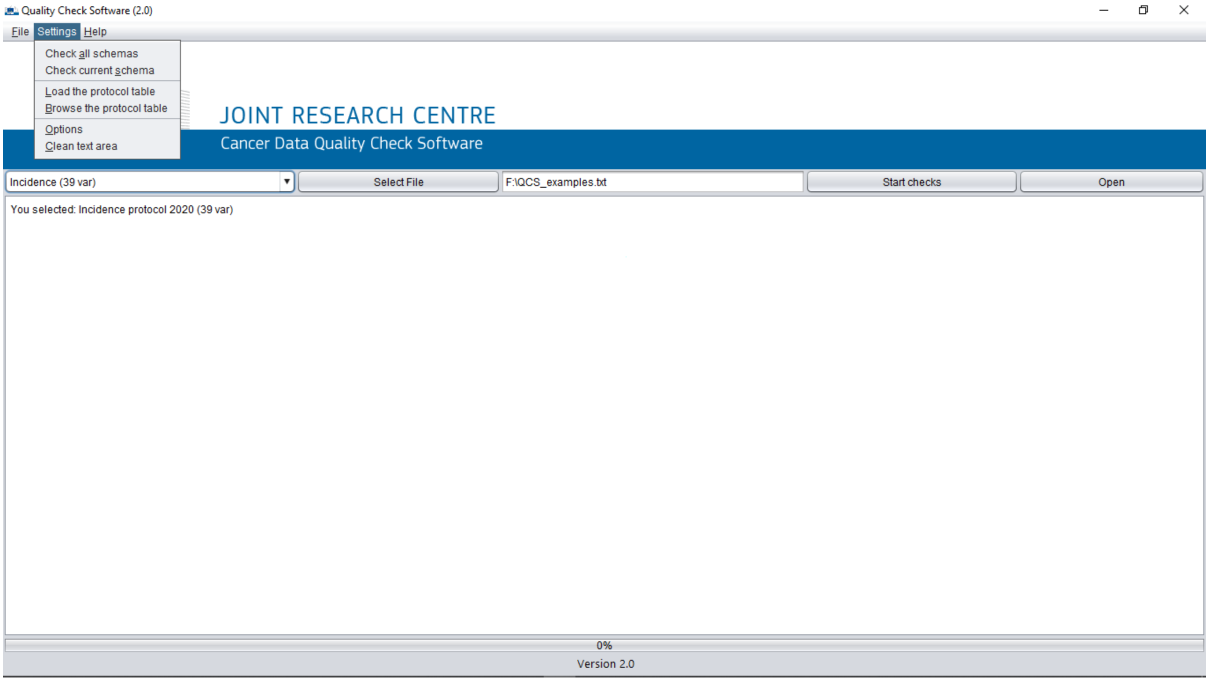
It is possible for the software to perform checks related to the previous data call protocol by selecting "Incidence 2014 (56 var)":



Check are performed according to *The JRC-ENCR Quality Check Software (QCS) for the validation of cancer registry data: user compendium – version 1.8.1* (https://encr.eu/sites/default/files/User_compendium_v1_8_1.pdf)

5.2.1 *Settings and options*

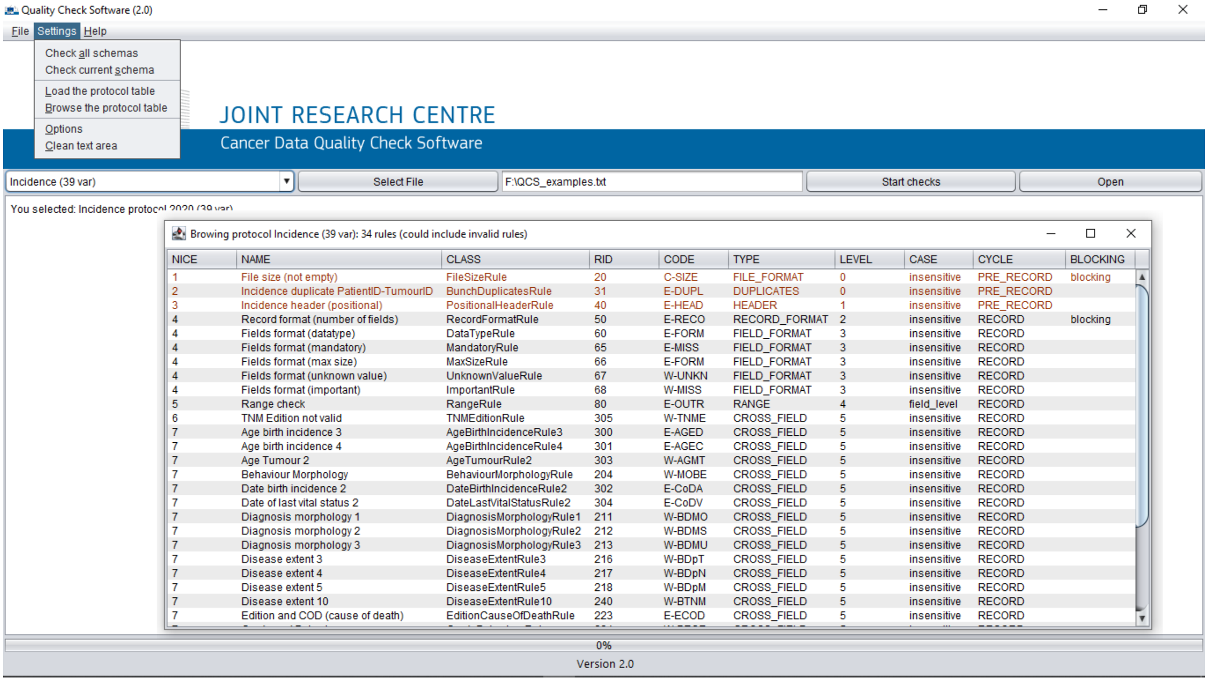
The “Settings” menu enables to select additional JRC-ENCR QCS functionalities.



You selected: Incidence protocol 2020 (39 var)

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 (see screenshot below);



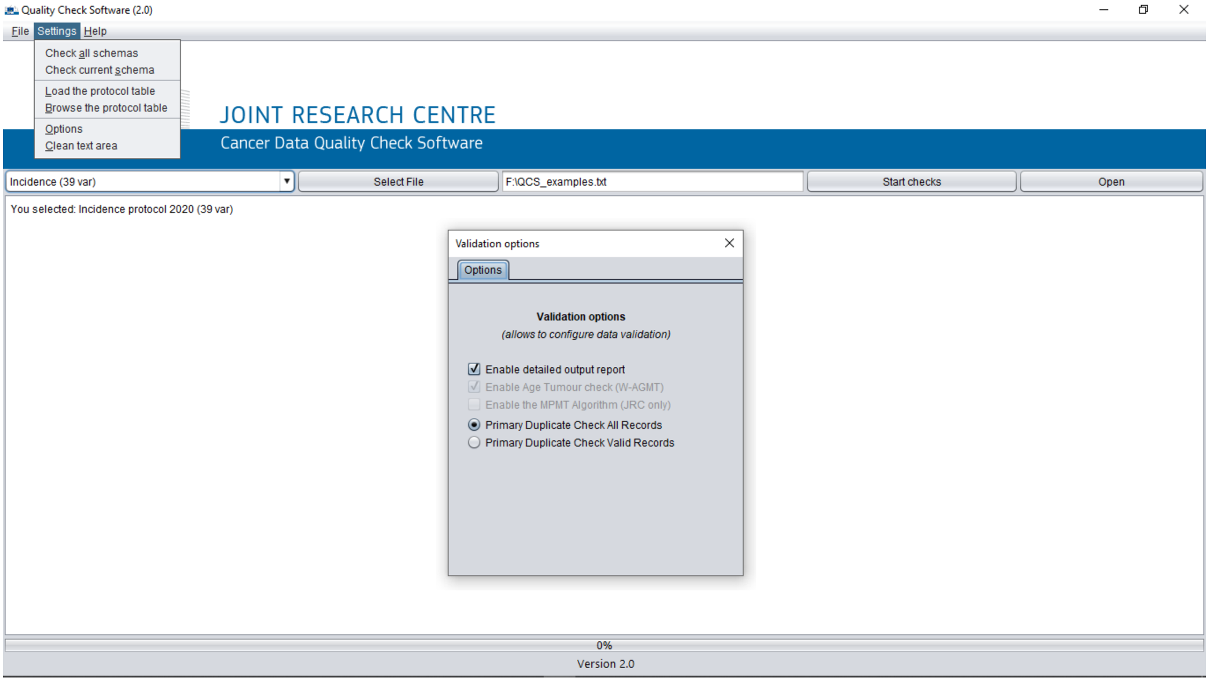
You selected: Incidence protocol 2020 (39 var)

Browing protocol Incidence (39 var): 34 rules (could include invalid rules)

NICE	NAME	CLASS	RID	CODE	TYPE	LEVEL	CASE	CYCLE	BLOCKING
1	File size (not empty)	FileSizeRule	20	C-SIZE	FILE_FORMAT	0	insensitive	PRE_RECORD	blocking
2	Incidence duplicate PatientID-TumourID	BunchDuplicatesRule	31	E-DUPL	DUPLICATES	0	insensitive	PRE_RECORD	
3	Incidence header (positional)	PositionalHeaderRule	40	E-HEAD	HEADER	1	insensitive	PRE_RECORD	
4	Record format (number of fields)	RecordFormatRule	50	E-RECO	RECORD_FORMAT	2	insensitive	RECORD	blocking
4	Fields format (datatype)	DataTypeRule	60	E-FORM	FIELD_FORMAT	3	insensitive	RECORD	
4	Fields format (mandatory)	MandatoryRule	65	E-MISS	FIELD_FORMAT	3	insensitive	RECORD	
4	Fields format (max size)	MaxSizeRule	66	E-FORM	FIELD_FORMAT	3	insensitive	RECORD	
4	Fields format (unknown value)	UnknownValueRule	67	W-UNKN	FIELD_FORMAT	3	insensitive	RECORD	
4	Fields format (important)	ImportantRule	68	W-MISS	FIELD_FORMAT	3	insensitive	RECORD	
5	Range check	RangeRule	80	E-OUTR	RANGE	4	field_level	RECORD	
6	TNM Edition not valid	TNMEditionRule	305	W-TNME	CROSS_FIELD	5	insensitive	RECORD	
7	Age birth incidence 3	AgeBirthIncidenceRule3	300	E-AGED	CROSS_FIELD	5	insensitive	RECORD	
7	Age birth incidence 4	AgeBirthIncidenceRule4	301	E-AGEC	CROSS_FIELD	5	insensitive	RECORD	
7	Age Tumour 2	AgeTumourRule2	303	W-AGMT	CROSS_FIELD	5	insensitive	RECORD	
7	Behaviour Morphology	BehaviourMorphologyRule	204	W-MOBE	CROSS_FIELD	5	insensitive	RECORD	
7	Date birth incidence 2	DateBirthIncidenceRule2	302	E-CoDA	CROSS_FIELD	5	insensitive	RECORD	
7	Date of last vital status 2	DateLastVitalStatusRule2	304	E-CoDV	CROSS_FIELD	5	insensitive	RECORD	
7	Diagnosis morphology 1	DiagnosisMorphologyRule1	211	W-BDMO	CROSS_FIELD	5	insensitive	RECORD	
7	Diagnosis morphology 2	DiagnosisMorphologyRule2	212	W-BDMS	CROSS_FIELD	5	insensitive	RECORD	
7	Diagnosis morphology 3	DiagnosisMorphologyRule3	213	W-BDMU	CROSS_FIELD	5	insensitive	RECORD	
7	Disease extent 3	DiseaseExtentRule3	216	W-BDpT	CROSS_FIELD	5	insensitive	RECORD	
7	Disease extent 4	DiseaseExtentRule4	217	W-BDpN	CROSS_FIELD	5	insensitive	RECORD	
7	Disease extent 5	DiseaseExtentRule5	218	W-BDpM	CROSS_FIELD	5	insensitive	RECORD	
7	Disease extent 10	DiseaseExtentRule10	240	W-BTNM	CROSS_FIELD	5	insensitive	RECORD	
7	Edition and COD (cause of death)	EditionCauseOfDeathRule	223	E-ECOD	CROSS_FIELD	5	insensitive	RECORD	

- *Options*. When selected, validation options are shown. Tick box *Enable detailed output report* allows the creation of either a detailed or aggregated report. A detailed report is created with the default option.

  Option *Primary Duplicate Check All Records/Valid Records* allows to have different conditions for the check of multiple primary tumours. With *Primary Duplicate Check All Records* the check is performed on valid records and on records with errors, except errors involving the tumour morphology value. By selecting *Primary Duplicate Check Valid Records*, multiple primary tumours checks are performed, except on records with the following errors/warnings: E-SETO, E-AGED, E-AGEC, E-CoDA, W-AGMT, W-MOTO and errors involving topography and morphology (see Annex 2 for the definition of error and warning codes)



You selected: Incidence protocol 2020 (39 var)

Validation options

(allows to configure data validation)

☑ Enable detailed output report

☑ Enable Age Tumour check (W-AGMT)

□ Enable the MPMT Algorithm (JRC only)

● Primary Duplicate Check All Records

○ Primary Duplicate Check Valid Records

- *Clear text area*. Deletes all the text from the dialog box.

5.2.2 Help menu

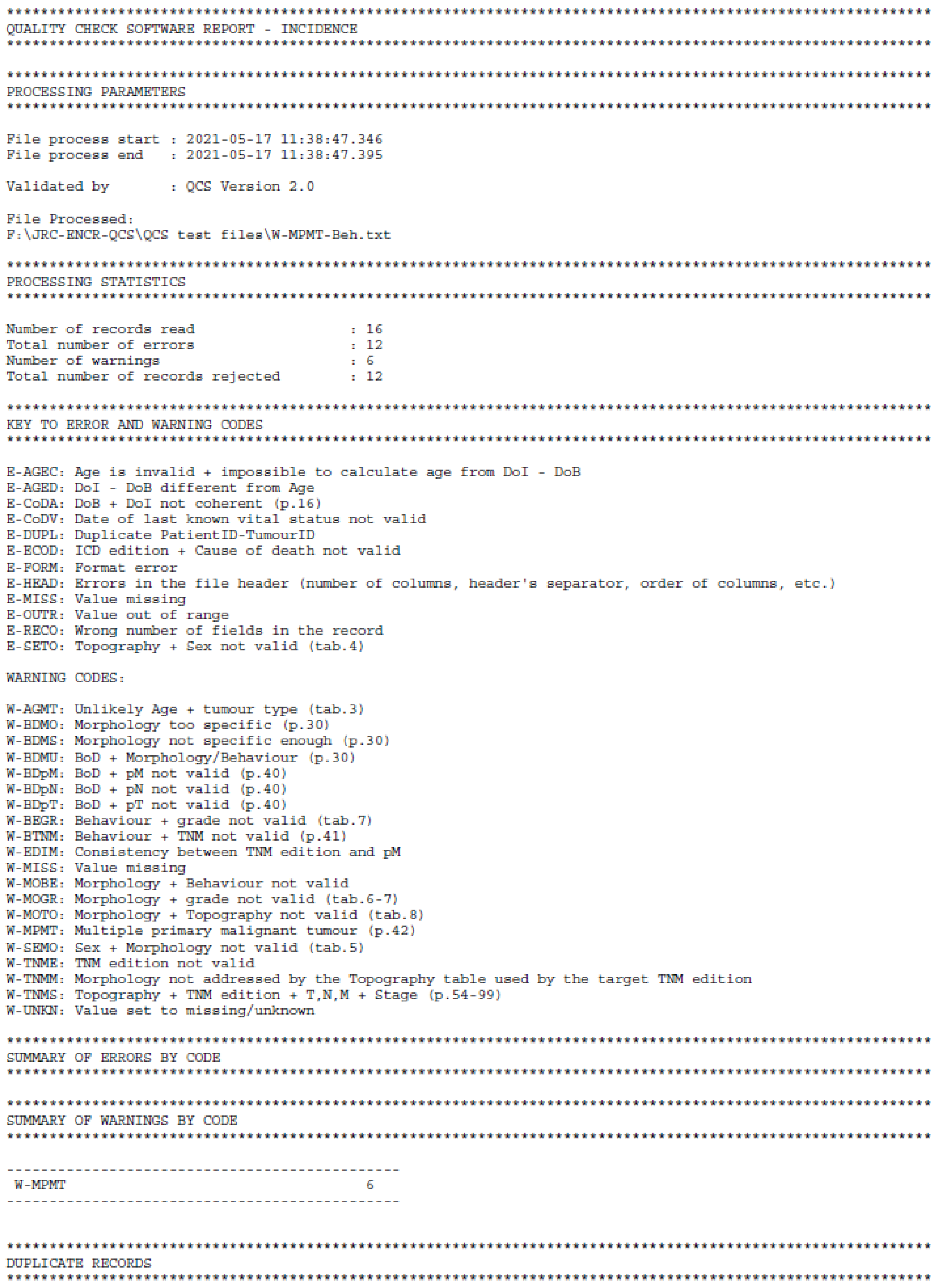
This functionality includes a link to the folder with information on the JRC-ENCR QCS, a contact e-mail and the JRC-ENCR QCS page on the ENCR website.

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

5.3 Output files

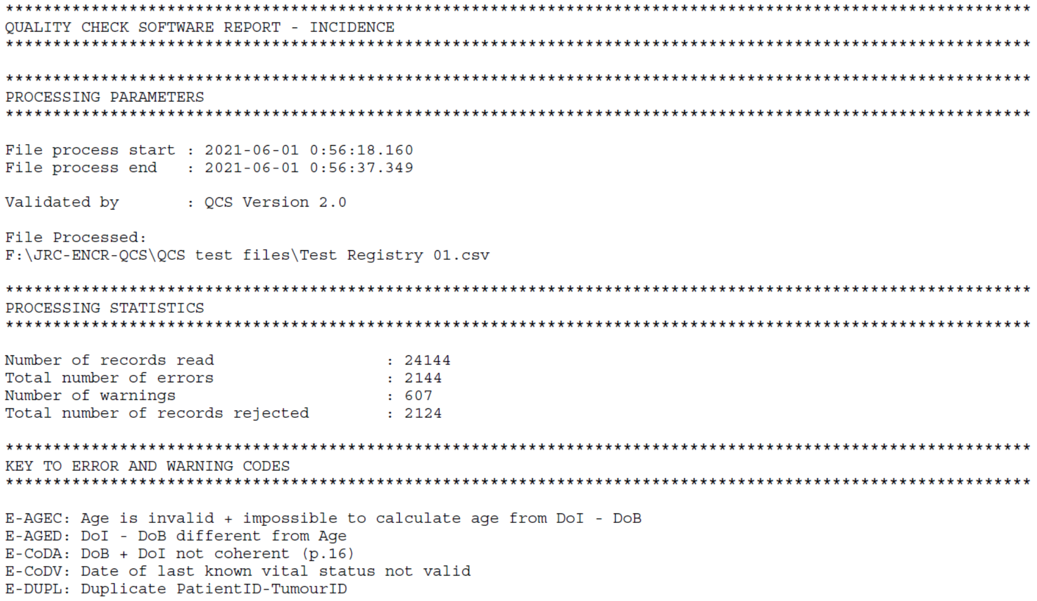
The output files are located in the subfolders inside the "*output*" folder, depending on the type of the file. For example, output files for an Incidence file are located in the "**\JRC-ENCR-QCS-V2.0\output\Incidence"** folder.

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



```
************************************************************************************************************
QUALITY CHECK SOFTWARE REPORT - INCIDENCE
************************************************************************************************************

************************************************************************************************************
PROCESSING PARAMETERS
************************************************************************************************************

File process start : 2021-05-17 11:38:47.346
File process end   : 2021-05-17 11:38:47.395

Validated by       : QCS Version 2.0

File Processed:
F:\JRC-ENCR-QCS\QCS test files\W-MPMT-Beh.txt

************************************************************************************************************
PROCESSING STATISTICS
************************************************************************************************************

Number of records read                  : 16
Total number of errors                  : 12
Number of warnings                      : 6
Total number of records rejected        : 12

************************************************************************************************************
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-CoDA: DoB + DoI not coherent (p.16)
E-CoDV: Date of last known vital status not valid
E-DUPL: Duplicate PatientID-TumourID
E-ECOD: ICD edition + Cause of death not valid
E-FORM: Format error
E-HEAD: Errors in the file header (number of columns, header's separator, order of columns, etc.)
E-MISS: Value missing
E-OUTR: Value out of range
E-RECO: Wrong number of fields in the record
E-SETO: Topography + Sex not valid (tab.4)

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 + Morphology/Behaviour (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: Behaviour + grade not valid (tab.7)
W-BTNM: Behaviour + TNM not valid (p.41)
W-EDIM: Consistency between TNM edition and pM
W-MISS: Value missing
W-MOBE: Morphology + Behaviour not valid
W-MOGR: Morphology + grade not valid (tab.6-7)
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-TNME: TNM edition not valid
W-TNMM: Morphology not addressed by the Topography table used by the target TNM edition
W-TNMS: Topography + TNM edition + T,N,M + Stage (p.54-99)
W-UNKN: Value set to missing/unknown

************************************************************************************************************
SUMMARY OF ERRORS BY CODE
************************************************************************************************************

************************************************************************************************************
SUMMARY OF WARNINGS BY CODE
************************************************************************************************************

----------------------------------------------
 W-MPMT                                   6
----------------------------------------------


************************************************************************************************************
DUPLICATE RECORDS
************************************************************************************************************
```

Detail: upper section



```
************************************************************************************************************
QUALITY CHECK SOFTWARE REPORT - INCIDENCE
************************************************************************************************************

************************************************************************************************************
PROCESSING PARAMETERS
************************************************************************************************************

File process start : 2021-06-01 0:56:18.160
File process end   : 2021-06-01 0:56:37.349

Validated by       : QCS Version 2.0

File Processed:
F:\JRC-ENCR-QCS\QCS test files\Test Registry 01.csv

************************************************************************************************************
PROCESSING STATISTICS
************************************************************************************************************

Number of records read                  : 24144
Total number of errors                  : 2144
Number of warnings                      : 607
Total number of records rejected        : 2124

************************************************************************************************************
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-CoDA: DoB + DoI not coherent (p.16)
E-CoDV: Date of last known vital status not valid
E-DUPL: Duplicate PatientID-TumourID
```

Processing parameters:

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

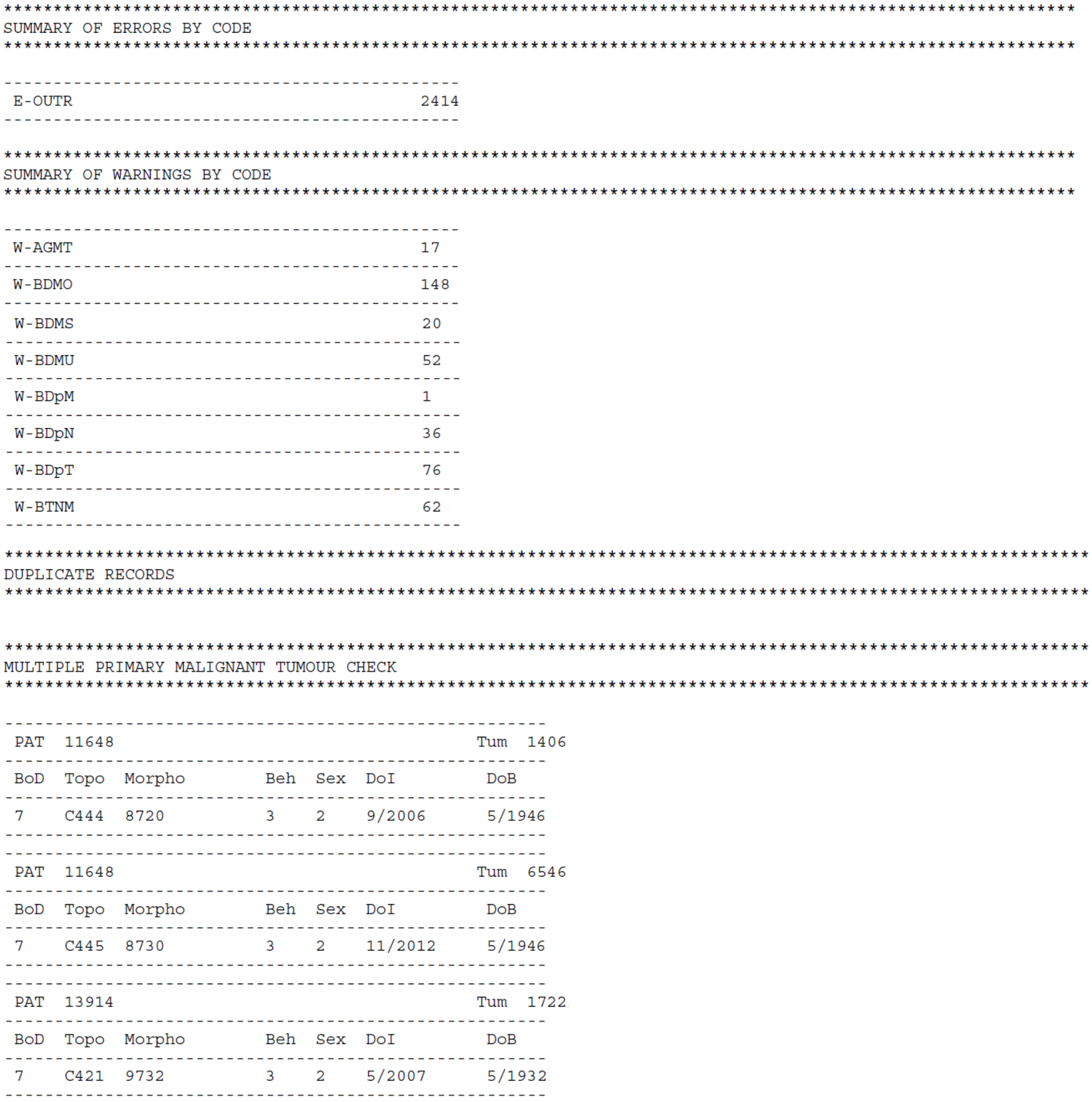
Processing statistics:

- *Number of records read, Total numbers of errors*;
- *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;

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 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 tumours)



```
****************************************************************************************************************
SUMMARY OF ERRORS BY CODE
****************************************************************************************************************

-----------------------------------------------
 E-OUTR                                    2414
-----------------------------------------------

****************************************************************************************************************
SUMMARY OF WARNINGS BY CODE
****************************************************************************************************************

-----------------------------------------------
 W-AGMT                                    17
-----------------------------------------------
 W-BDMO                                    148
-----------------------------------------------
 W-BDMS                                    20
-----------------------------------------------
 W-BDMU                                    52
-----------------------------------------------
 W-BDpM                                    1
-----------------------------------------------
 W-BDpN                                    36
-----------------------------------------------
 W-BDpT                                    76
-----------------------------------------------
 W-BTNM                                    62
-----------------------------------------------

*****************************************************************************************************************
DUPLICATE RECORDS
*****************************************************************************************************************

*****************************************************************************************************************
MULTIPLE PRIMARY MALIGNANT TUMOUR CHECK
*****************************************************************************************************************

--------------------------------------------------------
 PAT  11648                                    Tum  1406
--------------------------------------------------------
 BoD  Topo  Morpho         Beh  Sex  DoI          DoB
--------------------------------------------------------
 7    C444  8720           3    2    9/2006       5/1946
--------------------------------------------------------
--------------------------------------------------------
 PAT  11648                                    Tum  6546
--------------------------------------------------------
 BoD  Topo  Morpho         Beh  Sex  DoI          DoB
--------------------------------------------------------
 7    C445  8730           3    2    11/2012      5/1946
--------------------------------------------------------
--------------------------------------------------------
 PAT  13914                                    Tum  1722
--------------------------------------------------------
 BoD  Topo  Morpho         Beh  Sex  DoI          DoB
--------------------------------------------------------
 7    C421  9732           3    2    5/2007       5/1932
--------------------------------------------------------
```

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*

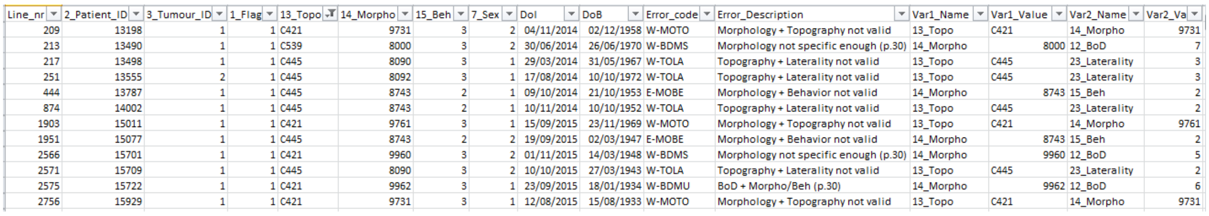
Multiple primary malignant tumour check: for each multiple primary tumour warning the following variables are reported: *PAT*, *Tum*, *BoD* (basis of diagnosis), *Topo* (topography), *Morpho* (morphology), *Beh* (behaviour), Sex, *DoI* (date of incidence), *DoB* (date of birth)

Detail: page(s) with errors and warnings

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

--------------------------------------------------------------------------------------------------
 PAT  317                                     Tum  316
--------------------------------------------------------------------------------------------------
 BoD       Topo      Morpho     Beh     Sex    DoI        DoB         Var_Name        Var_Value  Error_Code
--------------------------------------------------------------------------------------------------
 2         C724      9560       0       2      6/2005     5/1932      Autopsy         2          E-OUTR
                                                                      ----------------------------
                                                                      Morpho          9560       W-BDMO
                                                                      BoD             2          W-BDMO
--------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------------
 PAT  348                                     Tum  345
--------------------------------------------------------------------------------------------------
 BoD       Topo      Morpho     Beh     Sex    DoI        DoB         Var_Name        Var_Value  Error_Code
--------------------------------------------------------------------------------------------------
 5         C424      9871       3       2      5/2007     3/2000      Morpho          9871       W-MOTO
                                                                      Topo            C424       W-MOTO
--------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------------
```

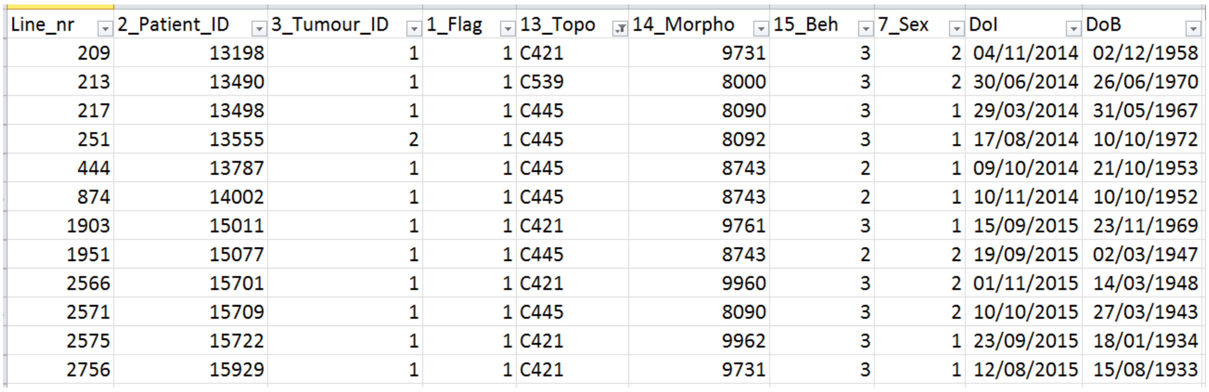
Errors and warnings: for each warning or error the following variables are reported: *PAT*, *Tum*, *Topo* (topography), *Morpho* (morphology), *Beh* (behaviour), Sex, *DoI* (date of incidence), *DoB* (date of birth), *Var_Name* and *Var_Value* (list of variables which caused the warning or error to be 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-Incidence-Output.csv* file:



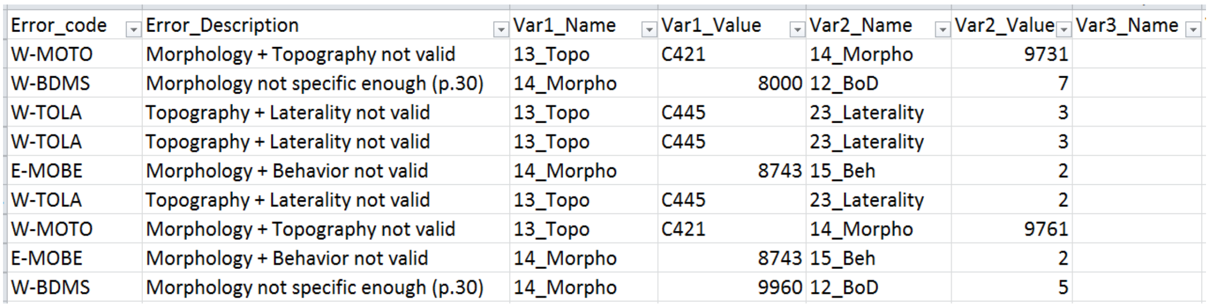
Line_nr	2_Patient_ID	3_Tumour_ID	1_Flag	13_Topo	14_Morpho	15_Beh	7_Sex	DoI	DoB	Error_code	Error_Description	Var1_Name	Var1_Value	Var2_Name	Var2_Va
209	13198	1	1	C421	9731	3	2	04/11/2014	02/12/1958	W-MOTO	Morphology + Topography not valid	13_Topo	C421	14_Morpho	9731
213	13490	1	1	C539	8000	3	2	30/06/2014	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 created by the QCS

This section describes how to interpret the outcomes of the JRC-ENCR QCS for some of the variables having an impact on the incidence estimation. Some examples of warnings on TNM and on multiple primary tumours are also reported.

The code of the errors starts 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 new *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.

```
----------------------------------------------------------------------------------------------------
 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 because topography C427 does not exist in the International Classification of Diseases for Oncology, third edition[15] (ICD-O-3).

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

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

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

```
--------------------------------------------------------------------------------------------------------------
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 error E-MISS because variable morphology (which impacts on incidence calculations) has a missing value.

- **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 error E-AGEC because variable *age* (which impacts on incidence calculations) is unknown and cannot be calculated.

- **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 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.

```
-----------------------------------------------------------------------------------------------------------
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 is gives error E-CoDA because the year of birth is later than the year of incidence.

```
-----------------------------------------------------------------------------------------------------------
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 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.

```
-----------------------------------------------------------------------------------------------------------
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 error E-CoDV because the date (year) of incidence occurs later than the date (year) of last known vital status.

```
-----------------------------------------------------------------------------------------------------------
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 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.

```
------------------------------------------------------------------------------------------------
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 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.

```
------------------------------------------------------------------------------------------------
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 warning W-AGMT because the morphology 9652 (Hodgkin lymphoma, mixed cellularity, NOS) is unlikely between ages 0-2.

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

4) **Errors and warnings due to tumour variables combinations.**

- **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, so the current version of the QCS reports as warnings such combinations. Previous versions of the QCS were reporting the morphology and behaviour combinations not listed in the ICD-O-3 as errors (E-MOBE).

```
---------------------------------------------------------------------------------------------------------------
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 W-MOBE because morphology=8621 (granulosa cell-theca cell tumour) with behaviour=3 (malignant tumour) is not listed in the ICD-O-3.

The combination of morphology and behaviour presented in the example above 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 a W-MOBE warning because morphology 9950 (polycythaemia vera) has behaviour=3 (malignant tumour) in ICD-O-3.

This term (polycythaemia vera) changed from borderline tumour (behaviour=1) in ICD-O-2[16], to malignant tumour (behaviour=3) in ICD-O-3.

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

```
---------------------------------------------------------------------------------------------------------------
PAT  000015                                        Tum  01
---------------------------------------------------------------------------------------------------------------
BoD       Topo      Morpho      Beh      Sex     DoI         DoB           Var_Name          Var_Value  Error_Code
---------------------------------------------------------------------------------------------------------------
6         C187      8210        2        2       11/1996     11/1922       BoD               6          W-BDMU
                                                                           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.

```
---------------------------------------------------------------------------------------------------------------
PAT  000016                                        Tum  01
---------------------------------------------------------------------------------------------------------------
BoD       Topo      Morpho      Beh      Sex     DoI         DoB           Var_Name          Var_Value  Error_Code
---------------------------------------------------------------------------------------------------------------
6         C421      9823        3        2       5/2013      7/1927        Morpho            9823       W-BDMU
                                                                           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).

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

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

```
----------------------------------------------------------------------------------------------------------
PAT  000017                                 Tum  01
----------------------------------------------------------------------------------------------------------
BoD      Topo     Morpho     Beh     Sex    DoI         DoB         Var_Name         Var_Value  Error_Code
----------------------------------------------------------------------------------------------------------
2        C209     8140       1       1      10/2014     11/1928     Morpho           8140       W-BDMO
                                                                    BoD              2          W-BDMO
```

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

```
----------------------------------------------------------------------------------------------------------
PAT  000018                                 Tum  01
----------------------------------------------------------------------------------------------------------
BoD      Topo     Morpho     Beh     Sex    DoI         DoB         Var_Name         Var_Value  Error_Code
----------------------------------------------------------------------------------------------------------
2        C199     8010       2       1      10/2017     1/1937      BoD              2          W-BDMO
                                                                    Beh              2          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

```
----------------------------------------------------------------------------------------------------------
PAT  000019                                 Tum  01
----------------------------------------------------------------------------------------------------------
BoD      Topo     Morpho     Beh     Sex    DoI         DoB         Var_Name         Var_Value  Error_Code
----------------------------------------------------------------------------------------------------------
7        C341     8000       3       2      10/2013     5/1943      Morpho           8000       W-BDMS
                                                                    BoD              7          W-BDMS
```

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

```
----------------------------------------------------------------------------------------------------------
PAT  000020                                 Tum  01
----------------------------------------------------------------------------------------------------------
BoD      Topo     Morpho     Beh     Sex    DoI         DoB         Var_Name         Var_Value  Error_Code
----------------------------------------------------------------------------------------------------------
7        C809     8001       3       2      5/2010      8/1934      Morpho           8001       W-BDMS
                                                                    BoD              7          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

```
-------------------------------------------------------------------------------------------------
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 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

```
-------------------------------------------------------------------------------------------------
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
```

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.

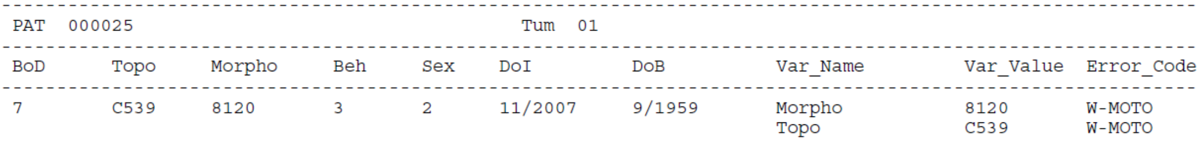
```
-------------------------------------------------------------------------------------------------
PAT  000023                                   Tum  01
-------------------------------------------------------------------------------------------------
BoD      Topo     Morpho     Beh     Sex    DoI        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 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

```
-------------------------------------------------------------------------------------------------
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
```

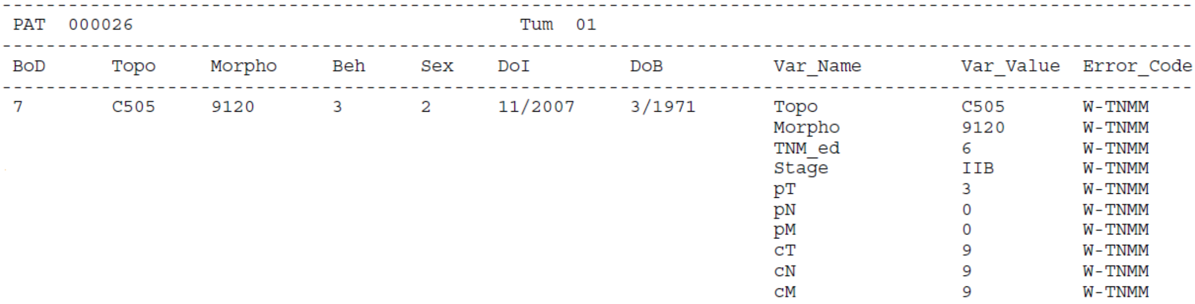
The QCS gives warning W-MOTO because topography=C779 (Lymph node, NOS) and morphology=8070 (squamous cell carcinoma, NOS); this combination is probably a metastasis and topography should be coded as C809.



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

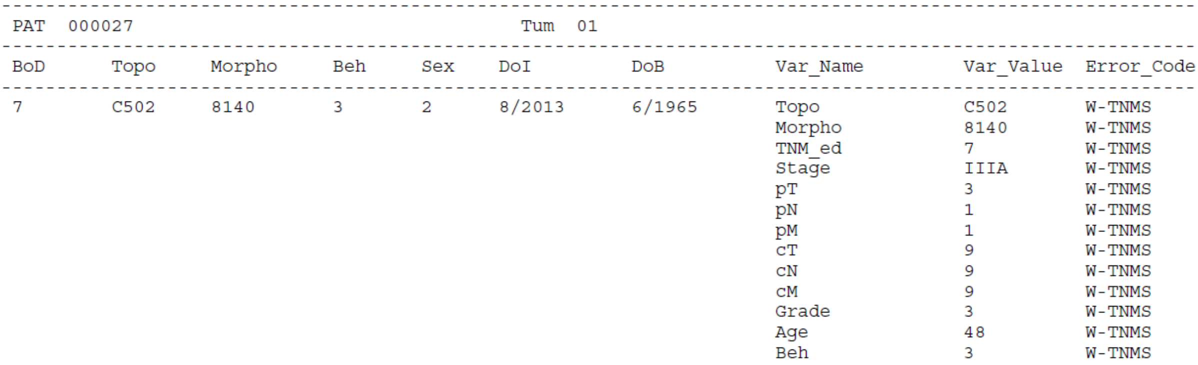
- **W-TNMM**: TNM and stage are present, but morphology is not included in the TNM



```
--------------------------------------------------------------------------------------------------------
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 warning W-TNMM because the case is a breast angiosarcoma (morphology=9120) with stage IIB. When topography=C50 (breast) only carcinomas should be staged.

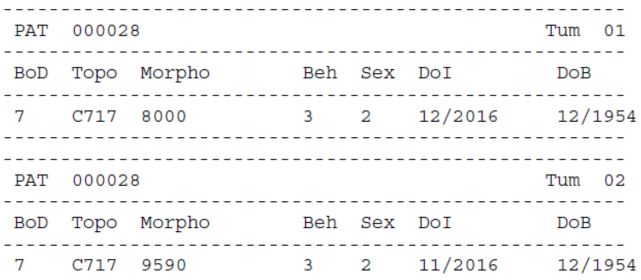
- **W-TNMS**: TNM and stage are not consistent



```
--------------------------------------------------------------------------------------------------------
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
```

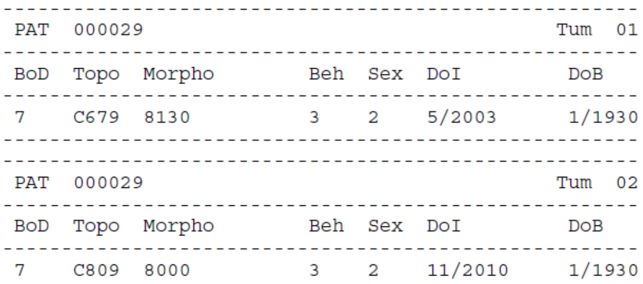
In the example above the QCS returns warning W-TNMS because the case is a breast carcinoma with pT=3, pN=1, pM=1 and Stage=IIIA. This combination is not consistent; perhaps either pM is actually 0, or stage is equal to IV.

5) Warnings for multiple primary tumours.



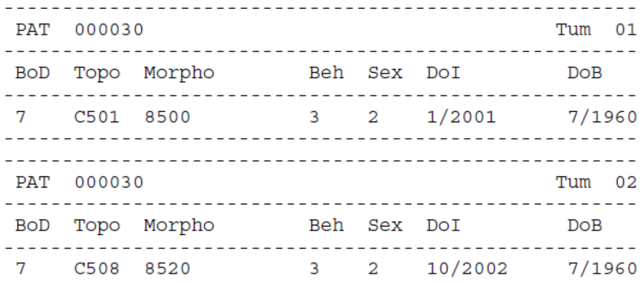
```
-----------------------------------------------------
 PAT   000028                                 Tum   01
-----------------------------------------------------
 BoD   Topo   Morpho          Beh   Sex   DoI          DoB
-----------------------------------------------------
 7     C717   8000            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 this example, the QCS gives warning for multiple primary tumours because probably the two records are the same tumour.



```
-----------------------------------------------------
 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
```

The QCS gives warning for multiple primary tumours because probably the two records are the same tumour.



```
-----------------------------------------------------
 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 this example, the QCS gives warning for multiple primary tumours 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 future improvements

The following is a list of the JRC-ENCR QCS issues that will be fixed at a later stage, and future improvements that are planned. See Annex 2 below for the definition of error and warning codes.

- The QCS accepts for variable Geo_code only NUTS codes for the following Countries: Belgium, Bulgaria, Czechia, Denmark, Germany, Estonia, Ireland, Greece, Spain, France, Croatia, Italy, Cyprus, Latvia, Lithuania, Luxembourg, Hungary, Malta, Netherlands, Austria, Poland, Portugal, Romania, Slovenia, Slovakia, Finland, Sweden, United Kingdom and Switzerland. If a value is included for this variable, and the Country is not included in the list above an E-OUTR error is raised. This can be avoided by leaving the value blank. In future releases of the QCS there will be the possibility to input Geo_code values also for the remaining European Countries.
- W-TNMS is raised incorrectly when topography is "C50", TNM edition is 6, pT is equal to "is", pN is "0", pM is "0" and stage is "0".
- Options *Enable detailed output report* and *Duplicate Check Valid Records* are not working correctly, and will be fixed in the next release of the JRC-ENCR QCS.

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-AGEC: Age is invalid or missing, and it is not possible to calculate the age by subtracting date of incidence from date of birth, since one or both dates are invalid or missing.

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

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

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-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) 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 variable *morphology* is unknown. This 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 2015 call for data protocol or the 2018 update (for instance, behaviour=6).

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

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).

[17] 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. Tumours included in the table below should also be graded[18]

W-BTNM: Invalid behaviour and TNM combination, e.g. Behaviour=3 and pT=Tis (see page 41).

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 empty. This applies to variables whose invalid/missing/unknown values don't have an impact on incidence statistics. For some of these variables is it enough to input the correct missing value (e.g. "9" for *Autopsy*) in order to avoid the warning at all.

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

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

[18] Non malignant tumours for which grade is allowed:

Topography	Morphology	Behaviour	Grade
C65-C68	8120-8131, 8020, 8031, 8082	1, 2	1-4
Any	9384, 9421, 9383, 9394, 9412, 9506	1	1
Any	9390, 9492, 9413, 9560, 9530	0	1
Any	9505	1	1, 2
Any	9361, 9539,	1	2

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

W-MPMT: Multiple primary tumour (p. 42) The quality checklist of warnings for Multiple Primary Tumours 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), with the inclusion of behaviour 2 (in situ) and behaviour 1 (uncertain and unknown behaviour) urological tumours (C65-C68) as well as behaviour 1 and behaviour 0 (benign tumours) central nervous systems tumours (C70-C72 and C751-C753) in the multiple primary tumour checks.

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

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: A variable with no impact on incidence calculations, which however could be important for quality evaluations (e.g. basis of diagnosis) or survival analysis (e.g. year of follow up) has a missing value.

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 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

The European Commission's science and knowledge service

Joint Research Centre

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

EU Science, Research and Innovation

EU Science Hub