A proposal on cancer data quality checks: one common procedure for European cancer registries Addendum to version 1.0 (November 2014)

During the last two years the European Network of Cancer Registries (ENCR) and The Joint Research Centre (JRC) have been working in preparing the 2015 ENCR-JRC call for data and developing the JRC-ENCR Quality Check (QC) software.

The JRC Technical Report '*A proposal on cancer data quality checks: one common procedure for European cancer registries*' was the basis for preparing the data call protocol and developing the Quality Check (QC) Software. Nevertheless, there are some small differences in variables considered and their respective formats between the JRC Technical Report and the call for data protocol and QC Software. The checks for multiple primary tumours were included in the QC software but they were not included in the JRC Technical report. In addition, a few errors in the report were identified.

The objective of this addendum is to update version 1.0 of the JRC Technical Report '*A proposal on cancer data quality checks: one common procedure for European cancer registries*' according to the 2015 ENCR-JRC call for data protocol, the QC Software and the feed-back obtained from the users.

Table 1. Quality checks for the variables and their format (pages 9-14)

The following is the updated version of *Table 1*; changes from the previous version are in *italics*. The 2015 ENCR-JRC call for data protocol has been integrated in *Table 1*, adding variable names and a few variables.

Variable name	Variable description	Format	Maximum length	Core	Missing /unknown values	Allowed values
1_Flag	Check flag	F	1	Y	Not allowed	Allowed values: 0 and 1 0 →Not checked 1 →Checked
2_Patient_ID	Patient identification code	A	*50*	Y	Not allowed	Not allowed to have duplicate combination of the two variables: 2_Patient_ID + 3_Tumour_ID in the same dataset
3_Tumour_ID	Tumour identification code	A	*50*	Y	Not allowed	
4_Day_DoB	Day of birth	F	2	Y	99	Range of allowed values: *from 1 to 31 and 99*
5_Month_DoB	Month of birth	F	2	Y	99	Range of allowed values: *from 1 to 12 and 99* **Warning for value = 99**
6_Year_DoB	Year of birth	F	4	Y	9999	Range of allowed values: > 1842 and ≤ the current year and 9999 **Warning for value = 9999**
7_Sex	Sex	F	1	Y	9	Allowed values: 1, 2, 3 and 9 1→Male 2→Female 3→Other **Warning for value = 9**
8_Day_DoI	Day: date of incidence	F	2	Y	99	Range of allowed values: *from 1 to 31 and 99*
9_Month_DoI	Month: date of incidence	F	2	Y	99	Range of allowed values: *from 1 to 12 and 99* **Warning for value = 99**
10_Year_DoI	Year: date of incidence ENCR recommendations: http://www.encr.eu/images/docs/recommendations/incideng.pdf	F	4	Y	Not allowed	Range of allowed values: > 1941 and ≤ the current year
11_Age	Age at diagnosis (incidence date) in years	F	3	Y*	999	Range of allowed values: ≥ 0 and < 121 **Warning for value = 999 if completed dates are not available**
12_BoD	Basis of diagnosis ENCR recommendations: http://www.encr.eu/images/docs/recommendations/basisd.pdf	F	1	Y	9	Allowed values: 0, 1, 2, 4, 5, 6, 7 and 9 0→Death certificate only (DCO) 1→Clinical 2→Clinical investigation 4→Specific tumour markers 5→Cytology 6→Histology of a metastasis 7→Histology of a primary tumour **Warning for value = 9**

F: Numeric variable; A: Alphanumeric variable; Y: yes; N: no

*If complete date of birth and/or date of incidence are missing or unknown

Table 1. (cont.)

Variable name	Variable description	Format	Maximum length	Core	Missing /unknown values	Allowed values
13_Topo	Topography (topography of the metastasis is not admitted)	A	4	Y	Not allowed	Valid code in ICD-O-3. **Warning for undefined topography when BoD is 5 or 7** C809; C76 (C760, C761, C762, C763, C764, C765, C767 and C768); C14 (C140, C148); C26 (C260, C268, C269); C39 (C390, C398, C399); C559; C579; C639; C689; C729; C759
14_Morpho	Morphology	F	4	Y	Not allowed	Valid code in ICD-O-3 **Warning for undefined morphology taking into account BoD** (See *Figure 2*, p.30)
15_Beh	ICD-O-3 behaviour	F	1	Y	Not allowed	Accepted values: 0, 1, 2, 3
16_Grade	Grade (ICD-O-3)	F	1	Y	9	Allowed values: 1, 2, 3, 4, 5, 6, 7, 8 and 9 1→Well differentiated, 2→Moderately differentiated 3→Poorly differentiated 4→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)
17_Autopsy	Incidental finding of cancer at autopsy	F	1	Y	9	Allowed values: 0, 1 and 9 0→No 1→Yes **Warning for value = 9**
18_Vital_status	The last known vital status	F	1	Y	9	Allowed values: 0, 1 and 9 1→Alive 2→Dead **Warning for value = 9**
19_Day_FU	Day of last known vital status	F	2	Y	99	Range of allowed values: **From 1 to 31 and 99**
20_Month_FU	Month of last known vital status	F	2	Y	99	Range of allowed values: from 1 to 12 and 99 **Warning for value = 99**
21_Year_FU	Year of last known vital status	F	4	Y	9999	Range of allowed values: > 1941 and ≤ the current year and 9999 **Warning for value = 9999**
22_Survival	Duration of survival in days	F	5	Y***	99999	≥ 0 **Warning for value = 99999**
23_Laterality	Laterality of paired organs	F	1	N	9	Allowed values: 1, 2, 3, 4 and 9 0→Not applicable 1→Right 2→Left 3→Unilateral NOS 4→Bilateral

F: Numeric variable; A: Alphanumeric variable; Y: yes; N: no

**If complete date of birth, data of incidence and/or date of end of follow-up are missing or unknown.

***If complete date of incidence and/or date of end of follow-up are missing or unknown.

Table 1. (cont.)

Variable name	Variable description	Format	Maximum length	Core	Missing /unknown values	Allowed values
24_Day_DoR	Day of case registration	F	2	N	99	Range of allowed values: from 1 to 31 and 99
25_Month_DoR	Month of case registration	F	2	N	99	Range of allowed values: from 1 to 12 and 99
26_Year_DoR	Year of case registration	F	4	N	9999	Range of allowed values: > 1941 and ≤ the current year and 9999
27_Cause_death	Official underlying cause of death	A	5	N	99999	Valid code in International Classification of Diseases (ICD) according to ICD edition and 99999
28_ICD_edition	ICD edition	F	2	N	99	Range of allowed values: ≥ 7 and ≤ 10 and 99
29_TNM_prefix	Additional descriptor for TNM	A	1	N	9	Allow values: y and a Prefix modifiers will be considered: **y**: stage assessed after neo-adjuvant therapy; **a**: stage determined at autopsy
30_pT	TNM stage, pathological primary site (pT)	A	10	N	9	Valid code in TNM Classification of Malignant Tumours (5th, 6th or 7th edition) and 9
31_pN	TNM stage, pathological lymph nodes (pN)	A	10	N	9	
32_pM	TNM stage, pathological metastases (pM)	A	10	N	9	
33_cT	TNM stage, clinical primary site (cT)	A	10	N	9	
34_cN	TNM stage, clinical lymph nodes (cN)	A	10	N	9	
35_cM	TNM stage, clinical metastases (cM)	A	10	N	9	
36_Stage	TNM stage grouping	A	4	N	9	Based on pathological TNM if available, or clinical TNM (when pathological TNM is not available) Valid code in TNM Classification of Malignant Tumours (5th, 6th or 7th edition) and 9
37_TNM_edition	TNM edition	F	2	N	99	Allowed values: 5, 6, 7 and 99
38_Cond_T	Condensed TNM, T	A	2	N	9	Allowed values: TL, TA, TX and 9 TL→Localised TA→Advanced TX→Unknown
39_Cond_N	Condensed TNM, N	A	2	N	9	Allowed values: N0, N1, NX and 9 N0→No regional nodes N1→Regional nodes NX→Unknown
40_Cond_M	Condensed TNM, M	A	2	N	9	Allowed values: M0, M1, MX and 9 M0→No distant metastasis M1→Distant metastasis MX→Unknown

F: Numeric variable; A: Alphanumeric variable; Y: yes; N: no

Table 1. (cont.)

Variable name	Variable description	Format	Maximum length	Core	Missing /unknown values	Allowed values
41_Dukes	Dukes' stage	A	1	N	9	Allowed values: A, B, C, D, 8 and 9 A→Dukes' stage A B→Dukes' stage B C→Dukes' stage C D→Dukes' stage D 8→not applicable
42_FIGO	FIGO stage	A	5	N	9	Allowed values: 0, I, IA, IA1, IA2, IB, IB1, IB2, IC, II, IIA, IIA1, IIA2, IIB, IIB1, IIB2, IIC, III, IIIA, IIIB, IIIC, IIIC1, IIIC2, *IV*, IVA, IVB, 8 and 9 0→FIGO stage 0 I→FIGO stage I [...] IVB→FIGO stage IVB 8→not applicable
43_AArbor	*ANN ARBOR stage*	*A*	*4*	*N*	*9*	*Allowed values: I,II,III,IV with suffixes 1,2 (for stage III), suffixes 2-9 (for stage II), 8 (not applicable) and 9*
44_Gleason	*GLEASON grading*	*F*	*2*	*N*	*9*	*Allowed values: 1, 2, 3, 8 and 9* *1→Gleason ≤ 6* *2→Gleason =7* *3→Gleason 8-10* *8→not applicable*
45_Breslow	*BRESLOW thickness*	*F*	*6*	*N*	*999.99*	*Tumour size in mm.* *>0.00-999.98*
46_EoD	Summary extent of disease	F	1	N	9	Allowed values: 1, 2, 3, 4, 5 and 9 1→Confined 2→Adjacent tissues, and/or regional lymph-nodes 3→Distant organs 4→Not confined but not specified whether code 2 or 3 applies 5→Not distant metastasis but not specified whether code 1 or 2 applies
47_Tsize	Tumour size in mm	F	3	N	999	Range of allowed values: >0 or 999
48_N_exam_nodes	Number examined nodes	F	2	N	99	Range of allowed values: 0 to 98 and 99
49_N_met_nodes	Number metastatic nodes	F	2	N	99	Range of allowed values: 0 to number of examined nodes and 99
50_Sent_nodes	Sentinel nodes	F	1	N	9	Allowed values: 1, 2, 3 and 9 1→ Done 2→ Not done 3→ Not applicable
51_Met_sent_nodes	Metastatic in sentinel nodes	F	1	N	9	Allowed values: 1, 2, 3 and 9 1→Yes 2→No 3→Not applicable

F: Numeric variable; A: Alphanumeric variable; Y: yes; N: no

Table 1. (cont.)

Variable name	Variable description	Format	Maximum length	Core	Missing /unknown values	Allowed values
52_Cfactor	C factor	F	1	N	9	Allowed values: 1, 2, 3, 4, 5 and 9 1→C1 Evidence from standard diagnostic methods only 2→C2 Evidence obtained by special diagnostic means 3→C3 Evidence from surgical exploration, including biopsy and cytology 4→C4 Evidence following definitive surgery and pathological examination of the resected specimen 5→C5 Evidence from autopsy
53_Surgery	Surgery	F	1	N	9	Allowed values: 1, 2, and 9 1→Yes 2→No
54_Systemic_th	Systemic therapy	F	1	N	9	
55_Radiotherapy	Radiotherapy	F	1	N	9	
56_BMtransp	Bone marrow transplantation	F	1	N	9	

F: Numeric variable; A: Alphanumeric variable; Y: yes; N: no

Figure 1. Range of values for the variable 'day' according to variables 'month' and 'year' (page 15)

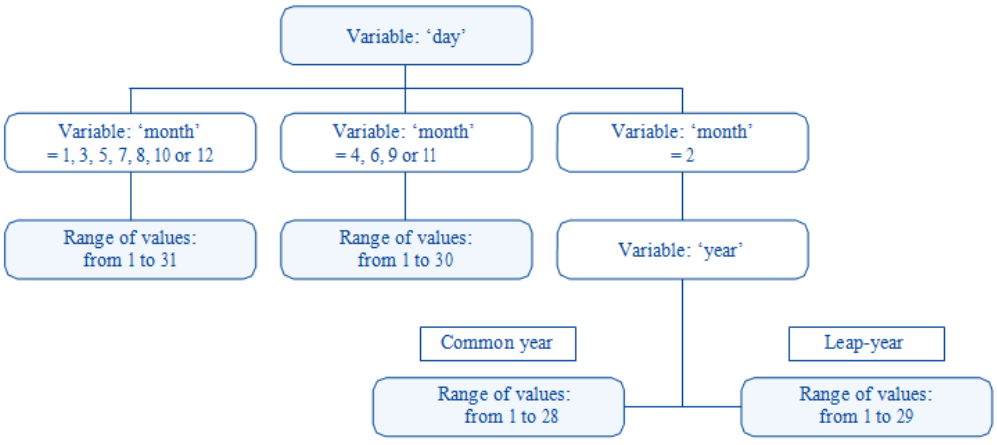


Table 3. Unlikely and rare combinations of age and tumour type (page 27)

Age group [years]	Morphology	Topography
> 7	Malignant extra-cranial and extra-gonadal germ cell: 9060-9065, 9070-*9072*, 9080-9085, 9100-9105	C00-C55, C57-C61, C63-C69, C73-*C750*, C754-C768, C80

Table 6 (Valid combinations for behaviour and topography/morphology) on page 31 will be removed in the next version of the report; it was not taken into account by the QC Software.

Table 9 Morphology codes and allowed/refused topography codes (page 36)

Morphology codes	Allowed topography codes	Not allowed topography codes
8744	C446, *C447*	

Section 3.4 (Other additional checks on the extent of the disease), page 41

The checks related to inconsistencies between behaviour and TNM were not taken into account in the QC Software when behaviour was equal to 6. These checks will be removed in the next version of the JRC Technical Report 'A proposal on cancer data quality checks: one common procedure for European cancer registries'.

Multiple primary tumours (MPT): quality checks included in the JRC-ENCR QC software

In order to include in the JRC-ENCR QC software a quality checklist of warning was developed by the JRC according to the current International Rules for Multiple Primary Cancers published in 2004 (http://www.encr.eu/images/docs/recommendations/MPrules_july2004.pdf).

This checklist was not included in the version 1.0 of the JRC Technical Report 'A proposal on cancer data quality checks'. Nevertheless, it is included in the addendum to clarify the checks and the warning message given by the QC Software.

A large variability of the MPT proportion was found among the European cancer registries which took part in the 2015 ENCR-JRC call for data. This common checklist, included in the JRC-ENCR QC Software, will improve data comparability among registries.

The steps for checking MPT each pair of malignant tumours are as following:

1. **How topographies are compared according to the 2004 MPT rules.**

   In addition to the groups of topography codes considered as a single site in the definition of MPT (table 1 of the 2004 MPT rules), **for checking** other groups are considering as a single topography:

Table A additional groups of topography codes considered as a single site

Topography code	Definition
C15 C26	Esophagus Other and ill-defined digestive organs

Table A. (cont.)

Topography code	Definition
C16 C26	Stomach Other and ill-defined digestive organs
C17 C26	Small intestine Other and ill-defined digestive organs
C18 C26	Colon Other and ill-defined digestive organs
C19 C26	Rectosigmoid junction Other and ill-defined digestive organs
C20 C26	Rectum Other and ill-defined digestive organs
C21 C26	Anus and anal canal Other and ill-defined digestive organs
C22 C26	Liver and intrahepatic bile ducts Other and ill-defined digestive organs
C23 C26	Gallbladder Other and ill-defined digestive organs
C24 C26	Other and unspecified parts of biliary tract Other and ill-defined digestive organs
C25 C26	Pancreas Other and ill-defined digestive organs
C30 C39	Nasal cavity and middle ear Other and ill-defined sites within respiratory system and intrathoracic organs
C31 C39	Accessory sinuses Other and ill-defined sites within respiratory system and intrathoracic organs
C32 C39	Larynx Other and ill-defined sites within respiratory system and intrathoracic organs
C33 C39	Trachea Other and ill-defined sites within respiratory system and intrathoracic organs
C34 C39	Bronchus and lung Other and ill-defined sites within respiratory system and intrathoracic organs
C37 C39	Thymus Other and ill-defined sites within respiratory system and intrathoracic organs
C38 C39	Heart, mediastinum, and pleura Other and ill-defined sites within respiratory system and intrathoracic organs
C51 C57	Vulva Other and unspecified female genital organs
C52 C57	Vagina Other and unspecified female genital organs
C53 C55 C57	Cervix uteri Uterus, NOS Other and unspecified female genital organs
C54 C55 C57	Corpus uteri Uterus, NOS Other and unspecified female genital organs
C56 C57	Ovary Other and unspecified female genital organs
C60 C63	Penis Other and unspecified female genital organs

Table A. (cont.)

Topography code	Definition
C61 C63	Prostate gland Other and unspecified female genital organs
C62 C63	Testis Other and unspecified female genital organs

C76 (other and ill-defined sites) and **C80** (unknown primary site) are considered as a single site with any topography.

a. **If the two topographies are in the same group, the two morphologies are compared (step 2).**
b. **If the two topographies are in different group,** no warning message is given.

2. **How morphologies are compared according to the 2004 MPT rules.**

Some unspecified morphologies were included **for checking** in the groups defined in table 2 of the 2004 MPT rules, and are considered as the same morphology. The groups of morphologies considered to be histologically "different" are:

Table B groups of morphologies

Morphology	Morphology code
Squamous and transitional cell carcinoma Unspecified carcinomas (NOS) Unspecified types of cancer	8051-8084, 8120-8131 8010-8015, 8020-8022, 8050 8000-8005
Basal cell carcinomas Unspecified carcinomas (NOS) Unspecified types of cancer	8090-8011 8010-8015, 8020-8022, 8050 8000-8005
Adenocarcinomas Unspecified carcinomas (NOS) Unspecified types of cancer	8140-8149, 8160-8162, 8190-8221, 8260-8337/3, 8350-8551, 8570-8576, 8940-8941 8010-8015, 8020-8022, 8050 8000-8005
Other specific carcinomas Unspecified carcinomas (NOS) Unspecified types of cancer	8030-8046, 8150-8157, 8170-8180, 8230-8255, 8340-8347, 8560-8562, 8580-8671 8010-8015, 8020-8022, 8050 8000-8005
Sarcomas and soft tissue tumours Unspecified types of cancer	8680-8713, 8800-8921, 8990-8991, 9040-9044, 9120-9125, 9130-9136, 9141-9252, 9370-9373, 9540-9582 8000-8005
Mesothelioma Unspecified types of cancer	9050-9055 8000-8005
Myeloid Unspecified types of haematopoietic and lymphoid tissues Unspecified types of cancer	9840, 9861-9931, 9945-9946, 9950, 9961-9964, 9980-9987 9590-9591, 9596, 9727, 9760, 9800-9801, 9805, 9820, 9832, 9835, 9860, 9960, 9970, 9975, 9989 8000-8005

Table B. (cont.)

Morphology	Morphology code
B–cell neoplasms Unspecified types of haematopoietic and lymphoid tissues Unspecified types of cancer	9670-9699, 9728, 9731-9734, 9761-9767, 9769, 9823-9826, 9833, 9836, 9940 9590-9591, 9596, 9727, 9760, 9800-9801, 9805, 9820, 9832, 9835, 9860, 9960, 9970, 9975, 9989 8000-8005
T–cell and NK–cell neoplasms Unspecified types of haematopoietic and lymphoid tissues Unspecified types of cancer	9700-9719, 9729, 9768, 9827-9831, 9834, 9837, 9948 9590-9591, 9596, 9727, 9760, 9800-9801, 9805, 9820, 9832, 9835, 9860, 9960, 9970, 9975, 9989 8000-8005
Hodgkin lymphoma Unspecified types of haematopoietic and lymphoid tissues Unspecified types of cancer	9650-9667 9590-9591, 9596, 9727, 9760, 9800-9801, 9805, 9820, 9832, 9835, 9860, 9960, 9970, 9975, 9989 8000-8005
Mast–cell tumours Unspecified types of haematopoietic and lymphoid tissues Unspecified types of cancer	9740-9742 9590-9591, 9596, 9727, 9760, 9800-9801, 9805, 9820, 9832, 9835, 9860, 9960, 9970, 9975, 9989 8000-8005
Histiocytes and Accessory Lymphoid cells Unspecified types of haematopoietic and lymphoid tissues Unspecified types of cancer	9750-9758 9590-9591, 9596, 9727, 9760, 9800-9801, 9805, 9820, 9832, 9835, 9860, 9960, 9970, 9975, 9989 8000-8005
Kaposi sarcoma Unspecified types of cancer	9140/3 8000-8005
Other specified types of cancer Unspecified types of cancer	8720-8790, 8930-8936, 8950-8983, 9000-9030, 9060-9110, 9260-9365, 9380-9539 8000-8005

a. **If the two topographies and the two morphologies are in the same group, a warning message for MPT is given; the two cases should be reviewed by the cancer registry staff.**

b. **If the two morphologies are in different group, the two cases are considered multiple primary tumours; no warning message is given.**