

The new DEPeditis
BETA version, distributed July 2008

New DEP edits brief help manual last update: 22/9/2008

What is the new DEPeditis software?

The new DEPeditis is a package that can assist registries in complying with the requirements of the EUROCIM 2, 2008 call for data. It has 5 independent modules that perform specific functions:

1. **Checks:** Performs all the edits for the EUROCIM call.
2. **Paediatric:** Performs all the ACCIS required edits, in addition to the regular edits
3. **Conversion:** Converts the ICD-9/10 based site/morphology codes to ICD-O-3 codes.
4. **Unduplicate:** Assists in the identification of multiple records for the same patient (Unduplicates the records at the patient level).
5. **Multiple Primary:** Assists in the identification of multiple tumors for the same patient, using the 2004 rules.

How do I use the new DEPeditis?

As DEPeditis is not a database management software, but rather a tool to assist you identify problems and correct them, a 'clean' data submission for the call involves an iterative process (Note that the program runs significantly faster if the data set is in the same drive as the DEPeditis. Put the data file on your machine):

Step 1

Extract from your database a tab-delimited file containing the variables names on the first line, following the EUROCIM 2 call for data record layout. Recode the records as needed to adhere to the coding schemes described in the document (<http://www.encr.com.fr>).

Step 2

Run the various modules, as needed. For example, if your registry has its own software to unduplicate records at the patient and tumor level, skip these modules and run the checks directly.

Step 3.

Use the output of each module to look up the cases and make the appropriate corrections in your database.

Step 4

Re-extract the data and re-run the same module to make sure all errors are corrected.

ALL registries have to run – at a minimum- the 'Checks' module and make the appropriate corrections to their data before submitting them for the EUROCIM 2008 call (assuming that they unduplicate their records with their own software. If not, then they should run the 'Unduplicate' and 'Multiple primary' modules as well).

Useful tips:

As with most computer programs the new DEPedit runs significantly faster when accessing the data from a local drive!

If it takes very long time to load and 'check' a data set, please check the following:

- Is your data file on your machine? If not, copy it on your machine prior to running the DEPedit.
- Make sure the data are in the format requested (tab delimited file with the variable names on top), as described in the call for document.
- This version of DEPedit requires a recent version of the Java Runtime Environment, check if you have it (download from <http://www.java.com/en/download/manual.jsp>)

We tested all modules in medium-large size files and the reported times are:

Loading files:

I tested a file with 115,000 records and it takes less than 30 seconds to load.

Conversions:

413,000 cases take 1 minute to load and just under 3 minutes to do the ICD10/02 to ICD02 conversion. The times for the ICD02 -> ICD03 are similar. (Courtesy of Andy Smith – Trent Cancer Registry – Many thanks!)

Checks:

A file with 115,000 records all of which have multiple errors (which means that they create unusually huge trafficking for outputting multiple error messages for each record), it took less than 90 minutes to run.

Frequently asked questions

What kind of edits does this call involve?

The old DEP edits, in addition to new edits added to accommodate the new data items requested. More specifically, the following data items are checked for validity:

Incidence Date
Birth Date
Age at Diagnosis
Date of Last Contact
Death Date
Sex
Topography
Morphology
Behaviour
Grade
Basis of Diagnosis
Condensed TNM
T/N/M
Grouped Stage
Laterality
Cause of Death

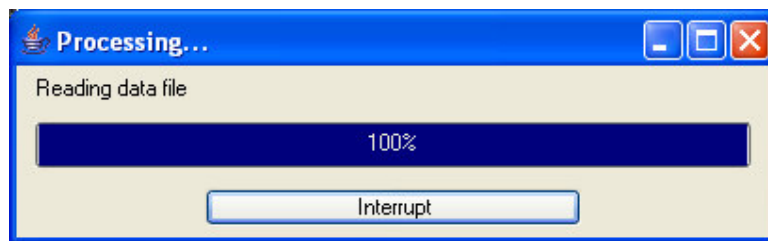
Inter-record edits are also included, to spot logical errors and to check the consistency between data items:

Incidence Date and Birth Date
Age at Diagnosis, Incidence Date and Birth Date
Incidence Date and Date Last Contact
Incidence Date and Death Date
Vital Status , Date of Last Contact and Death date
Vital Status and Cause Death
Incidence Date, Death date and Survival
Topography & Morphology
Topography & Laterality
Morphology & Behaviour
Sex and Topography
Sex and Histology
Age and Topography
Age and Histology
Age, Topography and Histology
Topography and Behavior
Topography and Laterality
Grouped Stage and Behavior
TNM Stage and Behavior
TNM stage and Morphology

1. What is the best way to start with this data request?

- Read very carefully the **call instructions** and especially the ‘**record layout-check list**’ which you need to print and have it handy at all times. **Most answers on the coding schemes and instructions for populating the fields correctly (to pass the ‘checks’)** are there. If you come across situations for which there are no standards, or the instructions are not clear, or the way you collect the data can not be translated to the coding schemes requested, please contact me immediately votil@iarc.fr
- Extract a small data set and try the new DEPEdits on it, before you run the edits on the complete data set.
- Install your data set on your machine
- Check if you have a recent version of the Java Runtime Environment. **This version of DEPEdits requires a recent version of the Java Runtime Environment (download from <http://www.java.com/en/download/manual.jsp>)**

2. The DEPEdits Check Program doesn't work correctly ! See the Pop Up Window below. This is the only message I get. What is the problem?



You version of the Java Runtime Environment. This version of DEPEdits requires a recent version of the Java Runtime Environment (download from <http://www.java.com/en/download/manual.jsp>)

3. My registry does not collect laterality, how do I populate this field?

Check the record layout for the values accepted. Code all non-paired organ as 0 and all paired organ as 9.

The following are paired organ sites and MUST have a code other than 0 for laterality. The code 9 is allowed if the registry doesn't collect or can not recode laterality to the coding scheme requested:

C079 Parotid gland
C080 Submandibular gland
C081 Sublingual gland
C090 Tonsillar fossa
C091 Tonsillar pillar
C098-C099 Tonsil, NOS
C301 Middle ear
C310 Maxillary sinus

C312 Frontal sinus
 C341-C349 Lung
 C384 Pleura
 C400 Long bones of upper limb, scapula and associated joints
 C401 Short bones of upper limb and associated joints
 C402 Long bones of lower limb and associated joints
 C403 Short bones of lower limb and associated joint
 C441 Skin of eyelid
 C442 Skin of external ear
 C443 Skin of other and unspecified parts of face (midline code '9')
 C445 Skin of trunk (midline code `9')
 C446 Skin of upper limb and shoulder
 C447 Skin of lower limb and hip
 C471 Peripheral nerves and autonomic nervous system of upper limb and shoulder
 C472 Peripheral nerves and autonomic nervous system of lower limb and hip
 C491 Connective, subcutaneous, and other soft tissues of upper limb and shoulder
 C492 Connective, subcutaneous, and other soft tissues of lower limb and hip
 C500-C509 Breast
 C570 Fallopian tube
 C620-C629 Testis
 C630 Epididymis
 C631 Spermatic cord
 C649 Kidney, NOS (exception if morphology=8960, behaviour=3: laterality=0 is possible)
 C659 Renal pelvis
 C669 Ureter
 C690-C699
 C740-C749 Adrenal gland
 C754 Carotid body

Note: Ovary C569 is not contained in the above list as there is a conflict with the multiple primary rules

4. My registry collects the month of birth of the patient but doesn't want to submit it, how do I populate this field for the data submission?

All components of the dates should be treated as separate. Use 99 for the month and the year of birth (example: 991968)

5. Do I use the letters T, N, M to report the 3 components of the TNM stage?

NO! There are 2 bytes allowed simply because the valid codes for T, N, M can be alphanumeric and of length 2.

The following codes are valid for TNM path T:

X,0,A,IS,SU,SD,1M,1,1A,A1,A2,1B,B1,B2,1C,2,2A,2B,2C,3,3A,3B,3C,4,4A,4B,4C,4D

The following codes are valid for TNM path N:

X,0,1,1A,1B,1C,1M,2,2A,2B,2C,3,3A,3B,3C,I+,I-,M+,M-

The following codes are valid for TNM clinical or path M:
X,0,1,1A,1B,1C,1M

6. My registry collects some in-situ cancers and the call asks to submit cancers of any behavior. How do I code the condensed TNM for them?

As the ENCR standards do not provide a specific code for the in-situs, leave this field blank to distinguish from the cases of unknown stage.

7. My registry collects cancers with behaviour 6 and 9. Do I submit them?

Yes, at the recommendation of the ENCR Steering Committee meeting held on 2-3/9/2008

8. My registry collects basal squamous cell skin cancers (non-melanomas), do I submit them?

No, only melanomas are collected. Basal cell and squamous cell carcinomas are excluded.

9. We don't use sequence number at the registry, how do I code it?

Use 00 if person has only one cancer, and 01 and 02 if person have two cancers etc.

10. Error messages are very detailed with every possible combination of errors. I'm lost with the output and don't know how to optimally use it....Help!

The way the software is written it outputs one record per error found. When you have multiple errors that need to be corrected in the record, it will output equally many lines. I will look into this and see if it can output only one error message at a time when multiple errors exist in the same record. In the mean time, sort the records by error message and try to solve problems in groups of records. You will find trends that you can correct globally in your database, or global recodes needed.

11. The definition of paired organs differs from we use in my registry. For instance Sinus C310, C312, Skin C441-C445 are not considered paired organs.

Sorry...I had to go by a standard in order to develop the edits and in the absence of IACR or ENCR lists of paired organs, I used the list developed by NAACR and excluded from the edits those sites that conflicted with the IACR multiple primary rules. Please code laterality based on the list provided for the call.

12. Why do you use SEER laterality instead of the ENCR?

Because there is no ENCR standard... In the absence of both IACR and ENCR standard for laterality, I had to use a standard and I picked the only existing one,

the SEER standard. During the pilot project for testing the record layout I heard from a few European registries that they have adopted the SEER standard as well.

13. In Variable 13, vital status, we would like to use the code 3 for 'emigrated/disappeared', is it possible?

Unless all registries follow the same coding scheme, it is impossible to use the same edits for checking the data... That is, we would have to write code for 183 registries to accommodate their coding schemes....So...sorry , I can't accept it. Please recode 3 to 9, as in essence the vital status is unknown.

14. Conversions:

We are using ICD-10 topography and ICD-0-2 Morphology, and we convert the data first to first to ICD02 and then to ICD03.

a) Some records failed for Morphology 8010 when combined with site C22.9 (Liver). We think this should be allowed...

We checked the code of DEPEdits and you found that there is a rule requiring morphology 8170 and 8171, and not allowing morphology 8010 for C22.9. We agree with you that it should be acceptable. Until we correct the code, please don't change any data, just ignore the error message.

b) Some record failed because of Morphology 8046 with behaviour 3 (site=C34) which our registration manager thinks should be valid as this is the code for 'Non small cell carcinoma'.

We checked the code and it looks that the morphology 8046 does not exist in the list of conversions. It needs to be added, in the mean time please ignore this error message.

Courtesy of Andy Smith, Trent Cancer Registry.
Many thanks for helping us improve the DEPEdits!