

DEPedit – Conversion and check programs for cancer registries

The *DEPedit* package is provided to cancer registries that wish to submit data to the Descriptive Epidemiology group (DEP) of IARC for various projects managed by the group, or for their internal use. It includes programs for conversion from various *International Classification of Diseases* (ICD) or various *International Classification of Diseases for Oncology* (ICD-O) systems into ICD-O-3 (if necessary). The data file can then be checked using a check program (ICD-O-3 based), which will perform edits on selected variables.

The *DEPedit* package is similar to *IARCcrgTools* (version 2.02, November 2005) available from the IACR web site (http://www.iacr.com.fr/iacr_iarcrcrgtools.htm). All the conversions and checks available in *IARCcrgTools* are included in *DEPedit*, but, in addition to these, *DEPedit*:

1. Performs specific checks for survival data
2. Performs specific checks for childhood data
3. Includes the SEER validation of ICD-O-3 (T) / ICD-O-3 (M) combinations (<http://seer.cancer.gov/>).

System requirements:

- Windows 95/98/Me/NT/2000/XP (NT/2000/XP recommended)
- 5 Mb hard disk space requested
- 64 Mb central memory recommended

The conversion and check programs work in batch mode, and can only process **text files which have a fixed field format** (information is aligned into columns of equal width, so that all records have the same length). The file format, i.e. the position of each variable within a record is **free**, but must be the same throughout the file. The first task you have to do is to create this text file with your data.

You can use the *File/Transfer* facility to convert a text file from delimited (such as .csv, tab delimited text .txt) to fixed field format. Typically, a fixed field format can be created using the 'Save as' (formatted text, space delimited *.prn) option in Microsoft Excel.

Input file example (1)

```
223612 2 56 19340808 19910503 C504 85003 1
601426 2 84 19052507 19892809 C180 81403 1
601476 2 51 19371304 19882212 C509 81403 1
601799 2 75 19143108 19901501 C343 82603 1
601975 1 66 19230505 19891809 C185 81403 1
603437 2 39 19512307 19911601 C504 85003 1
```

| Item | Columns |
|--------------------|----------------|
| Patient ID number | (1: 6) |
| Sex | (8: 8) |
| Age | (10:11) |
| Birth date | (13:20) |
| Incidence date | (22:29) |
| Topography | (31:34) |
| Histology | (36:39) |
| Behaviour | (40:40) |
| Basis of diagnosis | (42:42) |

Input file example (2)

The following file has both delimited and fixed field formats, and is suitable:

```
8501208,2,55,301092,1748,850039,7,2
8502004,2,75,201292,1450,807039,7,1
8600645,1,50,100491,1531,848039,7,1
8700936,2,48,110692,1737,807034,7,1
8701560,2,39,150791,1749,850039,7,1
8703231,2,64,010192,1450,807039,7,3
8800001,1,55,290588,1869,906139,7,2
8800002,1,45,280888,1419,807039,7,2
8800003,1,70,011188,1481,807039,7,2
8800004,1,17,250488,1479,807039,7,2
8800005,1,63,231288,1469,807039,7,2
8800006,1,58,100388,1450,807039,7,2
8800007,1,57,080288,1489,807039,7,2
8800008,1,55,210388,1619,807039,7,2
8800009,1,72,050388,1410,807033,7,2
8800064,1, 4,170888,1691,982139,5,1
8800065,1,31,120688,1999,999039,2,1
8800066,1,40,251088,1519,814039,7,1
8800067,1,12,190988,1691,982139,5,1
8800068,1, 7,151288,1691,982139,5,1
```

| Item | Columns |
|--------------------|---------|
| Patient ID number | (1: 7) |
| Sex | (9: 9) |
| Age | (11:12) |
| Incidence date | (14:19) |
| Topography | (21:24) |
| Histology | (26:29) |
| Behaviour | (30:30) |
| Grade | (31:31) |
| Basis of diagnosis | (33:33) |
| Vital status | (35:35) |

Please note that this file must be converted from ICD-O-1 to ICD-O-2, then from ICD-O-2 to ICD-O-3 before being handled by the check program.

Requested data set

Incidence data

All malignant cancers (/3), unspecified (/1) or in situ (/2) tumours of the bladder (C67), and, for the childhood data base, benign (/0) and unspecified (/1) tumours of the central nervous system (C70-72, C75) should be included.

Requested variables:

- Patient identification key (numeric or alphanumeric)
- Incidence date
- Birth date and/or age in years
- Sex
- Topography (**ICD-O-3**)*
- Histology (**ICD-O-3**)*
- Behaviour code
- Basis of diagnosis

*If you use another codification system, please convert your data file first using the appropriate conversion program.

Optional variables:

- Grade code (ICD-O)
- Sequence number
- Date when case first registered

Mandatory variables for the database of survival:

- Date of death
- Date of last known vital status
- Vital status (please provide the coding scheme)

Additional necessary variables for the ACCIS database:

- Laterality (please provide the coding scheme)

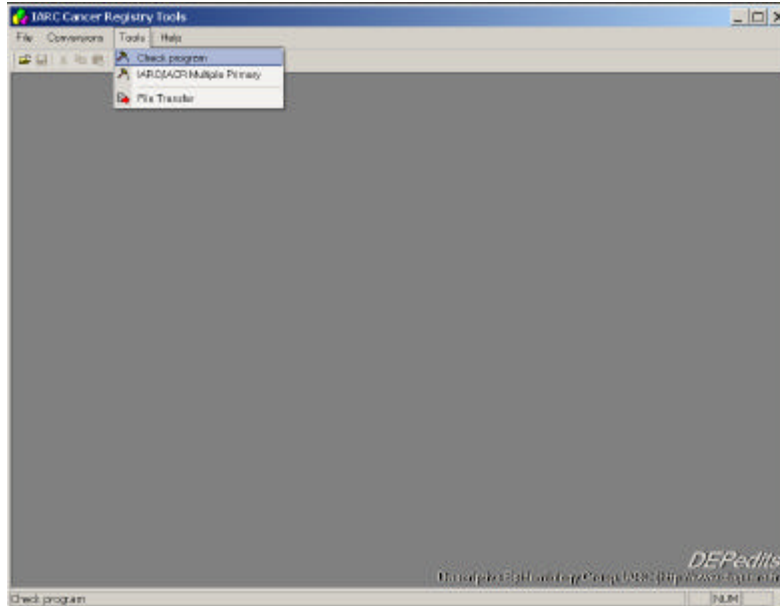
For survival and childhood databases, the date of diagnosis, date of birth and date of last known vital status should include the month, and if possible, the day.

The Descriptive Epidemiology group of IARC will be very grateful if you could pass your data through the CHECK program and correct all possible errors or unlikely combinations before submission. The complete dataset (incidence and population files) is accepted in any format (text file, EpiInfo, SAS, Microsoft Access or Excel etc.) and on any medium (CD-ROM, electronic mail).

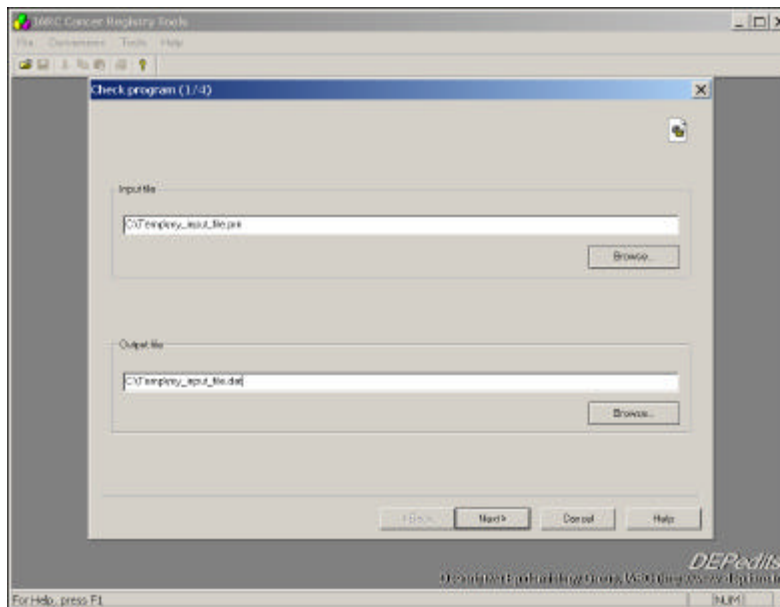
Processing

To check a data file using *DEPedits*, you perform the following steps using a the 'check wizard' process:

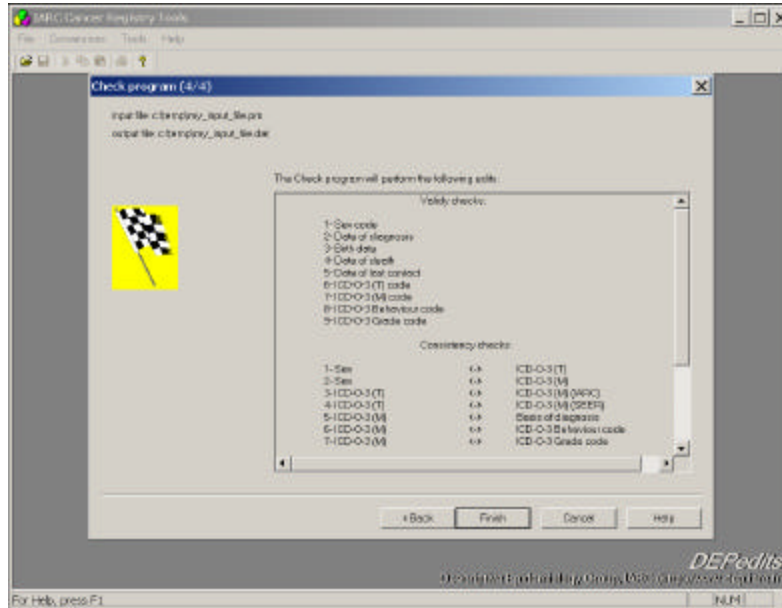
1 Select the check program:



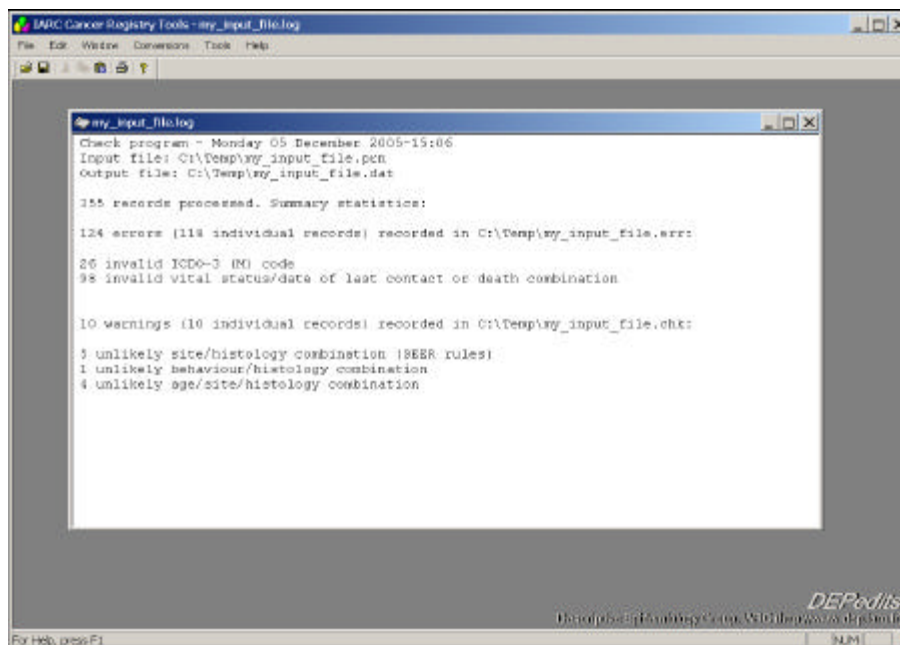
2 Select the input file you want to check, and the output file in which the new codes (conversion programs) or the result of the check process will be recorded.



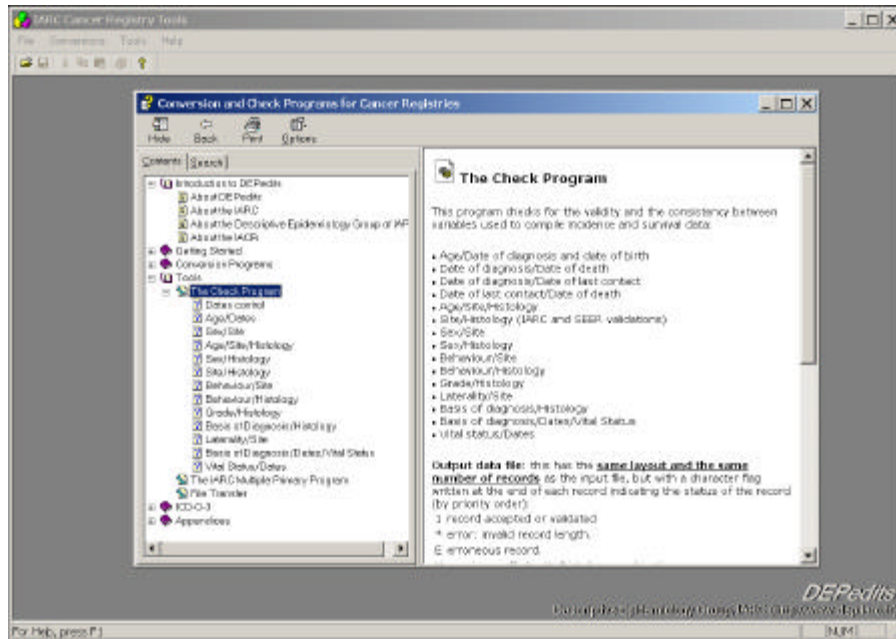
5 The final screen summarizes the checks that will be performed on your data. Click on the 'Finish' button to start processing.



At the end of the check process, a log file will be displayed which indicates the number and the type of errors or unlikely combinations found by the program. These are recorded in two different files for further analysis.



You can always have access to the on-line help by striking the 'F1' key at any time. In the Help/Topic options, you will find the detailed description of the edits performed by the CHECK program and an ICD-O-3 index.



If you need help regarding the use of the program, please contact J. Ferlay, IARC (ferlay@iarc.fr).