

IARCCrgTools Version2.05 使用说明

IARCCrgTools 是由国际癌症研究中心（IARC）描述流行病学工作组的 J. Ferlay 设计和编写的一款工具软件，用于 ICD-O 编码的转换以及肿瘤登记数据的逻辑核查。

第一部分：数据格式转换

在进行 ICD 编码转换及执行逻辑核查和多原发肿瘤检查的时候，IARCCrgTools 软件对数据格式有一定要求，如果格式不符合要求，软件无法读取文件内容。所以在进行逻辑核查及编码转换之前，要先对数据库文件的格式进行转换，IARCCrgTools 也提供了相应功能。

数据格式要求：以文本格式存放，固定列宽

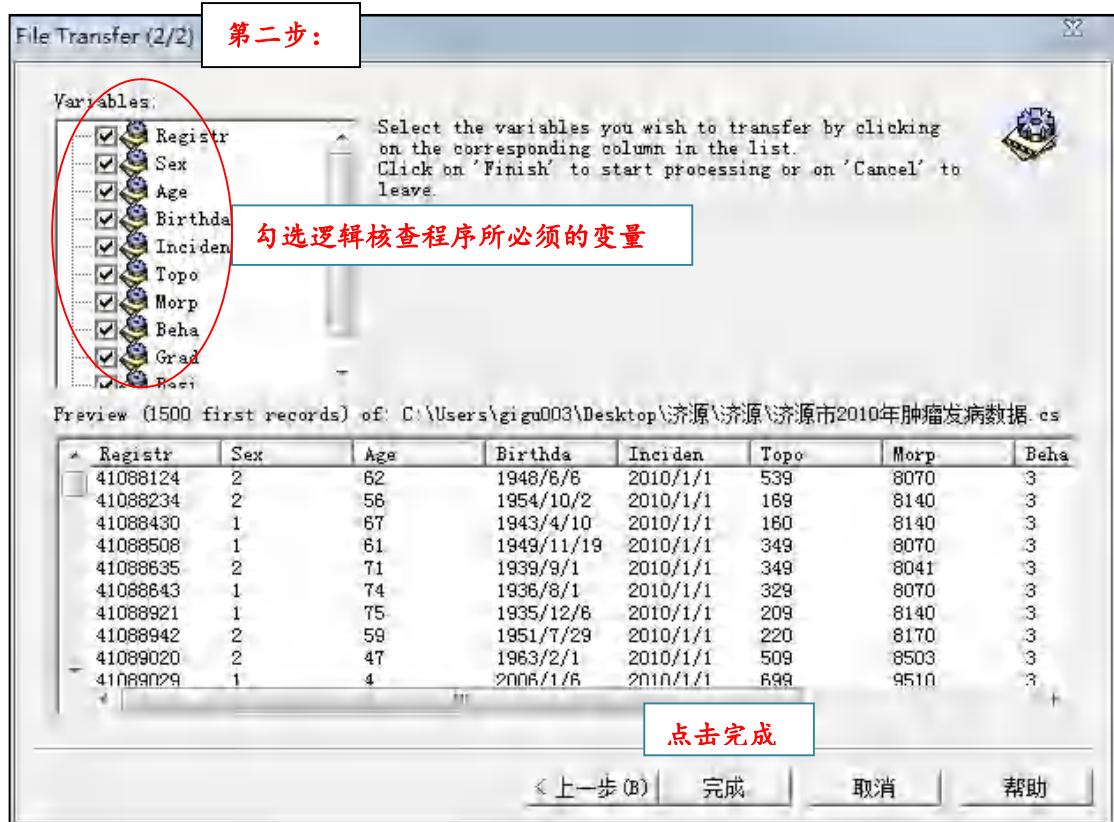
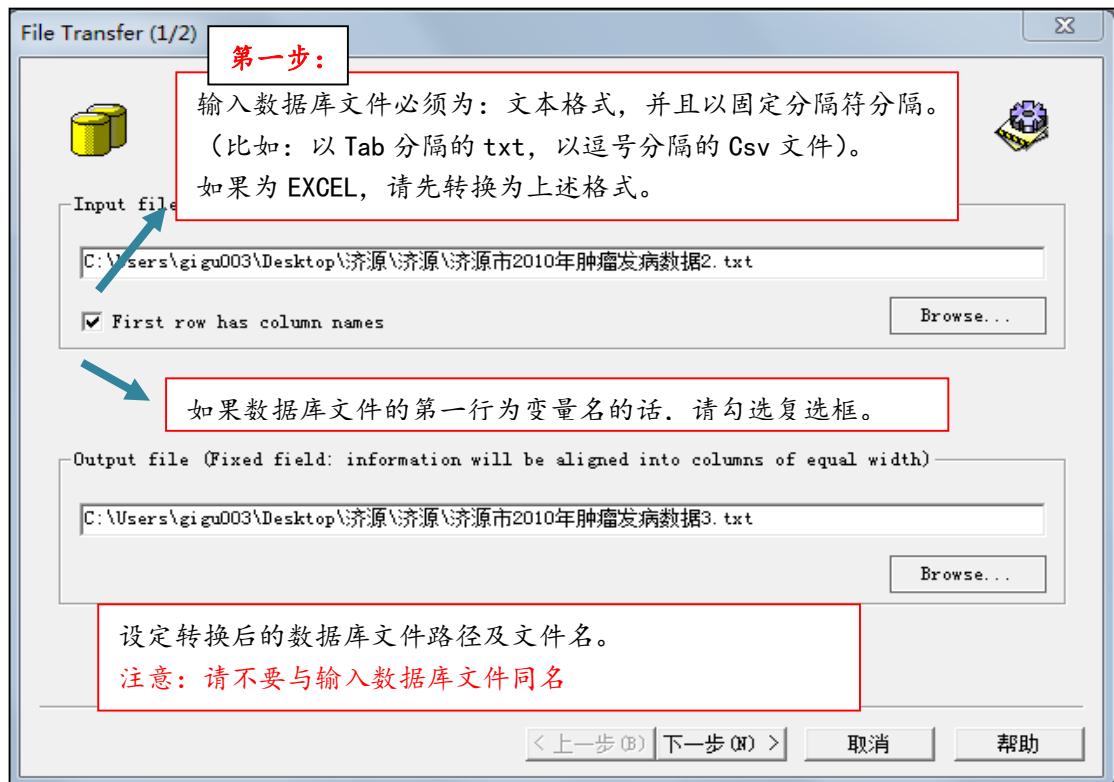
如果数据文件为 CanReg4 导出的文件，可以直接进行转换。

如果数据库文件为 EXCEL 格式，则要先转换为文本格式（如，以逗号分隔的 CSV 格式，以 Tab 分隔的文本格式）。具体方法如下：

打开 Excel 文件，点击文件—另存为，在弹出的窗口中，保存类型选择为 CSV（以逗号分隔）(*.csv)，输入文件名，保存即可。



IARCCrgTools 数据格式转换: (共两步)



第二部分：编码转换

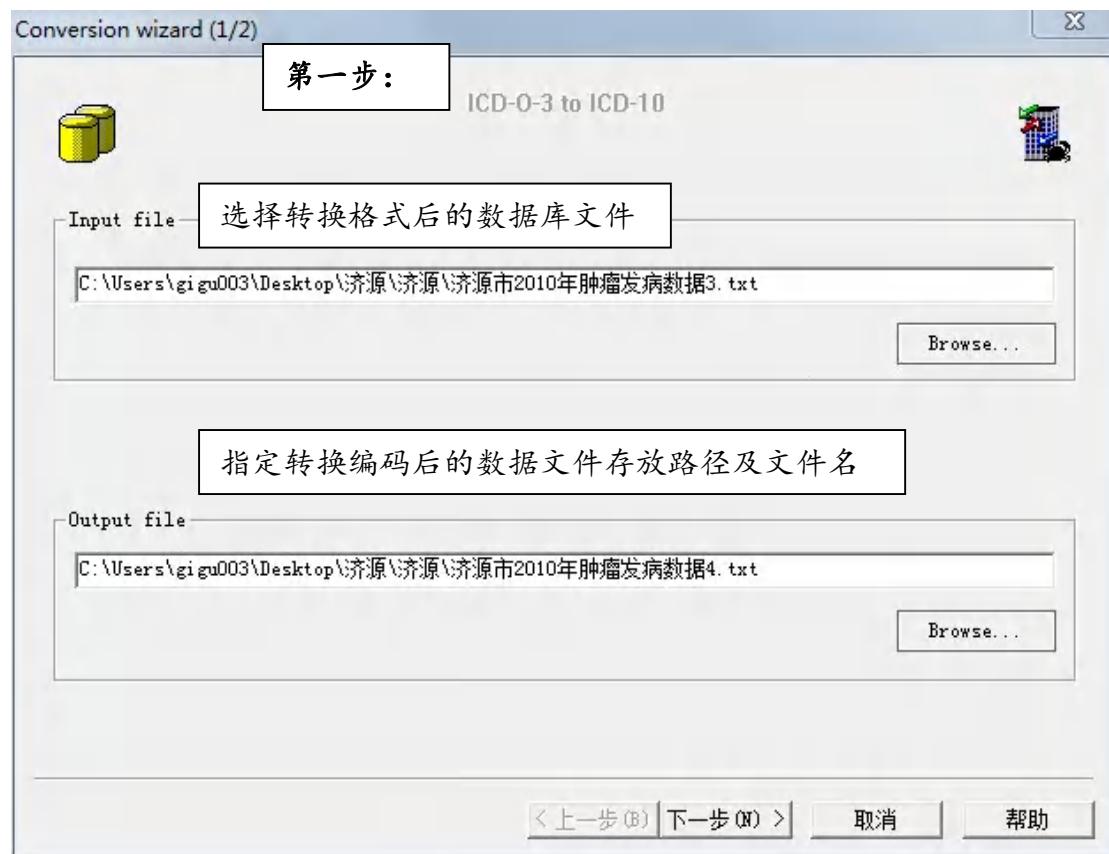
1、ICD-0-3 → ICD-10

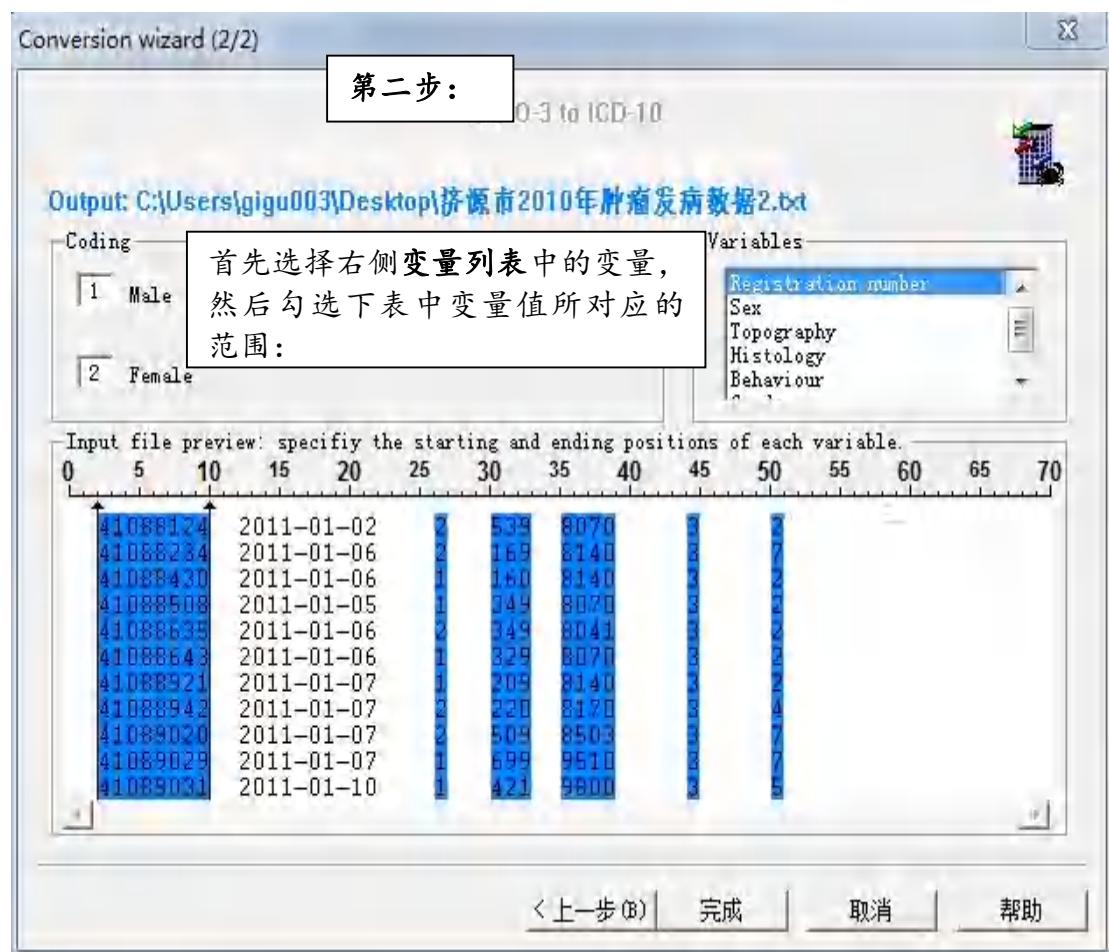
数据要求：

A. 每个病例一行，必须包含以下变量

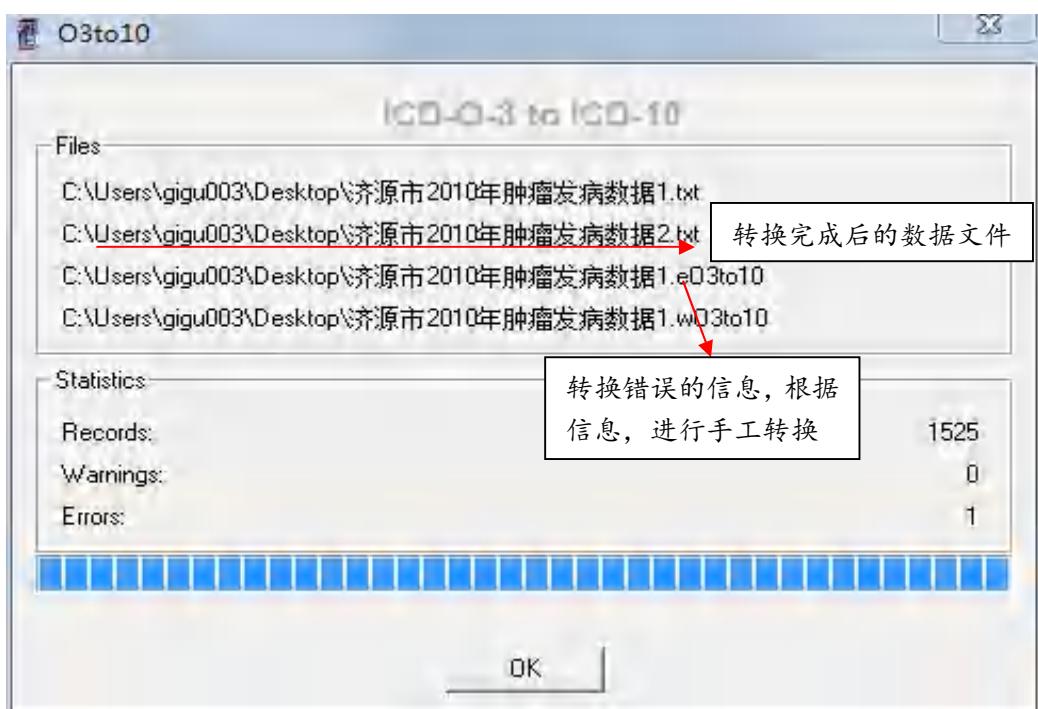
- (1) 登记号 (2) 性别 (3) 解剖学编码 (Topo)
- (4) 组织学编码 (5) 行为编码 (6) 分级

2、操作：打开 IARCCrgTools 软件，点击 conversions，在下拉菜单中选择 ICD-0-3 → ICD-10，出现以下界面，共需两步完成。





输出结果:



点击 File—Open 打开 .eo3to10 的错误信息文件：

ICD-0-3 to ICD-10 - Tuesday 15 November 2011-09:13						
C:\Users\gigu003\Desktop\济源市2010年肿瘤发病数据1.txt						
The following records cannot be converted to ICD-10						
41089246 2011-01-10 1 141 8070 3 7 Invalid ICD-0-3 code						
1 record(s)	一条记录由于无效的 ICD-0-3 编码，转换失败，需检查原始记录，进行修改后再转换。					

点击 File—Open 打开转换编码后的数据文件

41088124	2011-01-02	2	539	8070	3	2	C639
41088234	2011-01-06	2	169	8140	3	7	C169
41088430	2011-01-06	1	160	8140	3	2	C160
41088508	2011-01-05	1	349	8070	3	2	C349
41088635	2011-01-06	2	349	8041	3	2	C349
41088643	2011-01-06	1	329	8070	3	2	C329
41088921	2011-01-07	1	209	8140	3	2	C20
41088942	2011-01-07	2	220	8170	3	4	C220
41089020	2011-01-07	2	509	8503	3	7	C509
41089029	2011-01-07	1	699	9510	3	7	C692
41089031	2011-01-10	1	421	9800	3	5	C959
41089032	2011-01-10	2	421	9800	3	5	C959
41089035	2011-01-10	2	421	9800	3	5	C959
41089039	2011-01-10	1	421	9800	3	5	C959
41089044	2011-01-10	1	679	8120	3	7	C679
41089052	2011-01-10	1	679	8120	3	7	C679
41089063	2011-01-10	1	679	8120	3	7	C679
41089070	2011-01-10	1	160	8140	3	2	C160
41089072	2011-01-10	1	160	8140	3	2	C160
41089079	2011-01-10	1	160	8140	3	2	C160
41089088	2011-01-10	1	160	8140	3	2	C160
41089106	2011-01-10	1	160	8140	3	2	C160
41089109	2011-01-10	2	160	8140	3	2	C160
41089115	2011-01-10	1	160	8140	3	2	C160

输出结果的最后一列为
转换后的
ICD10 编码，
可以把它粘
贴到 Excel
中进行进一
步的编辑。

第三部分：逻辑核查

数据格式要求：

A. 数据库文件必须以文本格式存放，变量之间以固定格式分隔

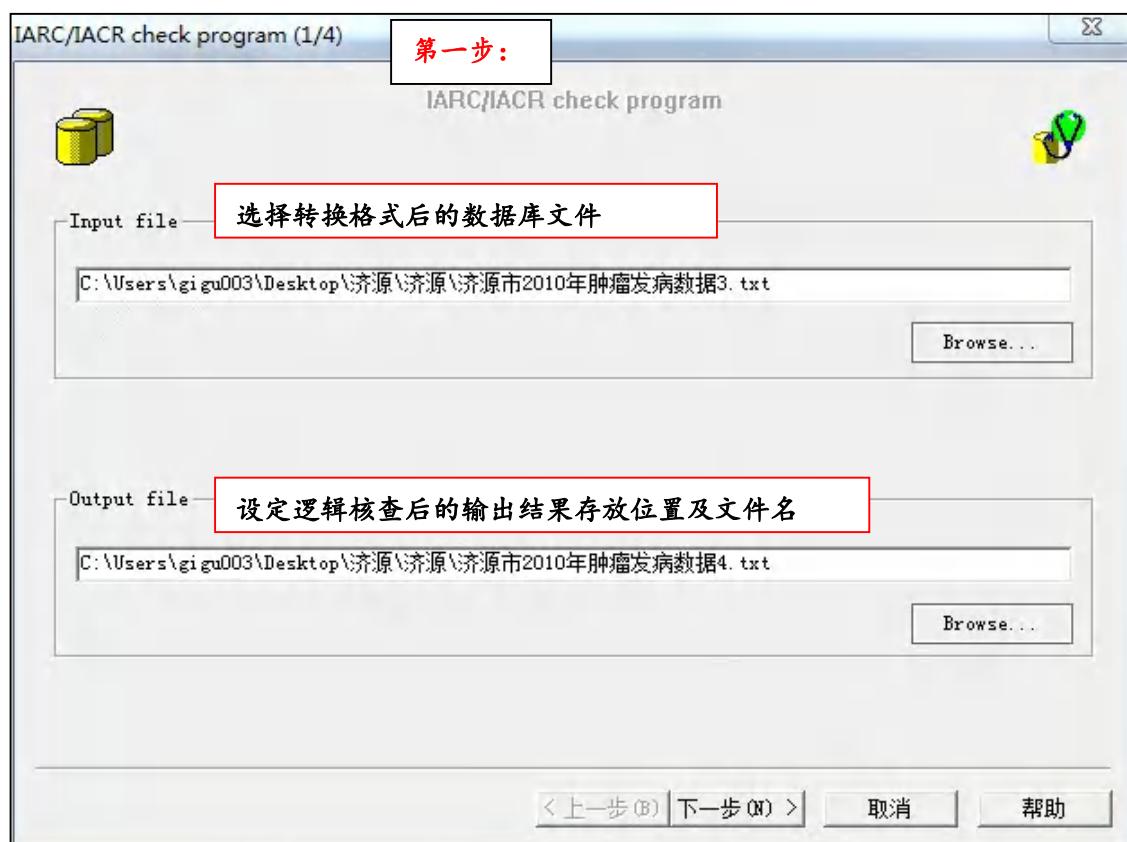
(按照第一部分介绍的数据转换方法对数据进行转换即可)

B. 每一病例用一行，要求必须包括以下变量：

- | | | | |
|----------------------|-----------------------|--------|----------|
| (1) 登记号 | (2) 性别 | (3) 年龄 | (4) 出生日期 |
| (5) 发病日期 | (6) 解剖学部位编码 (ICD-0-3) | | |
| (7) 组织学编码 (ICD-0-3) | (8) 肿瘤行为编码 (ICD-0-3) | | |
| (9) 肿瘤分级编码 (ICD-0-3) | (10) 诊断依据编码 | | |

逻辑核查流程：

Tools—IARC/IACR CHECK (共四步：)



IARC/IACR check program (2/4) 第二步:

Output: C:\Users\gigu003\Desktop\济源\济源\济源市2010年肿瘤发病数据4.txt

选择右侧变量列表中的变量（逻辑核查所必须），并勾选下表中变量值所对应的范围：
如：当前选择变量为：Topography (ICD0-3)
两个向上的箭头所包含的数据为 Topography

Variables

- Registration number
- Sex
- Topography (ICD0-3)**
- Histology (ICD0-3)
- Behaviour (ICD0-3)
- Grade (ICTN-3)

Input file preview: specify the starting and ending positions of the variables you wish to check

0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	
#	1088124		62	1948/6/6	2010-1-1	579	8070	3	2						
410888234			56	1954/10/2	2010-1-1	580	8140	3	3						
410888430			67	1943/4/10	2010-1-1	581	8140	3	2						
410888500			61	1949/11/19	2010-1-1	582	8070	3	2						
410888635			71	1939/9/1	2010-1-1	583	8041	3	2						
410888641			74	1936/8/1	2010-1-1	584	8070	3	2						
410888921			75	1935/12/6	2010-1-1	585	8140	3	3						
410888947			59	1951/7/29	2010-1-1	586	8170	3	3						
410889020			47	1963/2/1	2010-1-1	587	8503	3	3						
410889030			4	2006/1/6	2010-1-1	588	9510	3	2						
410889031			50	1960/11/25	2010-1-1	589	9800	3	3						
410889032			34	1976/11/15	2010-1-1	590	9800	3	3						

用鼠标左键点击变量值左界和右界，勾选变量范围

帮助

IARC/IACR check program (3/4) 第三步:

Output: C:\Users\gigu003\Desktop\技师\技师\技师市2010年肿瘤发病数据4.txt

Sex codes 填写数据中性别编码，默认男为1，女为2 Age unknown code 99

Basis of diagnosis codes for microscopic verification 5 6 7 7 数据中，未知年龄编码

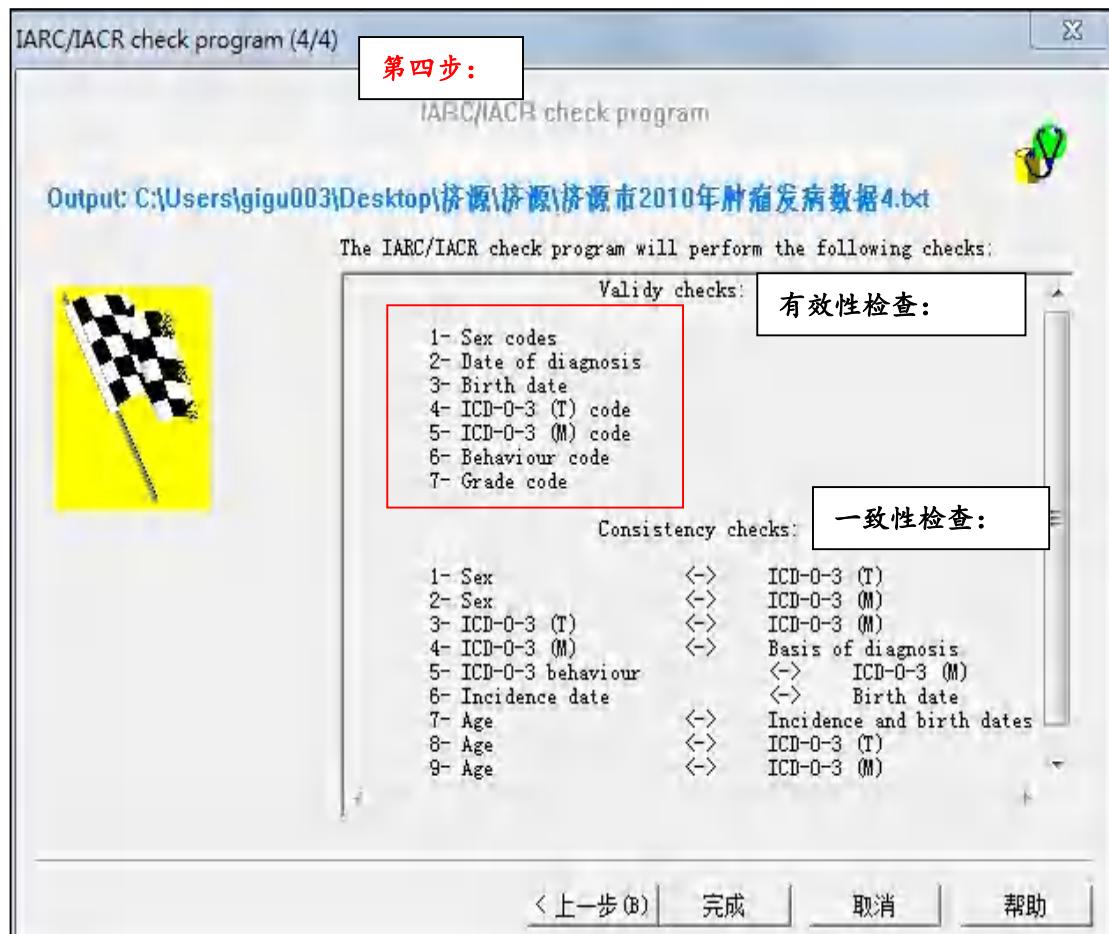
Incidence date format YYYYMMDD
选择数据中发病日期对应的格式 DDMMYY YY/MM/DD DD/MM/YY

Birth date format YYYY/MM/DD
选择数据中出生日期对应的格式 DD/MM/YYYY MM/DD/YYYY

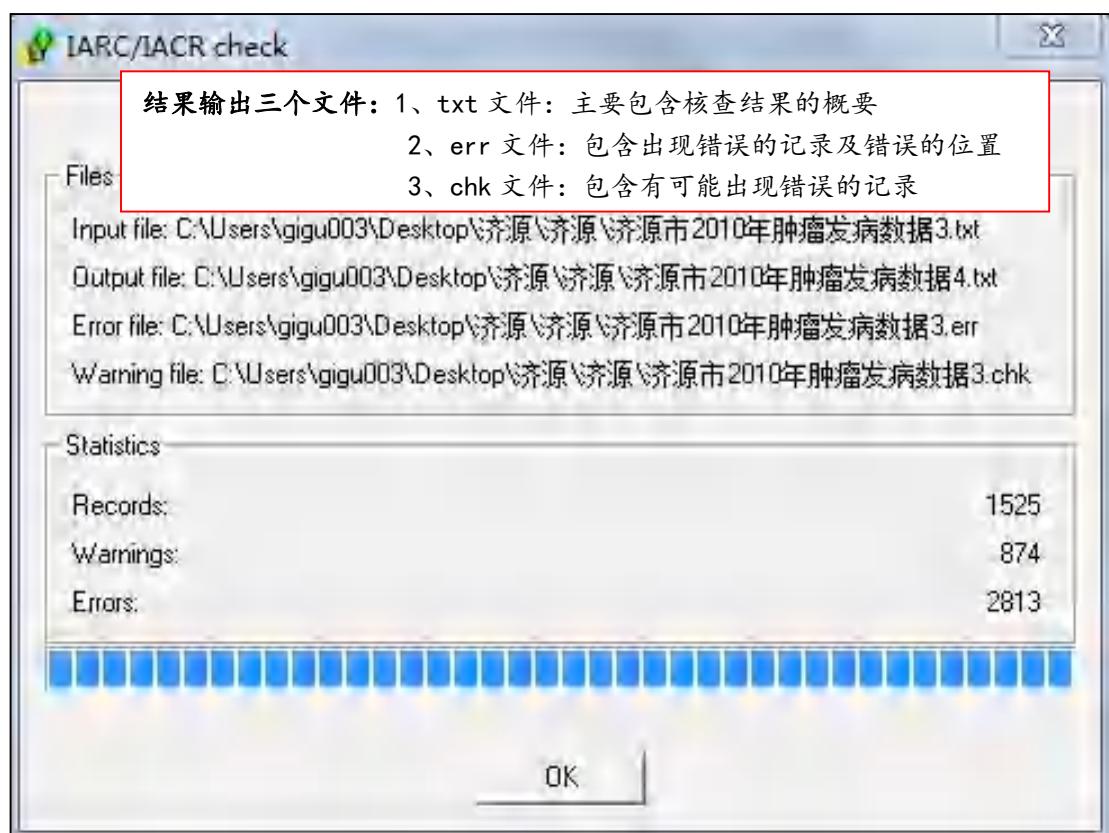
Error file Creates one record per type of error (ie the same case may appear several times)

如果勾选：每一种错误类型输出记录文件

<上一步(B) | 下一步(N) > 取消 帮助

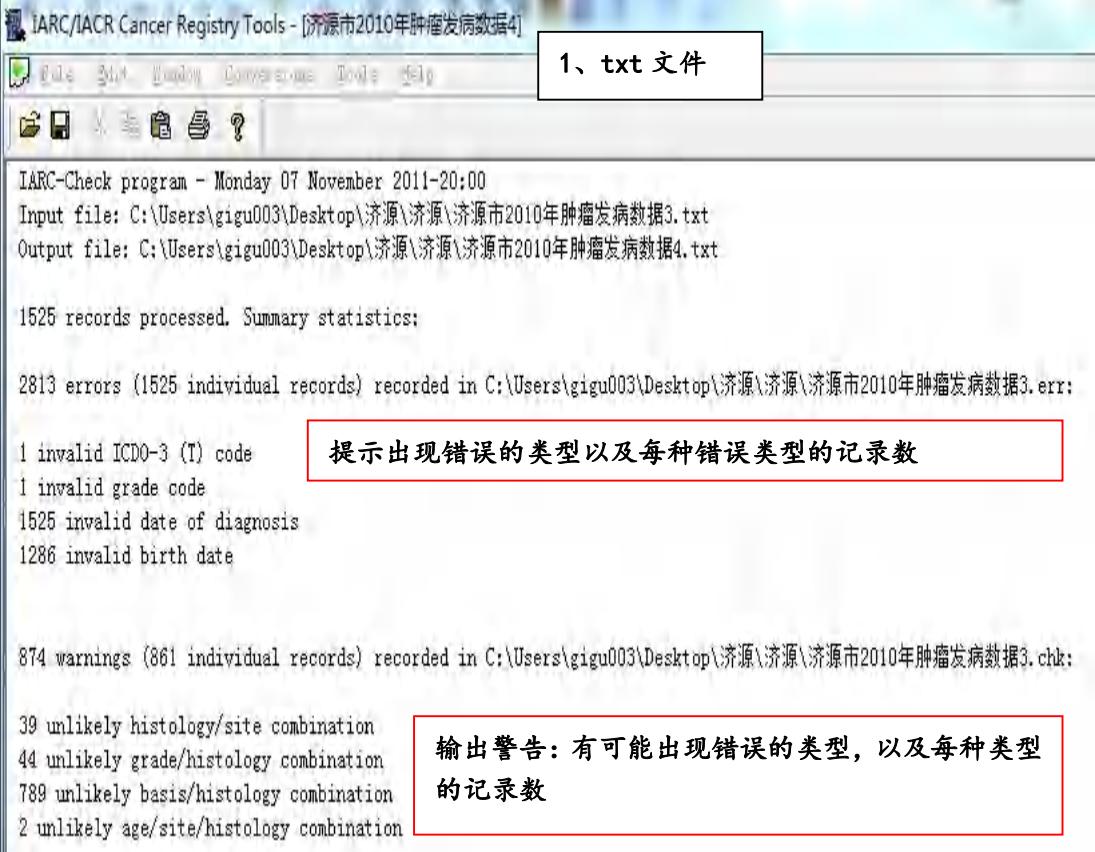


输出结果：



1、打开核查结果输出的 txt 文件：

File-Open 选择输出的 txt 文件



IARC-Check program - Monday 07 November 2011-20:00
Input file: C:\Users\gigu003\Desktop\济源\济源\济源市2010年肿瘤发病数据3.txt
Output file: C:\Users\gigu003\Desktop\济源\济源\济源市2010年肿瘤发病数据4.txt

1525 records processed. Summary statistics:

2813 errors (1525 individual records) recorded in C:\Users\gigu003\Desktop\济源\济源\济源市2010年肿瘤发病数据3.err:

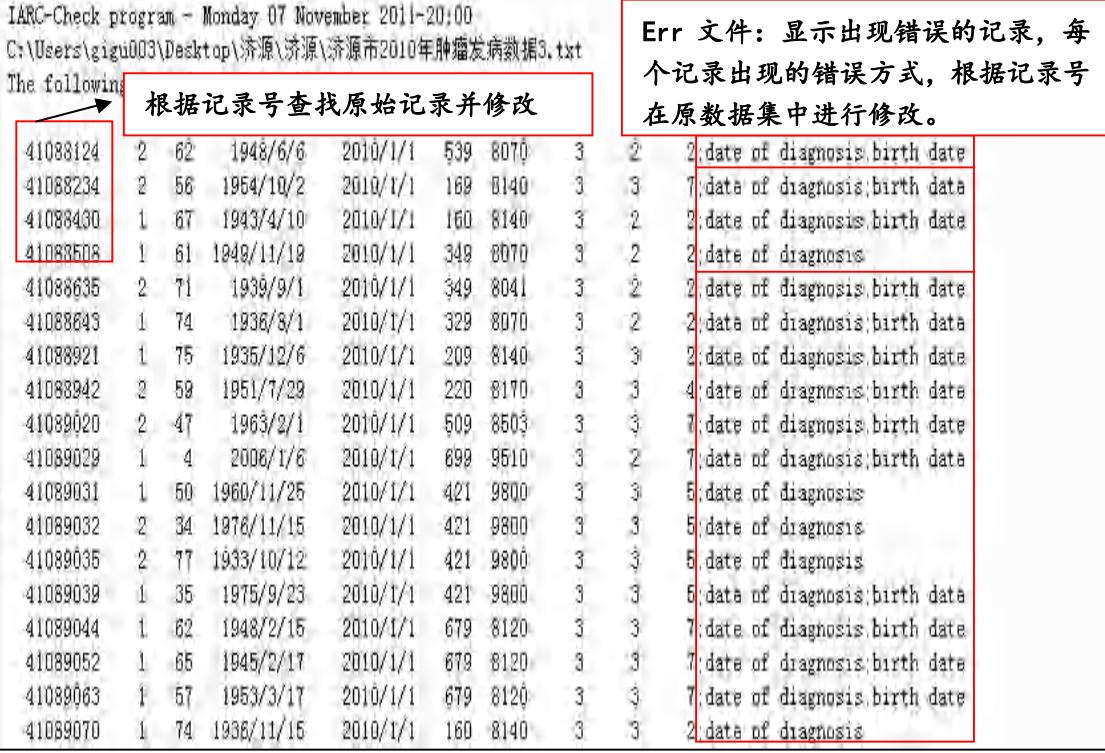
1 invalid ICDO-3 (T) code	提示出现错误的类型以及每种错误类型的记录数
1 invalid grade code	
1525 invalid date of diagnosis	
1286 invalid birth date	

874 warnings (861 individual records) recorded in C:\Users\gigu003\Desktop\济源\济源\济源市2010年肿瘤发病数据3.chk:

39 unlikely histology/site combination	输出警告：有可能出现错误的类型，以及每种类型的记录数
44 unlikely grade/histology combination	
789 unlikely basis/histology combination	
2 unlikely age/site/histology combination	

2、打开核查结果输出的 err 文件：

File-Open 选择输出的 err 文件



IARC-Check program - Monday 07 November 2011-20:00
C:\Users\gigu003\Desktop\济源\济源\济源市2010年肿瘤发病数据3.txt

The following

41088124	2	62	1948/6/6	2010/1/1	539	8070	3	2	2:date of diagnosis,birth date
41088234	2	56	1954/10/2	2010/1/1	169	8140	3	3	7:date of diagnosis,birth date
41088430	1	67	1943/4/10	2010/1/1	160	8140	3	2	2:date of diagnosis,birth date
41088508	1	61	1948/11/10	2010/1/1	349	8070	3	2	2:date of diagnosis
41088635	2	71	1939/9/1	2010/1/1	349	8041	3	2	2:date of diagnosis,birth date
41088643	1	74	1938/8/1	2010/1/1	329	8070	3	2	2:date of diagnosis,birth date
41088921	1	75	1935/12/6	2010/1/1	209	8140	3	3	2:date of diagnosis,birth date
41088942	2	59	1951/7/29	2010/1/1	220	8170	3	3	4:date of diagnosis,birth date
41089020	2	47	1963/2/1	2010/1/1	509	8503	3	3	7:date of diagnosis,birth date
41089029	1	4	2008/1/6	2010/1/1	699	9510	3	2	7:date of diagnosis,birth date
41089031	1	50	1960/11/25	2010/1/1	421	9800	3	3	5:date of diagnosis
41089032	2	34	1976/11/15	2010/1/1	421	9800	3	3	5:date of diagnosis
41089035	2	77	1933/10/12	2010/1/1	421	9800	3	3	5:date of diagnosis
41089039	1	35	1975/9/23	2010/1/1	421	9800	3	3	5:date of diagnosis,birth date
41089044	1	62	1948/2/15	2010/1/1	679	8120	3	3	7:date of diagnosis,birth date
41089052	1	65	1945/2/17	2010/1/1	679	8120	3	3	7:date of diagnosis,birth date
41089063	1	57	1953/3/17	2010/1/1	679	8120	3	3	7:date of diagnosis,birth date
41089070	1	74	1938/11/15	2010/1/1	160	8140	3	3	2:date of diagnosis

3、打开核查结果输出的 chk 文件：

File-Open 选择输出的 chk 文件

IARC-Check program - Monday 07 November 2011-20:00 C:\Users\gigu003\Desktop\济源\济源\济源市2010年肿瘤发病数据3.txt The following records should be checked											Chk 显示：有可能出现错误的记录，以及可能错误的类型
41088124	2	62	1948/6/6	2010/1/1	539	8070	3	2	2	basis/histology	
41088430	1	67	1943/4/10	2010/1/1				可能的错误类型	2	basis/histology	
41088508	1	6	根据记录号查找记录，并对记录进行核查		0/1/1	349	8070	3	2	2	basis/histology
41088635	2	7			0/1/1	349	8041	3	2	2	basis/histology
41088643	1	74	1936/8/1	2010/1/1	329	8070	3	2	2	basis/histology	
41088921	1	75	1935/12/6	2010/1/1	209	8140	3	3	2	basis/histology	
41089031	1	50	1960/11/25	2010/1/1	421	9800	3	3	5	grade/morphology	
41089032	2	34	1976/11/15	2010/1/1	421	9800	3	3	5	grade/morphology	
41089035	2	77	1933/10/12	2010/1/1	421	9800	3	3	5	grade/morphology	