

SEERaBomb Overview

Tom Radivoyevitch

May 31, 2012

Introduction

SEERaBomb is for SEER and Japanese A-bomb survivor data analysts. It contributes speed to SEER analyses by reducing file sizes to contain only items of interest. This document assumes that the data has been downloaded into folders off of the root called `/data/SEER` and `/data/abomb`. To obtain the data please see the links in `dataLinks.docx`. Use cases are given in the `blood2012` functions available from help and in the R scripts in the `papers` directory.

SEER Data R Binaries

The incidence directory of the SEER dataset contains a SAS file that defines the field names, their starting positions, and their fixed widths. This file can be used to read the SEER data into SAS, but it is used here to: 1) present the field choices (see `fieldNames.html` and the output of `getFields()`); and 2) given user choices, automatically determine the sequence of widths needed to extract the data of interest using the speedy R package `LaF`. `getFields()` has one parameter, `seerHome="/data/SEER"`, which should be over-ridden if the SEER data lives elsewhere. Its data.frame output and the SEER file `seerDic.pdf` in the SEER incidence directory must be thoroughly examined to determine which fields will be useful. Once this is determined, the output and list of field choices, the default of which is

```
picks=c("casenum","reg","race","sex","agedx","yrbrth",  
        "seqnum","yrdx","histo2","histo3","radiatn","agerec",  
        "ICD9","histrec","css03","numprims","survtm","COD"),
```

must then be inputted into `pickFields()`.

The output of `pickFields()` contains not only pulled rows from the input, but also inserted rows with widths computed to fill the gaps of no interest. Knowing these gap sizes enables fast file reading by `LaF` in `mkSEER()`. This function produces R Data binaries that can be found in SEER dataset subdirectories of `seerHome`, e.g. in directories such as `/data/SEER/yr2000_2009`.

This work was supported by the National Cancer Institute and Tufts Integrative Cancer Biology Program under U54CA149233-029689.

```

library(SEERaBomb)

## Loading required package: RSQLite
## Loading required package: DBI
## Loading required package: LaF
## Loading required package: Rcpp

(df = getFields())

```

	start	width	names	desc
## 1	1	8	casenum	Patient ID
## 2	9	10	reg	SEER registry
## 3	19	1	marstat	Marital status at diagnosis
## 4	20	2	race	Race/ethnicity
## 5	22	1	origin	Spanish surname or origin
## 6	23	1	nhia	NHIA Derived Hisp Origin
## 7	24	1	sex	Sex
## 8	25	3	agedx	Age at diagnosis
## 9	28	4	yrbirth	Year of birth
## 10	32	3	plcbirth	Place of birth
## 11	35	2	seqnum	Sequence number
## 12	37	2	modx	Month of diagnosis
## 13	39	4	yrdx	Year of diagnosis
## 14	43	4	siteo2	Primary site ICD-0-2 (1973+)
## 15	47	1	lateral	Laterality
## 16	48	4	histo2	Histologic Type ICD-0-2
## 17	52	1	beho2	Behavior Code ICD-0-2
## 18	53	4	histo3	Histologic Type ICD-0-3
## 19	57	1	beho3	Behavior code ICD-0-3
## 20	58	1	grade	Grade
## 21	59	1	dxconf	Diagnostic confirmation
## 22	60	1	reptsrc	Type of reporting source
## 23	61	3	eod10sz	EOD 10 - size (1988+)
## 24	64	2	eod10ex	EOD 10 - extension
## 25	66	2	eod10pe	EOD 10 - path extension
## 26	68	1	eod10nd	EOD 10 - lymph node
## 27	69	2	eod10pn	EOD 10 - positive lymph nodes examined
## 28	71	2	eod10ne	EOD 10 - number of lymph nodes examined
## 29	73	13	eod13	EOD--old 13 digit
## 30	86	2	eod2	EOD--old 2 digit
## 31	88	4	eod4	EOD--old 4 digit
## 32	92	1	eodcode	Coding system for EOD
## 33	93	1	tumor1	Tumor marker 1
## 34	94	1	tumor2	Tumor marker 2
## 35	95	1	tumor3	Tumor marker 3
## 36	96	3	cssize	Collaborative Stage (CS) Tumor Size

## 37	99	3	csext	CS Extension
## 38	102	3	csnode	CS Lymph Nodes
## 39	105	2	csmets	CS Mets at DX
## 40	107	3	csssf1	CS Site-Specific Factor 1
## 41	110	3	csssf2	CS Site-Specific Factor 2
## 42	113	3	csssf3	CS Site-Specific Factor 3
## 43	116	3	csssf4	CS Site-Specific Factor 4
## 44	119	3	csssf5	CS Site-Specific Factor 5
## 45	122	3	csssf6	CS Site-Specific Factor 6
## 46	125	3	csssf25	CS Site-Specific Factor 25
## 47	128	2	dajcct	Derived AJCC T
## 48	130	2	dajccn	Derived AJCC N
## 49	132	2	dajccm	Derived AJCC M
## 50	134	2	dajccs	Derived AJCC Stage Group
## 51	136	1	dssg77	Derived SS1977
## 52	137	1	dssg00	Derived SS2000
## 53	138	1	dajccf	Derived AJCC - flag
## 54	139	1	dssg77f	Derived SS1977 - flag
## 55	140	1	dssg00f	Derived SS2000 - flag
## 56	141	6	csvorg	CS Version Input Original
## 57	147	6	csvder	CS Version Derived
## 58	153	6	csvcur	CS Version Input Current
## 59	159	2	surgprim	RX Summ--surg prim site
## 60	161	1	scope	RX Summ--scope reg LN sur 2003+
## 61	162	1	surgoth	RX Summ--surg oth reg/dis
## 62	163	2	surgnode	Number of lymph nodes
## 63	165	1	reconst	Reconstruction
## 64	166	1	nosurg	Reason no cancer-directed surgery
## 65	167	1	radiatn	Radiation
## 66	168	1	radbrn	Radiation to Brain and/or CNS
## 67	169	1	radsurg	Radiation sequence with surgery
## 68	170	2	sssurg	Site specific surgery (1983-1997)
## 69	172	2	srprim02	Surgery to Primary Site
## 70	174	1	scope02	Scope of lymph node surgery 98-02
## 71	175	1	srgoth02	Surgery to other sites
## 72	176	2	recno	Record number
## 73	178	1	ositage	Age-site edit override
## 74	179	1	oseqcon	Sequence number-dx conf override
## 75	180	1	oseqlat	Site-type-lat-seq override
## 76	181	1	osurcon	Surgery-diagnostic conf override
## 77	182	1	osittp	Site-type edit override
## 78	183	1	hbenign	Histology edit override
## 79	184	1	orptsrc	Report source sequence override
## 80	185	1	odfsite	Seq-ill-defined site override
## 81	186	1	oleukdx	Leuk-Lymph dx confirmation override

## 82	187	1	ositbeh	Site-behavior override
## 83	188	1	oeoddt	Site-EOD-dx date override
## 84	189	1	ositeod	Site-laterality-EOD override
## 85	190	1	ositmor	Site-laterality-morph override
## 86	191	1	typefup	Type of followup expected
## 87	192	2	agerec	Age recode
## 88	194	5	siterec	Site recode
## 89	199	5	site2re	Site rec with Kaposis and mesothelioma
## 90	204	4	ICD9	Recode ICD-0-2 to 9
## 91	208	4	ICD10	Recode ICD-0-2 to 10
## 92	212	3	icccsite	ICCC site recode
## 93	215	3	icccsm	SEER modified ICC site recode
## 94	218	3	iccc3	ICCC ICD-0-3 Recode
## 95	221	3	iccc3ext	ICCC ICD-0-3 Extended Recode
## 96	224	1	behanal	Behavior recode for analysis
## 97	226	2	histrec	Broad Histology recode
## 98	228	2	brainrec	Brain recode
## 99	230	3	css03	Collaborative Stage (CS) Schema v0203
## 100	233	1	racreca	Race recode A
## 101	234	1	racrecy	Race recode Y
## 102	235	1	nhiares	Origin Recode NHIA
## 103	236	1	hststga	SEER historic stage A
## 104	237	2	ajccstg	AJCC stage 3rd edition (1988+)
## 105	239	2	aj3seer	SEER modified AJCC stage 3rd ed (1988+)
## 106	241	1	ssg77	SEER Summary Stage 1977 (1995-2000)
## 107	242	1	ssg2000	SEER Summary Stage 2000 2000 (2001-2003)
## 108	243	2	numprims	Number of primaries
## 109	245	1	firstprm	First malignant primary indicator
## 110	246	5	stcounty	State-county recode
## 111	251	4	survtm	Survival time recode
## 112	255	5	COD	Cause of death to SEER site recode
## 113	260	5	codkm	COD to site rec KM
## 114	265	1	statrec	Vital status recode (study cutoff used)
## 115	266	1	ihs	IHS link
## 116	267	1	histssg2000	Historic SSG 2000 Stage
## 117	268	2	ayarecode	AYA recode
## 118	270	2	lymphomarecode	Lymphoma recode
## 119	272	1	dthclass	SEER cause of death classification
## 120	273	1	odthclass	SEER other cause of death classification
## 121	274	1	exteval	CS EXT/Size Eval
## 122	275	1	nodeeval	CS Nodes Eval
## 123	276	1	metseval	CS Mets Eval
## 124	277	1	intprim	Primary by International Rules
## 125	278	1	erstatus	ER Status Recode Breast Cancer (1990+)
## 126	279	1	prstatus	PR Status Recode Breast Cancer (1990+)

```
## 127 280 2 csschema CS Schema - AJCC 6th Edition
## 128 282 3 csssf8 Cs Site-specific Factor 8
## 129 285 3 csssf10 CS Site-Specific Factor 10
## 130 288 3 csssf11 CS Site-Specific Factor 11
## 131 291 3 csssf13 CS Site-Specific Factor 13
## 132 294 3 csssf15 CS Site-Specific Factor 15
## 133 297 3 csssf16 CS Site-Specific Factor 16
## 134 300 1 vasin Lymph-vascular Invasion (2004+)
```

```
(df = pickFields(df))
```

```
##          start width  names          desc
## casenum      1    8 casenum      Patient ID
## reg          9   10 reg          SEER registry
## 3           19    1          race
## race         20    2 race          Race/ethnicity
## 5           22    2          sex
## sex          24    1 sex          Sex
## agedx        25    3 agedx        Age at diagnosis
## yrbrth       28    4 yrbrth       Year of birth
## 9           32    3
## seqnum       35    2 seqnum       Sequence number
## 11          37    2
## yrdx         39    4 yrdx         Year of diagnosis
## 13          43    5
## histo2       48    4 histo2      Histologic Type ICD-0-2
## 15          52    1
## histo3       53    4 histo3      Histologic Type ICD-0-3
## 17          57   110
## radiatn     167    1 radiatn      Radiation
## 19          168   24
## agerec      192    2 agerec      Age recode
## 21          194   10
## ICD9        204    4 ICD9        Recode ICD-0-2 to 9
## 23          208   18
## histrec     226    2 histrec     Broad Histology recode
## 25          228    2
## css03       230    3 css03 Collaborative Stage (CS) Schema v0203
## 27          233   10
## numprims    243    2 numprims    Number of primaries
## 29          245    6
## srvy        251    2 srvy        Survival time recode
## srvm        253    2 srvm        Survival time recode
## 32          255   46
##          type
## casenum integer
```

```
## reg      integer
## 3         string
## race      integer
## 5         string
## sex       integer
## agedx     integer
## yrbrth    integer
## 9         string
## seqnum    integer
## 11        string
## yrdx      integer
## 13        string
## histo2    integer
## 15        string
## histo3    integer
## 17        string
## radiatn   integer
## 19        string
## agerec    integer
## 21        string
## ICD9      integer
## 23        string
## histrec   integer
## 25        string
## css03     integer
## 27        string
## numprims  integer
## 29        string
## srvy      integer
## srvn      integer
## 32        string

# mkSEER(df,release='yr1992_2009')
```