Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Konrad-Zuse-Strasse 1 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
http://www.demogr.mpg.de

MPIDR TECHNICAL REPORT 2016-001
APRIL 2016

R Programs for Writing HMD Life Tables and HMD Death Rates to Pooled Data Files

Vladimir M. Shkolnikov (shkolnikov@demogr.mpg.de)
Dmitri A. Jdanov (jdanov@demogr.mpg.de)

For additional material see www.demogr.mpg.de/tr/

R Programs for Writing HMD Life Tables or HMD Death Rates to Pooled Data File

Vladimir M. Shkolnikov (shkolnikov@demogr.mpg.de)
Dmitri A. Jdanov (jdanov@demogr.mpg.de)

Abstract. The Human Mortality Database (HMD) website provides an opportunity for downloading output data as zipped files. Unzipping these files returns folders that contain numerous data files, which correspond to age-by-year data formats (1x1, 5x1, 1x5, 5x5, …), sexes, and dozens of country-populations. For many analyses, however, it would be more practical to operate with long data file containing many populations rather than large number of population-specific files. Herewith, we provide two simple programs for: 1) pooling together life tables' for many countries and years in a single file and 2) death rates for many countries and years in a single file.

Motivation

The world leading resource on mortality data the Human Mortality Database (HMD) provides detailed data series for more than 40 population and subpopulation. Despite on high quality of output data and easy-to-use interface the range of output statistics does not include files aggregated over countries. But the website provides an opportunity for downloading output data as zipped files by indicators. Unzipping these files returns folders that contain numerous data files, which correspond to age-by-year data formats (1x1, 5x1, 1x5, 5x5, …), sexes, and dozens of country-populations. In this technical report we present two R scripts for: 1) pooling together life tables for many countries and years in a single file and 2) pooling together death rates for many countries and years in a single file. Using these scripts is extremely simple and practically does not require any knowledge of R. Scripts were tested under R version 3.1.

The aggregation of HMD files requires two steps: 1) downloading data from the HMD and 2) running the respective R script. Below we describe both steps.

Preparation of input data: downloading zipped files from the HMD website

In addition to possibility of downloading single data files for chosen mortality indicator, sex, age-by-year format and country, the HMD provides possibility to download zip archives containing many data files. The page with zip files by all HMD output indicators is available at http://www.mortality.org/cgi-bin/hmd/hmd_download.php . There is also link to this page ("Zipped Data Files") in the "HMD Main Menu" on left-hand-side of the HMD face webpage. Clicking on this item brings user to the list of downloadable zipped data files (see Figure 1). Among these files, we will be interested in period and cohort Life Tables and in period and cohort Death Rates.

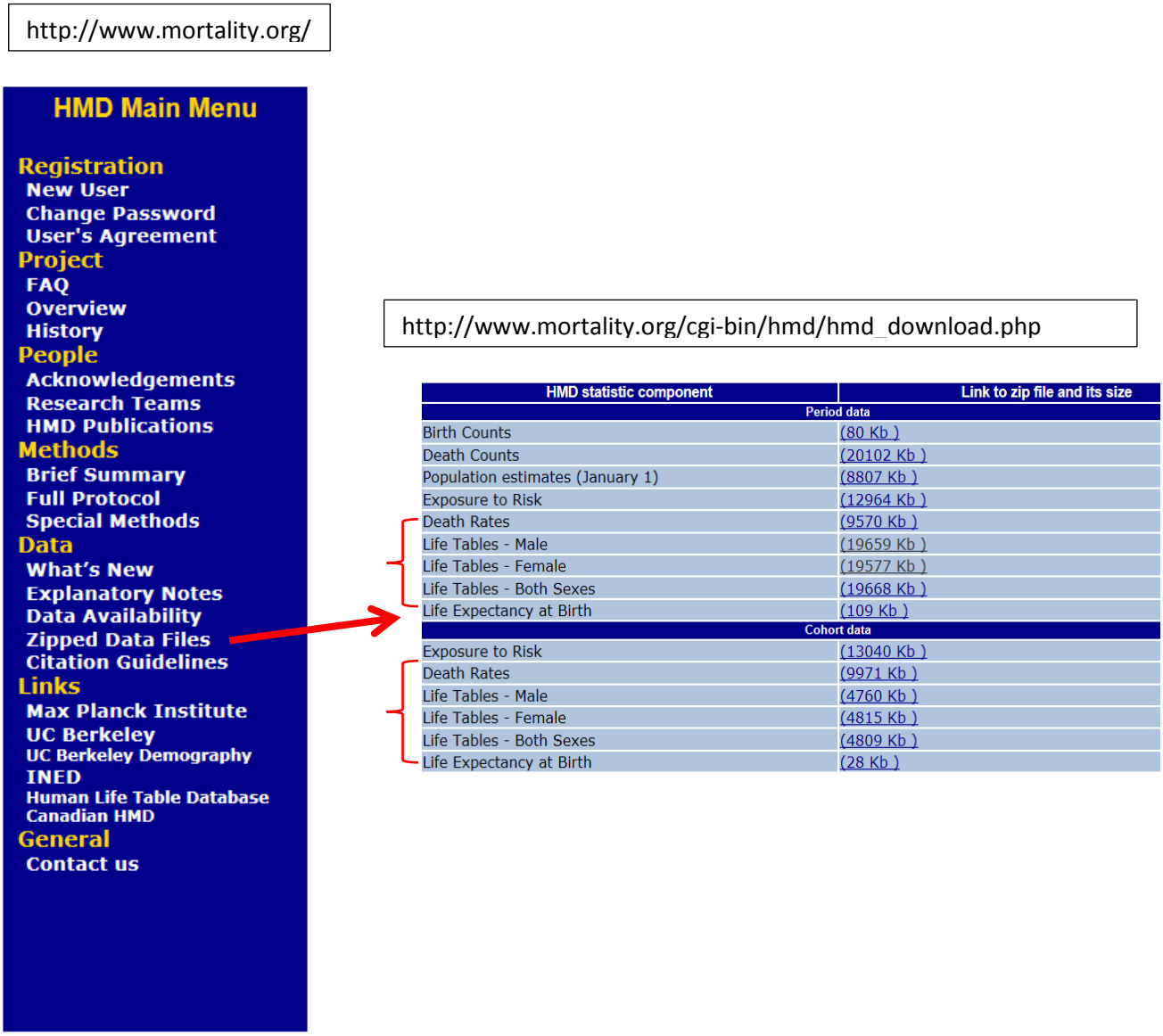

Figure 1. Getting to downloadable zipped files with period and cohort life tables and with period and cohort death rates from the HMD face page

Right clicking on links in the right column of the table (Figure 1) enables to download corresponding zipped files of death rates or life tables. To finish the process of data preparation it is enough just unzip the file in a new empty folder.

R scripts

Installation

The technical report includes the zip file with two R scripts. **Input and output data corresponding to the two examples given below can be provided upon request.** There is no need to install any packages before running these scripts. It is enough just to unzip the zip file and source both R files (Pool_HMD_LT.r and Pool_HMD_Mx.r) either using menu or by typing two following command in R console:

source("/path-to-the-script/Pool_HMD_LT.r")

source("/path-to-the-script/Pool_HMD_Mx.r")

where */path-to-the-script/* is the full path to the R scripts (in case they are not in the R working directory).

Usage

To aggregate life table files:

Pool.HMD.LT.files (DataDir = '.', OutputDir = './pooled', OutputFile = "agg.csv", Separator = ",")

To aggregate death rates

Pool.HMD.LT.files (DataDir = '.', OutputDir = './pooled', OutputFile = "agg.csv", Separator = ",")

Arguments:

DataDir – Folder with HMD files to be aggregated. Please check that it does not contain any other (non-HMD) files

OutputDir – Folder, where the output file will be saved

OutputFile – File name of the output file with aggregated data

Separator – Separator of data fields in the output file

Examples

Herewith, we illustrate use of the R scripts *Pool.HMD.Mx.files* and *Pool.HMD.LT.files* by two examples.

Example 1. Pooling country-specific death rates

Right-click on the link corresponding to "Death Rates" in the upper (period) part of the table on the webpage http://www.mortality.org/cgi-bin/hmd/hmd_download.php and download the file *death_rates.zip*. Unzipping of this file produces five folders *Mx_1x1*, *Mx_1x5*, *Mx_1x10*, *Mx_5x1*, *Mx_5x5*, *Mx_5x10*. Each of these folders contains country-files of death rates in respective age-by-year formats. For example, the folder *Mx_1x1* contains 47 country-files[1] of single-age and single-calendar-year death rates named:

AUS.Mx_1x1.txt, AUT.Mx_1x1.txt, BEL.Mx_1x1.txt, …, USA.Mx_1x1.txt.

To build the pooled file, which will contain single-age- and single-year-specific death rates for all country-populations, it is necessary to copy all the files (*AUS.Mx_1x1.txt, AUT.Mx_1x1.txt, BEL.Mx_1x1.txt, …, USA.Mx_1x1.txt*) into certain "input" folder (say) *c:\InputData*1\ , open the R-Studio or simple R-console and after sourcing of the file *Pool.HMD.Mx.files*.R (see section Instalation for details) call the function

[1] The actual number of data series as of April 2016. Might be more in future.

Pool.HMD.Mx.files("*c:/InputData*1", "*c:/OutputData*1", "*Death-rates*1x1.*csv*", *Separator*=",").
During execution of the code, the following lines will appear in the Console window:

```
  file # 1     AUS.Mx_1x1.txt
  file # 2     AUT.Mx_1x1.txt
  file # 3     BEL.Mx_1x1.txt
  file # 4     BGR.Mx_1x1.txt
  file # 5     BLR.Mx_1x1.txt
  file # 6     CAN.Mx_1x1.txt
  file # 7     CHE.Mx_1x1.txt
  file # 8     CHL.Mx_1x1.txt
  file # 9     CZE.Mx_1x1.txt
  file # 10     DEUTE.Mx_1x1.txt
  file # 11     DEUTNP.Mx_1x1.txt
  file # 12     DEUTW.Mx_1x1.txt
  file # 13     DNK.Mx_1x1.txt
  file # 14     ESP.Mx_1x1.txt
  file # 15     EST.Mx_1x1.txt
  file # 16     FIN.Mx_1x1.txt
  file # 17     FRACNP.Mx_1x1.txt
  file # 18     FRATNP.Mx_1x1.txt
  file # 19     GBR_NIR.Mx_1x1.txt
  file # 20     GBR_NP.Mx_1x1.txt
  file # 21     GBR_SCO.Mx_1x1.txt
  file # 22     GBRCENW.Mx_1x1.txt
  file # 23     GBRTENW.Mx_1x1.txt
  file # 24     GRC.Mx_1x1.txt
  file # 25     HUN.Mx_1x1.txt
  file # 26     IRL.Mx_1x1.txt
  file # 27     ISL.Mx_1x1.txt
  file # 28     ISR.Mx_1x1.txt
  file # 29     ITA.Mx_1x1.txt
  file # 30     JPN.Mx_1x1.txt
  file # 31     LTU.Mx_1x1.txt
  file # 32     LUX.Mx_1x1.txt
  file # 33     LVA.Mx_1x1.txt
  file # 34     NLD.Mx_1x1.txt
  file # 35     NOR.Mx_1x1.txt
  file # 36     NZL_MA.Mx_1x1.txt
  file # 37     NZL_NM.Mx_1x1.txt
  file # 38     NZL_NP.Mx_1x1.txt
  file # 39     POL.Mx_1x1.txt
  file # 40     PRT.Mx_1x1.txt
  file # 41     RUS.Mx_1x1.txt
  file # 42     SVK.Mx_1x1.txt
  file # 43     SVN.Mx_1x1.txt
  file # 44     SWE.Mx_1x1.txt
  file # 45     TWN.Mx_1x1.txt
  file # 46     UKR.Mx_1x1.txt
  file # 47     USA.Mx_1x1.txt
=====================================
```

The output file *Death-rates*1x1.*csv* appears in the output folder *c:\OutputData*1. The file contains death rates for all country-populations:

DType	Scale	Popul	Year	Age	Female	Male	Total
Mx	1x1	AUS	1921	0	0.059987	0.076533	0.068444
Mx	1x1	AUS	1921	1	0.012064	0.014339	0.013225
Mx	1x1	AUS	1921	2	0.005779	0.006047	0.005916
Mx	1x1	AUS	1921	3	0.002889	0.004197	0.003554
Mx	1x1	AUS	1921	4	0.003254	0.003254	0.003254
Mx	1x1	AUS	1921	5	0.002517	0.002509	0.002513
Mx	1x1	AUS	1921	6	0.002485	0.002218	0.00235
Mx	1x1	AUS	1921	7	0.001812	0.001924	0.001869
Mx	1x1	AUS	1921	8	0.001384	0.001914	0.001651

Mx	1x1	AUS	1921	9	0.001371	0.001419	0.001395
Mx	1x1	AUS	1921	10	0.001249	0.001508	0.00138
Mx	1x1	AUS	1921	11	0.001159	0.002104	0.001639
Mx	1x1	AUS	1921	12	0.001128	0.001639	0.001387
Mx	1x1	AUS	1921	13	0.001417	0.001547	0.001482
Mx	1x1	AUS	1921	14	0.001435	0.001809	0.001625
Mx	1x1	AUS	1921	15	0.001602	0.001463	0.001531
Mx	1x1	AUS	1921	16	0.00181	0.002022	0.001918
Mx	1x1	AUS	1921	17	0.002166	0.002518	0.002343
Mx	1x1	AUS	1921	18	0.002477	0.002569	0.002523
Mx	1x1	AUS	1921	19	0.002201	0.002415	0.002308
Mx	1x1	AUS	1921	20	0.002976	0.002496	0.002737
Mx	1x1	AUS	1921	21	0.002542	0.003257	0.002896
Mx	1x1	AUS	1921	22	0.002535	0.003347	0.002934
Mx	1x1	AUS	1921	23	0.003258	0.003237	0.003248
Mx	1x1	AUS	1921	24	0.003139	0.003657	0.003388
.....							
.....							
Mx	1x1	AUS	1921	103	0	0.404313	0.212164
Mx	1x1	AUS	1921	104	1.039861	0.436999	0.615385
Mx	1x1	AUS	1921	105	.	0	0
Mx	1x1	AUS	1921	106	.	0	0
Mx	1x1	AUS	1921	107	.	0	0
Mx	1x1	AUS	1921	108	.	0	0
Mx	1x1	AUS	1921	109	.	0	0
Mx	1x1	AUS	1921	110+	.	0	0
.....							
.....							
Mx	1x1	AUS	2011	0	0.003418	0.004187	0.003813
Mx	1x1	AUS	2011	1	0.000259	0.000345	0.000303
Mx	1x1	AUS	2011	2	0.000161	0.000166	0.000163
Mx	1x1	AUS	2011	3	0.000155	0.00012	0.000137
Mx	1x1	AUS	2011	4	0.000114	0.000094	0.000104
Mx	1x1	AUS	2011	5	0.000108	0.000143	0.000126
Mx	1x1	AUS	2011	6	0.000081	0.000111	0.000096
Mx	1x1	AUS	2011	7	0.000052	0.000113	0.000083
Mx	1x1	AUS	2011	8	0.000135	0.000086	0.00011
Mx	1x1	AUS	2011	9	0.000098	0.000072	0.000084
Mx	1x1	AUS	2011	10	0.000104	0.000092	0.000098
Mx	1x1	AUS	2011	11	0.00003	0.000056	0.000043
Mx	1x1	AUS	2011	12	0.000096	0.000091	0.000094
Mx	1x1	AUS	2011	13	0.000089	0.000091	0.00009
Mx	1x1	AUS	2011	14	0.000117	0.000167	0.000143
Mx	1x1	AUS	2011	15	0.000188	0.00024	0.000215
.....							
.....							

Mx	1x1	USA	2013	0	0.005371	0.006518	0.005958
Mx	1x1	USA	2013	1	0.000377	0.000459	0.000419
Mx	1x1	USA	2013	2	0.00022	0.00029	0.000255
Mx	1x1	USA	2013	3	0.000163	0.000212	0.000188
Mx	1x1	USA	2013	4	0.000138	0.000182	0.00016
Mx	1x1	USA	2013	5	0.000125	0.000151	0.000138
Mx	1x1	USA	2013	6	0.000112	0.00013	0.000121
Mx	1x1	USA	2013	7	9.50E-05	0.00012	0.000108
Mx	1x1	USA	2013	8	9.70E-05	0.000124	0.000111
Mx	1x1	USA	2013	9	9.10E-05	0.000135	0.000113
Mx	1x1	USA	2013	10	9.70E-05	0.000112	0.000105
Mx	1x1	USA	2013	11	9.90E-05	0.000124	0.000112
Mx	1x1	USA	2013	12	0.000114	0.000153	0.000134
Mx	1x1	USA	2013	13	0.000143	0.000179	0.000161
Mx	1x1	USA	2013	14	0.000149	0.000234	0.000192
.....							
.....							
Mx	1x1	USA	2013	101	0.395261	0.470728	0.407509
Mx	1x1	USA	2013	102	0.433161	0.47475	0.439277
Mx	1x1	USA	2013	103	0.464843	0.527056	0.473084
Mx	1x1	USA	2013	104	0.485677	0.52569	0.490556
Mx	1x1	USA	2013	105	0.543724	0.628523	0.553062
Mx	1x1	USA	2013	106	0.550437	0.553458	0.550762
Mx	1x1	USA	2013	107	0.557404	0.536039	0.555127
Mx	1x1	USA	2013	108	0.701961	0.833003	0.714516
Mx	1x1	USA	2013	109	0.583899	0.401284	0.565767
Mx	1x1	USA	2013	110+	0.654499	1.056471	0.698257

Example 2. Pooling country-specific abridged (5x1) cohort life tables

Right-click on the links corresponding to “Life Tables - Male” and “Life Tables - Female” in the lower (cohort) part of the table on the webpage http://www.mortality.org/cgi-bin/hmd/hmd_download.php and download the files *c_lt_male.zip* and *c_lt_female.zip*. Unzipping of these two archives produces ten folders: *mltcoh*_1x1, *mltcoh*_1x5, *mltcoh*_1x10, *mltcoh*_5x1, *mltcoh*_5x5, *mltcoh*_5x10, *fltcoh*_1x1, *fltcoh*_1x5, *fltcoh*_1x10, *fltcoh*_5x1, *fltcoh*_5x5, *fltcoh*_5x10. Country files from the folders *mltcoh*_5x1 and *fltcoh*_5x1 contain abridged life tables (5-year age intervals) by single calendar years for 13 populations. Copy 26 male and female life tables into folder *c:\InputData*1\ , open R-Studio and after sourcing of the file *Pool.HMD.Mx.files*.R (see the “Installation” section for details) call the function

Pool.HMD.LT.files("*c:/InputData*2", "*c:/OutputData*2", " *Cohort-LTs*-5x1.*txt*", *Separator*=",").

During execution of the code, the following lines appear in the Console window:

```
file # 1     CHE.fltcoh_5x1.txt
file # 2     CHE.mltcoh_5x1.txt
file # 3     DNK.fltcoh_5x1.txt
file # 4     DNK.mltcoh_5x1.txt
file # 5     FIN.fltcoh_5x1.txt
file # 6     FIN.mltcoh_5x1.txt
file # 7     FRACNP.fltcoh_5x1.txt
file # 8     FRACNP.mltcoh_5x1.txt
file # 9     FRATNP.fltcoh_5x1.txt
file # 10     FRATNP.mltcoh_5x1.txt
file # 11     GBR_SCO.fltcoh_5x1.txt
file # 12     GBR_SCO.mltcoh_5x1.txt
file # 13     GBRCENW.fltcoh_5x1.txt
file # 14     GBRCENW.mltcoh_5x1.txt
file # 15     GBRTENW.fltcoh_5x1.txt
file # 16     GBRTENW.mltcoh_5x1.txt
file # 17     ISL.fltcoh_5x1.txt
file # 18     ISL.mltcoh_5x1.txt
file # 19     ITA.fltcoh_5x1.txt
file # 20     ITA.mltcoh_5x1.txt
file # 21     NLD.fltcoh_5x1.txt
file # 22     NLD.mltcoh_5x1.txt
file # 23     NOR.fltcoh_5x1.txt
file # 24     NOR.mltcoh_5x1.txt
file # 25     SWE.fltcoh_5x1.txt
file # 26     SWE.mltcoh_5x1.txt
```

The output file Cohort-LTs-5x1.txt appears in the output folder *c:\OutputData*2. The file contains abridged life tables for males and females for all 13 populations:

DType	Scale	Popul	Sex	Year	Age	mx	qx	ax	lx	dx	LLx	Tx	ex
ltcoh	5x1	CHE	f	1876	0	0.2306	0.2005	0.35	100000	20050	86967	4706115	47.06
ltcoh	5x1	CHE	f	1876	1-4	0.0226	0.08545	1.4	79950	6831	302008	4619148	57.78
ltcoh	5x1	CHE	f	1876	5-9	0.0057	0.028	2.01	73118	2047	359461	4317140	59.04
ltcoh	5x1	CHE	f	1876	10-14	0.0034	0.01686	2.76	71071	1198	352668	3957679	55.69
ltcoh	5x1	CHE	f	1876	15-19	0.0052	0.02577	2.62	69873	1801	345074	3605011	51.59
ltcoh	5x1	CHE	f	1876	20-24	0.0056	0.02745	2.52	68072	1869	335728	3259937	47.89
ltcoh	5x1	CHE	f	1876	25-29	0.006	0.0293	2.54	66203	1940	326254	2924209	44.17
ltcoh	5x1	CHE	f	1876	30-34	0.0063	0.03077	2.48	64264	1977	316328	2597955	40.43
ltcoh	5x1	CHE	f	1876	35-39	0.0061	0.03008	2.54	62286	1873	306830	2281626	36.63
ltcoh	5x1	CHE	f	1876	40-44	0.0079	0.03866	2.52	60413	2335	296283	1974797	32.69
ltcoh	5x1	CHE	f	1876	45-49	0.0072	0.03559	2.58	58077	2067	285384	1678514	28.90
ltcoh	5x1	CHE	f	1876	50-54	0.0101	0.04933	2.61	56011	2763	273444	1393130	24.87
ltcoh	5x1	CHE	f	1876	55-59	0.0141	0.06807	2.55	53247	3625	257374	1119687	21.03
ltcoh	5x1	SWE	m	1923	50-54	0.0069	0.0339	2.68	82839	2808	407667	2328254	28.11
ltcoh	5x1	SWE	m	1923	55-59	0.0107	0.05239	2.66	80031	4192	390347	1920587	24
ltcoh	5x1	SWE	m	1923	60-64	0.0158	0.07594	2.65	75838	5759	365657	1530241	20.18
ltcoh	5x1	SWE	m	1923	65-69	0.0229	0.10857	2.59	70079	7609	332046	1164584	16.62
ltcoh	5x1	SWE	m	1923	70-74	0.034	0.15723	2.58	62471	9822	288571	832537	13.33
ltcoh	5x1	SWE	m	1923	75-79	0.0527	0.23366	2.57	52648	12302	233297	543966	10.33
ltcoh	5x1	SWE	m	1923	80-84	0.0829	0.34366	2.51	40347	13865	167273	310669	7.7
ltcoh	5x1	SWE	m	1923	85-89	0.1433	0.52449	2.44	26481	13889	96909	143395	5.41
ltcoh	5x1	SWE	m	1923	90-94	0.2404	0.72195	2.23	12592	9091	37811	46486	3.69
ltcoh	5x1	SWE	m	1923	95-99	0.386	0.87607	1.88	3501	3067	7945	8675	2.48
ltcoh	5x1	SWE	m	1923	100-104	0.5895	0.95959	1.49	434	416	706	730	1.68
ltcoh	5x1	SWE	m	1923	105-109	0.7467	1	1.34	18	18	23	23	1.34
ltcoh	5x1	SWE	m	1923	110+	.	.	.	0	0	0	0	.