Methods & software for calculating health expectancy.

Worked examples, results and cautionary tales.

Vikki O'Neill

REVES - 27th May 2014

Outline

- Worked example
  - iMaCh
  - ELECT
  - SPACE
- Comparisons
  - ADL
  - MMSE
  - Stroke
- Summary

Recap

- CFAS Data - Females
- No missing states at baseline
- No two events in same month

- IMaCh: A maximum likelihood computer program using Interpolation of Markov Chains
- ELECT: Estimation of Life Expectancies using Continuous-Time multi-state models
- SPACE: Stochastic Population Analysis for Complex Events

iMaCh website

- There are three main web-pages[1] for iMaCh:
  1. euroreves.ined.fr/imach/
  2. euroreves.ined.fr/imach/wiki/index.php/Main_Page
  3. euroreves.ined.fr/imach/wiki/index.php/Documentation
- The first website contains the download instructions for installation for different platforms; Windows (win9X, 2000, XP, Vista), MacOS/X Leopard and Linux.
- The second website contains an overview of the software, with the main documentation contained on the third website.

[1]Note: There is also euroreves.ined.fr/imach/doc/imach.htm, but this links to documentation on the older Version 0.97 (June 2004) of IMaCh

iMaCh data setup I

Each line of the data set is an individual record.
Fields are separated by blank spaces:

First Column	Unique index number	Must be a positive number
Covariate Columns	First Covariate, Second Covariate, Third Covariate, etc.	Boolean, i.e., should be 1 or 0
Weight	Individuals can be weighted to account for sample stratification	Must be a positive number

Even if you do not wish to include weights or covariates in the model, you must still fill the column with a number (e.g. 1) as all fields must be present.

iMaCh data setup II

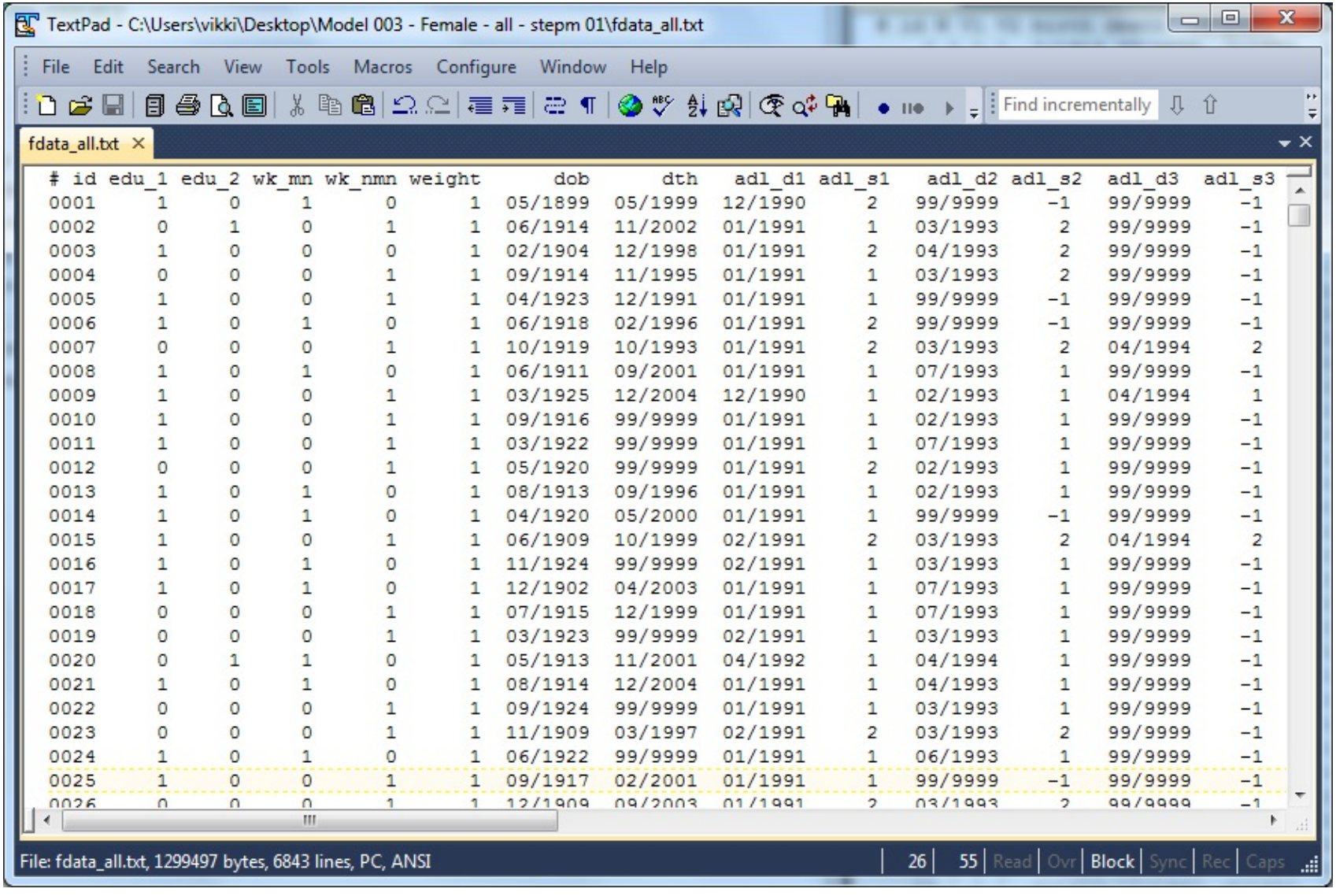
Date of birth	coded as mm/yyyy	Cannot be missing
Date of death	coded as mm/yyyy	If death occurs this cannot be missing. Otherwise censored observations are coded as 99/9999
Date of 1st interview	coded as mm/yyyy	Cannot be missing
Status at 1st interview	positive number	Cannot be missing

iMaCh data setup III

Date of 2nd interview	coded as mm/yyyy	Missing dates are coded as 99/9999
Status at 2nd interview	positive number	Missing values are coded -1
Date of nth interview	coded as mm/yyyy	Missing dates are coded as 99/9999
Status at nth interview	positive number	Missing values are coded -1

- If a death occurs before the *n*th interview: code the status at interview *n* to the death state and include a date of death in the 'date of death' column - it does not matter what date is included in the date column for this *n*th interview.
- All interview dates and states after this can be coded as either dead or missing - it doesn't make a difference.

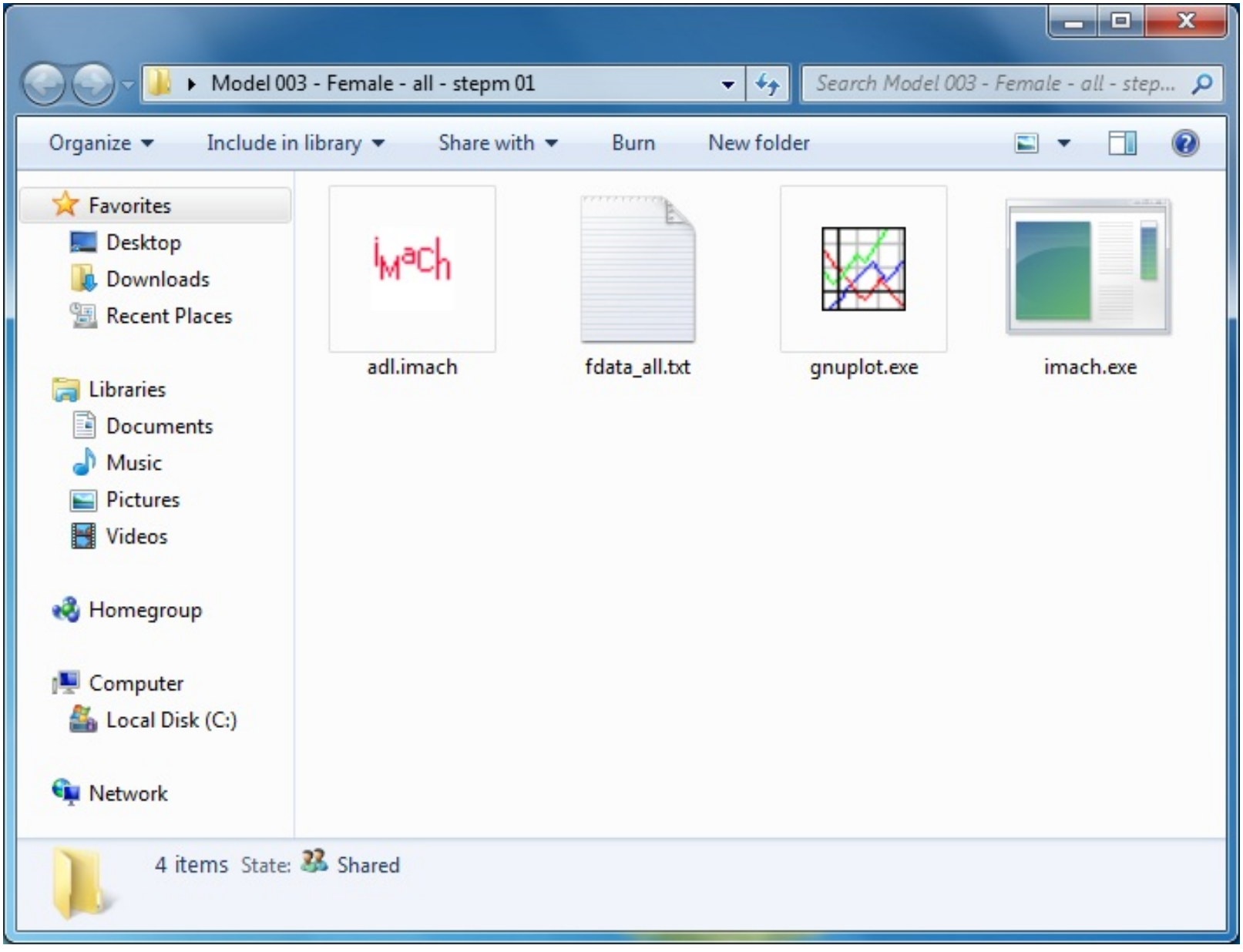
Data



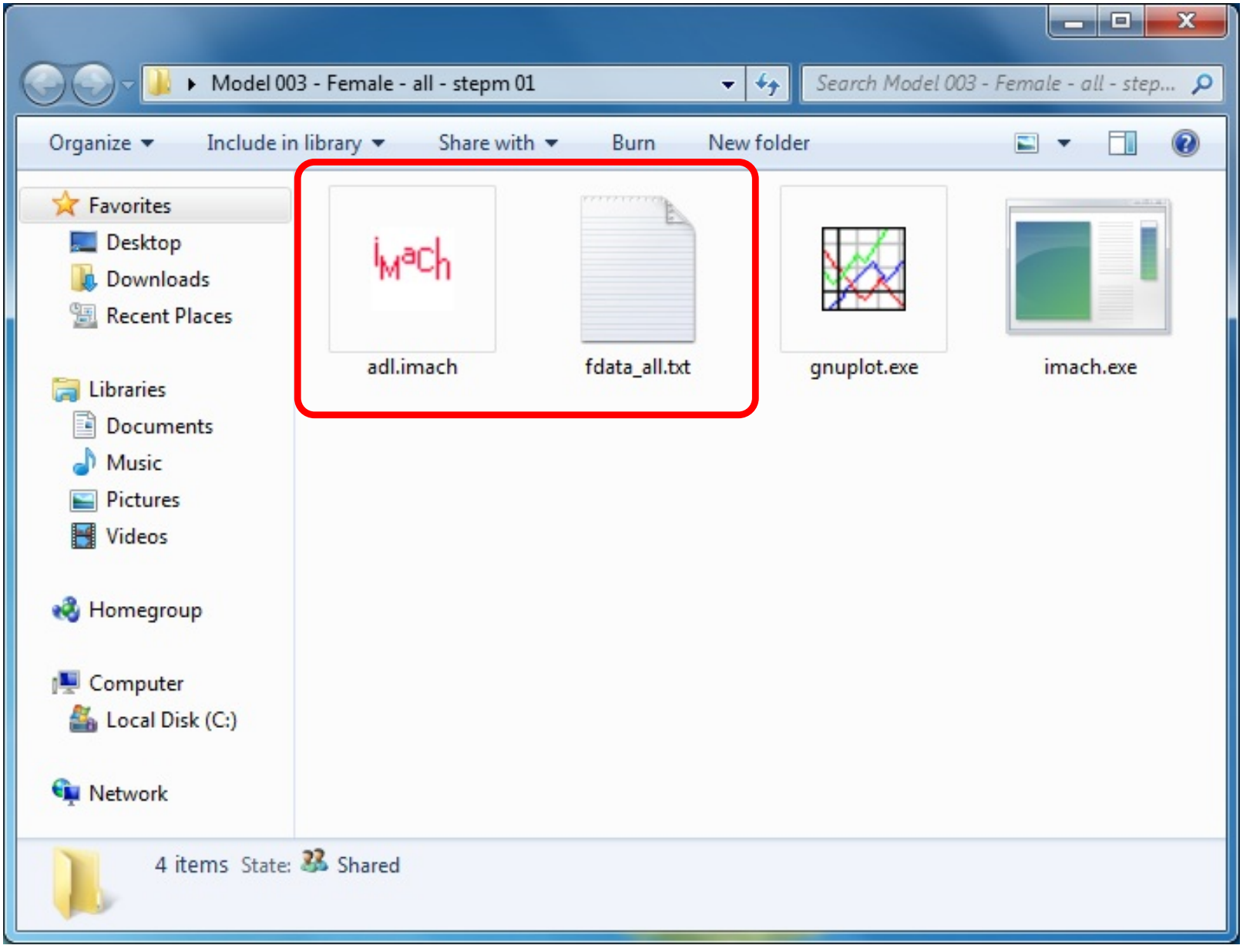
# id	edu_1	edu_2	wk_mn	wk_nmn	weight	dob	dth	adl_d1	adl_s1	adl_d2	adl_s2	adl_d3	adl_s3
0001	1	0	1	0	1	05/1899	05/1999	12/1990	2	99/9999	-1	99/9999	-1
0002	0	1	0	1	1	06/1914	11/2002	01/1991	1	03/1993	2	99/9999	-1
0003	1	0	0	0	1	02/1904	12/1998	01/1991	2	04/1993	2	99/9999	-1
0004	0	0	0	1	1	09/1914	11/1995	01/1991	1	03/1993	2	99/9999	-1
0005	1	0	0	1	1	04/1923	12/1991	01/1991	1	99/9999	-1	99/9999	-1
0006	1	0	1	0	1	06/1918	02/1996	01/1991	2	99/9999	-1	99/9999	-1
0007	0	0	0	1	1	10/1919	10/1993	01/1991	2	03/1993	2	04/1994	2
0008	1	0	1	0	1	06/1911	09/2001	01/1991	1	07/1993	1	99/9999	-1
0009	1	0	0	1	1	03/1925	12/2004	12/1990	1	02/1993	1	04/1994	1
0010	1	0	0	1	1	09/1916	99/9999	01/1991	1	02/1993	1	99/9999	-1
0011	1	0	0	1	1	03/1922	99/9999	01/1991	1	07/1993	1	99/9999	-1
0012	0	0	0	1	1	05/1920	99/9999	01/1991	2	02/1993	1	99/9999	-1
0013	1	0	1	0	1	08/1913	09/1996	01/1991	1	02/1993	1	99/9999	-1
0014	1	0	1	0	1	04/1920	05/2000	01/1991	1	99/9999	-1	99/9999	-1
0015	1	0	0	1	1	06/1909	10/1999	02/1991	2	03/1993	2	04/1994	2
0016	1	0	1	0	1	11/1924	99/9999	02/1991	1	03/1993	1	99/9999	-1
0017	1	0	1	0	1	12/1902	04/2003	01/1991	1	07/1993	1	99/9999	-1
0018	0	0	0	1	1	07/1915	12/1999	01/1991	1	07/1993	1	99/9999	-1
0019	0	0	0	1	1	03/1923	99/9999	02/1991	1	03/1993	1	99/9999	-1
0020	0	1	1	0	1	05/1913	11/2001	04/1992	1	04/1994	1	99/9999	-1
0021	1	0	1	0	1	08/1914	12/2004	01/1991	1	04/1993	1	99/9999	-1
0022	0	0	0	1	1	09/1924	99/9999	01/1991	1	03/1993	1	99/9999	-1
0023	0	0	0	1	1	11/1909	03/1997	02/1991	2	03/1993	2	99/9999	-1
0024	1	0	1	0	1	06/1922	99/9999	01/1991	1	06/1993	1	99/9999	-1
0025	1	0	0	1	1	09/1917	02/2001	01/1991	1	99/9999	-1	99/9999	-1
0026	0	0	0	1	1	12/1909	09/2003	01/1991	2	03/1993	2	99/9999	-1

Warning: iMaCh has a character limit on the number of columns!

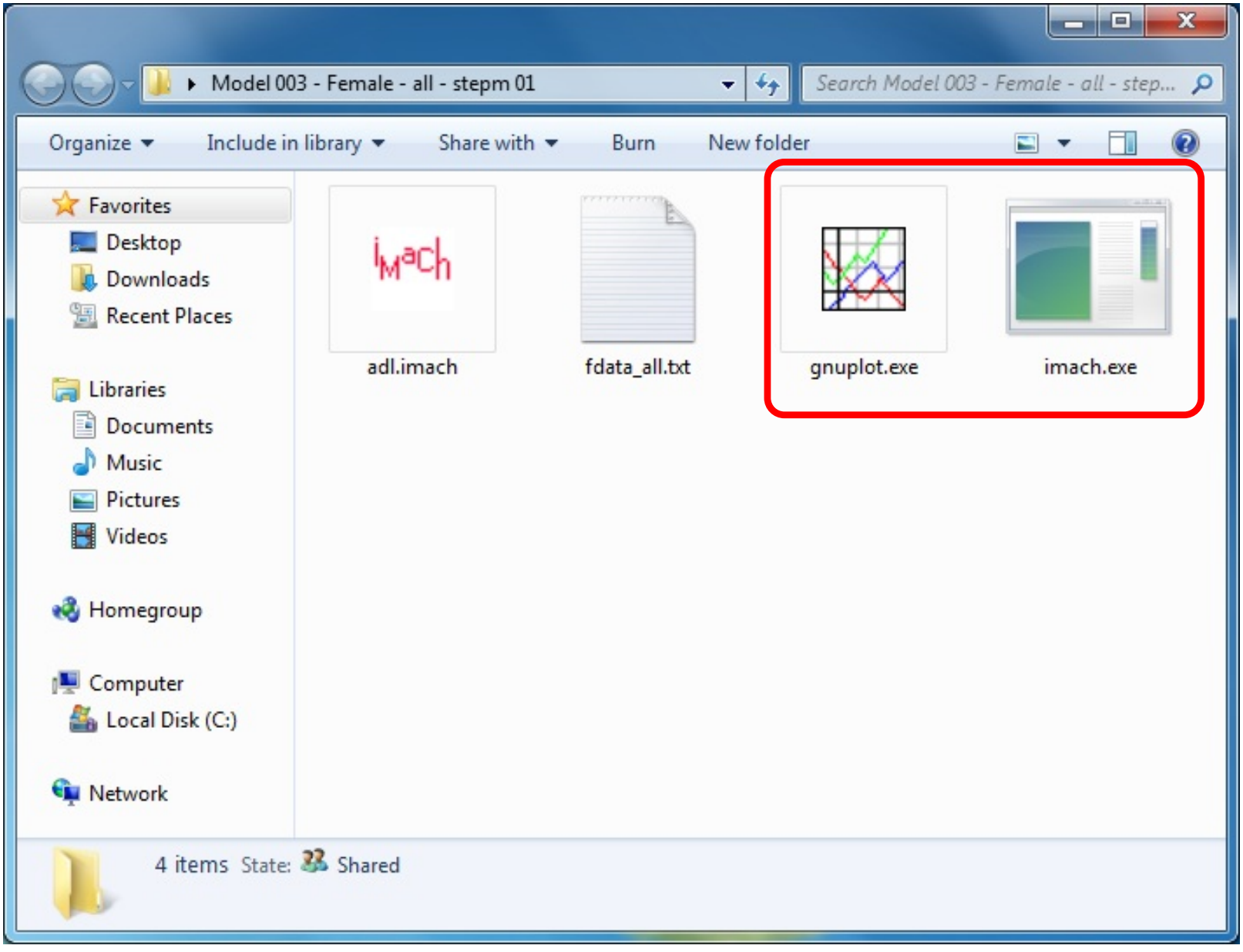
Running iMaCh - within folder



Running iMaCh - within folder



Running iMaCh - within folder



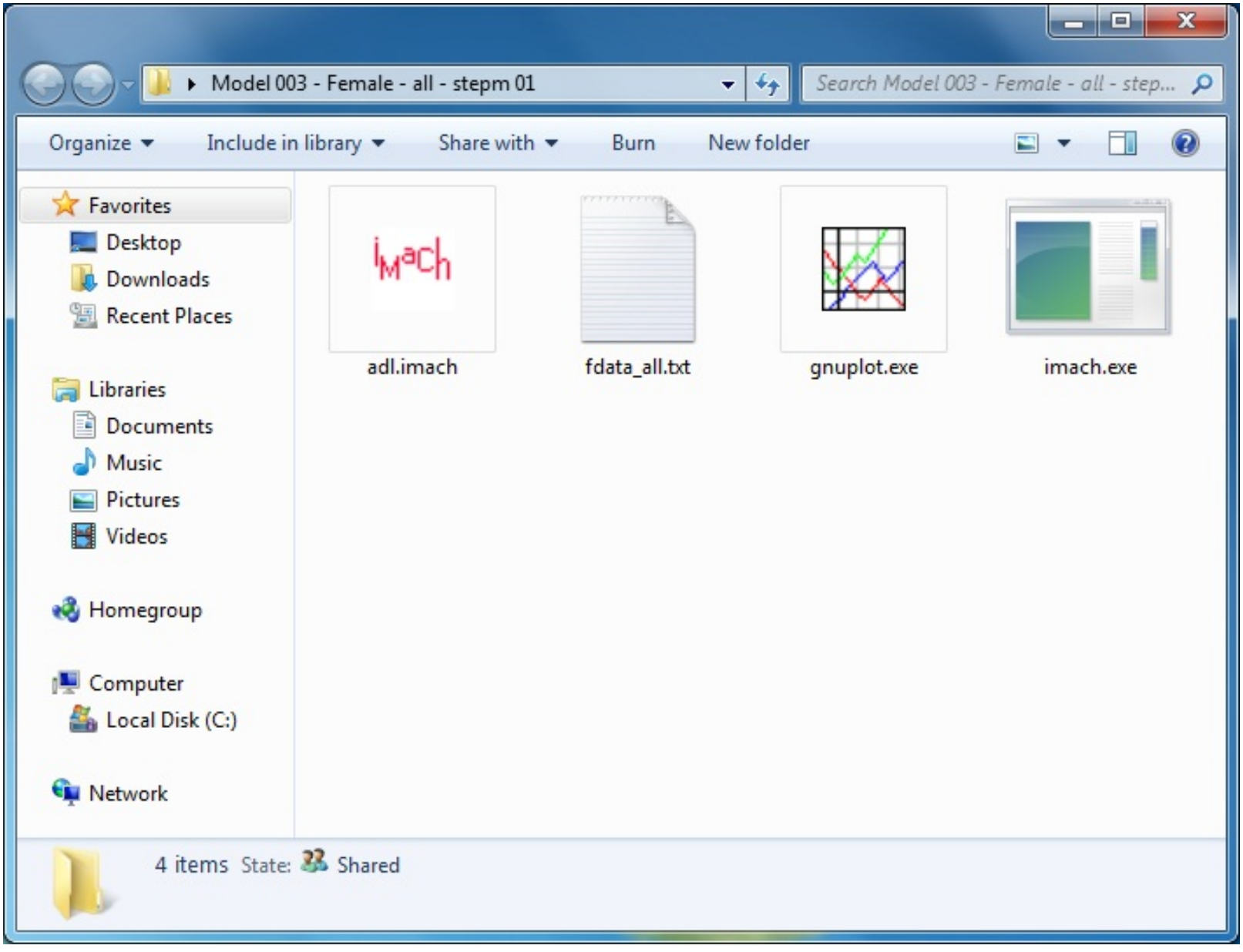
Running iMaCh - within folder

There are a number of files you need to have within your working folder in order to run analysis in iMaCh. These are:

- iMaCh executible file
- GNU Plot
- data
- .imach file containing model/parameter specifications

Within the .imach file there are a number of settings..

Running iMaCh - within folder

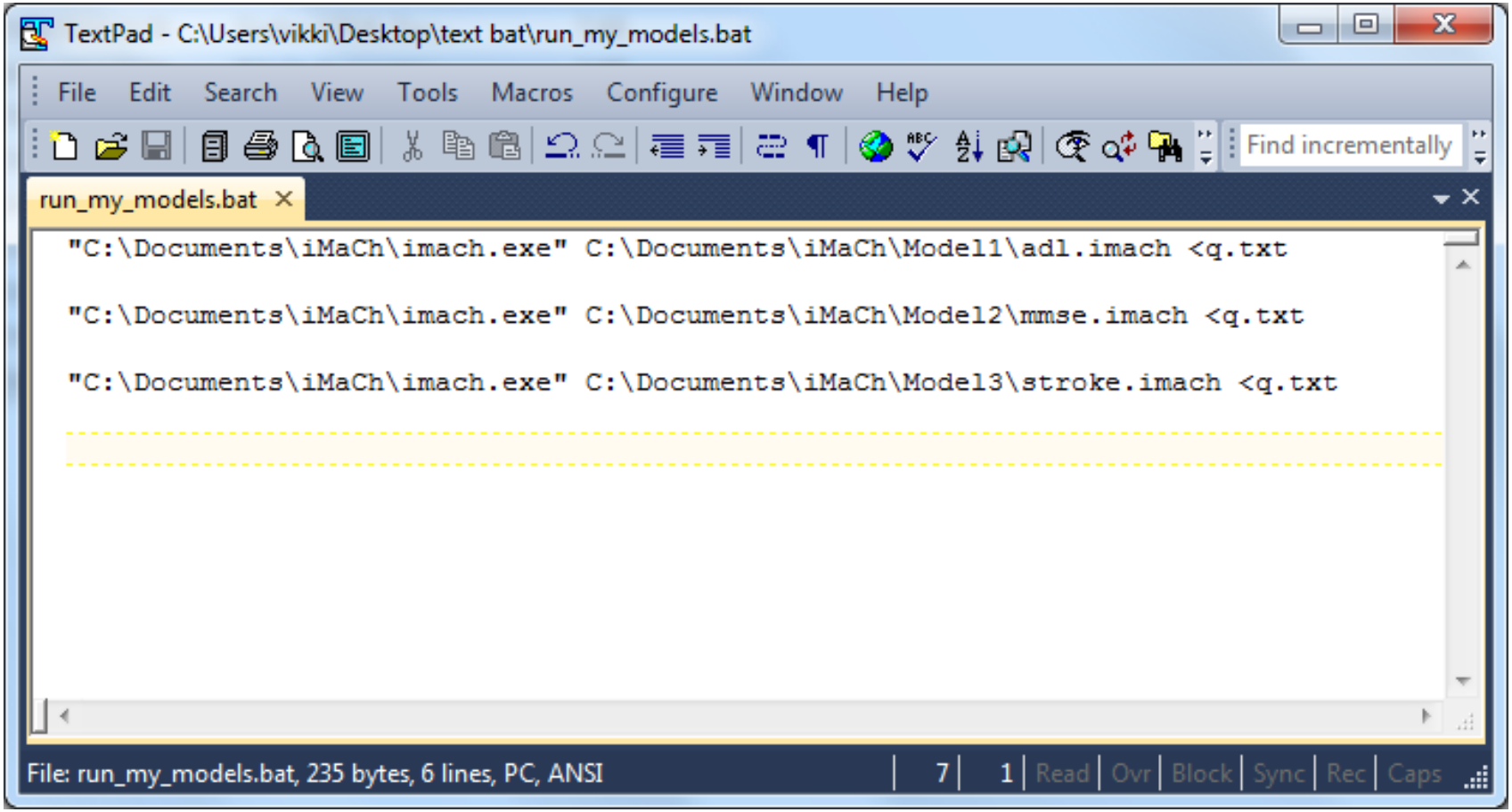


Running iMaCh - using a bat file

- Create **q.txt**: this is a .txt file containing only the letter q
- Create a .bat file using a text editor, listing the 3 items below, on one line separated by spaces:
  - Location in quote marks of the iMaCh.exe programme
  - Location not in quote marks of the iMaCh parameter file
  - <q.txt
- Double click on the .bat file when you would like it to run the models

- **Advantage:** You can run several models in sequence.
- **Warning:** If the model fails you will not see any error messages - the programme will just close - you can still find these in the .log files.

Running iMaCh - using a bat file



iMaCh parameter file I

adl.imach - Notepad

File Edit Format View Help

```
#Imach version 0.98k, June 2009, INED-EUROREVES-Institut de longevite
title=females datafile=fdata_all.txt lastobs=13000 firstpass=1 lastpass=8
ftol=1.000000e-008 stepm=12 ncovcol=4 nlstate=2 ndeath=1 maxwav=8 mle=1 weight=0
model=.
# Parameters nlstate*nlstate*ncov a12*1 + b12 * age + ...
12 0. 0.
13 0. 0.
21 0. 0.
23 0. 0.
# Scales (for hessian or gradient estimation)
12 0. 0.
13 0. 0.
21 0. 0.
23 0. 0.
#covariance matrix#
121 0.
122 0. 0.
131 0. 0. 0.
132 0. 0. 0. 0.
211 0. 0. 0. 0. 0.
212 0. 0. 0. 0. 0. 0.
231 0. 0. 0. 0. 0. 0. 0.
232 0. 0. 0. 0. 0. 0. 0. 0.
# agemin agemax for life expectancy, bage fage (if mle==0 ie no data nor Max likelihood).
agemin=64 agemax=120 bage=65 fage=100 estepm=12
# Observed prevalence period
begin-prev-date=01/01/1991 end-prev-date=31/12/1994 mov_average=0
# Health expectancies computed from stationary prevalence (pop_based=0) or population based (1)
pop_based=1
# Prevalence forecasting
prevforecast=0 starting-proj-date=01/01/1991 final-proj-date=31/12/1994 mobil_average=0
```

iMaCh parameter file II

```
#Imach version 0.98k, June 2009, INED-EUROREVES-Institut
de longevite
```

This first line is a comment. Comment lines start with a '#'.
You cannot add comments to the parameter file!

```
title=females datafile=fdata_all.txt lastobs=13000
firstpass=1 lastpass=8
```

- **title**=females is any given name for the model.
- **datafile**=fdata_all.txt is the name of the data set.
- **lastobs**=13000 the program is able to run on a subsample where the last observation number is lastobs. It can be set to be a bigger number than the true number of observations (e.g. 100000).
- **firstpass=1, lastpass=8** firstpass=1 means the first wave in the survey. lastpass=8 means the last wave to be included will be the 8th.

iMaCh parameter file III

```
ftol=1.000000e-008 stepm=12 ncovcol=4 nlstate=2 ndeath=1
maxwav=8 mle=1 weight=0
```

- **ftol=1e-8** Convergence tolerance
- **stepm=1** The time unit in months for interpolation.
- **ncovcol=2** Number of covariate columns included in the datafile before the column for the date of birth.
- **nlstate=2** Number of non-absorbing (alive) states.
- **ndeath=1** Number of absorbing states.
- **maxwav=4** Number of waves in the datafile.
- **mle=1** Option for the Maximisation Likelihood Estimation.
  - If mle=1 the program does the maximisation and the calculation of health expectancies
  - If mle=-1 you get a template for the number of parameters and the size of the variance-covariance matrix.
- **weight=0** Provides the possibility of adding weights.

iMaCh parameter file IV

```
model=.
```

Intercept and age are automatically included in the model.

Additional covariates can be included with the command:

- if **model=.** then no covariates are included
- if **model=V1** the model includes the first covariate
- if **model=V2** the model includes the second covariate
- if **model=V1+V2** the model includes the first and the second covariate
- if **model=V1*V2** the model includes the product of the first and the second covariate
- if **model=V1+V1*age** the model includes the product covariate*age

iMaCh parameter file V

- You must write the initial guess values of the parameters for optimization.
- The number of parameters, N depends on the number of absorbing states and non-absorbing states and on the number of covariates in the model (ncovmodel).
- Each of the four lines starts with indices "ij": **ij aij bij**

```
# Parameters nlstate*nlstate*ncov a12*1 + b12 * age + ...
12   0.0   0.0
13   0.0   0.0
21   0.0   0.0
23   0.0   0.0
```

- In order to speed up the convergence you can make a first run with a large stepm i.e stepm=12 or 24 and then decrease the stepm until stepm=1 month.

iMaCh parameter file VI

- The 'scales' are needed to compute the numerical derivatives. These derivatives are used to compute the hessian matrix of the parameters. They are often used for estimating variances and confidence intervals.
- Each line consists of indices "ij" followed by the initial scales associated with aij and bij:
  - If mle=1 you can enter zeros
  - If mle=0 (no maximisation of Likelihood) you must enter a covariance matrix (usually obtained from an earlier run).

```
# Scales (for hessian or gradient estimation)
12   0.   0.
13   0.   0.
21   0.   0.
23   0.   0.
```

iMaCh parameter file VII

- The covariance matrix is output if mle=1. But it can be also be used as an input to get the various output data files (Health expectancies, period prevalence etc.) and figures without rerunning the maximisation phase (mle=0).
- Each line starts with indices "ijk" followed by the covariances between aij and bij:

```
121 Var(a12)
122 Cov(b12,a12) Var(b12)
    ...
232 Cov(b23,a12) Cov(b23,b12) ... Var (b23)
```

iMaCh parameter file VIII

Age range for calculation of stationary prevalences and health expectancies:

```
# agemin agemax for life expectancy, bage fage (if mle==0
ie no data nor Max likelihood).
agemin=64 agemax=120 bage=65 fage=100 estepm=12
```

- **agemin=** Minimum age for calculation of the period prevalence
- **agemax=** Maximum age for calculation of the period prevalence
- **bage=** Minimum age for calculation of the health expectancies
- **fage=** Maximum age for calculation of the health expectancies
- **estepm=** Unit (in months). We compute the life expectancy from trapezoids spaced every estepm months. A more precise result will be obtained if estepm is as small as stepm.

iMaCh parameter file IX

Computing the cross-sectional prevalence

```
# Observed prevalence period
begin-prev-date=01/01/1991 end-prev-date=31/12/1994
mov_average=0
```

Statements 'begin-prev-date' and 'end-prev-date' allow you to select the period in which the observed prevalences in each state.

- **begin-prev-date=** Starting date (day/month/year)
- **end-prev-date=** Final date (day/month/year)
- **mov_average=0** *Necessary for population–based health expectancy calculations*

iMaCh parameter file X

Population- or status-based health expectancies

```
# Health expectancies computed from stationary prevalence
(pop_based=0) or population based (1)
pop_based=1
```

- The program can compute both population and status-based health expectancies, i.e health expectancies which do and do not depend on the initial health state.
- TRICKY!

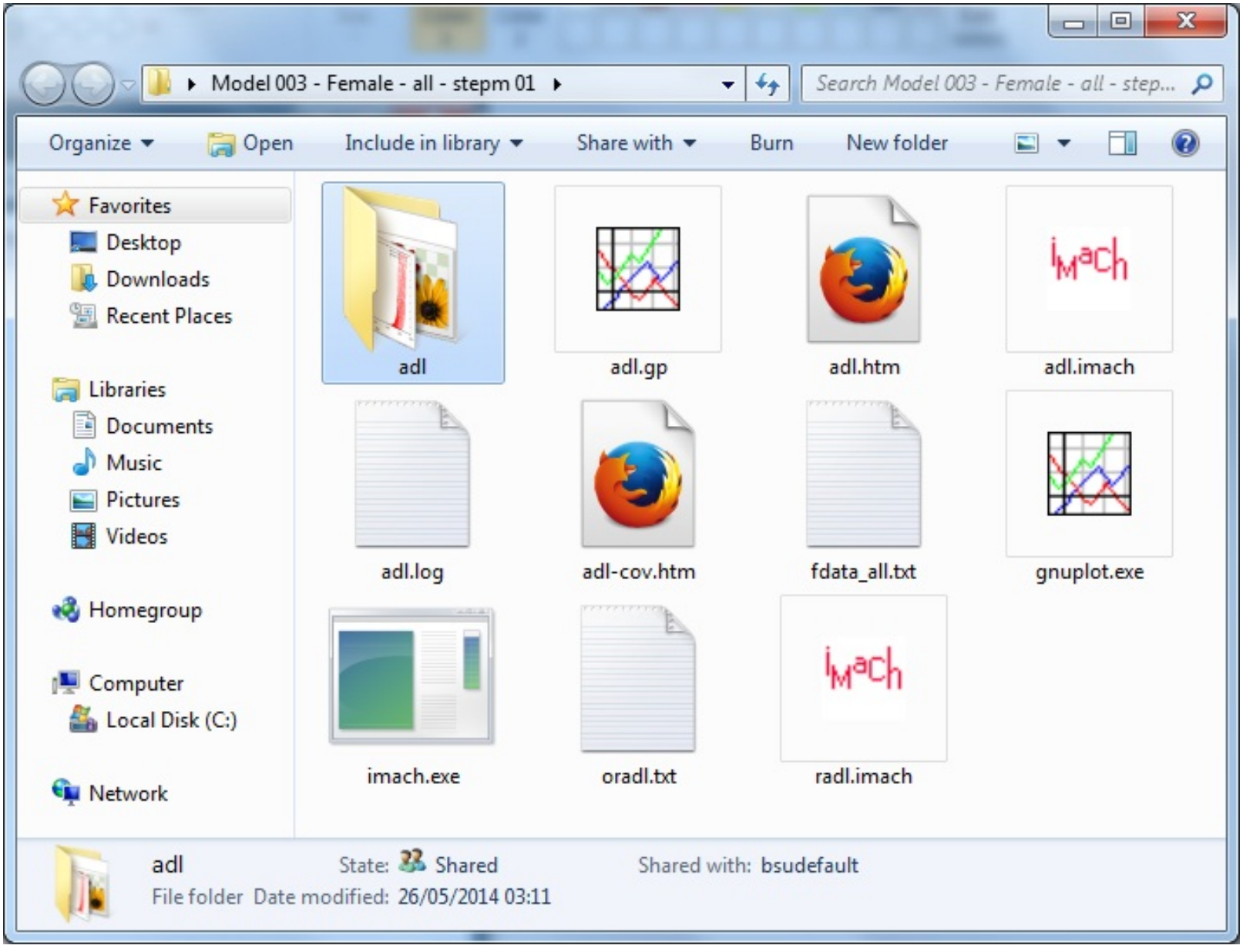
iMaCh parameter file XI

Prevalence forecasting (Experimental)

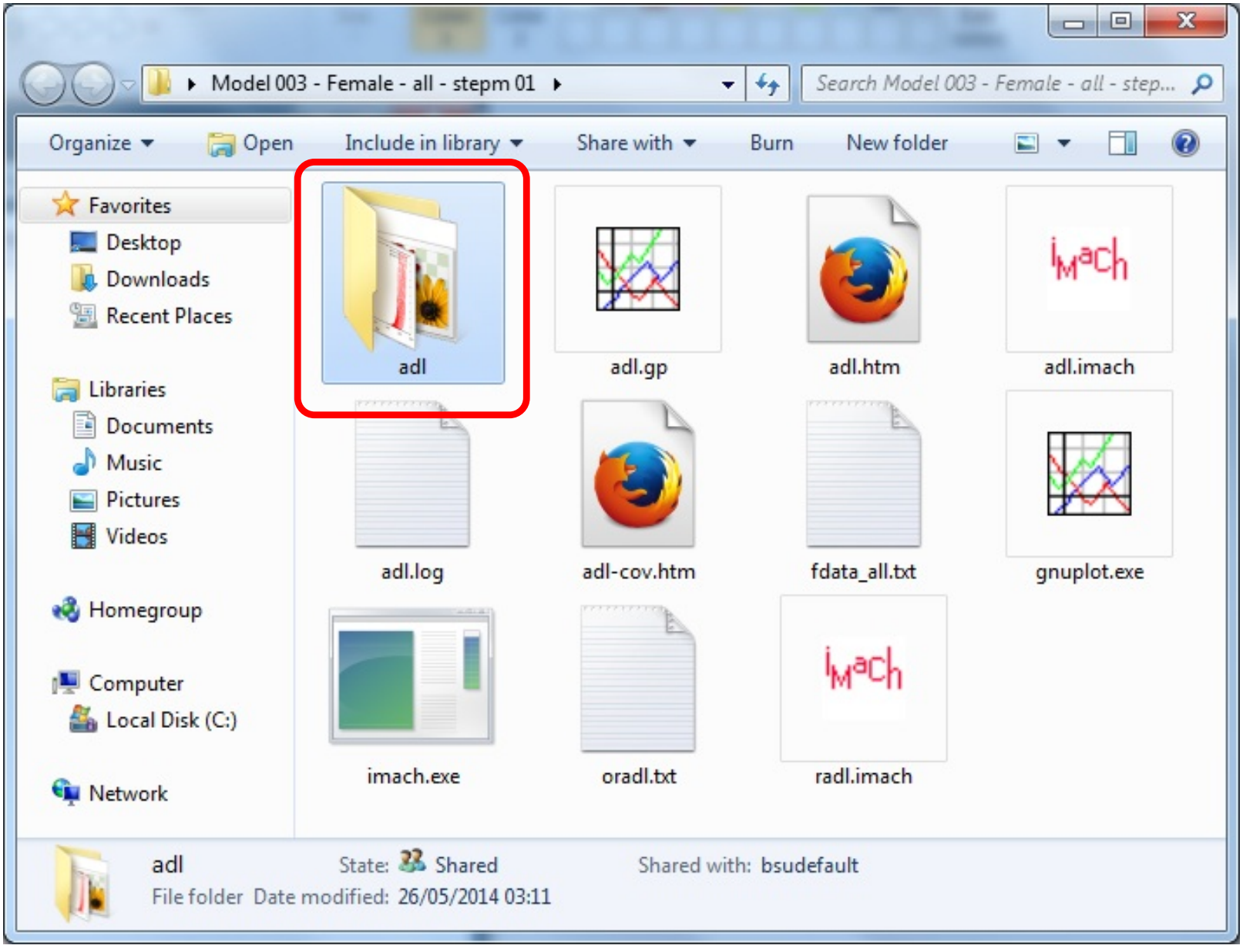
```
# Prevalence forecasting
prevforecast=0 starting-proj-date=01/01/1991
final-proj-date=31/12/1994 mobil_average=0
```

- Prevalence and population projections are only available if the interpolation unit is a month, i.e. stepm=1 and if there are no covariate

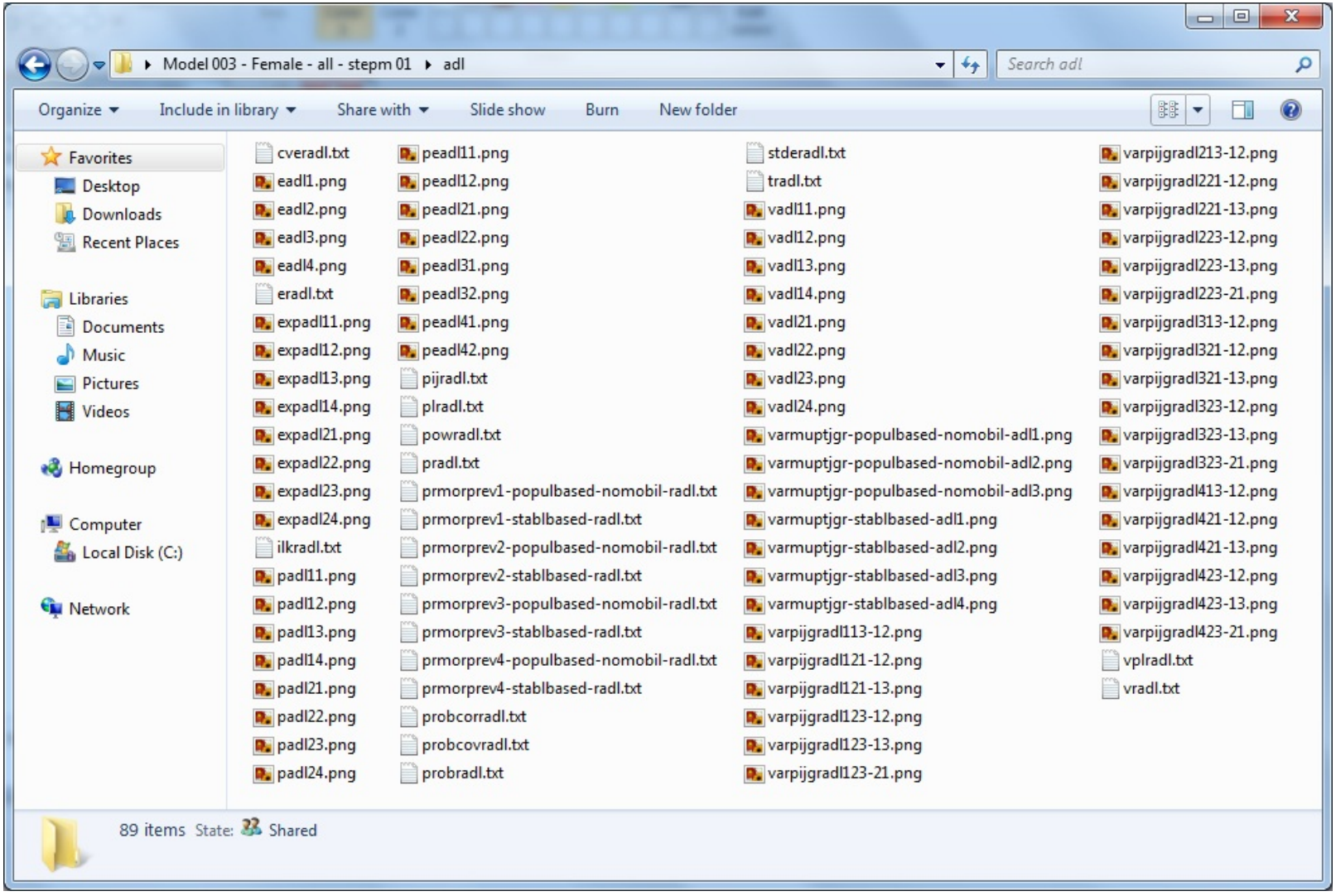
Output files - top level



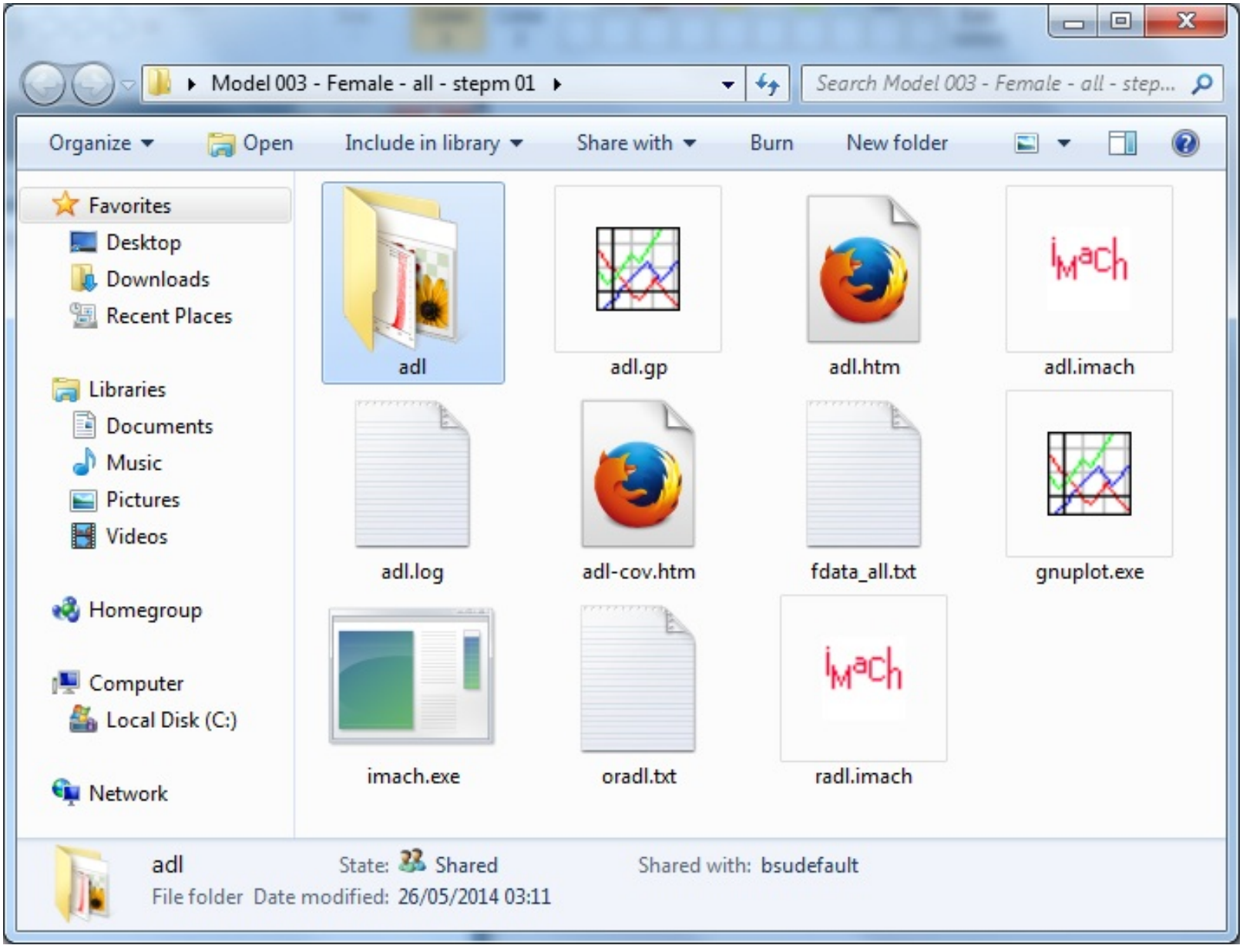
Output files - top level



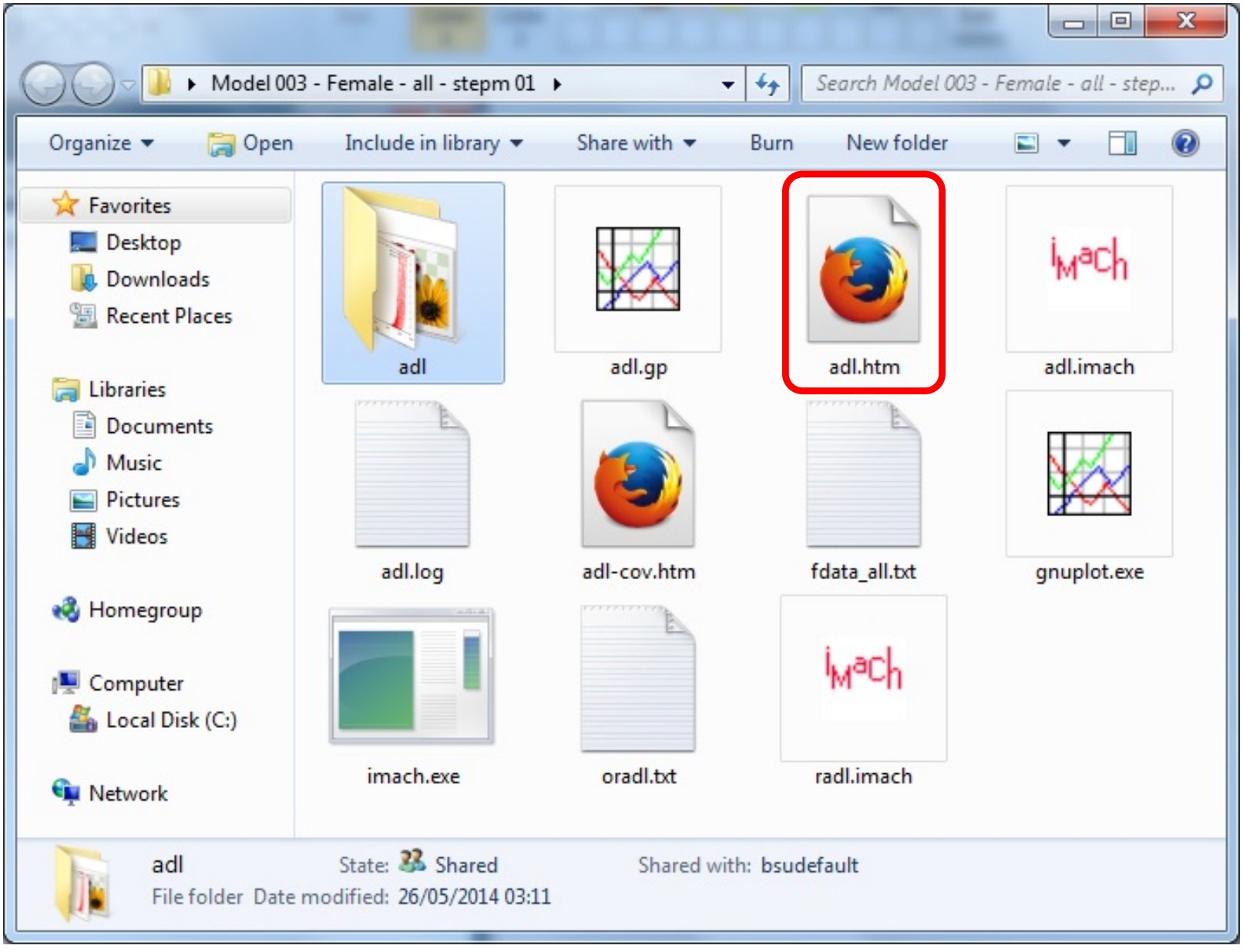
Output files - ADL folder



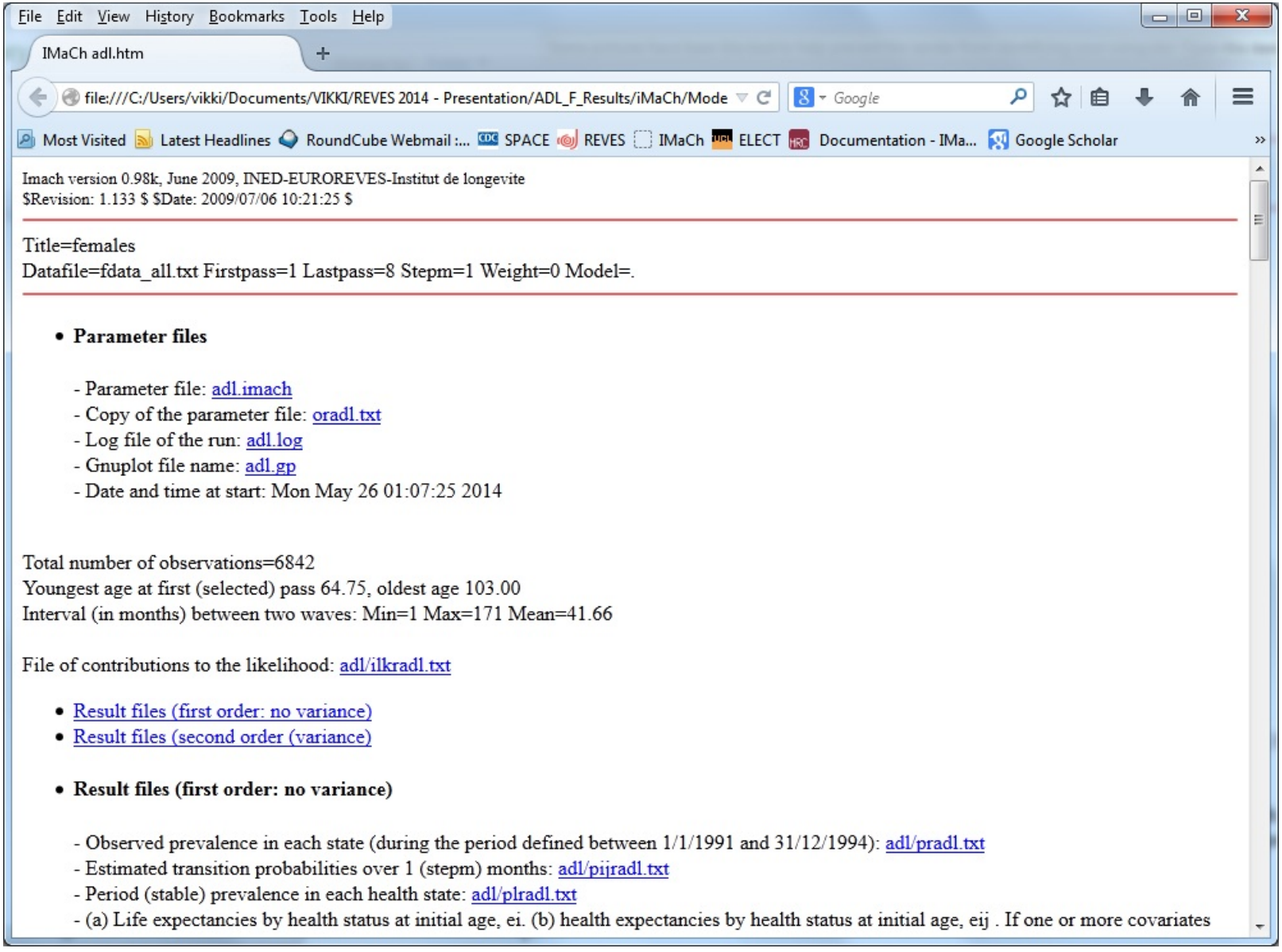
Output files - webpage



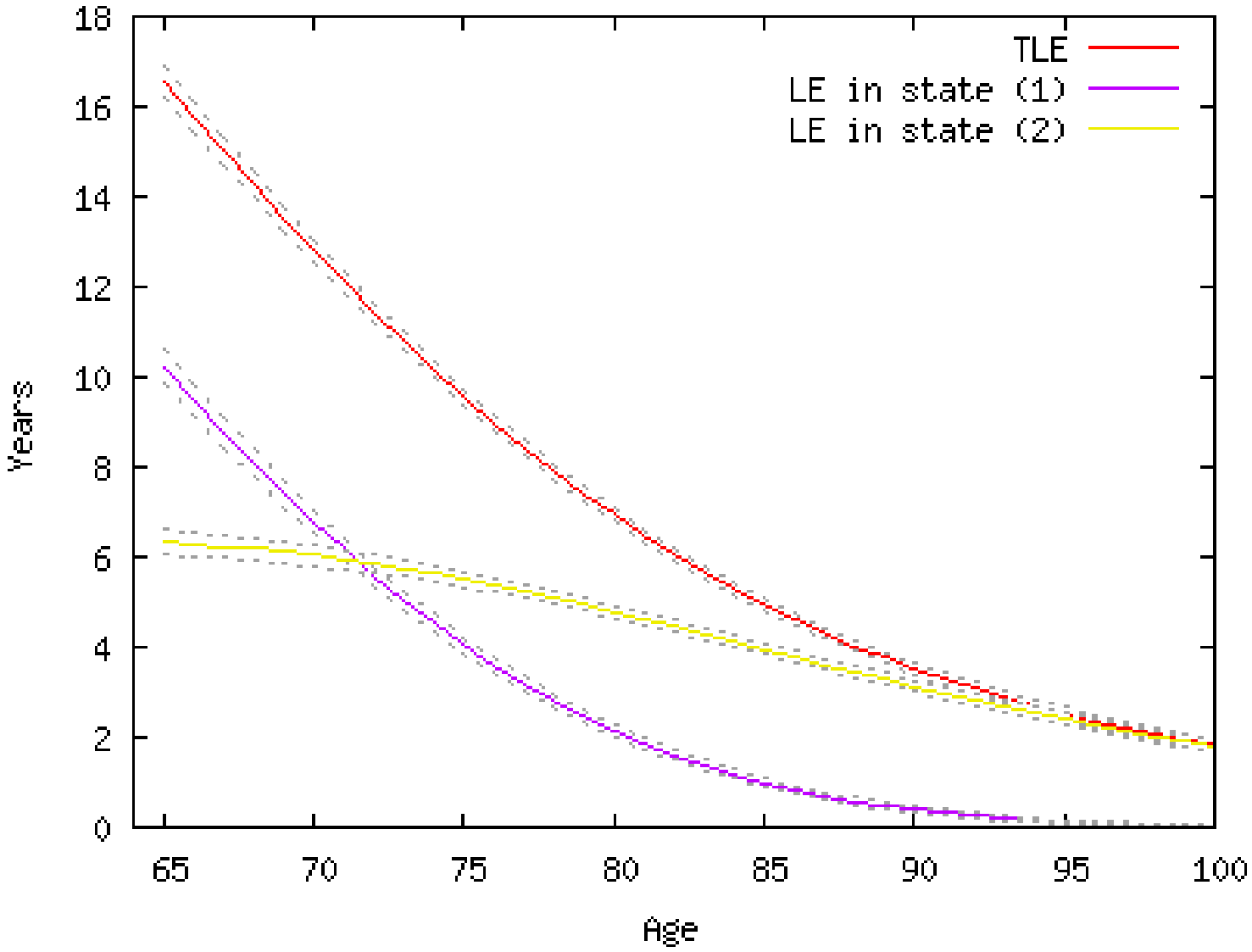
Output files - webpage



Results webpage

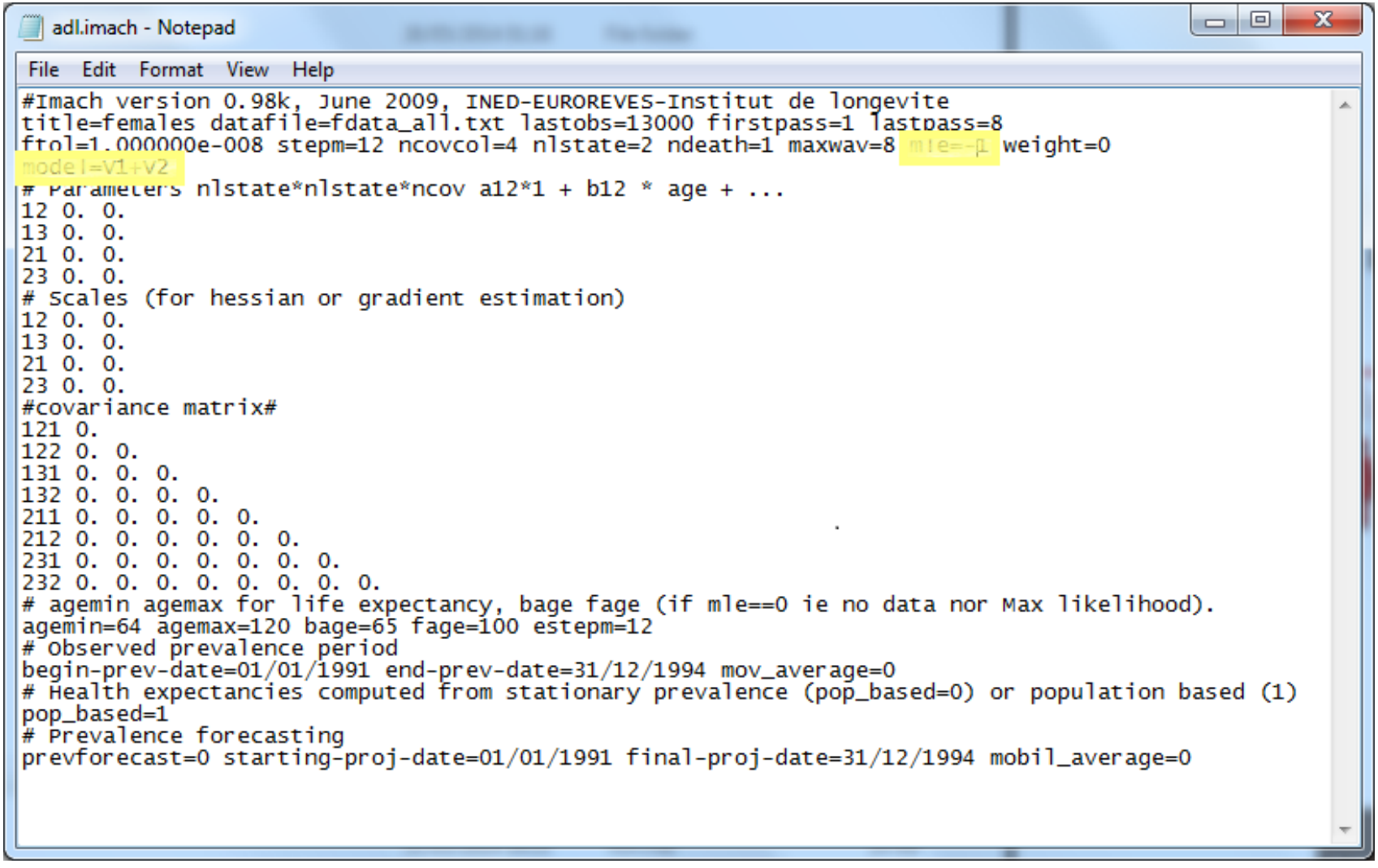


IMaCh results

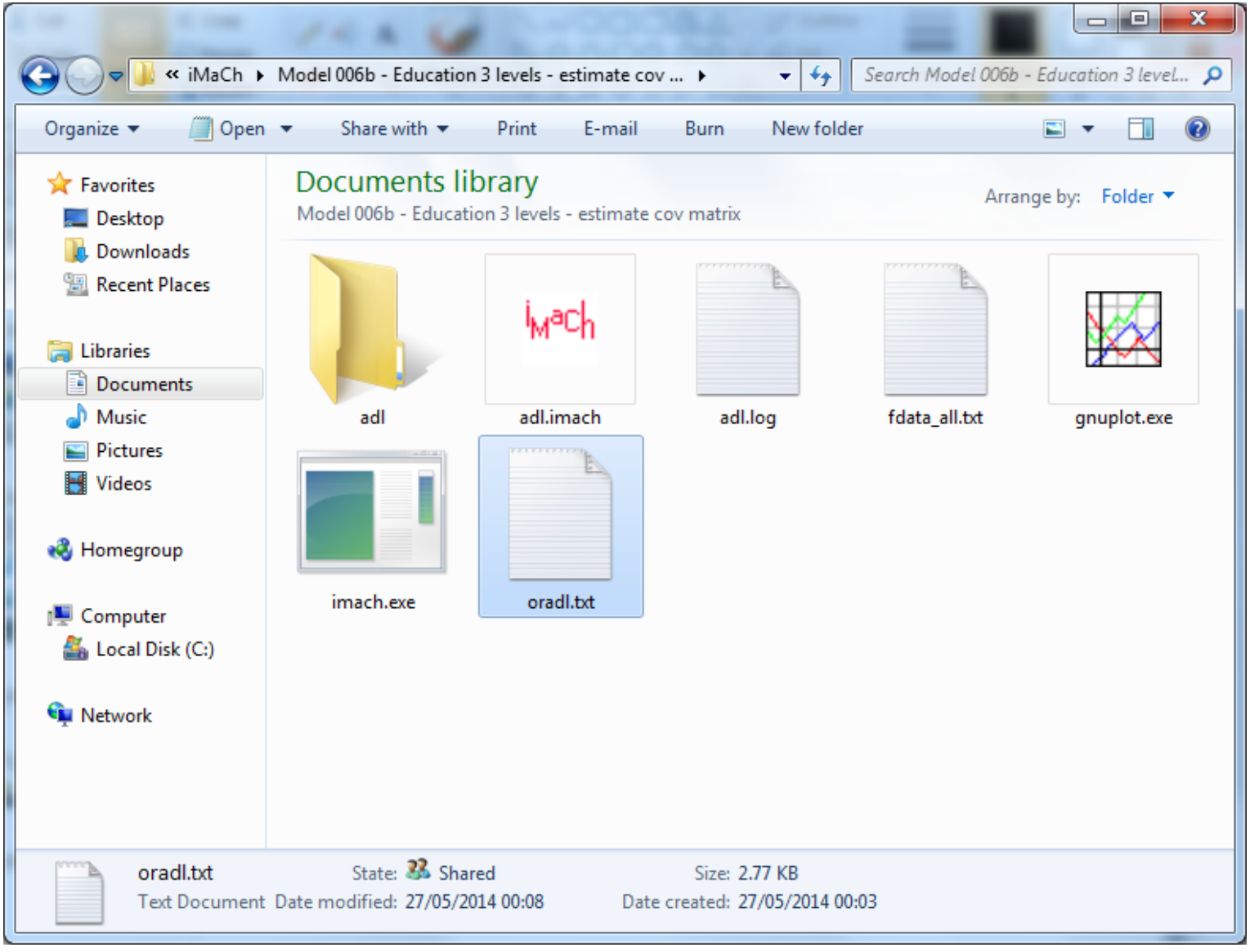


IMaCh covariates

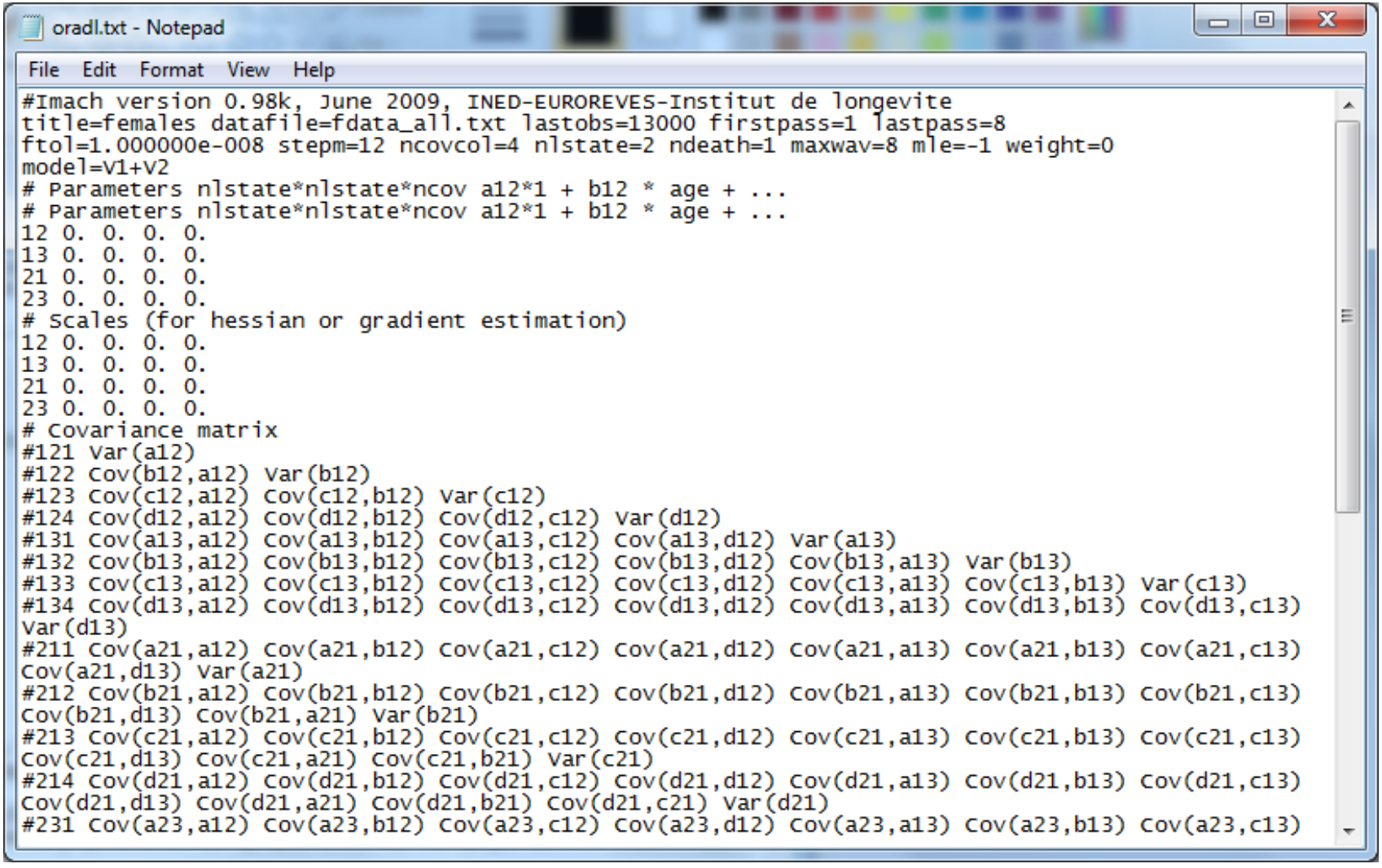
Set **mle=-1** and **model=V1+V2**



IMaCh covariates - oradl.txt



IMaCh covariates



IMaCh covariates

File Edit View History Bookmarks Tools Help

IMaCh adl.htm

file:///C:/Users/vikki/Documents/VIKKI/REVES 2014 - Presentation/ADL_F_Results/iMaCh/Mode

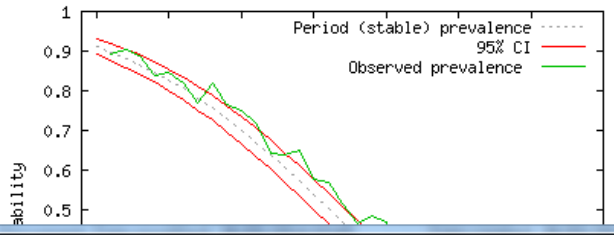
Most Visited Latest Headlines RoundCube Webmail :... SPACE REVES IMaCh ELECT Documentation - IMa... Google Scholar

- **Result files (second order: variances)**

  - Parameter file with estimated parameters and covariance matrix: radl.imach
  - Variance of one-step probabilities: adl/probradl.txt
  - Variance-covariance of one-step probabilities: adl/probcovradl.txt
  - Correlation matrix of one-step probabilities: adl/probcorradl.txt
  - Variances and covariances of health expectancies by age and **initial health status** ($cov(e^{ij},e^{kl})$(estepm= 1 months): adl/cveradl.txt
  - (a) Health expectancies by health status at initial age (e^{ij}) and standard errors (in parentheses) (b) life expectancies and standard errors ($e^{i.}=e^{i1}+e^{i2}+...$)(estepm= 1 months): adl/stderadl.txt
  - Variances and covariances of health expectancies by age. Status (i) based health expectancies (in state j), e^{ij} are weighted by the period prevalences in each state i (if popbased=1, an additional computation is done using the cross-sectional prevalences, i.e population based) (estepm=1 months): adl/vradl.txt
  - Total life expectancy and total health expectancies to be spent in each health state $e^{.j}$ with their standard errors (if popbased=1, an additional computation is done using the cross-sectional prevalences, i.e population based) (estepm=1 months): adl/tradl.txt
  - Standard deviation of period (stable) prevalences: adl/vplradl.txt
    - **Graphs**

********** Results for covariates V1=0 V2=0 **********

- Observed (cross-sectional) and period (incidence based) prevalence (with 95% confidence interval) in state (1): adl/vadl11.png



ELECT

- The web-page for downloading ELECT is:
  www.ucl.ac.uk/~ucakadl/indexELECT.html
- It is written in R and requires another R package, msm, to run.
- The data need to be arranged long format

ELECT data

Longitudinal panel data format means one row per observation.

An example for individual is given by:

id	state	age	education
5	1	66	9
5	1	67	9
5	1	69	9
5	2	73	9
5	2	74	9
5	3	78	9

Another example of data for an individual is:

id	state	age	education
7	1	81	11
7	2	82	11
7	1	84	11
7	-2	86	11

Note the last state is right-censored (denoted by the value -2). Right censoring here implies the person is still alive but in an unknown state.

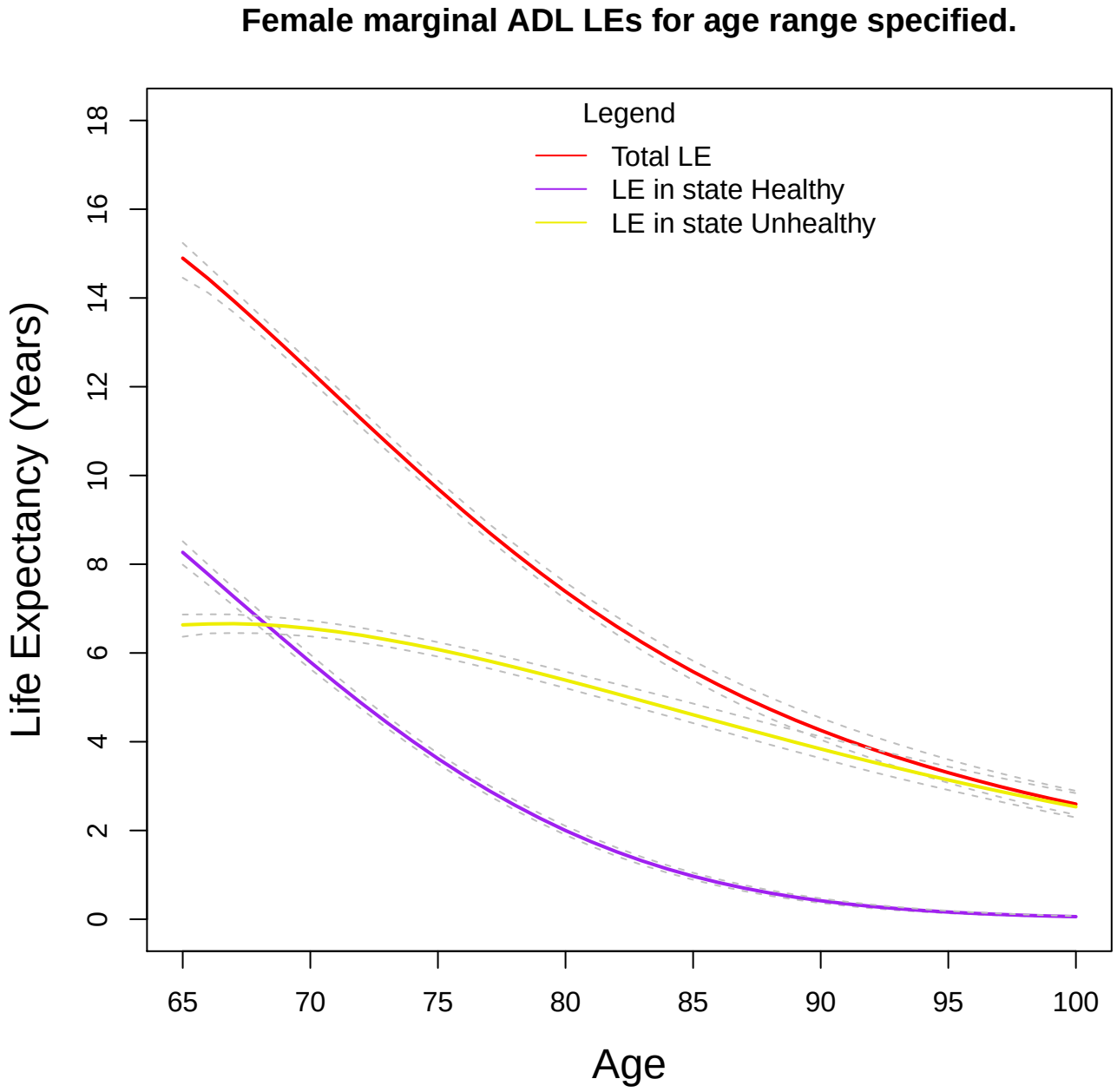
ELECT - msm model

```
# Fit the msm model:
model<-msm(state~age, subject=id, data=data,
           center=FALSE, qmatrix=Q, death=TRUE,
           covariates=~age+ybirth, censor=-2,
           censor.states=c(1,2), method="BFGS",
           control=list(trace=1, REPORT=1,
           maxit=1000, fnscale=100000))

# Fit the ELECT model:
LEs.pnt<-elect(model=model,
               b.covariates=list(age=0,ybirth=20),
               statedistdata=sddata,
               time.scale.msm="years",
               h=0.5, age.max=40, S=0)
```

ELECT - msm model

```
# Fit the msm model:
model<-msm(state~age, subject=id, data=data,
           center=FALSE, qmatrix=Q, death=TRUE,
           covariates=~age+ybirth, censor=-2,
           censor.states=c(1,2), method="BFGS",
           control=list(trace=1, REPORT=1,
           maxit=1000, fnscale=100000))

# Fit the ELECT model:
LEs.pnt<-elect(model=model,
               b.covariates=list(age=0,ybirth=20),
               statedistdata=sddata,
               time.scale.msm="years",
               h=0.5, age.max=40, S=0)
```

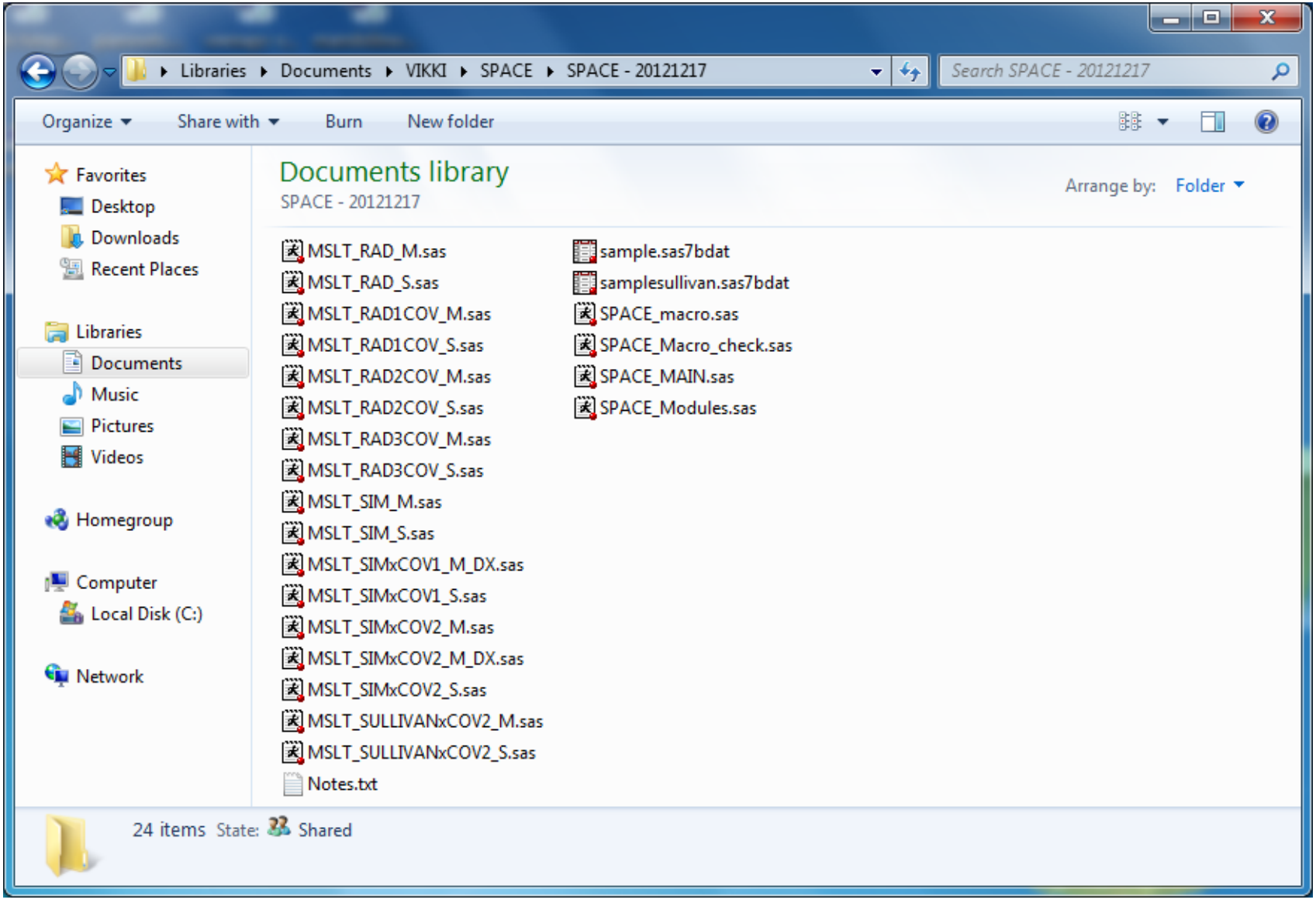
ELECT results



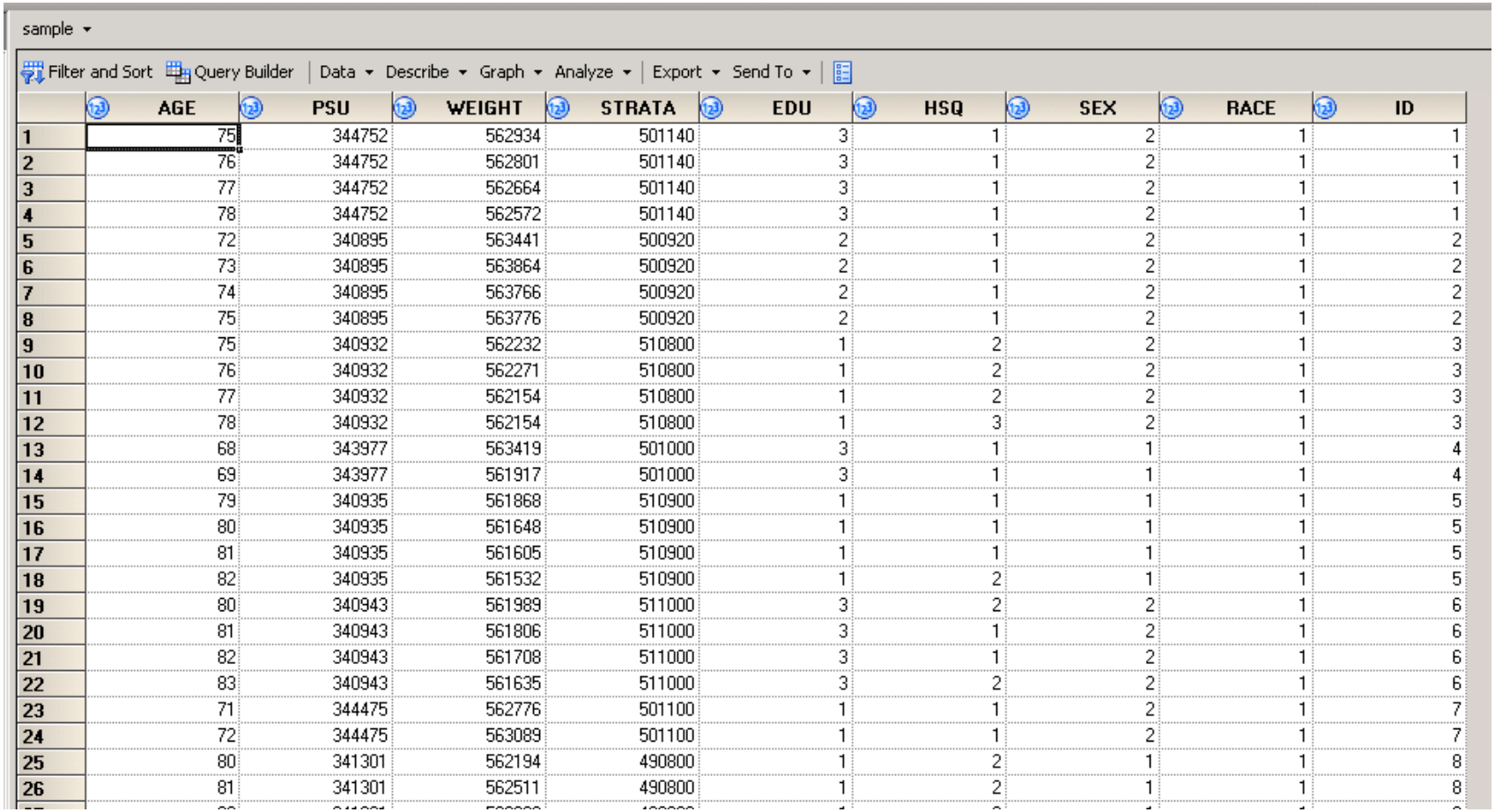
SPACE

- The web-page for downloading SPACE is:
  http://www.cdc.gov/nchs/data_access/space.htm
- SPACE is made up of many different SAS macro files
- However, a wrapper is available from its authors
- Designed for complex population sampling
- Can do semi-Markov Models (dependent on time in state)
- Data - Multiple lines, one per time point

SPACE files


Libraries ▸ Documents ▸ VIKKI ▸ SPACE ▸ SPACE - 20121217
Search SPACE - 20121217
Organize
Share with
Burn
New folder
Favorites
Desktop
Downloads
Recent Places
Libraries
Documents
Music
Pictures
Videos
Homegroup
Computer
Local Disk (C:)
Network
Documents library
SPACE - 20121217
Arrange by: Folder
MSLT_RAD_M.sas
MSLT_RAD_S.sas
MSLT_RAD1COV_M.sas
MSLT_RAD1COV_S.sas
MSLT_RAD2COV_M.sas
MSLT_RAD2COV_S.sas
MSLT_RAD3COV_M.sas
MSLT_RAD3COV_S.sas
MSLT_SIM_M.sas
MSLT_SIM_S.sas
MSLT_SIMxCOV1_M_DX.sas
MSLT_SIMxCOV1_S.sas
MSLT_SIMxCOV2_M.sas
MSLT_SIMxCOV2_M_DX.sas
MSLT_SIMxCOV2_S.sas
MSLT_SULLIVANxCOV2_M.sas
MSLT_SULLIVANxCOV2_S.sas
Notes.txt
sample.sas7bdat
samplesullivan.sas7bdat
SPACE_macro.sas
SPACE_Macro_check.sas
SPACE_MAIN.sas
SPACE_Modules.sas
24 items State: Shared

SPACE data

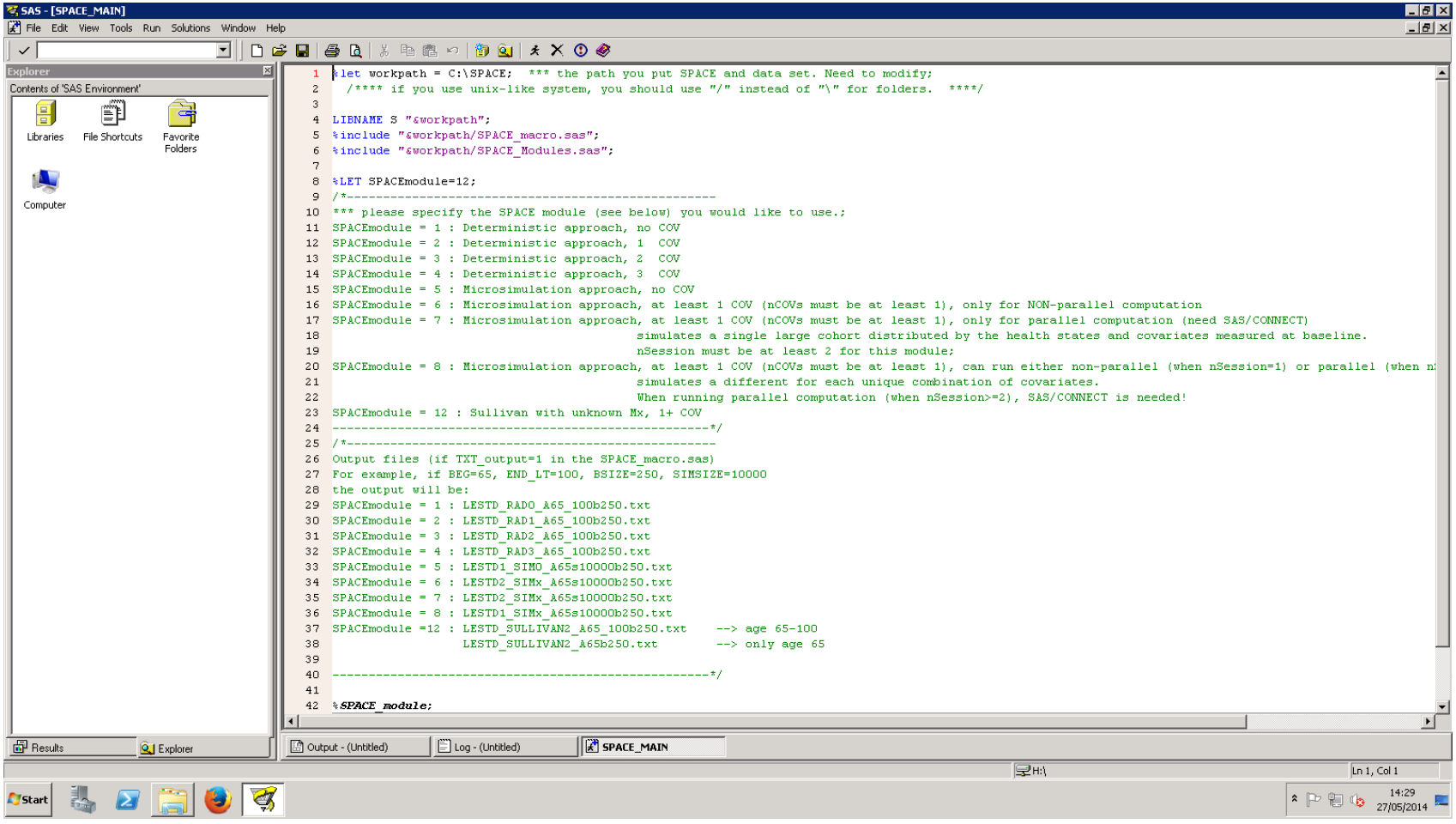


sample

Filter and Sort | Query Builder | Data | Describe | Graph | Analyze | Export | Send To

	AGE	PSU	WEIGHT	STRATA	EDU	HSQ	SEX	RACE	ID
1	75	344752	562934	501140	3	1	2	1	1
2	76	344752	562801	501140	3	1	2	1	1
3	77	344752	562664	501140	3	1	2	1	1
4	78	344752	562572	501140	3	1	2	1	1
5	72	340895	563441	500920	2	1	2	1	2
6	73	340895	563864	500920	2	1	2	1	2
7	74	340895	563766	500920	2	1	2	1	2
8	75	340895	563776	500920	2	1	2	1	2
9	75	340932	562232	510800	1	2	2	1	3
10	76	340932	562271	510800	1	2	2	1	3
11	77	340932	562154	510800	1	2	2	1	3
12	78	340932	562154	510800	1	3	2	1	3
13	68	343977	563419	501000	3	1	1	1	4
14	69	343977	561917	501000	3	1	1	1	4
15	79	340935	561868	510900	1	1	1	1	5
16	80	340935	561648	510900	1	1	1	1	5
17	81	340935	561605	510900	1	1	1	1	5
18	82	340935	561532	510900	1	2	1	1	5
19	80	340943	561989	511000	3	2	2	1	6
20	81	340943	561806	511000	3	1	2	1	6
21	82	340943	561708	511000	3	1	2	1	6
22	83	340943	561635	511000	3	2	2	1	6
23	71	344475	562776	501100	1	1	2	1	7
24	72	344475	563089	501100	1	1	2	1	7
25	80	341301	562194	490800	1	2	1	1	8
26	81	341301	562511	490800	1	2	1	1	8

SPACE syntax - MAIN wrapper



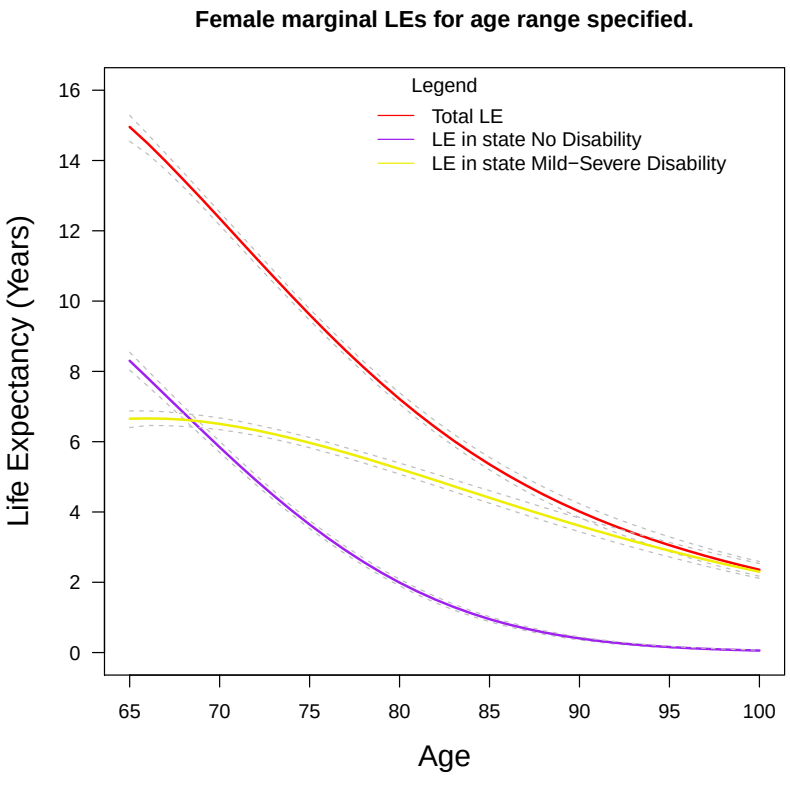
Disability free life expectancy

- Classifications of Disability:
  - State 1: No Disability
  - State 2: Mild to Severe Disability
  - State 3: Death

Disability

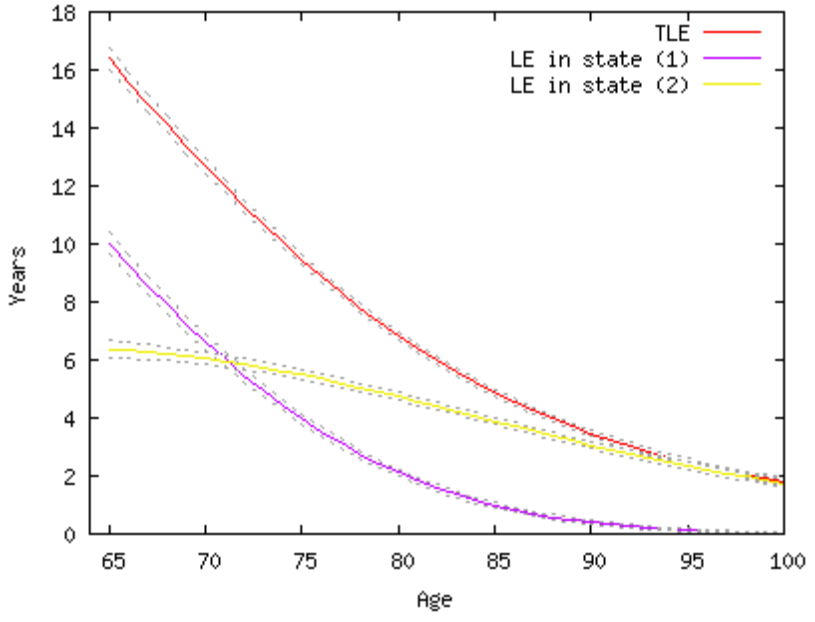
ELECT:

Estimation of Life Expectancies using Continuous-Time multi-state models



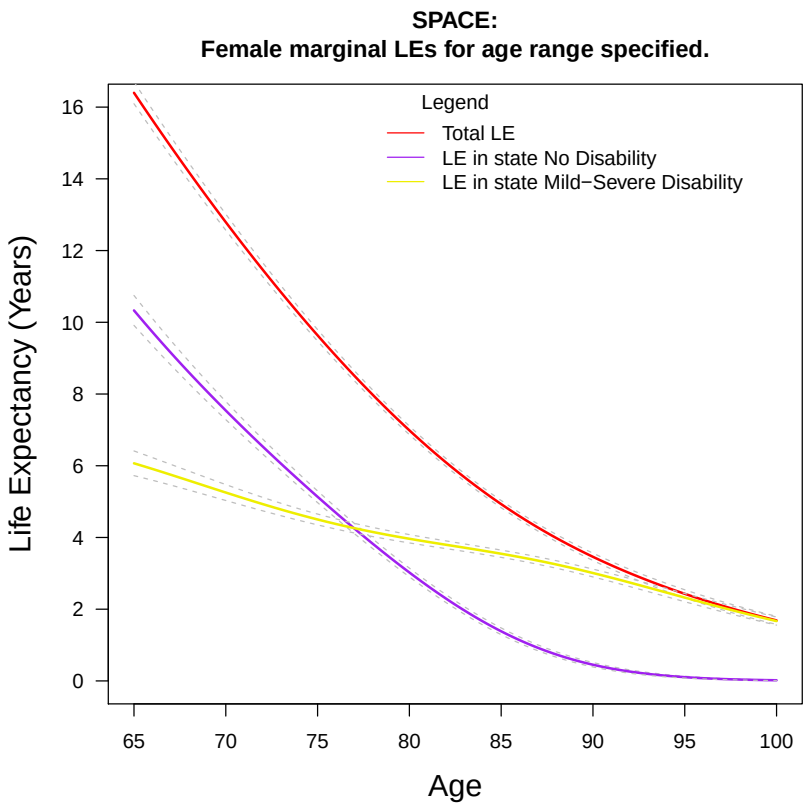
IMaCh:

A maximum likelihood computer program using Interpolation of Markov Chains



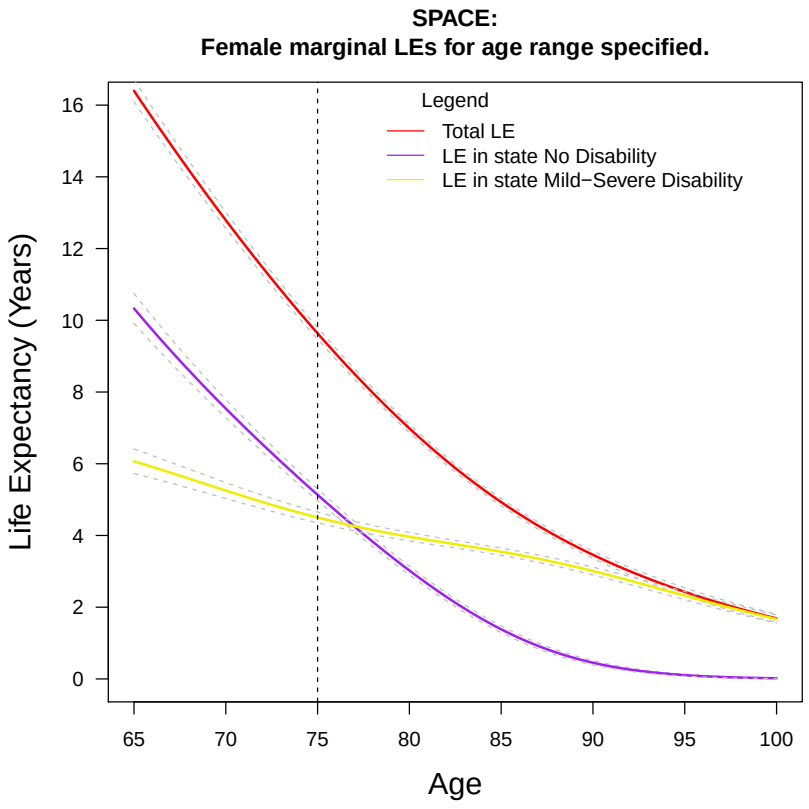
Disability

SPACE:
Stochastic Population Analysis for Complex Events



Disability

SPACE:
Stochastic Population Analysis for Complex Events



Disability: Comparison

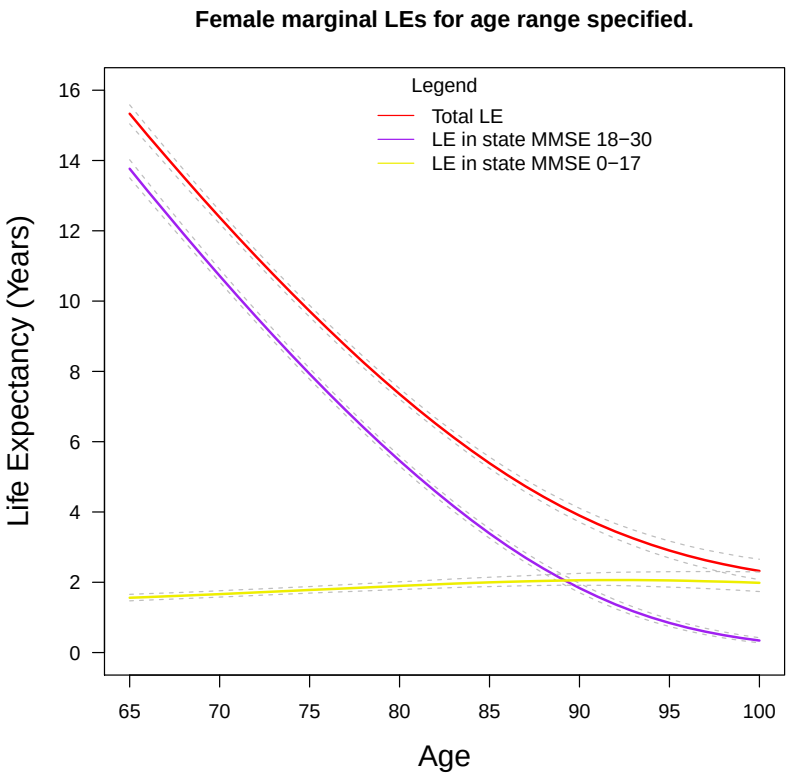
Cognitive Impairment

- Classifications of Cognitive Impairment:
  - State 1: MMSE $18 - 30$
  - State 2: MMSE $0 - 17$
  - State 3: Death

Cognitive Impairment

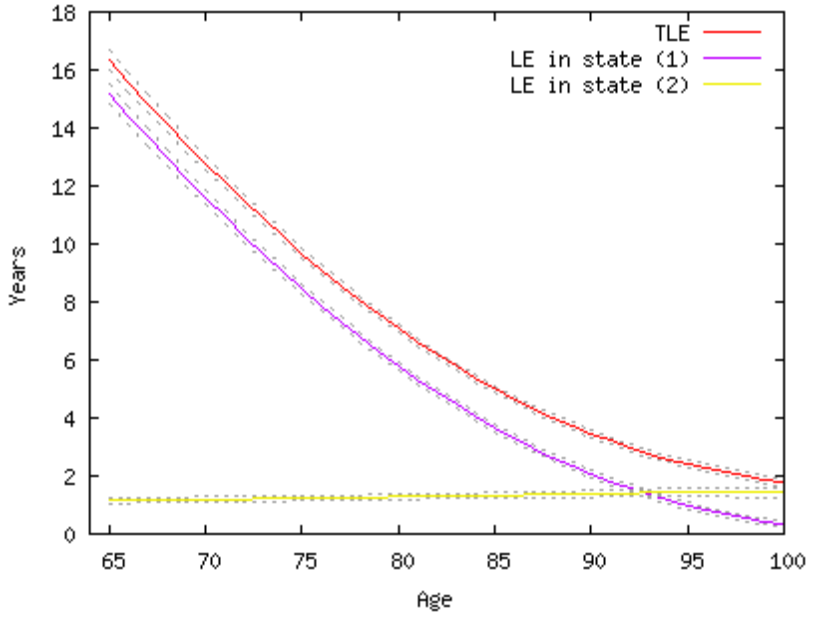
ELECT:

Estimation of Life Expectancies using Continuous-Time multi-state models



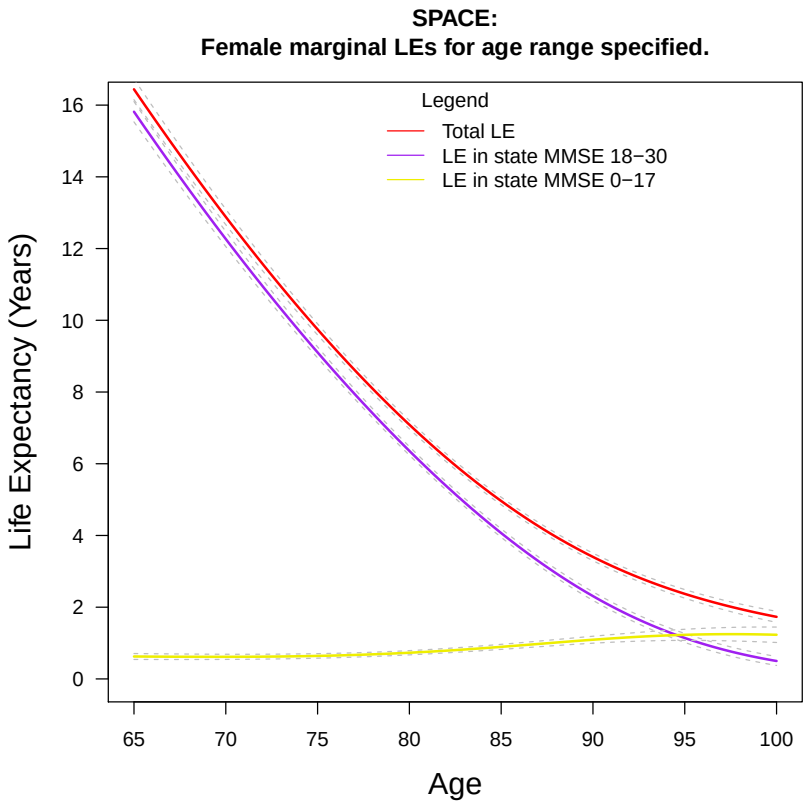
IMaCh:

A maximum likelihood computer program using Interpolation of Markov Chains



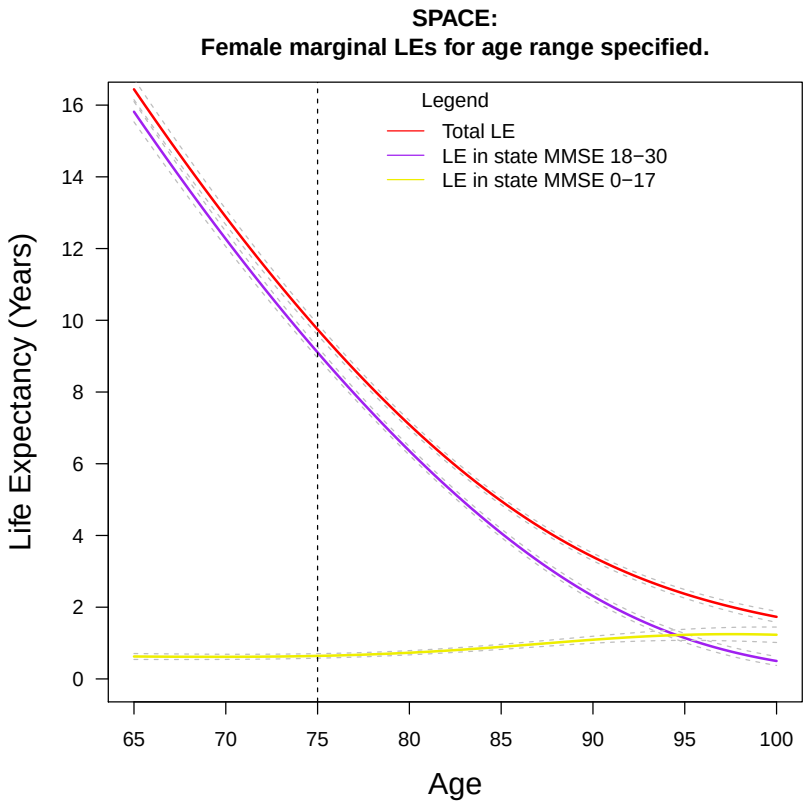
Cognitive Impairment

SPACE:
Stochastic Population Analysis for Complex Events

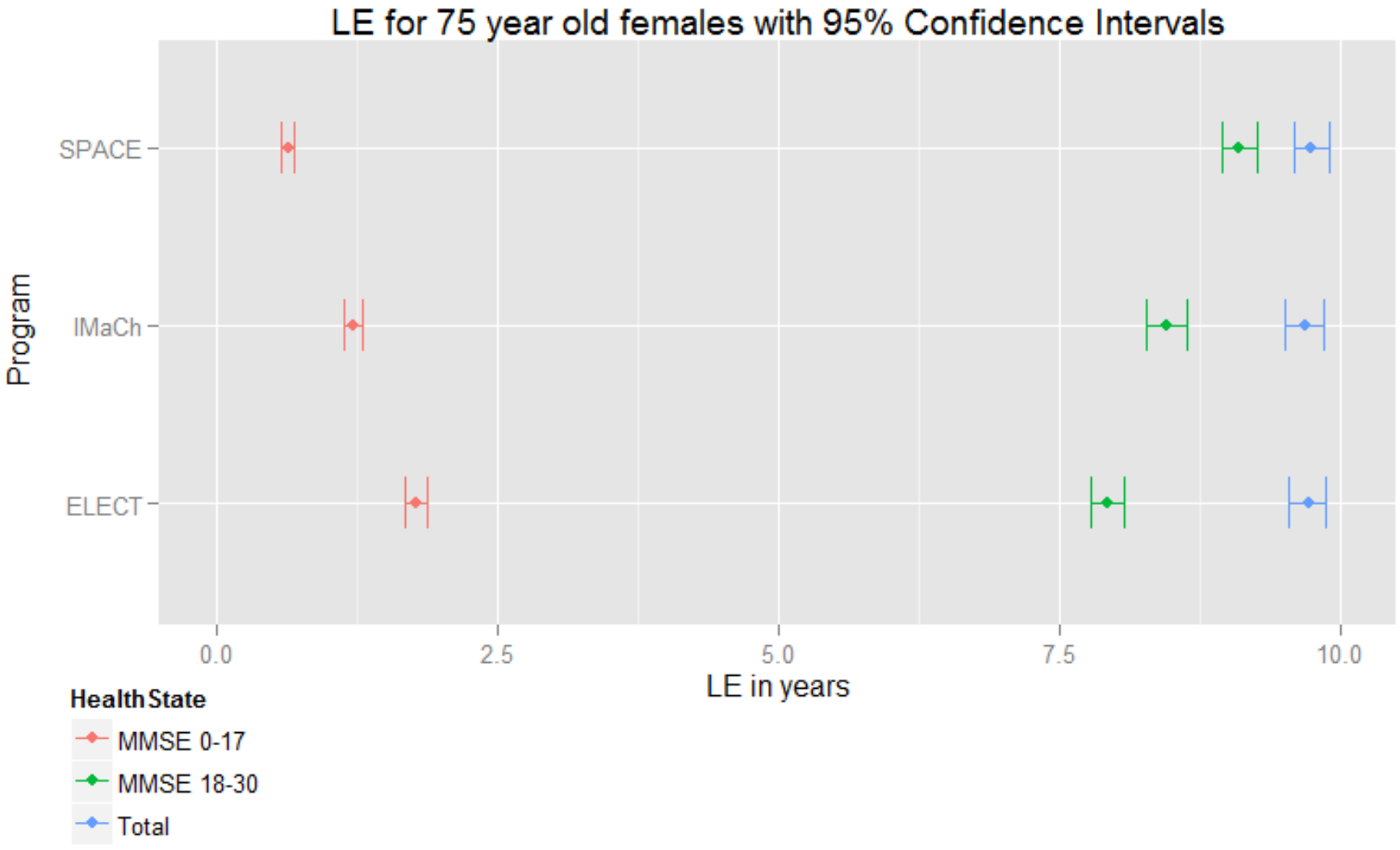


Cognitive Impairment

SPACE:
Stochastic Population Analysis for Complex Events



Cognitive Impairment: Comparison

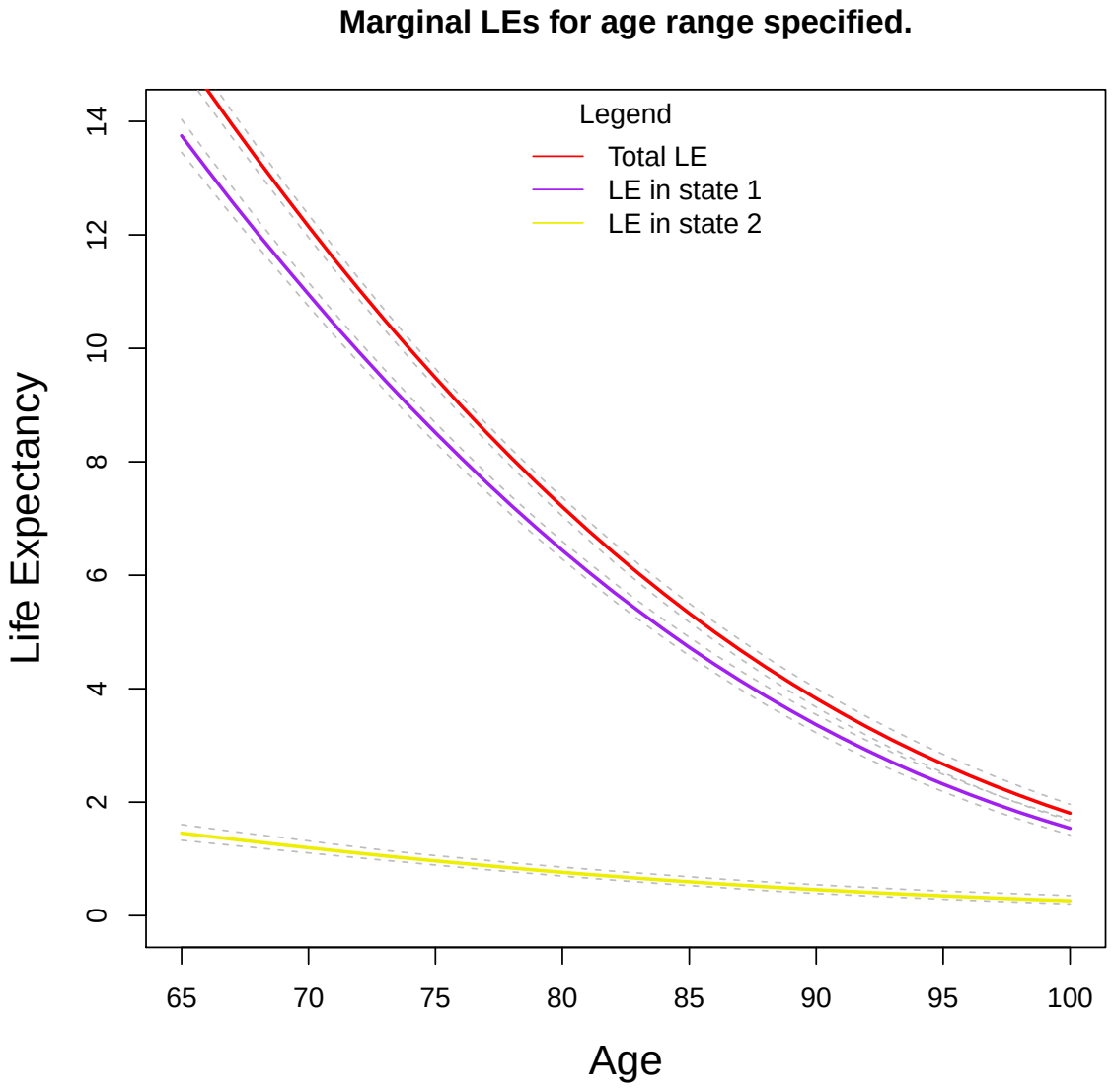


Stroke: iMaCh

- iMaCh cannot handle non recoverable conditions
- If you try to run a model on data with no recovery you will get an error message:

```
Error reading data around ‘12’ at line number 2
for individual 1, ‘0002 1 0 1 04/1899 05/1995
12/1990’ Should be a covariate value (=0 for the
reference or 1 for alternative).  Exiting.
Type q for exiting:
```

Stroke: ELECT results



Summary

Different strokes for different folks.

	Software				
	IMaCh	SPACE	ELECT	GLSMT	LxPct 2
Common disorder	★★	★★	★★	★	★
No recovery	–	★	★★	★	★
Rare recovery	★	★	★★	★	★
Uneven observations	★★	★	★★	★	★
Right censored data	★★	–	★★	–	★
Two waves, lots of covariates	★	★★	★★	★★	★
Flexibility	★★	★	★	★★	★
Useability	★★	★	★	★	★★

Questions

Questions?