Supplemental material for: Dean, D.O., Bauer, D.J., & Shanahan, M.J. A discrete-time multiple event process survival mixture (MEPSUM) model. *Psychological Methods*.

Example data (numbers substituted for confidentiality) for application in paper:

Id	Parent Age	Work Age	Marriage Age	College Age	Age at Survey
1	25	19	21	22	28
2	21	18	20	?	30
3	?	25	?	24	29
4	24	?	?	?	26
5	18	24	24	?	28
6	?	?	26	?	32
7	?	26	?	24	27

The key to the analysis is restructuring the data from the type of format above to a "wide" format with an indicator for each age (in this example, the ages studied were 18 to 30) for each event process. A zero is inserted for each age at which the person was eligible to experience the event but did not yet experience it. A one is inserted for the age at which the event occurred, if applicable. "999" is inserted as a missing indicator, for all time periods after event occurrence and for time periods where the data is "censored" (for example, if the person was age 26 at the time of the survey and had not yet been married, their data is censored for the marriage process after age 26).

For example, the data would be re-coded for the parent process like this:

	Parent - Age												
ID	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0	0	0	0	0	0	0	1	999	999	999	999	999
2	0	0	0	1	999	999	999	999	999	999	999	999	999
3	0	0	0	0	0	0	0	0	0	0	0	0	999
4	0	0	0	0	0	0	1	999	999	999	999	999	999
5	1	999	999	999	999	999	999	999	999	999	999	999	999
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	999	999	999

And for the work process like this:

	Work - Age												
ID	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0	1	999	999	999	999	999	999	999	999	999	999	999
2	1	999	999	999	999	999	999	999	999	999	999	999	999
3	0	0	0	0	0	0	0	1	999	999	999	999	999
4	0	0	0	0	0	0	0	0	0	999	999	999	999
5	0	0	0	0	0	0	1	999	999	999	999	999	999
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	1	999	999	999	999

The marriage and college processes would be recoded the same way. Additional variables (e.g. predictors, sample weights) can be added to the data file also (no special restructuring needed).

Danielle O. Dean, September 2013

Example Code to reformat data in R (http://cran.us.r-project.org/):

For this example, “Raw data.csv” looks like this:

Id	Par	Work	Marr	Col	Age	Female
1	25	19	21	22	28	1
2	21	18	20	NA	30	0
3	NA	25	NA	24	29	0
4	24	NA	NA	NA	26	0
5	18	24	24	NA	28	1
6	NA	NA	26	NA	32	1
7	NA	26	NA	24	27	1

```
##########################################
############## Read in raw data ##########
##########################################

setwd("C:/Users/dadean/Research/MEPSUM/Example Code + Data") #working directory - substitute location
raw=read.csv("Raw data.csv",header=T) #raw data - substitute name of file
print(raw[1:5,]) #viewing raw data after it has been read in (to check to make sure it's ok)

##########################################
############## Set-up packages ###########
##########################################

install.packages("reshape2") #do not need this line after first time running code
library("reshape2") #special reshaping package that makes it easy to reformat

##########################################
### Reformat event process data ##########
##########################################

#the column names in the code below must match the column names in the raw data
process=subset(raw,select=c(Id,Par,Work,Marr,Col,Age)) #select out the event process variables only
covariates=subset(raw,select=c(Id,Female)) #save the Id and other variables for re-merging later

#make the event process data into "long" format to help reshaping
process_long=melt(process,id=c("Id","Age"))
print(process_long[1:5,])

#create a new variable for each time period for each event process for each person (expanded data file)
expanded=NULL
for (time in 18:30){ #loop over all ages interested in (substitute time periods here)
  process_long$time=time
  expanded=rbind(expanded,process_long)
}
expanded$ProcTime=paste(expanded$variable,expanded$time,sep="") #new column name (process+event time)
print(expanded[1:5,])

# get the event indicator (0/1/999) for each time period where 999 = missing indicator
expanded$event=ifelse(is.na(expanded$value)==TRUE,
                      ifelse(expanded$time<=expanded$Age,0,999),
                        ifelse(expanded$time==expanded$value,1,
                             ifelse(expanded$Age<expanded$time,999,
                                    ifelse(expanded$time<expanded$value,0,999))))

expanded=subset(expanded,select=-c(Age,time,value,variable)) #remove extra columns no longeer needed
expanded=expanded[order(expanded$ProcTime),] #order dataset so column names will be in correct order
expanded_wide=reshape(expanded,idvar="Id",timevar="ProcTime",direction="wide") #reshape into wide format

##########################################
### Get other variables & output #########
##########################################

final=merge(expanded_wide,covariates,by="Id") #merge event indicators with the covariates saved earlier
print(final[1:5,]) #see order of column names and first 5 rows of data
write.table(final,"final.csv",row.names=F,col.names=F,sep=",") #neither row/column names will print
#this “final.csv” can be used for reading into Mplus
```

Danielle O. Dean, September 2013

Example Mplus syntax with unstructured hazard functions:

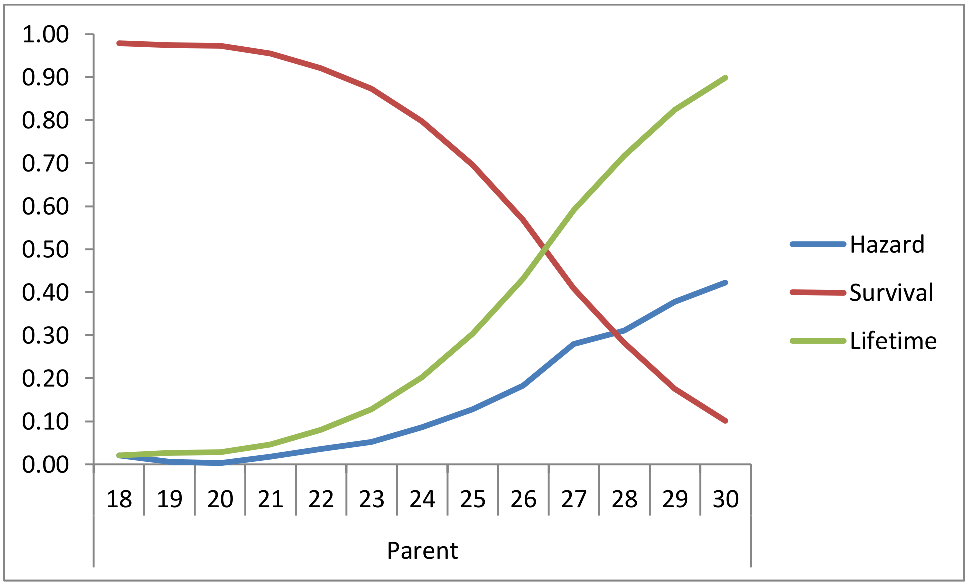
Comments after the explanation points

```
title: adult MEPSUM C1
data: file is 'C:\Users\deand\adult\adultwide.txt';

                                                             !variable names given here only (not in data file)
variable: names are aid female psuscid region weight par18-par30
work18-work30 marr18-marr30 col18-col30 black hispanic orace par_nohs par_col;
         ! female = dummy-coded indicator for females (v. males)
         ! psuscid, region, weight are used to account for sample design
         ! black, Hispanic, orace are dummy-coded indicators for African-America,Hispanic, & other race (v. Caucausians)
         ! par_nohs = dummy-coded indicator for parent had no high school degree
         ! par_col = dummy-coded indicator for parent had college degree (v. parent had high school degree)

                                                             !use variables tells Mplus which variables are used
usevariables = female black hispanic orace par_nohs par_col
psuscid region weight par18-par30
work18-work30 marr18-marr30 col18-col30;

categorical are par18-par30
work18-work30 marr18-marr30 col18-col30;

classes=c(1);                                                !1-class model. To fit a 2-class model, “classes=c(2)” instead, etc.
missing=all(999);                                            !missing indicator. We used “999” to indicate missing
weight=weight;                                               ! sample weights
stratification=region;                                       ! region & psuscid are used to account for sample design
cluster=psuscid;
                  ! for an unweighted analysis, delete the “weight”, “stratification” and “cluster” lines from the code

analysis: type=mixture complex;
starts=150 100;                                              !might need to increase in order to ensure global solution
stiterations=20;

model:
%overall%
c on female black hispanic orace par_nohs par_col;           !predictors of latent class membership
                                                             !remove the above line for an unconditional model

                                                             !with unstructured hazard functions, no other syntax needed
                                                             !may want to include start values

output:
tech11 tech14;
```

The probabilities that result give the hazard in probability scale (can also examine the parameters in logit scale, which are needed for example to get predicted probabilities depending on levels of covariates).

The hazard function in probability scale can then be used to compute the survival function and lifetime distribution function, as described in the paper. For example, the hazard functions within a given latent class for the parent process might look like the left column, and the middle and right columns are calculated based on this hazard function. These functions are given for each event process within each latent class.

Age	Hazard	Survival	Lifetime
18	0.02	0.98	0.02
19	0.01	0.97	0.03
20	0.00	0.97	0.03
21	0.02	0.95	0.05
22	0.04	0.92	0.08
23	0.05	0.87	0.13
24	0.09	0.80	0.20
25	0.13	0.70	0.30
26	0.18	0.57	0.43
27	0.28	0.41	0.59
28	0.31	0.28	0.72
29	0.38	0.18	0.82
30	0.42	0.10	0.90


Parent
Hazard
Survival
Lifetime