

Analysis of eGFR trajectories from Hong Kong Diabetes Registry

January 23, 2017

Contents

1	Description of data	1
1.1	Data overview	1
1.2	Sample selection	10
2	eGFR Trajectory Analysis	14
2.1	Modeling of trajectory	14
2.2	Plots of trajectory	27
2.3	Post-model analyses	30

Chapter 1

Description of data

1.1 Data overview

```
> rm(list=ls())
> start <- Sys.time()
> set.seed(1983)
> library(lcmm)
> library(Epi)
> library(lme4)
> library(nlme)
> library(splines)
> print(sessionInfo(), l=F)
```

```
R version 3.2.3 (2015-12-10)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 16.04.1 LTS
```

```
attached base packages:
```

```
[1] splines  stats      graphics  grDevices  utils      datasets  methods
[8] base
```

```
other attached packages:
```

```
[1] nlme_3.1-124    lme4_1.1-12     Matrix_1.2-3    Epi_2.0
[5] lcmm_1.7.5      survival_2.38-3
```

```
loaded via a namespace (and not attached):
```

```
[1] Rcpp_0.12.1      lattice_0.20-33 MASS_7.3-45      grid_3.2.3
[5] plyr_1.8.3       etm_0.6-2        minqa_1.2.4     nloptr_1.0.4
[9] tools_3.2.3      cmprsk_2.2-7     parallel_3.2.3
```

The data are comprised by two parts: the baseline data and the follow-up eGFR data. The baseline data, which has been extracted from the Hong Kong Diabetes

Registry(HKDR), including the information of clinical assessments and laboratory investigations at enrollment, and the well-defined complication outcomes censored on 30th, June 2014. As we focus on the ESRD outcome in this analysis, we only select those Chinese patients with no history of ESRD which is defined according to the ICD-9 codes and eGFR <15 . Therefore, we obtain a cohort consisted of 1218 ESRD events and 8336 event-free patients at cersoring date. Then we further discarded 8 patients with abnormal baseline eGFR (>300). The follow-up data include all the creatinine records from enrollment to 2014, and the eGFR is corresponding calculated using the Chinese-modified MDRD formula.

```
> path <- "/home/gzjiang/Work_Space/eGFR_Traj/"
> base_dat <- read.table(paste(path, "data/ESRD1_Prosp2014_CH-T2D_1218vs8336.csv", sep=""), hea
> follow_dat <- read.table(paste(path, "data/eGFR_19940714_20140630.csv", sep=""), header=TRUE
> gwas_id <- read.table(paste(path, "data/TBGWA_ID_6445.csv", sep=""), header=TRUE, sep=",")
> ckd_hist <- FALSE #Whether includes those with CKD history at baseline: no CKD hist AN
> cut_label <- TRUE #Whether cut the follow-up eGFR data based on quantile
> quant <- 3 #number of quantiles for each follow-up year
> egfr_grp <- c(2,3) #build the model using the sub-data of baseline eGFR groups
```

```
> dim(base_dat)
```

```
[1] 9554 112
```

```
> head(base_dat)
```

	Obs_id	date	ETHNICOD	SEX_str	SEX	DOB	AGE	YEAR_DIA	AGE_ONSE	DMAGE
1	1	2002-10-22	C	F	0	7/8/1940	62	2002	62	0
2	2	1996-10-04	C	M	1	13/3/1939	57	1995	56	1
3	3	1996-08-02	C	M	1	5/3/1935	61	1983	48	13
4	4	2001-03-16	C	M	1	1927	74	1980	53	21
5	5	1997-05-20	C	M	1	12/7/1924	73	1993	69	4
6	6	1999-03-23	C	F	0	7/5/1922	77	1991	69	8
	IDDM_UPD	FX	DAD_AF	MOM_AF	SIBLING1	SIBLING2	CHILD_A1	CHILD_A2	SMOKING_str	
1	N	0	0	0	0	4	0	3	N	
2	N	0	0	0	0	3	0	0	E	
3	N	1	0	0	1	4	0	0	C	
4	N	0	0	0	0	0	0	5	E	
5	N	1	0	0	0	3	1	5	N	
6	N	2	0	0	2	4	0	4	N	
	SMOKING	SMOKING_C	SMOKING_E	ALCOHOL_str	ALCOHOL	ALCOHOL_C	ALCOHOL_E	HEIGHT		
1	0	0	0	N	0	0	0	1.450		
2	1	0	1	C	2	1	1	1.785		

3	2	1	1	E	1	0	1	1.590					
4	1	0	1	E	1	0	1	1.550					
5	0	0	0	N	0	0	0	1.535					
6	0	0	0	N	0	0	0	1.445					
	WT	WAIST	HIP	SBP	DBP	BMI	WHR	HBA1C	FBG	TC	HDL	LDL	TG
1	55.5	81.0	93.5	131.0	73.5	26.39715	0.8663102	6.8	5.6	5.40	1.10	3.73	1.26
2	68.3	87.5	94.0	140.0	90.0	21.43603	0.9308510	6.3	5.0	6.20	1.04	4.60	1.19
3	46.4	69.0	82.0	140.0	68.0	18.35370	0.8414634	9.8	5.3	4.60	1.56	2.80	0.53
4	52.3	81.0	87.0	128.5	55.5	21.76899	0.9310345	6.2	9.3	6.09	1.11	4.58	0.88
5	62.6	86.5	92.5	125.0	70.0	26.56792	0.9351351	6.9	7.2	4.40	1.25	2.70	0.87
6	51.4	83.0	96.0	197.5	95.0	24.61656	0.8645833	5.6	8.7	4.20	1.22	2.60	0.74
	ACR	RBC	HB	WBC	eGFR_BASE	FORM_D_A	FORM_DM	HOME_BLO	HOME_MON	HOME_FRE			
1	0.17	4.45	13.8	8.4	72.61830	Y	Y	1		1		3	
2	138.01	4.40	13.8	9.3	80.16108	N	N	4		0		NA	
3	0.77	4.51	14.0	5.4	141.57826	N	N	2		1		5	
4	9.43	3.85	11.9	6.9	57.68408	Y	Y	1		1		3	
5	0.63	5.00	15.5	4.5	100.98319	Y	Y	1		1		3	
6	1.12	NA	NA	NA	101.95334	Y	Y	1		1		3	
	ED_CHIRO	RETINO_B	NEURO_B	MICROALB_B	MACROALB_B	DEATH_HIST	LLDs_Base						
1		N	0	0	0	0	0		0		0		
2		N	1	1	0	1	0		0		0		
3		N	0	0	0	0	0		0		0		
4		N	1	0	1	0	0		0		0		
5		N	0	1	0	0	0		0		0		
6		N	0	0	0	0	0		0		1		
	Hypert_Base	Oads_Base	Insulin_Base	ACEIARB_Base	Ln_TG	Ln_ACR	Ln_eGFR						
1		1	0	0	1	0.2311	0.1570	4.2852					
2		0	1	0	0	0.1740	4.9345	4.3840					
3		0	1	0	0	-0.6349	0.5710	4.9529					
4		1	1	0	0	-0.1278	2.3447	4.0550					
5		1	1	0	1	-0.1393	0.4886	4.6150					
6		1	1	0	1	-0.3011	0.7514	4.6245					
	MI_HIST	MI_END	MI_DATE	MI_TIME	IHD_HIST	IHD_END	IHD_DATE	IHD_TIME					
1	0	0	2014-06-30	11.687885	0	0	2014-06-30	11.6878850					
2	0	0	2014-06-30	17.735797	0	0	2014-06-30	17.7357974					
3	0	0	2014-06-30	17.908282	0	0	2014-06-30	17.9082820					
4	0	0	2004-08-25	3.444216	0	0	2004-08-25	3.4442163					
5	0	0	2014-06-30	17.111567	0	0	2014-06-30	17.1115674					
6	0	1	2010-07-20	11.326489	0	1	1999-06-02	0.1943874					
	CHD1_HIST	CHD1_END	CHD1_DATE	CHD1_TIME	STK_HIST	STK_END	STK_DATE						
1	0	0	2014-06-30	11.6878850	0	0	2014-06-30						
2	0	0	2014-06-30	17.7357974	0	0	2014-06-30						
3	0	0	2014-06-30	17.9082820	0	0	2014-06-30						
4	0	0	2004-08-25	3.4442163	0	0	2004-08-25						
5	0	0	2014-06-30	17.1115674	0	0	2014-06-30						
6	1	1	1999-06-02	0.1943874	0	0	2010-07-26						
	STK_TIME	PVD_HIST	PVD_END	PVD_DATE	PVD_TIME	CVD_HIST	CVD_END	CVD_DATE					
1	11.687885	0	0	2014-06-30	11.6878850	0	0	2014-06-30					

```

2 17.735797      0      0 2014-06-30 17.7357974      0      0 2014-06-30
3 17.908282      0      0 2014-06-30 17.9082820      0      0 2014-06-30
4  3.444216      1      1 2001-05-25  0.1916496      1      1 2001-05-25
5 17.111567      0      0 2014-06-30 17.1115674      0      0 2014-06-30
6 11.342916      0      0 2010-07-26 11.3429158      1      1 1999-06-02

```

```

      CVD_TIME CHF_HIST CHF_END   CHF_DATE   CHF_TIME HYPO_HIST HYPO_END

```

```

1 11.6878850      0      0 2014-06-30 11.6878850      0      0
2 17.7357974      0      1 1998-04-20  1.5414100      0      0
3 17.9082820      0      0 2014-06-30 17.9082820      0      0
4  0.1916496      0      1 2001-09-29  0.5393566      0      1
5 17.1115674      0      0 2014-06-30 17.1115674      0      0
6  0.1943874      0      0 2010-07-26 11.3429158      0      0

```

```

      HYPO_DATE HYPO_TIME CAN_HIST CAN_END   CAN_DATE   CAN_TIME CKD_HIST CKD_END

```

```

1 2014-06-30 11.687885      0      0 2014-06-30 11.687885      0      0
2 2014-06-30 17.735797      0      0 2014-06-30 17.735797      0      0
3 2014-06-30 17.908282      0      0 2014-06-30 17.908282      0      0
4 2002-06-18  1.256674      0      0 2004-08-25  3.444216      1      1
5 2014-06-30 17.111567      0      0 2014-06-30 17.111567      0      0
6 2010-07-26 11.342916      0      0 2010-07-26 11.342916      0      1

```

```

      CKD_DATE   CKD_TIME ESRD1_HIST ESRD1_END ESRD1_DATE ESRD1_TIME DEATH_HIST.1

```

```

1 2014-06-30 11.687885      0      0 2014-06-30  11.687885      0
2 2014-06-30 17.735797      0      0 2014-06-30  17.735797      0
3 2014-06-30 17.908282      0      0 2014-06-30  17.908282      0
4 2001-05-11  0.1533196      0      0 2004-08-25   3.444216      0
5 2014-06-30 17.111567      0      0 2014-06-30  17.111567      0
6 2010-07-21 11.3292266      0      0 2010-07-26  11.342916      0

```

```

      DEATH_END DEATH_DATE DEATH_TIME

```

```

1      0 2014-06-30  11.687885
2      0 2014-06-30  17.735797
3      0 2014-06-30  17.908282
4      1 2004-08-25   3.444216
5      0 2014-06-30  17.111567
6      1 2010-07-26  11.342916

```

```
> dim(follow_dat)
```

```
[1] 391551      4
```

```
> head(follow_dat)
```

```

      Obs_id test_date F_eGFR creatinine
1          1 2002-08-20 80.6474          84
2          1 2002-08-24 97.9131          71
3          1 2002-08-31 73.5245          91
4          1 2002-10-08 91.8750          75
5          1 2002-10-22 72.5693          92
6          1 2003-12-01 81.4311          83

```

```
> dim(gwas_id)
```

```
[1] 6445    3
```

```
> head(gwas_id)
```

```
      FID      IID Obs_id
1 CX007533 CX007533  3339
2 CX001628 CX001628  7114
3 CX001660 CX001660  8198
4 CX007614 CX007614  1337
5 CX001725 CX001725  5420
6 CX001749 CX001749  7976
```

As we divide the subjects into different groups according to their baseline eGFR, we require all subjects have normal baseline eGFR measures. We define the survival outcome of informative censoring: ESRD or Death event, whichever came first.

```
> length(which(base_dat$eGFR_BASE > 300))
```

```
[1] 8
```

```
> base_dat <- subset(base_dat, eGFR_BASE <= 300) #remove those patients with abnormal b
> base_dat <- transform(base_dat, CENSOR_END=as.numeric(ESRD1_END | DEATH_END), CENSOR_DATE=pmin(as.Da
+ CENSOR_TIME=pmin(ESRD1_TIME, DEATH_TIME), doin=cal.yr(date), dob=cal.yr(DOB, '
+ base_egfr_grp=as.numeric(cut(eGFR_BASE,c(60, 90, Inf, right=F))))
> with(base_dat, tapply(eGFR_BASE, base_egfr_grp, range)) #Three groups: 1 for <60; 2 for >=60-90; 3 for >90
```

```
$`1`
[1] 15.06820 59.99865
```

```
$`2`
[1] 60.06306 89.95094
```

```
$`3`
[1] 90.02061 296.19463
```

```
> table(base_dat$base_egfr_grp)
```

```

      1      2      3
978 2170 6398

```

```
> length(which(base_dat$eGFR_BASE<30))
```

```
[1] 163
```

```

> if(!ckd_hist) {                                     #no CKD hist AND in specific g
+   base_dat <- subset(base_dat, (CKD_HIST==0) & (base_egfr_grp %in% egfr_grp))
+ }else {                                             #with CKD OR in specific group
+   base_dat <- subset(base_dat, (CKD_HIST==1) | (base_egfr_grp %in% egfr_grp))
+ }
> dim(base_dat)

```

```
[1] 8235  119
```

```
> table(base_dat$CKD_HIST)
```

```

      0
8235

```

```

> base_dat <- transform(base_dat, dodm = pmin(YEAR_DIA + runif(length(YEAR_DIA)), doin))
> dob_na <- which(is.na(base_dat$dob))                #missing values of date of birth
> base_dat$dob[dob_na] <- (floor(base_dat$doin[dob_na]) - base_dat$AGE[dob_na]) + runif(length(dob_na))
> base_subdat <- subset(base_dat, select=c("Obs_id", "doin", "dob", "dodm", "dox", "base_egfr_g",
+                                          "SEX", "ESRD1_DATE", "ESRD1_END", "ESRD1_TIME", "DEATH_DATE",
+                                          "DEATH_END", "DEATH_TIME", "CENSOR_DATE", "CENSOR_END", "CENSOR_REASON"))
> dim(base_subdat)

```

```
[1] 8235  17
```

```
> head(base_subdat)
```

	Obs_id	doin	dob	dodm	dox	base_egfr_grp	AGE	SEX	ESRD1_DATE	
1	1	2002.805	1940.598	2002.153	2014.493		2	62	0	2014-06-30
2	2	1996.757	1939.194	1995.217	2014.493		2	57	1	2014-06-30
3	3	1996.585	1935.172	1983.747	2014.493		3	61	1	2014-06-30
5	5	1997.381	1924.527	1993.908	2014.493		3	73	1	2014-06-30
6	6	1999.221	1922.345	1991.221	2010.564		3	77	0	2010-07-26
7	7	1998.375	1927.624	1994.408	2014.493		2	71	1	2014-06-30

	ESRD1_END	ESRD1_TIME	DEATH_DATE	DEATH_END	DEATH_TIME	CENSOR_DATE	CENSOR_END
1	0	11.68789	2014-06-30	0	11.68789	2014-06-30	0
2	0	17.73580	2014-06-30	0	17.73580	2014-06-30	0
3	0	17.90828	2014-06-30	0	17.90828	2014-06-30	0
5	0	17.11157	2014-06-30	0	17.11157	2014-06-30	0
6	0	11.34292	2010-07-26	1	11.34292	2010-07-26	1
7	0	16.11773	2014-06-30	0	16.11773	2014-06-30	0

	CENSOR_TIME
1	11.68789
2	17.73580
3	17.90828
5	17.11157
6	11.34292
7	16.11773

```
> table(base_subdat$ESRD1_END)
```

```

  0    1
7566 669
```

```
> follow_dat <- transform(follow_dat, dolab=cal.yr(test_date))
> follow_dat <- subset(follow_dat, select=-c(test_date))
```

We merge the baseline data and follow-up eGFR data according to the id of subject, and remove those records with eGFR >300, which are considered to be errors, and also remove those with only 1 eGFR measurement.

```
> merged_dat <- merge(base_subdat, follow_dat, by=intersect("Obs_id", "Obs_id"), sort=F)
> dim(merged_dat)
```

```
[1] 288614    20
```

```
> head(merged_dat)
```

	Obs_id	doin	dob	dodm	dox	base_egfr_grp	AGE	SEX	ESRD1_DATE
1	1	2002.805	1940.598	2002.153	2014.493		2	62	0 2014-06-30
2	1	2002.805	1940.598	2002.153	2014.493		2	62	0 2014-06-30
3	1	2002.805	1940.598	2002.153	2014.493		2	62	0 2014-06-30
4	1	2002.805	1940.598	2002.153	2014.493		2	62	0 2014-06-30
5	1	2002.805	1940.598	2002.153	2014.493		2	62	0 2014-06-30
6	1	2002.805	1940.598	2002.153	2014.493		2	62	0 2014-06-30
	ESRD1_END	ESRD1_TIME	DEATH_DATE	DEATH_END	DEATH_TIME	CENSOR_DATE	CENSOR_END		
1	0	11.68789	2014-06-30	0	11.68789	2014-06-30	0		
2	0	11.68789	2014-06-30	0	11.68789	2014-06-30	0		
3	0	11.68789	2014-06-30	0	11.68789	2014-06-30	0		
4	0	11.68789	2014-06-30	0	11.68789	2014-06-30	0		
5	0	11.68789	2014-06-30	0	11.68789	2014-06-30	0		
6	0	11.68789	2014-06-30	0	11.68789	2014-06-30	0		
	CENSOR_TIME	F_eGFR	creatinine	dolab					
1	11.68789	93.2876	73	2007.843					
2	11.68789	90.7499	75	2006.657					
3	11.68789	81.4311	83	2003.914					
4	11.68789	80.6474	84	2002.632					
5	11.68789	73.5245	91	2002.663					
6	11.68789	103.2903	66	2012.812					

```
> addmargins(table(table(merged_dat$Obs_id)))
```

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
271	61	67	72	89	137	179	173	154	174	152	147	146	162	151	172
17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
161	202	200	184	205	149	168	173	163	151	160	151	140	142	141	101
33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
115	115	109	111	97	91	94	110	101	100	98	90	77	105	61	71
49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64
71	84	60	61	59	45	40	45	49	48	49	39	43	40	36	27
65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
24	27	26	38	30	29	22	24	22	23	26	21	22	17	21	15
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96
16	18	15	17	15	19	18	14	15	15	9	12	11	11	13	9
97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112
8	10	16	2	7	8	9	13	7	6	11	9	9	6	4	8
113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128
5	7	10	5	7	4	5	10	6	6	5	2	4	4	3	4
129	130	131	132	133	134	135	137	138	139	140	141	142	143	144	145
3	5	4	5	1	4	5	9	4	3	3	2	1	3	2	1
146	147	148	149	150	151	154	155	158	159	161	162	163	164	165	166
1	3	2	5	4	2	3	4	3	3	2	1	1	3	2	1
167	169	171	172	173	175	176	178	179	182	183	186	187	191	193	194
2	1	1	1	4	1	4	3	2	1	1	1	3	2	1	1

195	196	198	199	200	208	209	211	213	216	218	221	223	224	225	228
2	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1
231	232	238	240	243	249	251	252	254	267	284	288	293	301	431	725
1	1	1	1	1	1	2	1	1	3	1	1	1	1	1	1
Sum															
8235															

```
> with(merged_dat, table((F_eGFR>300) + (F_eGFR>1000)))
```

0	1	2
288341	264	9

```
> merged_dat <- subset(merged_dat, F_eGFR <= 300) #remove those abnormal eGFR records
> dim(merged_dat)
```

```
[1] 288341    20
```

```
> addmargins(table(table(merged_dat$Obs_id)))
```

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
271	62	66	73	88	137	179	173	154	174	152	148	145	163	151	172
17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
162	202	200	186	203	149	169	172	165	151	161	149	140	143	141	103
33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
116	112	109	111	98	90	92	109	104	99	100	90	76	105	62	73
49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64
72	82	59	60	58	43	40	46	48	48	50	41	40	40	36	28
65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
25	26	26	40	30	28	22	23	21	24	26	22	20	19	21	14
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96
16	17	15	17	15	19	17	16	15	15	9	11	10	11	13	10
97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112
8	11	16	3	6	8	8	13	7	6	10	10	8	6	4	8
113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128
5	7	10	6	7	4	5	10	6	6	5	2	4	4	4	3
129	130	131	132	133	134	135	137	138	139	140	141	143	144	145	146
3	5	4	5	1	4	5	9	4	3	3	2	3	2	1	1
147	148	149	150	151	154	155	158	159	161	162	163	164	165	166	167
3	2	5	4	2	3	4	3	3	2	1	1	3	2	1	2
169	171	173	175	176	178	179	182	183	186	187	191	193	194	195	196
1	2	4	1	4	3	2	1	1	1	3	2	1	1	2	1
198	199	200	208	209	211	213	216	218	221	223	224	225	228	231	232
2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
238	240	243	249	251	252	254	267	284	288	293	301	431	725	Sum	
1	1	1	1	2	1	1	3	1	1	1	1	1	1	8235	

```
> merged_dat <- merged_dat[order(merged_dat$Obs_id, merged_dat$dolab),] #sort by id & lab data
```

1.2 Sample selection

We only select those records between enrollment and event/censoring dates, including the baseline eGFR. That is, those eGFR records before enrollment or after event/censoring dates need to be removed. Moreover, we obtain the rate of eGFR change during the first 1,2,3 years and the whole period using two methods: linear mixed model and (last eGFR - first eGFR)/first eGFR*100%. And also calculate the follow-up age(F_AGE), duration of diabetes(F_DMAGE), and the forward time gap between the baseline date and eGFR measurement date (FW_TIME).

```
> update_base <- base_dat
> merged_dat <- subset(merged_dat, dolab>=doin & dolab<=dox) #update the merged data
> merged_dat <- transform(merged_dat, FW_TIME=dolab-doin, F_AGE=dolab-dob, F_DMAGE=dolab-dodm,
> dim(merged_dat)
```

```
[1] 239823      24
```

```
> head(merged_dat)
```

	Obs_id	doin	dob	dodm	dox	base_egfr_grp	AGE	SEX	ESRD1_DATE
7	1	2002.805	1940.598	2002.153	2014.493	2	62	0	2014-06-30
3	1	2002.805	1940.598	2002.153	2014.493	2	62	0	2014-06-30
16	1	2002.805	1940.598	2002.153	2014.493	2	62	0	2014-06-30
10	1	2002.805	1940.598	2002.153	2014.493	2	62	0	2014-06-30
2	1	2002.805	1940.598	2002.153	2014.493	2	62	0	2014-06-30
13	1	2002.805	1940.598	2002.153	2014.493	2	62	0	2014-06-30
	ESRD1_END	ESRD1_TIME	DEATH_DATE	DEATH_END	DEATH_TIME	CENSOR_DATE	CENSOR_END		
7	0	11.68789	2014-06-30	0	11.68789	2014-06-30	0		
3	0	11.68789	2014-06-30	0	11.68789	2014-06-30	0		
16	0	11.68789	2014-06-30	0	11.68789	2014-06-30	0		
10	0	11.68789	2014-06-30	0	11.68789	2014-06-30	0		
2	0	11.68789	2014-06-30	0	11.68789	2014-06-30	0		
13	0	11.68789	2014-06-30	0	11.68789	2014-06-30	0		
	CENSOR_TIME	F_eGFR	creatinine	dolab	FW_TIME	F_AGE	F_DMAGE	log_eGFR	
7	11.68789	72.5693	92	2002.805	0.000000	62.20671	0.6523099	4.284542	
3	11.68789	81.4311	83	2003.914	1.108830	63.31554	1.7611395	4.399757	
16	11.68789	81.2613	83	2004.568	1.763176	63.96988	2.4154858	4.397670	
10	11.68789	96.8863	71	2005.951	3.145791	65.35250	3.7981005	4.573538	
2	11.68789	90.7499	75	2006.657	3.852156	66.05886	4.5044660	4.508107	
13	11.68789	87.7935	77	2007.558	4.752909	66.95962	5.4052189	4.474987	

```

> sub_merged <- merged_dat
> if(cut_label) {
+   grp <- with(sub_merged, as.numeric(cut(FW_TIME, c(-Inf, 0:(ceiling(max(FW_TIME))*quant)/quant),
+   sub_merged <- cbind(sub_merged, grp)
+   stat <- with(sub_merged, by(sub_merged, list(Obs_id, grp), function(x) {
+     data.frame(Obs_id=x$Obs_id[1], F_eGFR=median(x$F_eGFR), FW_TIME=(x$grp[1]-1)/quant) #mean of
+   }))
+   stat <- do.call("rbind", stat)
+   stat <- stat[order(stat$Obs_id, stat$FW_TIME),]
+   sub_merged <- merge(base_subdat, stat, by=intersect("Obs_id", "Obs_id"), sort=F) #merged with the
+   sub_merged <- transform(sub_merged, F_AGE=doin-dob+FW_TIME, F_DMAGE=doin-dodm+FW_TIME, log_eGFR=
+ }
> addmargins(table(table(sub_merged$Obs_id)))

```

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
295	161	174	179	257	326	339	355	287	340	296	306	330	321	323	325
17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
304	304	295	262	260	221	207	178	191	141	105	119	96	106	95	94
33	34	35	36	37	38	39	40	41	42	43	Sum				
77	104	100	93	81	64	45	43	23	10	3	8235				

```

> head(sub_merged)

```

	Obs_id	doin	dob	dodm	dox	base_egfr_grp	AGE	SEX	ESRD1_DATE
1	1	2002.805	1940.598	2002.153	2014.493		2	62	0 2014-06-30
2	1	2002.805	1940.598	2002.153	2014.493		2	62	0 2014-06-30
3	1	2002.805	1940.598	2002.153	2014.493		2	62	0 2014-06-30
4	1	2002.805	1940.598	2002.153	2014.493		2	62	0 2014-06-30
5	1	2002.805	1940.598	2002.153	2014.493		2	62	0 2014-06-30
6	1	2002.805	1940.598	2002.153	2014.493		2	62	0 2014-06-30
	ESRD1_END	ESRD1_TIME	DEATH_DATE	DEATH_END	DEATH_TIME	CENSOR_DATE	CENSOR_END		
1	0	11.68789	2014-06-30		0	11.68789	2014-06-30		0
2	0	11.68789	2014-06-30		0	11.68789	2014-06-30		0
3	0	11.68789	2014-06-30		0	11.68789	2014-06-30		0
4	0	11.68789	2014-06-30		0	11.68789	2014-06-30		0
5	0	11.68789	2014-06-30		0	11.68789	2014-06-30		0
6	0	11.68789	2014-06-30		0	11.68789	2014-06-30		0
	CENSOR_TIME	F_eGFR	FW_TIME	F_AGE	F_DMAGE	log_eGFR			
1	11.68789	72.5693	0.000000	62.20671	0.6523099	4.284542			
2	11.68789	81.4311	1.333333	63.54004	1.9856432	4.399757			
3	11.68789	81.2613	2.000000	64.20671	2.6523099	4.397670			
4	11.68789	96.8863	3.333333	65.54004	3.9856432	4.573538			
5	11.68789	90.7499	4.000000	66.20671	4.6523099	4.508107			
6	11.68789	87.7935	5.000000	67.20671	5.6523099	4.474987			

```
> str(sub_merged)
```

```
'data.frame':      132735 obs. of  22 variables:
 $ Obs_id      : int   1 1 1 1 1 1 1 1 1 1 ...
 $ doin        : num   2003 2003 2003 2003 2003 ...
 $ dob         : num   1941 1941 1941 1941 1941 ...
 $ dodm        : num   2002 2002 2002 2002 2002 ...
 $ dox         : num   2014 2014 2014 2014 2014 ...
 $ base_egfr_grp: num    2 2 2 2 2 2 2 2 2 2 ...
 $ AGE         : int   62 62 62 62 62 62 62 62 62 ...
 $ SEX         : int    0 0 0 0 0 0 0 0 0 0 ...
 $ ESRD1_DATE  : Factor w/ 2162 levels "1995-10-04","1996-01-05",...: 2162 2162 2162 2162 2162 ...
 $ ESRD1_END   : int    0 0 0 0 0 0 0 0 0 0 ...
 $ ESRD1_TIME  : num   11.7 11.7 11.7 11.7 11.7 ...
 $ DEATH_DATE  : Factor w/ 1960 levels "1995-10-04","1996-01-05",...: 1960 1960 1960 1960 1960 ...
 $ DEATH_END   : int    0 0 0 0 0 0 0 0 0 0 ...
 $ DEATH_TIME  : num   11.7 11.7 11.7 11.7 11.7 ...
 $ CENSOR_DATE : Date, format: "2014-06-30" "2014-06-30" ...
 $ CENSOR_END  : num    0 0 0 0 0 0 0 0 0 0 ...
 $ CENSOR_TIME : num   11.7 11.7 11.7 11.7 11.7 ...
 $ F_eGFR      : num   72.6 81.4 81.3 96.9 90.7 ...
 $ FW_TIME     : num    0 1.33 2 3.33 4 ...
 $ F_AGE       : num   62.2 63.5 64.2 65.5 66.2 ...
 $ F_DMAGE     : num    0.652 1.986 2.652 3.986 4.652 ...
 $ log_eGFR    : num    4.28 4.4 4.4 4.57 4.51 ...
```

```
> range(sub_merged$FW_TIME)
```

```
[1] 0 20
```

We extract a complete-case cohort by removing those records with missing values on follow-up age/dmage. We then remove those with only 1 eGFR measurement again.

```
> sub_nomiss <- sub_merged[complete.cases(sub_merged), ]
> dim(sub_nomiss)
```

```
[1] 131247      22
```

```
> num_test <- data.frame(table(sub_nomiss$Obs_id))           #final number of eGFR measurements af
> names(num_test) <- c("Obs_id", "cut_num_eGFR")
> update_base <- merge(update_base, num_test, by.x="Obs_id", by.y="Obs_id", all.x=T, sort=F)
> dim(update_base)
```

```
[1] 8235 147
```

```
> id_keep <- subset(num_test, cut_num_eGFR>1)$Obs_id           #removed those patients with only
> sub_nomiss <- subset(sub_nomiss, Obs_id %in% id_keep)
> dim(sub_nomiss)
```

```
[1] 130963      22
```

```
> addmargins(table(table(sub_nomiss$Obs_id)))
```

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
156	172	173	252	322	335	347	284	336	294	302	324	314	321	319	301	
18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	
301	291	258	254	217	205	176	188	139	105	119	94	105	95	94	77	
34	35	36	37	38	39	40	41	42	43	Sum						
103	100	93	81	64	44	43	23	10	3	7834						

Chapter 2

eGFR Trajectory Analysis

2.1 Modeling of trajectory

We fit the joint model using `lcmm` package. We may only focus on the sub-group of baseline eGFR.

```
> uni_id <- unique(sub_nomiss$Obs_id)
> table(subset(base_dat, Obs_id %in% uni_id)$ESRD1_END)      #number of ESRD events and non-ev
```

```
      0      1
7173  661
```

```
> (kn <- seq(0, 20, , 5))      #build cubic spline data with 5 knots
```

```
[1]  0  5 10 15 20
```

```
> MM <- Ns(sub_nomiss$FW_TIME, knots=kn)
> dim(MM)
```

```
[1] 130963      4
```

```
> MM <- detrend(MM, sub_nomiss$FW_TIME)
> dim(MM)
```

```
[1] 130963      3
```

```
> head(MM)
```

```
      1      2      3
[1,] -0.09278708 0.24736837 -0.31475232
[2,] -0.12311989 0.12545566 -0.11089451
[3,] -0.13236037 0.06750430 -0.01798058
[4,] -0.12713762 -0.03637846 0.13178736
[5,] -0.10793365 -0.07990586 0.18142940
[6,] -0.05172029 -0.13129886 0.21419816
```

```
> (colnames(MM) <- paste("x", colnames(MM), sep=""))
```

```
[1] "x1" "x2" "x3"
```

```
> sub_nomiss <- cbind(sub_nomiss, MM)
```

```
> dim(sub_nomiss)
```

```
[1] 130963    25
```

```
> head(sub_nomiss)
```

```
  Obs_id   doin   dob   dodm   dox base_egfr_grp AGE SEX ESRD1_DATE
1      1 2002.805 1940.598 2002.153 2014.493      2  62  0 2014-06-30
2      1 2002.805 1940.598 2002.153 2014.493      2  62  0 2014-06-30
3      1 2002.805 1940.598 2002.153 2014.493      2  62  0 2014-06-30
4      1 2002.805 1940.598 2002.153 2014.493      2  62  0 2014-06-30
5      1 2002.805 1940.598 2002.153 2014.493      2  62  0 2014-06-30
6      1 2002.805 1940.598 2002.153 2014.493      2  62  0 2014-06-30
  ESRD1_END ESRD1_TIME DEATH_DATE DEATH_END DEATH_TIME CENSOR_DATE CENSOR_END
1         0   11.68789 2014-06-30         0   11.68789 2014-06-30         0
2         0   11.68789 2014-06-30         0   11.68789 2014-06-30         0
3         0   11.68789 2014-06-30         0   11.68789 2014-06-30         0
4         0   11.68789 2014-06-30         0   11.68789 2014-06-30         0
5         0   11.68789 2014-06-30         0   11.68789 2014-06-30         0
6         0   11.68789 2014-06-30         0   11.68789 2014-06-30         0
  CENSOR_TIME F_eGFR FW_TIME  F_AGE  F_DMAGE log_eGFR      x1
1   11.68789 72.5693 0.000000 62.20671 0.6523099 4.284542 -0.09278708
2   11.68789 81.4311 1.333333 63.54004 1.9856432 4.399757 -0.12311989
3   11.68789 81.2613 2.000000 64.20671 2.6523099 4.397670 -0.13236037
4   11.68789 96.8863 3.333333 65.54004 3.9856432 4.573538 -0.12713762
5   11.68789 90.7499 4.000000 66.20671 4.6523099 4.508107 -0.10793365
```

```

6      11.68789 87.7935 5.000000 67.20671 5.6523099 4.474987 -0.05172029
      x2      x3
1  0.24736837 -0.31475232
2  0.12545566 -0.11089451
3  0.06750430 -0.01798058
4 -0.03637846  0.13178736
5 -0.07990586  0.18142940
6 -0.13129886  0.21419816

> jlcm_c1 <- Jointlcmm(log_eGFR ~ FW_TIME + x1 + x2 + x3 + F_AGE + SEX + F_DMAGE,
+                      random = ~ FW_TIME,
+                      survival = Surv(CENSOR_TIME, CENSOR_END) ~ F_AGE + SEX + F_DMAGE,
+                      maxiter=100,
+                      subject = "Obs_id", ng=1, data=sub_nomiss)

Be patient, Jointlcmm is running ...
The program took 709.19 seconds

> jlcm_c1

Joint latent class model for quantitative outcome and competing risks
fitted by maximum likelihood method

Jointlcmm(fixed = log_eGFR ~ FW_TIME + x1 + x2 + x3 + F_AGE +
  SEX + F_DMAGE, random = ~FW_TIME, subject = "Obs_id", ng = 1,
  survival = Surv(CENSOR_TIME, CENSOR_END) ~ F_AGE + SEX +
  F_DMAGE, data = sub_nomiss, maxiter = 100)

Statistical Model:
  Dataset: sub_nomiss
  Number of subjects: 7834
  Number of observations: 130963
  Number of latent classes: 1
  Number of parameters: 17
  Event 1 :
    Number of events: 1806
    Weibull baseline risk function

Iteration process:
  Convergence criteria satisfied
  Number of iterations: 17
  Convergence criteria: parameters= 2e-08
                      : likelihood= 4.6e-06
                      : second derivatives= 3.2e-12

```

```

Goodness-of-fit statistics:
  maximum log-likelihood: 34390.12
  AIC: -68746.25
  BIC: -68627.82

> summary(jlcm_c1)

Joint latent class model for quantitative outcome and competing risks
  fitted by maximum likelihood method

Jointlcm(fixed = log_eGFR ~ FW_TIME + x1 + x2 + x3 + F_AGE +
  SEX + F_DMAGE, random = ~FW_TIME, subject = "Obs_id", ng = 1,
  survival = Surv(CENSOR_TIME, CENSOR_END) ~ F_AGE + SEX +
  F_DMAGE, data = sub_nomiss, maxiter = 100)

Statistical Model:
  Dataset: sub_nomiss
  Number of subjects: 7834
  Number of observations: 130963
  Number of latent classes: 1
  Number of parameters: 17
  Event 1:
    Number of events: 1806
    Weibull baseline risk function

Iteration process:
  Convergence criteria satisfied
  Number of iterations: 17
  Convergence criteria: parameters= 2e-08
                      : likelihood= 4.6e-06
                      : second derivatives= 3.2e-12

Goodness-of-fit statistics:
  maximum log-likelihood: 34390.12
  AIC: -68746.25
  BIC: -68627.82
  Score test statistic for CI assumption: 607.203 (p-value=0)

Maximum Likelihood Estimates:

Parameters in the proportional hazard model:

              coef      Se    Wald p-value
event1 +/-sqrt(Weibull1) 0.14458 0.00276 52.452 0.00000
event1 +/-sqrt(Weibull2) 1.26008 0.01338 94.200 0.00000

```

F_AGE	0.00215	0.00020	10.845	0.00000
SEX	0.32210	0.04739	6.797	0.00000
F_DMAGE	0.05397	0.00309	17.460	0.00000

Fixed effects in the longitudinal model:

	coef	Se	Wald	p-value
intercept	4.84230	0.00754	642.315	0.00000
FW_TIME	-0.01689	0.00082	-20.664	0.00000
x1	0.03862	0.00416	9.282	0.00000
x2	0.03885	0.00729	5.327	0.00000
x3	-0.02152	0.00666	-3.230	0.00124
F_AGE	-0.00257	0.00011	-22.890	0.00000
SEX	-0.00928	0.00535	-1.734	0.08291
F_DMAGE	-0.00618	0.00043	-14.449	0.00000

Variance-covariance matrix of the random-effects:

```

intercept FW_TIME
intercept  0.04753
FW_TIME    0.00031 0.00341

```

	coef	Se
Residual standard error	0.14179	0.00030

```
> #joint model with 3 class. We initialized the parameters based on 2 classes model
> #jlc_m_c3 <- gridsearch(rep=20, maxiter=50, minit=jlc_m_c1, m=Jointlcm(m~log_eGFR ~ FW_TIME + x1
> #                               mixture =~ FW_TIME + x1 + x2 + x3,
> #                               random =~ FW_TIME,
> #                               survival = Surv(CENSOR_TIME, CENSOR_END) ~ F_AGE + SEX + F_DMAGE,
> #                               subject = "Obs_id", ng=3, data=sub_nomiss))
> init_c3 <- c("intercept_class1"=-2.1, "intercept_class2"=-2.4, "weib_class1_1"=0.3, "weib_cla
+           "weib_class2_1"=0.3, "weib_class2_2"=1.1,"weib_class3_1"=1.5, "weib_class3_2"=1.
+           "sex"=0.22, "f_dmage"=0.034, "intercept_class1"=5.7, "intercept_class2"=4.4, "in
+           "fw_time1"=-0.34, "fw_time2"=-0.19, "fw_time3"=-0.0042, "x1_class1"=7.1, "x1 cla
+           "x1_class3"=2.5, "x2_class1"=8.2, "x2_class2"=1.5, "x2_class3"=0.025, "x3_class1
+           "x3_class3"=-0.071, "fage2"=-0.0035,"sex2"=-0.0012, "fdmage2"=-0.01,
+           "varcov1"=0.093, "varcov2"=-0.00013, "varcov3"=0.0019, "stderr"=0.13)
> jlc_m_c3 <- Jointlcm(m~log_eGFR ~ FW_TIME + x1 + x2 + x3 + F_AGE + SEX + F_DMAGE, #
+           mixture =~ FW_TIME + x1 + x2 + x3,
+           random =~ FW_TIME,
+           survival = Surv(CENSOR_TIME, CENSOR_END) ~ F_AGE + SEX + F_DMAGE,
+           maxiter=100, B=init_c3,
+           subject = "Obs_id", ng=3, data=sub_nomiss)
```

```
Be patient, Jointlcm is running ...
The program took 5271.05 seconds
```

```
> jlcm_c3
```

```
Joint latent class model for quantitative outcome and competing risks
  fitted by maximum likelihood method
```

```
Jointlcmmm(fixed = log_eGFR ~ FW_TIME + x1 + x2 + x3 + F_AGE +
  SEX + F_DMAGE, mixture = ~FW_TIME + x1 + x2 + x3, random = ~FW_TIME,
  subject = "Obs_id", ng = 3, survival = Surv(CENSOR_TIME,
    CENSOR_END) ~ F_AGE + SEX + F_DMAGE, data = sub_nomiss,
  maxiter = 100)
```

```
Statistical Model:
```

```
Dataset: sub_nomiss
Number of subjects: 7834
Number of observations: 130963
Number of latent classes: 3
Number of parameters: 33
Event 1 :
  Number of events: 1806
  Class-specific hazards and
  Weibull baseline risk function
```

```
Iteration process:
```

```
Convergence criteria satisfied
Number of iterations: 34
Convergence criteria: parameters= 9.5e-14
                      : likelihood= 5e-10
                      : second derivatives= 1.5e-11
```

```
Goodness-of-fit statistics:
```

```
maximum log-likelihood: 44744.8
AIC: -89423.6
BIC: -89193.72
```

```
> postprob(jlcm_c3)
```

```
Posterior classification based on longitudinal and time-to-event data:
```

```
class1 class2 class3
N 569.00 6433.00 832.00
% 7.26 82.12 10.62
```

Posterior classification table:

```
--> mean of posterior probabilities in each class
      prob1 prob2 prob3
class1 0.9388 0.0435 0.0177
class2 0.0031 0.9226 0.0743
class3 0.0068 0.1343 0.8589
```

Posterior probabilities above a threshold (%):

```
      class1 class2 class3
prob>0.7  91.21  93.21  76.20
prob>0.8  87.17  85.98  67.19
prob>0.9  80.84  70.76  58.29
```

Posterior classification based only on longitudinal data:

```
      class1 class2 class3
N 480.00 6637.00 717.00
%   6.13   84.72   9.15
```

```
> summary(jlcm_c3)
```

Joint latent class model for quantitative outcome and competing risks
fitted by maximum likelihood method

```
Jointlcm(fixed = log_eGFR ~ FW_TIME + x1 + x2 + x3 + F_AGE +
  SEX + F_DMAGE, mixture = ~FW_TIME + x1 + x2 + x3, random = ~FW_TIME,
  subject = "Obs_id", ng = 3, survival = Surv(CENSOR_TIME,
    CENSOR_END) ~ F_AGE + SEX + F_DMAGE, data = sub_nomiss,
  maxiter = 100)
```

Statistical Model:

```
Dataset: sub_nomiss
Number of subjects: 7834
Number of observations: 130963
Number of latent classes: 3
Number of parameters: 33
Event 1:
  Number of events: 1806
  Class-specific hazards and
  Weibull baseline risk function
```

Iteration process:

```
Convergence criteria satisfied
Number of iterations: 34
Convergence criteria: parameters= 9.5e-14
                     : likelihood= 5e-10
```

```
: second derivatives= 1.5e-11
```

Goodness-of-fit statistics:

```
maximum log-likelihood: 44744.8
```

```
AIC: -89423.6
```

```
BIC: -89193.72
```

```
Score test statistic for CI assumption: 180.127 (p-value=0)
```

Maximum Likelihood Estimates:

Fixed effects in the class-membership model:

(the class of reference is the last class)

	coef	Se	Wald	p-value
intercept class1	-0.76441	0.06271	-12.189	0.00000
intercept class2	1.61906	0.05177	31.273	0.00000

Parameters in the proportional hazard model:

	coef	Se	Wald	p-value
event1 +/-sqrt(Weibull1) class 1	0.33825	0.00656	51.568	0.00000
event1 +/-sqrt(Weibull2) class 1	1.32305	0.02775	47.680	0.00000
event1 +/-sqrt(Weibull1) class 2	0.13543	0.00465	29.098	0.00000
event1 +/-sqrt(Weibull2) class 2	1.33904	0.03672	36.464	0.00000
event1 +/-sqrt(Weibull1) class 3	0.24125	0.00254	94.810	0.00000
event1 +/-sqrt(Weibull2) class 3	1.85656	0.06602	28.123	0.00000
F_AGE	0.00239	0.00022	10.852	0.00000
SEX	0.21303	0.04883	4.362	0.00001
F_DMAGE	0.02957	0.00330	8.962	0.00000

Fixed effects in the longitudinal model:

	coef	Se	Wald	p-value
intercept class1	5.98451	0.02653	225.540	0.00000
intercept class2	4.82430	0.00725	665.334	0.00000
intercept class3	5.02966	0.01162	432.834	0.00000
FW_TIME class1	-0.49807	0.00664	-75.056	0.00000
FW_TIME class2	0.00244	0.00072	3.396	0.00068
FW_TIME class3	-0.07305	0.00196	-37.222	0.00000
x1 class1	6.02422	0.21107	28.542	0.00000
x1 class2	-0.04890	0.00520	-9.400	0.00000
x1 class3	1.45480	0.02434	59.782	0.00000
x2 class1	6.00295	0.32327	18.570	0.00000
x2 class2	0.00105	0.01080	0.097	0.92268
x2 class3	1.25631	0.03835	32.762	0.00000
x3 class1	6.87198	0.25154	27.320	0.00000
x3 class2	-0.16791	0.00981	-17.123	0.00000
x3 class3	1.44834	0.03208	45.147	0.00000

```

F_AGE          -0.00248 0.00011 -22.760 0.00000
SEX            -0.01126 0.00517  -2.179 0.02930
F_DMAGE        -0.00601 0.00041 -14.542 0.00000

```

Variance-covariance matrix of the random-effects:

```

            intercept FW_TIME
intercept    0.04425
FW_TIME      0.00036 0.00175

```

```

               coef      Se
Residual standard error 0.13006 0.00028

```

```
> (par_c3 <- jlcm_c3$best)
```

```

            intercept class1            intercept class2
            -0.7644072016                1.6190648056
event1 +/-sqrt(Weibull1) class 1 event1 +/-sqrt(Weibull2) class 1
            0.3382510520                1.3230479321
event1 +/-sqrt(Weibull1) class 2 event1 +/-sqrt(Weibull2) class 2
            0.1354311678                1.3390440624
event1 +/-sqrt(Weibull1) class 3 event1 +/-sqrt(Weibull2) class 3
            0.2412492457                1.8565575710
            F_AGE                      SEX
            0.0023916635                0.2130334850
            F_DMAGE                    intercept class1
            0.0295725511                5.9845132937
intercept class2                    intercept class3
            4.8243001388                5.0296571418
FW_TIME class1                    FW_TIME class2
            -0.4980748837                0.0024441032
FW_TIME class3                    x1 class1
            -0.0730477503                6.0242181856
            x1 class2                    x1 class3
            -0.0489045541                1.4548031899
            x2 class1                    x2 class2
            6.0029529120                0.0010483777
            x2 class3                    x3 class1
            1.2563127237                6.8719755096
            x3 class2                    x3 class3
            -0.1679125273                1.4483393088
            F_AGE                      SEX
            -0.0024790065                -0.0112569433
            F_DMAGE                    varcov 1
            -0.0060078041                0.0442461523
            varcov 2                    varcov 3

```

```

0.0003617225
      stderr
0.1300637874

> length(par_c3)

[1] 33

> #joint model with 4 class. We initialized the parameters based on 3 classes model
> #4 classes
> #jlc4 <- gridsearch(rep=20, maxiter=50, minit=jlc3, m=Jointlcmm(log_eGFR ~ FW_TIME + x1 + x2 +
> #
> #           mixture = ~ FW_TIME + x1 + x2 + x3,
> #           random = ~ FW_TIME,
> #           survival = Surv(CENSOR_TIME, CENSOR_END) ~ F_AGE + SEX + F_DMAGE,
> #           subject = "Obs_id", ng=4, data=sub_nomiss))
> init_c4 <- c("intercept_class1"=-0.7, "intercept_class2"=0.3, "intercept_class3"=2.2, "weib_class1_1"=
+           "weib_class2_1"=0.2, "weib_class2_2"=1.5, "weib_class3_1"=0.2, "weib_class3_2"=2.2, "we
+           "f_age"=0.003, "sex"=0.3, "f_dmage"=0.02, "intercept_class1"=5.7, "intercept_class2"=4.
+           "fw_time1"=-0.46, "fw_time2"=-0.02, "fw_time3"=0.0002, "fw_time4"=-0.14, "x1_class1"=-
+           "x1_class3"=-0.08, "x1_class4"=3.2, "x2_class1"=-5.9, "x2_class2"=0.54, "x2_class3"=-0.
+           "x3_class3"=-0.15, "x3_class4"=3.3, "fage2"=-0.001, "sex2"=-0.02, "fdmage2"=-0.004,
+           "varcov1"=0.03, "varcov2"=-0.0001, "varcov3"=0.0012, "stderr"=0.12)
> jlc4 <- Jointlcmm(log_eGFR ~ FW_TIME + x1 + x2 + x3 + F_AGE + SEX + F_DMAGE, #
+           mixture = ~ FW_TIME + x1 + x2 + x3,
+           random = ~ FW_TIME,
+           survival = Surv(CENSOR_TIME, CENSOR_END) ~ F_AGE + SEX + F_DMAGE,
+           maxiter=100, B=init_c4,
+           subject = "Obs_id", ng=4, data=sub_nomiss)

Be patient, Jointlcmm is running ...
The program took 6989.42 seconds

> jlc4

Joint latent class model for quantitative outcome and competing risks
fitted by maximum likelihood method

Jointlcmm(fixed = log_eGFR ~ FW_TIME + x1 + x2 + x3 + F_AGE +
SEX + F_DMAGE, mixture = ~FW_TIME + x1 + x2 + x3, random = ~FW_TIME,
subject = "Obs_id", ng = 4, survival = Surv(CENSOR_TIME,
CENSOR_END) ~ F_AGE + SEX + F_DMAGE, data = sub_nomiss,
maxiter = 100)

```

Statistical Model:

```

Dataset: sub_nomiss
Number of subjects: 7834
Number of observations: 130963
Number of latent classes: 4
Number of parameters: 41
Event 1 :
    Number of events: 1806
    Class-specific hazards and
    Weibull baseline risk function

```

Iteration process:

```

Convergence criteria satisfied
Number of iterations: 28
Convergence criteria: parameters= 3e-06
                     : likelihood= 5.2e-06
                     : second derivatives= 3e-08

```

Goodness-of-fit statistics:

```

maximum log-likelihood: 47027.28
AIC: -93972.56
BIC: -93686.95

```

```
> postprob(jlcm_c4)
```

Posterior classification based on longitudinal and time-to-event data:

```

class1 class2 class3 class4
N 342.00 1208.00 5845.00 439.0
% 4.37 15.42 74.61 5.6

```

Posterior classification table:

```

--> mean of posterior probabilities in each class
      prob1 prob2 prob3 prob4
class1 0.9457 0.0323 0.0038 0.0182
class2 0.0074 0.8046 0.1458 0.0422
class3 0.0000 0.0958 0.9009 0.0032
class4 0.0064 0.0696 0.0303 0.8937

```

Posterior probabilities above a threshold (%):

```

      class1 class2 class3 class4
prob>0.7  92.40  69.54  91.72  82.92
prob>0.8  87.72  54.55  78.82  77.45
prob>0.9  83.33  37.91  60.02  70.39

```

Posterior classification based only on longitudinal data:

	class1	class2	class3	class4
N	286.00	811.00	6365.00	372.00
%	3.65	10.35	81.25	4.75

```
> summary(jlcm_c4)
```

Joint latent class model for quantitative outcome and competing risks
fitted by maximum likelihood method

```
Jointlcm(fixed = log_eGFR ~ FW_TIME + x1 + x2 + x3 + F_AGE +  
  SEX + F_DMAGE, mixture = ~FW_TIME + x1 + x2 + x3, random = ~FW_TIME,  
  subject = "Obs_id", ng = 4, survival = Surv(CENSOR_TIME,  
    CENSOR_END) ~ F_AGE + SEX + F_DMAGE, data = sub_nomiss,  
  maxiter = 100)
```

Statistical Model:

Dataset: sub_nomiss
Number of subjects: 7834
Number of observations: 130963
Number of latent classes: 4
Number of parameters: 41
Event 1:
 Number of events: 1806
 Class-specific hazards and
 Weibull baseline risk function

Iteration process:

Convergence criteria satisfied
Number of iterations: 28
Convergence criteria: parameters= 3e-06
 : likelihood= 5.2e-06
 : second derivatives= 3e-08

Goodness-of-fit statistics:

maximum log-likelihood: 47027.28
AIC: -93972.56
BIC: -93686.95
Score test statistic for CI assumption: 135.469 (p-value=0)

Maximum Likelihood Estimates:

Fixed effects in the class-membership model:
(the class of reference is the last class)

	coef	Se	Wald p-value
--	------	----	--------------

```

intercept class1 -0.33457  0.08122  -4.120  0.00004
intercept class2  1.21147  0.07177  16.880  0.00000
intercept class3  2.45503  0.05850  41.970  0.00000

```

Parameters in the proportional hazard model:

	coef	Se	Wald	p-value
event1 +/-sqrt(Weibull1) class 1	0.39767	0.00891	44.622	0.00000
event1 +/-sqrt(Weibull2) class 1	1.31077	0.03155	41.544	0.00000
event1 +/-sqrt(Weibull1) class 2	0.21618	0.00441	49.068	0.00000
event1 +/-sqrt(Weibull2) class 2	1.41911	0.03482	40.761	0.00000
event1 +/-sqrt(Weibull1) class 3	0.16645	0.00558	29.845	0.00000
event1 +/-sqrt(Weibull2) class 3	1.79791	0.08624	20.847	0.00000
event1 +/-sqrt(Weibull1) class 4	0.29457	0.00291	101.161	0.00000
event1 +/-sqrt(Weibull2) class 4	1.92491	0.05936	32.427	0.00000
F_AGE	0.00279	0.00024	11.612	0.00000
SEX	0.18336	0.04956	3.700	0.00022
F_DMAGE	0.03058	0.00343	8.924	0.00000

Fixed effects in the longitudinal model:

	coef	Se	Wald	p-value
intercept class1	7.09302	0.83190	8.526	0.00000
intercept class2	4.95149	0.01100	450.070	0.00000
intercept class3	4.82436	0.00724	666.435	0.00000
intercept class4	5.76886	0.02112	273.091	0.00000
FW_TIME class1	-0.97263	0.18033	-5.394	0.00000
FW_TIME class2	-0.04444	0.00157	-28.339	0.00000
FW_TIME class3	0.00241	0.00069	3.509	0.00045
FW_TIME class4	-0.30648	0.00426	-72.008	0.00000
x1 class1	1.47613	7.75935	0.190	0.84912
x1 class2	1.14058	0.02391	47.701	0.00000
x1 class3	-0.07202	0.00451	-15.956	0.00000
x1 class4	5.53620	0.15730	35.194	0.00000
x2 class1	-9.18951	10.59291	-0.868	0.38566
x2 class2	0.95806	0.03303	29.002	0.00000
x2 class3	-0.00963	0.00747	-1.290	0.19720
x2 class4	5.48670	0.24141	22.727	0.00000
x3 class1	-0.20405	8.62844	-0.024	0.98113
x3 class2	1.01707	0.03062	33.220	0.00000
x3 class3	-0.18845	0.00851	-22.150	0.00000
x3 class4	5.98294	0.19018	31.459	0.00000
F_AGE	-0.00241	0.00011	-22.193	0.00000
SEX	-0.00921	0.00514	-1.791	0.07331
F_DMAGE	-0.00596	0.00041	-14.484	0.00000

Variance-covariance matrix of the random-effects:

```

            intercept FW_TIME
intercept    0.04367
FW_TIME      0.00037 0.0013

            coef      Se
Residual standard error 0.12821 0.00028

> (sum_tab <- summarytable(jlcm_c1, jlcm_c3, jlcm_c4))

      G  loglik npm      BIC    %class1 %class2 %class3 %class4
jlcm_c1 1 34390.12 17 -68627.82 100.000000
jlcm_c3 3 44744.80 33 -89193.72  7.263212 82.11642 10.62037
jlcm_c4 4 47027.28 41 -93686.95  4.365586 15.41996 74.61067 5.603778
      G  loglik npm      BIC    %class1 %class2 %class3 %class4
jlcm_c1 1 34390.12 17 -68627.82 100.000000      NA      NA      NA
jlcm_c3 3 44744.80 33 -89193.72  7.263212 82.11642 10.62037      NA
jlcm_c4 4 47027.28 41 -93686.95  4.365586 15.41996 74.61067 5.603778

> write.table(sum_tab, "JLCM_Summary_Tab.csv", row.names=F, sep=",")

```

2.2 Plots of trajectory

Plot the trajectory of fitting and survival curves (goodness-of-fit)

```

> plot_model <- jlcm_c3
> (ng <- plot_model$ng)

[1] 3

> #par(mfrow=c(2,2))
> pdf(paste("Goodness_of_fit", ng, "classes.pdf", sep=""))
> plot(plot_model, which="survival", lwd=3)
> plot(plot_model, which="fit", var.time="FW_TIME", marg=F, break.times=10,
+      bty="l", ylab="Log-eGFR", xlab="Follow-up period (years)", lwd=2)      #subject-specific predictions
> plot(plot_model, which="fit", var.time="FW_TIME", marg=T, break.times=10,
+      bty="l", ylab="Log-eGFR", xlab="Follow-up period (years)", lwd=2)      #marginal predictions
> plot(plot_model, which="residuals")
> plot(plot_model, which="postprob")
> dev.off()

```

```

null device
      1

```

Plot of posterior probability

```

> pdf(paste("PProb", ng, "classes.pdf", sep=""))
> ylim <- c(0, 120)
> if(ng==3) {
+   clr <- c("black", "red", "blue")
+   label <- c(1, 3, 2) #to change the order of label (high risk -> middle risk)
+   xlim <- c(0, 20)
+ } else if(ng==4) {
+   clr <- c("black", "red", "blue", "brown")
+   label <- c(1, 2, 4, 3)
+   xlim <- c(0, 15)
+ }
> par(mfrow=c(1,ng), mar=c(3,0,1,1), oma=c(2,6,0,0),
+     las=1, bty="n", mgp=c(3,1,0)/1.6)
> ppr <- plot_model$pprob
> num <- table(ppr$class)
> for(i in 1:ng)
+ {
+   hist(ppr[ppr$class==i,i+2], breaks=0:50/50,
+        col=clr[i], border=clr[i], ylim=c(0,100),
+        main="", xlab="", yaxt="n", yaxs="i", ylab="")
+   if(i==1) axis(side=2)
+   text(0.4,30,paste("Class",i," n=",num[i]),
+        font=2,col=clr[i], cex=1)
+ }
> mtext("Posterior probability", side=1, line=0, outer=T)
> par(las=3)
> mtext("Number of subjects", side=2, line=3, outer=T)
> pairs(ppr[, 2+1:ng], pch=16, col=clr[ppr$class],
+       cex=0.4, gap=0)
> dev.off()

```

```

pdf
      2

```

```

> length(table(sub_nomiss$FW_TIME))

```

```

[1] 61

```

```

> x <- sort(unique(sub_nomiss$FW_TIME))
> length(x)

```

```
[1] 61
```

```
> wh <- match(x, sub_nomiss$FW_TIME)[1:length(x)]
> length(wh)
```

```
[1] 61
```

```
> summary(sub_nomiss$FW_TIME[wh])
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	5	10	10	15	20

```
> plotdata <- data.frame(FW_TIME = sub_nomiss$FW_TIME[wh],
+                         MM[wh,],
+                         F_AGE = 55 + sub_nomiss$FW_TIME[wh],
+                         SEX = 1,
+                         F_DMAGE = 7 + sub_nomiss$FW_TIME[wh])
> head(plotdata)
```

	FW_TIME	x1	x2	x3	F_AGE	SEX	F_DMAGE
1	0.0000000	-0.09278708	0.24736837	-0.31475232	55.00000	1	7.000000
2	0.3333333	-0.10111103	0.21651457	-0.26266099	55.33333	1	7.333333
3	0.6666667	-0.10913867	0.18581102	-0.21102042	55.66667	1	7.666667
4	1.0000000	-0.11657373	0.15540796	-0.16028134	56.00000	1	8.000000
5	1.3333333	-0.12311989	0.12545566	-0.11089451	56.33333	1	8.333333
6	1.6666667	-0.12848087	0.09610435	-0.06331067	56.66667	1	8.666667

```
> range(plotdata$FW_TIME)
```

```
[1] 0 20
```

```
> range(plotdata$F_DMAGE)
```

```
[1] 7 27
```

```
> pdf(paste("Prediction_plot", ng, "classes.pdf", sep=""))
> lwd_main <- 2
> ltys <- 1:ng
> pred_plot <- predictY(plot_model, plotdata, var.time="FW_TIME", draws=TRUE)
> pred_ori <- pred_plot
> pred_ori$pred <- exp(pred_plot$pred)
> pred_data <- pred_ori$pred
> dim(pred_data)
```

```
[1] 61 9
```

```
> #Modify the class label for log-eGFR trajectory curves.
> plot(pred_plot, xlab="Follow-up period (years)", ylab="log-eGFR", bty="l", lwd=lwd_main,
+       xlim=xlim, legend=NULL, col=clrs, lty=ltys, legend.loc="bottomleft")
> legend("bottomleft", legend = paste("class", 1:ng), col=clrs[label], lty=ltys[label], lwd=lwd_main,
+       cex=0.7)
> #Modify the class label for eGFR trajectory curves.
> plot(pred_ori, xlab="Follow-up period (years)", ylab="eGFR", bty="l", lwd=lwd_main, legend=NULL,
+       ylim=ylim, xlim=xlim, col=clrs, lty=ltys)
> legend("bottomleft", legend = paste("class", 1:ng), col=clrs[label], lty=ltys[label], lwd=lwd_main,
+       cex=0.7)
> abline(h=15, lty="dashed")
> dev.off()
```

```
null device
      1
```

2.3 Post-model analyses

We further investigate the characteristics of subjects in each classes, including the baseline characteristics, summary of eGFR measures!

```
> #final_dat <- merge(update_base, ppr[1:2], by.x="Obs_id", by.y="Obs_id", all.x=T, sort=F)
> final_dat <- merge(update_base, ppr[1:2], by=intersect("Obs_id", "Obs_id"), sort=F) #only the
> final_dat <- merge(final_dat, gwas_id, by.x="Obs_id", by.y="Obs_id", all.x=T, sort=F)
> (dim_dat <- dim(final_dat))
```

```
[1] 7834 150
```

```
> addmargins(table(final_dat$class, useNA="ifany"))
```

```

      1      2      3 Sum
569 6433  832 7834

> (total_esrd <- nrow(subset(final_dat, ESRD1_END==1)))

[1] 661

> (total_death <- nrow(subset(final_dat, DEATH_END==1)))

[1] 1590

> (total_censor <- nrow(subset(final_dat, CENSOR_END==1)))

[1] 1806

> (total_eGFR30 <- nrow(subset(final_dat, min_eGFR<30)))

[1] 1234

> stat <- by(final_dat, final_dat$class, function(x) {
+   num_obj <- nrow(x)
+   esrd_end <- nrow(subset(x, ESRD1_END==1))
+   death_end <- nrow(subset(x, DEATH_END==1))
+   censor_end <- nrow(subset(x, CENSOR_END==1))
+   min_eGFR30 <- nrow(subset(x, min_eGFR<30))
+   stat_obj <- paste(num_obj, " (", round(num_obj/dim_dat[1], 4)*100, "%)", sep="")
+   stat_esrd <- paste(esrd_end, " (", round(esrd_end/num_obj, 4)*100, "%)", sep="")
+   min_eGFR30 <- paste(min_eGFR30, " (", round(min_eGFR30/num_obj, 4)*100, "%)", sep="")
+   stat_death <- paste(death_end, " (", round(death_end/num_obj, 4)*100, "%)", sep="")
+   stat_censor <- paste(censor_end, " (", round(censor_end/num_obj, 4)*100, "%)", sep="")
+   stat_eGFR_IQR <- paste(median(x$cut_num_eGFR), " (", paste(quantile(x$cut_num_eGFR,c(0.25,0.75)),
+   stat_ftime_IQR <- paste(round(median(x$ESRD1_TIME),2), " (", paste(round(quantile(x$ESRD1_TIME),2),
+   stat_eGFR_range <- paste(median(x$cut_num_eGFR), " (", paste(range(x$cut_num_eGFR), collapse="-",
+   stat_ftime_range <- paste(round(median(x$ESRD1_TIME),2), " (", paste(round(range(x$ESRD1_TIME),2),
+   avg_slope <- round(median(x$slope_pct_nonadj_yInf, na.rm=T), 2)
+   num_gwas <- length(which(!is.na(x$FID)))
+   return(cbind(x$class[1], stat_obj, stat_esrd, min_eGFR30, stat_death, stat_censor, stat_eGFR_IQR,
+               stat_eGFR_range, stat_ftime_range, avg_slope, num_gwas))
+ })
> stat <- do.call("rbind", stat)
> colnames(stat) <- c("Class", "#Subjects", "#ESRD events", "#min eGFR30", "#Death events", "#Censor
+                   "Follow-up years[median(IQR)]", "#cutted eGFR measures[Median(range)]", "Follow-u
+                   "#GWAS subjects")
> stat

```

```

      Class #Subjects      #ESRD events  #min eGFR30    #Death events
[1,] "1"      "569 (7.26%)"  "354 (62.21%)" "455 (79.96%)" "432 (75.92%)"
[2,] "2"      "6433 (82.12%)" "84 (1.31%)"   "384 (5.97%)"  "772 (12%)"
[3,] "3"      "832 (10.62%)"  "223 (26.8%)"  "395 (47.48%)" "386 (46.39%)"
      #Censor events #cutted eGFR measures[Median(IQR)]
[1,] "540 (94.9%)"  "9 (5-14)"
[2,] "803 (12.48%)" "16 (9-23)"
[3,] "463 (55.65%)" "19 (13-26)"
      Follow-up years[median(IQR)] #cutted eGFR measures[Median(range)]
[1,] "5.18 (3.17-7.39)"           "9 (2-34)"
[2,] "12.87 (8.91-16.29)"         "16 (2-43)"
[3,] "11.62 (9.92-13.56)"         "19 (2-43)"
      Follow-up years[median(range)] eGFR change percentage(median)
[1,] "5.18 (0.12-17.7)"           "-18.48"
[2,] "12.87 (0.14-19.94)"         "-0.49"
[3,] "11.62 (3.93-19.76)"         "-5.66"
      #GWAS subjects
[1,] "370"
[2,] "3829"
[3,] "543"

```

```

> write.table(stat, paste("Grp", paste(egfr_grp, collapse="-")), "_Summary_", ng,"classes.csv",
> write.table(final_dat, paste("Grp", paste(egfr_grp, collapse="-")), "_Classified_Data", ng, "c

```

Time difference of 3.875077 hours