A practical guide to fit a recurrent episode model for episodic diseases by using *RecrEpis*

Based on the methodology developed in paper titled "A parametric model could jointly characterize rate, duration, and severity of exacerbations in episodic diseases" by A. Safari, J. M. FitzGerald, J. Petkau, and M. Sadatsafavi

Abstract

Most of episodic diseased can be characterized by periods of intensified disease activity often referred to as exacerbations. Three main features of such exacerbations are: rate, duration, and severity. We provided a commented flexible scripted SAS macro (called *RecrEpis*) that can be employed to model these features given patients characteristics (time-dependent/independent covariates). The SAS macro can be fined here along with an example dataset (sim_data.csv).

A brief introduction to our recurrent episode model

In recurrent episode models, we have a two-state alternate process where the first state is waiting time to the onset of an exacerbation (between exacerbations) and the second state is waiting time to the termination of an exacerbation (within exacerbation). We use two frailty models to characterize these two features of the data (rate and duration submodels). In addition, since severity of each exacerbation (binary outcome: severe or not) is another important feature in episodic diseases, we add a logistic regression model to characterize the severity outcome. At the end, in many episodic disease studies, there is a considerable portion of patients with no exacerbations during the follow-up period. In presence of such patients, especially when their portion is larger than usual, it may be easier and even more accurate to add another component into the model to characterize those patients separately. Specifically, we use another logistic regression model to characterize patients who are not likely to experience any exacerbations during follow-up period (non-susceptible submodel).

Steps to perform a recurrent episodic analysis using *RecrEpis* SAS macro

The purpose of this manual is to give a practical guide to fit a recurrent episode model by using *RecrEpis* SAS macro. This section will go over each required step to carry out a simple but complete analysis using SAS®.

To fit the model with its all four submodels (rate, duration, severity, and non-susceptibility), the following columns are required in the data:

i. Id: patient id
ii. Period: exacerbation index number (0 means no exacerbation)
iii. Event: mild exacerbation onset (1), severe exacerbation onset (2), exacerbation termination (3), out of exacerbation censoring (-1), within exacerbation censoring (-2)
iv. Time: time to events since time of the previous event (gap times)
v. Covariates: any available covariates

The SAS macro *RecrEpis* has 4 inputs:

- FileDir: the local directory of the data file.
- Dur_SubMol: if “YES”, the duration submodel will be included in the final model and “NO” otherwise.
- Sev_SubMod: if “YES”, the severity submodel will be included in the final model and “NO” otherwise.
- NoExac_SubMod: if “YES”, the non-susceptible submodel will be included in the final model and “NO” otherwise.

After adding the SAS macro into your SAS session, the following lines fit the model with all the submodels:

```
%LET FileDir="C:\Users\Desktop\sim_data.csv";

%RecrEpis(FileDir=&FileDir, Dur_SubMod="YES", Sev_SubMod="YES", NoExac_SubMod="YES")
```

The fit produces the following outputs:

- Model specifications:

Specifications	
Data Set	WORK.SIM_DATA
Dependent Variable	event
Distribution for Dependent Variable	General
Random Effects	Z_r Z_d Z_s
Distribution for Random Effects	Normal
Subject Variable	id
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

- Summary of number of observations and parameters used in the model:

Dimensions	
Observations Used	1109
Observations Not Used	0
Total Observations	1109
Subjects	500
Max Obs per Subject	10
Parameters	20
Quadrature Points	3

- Initial values for the parameters:

Initial Parameters																				
gamma1	gamma2	br_0	br_x1	br_x2	bp_0	bp_x1	bp_x2	bd_0	bd_x1	bd_x2	bs_0	bs_x1	bs_x2	v_r	v_d	v_s	v_rd	v_rs	v_ds	Negative Log Likelihood
1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	1343.03203

- Optimization iteration history:

Iteration History

Iteration	Calls	Negative Log Likelihood	Difference	Maximum Gradient	Slope
1	4	877.9633	465.0688	238.068	-7740.69
2	8	565.6299	312.3333	46.1277	-307.656
3	12	504.5977	61.03224	27.2030	-49.7556
4	16	491.6284	12.96934	12.8960	-12.2211
5	20	482.4737	9.154632	9.07798	-6.38344
6	23	479.1617	3.312072	11.4076	-2.82536
7	25	478.1270	1.034622	18.1492	-1.23642
8	29	475.6373	2.489697	5.56535	-4.28903
9	90	474.7284	0.908922	3.06628	-0.71568
10	154	474.4300	0.298422	4.05852	-52.2701

- Goodness of fit results:

Fit Statistics

-2 Log Likelihood	944.1
AIC (smaller is better)	984.1
AICC (smaller is better)	984.8
BIC (smaller is better)	1068.4

- Original parameter estimates used in the model:

Parameter Estimates

	Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	95% Confidence Limits		Gradient
log-normal shape parameters	gamma1	1.2591	0.08675	497	14.51	<.0001	1.0886	1.4295	0.002440
	gamma2	1.5080	0.1225	497	12.31	<.0001	1.2672	1.7487	0.001988
regression coefficients in rate submodel	br_0	-1.2960	0.1148	497	-11.29	<.0001	-1.5216	-1.0704	0.000503
	br_x1	0.6373	0.07624	497	8.36	<.0001	0.4876	0.7871	-0.00235
	br_x2	0.3036	0.1567	497	1.94	0.0532	-0.00420	0.6115	0.001604
regression coefficients in non-susceptible submodel	bp_0	-19.6987	1290.03	497	-0.02	0.9878	-2554.29	2514.89	7.51E-7
	bp_x1	-0.1995	1173.00	497	-0.00	0.9999	-2304.84	2304.44	-3.55E-7
	bp_x2	-3.8190	8441.80	497	-0.00	0.9996	-16590	16582	1.43E-8
regression coefficients in duration submodel	bd_0	0.2126	0.1779	497	1.19	0.2327	-0.1370	0.5623	0.000260
	bd_x1	0.6096	0.1130	497	5.40	<.0001	0.3876	0.8316	0.000153
	bd_x2	0.6261	0.2174	497	2.88	0.0042	0.1989	1.0532	0.000856
regression coefficients in severity submodel	bs_0	-1.1106	0.2795	497	-3.97	<.0001	-1.6597	-0.5615	0.006791
	bs_x1	0.6252	0.1648	497	3.79	0.0002	0.3014	0.9490	-0.00243
	bs_x2	-0.4907	0.2887	497	-1.70	0.0898	-1.0578	0.07643	0.006191
variances and covariances of random effects	v_r	1.4575	0.2795	497	5.21	<.0001	0.9084	2.0067	-0.00299
	v_d	0.6732	0.2976	497	2.26	0.0241	0.08844	1.2579	0.001941
	v_s	1.1496	1.2361	497	0.93	0.3528	-1.2791	3.5783	0.000362
	v_rd	-0.3412	0.2112	497	-1.62	0.1068	-0.7561	0.07371	0.001163
	v_rs	-0.1832	0.2967	497	-0.62	0.5373	-0.7662	0.3998	-0.00378
	v_ds	0.7861	0.2988	497	2.63	0.0088	0.1991	1.3732	-0.00106

- Standard deviations and correlation coefficients estimate of the random effect:

Additional Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Z_r_SD	1.2073	0.1158	497	10.43	<.0001	0.05	0.9798	1.4347
Z_d_SD	0.8205	0.1814	497	4.52	<.0001	0.05	0.4641	1.1768
Correlation coefficient (Zr_Zd_RHO)	-0.3445	0.1908	497	-1.81	0.0716	0.05	-0.7194	0.03039
Z_s_SD	1.0722	0.5764	497	1.86	0.0635	0.05	-0.06037	2.2048
Correlation coefficient (Zr_Zs_RHO)	-0.1415	0.2295	497	-0.62	0.5377	0.05	-0.5924	0.3093
Correlation coefficient (Zd_Zs_RHO)	0.8936	0.3697	497	2.42	0.0160	0.05	0.1672	1.6201

A few notes on the online version of *RecrEpis* macro:

- Considers only two covariates (var1 and var2) for all the submodels. To choose any arbitrary set of covariates for each submodel:
    - For rate submodel update lines 61, 89, 117, 143, 171, 196, 222, 248, 282
    - For non-susceptible submodel update lines 62, 90, 118, 144, 197, 223, 294
    - For duration submodel update lines 63, 118, 145, 223, 304
    - For severity submodel update lines 64, 91, 119, 172, 317
- Considers log-normal distribution for both between and within exacerbation times. To change those to any other distribution:
    - For between exacerbation times update lines 285-286
    - For within exacerbation times update lines 307-308