Final Report

randy

2024-06-19

Table of Contents

Goal of the analysis

- ☒ Fit the final joint model with rate with `JMbayes2` based on model selection results

- ☒ Table1

- ☒ Life Tables for Training and Testing datasets

- ☒ Dynamic prediction for the joint model

- ☒ Training dataset

  - ☒ Calculate the AUC for the joint model for $Dt = 1,2,3,4,5$ (year)

  - ☒ Calculate the Brier score for the joint model for $Dt = 1,2,3,4,5$ (year)

- ☒ Testing dataset

  - ☒ Calculate the AUC for the joint model for $Dt = 1,2,3,4,5$ (year)

  - ☒ Calculate the Brier score for the joint model for $Dt = 1,2,3,4,5$ (year)

Final Table1

For testing dataset there are missings for amh and the visit numbers are different from the training.

Characteristics	Overall, N = 901[1]	0Training, N = 600[1]	1Testing, N = 301[1]
Baseline Age (years)			
Mean (SD)	46.38 (2.41)	46.42 (2.41)	46.29 (2.42)
Median	46.00	46.00	46.00
[Range]	[42.00, 53.00]	[42.00, 53.00]	[42.00, 53.00]
Ethnicity / Race			
BLACK	242 / 901 (27%)	170 / 600 (28%)	72 / 301 (24%)
CAUCA	418 / 901 (46%)	265 / 600 (44%)	153 / 301 (51%)
CHINE	99 / 901 (11%)	62 / 600 (10%)	37 / 301 (12%)
HISPA	24 / 901 (2.7%)	18 / 600 (3.0%)	6 / 301 (2.0%)
JAPAN	118 / 901 (13%)	85 / 600 (14%)	33 / 301 (11%)
Marital Status			
Not Married/Partnered	285 / 899 (32%)	191 / 600 (32%)	94 / 299 (31%)
Married/Partnered	614 / 899 (68%)	409 / 600 (68%)	205 / 299 (69%)
Unknown	2	0	2
Alcohol Use			
None	467 / 901 (52%)	308 / 600 (51%)	159 / 301 (53%)
<1/wk	111 / 901 (12%)	71 / 600 (12%)	40 / 301 (13%)
1-7/wk	222 / 901 (25%)	161 / 600 (27%)	61 / 301 (20%)
>7/wk	101 / 901 (11%)	60 / 600 (10%)	41 / 301 (14%)
Smoking History			
Never Smoked	562 / 901 (62%)	373 / 600 (62%)	189 / 301 (63%)
Past Only	220 / 901 (24%)	149 / 600 (25%)	71 / 301 (24%)
Current Smoker	119 / 901 (13%)	78 / 600 (13%)	41 / 301 (14%)
Ever use of BC Pills			

Characteristics	Overall, N = 901[1]	0Training, N = 600[1]	1Testing, N = 301[1]
Never	235 / 897 (26%)	155 / 600 (26%)	80 / 297 (27%)
Ever use BC pills	662 / 897 (74%)	445 / 600 (74%)	217 / 297 (73%)
Unknown	4	0	4
BMI			
Mean (SD)	28 (7)	28 (7)	28 (7)
Median	26	26	26
[Range]	[16, 57]	[16, 56]	[16, 57]
Overall Health			
Excellent	224 / 901 (25%)	147 / 600 (25%)	77 / 301 (26%)
Very Good	353 / 901 (39%)	233 / 600 (39%)	120 / 301 (40%)
Good	224 / 901 (25%)	151 / 600 (25%)	73 / 301 (24%)
Fair/Poor	100 / 901 (11%)	69 / 600 (12%)	31 / 301 (10%)
Physical Activity Score			
Mean (SD)	7.81 (1.75)	7.75 (1.77)	7.93 (1.70)
Median	7.70	7.60	7.90
[Range]	[3.40, 12.80]	[3.40, 12.80]	[3.80, 12.30]
Baseline AMH			
Mean (SD)	538 (687)	540 (702)	534 (658)
Median	271	273	266
[Range]	[1, 4,652]	[1, 4,652]	[1, 3,621]
Menopausal Status			
Pre	333 / 901 (37%)	222 / 600 (37%)	111 / 301 (37%)
Early Peri	568 / 901 (63%)	378 / 600 (63%)	190 / 301 (63%)
Late Peri	0 / 901 (0%)	0 / 600 (0%)	0 / 301 (0%)
Post	0 / 901 (0%)	0 / 600 (0%)	0 / 301 (0%)
Event Time			
Mean (SD)	10.97 (2.48)	10.98 (2.51)	10.94 (2.42)

Characteristics	Overall, N = 901[1]	0Training, N = 600[1]	1Testing, N = 301[1]
Median	10.96	11.05	10.82
[Range]	[3.59, 20.10]	[3.59, 20.10]	[4.42, 17.69]
Employment Status			
Unemployed	158 / 901 (18%)	104 / 600 (17%)	54 / 301 (18%)
Employed	743 / 901 (82%)	496 / 600 (83%)	247 / 301 (82%)
Educational Attainment			
High School or Less	186 / 900 (21%)	125 / 600 (21%)	61 / 300 (20%)
Some College	272 / 900 (30%)	190 / 600 (32%)	82 / 300 (27%)
College degree or higher	442 / 900 (49%)	285 / 600 (48%)	157 / 300 (52%)
Unknown	1	0	1
Study Site			
B	146 / 901 (16%)	96 / 600 (16%)	50 / 301 (17%)
C	125 / 901 (14%)	76 / 600 (13%)	49 / 301 (16%)
M	144 / 901 (16%)	103 / 600 (17%)	41 / 301 (14%)
NJ	35 / 901 (3.9%)	27 / 600 (4.5%)	8 / 301 (2.7%)
P	124 / 901 (14%)	74 / 600 (12%)	50 / 301 (17%)
UCD	159 / 901 (18%)	102 / 600 (17%)	57 / 301 (19%)
UCLA	168 / 901 (19%)	122 / 600 (20%)	46 / 301 (15%)
Visiting Numbers			
Mean (SD)	5.54 (2.10)	5.54 (2.14)	5.54 (2.02)
Median	5.00	5.00	5.00
[Range]	[3.00, 11.00]	[3.00, 11.00]	[3.00, 10.00]

[1]n / N (%)

Contingency Table

Training

1 Year Interval

Tstart   Tend   Events   AtRisks

Tstart	Tend	Events	AtRisks
0	1	0	601
1	2	0	601
2	3	0	601
3	4	2	600
4	5	4	598
5	6	6	594
6	7	22	588
7	8	35	566
8	9	58	531
9	10	81	473
10	11	90	392
11	12	93	302
12	13	84	209
13	14	59	125
14	15	32	66
15	16	16	34
16	17	11	18
17	18	6	7
20	21	1	1

2 Year Interval

Tstart	Tend	Events	AtRisks
0	2	0	601
2	4	2	601
4	6	10	598
6	8	57	588
8	10	139	531
10	12	183	392
12	14	143	209
14	16	48	66
16	18	17	18
20	22	1	1

3 Year Interval

Tstart	Tend	Events	AtRisks
0	3	0	601

Tstart	Tend	Events	AtRisks
3	6	12	600
6	9	115	588
9	12	264	473
12	15	175	209
15	18	33	34
18	21	1	1

4 Year Interval

Tstart	Tend	Events	AtRisks
0	4	2	601
4	8	67	598
8	12	322	531
12	16	191	209
16	20	17	18

5 Year Interval

Tstart	Tend	Events	AtRisks
0	5	6	601
5	10	202	594
10	15	358	392
15	20	33	34

Testing

1 Year Interval

Tstart	Tend	Events	AtRisks
0	1	0	302
1	2	0	302
2	3	0	302
4	5	1	301
5	6	2	300
6	7	12	298
7	8	21	286
8	9	34	265
9	10	33	231
10	11	59	198
11	12	49	139
12	13	29	90
13	14	29	61
14	15	15	32
15	16	12	17
16	17	2	5
17	18	3	3

2 Year Interval

Tstart	Tend	Events	AtRisks
0	2	0	302
2	4	0	302
4	6	3	301
6	8	33	298
8	10	67	265
10	12	108	198
12	14	58	90
14	16	27	32
16	18	5	5

3 Year Interval

Tstart	Tend	Events	AtRisks
0	3	0	302
3	6	3	301

Tstart	Tend	Events	AtRisks
6	9	67	298
9	12	141	231
12	15	73	90
15	18	17	17

4 Year Interval

Tstart	Tend	Events	AtRisks
0	4	0	302
4	8	36	301
8	12	175	265
12	16	85	90
16	20	5	5

5 Year Interval

Tstart	Tend	Events	AtRisks
0	5	1	302
5	10	102	300
10	15	181	198
15	20	17	17

Final Model

Value + Rate model

$jmb2.119.value.rate$:

$$lme1: \quad \mathbf{m}(time) = (fixed = lamh \sim time,$$
$$random = time|id)$$
$$surv19: \quad Surv(etime, event) \sim factor(bmi) + site + ethnic + smoke$$
$$+\mathbf{m}(time) + \mathbf{m}'(time)$$

Fitting the model with:

- Chain 5
- Iteration 500k
- Thinning 50
- Total iterations of 50k

Summarization

```
#>
#> Call:
#> Data Descriptives:
#> Number of Groups: 600        Number of events: 600 (100%)
#> Number of Observations:
#>   lamh: 3326
#>
#>                 DIC          WAIC           LPML
#> marginal    1582856.93 3.098274e+14 -1.972734e+09
#> conditional   12568.56 1.243650e+04 -7.035887e+03
#>
#> Random-effects covariance matrix:
#>
#>        StdDev    Corr
#> (Intr) 1.4337 (Intr)
#> time   0.0554 0.5626
#>
#> Survival Outcome:
#>                          Mean    StDev     2.5%     97.5%       P   Rhat
#> as.factor(bmi_cat)2    -1.7653   0.9776  -3.8803   -0.0204 0.0478 1.0915
#> as.factor(bmi_cat)3    -4.1637   1.2015  -6.7940   -2.1801 0.0000 1.3149
#> ethnic_black           -1.8183   0.9090  -3.8019   -0.2063 0.0270 1.1094
#> ethnic_chine           -0.5666   1.4562  -3.5373    2.2573 0.6963 1.0191
#> ethnic_hispa           -2.9555   3.0372  -9.2317    2.8380 0.3164 1.0236
#> ethnic_japan           -0.0405   1.3552  -2.7209    2.6614 0.9690 1.0154
#> site_c                  0.8305   1.1303  -1.3235    3.1756 0.4524 1.0208
#> site_m                  1.4957   1.0779  -0.5081    3.7606 0.1423 1.0487
#> site_nj                 3.1132   2.5748  -1.7007    8.5777 0.2066 1.0306
#> site_p                  2.1477   1.2220  -0.0466    4.7813 0.0557 1.0724
#> site_ucd                0.2790   1.3536  -2.3839    3.0123 0.8378 1.0141
#> site_ucla               2.1235   1.4331  -0.5253    5.0853 0.1219 1.0680
#> smoke_current_smoker    1.9266   1.0421   0.0660    4.1468 0.0430 1.1092
#> smoke_past_only        -0.1978   0.7382  -1.6563    1.2888 0.7808 1.0199
#> value(lamh)            -1.5168   0.6703  -2.7496   -0.1883 0.0226 1.0634
#> slope(lamh)          -168.5483  36.0005 -249.3613 -108.8513 0.0000 1.5492
#>
#> Longitudinal Outcome: lamh (family = gaussian, link = identity)
#>               Mean  StDev    2.5%   97.5% P   Rhat
#> (Intercept)  7.8514 0.0981  7.6627  8.0449 0 1.0480
#> time        -0.5953 0.0116 -0.6184 -0.5728 0 1.1215
#> sigma        1.3194 0.0179  1.2848  1.3551 0 1.0000
#>
#> MCMC summary:
#> chains: 5
#> iterations per chain: 7e+05
#> burn-in per chain: 2e+05
#> thinning: 50
#> time: 14.4 hours
```

SD Calculation for slope term

Hence, we have the final results for the variability of the slope term.

- The random effect slope term based on MCMC is $\sigma_{\beta_1} = 0.0146757$;

- The fixed effect slope term $\sigma_{b_1} = 0.0116406$;

- We have the pooled $\sigma_{b1+\beta1} = 0.0187318$.

- The 95% CI for the slope term is -0.5984049 (-0.6320562, -0.5586275)

Prediction Performance

Training set

AUC

	Dt1	Dt2	Dt3	Dt4	Dt5
Tstart2	NA	0.4739	0.7701	0.7270	0.7272
Tstart3	0.7062	0.8250	0.7565	0.7925	0.7755
Tstart4	0.8227	0.7686	0.7975	0.7682	0.7576
Tstart5	0.7517	0.7693	0.7524	0.7580	0.8016
Tstart6	0.7840	0.8101	0.7918	0.8030	0.7935
Tstart8	0.7892	0.8028	0.7814	0.7746	0.8105

Brier Scores

	Dt1	Dt2	Dt3	Dt4	Dt5
Tstart2	NA	0.0138	0.0324	0.0630	0.1196
Tstart3	0.0091	0.0284	0.0574	0.1116	0.1543
Tstart4	0.0118	0.0296	0.0837	0.1307	0.1847
Tstart5	0.0234	0.0685	0.1208	0.1716	0.1851
Tstart6	0.0542	0.0964	0.1491	0.1784	0.1871
Tstart8	0.1151	0.1606	0.1922	0.1853	0.1315

Testing set

AUC

	Dt1	Dt2	Dt3	Dt4	Dt5
Tstart2	NA	NA	0.9615	0.6820	0.7512
Tstart3	NA	0.9714	0.9432	0.9238	0.7250
Tstart4	0.9953	0.9163	0.8876	0.7834	0.7842
Tstart5	0.9321	0.8661	0.8346	0.7905	0.7184
Tstart6	0.8628	0.8537	0.7837	0.7113	0.7067
Tstart8	0.7681	0.6955	0.7328	0.7485	0.7717

Brier Scores

	Dt1	Dt2	Dt3	Dt4	Dt5
Tstart2	NA	NA	0.0078	0.0290	0.0897
Tstart3	NA	0.0057	0.0186	0.0521	0.1401
Tstart4	0.0041	0.0154	0.0562	0.1182	0.1685
Tstart5	0.0174	0.0512	0.0921	0.1541	0.2102
Tstart6	0.0444	0.0839	0.1433	0.2062	0.2282
Tstart8	0.1073	0.1903	0.2156	0.1974	0.1469