

# MERLIN - multivariate Mixed-Effects Regression for LInear, Non-linear and user defined models

Emma C. Martin,  
Alessandro Gasparini, Michael J. Crowther

*Department of Health Sciences,  
University of Leicester, UK,  
emma.martin@le.ac.uk*

Funding: MRC (MR/P015433/1)

# The motivation

# The motivation

Increasing access to big data such as electronic health records (EHRs)

# The motivation

Increasing access to big data such as electronic health records (EHRs)

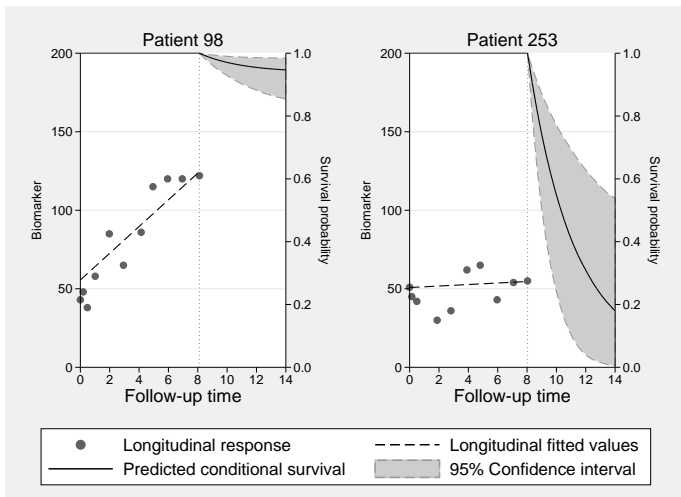
- ▶ multi-level (biomarkers < patients < GP practice area < geographical regions...)

# The motivation

Increasing access to big data such as electronic health records (EHRs)

- ▶ multi-level (biomarkers  $<$  patients  $<$  GP practice area  $<$  geographical regions...)
- ▶ multiple related outcomes (biomarkers, survival endpoints)

# Joint longitudinal survival models



# Joint longitudinal survival models

- ▶ `stjm` in Stata (Crowther et al., 2013)
- ▶ `gsem` in Stata
- ▶ `frailtypack` in R (Rondeau et al., 2012)
- ▶ `joiner` and `joinerML` in R (Philipson et al., 2018; Hickey et al., 2018)
- ▶ `JM` and `JMBayes` in R (Rizopoulos, 2016)
- ▶ ...

# The Goal

We want to increase the flexibility of joint longitudinal survival models, including extensions:

- ▶ competing risks (Li et al., 2009)
- ▶ different types of outcomes (Rizopoulos et al., 2008)
- ▶ multiple continuous outcomes (Lin et al., 2002)
- ▶ delayed entry (Crowther et al., 2016)
- ▶ recurrent events and a terminal event (Krol et al., 2016)
- ▶ predictions (Barrett and Su, 2017)



# MERLIN

- ▶ Tutorial paper in Stata (Crowther, 2018)
- ▶ [www.mjcrowther.co.uk/software/merlin](http://www.mjcrowther.co.uk/software/merlin)

MICHAEL J. CROWTHER

Home

Publications

Software

Posts

Projects

Teaching

## Survival (time-to-event) analysis

- [Parametric survival model with a frailty/random intercept](#) [Draft]
- [Parametric survival model with random coefficients](#) [TBA]
- [Three-level survival models - IPD meta-analysis of recurrent event data](#) [Draft, Sim]
- [Royston-Parmar multilevel survival models](#) [TBA]
- [User-defined hazard models - an example with fractional polynomials](#) [Tutorial]
- [Interval-censored survival analysis](#) [Tutorial]
- [Individual patient data network meta-analysis of survival data](#) [TBA]

# Primary biliary cirrhosis

To illustrate the flexibility we will use a single dataset of patients with primary biliary cirrhosis

# Primary biliary cirrhosis

To illustrate the flexibility we will use a single dataset of patients with primary biliary cirrhosis

- ▶ 312 patients with PBC collected at the Mayo Clinic 1974-1984 (Murtaugh et al., 1994)

# Primary biliary cirrhosis

To illustrate the flexibility we will use a single dataset of patients with primary biliary cirrhosis

- ▶ 312 patients with PBC collected at the Mayo Clinic 1974-1984 (Murtaugh et al., 1994)
- ▶ 158 randomised to receive D-penicillamine and 154 to placebo

# Primary biliary cirrhosis

To illustrate the flexibility we will use a single dataset of patients with primary biliary cirrhosis

- ▶ 312 patients with PBC collected at the Mayo Clinic 1974-1984 (Murtaugh et al., 1994)
- ▶ 158 randomised to receive D-penicillamine and 154 to placebo
- ▶ survival outcome is all-cause death, with 140 events

# Primary biliary cirrhosis

To illustrate the flexibility we will use a single dataset of patients with primary biliary cirrhosis

- ▶ 312 patients with PBC collected at the Mayo Clinic 1974-1984 (Murtaugh et al., 1994)
- ▶ 158 randomised to receive D-penicillamine and 154 to placebo
- ▶ survival outcome is all-cause death, with 140 events
  - ▶ We will simulate competing risks of death for illustration

# Primary biliary cirrhosis

To illustrate the flexibility we will use a single dataset of patients with primary biliary cirrhosis

- ▶ 312 patients with PBC collected at the Mayo Clinic 1974-1984 (Murtaugh et al., 1994)
- ▶ 158 randomised to receive D-penicillamine and 154 to placebo
- ▶ survival outcome is all-cause death, with 140 events
  - ▶ We will simulate competing risks of death for illustration
- ▶ 1945 repeated measurements of serum bilirubin, as well as other longitudinal biomarkers

# Primary biliary cirrhosis

To illustrate the flexibility we will use a single dataset of patients with primary biliary cirrhosis

- ▶ 312 patients with PBC collected at the Mayo Clinic 1974-1984 (Murtaugh et al., 1994)
- ▶ 158 randomised to receive D-penicillamine and 154 to placebo
- ▶ survival outcome is all-cause death, with 140 events
  - ▶ We will simulate competing risks of death for illustration
- ▶ 1945 repeated measurements of serum bilirubin, as well as other longitudinal biomarkers
- ▶ a formatted version of this data set is included in the `merlin` package in R



# Primary biliary cirrhosis data

	id	stime	died	cancer	other	trt	time	logb	logp
1	1	1.095	1	1	0	1	0.000	2.674	2.501
2	1	NA	NA	NA	NA	1	0.526	3.059	2.416
3	3	2.771	1	0	1	1	0.000	0.336	2.485
4	3	NA	NA	NA	NA	1	0.482	0.095	2.485
5	3	NA	NA	NA	NA	1	0.997	0.405	2.485
6	3	NA	NA	NA	NA	1	2.034	0.588	2.588
7	7	6.848	0	0	0	0	0.000	0.000	2.272
8	7	NA	NA	NA	NA	0	1.073	0.182	2.370
9	7	NA	NA	NA	NA	0	1.492	-0.223	2.370
10	7	NA	NA	NA	NA	0	2.081	0.000	2.332
11	7	NA	NA	NA	NA	0	3.083	0.182	2.389
12	7	NA	NA	NA	NA	0	4.077	0.182	2.434
13	7	NA	NA	NA	NA	0	6.193	0.336	2.485

# Primary biliary cirrhosis data

	id	stime	died	cancer	other	trt	time	logb	logp
1	1	1.095	1	1	0	1	0.000	2.674	2.501
2	1	NA	NA	NA	NA	1	0.526	3.059	2.416
3	3	2.771	1	0	1	1	0.000	0.336	2.485
4	3	NA	NA	NA	NA	1	0.482	0.095	2.485
5	3	NA	NA	NA	NA	1	0.997	0.405	2.485
6	3	NA	NA	NA	NA	1	2.034	0.588	2.588
7	7	6.848	0	0	0	0	0.000	0.000	2.272
8	7	NA	NA	NA	NA	0	1.073	0.182	2.370
9	7	NA	NA	NA	NA	0	1.492	-0.223	2.370
10	7	NA	NA	NA	NA	0	2.081	0.000	2.332
11	7	NA	NA	NA	NA	0	3.083	0.182	2.389
12	7	NA	NA	NA	NA	0	4.077	0.182	2.434
13	7	NA	NA	NA	NA	0	6.193	0.336	2.485

# Primary biliary cirrhosis data

	id	stime	died	cancer	other	trt	time	logb	logp
1	1	1.095	1	1	0	1	0.000	2.674	2.501
2	1	NA	NA	NA	NA	1	0.526	3.059	2.416
3	3	2.771	1	0	1	1	0.000	0.336	2.485
4	3	NA	NA	NA	NA	1	0.482	0.095	2.485
5	3	NA	NA	NA	NA	1	0.997	0.405	2.485
6	3	NA	NA	NA	NA	1	2.034	0.588	2.588
7	7	6.848	0	0	0	0	0.000	0.000	2.272
8	7	NA	NA	NA	NA	0	1.073	0.182	2.370
9	7	NA	NA	NA	NA	0	1.492	-0.223	2.370
10	7	NA	NA	NA	NA	0	2.081	0.000	2.332
11	7	NA	NA	NA	NA	0	3.083	0.182	2.389
12	7	NA	NA	NA	NA	0	4.077	0.182	2.434
13	7	NA	NA	NA	NA	0	6.193	0.336	2.485

# Longitudinal biomarker - linear model

```
merlin(  
  model = logb ~ time,  
  timevar = "time",  
  family = "gaussian",  
  data = pbc)
```

# Longitudinal biomarker - restricted cubic splines

```
merlin(  
  model = logb ~ rcs(time, df = 3),  
  timevar = "time",  
  family = "gaussian",  
  data = pbc)
```

# Longitudinal biomarker - random intercept

```
merlin(  
  model = logb ~ rcs(time, df = 3) + M1[id]*1,  
  level = "id",  
  timevar = "time",  
  family = "gaussian",  
  data = pbc)
```

# Longitudinal biomarker - random slope

```
merlin(  
  model = logb ~ rcs(time, df = 3) + M1[id]*1  
           + time:M2[id]*1,  
  level = "id",  
  timevar = "time",  
  family = "gaussian",  
  data = pbc)
```

# Longitudinal biomarker - covariance structure

```
merlin(  
  model = logb ~ rcs(time, df = 3) + M1[id]*1  
           + time:M2[id]*1,  
  level = "id",  
  timevar = "time",  
  family = "gaussian",  
  covariance = "unstructured",  
  data = pbc)
```



# Joint longitudinal-survival model

A number of time-to-event models are available in `merlin`, including standard models such as

- ▶ Weibull
- ▶ Exponential
- ▶ Gompertz

# Joint longitudinal-survival model

A number of time-to-event models are available in `merlin`, including standard models such as

- ▶ Weibull
- ▶ Exponential
- ▶ Gompertz

Additionally a range of more flexible models are also available including

- ▶ Royston-Parmar - restricted cubic splines on log cumulative hazard scale
- ▶ Restricted cubic splines on log hazard scale

## Joint model - survival submodel

```
merlin(  
  model = list(logb ~ rcs(time, df = 3) + M1[id]*1  
               + time:M2[id]*1,  
               Surv(stime, died) ~ trt),  
  level = "id",  
  timevar = "time",  
  family = c("gaussian", "weibull"),  
  covariance = "unstructured",  
  data = pbc)
```

## Joint model - links

```
merlin(  
  model = list(logb ~ rcs(time, df = 3) + M1[id]*1  
               + time:M2[id]*1,  
               Surv(stime, died) ~ trt + EV[logb]),  
  level = "id",  
  timevar = c("time", "stime"),  
  family = c("gaussian", "weibull"),  
  covariance = "unstructured",  
  data = pbc)
```

## Joint model - links

```
merlin(  
  model = list(logb ~ rcs(time, df = 3) + M1[id]*1  
              + time:M2[id]*1,  
              Surv(stime, died) ~ trt + dEV[logb]),  
  level = "id",  
  timevar = c("time", "stime"),  
  family = c("gaussian", "weibull"),  
  covariance = "unstructured",  
  data = pbc)
```

## Joint model - links

```
merlin(  
  model = list(logb ~ rcs(time, df = 3) + M1[id]*1  
              + time:M2[id]*1,  
              Surv(stime, died) ~ trt + iEV[logb]),  
  level = "id",  
  timevar = c("time", "stime"),  
  family = c("gaussian", "weibull"),  
  covariance = "unstructured",  
  data = pbc)
```

## Joint model - links

```
merlin(  
  model = list(logb ~ rcs(time, df = 3) + M1[id]*1  
              + time:M2[id]*1,  
              Surv(stime, died) ~ trt + EV[logb]  
              + dEV[logb] + iEV[logb]),  
  level = "id",  
  timevar = c("time", "stime"),  
  family = c("gaussian", "weibull"),  
  covariance = "unstructured",  
  data = pbc)
```

## Joint model - links

```
merlin(  
  model = list(logb ~ rcs(time, df = 3) + M1[id]*1  
               + time:M2[id]*1,  
               Surv(stime, died) ~ trt + EV[logb]),  
  level = "id",  
  timevar = c("time", "stime"),  
  family = c("gaussian", "weibull"),  
  covariance = "unstructured",  
  data = pbc)
```



# Joint model - time dependent effects

```
merlin(  
  model = list(logb ~ rcs(time, df = 3) + M1[id]*1  
               + time:M2[id]*1,  
               Surv(stime, died) ~ trt + EV[logb]  
               + trt:fp(stime, powers = c(0))),  
  level = "id",  
  timevar = c("time", "stime"),  
  family = c("gaussian", "weibull"),  
  covariance = "unstructured",  
  data = pbc)
```

## Joint model - competing risks

```
merlin(  
  model = list(logb ~ rcs(time, df = 3) + M1[id]*1  
              + time:M2[id]*1,  
              Surv(stime, cancer) ~ trt + EV[logb]  
              + trt:fp(stime, powers = c(0)),  
              Surv(stime, other) ~ trt + dEV[logb]  
              + rcs(stime, df = 3, log = T)),  
  level = "id",  
  timevar = c("time", "stime"),  
  family = c("gaussian", "weibull", "rp"),  
  covariance = "unstructured",  
  data = pbc)
```

## Joint model - competing risks

```
merlin(  
  model = list(logb ~ rcs(time, df = 3) + M1[id]*1  
              + time:M2[id]*1,  
              Surv(stime, cancer) ~ trt + EV[logb]  
              + trt:fp(stime, powers = c(0)),  
              Surv(stime, other) ~ trt + dEV[logb]  
              + rcs(stime, df = 3, log = T)),  
  level = "id",  
  timevar = c("time", "stime"),  
  family = c("gaussian", "weibull", "rp"),  
  covariance = "unstructured",  
  data = pbc)
```

## Joint model - multiple biomarkers

```
merlin(  
  model = list(logb ~ rcs(time, df = 3) + M1[id]*1  
              + time:M2[id]*1,  
              logp ~ rcs(time, df = 3) + M3[id]*1,  
              Surv(stime, cancer) ~ trt + EV[logb]  
              + trt:fp(stime, powers = c(0)),  
              Surv(stime, other) ~ trt + dEV[logb]  
              + rcs(stime, df = 3, log = T)),  
  level = "id",  
  timevar = c("time", "stime"),  
  family = c("gaussian", "gaussian", "weibull", "rp"),  
  covariance = "unstructured",  
  data = pbc)
```

## Joint model - multiple biomarkers

```
merlin(  
  model = list(logb ~ rcs(time, df = 3) + M1[id]*1  
              + time:M2[id]*1,  
              logp ~ rcs(time, df = 3) + M3[id]*1,  
              Surv(stime, cancer) ~ trt + EV[logb]  
                + EV[logp] + iEV[logp]  
              + trt:fp(stime, powers = c(0)),  
              Surv(stime, other) ~ trt + dEV[logb]  
                + rcs(stime, df = 3, log = T)),  
  level = "id",  
  timevar = c("time", "stime"),  
  family = c("gaussian", "gaussian", "weibull", "rp"),  
  covariance = "unstructured",  
  data = pbc)
```

## Joint model - final model

```
merlin(  
  model = list(logb ~ rcs(time, df = 3) + M1[id]*1  
              + time:M2[id]*1,  
              logp ~ rcs(time, df = 3) + M3[id]*1,  
              Surv(stime, cancer) ~ trt + EV[logb]  
              + EV[logp] + iEV[logp]  
              + trt:fp(stime, powers = c(0)),  
              Surv(stime, other) ~ trt + dEV[logb]  
              + rcs(stime, df = 3, log = T)),  
  level = "id",  
  timevar = c("time", "stime"),  
  family = c("gaussian", "gaussian", "weibull", "rp"),  
  covariance = "unstructured",  
  data = pbc)
```

## Clinically meaningful predictions

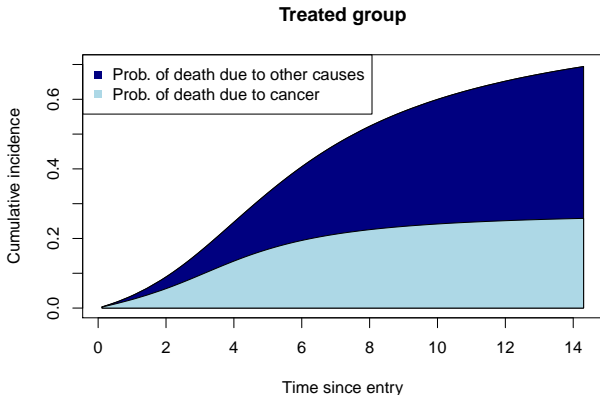
e.g. using the `predict` function we can calculate the **marginal cause-specific cumulative incidence function**, which tells us the probability of an event in the presence of competing events,

```
predict(model, stat = "cif", predmodel = 3,  
        type = "marginal")
```

Using the `marginal` option allows us to interpret them as population-average predictions.

# Clinically meaningful predictions

We can specify the level of a covariate, in order to investigate the effect of covariates (such as treatment) on predictions





# The future

- ▶ Dynamic risk prediction
- ▶ Timing of observations - informative observation process (Gasparini et al., 2018)
- ▶ `merlin` is very flexible, and hence it can be slow(er)
- ▶ Penalisation
- ▶ Scalability - sample weights
- ▶ Updates and tutorials are available on the website  
[www.mjcrowther.co.uk/software/merlin](http://www.mjcrowther.co.uk/software/merlin)

# References I

- Barrett, J. and Su, L. Dynamic predictions using flexible joint models of longitudinal and time-to-event data. *Statistics in Medicine*, 36(9):1447–1460, 2017.
- Crowther, M. J. merlin - a unified modelling framework for data analysis and methods development in Stata. *arXiv e-prints*, art. arXiv:1806.01615, Jun 2018.
- Crowther, M. J., Abrams, K. R., and Lambert, P. C. Joint modeling of longitudinal and survival data. *Stata Journal*, 13(1):165–184, 2013.
- Crowther, M. J., Andersson, T. M., Lambert, P. C., Abrams, K. R., and Humphreys, K. Joint modelling of longitudinal and survival data: incorporating delayed entry and an assessment of model misspecification. *Stat Med*, 35(7):1193–209, 2016.
- Gasparini, A., Abrams, K. R., Barrett, J. K., Major, R. W., Sweeting, M. J., Brunskill, N. J., and Crowther, M. J. Mixed effects models for healthcare longitudinal data with an informative visiting process: a Monte Carlo simulation study. *arXiv e-prints*, art. arXiv:1808.00419, Aug 2018.
- Hickey, G. L., Philipson, P., Jorgensen, A., and Kolamunnage-Dona, R. joineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC Medical Research Methodology*, 18(1):50, 2018.

## References II

- Krol, A., Ferrer, L., Pignon, J. P., Proust-Lima, C., Ducreux, M., Bouche, O., Michiels, S., and Rondeau, V. Joint model for left-censored longitudinal data, recurrent events and terminal event: Predictive abilities of tumor burden for cancer evolution with application to the ffcd 2000-05 trial. *Biometrics*, 72(3):907–16, 2016.
- Li, N., Elashoff, R. M., and Li, G. Robust joint modeling of longitudinal measurements and competing risks failure time data. *Biom J*, 51(1):19–30, 2009.
- Lin, H., McCulloch, C. E., and Mayne, S. T. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Stat Med*, 21(16):2369–82, 2002.
- Murtaugh, P. A., Dickson, E. R., Vandam, G. M., Malinchoc, M., Grambsch, P. M., Langworthy, A. L., and Gips, C. H. Primary biliary-cirrhosis - prediction of short-term survival based on repeated patient visits. *Hepatology*, 20(1):126–134, 1994.
- Phillipson, P., Sousa, I., Diggle, P. J., Williamson, P., Kolamunnage-Dona, R., Henderson, R., and Hickey, G. L. *joiner: Joint Modelling of Repeated Measurements and Time-to-Event Data*, 2018.
- Rizopoulos, D. The r package jmbayes for fitting joint models for longitudinal and time-to-event data using mcmc. *Journal of Statistical Software*, 72(7):1–46, 2016.

# References III

- Rizopoulos, D., Verbeke, G., Lesaffre, E., and Vanrenterghem, Y. A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics*, 64(2):611–9, 2008.
- Rondeau, V., Mazroui, Y., and Gonzalez, J. R. frailtypack: An r package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software*, 47(4):1–28, 2012.