

Extended multivariate generalised linear and non-linear mixed effects models: One model might fit all

International Society for Clinical Biostatistics
Melbourne, 27th August 2018

Michael J. Crowther

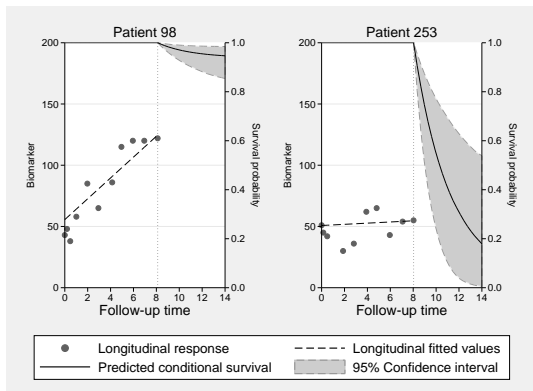
*Biostatistics Research Group,
Department of Health Sciences,
University of Leicester, UK,
michael.crowther@le.ac.uk
@Crowther.MJ*

Funding: MRC (MR/P015433/1)

the motivation

- More data → more questions
 - need for appropriate statistical modelling techniques, and implementations
- Growth in access to EHR
 - biomarkers < patients < GP practice area < geographical regions...
- The standard challenges
 - time-dependent effects, non-linear covariate effects
- The neglected challenges
 - Within-patient variability
 - Informative observations times

joint longitudinal-survival models



Linking via - current value, gradient, AUC, random effects...

joint longitudinal-survival models - extensions

- competing risks [1]
- different types of outcomes [2]
- multiple continuous outcomes [3]
- delayed entry [4]
- recurrent events and a terminal event [5]
- prediction [6]
- ...

joint longitudinal-survival models - software

- stjlm in Stata [7]
- gsem in Stata
- frailtypack in R [8]
- joineR in R [9]
- JM and JMBayes in R [10, 11]
- Many others...

the goal

- multiple outcomes of varying types
- measurement schedule can vary across outcomes
- any number of levels and random effects
- sharing and linking random effects between outcomes
- sharing functions of the expected value of other outcomes
- a reliable estimation engine
- easily extendable by the user
- ...

**a unified framework for data analysis and methods
development**

Mixed Effects Regression for Linear, Non-linear and user-defined models

`merlin`

- We need modelling frameworks that can accommodate all sorts of types of data
- We need usable software implementations

the example

- there's no equations in this talk
- there's 13 models
- each of them is applied to the same dataset
- most of them can be considered *new* models
- we can fit all of them with a single line of code

- data from 312 patients with PBC collected at the Mayo Clinic 1974-1984 [12]
- 158 randomised to receive D-penicillamine and 154 to placebo
- survival outcome is all-cause death, with 140 events observed
 - we're going to pretend we have competing causes of death - cancer and other causes
- 1945 measurements of serum bilirubin, among other things

the data

id	time	logb	prothr~n	trt	stime	cancer	other
1	0	2.674149	12.2	D-penicil	1.09517	1	0
1	.525682	3.058707	11.2	D-penicil	.	.	.
2	0	.0953102	10.6	D-penicil	14.1523	0	1
2	.498302	-.2231435	11	D-penicil	.	.	.
2	.999343	0	11.6	D-penicil	.	.	.
2	2.10273	.6418539	10.6	D-penicil	.	.	.
2	4.90089	.9555114	11.3	D-penicil	.	.	.
2	5.88928	1.280934	11.5	D-penicil	.	.	.
2	6.88588	1.435084	.	D-penicil	.	.	.
2	7.8907	1.280934	.	D-penicil	.	.	.
2	8.83255	1.526056	.	D-penicil	.	.	.

a model

```
merlin (logb          /// log serum bilirubin
        time          /// covariate
        ,             /// options
        family(gaussian) /// distribution
    )
```

a model

```
merlin (logb                                     /// log serum bilirubin
      time                                       /// covariate
      time#trt                                  /// interaction
      ,                                         /// options
      family(gaussian)                         /// distribution
      )                                         ///
```

a model

```
merlin (logb                                     /// log serum bilirubin
      time                                       /// covariate
      time#trt                                  /// interaction
      M1[id]@1                                  /// random intercept
      ,                                         /// options
      family(gaussian)                         /// distribution
      )                                         ///
```

a model

```
merlin (logb                                     /// log serum bilirubin
      time                                       /// covariate
      time#trt                                  /// interaction
      M1[id]@1                                  /// random intercept
      time#M2[id]@1                             /// random slope
      ,                                         /// options
      family(gaussian)                         /// distribution
    )
```

a model

```
merlin (logb                                     /// log serum bilirubin
      time                                       /// covariate
      time#trt                                  /// interaction
      M1[id]@1                                  /// random intercept
      time#M2[id]@1                             /// random slope
      ,                                         /// options
      family(gaussian)                         /// distribution
    )
(logb                                          ///
  rcs(time, df(3))                             /// prothrombin index
  , family(gamma)                               /// covariate
  )                                             /// distribution
  )                                             ///
```

a model

```

merlin (logb
    time
    time#trt
    M1[id]@1
    time#M2[id]@1
    ,
    family(gaussian)
)
(logb
    rcs(time, df(3))
    M3[id]@1
    , family(gamma)
)
    /// log serum bilirubin
    /// covariate
    /// interaction
    /// random intercept
    /// random slope
    /// options
    /// distribution
    ///
    /// prothrombin index
    /// covariate
    /// random effect
    /// distribution
    ///

```


a model

```

merlin (logb                                     /// log serum bilirubin
      time                                       /// covariate
      time#trt                                  /// interaction
      M1[id]@1                                  /// random intercept
      time#M2[id]@1                             /// random slope
      ,                                         /// options
      family(gaussian)                         /// distribution
    )
(logb                                           ///
  rcs(time, df(3))                             /// prothrombin index
  M3[id]@1                                       /// covariate
  , family(gamma)                               /// random effect
)                                                 /// distribution
,                                                 ///
covariance(unstructured)                       /// main options
                                                // vcv

```

a model

```

merlin (logb                                     /// log serum bilirubin
        time                                     /// covariate
        time#trt                                 /// interaction
        M1[id]@1                                 /// random intercept
        time#M2[id]@1                           /// random slope
        ,                                        /// options
        family(gaussian)                        /// distribution
    )
(logb                                           ///
    rcs(time, df(3))                             /// prothrombin index
    M3[id]@1                                     /// covariate
    , family(gamma)                             /// random effect
    )                                           /// distribution
,
covariance(unstructured)                       /// main options
redistribution(t) df(5)                         /// vcv
                                              /// re dist.

```

a model

```

merlin (logb                                     /// log serum bilirubin
      time                                       /// covariate
      time#trt                                  /// interaction
      M1[id]@1                                  /// random intercept
      time#M2[id]@1                             /// random slope
      ,                                         /// options
      family(gaussian)                         /// distribution
)
(logb                                          ///
  rcs(time, df(3))                             /// prothrombin index
  M3[id]@1                                     /// covariate
  , family(gamma)                             /// random effect
)                                              /// distribution
(stime trt                                     /// response + covariate
  , family(rp, df(3))                          /// distribution
      failure(other))                          /// event indicator
)                                              ///
,                                             /// main options
covariance(unstructured)                       /// vcv
redistribution(t) df(5)                         // re dist.

```

a model

```

merlin (logb                                     /// log serum bilirubin
      time                                       /// covariate
      time#trt                                  /// interaction
      M1[id]@1                                  /// random intercept
      time#M2[id]@1                             /// random slope
      ,                                         /// options
      family(gaussian)                         /// distribution
)
(logb                                          ///
  rcs(time, df(3))                            /// prothrombin index
  M3[id]@1                                     /// covariate
  , family(gamma)                             /// random effect
)                                              /// distribution
(stime trt                                     /// response + covariate
  dEV[logb] EV[pro]                           /// associations
  , family(rp, df(3))                         /// distribution
      failure(other))                        /// event indicator
)                                              ///
,                                             /// main options
covariance(unstructured)                     /// vcv
redistribution(t) df(5)                       /// re dist.

```

a model

```

merlin (logb                                     /// log serum bilirubin
      time                                     /// covariate
      time#trt                                 /// interaction
      M1[id]@1                                 /// random intercept
      time#M2[id]@1                           /// random slope
      ,                                        /// options
      family(gaussian)                       /// distribution
)
(logb                                          ///
  rcs(time, df(3))                          /// prothrombin index
  M3[id]@1                                   /// covariate
  , family(gamma)                           /// random effect
)                                             /// distribution
(stime trt                                   /// response + covariate
  trt#fp(stime, power(0))                   /// tde
  dEV[logb] EV[pro]                         /// associations
  , family(rp, df(3))                       /// distribution
      failure(other))                       /// event indicator
)                                             ///
,                                             /// main options
covariance(unstructured)                   /// vcV
redistribution(t) df(5)                     /// re dist.

```

a model

```

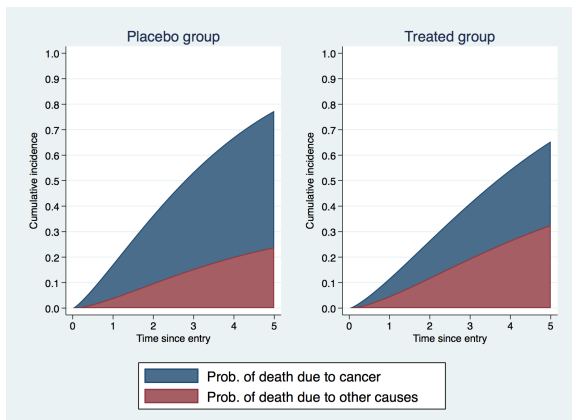
merlin (logb time time#trt M1[id]@1          /// model 1
        time#M2[id]@1 ,                      ///
        family(gaussian)                    ///
    )                                         ///
    (logb rcs(time, df(3)) M3[id]@1          /// model 2
      , family(gamma)                       ///
    )                                         ///
    (stime trt                               ///
      trt#fp(stime, power(0))                /// model 3 - cause 1
      dEV[logb] EV[pro]                      /// tde
      , family(rp, df(3))                   /// distribution
        failure(other))                    /// event indicator
    )                                         ///
    (stime trt                               /// model 4 - cause 2
      trt#rcs(stime, df(3) log)              /// tde
      EV[logb] iEV[pro]                     /// associations
      , family(weibull,                     /// distribution
        failure(cancer))                   /// event indicator
    )                                         ///
    ,                                         ///
    covariance(unstructured)

```

predictions

`predict cif1, cif marginal outcome(3) at(trt 0)`

`predict cif1, cif marginal outcome(4) at(trt 0)`



a user-defined model

```
real matrix gauss_logl(gml)
{
    y          = merlin_util_depvar(gml)           // dep. var.
    linpred    = merlin_util_xzb(gml)             // lin. pred.
    sdre       = exp(merlin_util_ap(gml,1))       // anc. param.
    return(lnnormalden(y,linpred,sdre))          // logl
}

merlin (logb ... , family(user, llfunction(gauss_logl) nap(1)))
...
...
...
```


stuff I didn't show

- random effects at arbitrary levels - `M4[centre>id]@1`
- B-splines - `bs(time, df(3) order(4))`
- `d2EV[]`, `?XB[]`
- `linterval(varname)` - interval censoring
- `ltruncated(varname)` - left-truncation
- 9 (so far) other inbuilt families, e.g. beta, ologit
- `bhazard(varname)` - relative survival
- `mf(func_name)` - user-defined element function

summary

- merlin can do a lot of things, hopefully in a usable way
- merlin is easily extended
- merlin's prediction tools are arguably the most important thing
- software is the future

the future

- dynamic risk prediction - predictions are a key focus of the `merlin` engine
- timing of observations
 - Informative observation process [13]
- penalisation
- `merlin` is very general, and hence it can be slow(er)
- scalability - sample weights (Emma Martin)
- updates and tutorials here:

www.mjcrowther.co.uk/software/merlin

the code

- `ssc install merlin` in Stata
- `merlin` also went on CRAN last week

the papers

- Crowther MJ. Extended multivariate generalised linear and non-linear mixed effects models.
<https://arxiv.org/abs/1710.02223>
- Crowther MJ. `merlin` - a unified framework for data analysis and methods development in Stata.
<https://arxiv.org/abs/1806.01615>

References I

- [1] Li N, Elashoff RM, Li G. Robust joint modeling of longitudinal measurements and competing risks failure time data. *Biom J* Feb 2009; **51**(1):19–30, doi:10.1002/bimj.200810491. URL <http://dx.doi.org/10.1002/bimj.200810491>.
- [2] Rizopoulos D, Verbeke G, Lesaffre E, Vanrenterghem Y. A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics* 2008; **64**(2):pp. 611–619. URL <http://www.jstor.org/stable/25502097>.
- [3] Lin H, McCulloch CE, Mayne ST. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Stat Med* Aug 2002; **21**(16):2369–2382, doi:10.1002/sim.1179. URL <http://dx.doi.org/10.1002/sim.1179>.
- [4] Crowther MJ, Andersson TML, Lambert PC, Abrams KR, Humphreys K. Joint modelling of longitudinal and survival data: incorporating delayed entry and an assessment of model misspecification. *Statistics in medicine* 2016; **35**(7):1193–1209.

References II

- [5] Król A, Ferrer L, Pignon JP, Proust-Lima C, Ducreux M, Bouché O, Michiels S, 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* 2016; **72**(3):907–916.
- [6] Barrett J, Su L. Dynamic predictions using flexible joint models of longitudinal and time-to-event data. *Statistics in Medicine* 2017; :n/a–n/adoi:10.1002/sim.7209. URL <http://dx.doi.org/10.1002/sim.7209>, sim.7209.
- [7] Crowther MJ, Abrams KR, Lambert PC, *et al.*. Joint modeling of longitudinal and survival data. *Stata J* 2013; **13**(1):165–184.
- [8] Król A, Mauguen A, Mazroui Y, Laurent A, Michiels S, Rondeau V. Tutorial in joint modeling and prediction: a statistical software for correlated longitudinal outcomes, recurrent events and a terminal event. *arXiv preprint arXiv:1701.03675* 2017; .
- [9] Philipson P, Sousa I, Diggle P, Williamson P, Kolamunnage-Dona R, Henderson R. *joineR - Joint Modelling of Repeated Measurements and Time-to-Event Data* 2012. URL <http://cran.r-project.org/web/packages/joineR/index.html>.

References III

- [10] Rizopoulos D. JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data. *J Stat Softw* 7 2010; **35**(9):1–33. URL <http://www.jstatsoft.org/v35/i09>.
- [11] Rizopoulos D. Jmbayes: joint modeling of longitudinal and time-to-event data under a bayesian approach 2015.
- [12] Murtaugh P, Dickson E, Van Dam M G Malincho, Grambsch P. Primary biliary cirrhosis: Prediction of short-term survival based on repeated patient visits. *Hepatology* 1994; **20**:126–134.
- [13] Gasparini A, Abrams KR, Barrett JK, Major RW, Sweeting MJ, Brunskill NJ, Crowther MJ. Mixed effects models for healthcare longitudinal data with an informative visiting process: a monte carlo simulation study ; .