

# Multilevel mixed effects parametric survival analysis

Michael J. Crowther  
Department of Health Sciences  
University of Leicester, UK  
michael.crowther@le.ac.uk

**Abstract.** With the release of Stata 14 came the `mestreg` command to fit multilevel mixed effects parametric survival models, assuming normally distributed random effects, estimated with maximum likelihood utilising Gaussian quadrature. In this article, I present the user written `stmixed` command, which serves as both an alternative and a complimentary program for the fitting of multilevel parametric survival models, to `mestreg`. The key extensions include incorporation of the flexible parametric Royston-Parmar survival model, and the ability to fit multilevel relative survival models. The methods are illustrated with a commonly used dataset of patients with kidney disease suffering recurrent infections, and a simulated example, illustrating a simple approach to simulating clustered survival data using `survsim` (Crowther and Lambert (2012, 2013)).

**Keywords:** `st0001`, `stmixed`, clustered survival data, multilevel models, adaptive Gauss-Hermite quadrature, relative survival

## 1 Introduction

Clustered survival data is often observed in a variety of settings. Within medical research, a common example is the analysis of recurrent event data, where individual patients can experience the event of interest multiple times throughout the follow-up period, and we wish to account for the inherent correlation within patients using a frailty term (Gutierrez (2002)).

In the field of meta-analysis, the individual patient data (IPD) meta-analysis of survival data is growing in use, as this form of analysis is recognised as the gold standard approach (Simmonds et al. (2005)). Analysing the IPD simultaneously within a hierarchical structure, allows direct adjustment for confounders and incorporation of non-proportional hazards in covariate effects (Tudur-Smith et al. (2005); Crowther et al. (2012, 2014)). Often a random treatment effect is assumed to account for heterogeneity present in treatment effects across the pooled trials.

The analysis of IPD within a one-stage meta-analysis draws parallels with the analysis of multi-centre clinical trial data (Komárek et al. (2007)). Despite random treatment effects being rarely assumed in this context, due to much greater control over inclusion/exclusion criteria, multilevel survival models can be used to assess the presence of heterogeneity between centres. Yamaguchi et al. (2002) and Legrand et al. (2005) have considered centre and treatment-by-centre effects to be random in this context.

A further area of interest is relative survival. Particularly prevalent in cancer survival studies, relative survival allows us to model the excess mortality associated with a diseased population compared to that of the general population (Dickman et al. (2004)). Such data often exhibits a hierarchical structure, with patients nested within geographical regions such as counties. Patients living in the same area may share unobserved characteristics, such as environmental aspects or medical care access (Charvat et al. (2016)).

With the release of Stata 14 came the `mestreg` command to fit multilevel mixed effects parametric survival models, assuming normally distributed random effects, estimated with maximum likelihood utilising Gaussian quadrature. In this article, I describe the user written `stmixed` command, first presented at the 2013 UK Stata meeting (Crowther (2013)), which serves as both an alternative and a complimentary program for the fitting of multilevel parametric survival models, to `mestreg`. The key extensions include incorporation of the flexible parametric Royston-Parmar survival model, and the ability to fit multilevel relative survival models. However, unlike `mestreg`, `stmixed` is currently limited to 2-level models.

The use of parametric models is growing in medical research, with the benefits of a parametric approach being extolled in a wide range of clinical areas (Miladinovic et al. (2012); Rooney et al. (2013)). Through direct modelling of the baseline hazard function, we can obtain clinically useful measures of absolute risk (King et al. (2012)). Parametric survival models available include the exponential, Weibull and Gompertz proportional hazards models, the flexible parametric Royston-Parmar model (Royston and Parmar (2002); Royston and Lambert (2011)), and the log-logistic, log normal and generalised gamma accelerated failure time model. The random effects are assumed to be normally distributed on the scale of the linear predictor, for example, the log hazard scale for proportional hazards models. The models are estimated using maximum likelihood utilising Gaussian quadrature. Each of the survival models available can be used in a relative survival modelling framework, except the log-logistic, log normal and generalised gamma accelerated failure time models.

The article is arranged as follows. Section 2 describes the multilevel parametric survival framework, and derives the likelihood used to estimate the models, including the extension to relative survival models. Section 3 details the model syntax of `stmixed`, describing the available options, and Section 4 describes the post-estimation tools available. I illustrate the command in Section 5 with a dataset of patients with kidney disease who are followed up for recurrent infection at the catheter insertion point, and show how to simulate clustered survival data using the `survsim` command, representing an IPD meta-analysis scenario, with a random treatment effect. I conclude the paper in Section 6.

## 2 Multilevel mixed effects parametric survival models

I begin with some notation. I define  $i = 1, \dots, N$  clusters (e.g. trials or centres), with each cluster having  $j = 1, \dots, n_i$  patients. Let  $S_{ij}$  be the true survival time of the  $j^{\text{th}}$  patient in the  $i^{\text{th}}$  cluster,  $T_{ij} = \min(S_{ij}, C_{ij})$  the observed survival time, with  $C_{ij}$  the censoring time. I define an event indicator  $d_{ij}$ , which takes the value of 1 if  $S_{ij} \leq C_{ij}$  and 0 otherwise.

### 2.1 Proportional hazards parametric survival models

I define the proportional hazards mixed effect survival model

$$h_{ij}(t) = h_0(t) \exp [x_{ij}^T \beta + z_{ij}^T b_j] \quad (1)$$

where  $h_0(t)$  is the baseline hazard function of either the exponential, Weibull or Gompertz distributions. I define design matrices  $x_{ij}$  and  $z_{ij}$  for the fixed ( $\beta$ ) and random ( $b_j$ ) effects, respectively. I assume the random effects follow a multivariate normal distribution, with  $b_j \sim N(0, \Sigma)$ . If  $z_j = 1$ , then Equation (1) reduces to a shared frailty model, such as those available in `streg`, albeit with a different choice of frailty distribution.

### 2.2 Flexible parametric model

An alternative to the standard proportional hazards distributions is the flexible parametric model of Royston and Parmar (2002), modelled on the cumulative hazard scale, which has recently been extended to incorporate random effects by Crowther et al. (2014). Therefore, we have

$$H_{ij}(t) = H_0(t) \exp [x_{ij}^T \beta + z_{ij}^T b_j] \quad (2)$$

where  $H_0(t)$  is the cumulative baseline hazard function. The spline basis for this specification is derived from the log cumulative hazard function of a Weibull proportional hazards model. The linear relationship with log time is relaxed through the use of restricted cubic splines. Further details can be found in Royston and Parmar (2002) and Royston and Lambert (2011). We can therefore write a restricted cubic spline function of  $\log(t)$ , with knots  $\mathbf{k}_0$ , as  $s\{\log(t)|\gamma, \mathbf{k}_0\}$ . For example, with  $K$  knots and letting  $x = \log(t)$ , a restricted cubic spline function can be expressed as:

$$s(x) = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \dots + \gamma_{K-1} z_{K-1} \quad (3)$$

and

$$z_1 = x \quad (4)$$

$$z_p = (x - k_p)_+^3 - \kappa_p (x - k_p)_+^3 - (1 - \kappa_p) (x - k_K)_+^3 \quad p = 2, \dots, K - 1 \quad (5)$$

This is now substituted for the [log] cumulative baseline hazard in Equation (2).

$$\log\{H_{ij}(t)\} = \eta_{ij} = s\{\log(t)|\gamma, \mathbf{k}_0\} + x_{ij}^T \beta + z_{ij}^T b_j \quad (6)$$

Transforming to the hazard and survival scales, we have

$$h_{ij}(t) = \left[ \frac{1}{t} \frac{ds\{\log(t)|\gamma, \mathbf{k}_0\}}{d\log(t)} \right] \exp(\eta_{ij}), \quad S_{ij}(t) = \exp[-\exp(\eta_{ij})] \quad (7)$$

In this framework we assume proportional cumulative hazards; however, this in fact implies proportional hazards, as in the models described in Section 2.1.

### 2.2.1 Non-proportional (cumulative) hazards

Relaxing the assumption of proportional hazards allows the investigation of whether the effect of a covariate changes with time. Termed non-proportional hazards, or time-dependent effects, the occurrence of which is commonplace in the analysis of survival data. Examples include treatment effects which vary over time Mok et al. (2009), and in registry based studies, where follow-up can be substantial, covariate effects have been found to vary Lambert et al. (2011).

Non-proportional cumulative hazards have been incorporated into the flexible parametric framework by Royston and Parmar (2002), achieved by interacting covariates with spline functions of log time and including them in the linear predictor Lambert and Royston (2009). This provides even greater flexibility in capturing complex effects, not restricted to linear functions of time. Equation (6) becomes

$$\log\{H_{ij}(t)\} = \eta_{ij} = s\{\log(t)|\gamma, k_0\} + x_{ij}^T\beta + z_{ij}^T b_j + \sum_{p=1}^P s\{\log(t)|\delta_p, k_p\} x_{ijp} \quad (8)$$

Each time-dependent effect can have varying number of spline terms, depending on the number of knots,  $\mathbf{k}_p$ . Quite often, fewer knots can be used than chosen for the baseline cumulative hazard function,  $s\{\log(t)|\gamma, k_0\}$ .

## 2.3 Accelerated failure time models

Mixed effects can be incorporated into the accelerated failure time framework. In `stmixed`, three choices are available, namely, the log-logistic, log-normal, and generalised gamma. The parameterisations are consistent with those in `mestreg`, with random effects incorporated into the same linear predictor as the fixed effects.

### 2.3.1 Log-logistic

I define the survival function for the log-logistic accelerated failure time model

$$S(t) = \{1 + (\lambda t)^{1/\gamma}\}^{-1} \quad (9)$$

and the density function

$$\frac{\lambda^{1/\gamma} t^{1/\gamma-1}}{\gamma \{1 + (\lambda t)^{1/\gamma}\}^2} \quad (10)$$

where

$$\log(\lambda) = -(x_{ij}^T\beta + z_{ij}^T b_j) \quad (11)$$

and  $\gamma$  treated as an ancillary parameter.

### 2.3.2 Log-normal

I define the survival function for the log-normal accelerated failure time model

$$S(t) = 1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right) \quad (12)$$

and the density function

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\{\log(t) - \mu\}^2\right] \quad (13)$$

where

$$\mu = x_{ij}^T\beta + z_{ij}^T b_i$$

and  $\sigma$  treated as an ancillary parameter.

### 2.3.3 Generalised gamma

Finally, I define the survival function for the generalised gamma accelerated failure time model

$$S(t) = \begin{cases} 1 - I(\gamma, u), & \text{if } \kappa > 0 \\ 1 - \Phi(z), & \text{if } \kappa = 0 \\ I(\gamma, u), & \text{if } \kappa < 0 \end{cases} \quad (14)$$

and the density function

$$f(t) = \begin{cases} 1 - I(\gamma, u), & \text{if } \kappa > 0 \\ 1 - \Phi(z), & \text{if } \kappa = 0 \end{cases} \quad (15)$$

where  $\gamma = |\kappa|^{-2}$ ,  $z = \text{sign}(\kappa)\{\log(t) - \mu\}/\sigma$ ,  $u = \gamma \exp(|\kappa|z)$ ,  $\Phi(z)$  is the standard normal cumulative distribution function, and  $I(a, x)$  is the incomplete gamma function. The model is parameterised as follows

$$\mu = x_{ij}^T\beta + z_{ij}^T b_i$$

with  $\kappa$  and  $\sigma$  treated as ancillary parameters.

An accelerated failure time framework can be seen as an attractive alternative to proportional hazards models, when interest is more specifically related to the time of an event occurring, and factors that increase or decrease this time, which may be of more clinical use.

## 2.4 Relative survival

Relative survival allows us to model the excess mortality associated with a diseased population compared to that of the general population, matched appropriately on the main factors associated with patient survival, such as age and gender (Dickman et al. (2004)). For a recent extensive description and implementation of the tools available for relative survival analysis in Stata, along

with a description of the differing approaches, I refer the reader to Dickman and Coviello (2015) and references therein.

Concentrating on applications of relative survival to cancer settings, the data generally comes from population based registries. Such data often exhibits a hierarchical structure, with patients nested within geographical regions such as counties. Patients living in the same area may share unobserved characteristics, such as environmental aspects or medical care access. Charvat et al. (2016) recently described a flexible relative survival model allowing a random intercept, with the baseline log hazard function modelled with B-splines, or restricted cubic splines. In this article, I extend the multilevel Royston-Parmar survival model described in Crowther et al. (2014), which uses restricted cubic splines on the log cumulative hazard scale, to the relative survival setting, further allowing any number of random effects, including random coefficients. Modelling on the log cumulative hazard scale avoids the need for numerical integration which is required when modeling on the log hazard scale with splines, and will generally require fewer spline terms than when modelling on the log hazard scale.

Within a multilevel modeling framework, we therefore define the total hazard at the time since diagnosis,  $t$ , for the  $j^{th}$  patient in the  $i^{th}$  cluster (area) to be  $h_{ij}(t)$ , with

$$h_{ij}(t) = h_{ij}^*(t) + \lambda_{ij}(t)$$

where

- $h_{ij}^*(t)$  is the expected mortality for the  $j^{th}$  patient in the  $i^{th}$  cluster (area)
- $\lambda_{ij}(t)$  is the excess mortality for the  $j^{th}$  patient in the  $i^{th}$  cluster (area)

and we model

$$\lambda_{ij}(t) = \lambda_0(t) \exp(X_{ij}^T \beta + Z_{ij}^T b_i)$$

where  $\lambda_0(t)$  is the baseline hazard function, with choices available including the exponential, Weibull or Gompertz.

Alternatively, we can model on the (log) cumulative excess hazard scale, using the flexible parametric model of Royston and Parmar (2002), where I define the total cumulative hazard at the time since diagnosis,  $t$ , for the  $j^{th}$  patient in the  $i^{th}$  cluster (area) to be  $H_{ij}(t)$ , with

$$H_{ij}(t) = H_{ij}^*(t) + \Lambda_{ij}(t)$$

where

- $H_{ij}^*(t)$  is the expected cumulative mortality for the  $j^{th}$  patient in the  $i^{th}$  cluster (area)
- $\Lambda_{ij}(t)$  is the excess cumulative mortality for the  $j^{th}$  patient in the  $i^{th}$  cluster (area)

and we model

$$\Lambda_{ij}(t) = \Lambda_0(t) \exp(X_{ij}^T \beta + Z_{ij}^T b_i)$$

where  $\Lambda_0(t)$  is the baseline cumulative hazard function, modeled with restricted cubic splines.

## 2.5 Likelihood and estimation

Defining the likelihood for the  $i^{th}$  cluster under the mixed effects framework, we have

$$L_i = \int_{-\infty}^{\infty} \left[ \prod_{j=1}^{n_i} p(T_{ij}, d_{ij} | b_i, \theta) \right] p(b_i | \theta) db_i \quad (16)$$

with parameter vector,  $\theta$ . Under a hazard scale model

$$p(T_{ij}, d_{ij} | b_i, \theta) = h(T_{ij})^{d_{ij}} \exp \left[ - \int_0^{T_{ij}} h(T_{ij}) \right] \quad (17)$$

with  $h(T_{ij})$  defined in Equation (1). Under the flexible parametric survival model

$$p(T_{ij}, d_{ij} | b_i, \theta) = \left[ \left\{ \frac{1}{T_{ij}} \frac{ds\{\log(T_{ij})|\gamma, k_0\}}{d\log(T_{ij})} \right\} \exp(\eta_{ij}) \right]^{d_{ij}} \exp\{-\exp(\eta_{ij})\} \quad (18)$$

and under an accelerated failure time model we have

$$p(T_{ij}, d_{ij} | b_i, \theta) = f(T_{ij})^{d_{ij}} \times S(T_{ij})^{1-d_{ij}} \quad (19)$$

with  $f(T_{ij})$  and  $S(T_{ij})$  defined in Section 2.3. Finally, we assume the random effects follow a multivariate normal distribution

$$p(b_i | \theta) = (2\pi|\Sigma|)^{-q/2} \exp \left\{ - \frac{b_i' \Sigma^{-1} b_i}{2} \right\}, \quad (20)$$

with variance-covariance matrix,  $\Sigma$ , and  $q$  the number of random effects. The (possibly multi-dimensional) integral in Equation (16) is analytically intractable, requiring numerical techniques to evaluate. In this article, we propose to use  $m$ -point mean-variance Gauss-Hermite quadrature (Pinheiro and Bates (1995); Rabe-Hesketh et al. (2002); Liu and Huang (2008)). The default estimation method is entirely consistent with that of in built Stata 12 commands, such as `xtmepoisson` and `xtmelogit`. This involves a two-stage procedure, where in the first stage, the appropriate fixed effects model is fitted, with variance and covariance parameters given starting values of 0 and 1, respectively, combined with two iterations of non-adaptive Gauss-Hermite quadrature are conducted to improve starting values, followed by the second stage which switches to fully adaptive Gauss-Hermite quadrature until convergence. Alternatively, only the first stage of non-adaptive quadrature can be used until standard convergence criteria is met.

## 2.6 Relative survival likelihood

The adaption to the likelihood in Equation (16) to turn it into a relative survival model is relatively simple. All that is needed is the expected mortality rate at

each event time, which are usually obtained from national/regional life tables. Under a hazard scale model, Equation (17) becomes,

$$p(T_{ij}, d_{ij} | b_i, \theta) = [h^*(T_{ij}) + \lambda(T_{ij})]^{d_{ij}} \exp \left[ - \int_0^{T_{ij}} \lambda(T_{ij}) \right] \quad (21)$$

and under a cumulative hazard scale model, Equation (22) becomes,

$$p(T_{ij}, d_{ij} | b_i, \theta) = \left[ \left\{ \frac{1}{T_{ij}} \frac{ds\{\log(T_{ij})|\gamma, k_0\}}{d \log(T_{ij})} \right\} \exp(\eta_{ij}) \right]^{d_{ij}} \exp \{-\exp(\eta_{ij})\} \quad (22)$$

Multilevel relative survival models are not currently available in the AFT framework.

## 3 The stmixed command

### 3.1 Syntax

```
stmixed [ fe_equation ] || re_equation [ , distribution(string) df(#)
      dftvc(dflist) knots(numlist) knotstvc(knotslist) tvc(varlist)
      bhazard(varname) keepcons level(#) showinit reiterate(#)
      retolerance(#) gh(#) initmatrix(matname) nonadapt showadapt
      variance vcvinitmatrix(matname) maximize_options ]
```

where the syntax of *fe\_equation* is

```
[ varlist ] [ if ] [ in ] [ , noconstant ]
```

and the syntax of *re\_equation* is

```
levelvar: [ varlist ] [ , covariance(vartype) noconstant ]
```

### 3.2 Options

#### 3.2.1 Model

noconstant suppresses the constant term in either the fixed or random effect equation, as appropriate.

covariance(*vartype*) specifies the variance-covariance structure of the random effects. covariance(independent) specifies a distinct variance for each random effect, with all covariances zero. covariance(exchangeable) specifies equal variances for all random effects, and one common pairwise covariance. covariance(identity) specifies equal variances for all random effects, with all covariances zero.

covariance(unstructured) specifies that all variances and covariances are distinctly estimated, which is the default.

distribution(*string*) specifies the survival distribution. Choices include exponential, gompertz and weibull proportional hazards models, fpm, the Royston-Parmar flexible parametric model, and loglogistic, lognormal and gamma accelerated failure time models.



`df(#)` specifies the degrees of freedom for the restricted cubic spline function used for the log baseline cumulative hazard under a flexible parametric survival model. `#` must be between 1 and 10, but usually a value between 1 and 5 is sufficient. The `knots()` option is not applicable if the `df()` option is specified.

`dftvc(dflist)` gives the degrees of freedom for time-dependent effects in `numlist`. With 1 degree of freedom a linear effect of log time is fitted. If there is more than one time-dependent effect and different degrees of freedom are requested for each time-dependent effect then the following syntax applies: `dftvc(x1:3 x2:2 1)`. This will use 3 degrees of freedom for `x1`, 2 degrees of freedom for `x2` and 1 degree of freedom for all remaining time-dependent effects.

`knots(numlist)` specifies knot locations for the baseline distribution function under a flexible parametric survival submodel, as opposed to the default locations set by `df()`. Note that the locations of the knots are placed on the standard time scale. However, the scale used by the restricted cubic spline function is always log time. Default knot positions are determined by the `df` option.

`knotstvc(knotslist)` defines `numlist knotslist` as the location of the interior knots for time-dependent effects. If different knots are required for different time-dependent effects the option is specified, for example, as follows: `knotstvc(x1 1 2 3 x2 1.5 3.5)`.

`tvv(varlist)` gives the name of the variables that are time-dependent when fitting a flexible parametric model. Time-dependent effects are fitted using restricted cubic splines. The degrees of freedom are specified using the `dftvc()` option.

`bhazard(varname)` is used when fitting multilevel relative survival models. `varname` gives the expected mortality rate at the time of death/censoring.

### 3.2.2 Reporting

`keepcons` do not drop constraints used by `ml` when fitting a flexible parametric model.

`level(#)` specifies the confidence level, with the default being that defined by `set level`.

`showinit` displays the output from the default initial values model fit, using `streg`.

### 3.2.3 Maximisation

`reiterate(#)` specifies the maximum number of iterations used when estimating the random effects to be used in adapting the Gaussian quadrature points; see the `retolerance()` option. The default is `reiterate(200)`. You should seldom have to use this option.

`retolerance(#)` specifies the convergence tolerance for the estimated random effects used by adaptive Gaussian quadrature. Gaussian quadrature points are adapted to be centered at the estimated random effects given a current

set of model parameters. Estimating these random effects is an iterative procedure, with convergence declared when the maximum relative change in the random effects is less than `retolerance()`. The default `retolerance()` is 1e-8. You should seldom have to use this option.

`gh(#)` specifies the number of Gauss-Hermite quadrature nodes used to evaluate the integrals over the random effects. Minimum number of quadrature points is 2. Default is 7.

`initmatrix(matname)` pass a matrix of initial values to `ml`, instead of the fixed effect models used to obtain starting values.

`nonadapt` use non-adaptive Gauss-Hermite quadrature to evaluate the likelihood. This will generally require a much higher number of nodes, `gh()`, to ensure accurate estimates and standard errors, resulting in much greater computation time.

`showadapt` display of the log-likelihood values under the sub-iterations used to assess convergence of the adaptive quadrature implemented at the beginning of each full Newton-Raphson iteration.

`vcvinitmatrix(matname)` pass a matrix of initial values for the variance-covariance parameters. They should be on the log(standard deviation) and tanh(correlation) scales.

*maximization\_options*: `difficult`, `technique(algorithm_spec)`, `iterate(#)`, `[no]log`, `trace`, `gradient`, `showstep`, `hessian`, `shownrtolerance`, `tolerance(#)`, `ltolerance(#)` `gtolerance(#)`, `nrtolerance(#)`, `nonrtolerance`, `from(init_specs)`; see [R] `maximize`. These options are seldom used, but the `difficult` option may be useful if there are convergence problems.

## 4 stmixed postestimation

### 4.1 Syntax for obtaining best linear unbiased predictions (BLUPs) of random effects, or the BLUPs' standard errors

```
predict {stub*|newvarlist} , {reffects | reses}
```

### 4.2 Syntax for obtaining other predictions

```
predict newvarname [if] [in] [, hazard survival cumhazard xb
    fitted at(varname #[varname #...]) ci stdp
    timevar(varname) zeros level(#) ]
```

## 4.3 Options

### 4.3.1 Survival

`hazard` calculates the predicted hazard. Default prediction, `xb`, includes only the fixed portion of the model. If `fitted` is specified then individual specific estimates of the random effects are included with the fixed portion of the model.

**survival** calculates each observation's predicted survival probability. Default prediction, **xb**, includes only the fixed portion of the model. If **fitted** is specified then individual specific estimates of the random effects are included with the fixed portion of the model.

**cumhazard** calculates the predicted cumulative hazard. Default prediction, **xb**, includes only the fixed portion of the model. If **fitted** is specified then individual specific estimates of the random effects are included with the fixed portion of the model.

### 4.3.2 Random effects

**reffects** calculates best linear unbiased predictions (BLUPs) of the random effects. You must specify *q* new variables, where *q* is the number of random effects terms in the model (or level). However, it is much easier to just specify **stub\*** and let Stata name the variables **stub1...stubq** for you.

**reses** calculates the standard errors of the best linear unbiased predictions (BLUPs) of the random effects. You must specify *q* new variables, where *q* is the number of random effects terms in the model (or level). However, it is much easier to just specify **stub\*** and let Stata name the variables **stub1...stubq** for you.

### 4.3.3 Subsidiary

**xb** specifies predictions based on fixed portion of the model.

**fitted** subject level predictions based on the fixed portion plus contributions of the predicted random effects.

**at**(*varname # [varname # ...]*) requests that the covariates specified by the listed *varname*(s) be set to the listed *#* values. For example, **at(x1 1 x3 50)** would evaluate predictions at  $x_1 = 1$  and  $x_3 = 50$ . This is a useful way to obtain out of sample predictions. Note that if **at()** is used together with zeros all covariates not listed in **at()** are set to zero. If **at()** is used without zeros then all covariates not listed in **at()** are set to their sample values. See also **zeros**.

**ci** calculate confidence interval and store in *newvarname\_lci* and *newvarname\_uci*.

**stdp** calculates the standard error of the fixed-portion linear prediction of the first *m1* equation.

**timevar**(*varlist*) defines the variable used as time in the predictions.

**zeros** sets all covariates to zero (baseline prediction).

**level**(*#*) specifies the confidence level, as a percentage, for confidence intervals. The default is as set by **set level**.

## 5 Examples

In this section, I illustrate the command in two areas of research, namely, recurrent events analysis, and the individual participant data meta-analysis of survival data.

## 5.1 Recurrent event data

I consider the commonly used `catheter` dataset consisting of 38 patients with kidney disease (McGilchrist and Aisbett (1991)). The outcome of interest is infection at the catheter insertion point, with our baseline being time of initial catheter insertion. Patients can experience up to two recurrences of infection, resulting in a total of 58 events. To show the basic syntax and use of `stmixed`, I fit a Weibull proportional hazards model with a normally distributed frailty

```
. webuse catheter, clear
(Kidney data, McGilchrist and Aisbett, Biometrics, 1991)
. stset time, fail(infect)
      failure event:  infect != 0 & infect < .
obs. time interval:  (0, time]
exit on or before:  failure
```

---

```
76 total obs.
0 exclusions
```

---

```
76 obs. remaining, representing
58 failures in single record/single failure data
7424 total analysis time at risk, at risk from t = 0
      earliest observed entry t = 0
      last observed exit t = 562
```

```
. stmixed age female || patient:, dist(weib)
Note: single-variable random-effects specification; covariance structure set to
> identity
Refining starting values:
Iteration 0: log likelihood = -104.84666
Iteration 1: log likelihood = -102.04333
Iteration 2: log likelihood = -98.903029
Performing gradient-based optimization:
Iteration 0: log likelihood = -98.924817
Iteration 1: log likelihood = -98.745856
Iteration 2: log likelihood = -98.744105
Iteration 3: log likelihood = -98.744112
Mixed effects survival regression          Number of obs. = 76
Panel variable: patient                    Number of panels = 38
Log-likelihood = -98.744112
```

	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
ln_lambda					
age	1.007329	.0137957	0.53	0.594	.9806494 1.034734
female	.1910468	.0999767	-3.16	0.002	.0685011 .5328217
_cons	.0073337	.007239	-4.98	0.000	.0010595 .050761
ln_p					
_cons	.2228571	.1387695	1.61	0.108	-.0491261 .4948402

---

Random effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
patient: Identity			
sd(_cons)	.9074878	.2654714	.5114876 1.610077

---

```
Survival submodel: Weibull proportional hazards model
Integration method: Adaptive Gauss-Hermite quadrature using 9 nodes
```

The default estimation routine proceeds with two full Newton-Raphson iterations using non-adaptive quadrature to refine the starting values, before switching to fully adaptive quadrature and proceeding until standard convergence criteria are met, which in this case took 3 iterations. By default the first `m1` equation is presented as hazard ratios. This can be switched off using the `nohr` option. We observe a hazard ratio of 0.191 (95% CI: 0.069, 0.533) for a female compared to a male of the same age, and a non-statistically significant age effect. The estimated frailty standard deviation is 0.907 (95% CI: 0.511, 1.610), indicating a highly heterogenous baseline hazard function.

As with all estimation routines which use numerical integration, it is important to check the consistency of results using an increasing number of quadrature points. This can be done, for example, using a `while` loop and increasing the points and refitting the model until the maximum relative difference between estimates using  $m$  and  $m + 1$  points is less than a specified tolerance, say  $1E-05$ :

```
. //store results with default (9) quadrature points
. mat est1 = e(b)
. //fit model with 10 points
. local gh = 10
. quietly stmixed age female || patient:, dist(weib) gh(`gh`)
. mat est2 = e(b)
. //continue until the maximum relative difference is < 1E-05
. while mreldif(est1,est2)>1E-05 {
2.     mat est1 = est2
3.     local gh = `gh` + 1
4.     quietly stmixed age female || patient:, dist(weib) gh(`gh`)
5.     mat est2 = e(b)
6. }
. di as txt "Converged with `gh` nodes"
Converged with 13 nodes
```

In this case, we found consistent estimates to 5 decimal places between 12 and 13 quadrature points.

### 5.1.1 Predictions

A variety of predictions can be obtained following the fitting of a model. We can obtain our predicted survival function with 95% confidence interval, based on the fixed portion of the model, for a female aged 45, through use of the `at()` option, as follows

```
. predict s1, survival ci at(age 45 female 1)
note: Confidence intervals calculated using Z critical values
```

Which can be plotted by

```
. twoway rarea s1_lci s1_uci _t, sort || line s1 _t, sort ylabel(,angle(h) form
```

```

> at(%2.1f) ///
>     xtitle("Follow-up time (days)") ytitle("Survival probability") ///
>     legend(order(2 "Predicted survival" 1 "95% CI") ring(0) pos(1) cols(1)
> ))

```

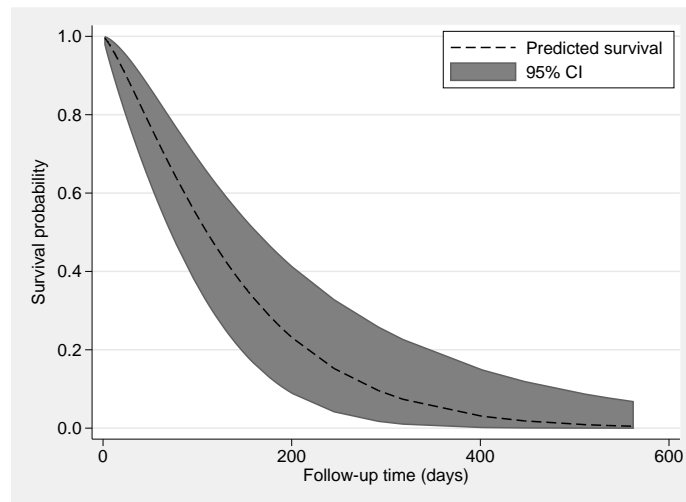


Figure 1: Predicted survival for a female, aged 45, based on the fixed portion of the model.

Predictions of the random effects can be obtained following the fit of a `stmixed` model. They can be used to assess the normality assumption of the random effects distribution, and also used to obtain cluster specific predictions. We obtain the best linear unbiased predictions (BLUPs) of the random effects using,

```
. predict refs1, reffects
```

We can then produce a quantile-normal plot to visually assess any deviations from normality, making sure to only plot one predicted value per cluster,

```
. bys patient: gen ind = _n==1
. qnorm refs1 if ind==1
```

Figure 5.1.1 indicates a reasonably normal approximation, with some indication of deviation in the left tail. With a sample size of only 38, this must be interpreted with caution.

## 5.2 Individual participant data meta-analysis of survival data

In this example I will illustrate a simple approach to simulating clustered survival data, in the setting of an IPD meta-analysis of survival data, through the use of the `survsim` command (Crowther and Lambert (2012, 2013)). I assume

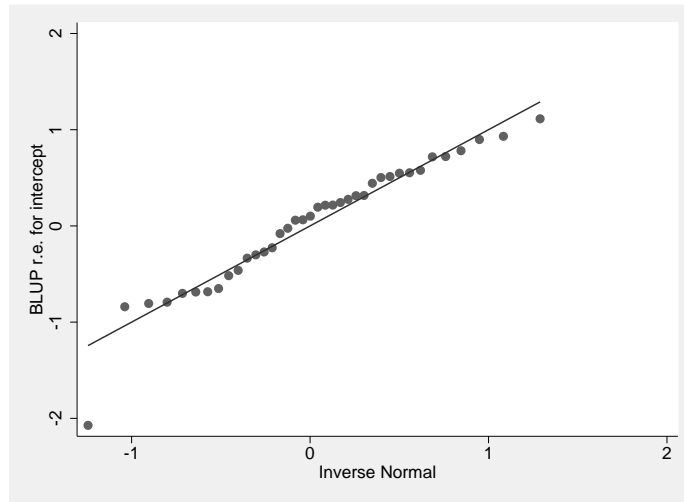


Figure 2: Plot of the quantiles of the BLUP predictions of the random frailty against quantiles of the normal distribution.

a scenario where I have data from 50 trials, each with 100 patients. Each trial compared a treatment to a control, with the probability of being assigned to each arm being 50%. I assume that the treatment effect for each trial comes from a normal distribution,  $N(-0.5, 0.5^2)$ , i.e. an average log hazard ratio of -0.5 (hazard ratio =  $\exp(-0.5) = 0.607$ ), with heterogeneity standard deviation of 0.5. I assume a Weibull baseline hazard function, with scale and shape parameter values of  $\lambda = 0.1$  and  $\gamma = 1.2$ , indicating 50.2% survival in the control group after 5 years, at which time administrative censoring is assumed.

```
. clear
. // Set seed for reproducibility
. set seed 278945
. // Assume we have 50 trials
. set obs 50
number of observations (_N) was 0, now 50
. // Generate a trial id variable
. gen trialid = _n
. // Generate trial specific treatments effect (log hazard ratio) from a normal
> distribution with mean 0.5, and std. dev. 0.5
. gen trteffect = rnormal(-0.5,0.5)
. // Assume 100 patients in each trial
. expand 100
(4,950 observations created)
. // Generate 0/1 patient level treatment group indicator
. gen trt = runiform()>0.5
. // Generate a variable containing patient specific treatment effects for use
> in simulation
. gen trteffectsim = trt*trteffect
```

```

. // Simulate survival times from a Weibull distribution, incorporating the ran
> dom treatment effect
. survsim stime died, dist(weibull) lambda(0.1) gamma(1.2) covariates(trteffect
> sim 1) maxt(5)
. stset stime, f(died)
      failure event: died != 0 & died < .
obs. time interval: (0, stime]
exit on or before: failure

```

---

```

5000 total observations
  0 exclusions

```

---

```

5000 observations remaining, representing
2131 failures in single-record/single-failure data
19602.039 total analysis time at risk and under observation
              at risk from t =          0
              earliest observed entry t =          0
              last observed exit t =          5

```

```

. stmixed trt || trialid: trt, nocons dist(weib)
Note: single-variable random-effects specification; covariance structure set to
> identity
Refining starting values:
Iteration 0: log likelihood = -5362.8282 (not concave)
Iteration 1: log likelihood = -5353.1551
Iteration 2: log likelihood = -5344.2178
Performing gradient-based optimization:
Iteration 0: log likelihood = -5344.341
Iteration 1: log likelihood = -5343.5106
Iteration 2: log likelihood = -5343.5091
Iteration 3: log likelihood = -5343.5091
Mixed effects survival regression          Number of obs. =          5000
Panel variable: trialid                   Number of panels =           50
Log-likelihood = -5343.5091

```

	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
ln_lambda						
trt	.551769	.0502547	-6.53	0.000	.4615624	.6596053
_cons	.1038053	.0046622	-50.44	0.000	.0950582	.1133574
ln_p						
_cons	.1761063	.020175	8.73	0.000	.1365641	.2156485

---

Random effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
trialid: Identity				
sd(trt)	.551641	.0687027	.4321613	.7041534

```

Survival submodel: Weibull proportional hazards model
Integration method: Adaptive Gauss-Hermite quadrature using 9 nodes

```

A key trick to note here is in the `survsim` command, I included the variable `trteffectsim` and assigned it a coefficient value of 1. This allows you to seamlessly incorporate random effects on covariates which are included in the linear



predictor, multiplied by a coefficient of 1. Then when we fit the model using `stmixed` I use the `trt` variable which indicates treatment group.

In the `stmixed` model fit, I enter `trt` as both a fixed and random effect, but using the `nocons` option to indicate no random intercept. This is a rather restrictive model as it assumes that each trial has the same baseline hazard function. In practice, we may include the trial id variable in the linear predictor, to allow proportional trials effects, or indeed stratify by trial membership to allow separate trials effects (Crowther et al. (2012, 2014)).

## 6 Conclusion

In this paper I have described the `stmixed` command for multilevel mixed effects parametric analysis. I hope the wide range of survival models available will be found useful in applied research. Future developments will involve extension to higher levels of clustering.

## 7 References

- Charvat, H., L. Remontet, N. Bossard, L. Roche, O. Dejardin, B. Rachet, G. Launoy, and A. Belot. 2016. A multilevel excess hazard model to estimate net survival on hierarchical data allowing for non-linear and non-proportional effects of covariates. *Statistics in medicine* .
- Crowther, M. J. 2013. Multilevel mixed-effects parametric survival analysis. United kingdom stata users' group meetings 2013, Stata Users Group. URL <http://EconPapers.repec.org/RePEc:boc:usug13:05>.
- Crowther, M. J., and P. C. Lambert. 2012. Simulating complex survival data. *Stata J* 12(4): 674–687.
- . 2013. Simulating biologically plausible complex survival data. *Stat Med* 32(23): 4118–4134. URL <http://dx.doi.org/10.1002/sim.5823>.
- Crowther, M. J., M. P. Look, and R. D. Riley. 2014. Multilevel mixed effects parametric survival models using adaptive Gauss-Hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Stat Med* 33(22): 3844–3858. URL <http://dx.doi.org/10.1002/sim.6191>.
- Crowther, M. J., R. D. Riley, J. A. Staessen, J. Wang, F. Gueyffier, and P. C. Lambert. 2012. Individual patient data meta-analysis of survival data using Poisson regression models. *BMC Med Res Methodol* 12(34). URL <http://dx.doi.org/10.1186/1471-2288-12-34>.
- Dickman, P. W., and E. Coviello. 2015. Estimating and modeling relative survival. *Stata Journal* 15(1): 186–215(30). URL <http://www.stata-journal.com/article.html?article=st0376>.
- Dickman, P. W., A. Sloggett, M. Hills, and T. Hakulinen. 2004. Regression models for relative survival. *Stat Med* 23(1): 51–64. URL <http://dx.doi.org/10.1002/sim.1597>.

- Gutierrez, R. G. 2002. Parametric frailty and shared frailty survival models. *The Stata Journal* 2(1): 22–44.
- King, N. B., S. Harper, and M. E. Young. 2012. Use of relative and absolute effect measures in reporting health inequalities: structured review. *BMJ* 345: e5774.
- Komárek, A., E. Lesaffre, and C. Legrand. 2007. Baseline and treatment effect heterogeneity for survival times between centers using a random effects accelerated failure time model with flexible error distribution. *Stat Med* 26(30): 5457–5472. URL <http://dx.doi.org/10.1002/sim.3083>.
- Lambert, P. C., L. Holmberg, F. Sandin, F. Bray, K. M. Linklater, A. Purushotham, D. Robinson, and H. Møller. 2011. Quantifying differences in breast cancer survival between England and Norway. *Cancer Epidemiol* 35(6): 526–533. URL <http://dx.doi.org/10.1016/j.canep.2011.04.003>.
- Lambert, P. C., and P. Royston. 2009. Further development of flexible parametric models for survival analysis. *Stata J* 9: 265–290.
- Legrand, C., V. Ducrocq, P. Janssen, R. Sylvester, and L. Duchateau. 2005. A Bayesian approach to jointly estimate centre and treatment by centre heterogeneity in a proportional hazards model. *Stat Med* 24(24): 3789–3804.
- Liu, L., and X. Huang. 2008. The use of Gaussian quadrature for estimation in frailty proportional hazards models. *Stat Med* 27(14): 2665–2683. URL <http://dx.doi.org/10.1002/sim.3077>.
- McGilchrist, C. A., and C. W. Aisbett. 1991. Regression with frailty in survival analysis. *Biometrics* 47(2): 461–466.
- Miladinovic, B., A. Kumar, R. Mhaskar, S. Kim, R. Schonwetter, and B. Djulbegovic. 2012. A flexible alternative to the Cox proportional hazards model for assessing the prognostic accuracy of hospice patient survival. *PLoS One* 7(10): e47804. URL <http://dx.doi.org/10.1371/journal.pone.0047804>.
- Mok, T. S., Y.-L. Wu, S. Thongprasert, C.-H. Yang, D.-T. Chu, N. Saijo, P. Sunpaweravong, B. Han, B. Margono, Y. Ichinose, Y. Nishiwaki, Y. Ohe, J.-J. Yang, B. Chewaskulyong, H. Jiang, E. L. Duffield, C. L. Watkins, A. A. Armour, and M. Fukuoka. 2009. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 361(10): 947–957. URL <http://dx.doi.org/10.1056/NEJMoa0810699>.
- Pinheiro, J. C., and D. M. Bates. 1995. Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *J Comput Graph Statist* 4(1): pp. 12–35.
- Rabe-Hesketh, S., A. Skrondal, and A. Pickles. 2002. Reliable estimation of generalized linear mixed models using adaptive quadrature. *Stata J* 2: 1–21.

- Rooney, J., S. Byrne, M. Heverin, B. Corr, M. Elamin, A. Staines, B. Goldacre, and O. Hardiman. 2013. Survival analysis of irish amyotrophic lateral sclerosis patients diagnosed from 1995-2010. *PLoS One* 8(9): e74733. URL <http://dx.doi.org/10.1371/journal.pone.0074733>.
- Royston, P., and P. C. Lambert. 2011. *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. Stata Press.
- Royston, P., and M. K. B. Parmar. 2002. Flexible Parametric Proportional Hazards and Proportional Odds Models for Censored Survival Data, with Application to Prognostic Modelling and Estimation of Treatment Effects. *Stat Med* 21(15): 2175–2197.
- Simmonds, M. C., J. P. T. Higgins, L. A. Stewart, J. F. Tierney, M. J. Clarke, and S. G. Thompson. 2005. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clin Trials* 2(3): 209–217.
- Tudur-Smith, C., P. R. Williamson, and A. G. Marson. 2005. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Stat Med* 24(9): 1307–1319.
- Yamaguchi, T., Y. Ohashi, and Y. Matsuyama. 2002. Proportional hazards models with random effects to examine centre effects in multicentre cancer clinical trials. *Stat Methods Med Res* 11(3): 221–236.

#### **About the authors**

Michael Crowther is a lecturer in biostatistics at the University of Leicester. His main research interests centre on the development of methods for the analysis of complex survival data.