The Stata Journal (2018)
18, Number 2, pp. 477–484

Frailty models and frailty-mixture models for recurrent event times: Update

Ying Xu
Center for Quantitative Medicine
Duke–NUS Graduate Medical School
and Procter & Gamble Co.
Singapore, Singapore
tinayxu@gmail.com

Yin Bun Cheung
Center for Quantitative Medicine
Duke–NUS Graduate Medical School
Singapore, Singapore
and Department for International Health
University of Tampere
Tampere, Finland
yinbun.cheung@duke-nus.edu.sg

Abstract. Xu and Cheung (2015, *Stata Journal* 15: 135–154) introduced the `strmcure` command, which fits frailty models and frailty-mixture models in the analysis of recurrent event times. In this article, we provide an update to `strmcure`. The update implements a two-step estimation procedure for a frailty-mixture model that allows the estimation of the effect of an intervention on the probability of cure and on the total effect on event rate in the noncured. To illustrate, we will use the same example dataset on respiratory exacerbations from the original article.

Keywords: st0374_1, strmcure, frailty-mixture model, primary effect, total effect, two-step estimation procedure

1 Introduction

Xu et al. (2012) and Xu and Cheung (2015) described a class of frailty models and frailty-mixture models for regression analysis of recurrent event times. Xu and Cheung (2015) also described the `strmcure` command, which implements these models. Briefly, this command fits frailty models for recurrent event times with or without stratification for event order and with or without a cured fraction. Introducing frailty into recurrent event-times modeling enables differentiation between cured (nonsusceptible) subjects and low-risk subjects. Both frailty and event dependence generate within-subject correlation in event times. Therefore, frailty cannot be accurately estimated without controlling for event dependence. As such, when frailty-mixture models are fit to the data, analysts need to control for the effect of event dependence on the event hazard, using the event-order stratification so that the effect of event dependence is incorporated into the unspecified event-specific baseline hazard. Otherwise, event dependence confounds heterogeneity, which in turn will distort the estimation of the cured fraction. The `strmcure` command has been coded following this rationale to require users to specify the `strata()` option for a frailty-mixture model (Xu and Cheung 2015).

"Event dependence" is when one's past event history may affect one's present and future event rate. As such, an intervention may have a primary effect on a person's event

rate that is not mediated by his or her event history and may have a secondary effect that is mediated by the event history. "Primary" and "secondary" are used here to refer to first and second in order, respectively, without referring to degree of importance. In addition, the two effects are not necessarily in the same direction; that is, one may be beneficial while the other is harmful. (We do not call the effects "direct" and "indirect" to avoid confusion in some contexts where these terms have other meanings, for example, regarding herd immunity.)

In the biostatistics literature, a recurrent event model with frailty and stratification for event order is known as the conditional frailty model (Box-Steffensmeier and De Boef 2006). Because in this model the comparison is between event times with the same event history, the effect of event dependence is isolated out and the coefficients are estimates of the primary effect. The primary and secondary effects reflect two pathways that an intervention impacts on the event rate. The primary effect combined with the secondary effect is the total effect (Cheung et al. 2010; Xu, Lam, and Cheung 2014). The primary effect is one mechanism through which an intervention may affect the outcome. Its evaluation is important for biological or product development, and the `strmcure` command with the `strata()` option fits this purpose.

On the other hand, the total effect examines what happens in a pragmatic study (Hand 1994). In the pragmatic paradigm, an intervention effect is the sum of all the effects the intervention may have, whether the effects were intended or not. For example, people acquire natural immunity from malaria disease episodes. A vaccine that prevents malaria disease episodes may also prevent the acquisition of natural immunity, which in turn may increase future malaria disease risk, that is, a secondary effect that is harmful. The beneficial primary effect may be canceled out by the harmful secondary effect and result in null total effect. In the early phase of developing a malaria vaccine, the investigators may mainly want to find out whether the body responds to the vaccine candidate. In this case, the secondary effect arising from disease episode history is distractive. However, to policy makers, a null total effect means the intervention is not worth deploying. Therefore, the total effect is a useful parameter from a public health viewpoint and can be obtained from the two-step estimation procedure for the frailty-mixture model proposed by Xu, Lam, and Cheung (2014). In essence, in the first step, the effect of intervention on the probability of cure, as well as the conditional probabilities of being noncured for all subjects given their observed data, is estimated. In the second step, these estimated conditional probabilities are plugged into a model that estimates the total effect in the noncured.

In this article, we introduce an update to the `strmcure` command: the new `twostep` option, which implements the aforementioned two-step estimation procedure. To illustrate the new option, we will use the same example dataset from the rhDNase trial of respiratory exacerbations as in the main article (Xu and Cheung 2015).

2 The frailty-mixture model with the two-step estimation procedure

Details about frailty-mixture models for estimating a cured fraction and the total effect on event rate in the noncured can be found in Xu, Lam, and Cheung (2014). Briefly, consider the setting of a randomized controlled trial with n subjects. For subject i ($i = 1, \ldots, n$), let x_i denote the binary intervention variable (1 for intervention and 0 for control), let t_{ij} denote the jth event time ($j = 1, \ldots, n_i$), let n_i denote the number of observed event times, and let δ_{ij} denote the event indicator, which takes a value of 0 if the jth event time is right-censored and takes a value of 1 otherwise.

We further define a binary latent variable k_i such that $k_i = 0$ if subject i is cured, and $k_i = 1$ if subject i is noncured and will eventually experience the event. To accommodate the effect of intervention on the cured fraction, the latent variable k_i can be modeled by

$$\pi_i = \Pr(k_i = 1|\mathbf{Z}_i) = g(\mathbf{Z}_i\boldsymbol{\theta}^T) \tag{1}$$

where $g(\cdot)$ is a link function, $\mathbf{Z}_i = (1, x_i)$, and $\boldsymbol{\theta} = (\theta_0, \theta_1)$ is the regression coefficient vector. The `strmcure` command supports three link functions: the logit, probit, and complementary log-log links.

For subject i, conditional on k_i, the intensity at time t is defined as

$$\lambda_i\{t|Y_i(t), x_i, k_i\} = k_i Y_i(t)\lambda_0(t)\ \exp(\beta_T x_i) \tag{2}$$

where the at-risk indicator $Y_i(t) = 1$ if subject i is at risk for the event at time t since study entry and equals 0 otherwise, $\lambda_0(t)$ is the unspecified baseline hazard function, and the log hazard-ratio parameter β_T represents the total effect of intervention on the event rate in the noncured subjects (Cheung et al. 2010; Xu, Lam, and Cheung 2014).

Estimation of regression parameters in (1) and (2) is not as straightforward as that for a typical frailty-mixture model that jointly models frailty, event dependence, and the cured fraction (Xu et al. 2012; Xu and Cheung 2015). Here, to estimate the total effect in the noncured fraction, event-order stratification is not used. Otherwise, the intensity equation would estimate the primary effect instead. Without stratification, frailty cannot be accurately estimated. Without the frailty term, accurate estimation of the cured fraction is difficult because of poor differentiation between cured and low-risk subjects (Xu, Lam, and Cheung 2014). To enable the estimation of both the total effect (among the noncured) and the cured fraction, Xu, Lam, and Cheung (2014) proposed a two-step estimation procedure that splits the parameter estimation for (1) and (2) into the following two steps.

2.1 Step 1: Estimation of the cured fraction

In step 1, we build a frailty-mixture model with a cured fraction and event-order stratification, using the same time scale as (2). The model for the cured fraction is the same as (1). For a noncured subject i, the hazard for the jth event time at time t is formulated following Xu, Lam, and Cheung (2014),

$$h_{ij}\{t|Y_{ij}^I(t),\omega_i,x_i,k_i=1\} = Y_{ij}^I(t)\omega_i h_{0j}(t)\ \exp(\beta_P x_i) \tag{3}$$

where $Y_{ij}^I(t)$ is the at-risk process for the step-1 model [$Y_{ij}^I(t)=1$ if subject i has experienced $(j-1)$ events prior to time t and is still at risk for the event at time t, and $Y_{ij}^I(t)=0$ otherwise]; ω_i is the subject-specific frailty term that is gamma distributed with a mean of 1 and a variance of ψ; and $h_{0j}(t)$ is the unspecified baseline hazard function specific to the jth event. Note that (3) is formulated to control for the effects of heterogeneity and event dependence via the frailty term and event-order stratification. This entails an interpretation of "primary effect" to the regression parameter β_P in (3).

The iterative expectation-maximization algorithm is used to estimate the regression parameters $\boldsymbol{\theta}=(\theta_0,\theta_1)$ and β_P in (1) and (3), as well as the frailty variance parameter ψ. Furthermore, the conditional probability of being noncured for each subject, given the observed data, can be estimated based on the formula

$$\Pr(k_i=1) = I\left(\sum_{j=1}^{n_i}\delta_{ij}>0\right) + I\left(\sum_{j=1}^{n_i}\delta_{ij}=0\right)\frac{\pi_i\{1+\psi H_{i1}(t_{i1})\}^{-1/\psi}}{1-\pi_i+\pi_i\{1+\psi H_{i1}(t_{i1})\}^{-1/\psi}} \tag{4}$$

where $H_{i1}(t_{i1}) = \int_0^{t_{i1}} Y_{i1}^I(s)h_{01}(s)\ \exp(\beta_P x_i)ds$.

2.2 Step 2: Estimation of the total effect

The regression parameter β_T is estimated by solving the following score function derived for (2):

$$U(\beta_T) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\delta_{ij}\left\{x_i - \frac{\sum_{l=1}^{n} Y_l(t_{ij})k_l\ \exp(\beta_T x_i)x_l}{\sum_{l=1}^{n} Y_l(t_{ij})k_l\ \exp(\beta_T x_i)}\right\} = 0$$

where the value of k_l is substituted by its conditional probability (or conditional expectation) given in (4).

The standard errors of the parameter estimates are estimated using the bootstrap method, and the 95% confidence intervals are obtained using the asymptotic normal distribution theory. For more details on the estimation procedure, see Xu, Lam, and Cheung (2014). Note that the two-step estimation uses bootstrapping instead of Louis's (1982) formula because there is uncertainty at both steps. The model not using the two-step procedure can possibly use either bootstrapping or Louis's formula, but the latter is faster.

3 Update to the strmcure command

3.1 Syntax

`strmcure` *varlist* [*if*] [*in*], `shared(`*varname*`)` [*original_options* `twostep`]

Users are referred to Xu and Cheung (2015) for the list of original options for the `strmcure` command. Users must now specify the `strata()` option together with the `zlist()` option when they fit a frailty-mixture model to the data. In particular, the `strata()` and `zlist()` options have been updated as follows:

`strata(`*varname*`)` stratifies the recurrent event times according to the specified event-order variable *varname*. Observations with the same value belong to the same stratum, and the baseline hazard function is unique to each stratum. Not specifying `strata()` means fitting a model without stratification when the model does not involve a cured fraction. `strata()` must be specified for the frailty-mixture model.

`zlist(`*varlist*`)` specifies the list of variable names to be included in modeling the probability of being noncured in the frailty-mixture model. When you specify `zlist()`, you must also specify `strata()`; otherwise, the error message "`strata() option must be specified for estimation of frailty-mixture models`" will be displayed.

The new twostep option

`twostep` invokes the two-step estimation procedure for the estimation of total effect (that is, primary effect plus secondary effect) in the noncured population for a frailty-mixture model. Specification of the `twostep` option will void options including `log`, `dots`, `saving`, and all those related to variance based on Louis's formula.

4 Application

We use the dataset from the rhDNase trial of respiratory exacerbations described in Cook and Lawless (2007) to illustrate the use of `strmcure` with the new `twostep` option. Xu, Lam, and Cheung (2014) and Xu and Cheung (2015) used the same dataset in their application sections. Users are referred to the previous articles for details about the trial; we skip that here to avoid redundancy.

There are some suspected errors in the data (Xu and Cheung 2015), so we begin by correcting these data points.

```
. use data_rhdnase
. recode etype 1=2 2=1 if id==951319 | id==985308 | id==985316 | id==986310
(etype: 14 changes made)
```

We declare the data to be survival-time data, adopting the counting-process time scale.

```
. stset time2 if etype==1, fail(status) id(id) time0(time1) exit(time .)
```

In the `stset` command, we conditioned out the data records with `etype` = 2 because they represent the acute treatment duration initiated by exacerbation, during which the patient was, in principle, not at risk of another exacerbation. We drop records with `etype` = 2 from now on. With the dataset at hand, we fit a frailty-mixture model with stratification (with the fourth and higher event strata being collapsed into one stratum because of data sparsity) and a cured fraction. The logit link function and zero-tail constraint for tail completion are used. The `twostep` option is specified for estimation of the total effect of the intervention in the noncured, which is our primary interest. One hundred bootstrapping replicates were used for variance estimation. We execute the following commands:

```
. stgen event_order=nfailures()
(361 missing values generated)
. replace event_order=event_order+1
(965 real changes made)
. bootstrap, seed(1234) cluster(id) idcluster(newid) reps(100): strmcure trt,
> shared(newid) strata(event_order) lastpool(4) zlist(trt) link(logistic)
> tailcm(zerotail) iter(500) tolerance(0.0001) twostep
```

(*output omitted*)

_t	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
Theta_step1						
trt	-.4405835	.179392	-2.46	0.014	-.7921853	-.0889817
_cons	-.1565902	.1388646	-1.13	0.259	-.4287597	.1155794
Beta_step1						
trt	-.0086713	.1959342	-0.04	0.965	-.3926954	.3753527
frailty_step1						
_cons	1.208458	.5931335	2.04	0.042	.045938	2.370979
Beta_step2						
trt	.0367756	.0986752	0.37	0.709	-.1566242	.2301753

The above table reproduces part of the results from table V in Xu, Lam, and Cheung (2014). The first three rows of the table show the estimation results from step 1, and the last row shows the estimation result for total effect from step 2. To summarize, the estimate for the log odds-ratio is about -0.4406, which is statistically significant ($P = 0.014$ and the estimated 95% confidence interval excludes the null value of 0), suggesting that subjects in the intervention (rhDNase) group are more likely to be noncured than those in the control group. On the other hand, the estimated primary effect from step 1 and total effect from step 2 are, respectively, -0.0087 and 0.0368, both practically 0 and statistically insignificant. The intervention appears to have been impactful mainly by curing a fraction of the patients instead of reducing the event rate in the noncured.

5 Conclusion

We presented an update to the recently published `strmcure` command that encompasses a new option, `twostep`. This update implements the previously introduced methodology in Xu, Lam, and Cheung (2014) and Xu and Cheung (2015). This method can provide insight into the details of how an intervention impacts a population. The method was evaluated for robustness by simulation. Despite model complexity, the method was shown to be robust to misspecifications of frailty distribution and the event dependence function (Xu, Lam, and Cheung 2014). The impact of misspecifications on other aspects, for example, the choice of link function in the cured fraction equation, has not yet been assessed.

The update should also be useful to fit simpler models, such as the Andersen–Gill model, which estimates the total effect in the whole population without intending to obtain details about cured or noncured (Cheung et al. 2010). Furthermore, improvement in data collection may help. For example, clinical evaluation or biomarkers of cured status or nonsusceptibility can be collected, if feasible. They may have the potential to remove the need for complex modeling involving unobserved characteristics.

6 Acknowledgments

This work was supported by the National Research Foundation, Singapore, under its Clinician Scientist Award (Award No. NMRC/CSA/039/2011) administered by the Singapore Ministry of Health's National Medical Research Council.

7 References

Box-Steffensmeier, J. M., and S. De Boef. 2006. Repeated events survival models: The conditional frailty model. *Statistics in Medicine* 25: 3518–3533.

Cheung, Y. B., Y. Xu, S. H. Tan, F. Cutts, and P. Milligan. 2010. Estimation of intervention effects using first or multiple episodes in clinical trials: The Andersen-Gill model re-examined. *Statistics in Medicine* 29: 328–336.

Cook, R. J., and J. F. Lawless. 2007. *The Statistical Analysis of Recurrent Events*. New York: Springer.

Hand, D. J. 1994. Deconstructing statistical questions. *Journal of the Royal Statistical Society, Series A* 157: 317–356.

Louis, T. A. 1982. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B* 44: 226–233.

Xu, Y., and Y. B. Cheung. 2015. Frailty models and frailty-mixture models for recurrent event times. *Stata Journal* 15: 135–154.

Xu, Y., Y. B. Cheung, K. F. Lam, and P. Milligan. 2012. Estimation of summary protective efficacy using a frailty mixture model for recurrent event time data. *Statistics in Medicine* 31: 4023–4039.

Xu, Y., K. F. Lam, and Y. B. Cheung. 2014. Estimation of intervention effects using recurrent event time data in the presence of event dependence and a cured fraction. *Statistics in Medicine* 33: 2263–2274.

About the authors

Ying Xu is a statistician/data scientist working at Procter & Gamble. Her current research interests include statistical methodology related to high-dimensional omics data as well as design and analysis of clinical trials.

Yin Bun Cheung is a medical statistician and pediatric epidemiologist. He is a professor at the Center for Quantitative Medicine at Duke–NUS Graduate Medical School in Singapore and is an adjunct professor in the Department of International Health at the University of Tampere in Finland. His current research areas include statistical methods for analysis of censored data and excess zeros and a nonsusceptible fraction, as well as the impact and interplay of infection and undernutrition on maternal and child health in developing countries.