Introduction to joint modelling of longitudinal and survival data

Recent advances in joint models for cancer and the new statistical challenge of immunotherapy clinical studies

Bordeaux, 24-25th January 2019

Michael Sweeting

Department of Health Sciences,
University of Leicester, UK,
michael.sweeting@le.ac.uk

Huge thanks to Michael Crowther

Outline

Outline

Background

Biomarkers are often collected repeatedly over time, in parallel to the time to an event of interest. Some examples from the clinical literature include:

- CD4 cell counts in patients with HIV, and the time to progression of AIDS
- Prostate specific antigen and risk of prostate cancer recurrence
- Serum bilirubin and primary biliary cirrhosis of the liver
- Abdominal aortic aneurysm diameter and time to aneurysm rupture

Research questions

- How does the trajectory of the biomarker over time impact the risk of the clinical event?
- If patients with higher biomarker levels are more likely to die, will this affect our estimates of the trajectory of the biomarker?
- Can we predict who will have the clinical event in the future from repeated measurements of the biomarker?

Background

Such biomarkers have inherent features which must be taken into account in any analysis

- These biomarkers are often measured with error
- Measurements taken on the same individual are generally correlated
- Measured intermittently throughout follow-up
- The value of the biomarker may be related to prognosis

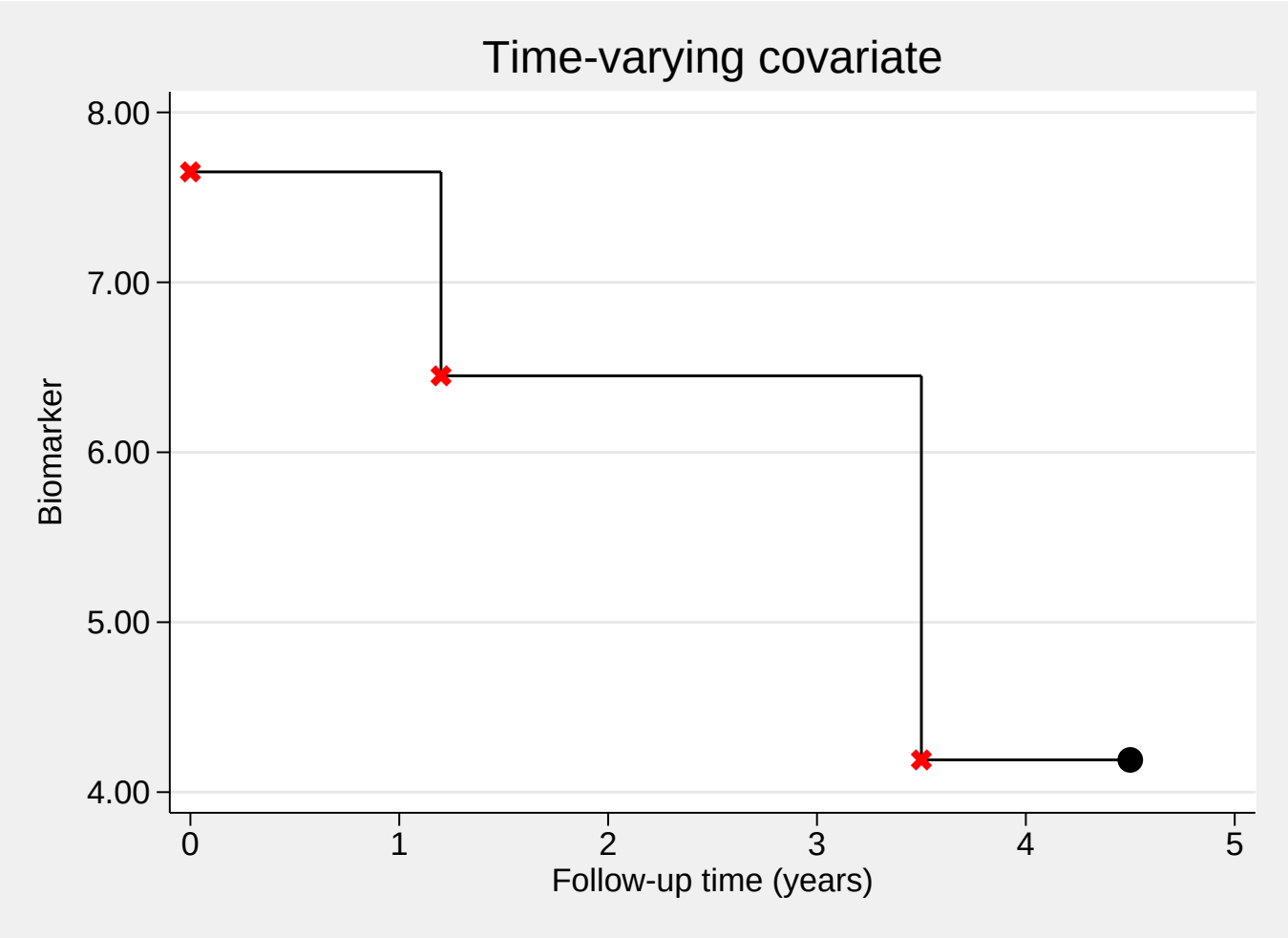
Survival analysis with a time-varying biomarker

- We could consider fitting a survival model with a time-varying covariate (TVC)

$$h_i(t) = h_0(t) \exp\left[\boldsymbol{\phi}^T \boldsymbol{v_i} + \alpha y_i(t)\right]$$

  where $y_i(t)$ is the *observed* biomarker value for the i^{th} patient at time t, $\boldsymbol{v_i}$ are baseline covariates, $h_0(t)$ is a baseline hazard function
- But, we assume the value of the biomarker doesn't change until a new measurement is taken.
- We are ignoring measurement error in the biomarker

Survival analysis with a time-varying biomarker



Two-stage models

- In a survival analysis with a time-varying covariate, we are assuming that the covariate is observed error-free, and only changes value at observation points.

Two-stage models

- In a survival analysis with a time-varying covariate, we are assuming that the covariate is observed error-free, and only changes value at observation points.
- If we model the biomarker using a linear mixed effects model, we are creating a model for the outcome at any time-point t, and furthermore, we were attempting to remove the measurement error.

Two-stage models

- In a survival analysis with a time-varying covariate, we are assuming that the covariate is observed error-free, and only changes value at observation points.
- If we model the biomarker using a linear mixed effects model, we are creating a model for the outcome at any time-point t, and furthermore, we were attempting to remove the measurement error.
- Instead of using the observed biomarker values, we can fit a linear mixed effects model, and obtain subject-specific predictions of the true, unobserved biomarker values, at the observation times and use these instead.

Two-stage models

Mathematically,

$$y_i(t) = m_i(t) + e_i(t), \qquad e_i(t) \sim \mathrm{N}(0, \sigma^2)$$

where

$$m_i(t) = \boldsymbol{X_i}^T(t)\boldsymbol{\beta} + \boldsymbol{Z_i}^T(t)\boldsymbol{b_i}$$

Two-stage models

Mathematically,

$$y_i(t) = m_i(t) + e_i(t), \qquad e_i(t) \sim \mathrm{N}(0, \sigma^2)$$

where

$$m_i(t) = \boldsymbol{X_i}^T(t)\boldsymbol{\beta} + \boldsymbol{Z_i}^T(t)\boldsymbol{b_i}$$

We then obtain our subject-specific predictions, $\hat{m}_i(t)$, and use these as our time-varying covariate

$$h_i(t) = h_0(t) \exp\left[\boldsymbol{\phi}^T \boldsymbol{v_i} + \alpha \hat{m}_i(t)\right]$$

However, there are still issues with the two-stage approach

However, there are still issues with the two-stage approach

- The uncertainty in our estimates from the first stage are not carried through to the second stage (Sweeting and Thompson, 2011). This means our estimates of association are too precise.
- In terms of how the survival model is estimated, we're still assuming the values do not change between observations

However, there are still issues with the two-stage approach

- The uncertainty in our estimates from the first stage are not carried through to the second stage (Sweeting and Thompson, 2011). This means our estimates of association are too precise.
- In terms of how the survival model is estimated, we're still assuming the values do not change between observations

However,

- It has been shown to greatly reduce bias compared to the TVC approach
- It allows us to fit complex models very quickly

Outline

Joint modelling of longitudinal and survival data

- Arose primarily in the field of AIDS, relating CD4 trajectories to progression to AIDS in HIV positive patients (Faucett and Thomas, 1996)
- Further developed in cancer, particularly modelling PSA levels and their association with prostate cancer recurrence (Proust-Lima and Taylor, 2009)
- Think of it as two component models:
  - Longitudinal part - linear mixed effects model (`mixed`)
  - Survival part - proportional hazards model (`streg`)
  - The component parts then share some parameter - dependence through shared random effects (Wulfsohn and Tsiatis, 1997; Henderson et al., 2000; Rizopoulos, 2012)

Joint modelling of longitudinal and survival data

Longitudinal submodel

Assume we observe continuous longitudinal marker:

$$y_i(t) = m_i(t) + e_i(t), \qquad e_i(t) \sim \mathrm{N}(0, \sigma^2)$$

where

$$m_i(t) = \boldsymbol{X_i}^T(t)\boldsymbol{\beta} + \boldsymbol{Z_i}^T(t)\boldsymbol{b_i}, \qquad \boldsymbol{b_i} \sim \mathrm{N}(0, \Sigma)$$

We call $m_i(t)$ the trajectory function, i.e. the true unobserved value of the biomarker for the i^{th} patient at time t.

Joint modelling of longitudinal and survival data

Longitudinal submodel

Assume we observe continuous longitudinal marker:

$$y_i(t) = m_i(t) + e_i(t), \qquad e_i(t) \sim \mathrm{N}(0, \sigma^2)$$

where

$$m_i(t) = \boldsymbol{X_i}^T(t)\boldsymbol{\beta} + \boldsymbol{Z_i}^T(t)\boldsymbol{b_i}, \qquad \boldsymbol{b_i} \sim \mathrm{N}(0, \Sigma)$$

We call $m_i(t)$ the trajectory function, i.e. the true unobserved value of the biomarker for the i^{th} patient at time t.

The basic framework

Survival submodel

Define $M_i(t) = \{m_i(s), 0 \leq s \leq t\}$, to be the true unobserved longitudinal profile up to time t. We assume a proportional hazards survival submodel

$$h(t|M_i(t), \boldsymbol{v_i}) = h_0(t) \exp\left[\boldsymbol{\phi}^T \boldsymbol{v_i} + \alpha m_i(t)\right]$$

where $h_0(t)$ is the baseline hazard function, and $\boldsymbol{v_i}$ a set of baseline time-independent covariates with associated vector of log hazard ratios, $\boldsymbol{\phi}$.

The basic framework

Survival submodel

Define $M_i(t) = \{m_i(s), 0 \leq s \leq t\}$, to be the true unobserved longitudinal profile up to time t. We assume a proportional hazards survival submodel

$$h(t|M_i(t), \boldsymbol{v_i}) = h_0(t) \exp\left[\boldsymbol{\phi}^T \boldsymbol{v_i} + \alpha m_i(t)\right]$$

where $h_0(t)$ is the baseline hazard function, and $\boldsymbol{v_i}$ a set of baseline time-independent covariates with associated vector of log hazard ratios, $\boldsymbol{\phi}$.

Linking the component models

Our key question here is how are changes in the biomarker trajectory associated with survival?

$$h(t|M_i(t), \boldsymbol{v_i}) = h_0(t) \exp\left[\phi^T \boldsymbol{v_i} + \alpha m_i(t)\right]$$

- $\alpha m_i(t)$ is termed the current value parameterisation

Linking the component models

Our key question here is how are changes in the biomarker trajectory associated with survival?

$$h(t|M_i(t), \boldsymbol{v_i}) = h_0(t) \exp\left[\boldsymbol{\phi}^T \boldsymbol{v_i} + \alpha m_i'(t)\right]$$

- $\alpha m_i(t)$ is termed the current value parameterisation
- $\alpha m_i'(t) = \alpha \frac{\mathrm{d}m_i(t)}{\mathrm{d}t}$ relates the hazard to the rate of change of the biomarker

Linking the component models

Our key question here is how are changes in the biomarker trajectory associated with survival?

$$h(t|M_i(t), \boldsymbol{v_i}) = h_0(t) \exp\left[\boldsymbol{\phi}^T \boldsymbol{v_i} + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\right]$$

- $\alpha m_i(t)$ is termed the current value parameterisation
- $\alpha m_i'(t) = \alpha \frac{\mathrm{d}m_i(t)}{\mathrm{d}t}$ relates the hazard to the rate of change of the biomarker
- $\alpha_1 m_i(t) + \alpha_2 m_i'(t)$ - both current value and rate of change

Linking the component models

Our key question here is how are changes in the biomarker trajectory associated with survival?

$$h(t|M_i(t), \boldsymbol{v_i}) = h_0(t) \exp\left[\boldsymbol{\phi}^T \boldsymbol{v_i} + \alpha(\beta_0 + b_{0i})\right]$$

- $\alpha m_i(t)$ is termed the current value parameterisation
- $\alpha m_i'(t) = \alpha \frac{\mathrm{d}m_i(t)}{\mathrm{d}t}$ relates the hazard to the rate of change of the biomarker
- $\alpha_1 m_i(t) + \alpha_2 m_i'(t)$ - both current value and rate of change
- $\alpha(\beta_0 + b_{0i})$ - the subject-specific intercept

Linking the component models

Our key question here is how are changes in the biomarker trajectory associated with survival?

$$h(t|M_i(t), \boldsymbol{v_i}) = h_0(t) \exp\left[\boldsymbol{\phi}^T \boldsymbol{v_i} + \boldsymbol{\alpha}^\top \boldsymbol{W}_i(t|\boldsymbol{b}_i; \boldsymbol{\beta})\right]$$

- $\alpha m_i(t)$ is termed the current value parameterisation
- $\alpha m_i'(t) = \alpha \frac{\mathrm{d}m_i(t)}{\mathrm{d}t}$ relates the hazard to the rate of change of the biomarker
- $\alpha_1 m_i(t) + \alpha_2 m_i'(t)$ - both current value and rate of change
- $\alpha(\beta_0 + b_{0i})$ - the subject-specific intercept
- $\boldsymbol{\alpha}^\top \boldsymbol{W}_i(t|\boldsymbol{b}_i; \boldsymbol{\beta})$ in general any (multivariate) function of the random coefficients

Outline

Joint likelihood (for those interested...)

Our full joint likelihood relies on conditional independence:

$$\prod_{i=1}^{N}\left[\int_{-\infty}^{\infty}\left(\prod_{j=1}^{n_i} p(y_i(t_{ij})|b_i,\theta)\right) p(b_i|\theta)p(T_i,d_i|b_i,\theta)\ \mathrm{d}b_i\right]$$

Joint likelihood (for those interested...)

Our full joint likelihood relies on conditional independence:

$$\prod_{i=1}^{N}\left[\int_{-\infty}^{\infty}\left(\prod_{j=1}^{n_i} p(y_i(t_{ij})|b_i,\theta)\right) p(b_i|\theta)p(T_i,d_i|b_i,\theta)\ \mathrm{d}b_i\right]$$

where we have our continuous longitudinal outcome,

$$p(y_i(t_{ij})|b_i,\theta) = (2\pi\sigma_e^2)^{-1/2}\exp\left\{-\frac{[y_i(t_{ij})-m_i(t_{ij})]^2}{2\sigma_e^2}\right\}$$

Joint likelihood (for those interested...)

Our full joint likelihood relies on conditional independence:

$$\prod_{i=1}^{N}\left[\int_{-\infty}^{\infty}\left(\prod_{j=1}^{n_i} p(y_i(t_{ij})|b_i,\theta)\right) p(b_i|\theta)p(T_i,d_i|b_i,\theta)\ \mathrm{d}b_i\right]$$

our multivariate normally distributed random effects,

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

Joint likelihood (for those interested...)

Our full joint likelihood relies on conditional independence:

$$\prod_{i=1}^{N}\left[\int_{-\infty}^{\infty}\left(\prod_{j=1}^{n_i} p(y_i(t_{ij})|b_i,\theta)\right) p(b_i|\theta)p(T_i, d_i|b_i,\theta)\ \mathrm{d}b_i\right]$$

and our survival outcome,

$$p(T_i, d_i|b_i,\theta) = [h_0(T_i)\exp(\alpha m_i(t) + \phi v_i)]^{d_i}$$

$$\times \exp\left\{-\int_0^{T_i} h_0(u)\exp(\alpha m_i(u) + \phi v_i)du\right\}$$

Joint likelihood (for those interested...)

Our full joint likelihood relies on conditional independence:

$$\prod_{i=1}^{N}\left[\int_{-\infty}^{\infty}\left(\prod_{j=1}^{n_i} p(y_i(t_{ij})|b_i,\theta)\right) p(b_i|\theta)p(T_i,d_i|b_i,\theta)\ \mathrm{d}b_i\right]$$

and our survival outcome,

$$p(T_i,d_i|b_i,\theta) = [h_0(T_i)\exp(\alpha m_i(t)+\phi v_i)]^{d_i}$$

$$\times \exp\left\{-\int_0^{T_i} h_0(u)\exp(\alpha m_i(u)+\phi v_i)du\right\}$$

Gauss-Hermite quadrature needed to approximate analytically intractable integrals (Pinheiro and Bates, 1995)

Outline

Example: Primary biliary cirrhosis

- 312 patients with primary biliary cirrhosis
- Cirrhosis is a slowly progressing disease in which healthy liver tissue is replaced with scar tissue, eventually preventing the liver from functioning properly
- 1945 repeated measures of serum bilirubin, a measure of liver function
- Treated with D-penicillamine or a placebo
- Outcome of all-cause death, where 140 (44.8%) patients died

Research question: How does serum bilirubin change over time, and are those changes associated with survival?

Data structure (Stata)

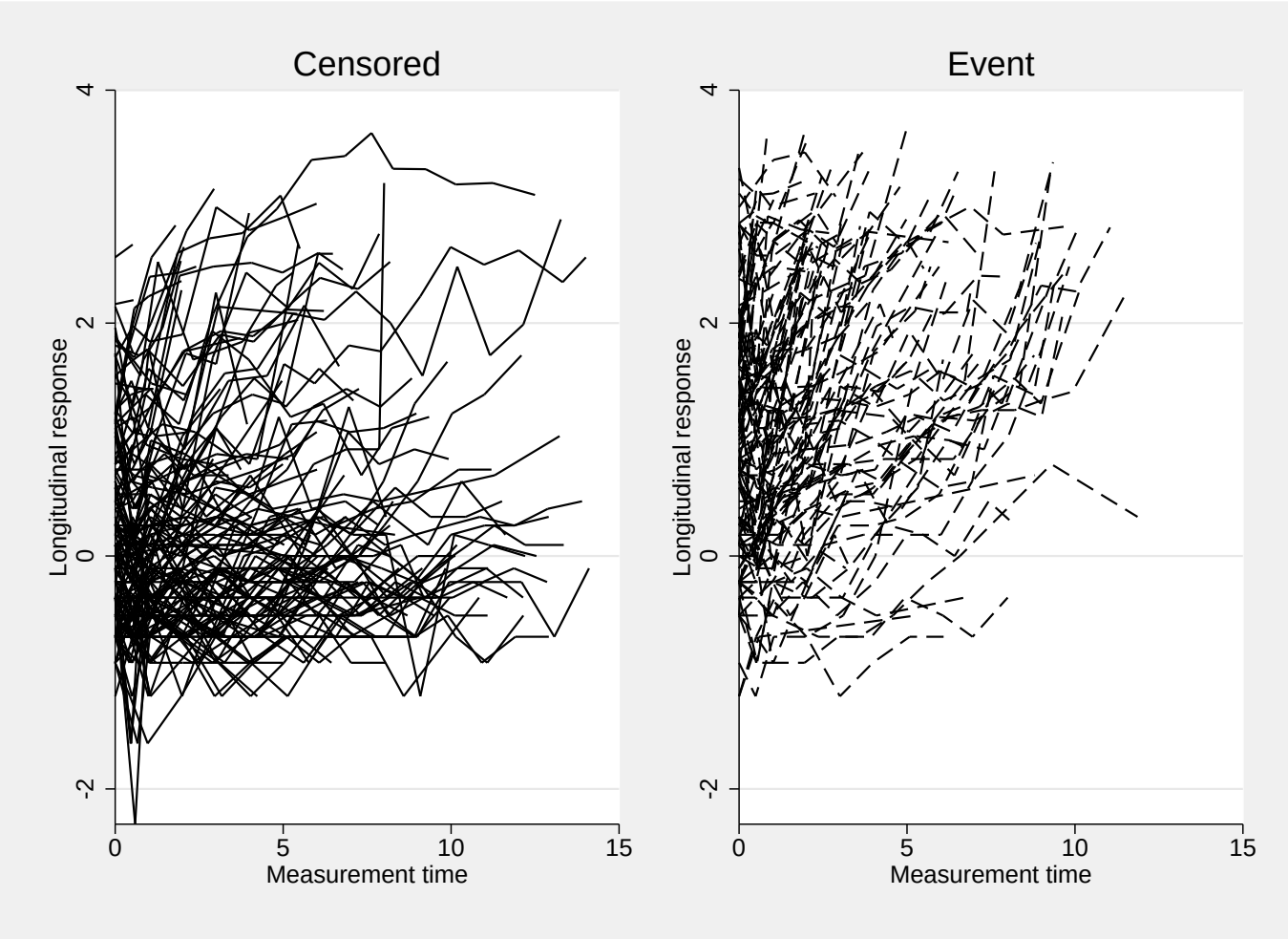
```
. use http://fmwww.bc.edu/repec/bocode/s/stjm_pbc_example_data, clear
. stset stop, enter(start) failure(event=1) id(id)

. list id logb trt start stop event if id==4, table noobs sepby(id)
```

id	logb	trt	start	stop	event
4	.5877866	D-penicil	0	.51473	0
4	.4700036	D-penicil	.51473	1.018508	0
4	.5306283	D-penicil	1.018508	1.995948	0
4	1.163151	D-penicil	1.995948	3.433359	0
4	1.308333	D-penicil	3.433359	4.002848	0
4	1.386294	D-penicil	4.002848	4.993977	0
4	1.667707	D-penicil	4.993977	5.270507	1

Lots of software now available to fit joint models

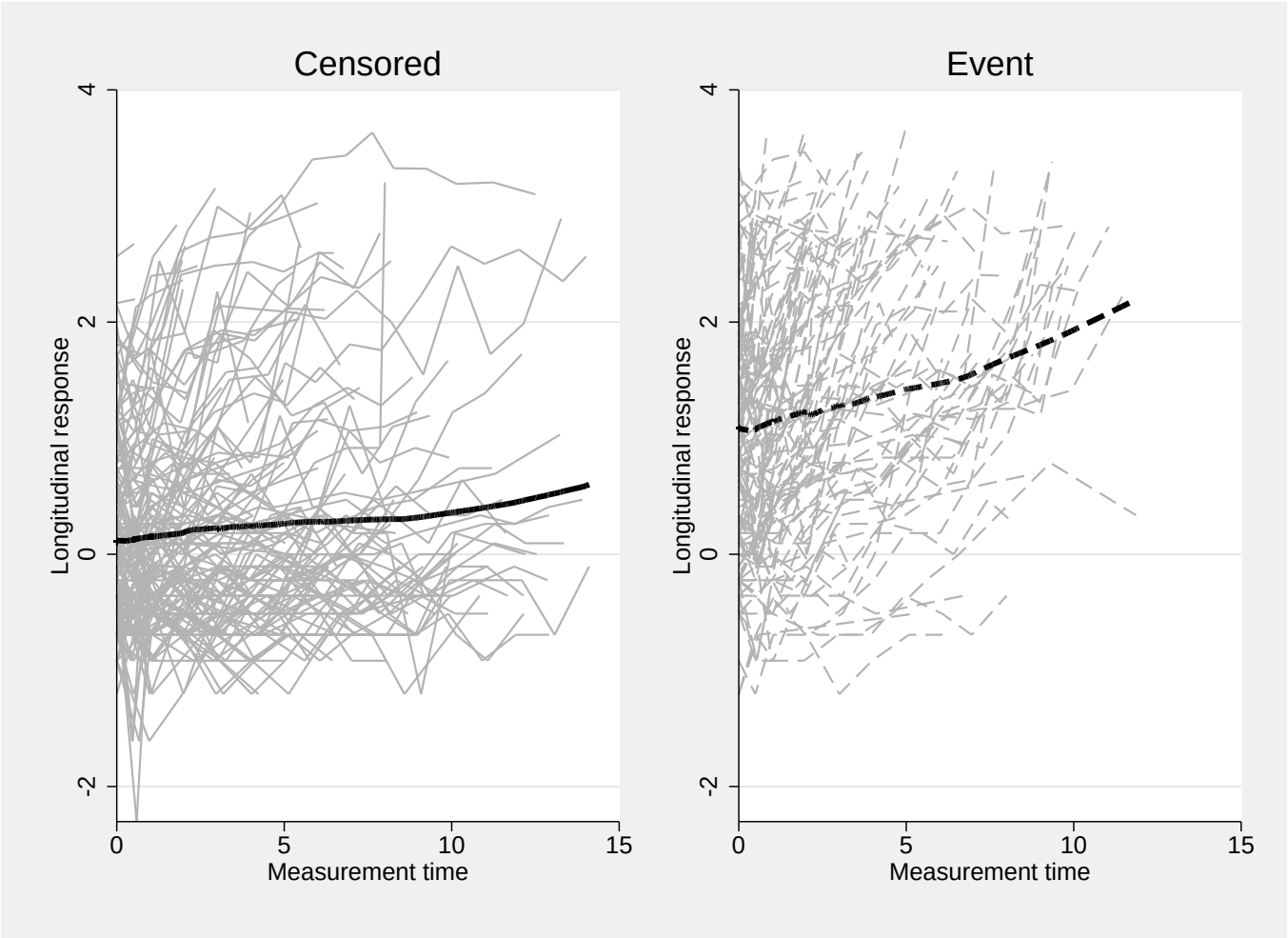
- `stjm` in Stata (Crowther et al., 2013)
- `JM` and `JMbayes` in R (Rizopoulos, 2012)
- `joineR` in R

Exploratory trajectory plots



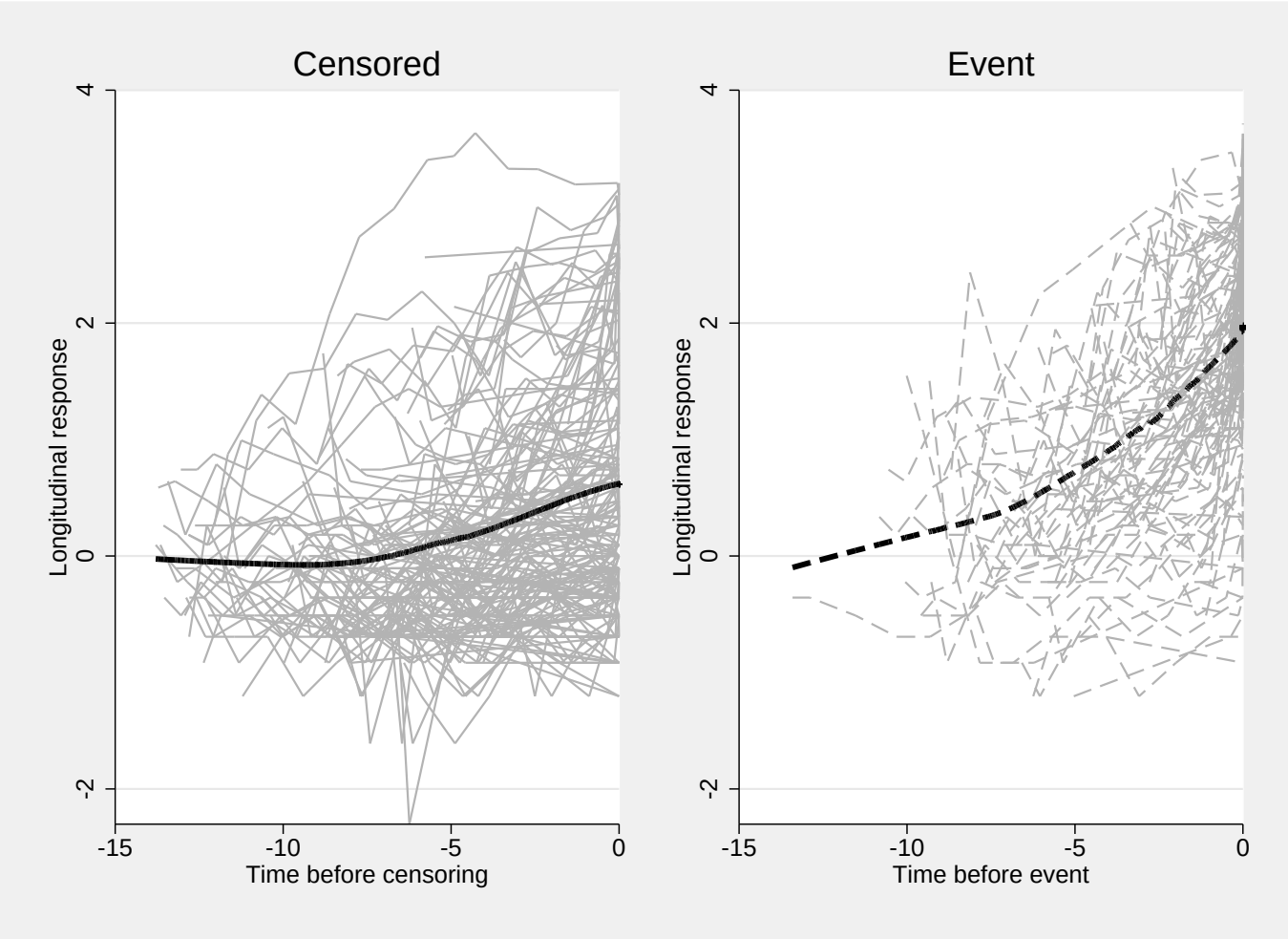
`stjmgraph logb, panel(id)`
(Crowther et al., 2013)

Exploratory trajectory plots



`stjmgraph logb, panel(id) lowess`
(Crowther et al., 2013)

Exploratory trajectory plots



`stjmgraph logb, panel(id) lowess adjust`
(Crowther et al., 2013)

Stata code for fitting TVC, two-stage and joint model

- Time-varying covariate

```
. streg logb trt, distribution(weibull) nohr
```

- Two-stage

```
. mixed logb time || id: time, covariance(unstructured)
. predict fitvals, fitted
. streg fitvals trt, distribution(weibull) nohr
```

- Joint model

```
. stjm logb , panel(id) survmodel(weibull) rfp(1) survcov(trt)
```

JMbayes code for joint model in R

```
> library(JMbayes)
```

```
# linear mixed model fit (random intercepts + random slopes)
> fitLME <- lme(log(serBilir) ~  year, random = ~ year | id, data = pbc2)
```

```
# survival Cox-PH fit
> fitSURV.cox <- coxph(Surv(years, status2) ~ drug, data = pbc2.id,
x = TRUE)
```

```
# joint model
> fitJOINTBayes <- jointModelBayes(fitLME, fitSURV.cox, timeVar="year",
param="td-value")
```

Model results

Comparing approaches,

- Per unit increase in log Bilirubin

Model	log HR	SE	95% CI	
TVC	1.308	0.085	1.142	1.475
2-stage	1.221	0.082	1.060	1.382
JM (`stjm`)	1.241	0.093	1.058	1.423
JM (`JMbayes`)	1.269	0.097	1.087	1.463

- Treatment effect (D-penicillamine vs. placebo)

Model	log HR	SE	95% CI	
TVC	-0.021	0.170	-0.355	0.313
2-stage	0.029	0.170	-0.304	0.363
JM (`stjm`)	0.044	0.179	-0.307	0.395
JM (`JMbayes`)	0.049	0.185	-0.312	0.409

Comparing association structures for joint model

Model	AIC	BIC
Current	3858.407	3914.137
Slope	3900.301	3956.032
Both	3850.974	3912.277

Outline

Estimating treatment effects

Suppose we have a treatment, u_i, that effects both the longitudinal outcome, and survival outcome. Let's assume,

$$\begin{aligned} y_i(t) &= m_i(t) + e_i(t) \\ &= (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t + \beta u_i + e_i(t) \end{aligned}$$

and

$$h(t) = h_0(t) \exp\left[\phi u_i + \alpha m_i(t)\right]$$

Estimating treatment effects

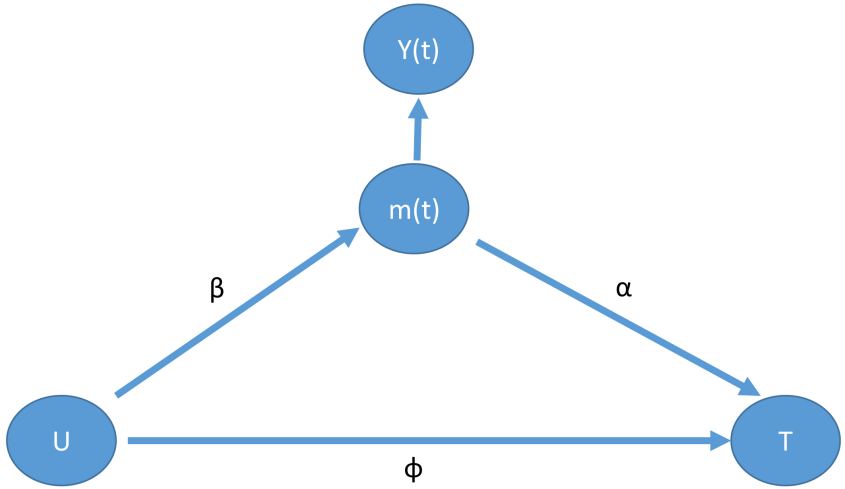
Suppose we have a treatment, u_i, that effects both the longitudinal outcome, and survival outcome. Let's assume,

$$\begin{aligned} y_i(t) &= m_i(t) + e_i(t) \\ &= (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t + \beta u_i + e_i(t) \end{aligned}$$

and

$$h(t) = h_0(t) \exp\left[\phi u_i + \alpha m_i(t)\right]$$

Because the models are linked, we have direct and indirect treatment effects on survival



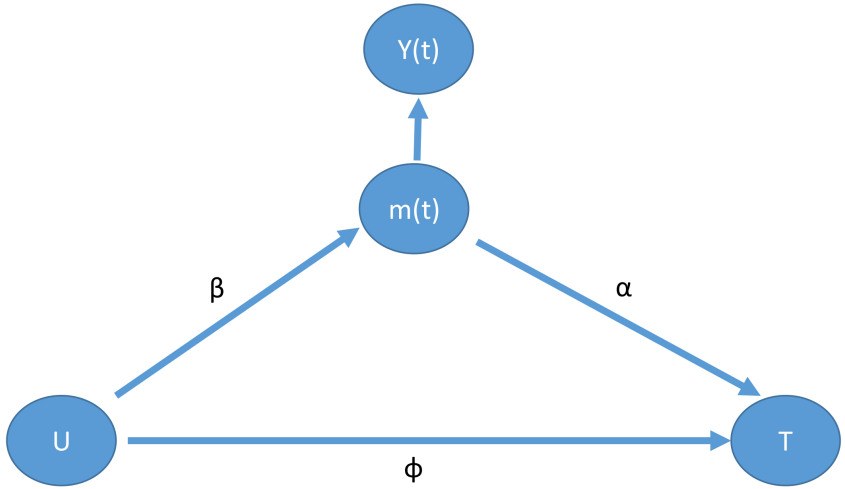
$$y_i(t) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t + \beta u_i + e_i(t)$$

and

$$h(t) = h_0(t) \exp\left[\phi u_i + \alpha m_i(t)\right]$$

We have,

- β: the direct effect of treatment on the longitudinal outcome
- ϕ: the direct effect of treatment on survival
- $\alpha\beta + \phi$: the overall treatment effect on survival



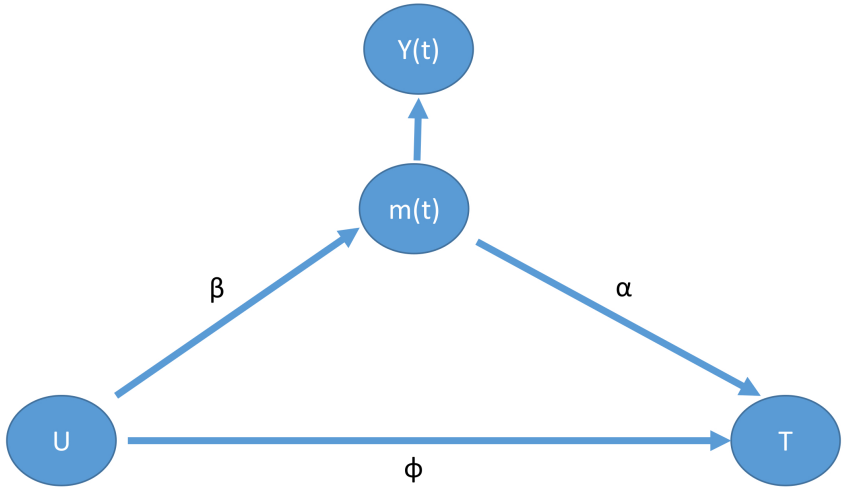
$$y_i(t) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t + \beta u_i + e_i(t)$$

and

$$h(t) = h_0(t) \exp\left[\phi u_i + \alpha m_i(t)\right]$$

We have,

- β: the direct effect of treatment on the longitudinal outcome
- ϕ: the direct effect of treatment on survival
- $\alpha\beta + \phi$: the overall treatment effect on survival



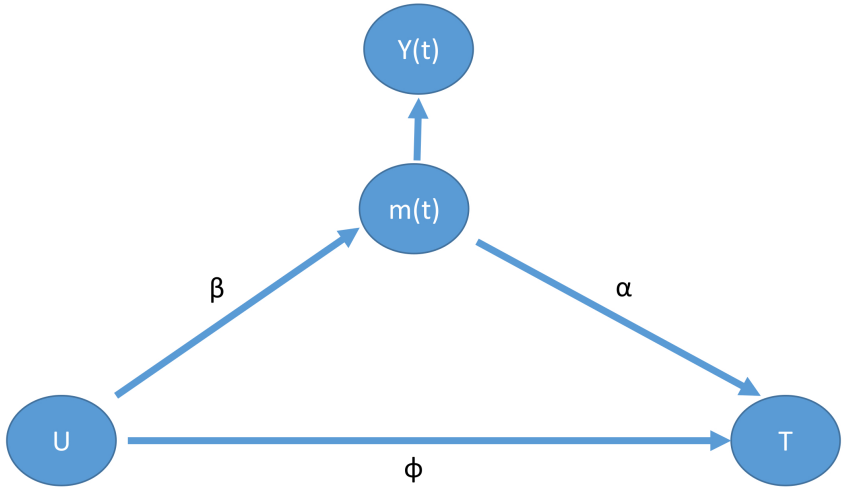
$$y_i(t) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t + \beta u_i + e_i(t)$$

and

$$h(t) = h_0(t) \exp\left[\phi u_i + \alpha m_i(t)\right]$$

We have,

- β: the direct effect of treatment on the longitudinal outcome
- ϕ: the direct effect of treatment on survival
- $\alpha\beta + \phi$: the overall treatment effect on survival



$$y_i(t) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t + \beta u_i + e_i(t)$$

and

$$h(t) = h_0(t) \exp\left[\phi u_i + \alpha m_i(t)\right]$$

We have,

- β: the direct effect of treatment on the longitudinal outcome
- ϕ: the direct effect of treatment on survival
- $\alpha\beta + \phi$: the overall treatment effect on survival

Outline

Dynamic prediction from a joint model

- Conditional on a set of biomarker measurements

$$\mathcal{Y}_i(t) = \{y_i(s), 0 \leq s < t\}$$

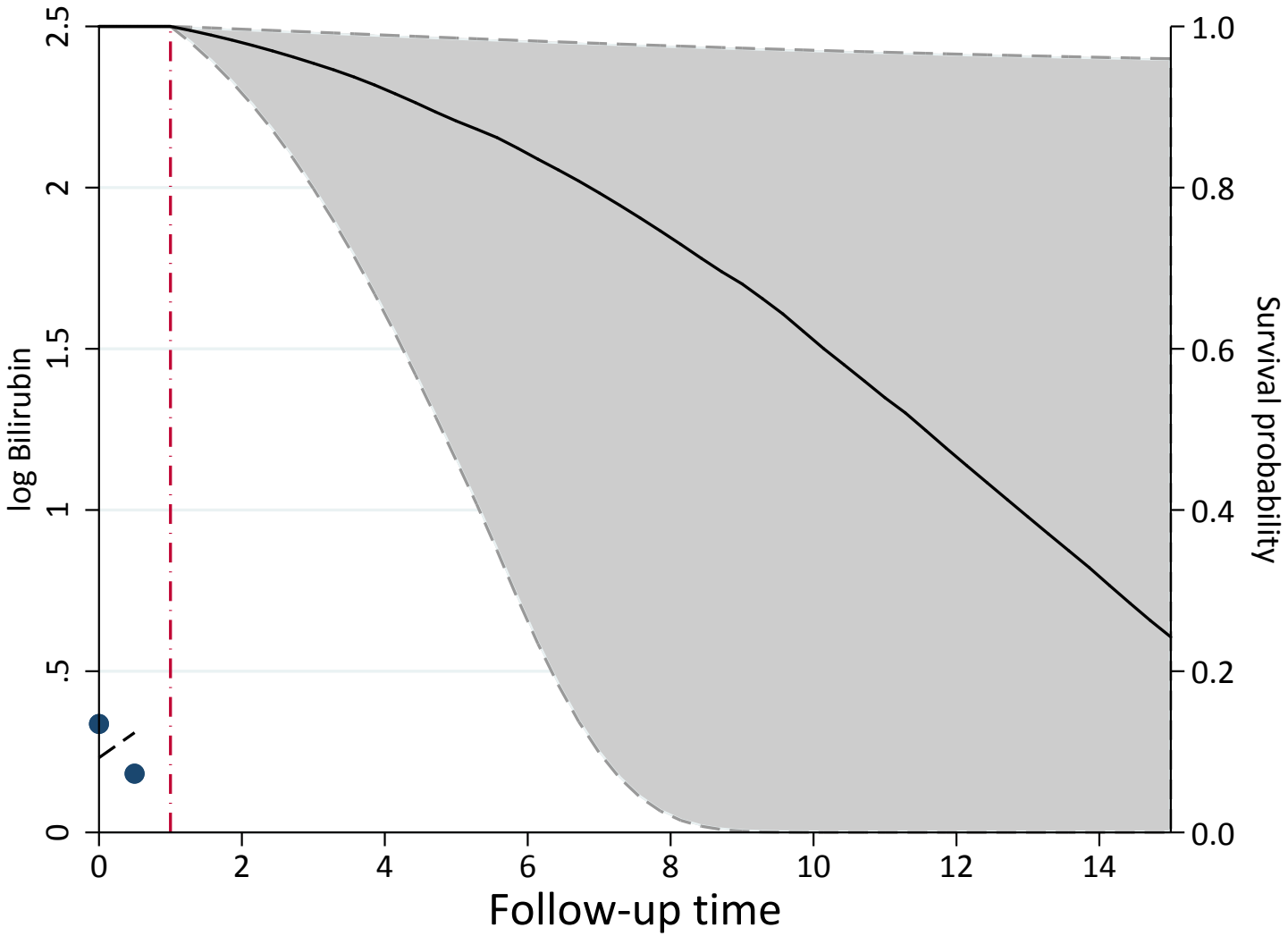
we are interested in predicting survival

$$P\{T_i^* \geq u | T_i^* > t, \mathcal{Y}_i(t), D_n\}$$

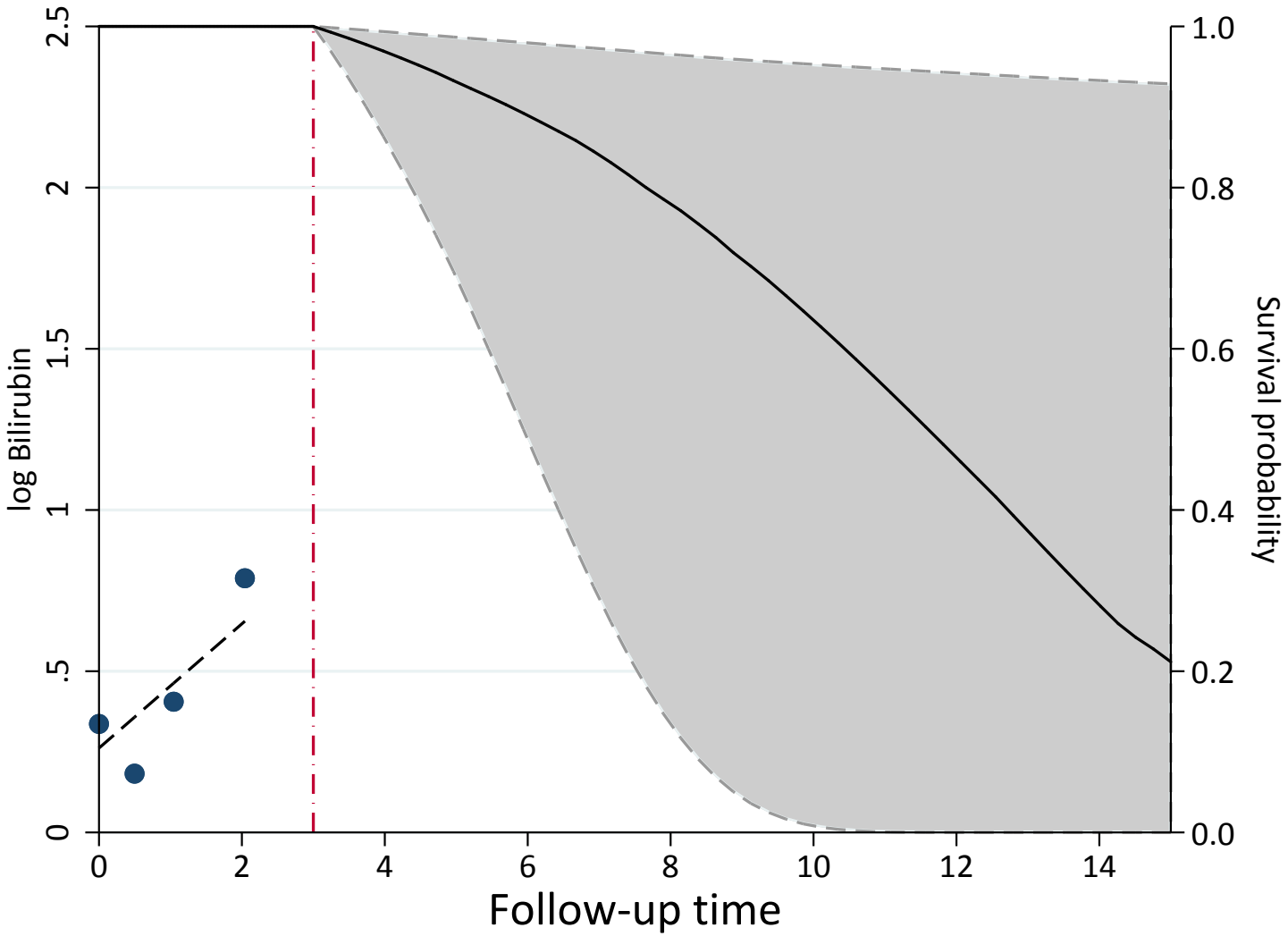
where, $u > t$, and D_n is our sample which the joint model was fitted

- Further info in (Rizopoulos, 2011)

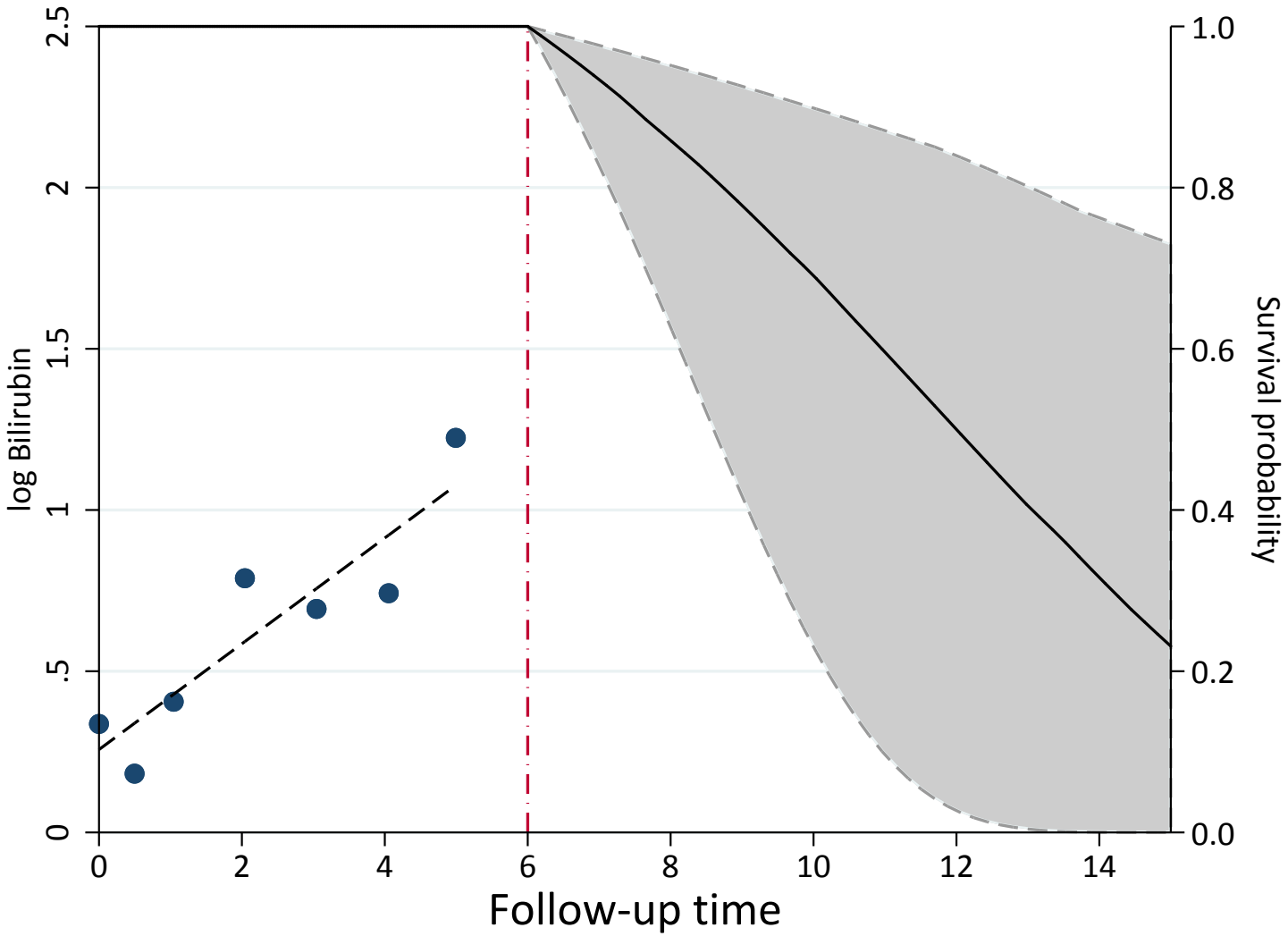
Conditional survival predictions



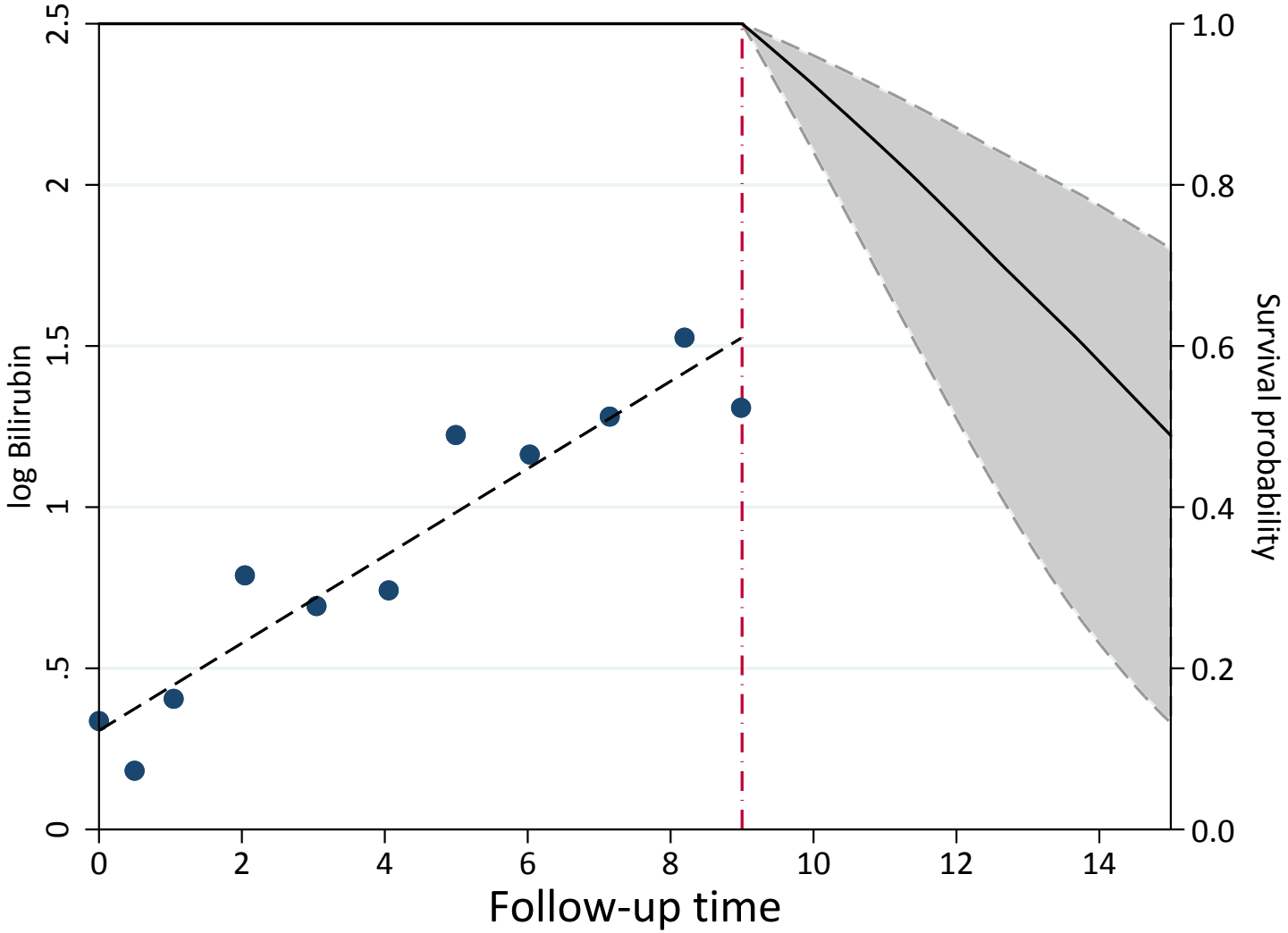
Conditional survival predictions



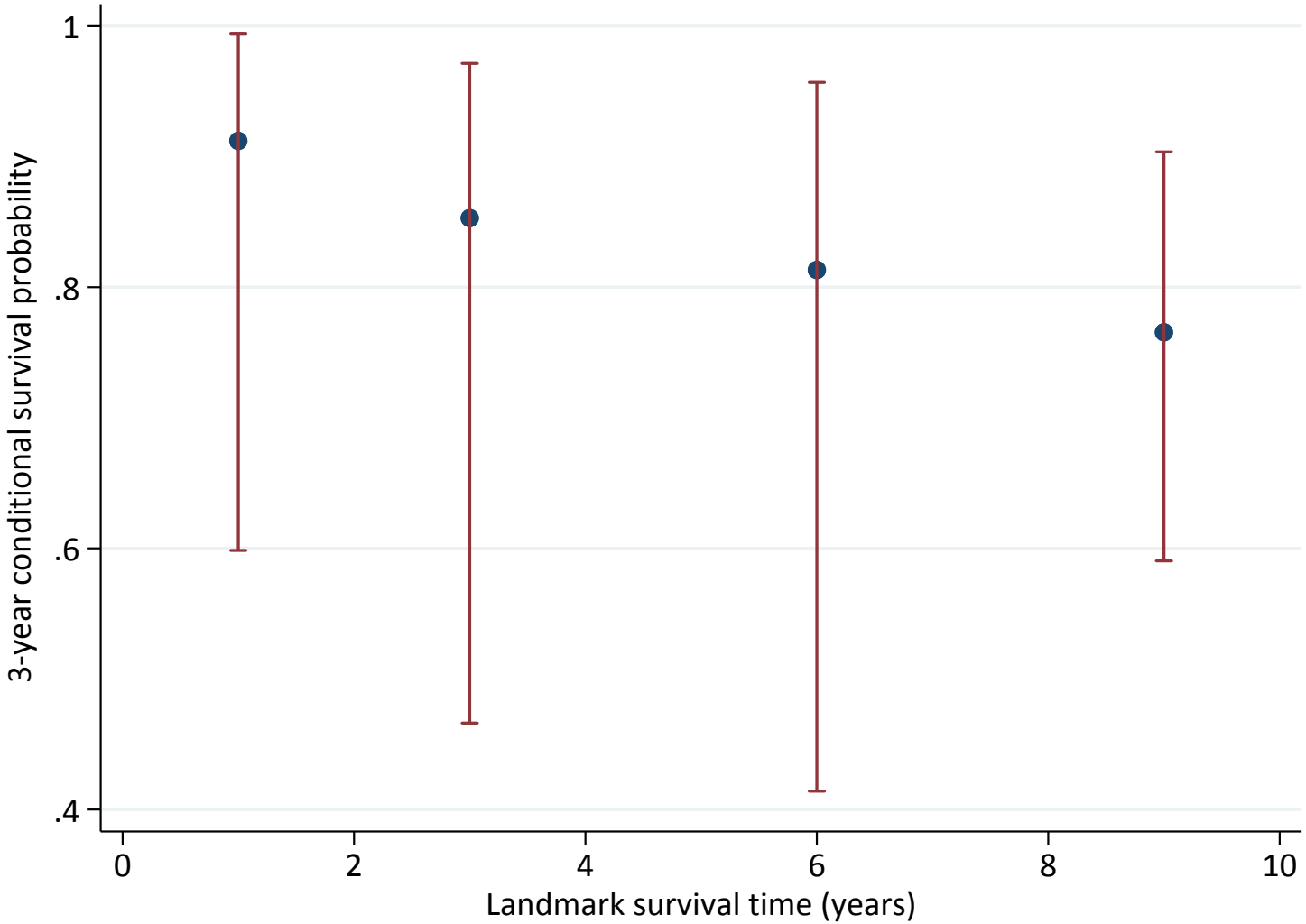
Conditional survival predictions



Conditional survival predictions



3-year conditional survival predictions



Outline

Summary

- Joint modelling provides us with a method of linking a longitudinal outcome, measured with error, to the time to an event of interest

Summary

- Joint modelling provides us with a method of linking a longitudinal outcome, measured with error, to the time to an event of interest
- It has been shown to reduce bias and maximise efficiency compared to naive approaches

Summary

- Joint modelling provides us with a method of linking a longitudinal outcome, measured with error, to the time to an event of interest
- It has been shown to reduce bias and maximise efficiency compared to naive approaches
- Failing to account for the longitudinal process causes bias in covariate effects on survival when there is a true association between outcomes

Summary

- Joint modelling provides us with a method of linking a longitudinal outcome, measured with error, to the time to an event of interest
- It has been shown to reduce bias and maximise efficiency compared to naive approaches
- Failing to account for the longitudinal process causes bias in covariate effects on survival when there is a true association between outcomes
- Ignoring the informative drop-out process leads to bias in estimates of the longitudinal trajectory

Summary

- Joint modelling provides us with a method of linking a longitudinal outcome, measured with error, to the time to an event of interest
- It has been shown to reduce bias and maximise efficiency compared to naive approaches
- Failing to account for the longitudinal process causes bias in covariate effects on survival when there is a true association between outcomes
- Ignoring the informative drop-out process leads to bias in estimates of the longitudinal trajectory
- Opportunities to utilise the joint model framework in prognostic modelling are substantial
  - Applications so far have been to datasets < 2000 patients

Extensions

- Multiple longitudinal outcomes, of different type;
- Choice of the survival submodel;
- Delayed entry;
- Competing risks;
- Recurrent and terminal events;
- Complex correlation structures for LME models;
- Many more...

See `merlin` package in Stata and R (Crowther, 2018) for general mixed effects regression of multivariate outcomes

References I

Michael J Crowther. merlin-a unified modelling framework for data analysis and methods development in stata. *arXiv preprint arXiv:1806.01615*, 2018.

Michael J. Crowther, Keith R. Abrams, and Paul C. Lambert. Joint modeling of longitudinal and survival data. *Stata Journal*, 13(1):165–184(20), 2013.

Cheryl L. Faucett and Duncan C. Thomas. Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in Medicine*, 15(15):1663–1685, 1996. doi: 10.1002/(SICI)1097-0258(19960815)15:15⟨1663::AID-SIM294⟩3.0.CO;2-1.

Robin Henderson, Peter Diggle, and Angela Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000. doi: 10.1093/biostatistics/1.4.465.

Josè C. Pinheiro and Douglas M. Bates. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4(1):12–35, 1995. doi: 10.1080/10618600.1995.10474663.

Cécile Proust-Lima and Jeremy M. G. Taylor. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics*, 10(3):535–549, 2009. doi: 10.1093/biostatistics/kxp009.

Dimitris Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–829, 2011. doi: 10.1111/j.1541-0420.2010.01546.x.

Dimitris Rizopoulos. *Joint models for longitudinal and time-to-event data with applications in R*. Chapman & Hall, 2012.

Michael J. Sweeting and Simon G. Thompson. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5):750–763, 2011. doi: 10.1002/bimj.201100052.

Michael S. Wulfsohn and Anastasios A. Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1):330–339, 1997.