How to impute missing data in Cox regression

New developments incorporating non-proportional hazards

Ruth Keogh

Department of Medical Statistics,
London School of Hygiene and Tropical Medicine

RSS Conference 2018

STRATOS Initiative

http://www.stratos-initiative.org/

Objective

To provide accessible and accurate guidance in the design and analysis of observational studies

- Providing evidence-based guidance regarding (new or existing) methods
- Identifying unmet (analytical) needs i.e. those challenges that need further methodological developments
- Stimulating collaboration between different Topic Groups (TG) and/or Panels whose joint expertise will be necessary to address such new analytical challenges

Sauerbrei et al, Stats Med 2014

STRATOS Initiative: targeting 3 types of researchers

Level 1: Applied analysts

- provide guidance on usable and appropriate methods for routine analysis

Level 2: Experienced analysts

- provide guidance on advantages and disadvantages of competing approaches

Level 3: Expert statisticians in specific areas

- improve statistical methods where needed and provide comparisons of state of the art methods

The work in this talk is aimed at level 3 researchers.

Connection of this work with STRATOS topic groups

1. Missing data
2. Selection of variables and functional forms in multivariable analysis
3. Initial data analysis
4. Measurement error and misclassification
5. Study design
6. Evaluating diagnostic tests and prediction models
7. Causal inference
8. Survival analysis
9. High-dimensional data

Background

- Cox regression is the most widely used analysis in time-to-event studies and missing data are common in these studies
- Two methods for multiple imputation (MI) of missing covariate data in Cox regression have been described

Problem

- We typically want to assess the proportional hazards assumption
- Sometimes we want to estimate time-varying effect of an exposure

Is it OK to use the existing imputation methods, or are extensions needed?

Background

- Cox regression is the most widely used analysis in time-to-event studies and missing data are common in these studies
- Two methods for multiple imputation (MI) of missing covariate data in Cox regression have been described

Problem

- We typically want to assess the proportional hazards assumption
- Sometimes we want to estimate time-varying effect of an exposure

Is it OK to use the existing imputation methods, or are extensions needed?

RESEARCH ARTICLE

WILEY Statistics in Medicine

Multiple imputation in Cox regression when there are time-varying effects of covariates

Ruth H. Keogh[1] | Tim P. Morris[2]

[1]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

[2]London Hub for Trials Methodology Research, MRC Clinical Trials Unit at UCL, Aviation House, London, UK

Correspondence
Ruth H. Keogh, Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.
Email: ruth.keogh@lshtm.ac.uk

In Cox regression, it is important to test the proportional hazards assumption and sometimes of interest in itself to study time-varying effects (TVEs) of covariates. TVEs can be investigated with log hazard ratios modelled as a function of time. Missing data on covariates are common and multiple imputation is a popular approach to handling this to avoid the potential bias and efficiency loss resulting from a "complete-case" analysis. Two multiple imputation methods have been proposed for when the substantive model is a Cox proportional hazards regression: an approximate method (Imputing missing covariate values for the Cox model in Statistics in Medicine (2009) by White and Royston) and a substantive-model-compatible method (Multiple imputation of covariates by fully conditional specification: accommodating the substantive model in Statis-

Background to multiple imputation (MI)

Multiple imputation in general

Aim: To fit an analysis model $Y \sim X_1, X_2$

Simple set-up:

- X_1 has missing data
- X_2 is fully observed

Naive approach: Complete case analysis

Multiple imputation (MI)

For a partially missing exposure X_1, fully observed covariates X_2

1. Draw values of X_1 from $X_1 | X_2, Y$
2. Obtain several imputed data sets
3. Fit the analysis model in each imputed data set and combine parameter estimates using Rubin's Rules

Multiple imputation in general

Aim: To fit an analysis model $Y \sim X_1, X_2$

Simple set-up:

- X_1 has missing data
- X_2 is fully observed

Naive approach: Complete case analysis

Multiple imputation (MI)

For a partially missing exposure X_1, fully observed covariates X_2

1. Draw values of X_1 from $X_1 | X_2, Y$
2. Obtain several imputed data sets
3. Fit the analysis model in each imputed data set and combine parameter estimates using Rubin's Rules

Multiple imputation in Cox Regression

Main challenge

What is the distribution of $X_1|X_2, Y$?

Cox proportional hazards model

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1 + \beta_{X2}X_2}$$

- T: Event or censoring time
- D: Event indicator

Distribution of interest for the imputation:

$$X_1|X_2, T, D$$

How do we draw from this distribution?

Multiple imputation in Cox Regression

Main challenge

What is the distribution of $X_1|X_2, Y$?

Cox proportional hazards model

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1+\beta_{X2}X_2}$$

- T: Event or censoring time
- D: Event indicator

Distribution of interest for the imputation:

$$X_1|X_2, T, D$$

How do we draw from this distribution?

Multiple imputation in Cox Regression

Main challenge

What is the distribution of $X_1|X_2, Y$?

Cox proportional hazards model

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1 + \beta_{X2}X_2}$$

- T: Event or censoring time
- D: Event indicator

Distribution of interest for the imputation:

$$X_1|X_2, T, D$$

How do we draw from this distribution?

Multiple imputation in Cox Regression

Main challenge

What is the distribution of $X_1|X_2, Y$?

Cox proportional hazards model

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1 + \beta_{X2}X_2}$$

- T: Event or censoring time
- D: Event indicator

Distribution of interest for the imputation:

$$X_1|X_2, T, D$$

How do we draw from this distribution?

Multiple imputation in Cox Regression

Cox proportional hazards model

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1 + \beta_{X2}X_2}$$

- T: Event or censoring time
- D: Event indicator

We might consider the imputation model

$$X_1|X_2, T, D \sim N(\alpha_0 + \alpha_1 X_2 + \alpha_2 D + \alpha_4 T, \sigma^2)$$

- But both models cannot be true.
- The models are incompatible

Two conditional models are said to be incompatible if there exists no joint model for which the conditionals (for the relevant variables) equal these conditional models. [Bartlett et al. 2015]

Multiple imputation in Cox Regression

Cox proportional hazards model

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1 + \beta_{X2}X_2}$$

- T: Event or censoring time
- D: Event indicator

We might consider the imputation model

$$X_1|X_2, T, D \sim N(\alpha_0 + \alpha_1 X_2 + \alpha_2 D + \alpha_4 T, \sigma^2)$$

- But both models cannot be true.
- The models are incompatible

Two conditional models are said to be incompatible if there exists no joint model for which the conditionals (for the relevant variables) equal these conditional models. [Bartlett et al. 2015]

Multiple imputation in Cox Regression

Cox proportional hazards model

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1 + \beta_{X2}X_2}$$

- T: Event or censoring time
- D: Event indicator

We might consider the imputation model

$$X_1|X_2, T, D \sim N(\alpha_0 + \alpha_1 X_2 + \alpha_2 D + \alpha_4 T, \sigma^2)$$

- But both models cannot be true.
- The models are incompatible

Two conditional models are said to be incompatible if there exists no joint model for which the conditionals (for the relevant variables) equal these conditional models. [Bartlett et al. 2015]

Statistical Science
1994, Vol. 9, No. 4, 538–573

Multiple-Imputation Inferences with Uncongenial Sources of Input

Xiao-Li Meng

Abstract. Conducting sample surveys, imputing incomplete observations, and analyzing the resulting data are three indispensable phases of modern practice with public-use data files and with many other statistical applications. Each phase inherits different input, including the information preceding it and the intellectual assessments available, and aims to provide output that is one step closer to arriving at statistical inferences with scientific relevance. However, the role of the imputation phase has often been viewed as merely providing computational convenience for users of data. Although facilitating computation is very important, such a viewpoint ignores the imputer's assessments and information inaccessible to the users. This view underlies the recent controversy over the validity of multiple-imputation inference when a procedure for analyzing multiply imputed data sets cannot be derived from (is "uncongenial" to) the model adopted for multiple imputation. Given sensible imputations and complete-data analysis procedures, inferences from standard multiple-imputation combining rules are typically superior to, and thus different from, users' incomplete-data analyses. The latter may suffer from serious

"The imputer's task is easy to state but hard to implement"

Existing methods for imputation in Cox regression

- White & Royston (2009)
- Bartlett et al. (2015)

STATISTICS IN MEDICINE
Statist. Med. 2009; **28**:1982–1998
Published online 19 May 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3618

Imputing missing covariate values for the Cox model

Ian R. White[1,*,†] and Patrick Royston[2]

[1]*MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, U.K.*
[2]*MRC Clinical Trials Unit, Cancer Group, London, U.K.*

White & Royston's method: MI-Approx

Cox proportional hazards model

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1 + \beta_{X2}X_2}$$

Imputation model arises from an approximation to the distribution

$$p(X_1|X_2, T, D)$$

The imputation model: MI-Approx

$$X_1 \sim X_2 + D + \widehat{H}(T)$$

e.g. linear or logistic regression

$\widehat{H}(T)$ is the Nelson-Aalen estimate of the cumulative hazard

White & Royston's method: MI-Approx

Cox proportional hazards model

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1 + \beta_{X2}X_2}$$

Imputation model arises from an approximation to the distribution

$$p(X_1|X_2, T, D)$$

The imputation model: MI-Approx

$$X_1 \sim X_2 + D + \widehat{H}(T)$$

e.g. linear or logistic regression

$\widehat{H}(T)$ is the Nelson-Aalen estimate of the cumulative hazard

Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–26

Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0962280214521348
smm.sagepub.com
SAGE

Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model

Jonathan W Bartlett,[1] Shaun R Seaman,[2] Ian R White[2] and James R Carpenter[1,3] for the Alzheimer's Disease Neuroimaging Initiative*

The basic idea...

- Draw *potential* values of X_1 from a proposal distribution $p(X_1|X_2)$
- Use a rejection rule to decide whether or not to accept the potential imputed values of X_1 as imputed values from the desired distribution $p(X_1|X_2, T, D)$

Bartlett et al's method: MI-SMC

Cox proportional hazards model

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1 + \beta_{X2}X_2}$$

1. Obtain initial estimates for β_{X_1}, β_{X_2}
2. Draw values $\beta_{X_1}^{(m)}, \beta_{X_2}^{(m)}$, and calculate $H_0^{(m)}(t)$
3. Fit the proposal distribution $p(X_1|X_2)$ and take draws of parameter values from their approx joint posterior
4. Draw a value X_1^* from the proposal distribution
5. Draw a value $U \sim \text{Uniform}(0,1)$. Accept X_1^* if

$$\begin{cases} U \leq \exp\{-H_0^{(m)}(t)e^{\beta_{X1}^{(m)}X_1^* + \beta_{X1}^{(m)}}X_2\} & \text{if } D = 0 \\ U \leq H_0^{(m)}(t)\exp\{1 + \beta_{X1}^{(m)}X_1^* + \beta_{X2}^{(m)}X_2 - H_0^{(m)}(t)e^{\beta_{X1}^{(m)}X_1^* + \beta_{X2}^{(m)}X_2}\} & \text{if } D = 1 \end{cases}$$

6. Repeat until the imputed X_1 values have converged to a stationary distribution.

Various extensions

- Missingness in multiple covariates
- Competing risks and censoring depending on covariates
- Left-truncation

White, Royston, Wood. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011.

Borgan & Keogh. Nested case-control studies: should one break the matching? *Lifetime Data Analysis* 2015.

Bartlett & Taylor. Missing covariates in competing risks analysis. *Biostatistics* 2016.

Cox regression with Time-Varying Effects (TVE)

Cox regression with Time-Varying Effects (TVE)

Cox proportional hazards model

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1} X_1 + \beta_{X2} X_2}$$

- Cox regression analyses usually incorporates assessment of the proportional hazards assumption
- If the proportional hazards assumption is not met, we may allow time-varying effects (TVE)
- Sometimes we are interested in TVEs from the outset

Cox regression with Time-Varying Effects (TVE)

Cox proportional hazards model

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1 + \beta_{X2}X_2}$$

Extended Cox models with TVEs

$$h(t|X_1, X_2) = h_0(t)e^{f_{X1}(t;\beta_{X1})X_1 + f_{X2}(t;\beta_{X2})X_2}$$

Example

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1 + \gamma_{X1}X_1 \times t + \beta_{X2}X_2 + \gamma_{X2}X_2 \times t}$$

A test of $\gamma_{X1} = 0$ is a test of the proportional hazards assumption for X_1.

Cox regression with Time-Varying Effects (TVE)

Cox proportional hazards model

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1 + \beta_{X2}X_2}$$

Extended Cox models with TVEs

$$h(t|X_1, X_2) = h_0(t)e^{f_{X1}(t;\beta_{X1})X_1 + f_{X2}(t;\beta_{X2})X_2}$$

Example

$$h(t|X_1, X_2) = h_0(t)e^{\beta_{X1}X_1 + \gamma_{X1}X_1 \times t + \beta_{X2}X_2 + \gamma_{X2}X_2 \times t}$$

A test of $\gamma_{X1} = 0$ is a test of the proportional hazards assumption for X_1.

Cox regression with Time-Varying Effects (TVE)

Extended Cox models with TVEs

$$h(t|X_1, X_2) = h_0(t)e^{f_{X1}(t;\beta_{X1})X_1 + f_{X2}(t;\beta_{X2})X_2}$$

What is $p(X_1|X_2, T, D)$?

Aims

1. To extend the two MI methods to accommodate TVEs
2. To investigate their performance in simulation studies

MI-Approx extended for TVEs: MI-TVE-Approx

Extended Cox model with TVEs

$$h(t|X_1, X_2) = h_0(t)e^{f_{X1}(t,\beta_{X1})X_1 + f_{X2}(t,\beta_{X2})X_2}$$

The imputation model: MI-TVE-Approx

$$X_1 \sim X_2 + f_{X1}(T)D + \widehat{H}(T)$$

e.g. linear or logistic regression

MI-SMC extended for TVEs: MI-TVE-SMC

Extended Cox model with TVEs

$$h(t|X_1, X_2) = h_0(t)e^{f_{X1}(t,\beta_{X1})X_1 + f_{X2}(t,\beta_{X2})X_2}$$

1. Obtain initial estimates for β_{X_1}, β_{X_2}
2. Draw values $\beta_{X_1}^{(m)}, \beta_{X_2}^{(m)}$, and calculate $H_0^{(m)}(t)$
3. Fit the proposal distribution $p(X_1|X_2)$ and take draws of parameter values from their approx joint posterior
4. Draw a value X_1^* from the proposal distribution
5. Draw a value $U \sim \mathrm{Uniform}(0,1)$. Accept X_1^* if

$$\begin{aligned}
U &\leq \exp\{-H_0^{(m)}(t)e^{f_{X1}(t,\beta_{X1}^{(m)})X_1^* + f_{X2}(t,\beta_{X2}^{(m)})X_2}\} \qquad \text{if } D = 0 \\
U &\leq h_0^{(m)}(t)\exp\{1 + f_{X1}(t,\beta_{X1}^{(m)})X_1^* + f_{X2}(t,\beta_{X2}^{(m)})X_2 - \\
&\quad \int_0^t h_0^{(m)}(u)e^{f_{X1}(u,\beta_{X1}^{(m)})X_1^* + f_{X2}(u,\beta_{X2}^{(m)})X_2} du\} \qquad \text{if } D = 1
\end{aligned}$$

6. Repeat until the imputed X_1 values have converged to a stationary distribution.

Practical considerations

- Functional form for time-varying effects (TVE)
- How to testing the proportional hazards assumption after MI?

Form of the TVEs

Extended Cox model with TVEs

$$h(t|X_1, X_2) = h_0(t)e^{f_{X1}(t,\beta_{X1})X_1 + f_{X2}(t,\beta_{X2})X_2}$$

- Simple pre-specified forms (e.g. Quantin 1999), e.g.

$$f_X(t) = \beta_{X0} + \beta_{X1} t$$

- Step function (e.g. Gore et al 1984)
- Fractional polynomials (e.g. Royston & Sauerbrei 2007)
- Restricted cubic splines (e.g. Hess 1994)

Form of the TVEs

Extended Cox model with TVEs

$$h(t|X_1, X_2) = h_0(t)e^{f_{X1}(t,\beta_{X1})X_1 + f_{X2}(t,\beta_{X2})X_2}$$

- Simple pre-specified forms (e.g. Quantin 1999), e.g.

$$f_X(t) = \beta_{X0} + \beta_{X1}t$$

- Step function (e.g. Gore et al 1984)
- Fractional polynomials (e.g. Royston & Sauerbrei 2007)
- Restricted cubic splines (e.g. Hess 1994)

Restricted cubic spline form for the TVEs

Extended Cox model with TVEs

$$h(t|X_1, X_2) = h_0(t)e^{f_{X1}(t,\beta_{X1})X_1 + f_{X2}(t,\beta_{X2})X_2}$$

Restricted cubic spline with L knots at $u_1, \ldots, u_L$:

$$f_X(t;\beta_X) = \beta_{X0} + \beta_{X1}t + \sum_{i=1}^{L-2} \theta_{Xi} \left\{ (t-u_i)_+^3 - \left(\frac{(t-u_{L-1})_+^3 (u_L - u_i)}{(u_L - u_{L-1})} \right) + \left(\frac{(t-u_L)_+^3 (u_{L-1} - u_i)}{(u_L - u_{L-1})} \right) \right\}$$

where $(t-u_i)_+$ takes value $(t-u_i)$ if $(t-u_i) > 0$ and 0 otherwise.

We used 5 knots at percentiles of the event time distribution:

(5, 25, 50, 75, 95)

Testing the proportional hazards assumption

Extended Cox model with TVEs

$$h(t|X_1, X_2) = h_0(t)e^{f_{X1}(t,\beta_{X1})X_1 + f_{X2}(t,\beta_{X2})X_2}$$

$$f_X(t;\beta_X) = \beta_{X0} + \beta_{X1}t + \sum_{i=1}^{3}\theta_{Xi}\left\{(t-u_i)_+^3 - \left(\frac{(t-u_4)_+^3(u_5-u_i)}{(u_5-u_4)}\right) + \left(\frac{(t-u_5)_+^3(u_4-u_i)}{(u_5-u_4)}\right)\right\}.$$

We can test the proportional hazards assumption by a joint Wald test of the relevant parameters:

$$\beta_{X1} = \theta_{X1} = \theta_{X2} = \theta_{X3} = 0$$

Testing the proportional hazards assumption

Testing the PH assumption in the context of MI:

1. Perform the imputation
2. Fit the substantive model (Cox model with TVEs for all covariates) to each imputed data set
3. Combine estimates using Rubin's Rules
4. Perform a joint Wald test using the pooled estimates

Wood et al. How should variable selection be performed with multiply imputed data? *Stat. Med.* 2008.

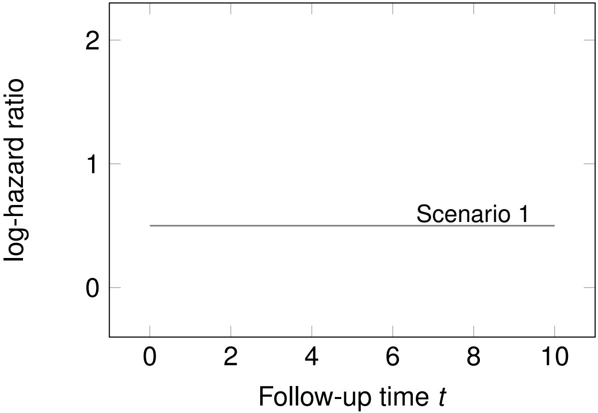
Morris et al. Combining fractional polynomial model building with multiple imputation. *Stat. Med.* 2016.

Simulation study

Simulation study

Extended Cox model with TVEs

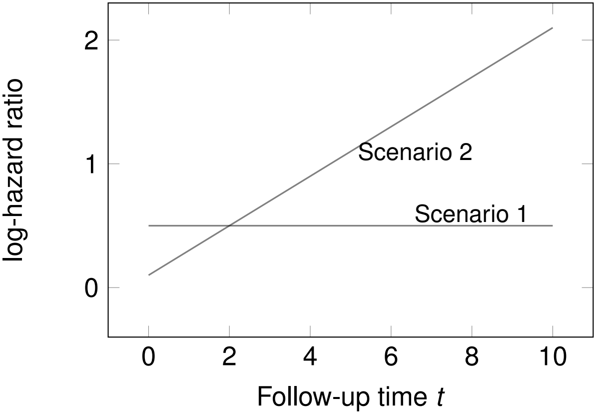
$$h(t|X_1, X_2) = h_0(t)e^{f_{X1}(t,\beta_{X1})X_1 + \beta_{X2}X_2}$$



Simulation study

Extended Cox model with TVEs

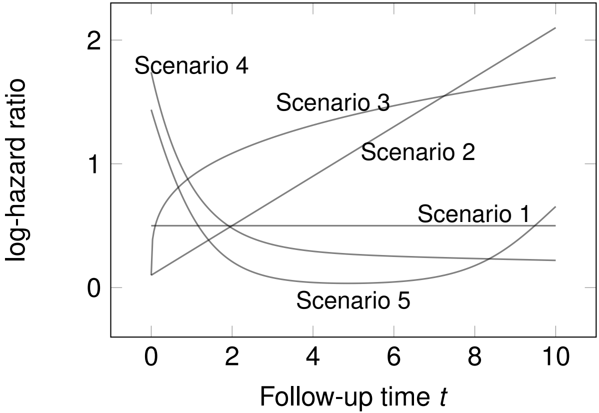
$$h(t|X_1, X_2) = h_0(t)e^{f_{X1}(t,\beta_{X1})X_1 + \beta_{X2}X_2}$$



Simulation study

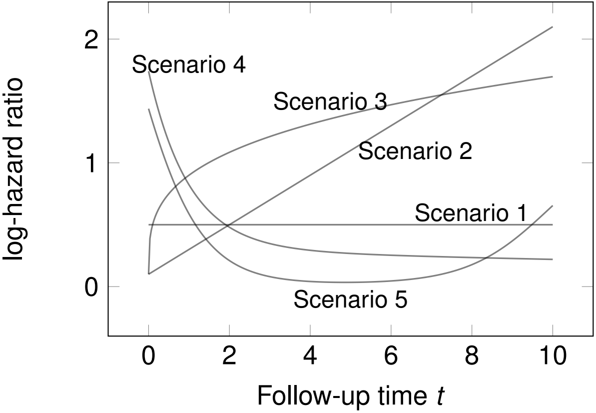
Extended Cox model with TVEs

$$h(t|X_1, X_2) = h_0(t)e^{f_{X1}(t,\beta_{X1})X_1+\beta_{X2}X_2}$$



Simulation study

- $n = 2000$
- X_1, X_2 both binary or bivariate normal
- MAR in 30% of X_1 and X_2



Simulation study: Methods

Methods performed

- Complete-data analysis (before missing data introduced)
- Complete-case analysis
- Existing methods: MI-Approx and MI-SMC
- Extended methods: MI-TVE-Approx and MI-TVE-SMC

Functional form for the TVE:

- In MI-TVE-Approx and MI-TVE-SMC we assume that the TVEs are restricted cubic splines with 5 knots
- The Cox model is fitted with TVEs of the same functional form

Simulation study: Methods

Methods performed

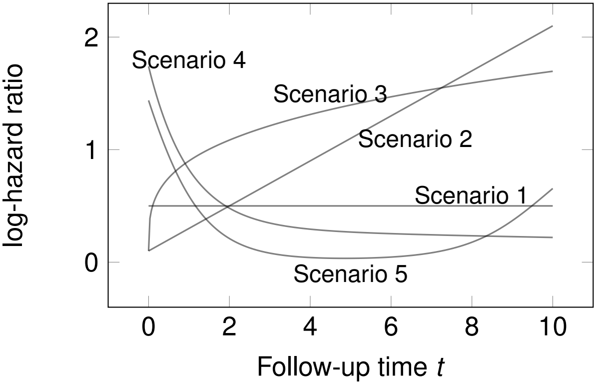
- Complete-data analysis (before missing data introduced)
- Complete-case analysis
- Existing methods: MI-Approx and MI-SMC
- Extended methods: MI-TVE-Approx and MI-TVE-SMC

Functional form for the TVE:

- In MI-TVE-Approx and MI-TVE-SMC we assume that the TVEs are restricted cubic splines with 5 knots
- The Cox model is fitted with TVEs of the same functional form

Simulation study: performance measures

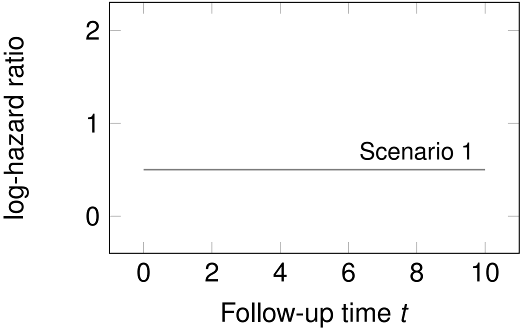
1. Test for proportional hazards: Type I error, Power
2. Mean estimated curve for the TVE: comparison with true curve



Test for proportional hazards: Scenario 1

- Percentage of simulations in which the null hypothesis of no time-varying effect is rejected
- Type I error

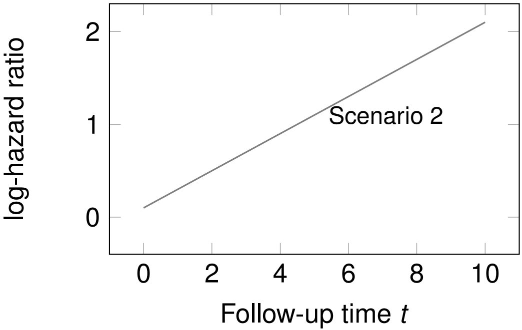
	X1	X2
Complete data	3	3
Complete case	2	3
MI-Approx	0	0
MI-SMC	0	0
MI-TVE-Approx	2	3
MI-TVE-SMC	3	4



Test for proportional hazards: Scenario 2

- Percentage of simulations in which the null hypothesis of no time-varying effect is rejected
- Power

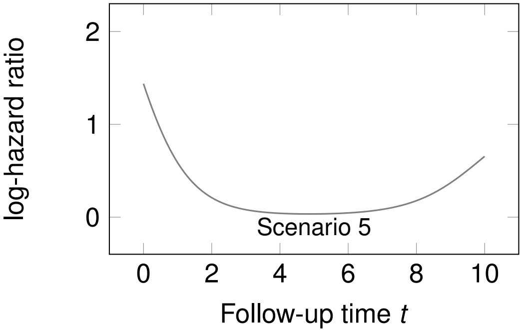
	X1	X2
Complete data	89	3
Complete case	42	3
MI-Approx	21	0
MI-SMC	17	0
MI-TVE-Approx	67	3
MI-TVE-SMC	68	6



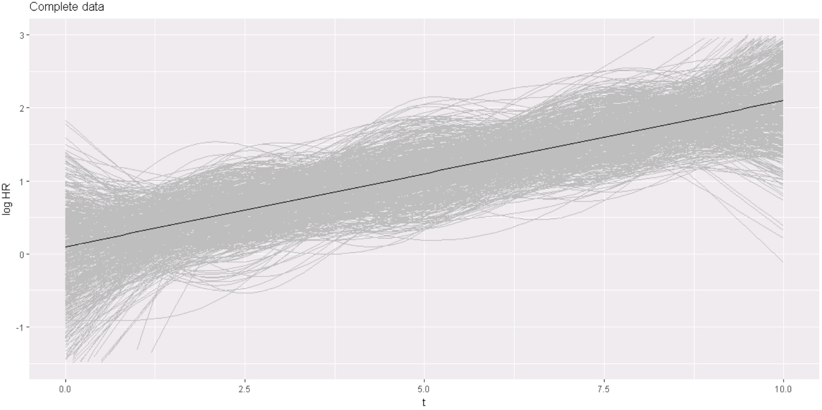
Test for proportional hazards: Scenario 5

- Percentage of simulations in which the null hypothesis of a time-varying effect is rejected
- Power

	X1	X2
Complete data	45	4
Complete case	14	3
MI-Approx	2	0
MI-SMC	1	0
MI-TVE-Approx	21	2
MI-TVE-SMC	27	5

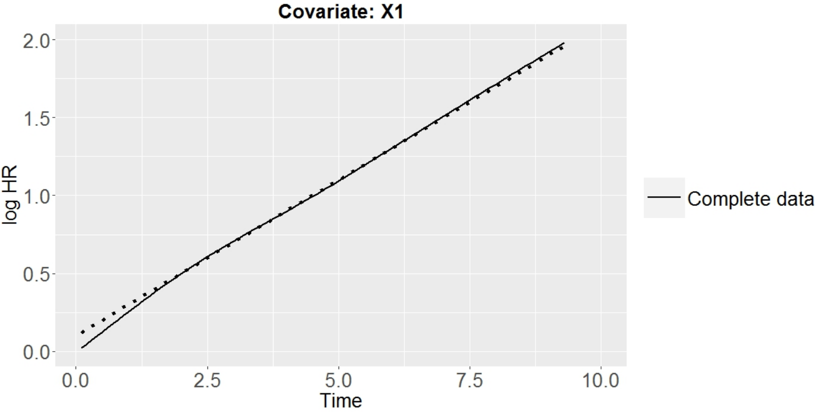


Simulation results: Mean estimated curve
Binary X, scenario 2



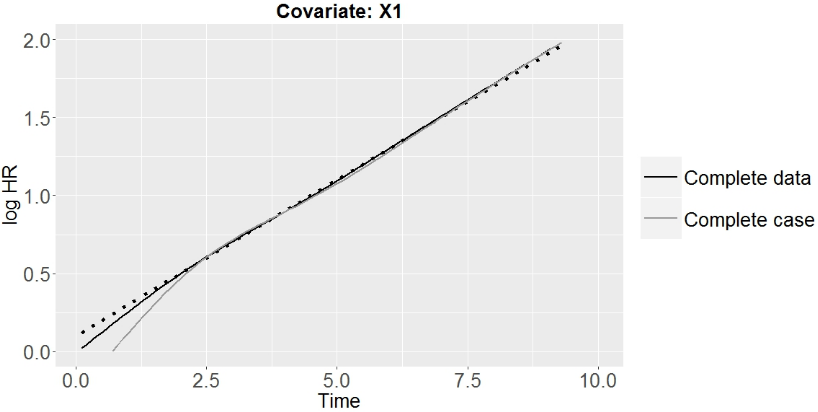
Simulation results: Mean estimated curve

Binary X



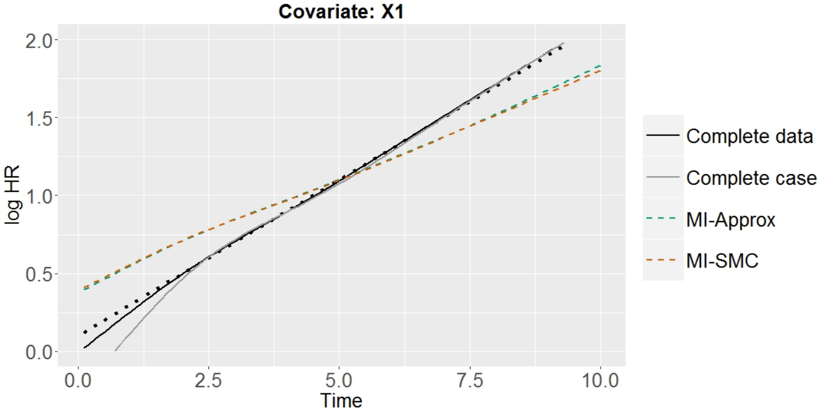
Simulation results: Mean estimated curve

Binary X



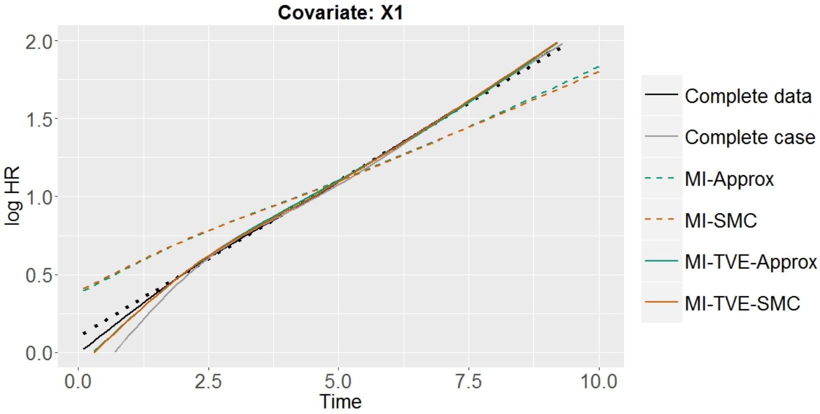
Simulation results: Mean estimated curve

Binary X



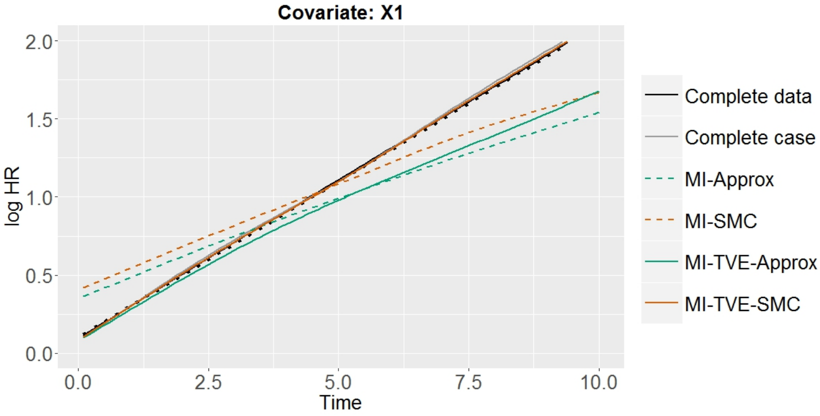
Simulation results: Mean estimated curve

Binary X



Simulation results: Mean estimated curve

Continuous X



Summary of simulation results

Ignoring TVEs in the imputation results in...

- incorrect tests for proportional hazards
- a big loss of power to detect TVEs
- biased estimates of the shape of the time-varying association

MI-TVE-Approx or MI-TVE-SMC?

- Both methods work well for binary exposures with missing data
- MI-TVE-SMC works better for continuous variables and has further advantages

Summary of simulation results

Ignoring TVEs in the imputation results in...

- incorrect tests for proportional hazards
- a big loss of power to detect TVEs
- biased estimates of the shape of the time-varying association

MI-TVE-Approx or MI-TVE-SMC?

- Both methods work well for binary exposures with missing data
- MI-TVE-SMC works better for continuous variables and has further advantages

Practical implementation

MI-Approx and MI-TVE-Approx

- `mice` in R
- `mi impute` in Stata

MI-SMC

- `smcfcs` in R and Stata

MI-TVE-SMC

- We have extended the `smcfcs` code in R to accommodate TVEs
- Available on github
- `https://github.com/ruthkeogh/MI-TVE`

Practical implementation

MI-Approx and MI-TVE-Approx

- `mice` in R
- `mi impute` in Stata

MI-SMC

- `smcfcs` in R and Stata

MI-TVE-SMC

- We have extended the `smcfcs` code in R to accommodate TVEs
- Available on github
- `https://github.com/ruthkeogh/MI-TVE`

Practical implementation

MI-Approx and MI-TVE-Approx

- `mice` in R
- `mi impute` in Stata

MI-SMC

- `smcfcs` in R and Stata

MI-TVE-SMC

- We have extended the `smcfcs` code in R to accommodate TVEs
- Available on github
- `https://github.com/ruthkeogh/MI-TVE`

Further work

- We also proposed a model selection algorithm...
- The model selection does not incorporate selection of functional forms for the covariates
- MI-TVE-SMC can be extended to accommodate this
- Drawing on the work of
  - Sauerbrei, Royston & Look (*Biometrical Journal* 2007)
  - Wood et al (*Stat. Med.* 2008)
  - Morris et al (*Stat. Med.* 2016)