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

Ruth H. Keogh1, Tim P. Morris2

1Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK
2London 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

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

- $T$: 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

\[
h(t|X_1, X_2) = h_0(t)e^{\beta X_1 X_1 + \beta X_2 X_2}
\]

- \( T \): 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**

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

- $T$: 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

$$h(t|X_1, X_2) = h_0(t)e^{\beta_1 X_1 + \beta_2 X_2}$$

- $T$: 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

\[ h(t|X_1, X_2) = h_0(t)e^{\beta X_1 X_1 + \beta X_2 X_2} \]

- \( T \): 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]
Cox proportional hazards model

\[ h(t|X_1, X_2) = h_0(t)e^{\beta_1 X_1 + \beta_2 X_2} \]

- \( T \): 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

\[ h(t|X_1, X_2) = h_0(t)e^{\beta_1 X_1 + \beta_2 X_2} \]

- \( T \): 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 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)
Imputing missing covariate values for the Cox model

Ian R. White\textsuperscript{1, *}, † and Patrick Royston\textsuperscript{2}

\textsuperscript{1}MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, U.K.
\textsuperscript{2}MRC Clinical Trials Unit, Cancer Group, London, U.K.
White & Royston’s method: MI-Approx

Cox proportional hazards model

\[
h(t|X_1, X_2) = h_0(t)e^{\beta_1 X_1 + \beta_2 X_2}
\]

Imputation model arises from an approximation to the distribution

\[
p(X_1|X_2, T, D)
\]

The imputation model: MI-Approx

\[
X_1 \sim X_2 + D + \hat{H}(T)
\]

e.g. linear or logistic regression

\(\hat{H}(T)\) is the Nelson-Aalen estimate of the cumulative hazard
White & Royston’s method: MI-Approx

**Cox proportional hazards model**

\[ h(t|X_1, X_2) = h_0(t)e^{\beta_1 X_1 + \beta_2 X_2} \]

Imputation model arises from an approximation to the distribution

\[ p(X_1|X_2, T, D) \]

**The imputation model: MI-Approx**

\[ X_1 \sim X_2 + D + \hat{H}(T) \]

e.g. linear or logistic regression

\( \hat{H}(T) \) is the Nelson-Aalen estimate of the cumulative hazard
Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model

Jonathan W Bartlett,¹ Shaun R Seaman,² Ian R White² and James R Carpenter¹,³ 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

\[ h(t|X_1, X_2) = h_0(t)e^{\beta X_1 X_1 + \beta X_2 X_2} \]

1. Obtain initial estimates for \( \beta X_1, \beta X_2 \)
2. Draw values \( \beta^{(m)} X_1, \beta^{(m)} X_2 \), 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

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

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


Cox regression with Time-Varying Effects (TVE)
Cox regression with Time-Varying Effects (TVE)

Cox proportional hazards model

\[ h(t|X_1, X_2) = h_0(t)e^{\beta X_1 X_1 + \beta X_2 X_2} \]

- 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

\[
h(t|X_1, X_2) = h_0(t)e^{\beta X_1 X_1 + \beta X_2 X_2}
\]

Extended Cox models with TVEs

\[
h(t|X_1, X_2) = h_0(t)e^{f_{X_1}(t;\beta X_1)X_1 + f_{X_2}(t;\beta X_2)X_2}
\]

Example

\[
h(t|X_1, X_2) = h_0(t)e^{\beta X_1 X_1 + \gamma X_1 X_1 \times t + \beta X_2 X_2 + \gamma X_2 X_2 \times t}
\]

A test of \( \gamma_{X_1} = 0 \) is a test of the proportional hazards assumption for \( X_1 \).
Cox regression with Time-Varying Effects (TVE)

Cox proportional hazards model

\[ h(t|X_1, X_2) = h_0(t)e^{\beta X_1 X_1 + \beta X_2 X_2} \]

Extended Cox models with TVEs

\[ h(t|X_1, X_2) = h_0(t)e^{f_{X_1}(t;\beta X_1) X_1 + f_{X_2}(t;\beta X_2) X_2} \]

Example

\[ h(t|X_1, X_2) = h_0(t)e^{\beta X_1 X_1 + \gamma X_1 X_1 \times t + \beta X_2 X_2 + \gamma X_2 X_2 \times t} \]

A test of \( \gamma_{X_1} = 0 \) is a test of the proportional hazards assumption for \( X_1 \).
Cox regression with Time-Varying Effects (TVE)

Extended Cox models with TVEs

\[ h(t|X_1, X_2) = h_0(t)e^{f_{X_1}(t;\beta_{X_1})X_1 + f_{X_2}(t;\beta_{X_2})X_2} \]

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

\[ h(t|X_1, X_2) = h_0(t)e^{f_{X_1}(t, \beta_{X_1})X_1 + f_{X_2}(t, \beta_{X_2})X_2} \]

The imputation model: MI-TVE-Approx

\[ X_1 \sim X_2 + f_{X_1}(T)D + \hat{H}(T) \]

e.g. linear or logistic regression
MI-SMC extended for TVEs: MI-TVE-SMC

Extended Cox model with TVEs

\[ h(t|X_1, X_2) = h_0(t)e^{f_{X_1}(t, \beta_{X_1})X_1} + f_{X_2}(t, \beta_{X_2})X_2 \]

1. Obtain initial estimates for \( \beta_{X_1}, \beta_{X_2} \)
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
   \[
   U \leq \exp\{-H_0^{(m)}(t)e^{f_{X_1}(t, \beta_{X_1}^{(m)})X_1^*} + f_{X_2}(t, \beta_{X_2}^{(m)})X_2\} \quad \text{if } D = 0
   \]
   \[
   U \leq h_0^{(m)}(t)e^{f_{X_1}(t, \beta_{X_1}^{(m)})X_1^*} + f_{X_2}(t, \beta_{X_2}^{(m)})X_2 - \int_0^t h_0^{(m)}(u)e^{f_{X_1}(u, \beta_{X_1}^{(m)})X_1^*} + f_{X_2}(u, \beta_{X_2}^{(m)})X_2 du \} \quad \text{if } D = 1
   \]
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?
Extended Cox model with TVEs

\[ h(t|X_1, X_2) = h_0(t)e^{f_{X_1}(t,\beta_{X_1})X_1 + f_{X_2}(t,\beta_{X_2})X_2} \]

- Simple pre-specified forms (e.g. Quantin 1999), e.g.
  \[ f_X(t) = \beta_{X_0} + \beta_{X_1}t \]
- 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

\[ h(t|X_1, X_2) = h_0(t)e^{f_{X_1}(t, \beta_{X_1})X_1 + f_{X_2}(t, \beta_{X_2})X_2} \]

- Simple pre-specified forms (e.g. Quantin 1999), e.g.
  \[ f_X(t) = \beta_{X_0} + \beta_{X_1}t \]
- 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

\[ h(t | X_1, X_2) = h_0(t) e^{f_{X_1}(t; \beta_{X_1})X_1 + f_{X_2}(t; \beta_{X_2})X_2} \]

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

\[
f_X(t; \beta_X) = \beta_{X_0} + \beta_{X_1} t + \sum_{i=1}^{L-2} \theta X_i \left\{ (t - u_i)_+^3 - \left( \frac{(t - u_{L-1})^3 (u_L - u_i)}{(u_L - u_{L-1})} \right) + \left( \frac{(t - u_L)^3 (u_{L-1} - u_i)}{(u_L - u_{L-1})} \right) \right\}
\]

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

\[ h(t | X_1, X_2) = h_0(t) e^{\sum_{i=1}^{3} \theta X_i \left( \left( t - u_{i+4} \right)_{+} \left( u_5 - u_{i+4} \right) \right) + \left( \frac{(t-u_5)^3 (u_4 - u_i)}{(u_5-u_4)} \right)} \]

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

\[ \beta_{X_1} = \theta_{X_1} = \theta_{X_2} = \theta_{X_3} = 0 \]
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


Simulation study
Simulation study

Extended Cox model with TVEs

\[ h(t|X_1, X_2) = h_0(t)e^{f_{X_1}(t, \beta X_1)X_1 + \beta X_2 X_2} \]

Scenario 1
Simulation study

Extended Cox model with TVEs

\[ h(t|X_1, X_2) = h_0(t) e^{f_{X_1}(t, \beta_{X_1})X_1 + \beta_{X_2}X_2} \]
Simulation study

Extended Cox model with TVEs

\[ h(t | X_1, X_2) = h_0(t) e^{f_{X_1}(t, \beta X_1) X_1 + \beta X_2 X_2} \]
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
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

![Graph showing log-hazard ratio against follow-up time for five scenarios]

Scenario 1
Scenario 2
Scenario 3
Scenario 4
Scenario 5
Test for proportional hazards: Scenario 1

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

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete data</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Complete case</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>MI-Approx</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MI-SMC</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MI-TVE-Approx</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>MI-TVE-SMC</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Test for proportional hazards: Scenario 2

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

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete data</td>
<td>89</td>
<td>3</td>
</tr>
<tr>
<td>Complete case</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td>MI-Approx</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>MI-SMC</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>MI-TVE-Approx</td>
<td>67</td>
<td>3</td>
</tr>
<tr>
<td>MI-TVE-SMC</td>
<td>68</td>
<td>6</td>
</tr>
</tbody>
</table>

Follow-up time $t$ vs. log-hazard ratio

Scenario 2
Test for proportional hazards: Scenario 5

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

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete data</td>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td>Complete case</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>MI-Approx</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>MI-SMC</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>MI-TVE-Approx</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>MI-TVE-SMC</td>
<td>27</td>
<td>5</td>
</tr>
</tbody>
</table>

Graph showing log-hazard ratio against follow-up time for Scenario 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**

![Graph showing the mean estimated curve for Covariate X1 with different lines representing Complete data, Complete case, MI-Approx, and MI-SMC.](image)
Simulation results: Mean estimated curve

Binary X
Simulation results: Mean estimated curve

Covariate: X1

- Complete data
- Complete case
- MI-Approx
- MI-SMC
- MI-TVE-Approx
- MI-TVE-SMC
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
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
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/ruthkeoghs/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/ruthkeoghi/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)