The Stata Journal

Editors
H. JOSEPH NEWTON
Department of Statistics
Texas A&M University
College Station, Texas
editors@stata-journal.com

NICHOLAS J. COX
Department of Geography
Durham University
Durham, UK
editors@stata-journal.com

Associate Editors
CHRISTOPHER F. BAUM, Boston College
NATHANIEL BECK, New York University
RINO BELLOCCO, Karolinska Institutet, Sweden, and University of Milano-Bicocca, Italy
MAARTEN L. BUSS, WZB, Germany
A. COLIN CAMERON, University of California–Davis
MARIO A. CLEVES, University of Arkansas for Medical Sciences
WILLIAM D. DUPONT, Vanderbilt University
PHILIP ENDER, University of California–Los Angeles
DAVID EPNSTEIN, Columbia University
ALLAN GREGORY, Queen’s University
JAMES HARDIN, University of South Carolina
BEN JANN, University of Bern, Switzerland
STEPHEN JENKINS, London School of Economics and Political Science
ULRICH KOHLER, University of Potsdam, Germany

FRAUKE KREUTER, Univ. of Maryland–College Park
PETER A. LACHENBRUCH, Oregon State University
JENS LAURITSEN, Odense University Hospital
STANLEY LEMESHOW, Ohio State University
J. SCOTT LONG, Indiana University
ROGER NEWSOM, Imperial College, London
AUSTIN NICHOLS, Urban Institute, Washington DC
MARCELLO PAGANO, Harvard School of Public Health
SOPHIA RABE-HESKETH, Univ. of California–Berkeley
J. PATRICK ROYSTON, MRC Clinical Trials Unit, London
PHILIP RYAN, University of Adelaide
MARK E. SCHAFFER, Heriot-Watt Univ., Edinburgh
JEROEN WESSEY, Utrecht University
IAN WHITE, MRC Biostatistics Unit, Cambridge
NICHOLAS J. G. WINTER, University of Virginia
JEFFREY WOODBRIDGE, Michigan State University

Stata Press Editorial Manager
LISA GILMORE

Stata Press Copy Editors
DAVID CULWELL and DEIRDRE SKAGGS

The Stata Journal publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The Stata Journal is indexed and abstracted by CompuMath Citation Index, Current Contents/Social and Behavioral Sciences, RePEc: Research Papers in Economics, Science Citation Index Expanded (also known as SciSearch, Scopus, and Social Sciences Citation Index).

For more information on the Stata Journal, including information for authors, see the webpage

http://www.stata-journal.com
**Subscriptions** are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at


**Subscription rates** listed below include both a printed and an electronic copy unless otherwise mentioned.

<table>
<thead>
<tr>
<th></th>
<th>U.S. and Canada</th>
<th>Elsewhere</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Printed &amp; electronic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-year subscription</td>
<td>$98</td>
<td>$138</td>
</tr>
<tr>
<td>2-year subscription</td>
<td>$165</td>
<td>$245</td>
</tr>
<tr>
<td>3-year subscription</td>
<td>$225</td>
<td>$345</td>
</tr>
<tr>
<td>1-year student subscription</td>
<td>$75</td>
<td>1-year student subscription</td>
</tr>
<tr>
<td>1-year university library subscription</td>
<td>$125</td>
<td>1-year university library subscription</td>
</tr>
<tr>
<td>2-year university library subscription</td>
<td>$215</td>
<td>2-year university library subscription</td>
</tr>
<tr>
<td>3-year university library subscription</td>
<td>$315</td>
<td>3-year university library subscription</td>
</tr>
<tr>
<td>1-year institutional subscription</td>
<td>$245</td>
<td>1-year institutional subscription</td>
</tr>
<tr>
<td>2-year institutional subscription</td>
<td>$445</td>
<td>2-year institutional subscription</td>
</tr>
<tr>
<td>3-year institutional subscription</td>
<td>$645</td>
<td>3-year institutional subscription</td>
</tr>
<tr>
<td><strong>Electronic only</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-year subscription</td>
<td>$75</td>
<td>$75</td>
</tr>
<tr>
<td>2-year subscription</td>
<td>$125</td>
<td>$125</td>
</tr>
<tr>
<td>3-year subscription</td>
<td>$165</td>
<td>$165</td>
</tr>
<tr>
<td>1-year student subscription</td>
<td>$45</td>
<td>1-year student subscription</td>
</tr>
</tbody>
</table>

Back issues of the *Stata Journal* may be ordered online at


Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.


The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

Copyright © 2013 by StataCorp LP

**Copyright Statement**: The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible websites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, *stata*, Stata Press, Mata, *mata*, and NetCourse are registered trademarks of StataCorp LP.
Extending the flexible parametric survival model for competing risks

Sally R. Hinchliffe  Paul C. Lambert
Department of Health Sciences  Department of Health Sciences
University of Leicester  University of Leicester
Leicester, UK  Leicester, UK
srh20@leicester.ac.uk  paul.lambert@le.ac.uk

Abstract. Competing risks are present when the patients within a dataset could experience one or more of several exclusive events and the occurrence of any one of these could impede the event of interest. One of the measures of interest for analyses of this type is the cumulative incidence function. stpm2cif is a postestimation command used to generate predictions of the cumulative incidence function after fitting a flexible parametric survival model using stpm2. There is also the option to generate confidence intervals, cause-specific hazards, and two other measures that will be discussed in further detail. The new command is illustrated through a simple example.

Keywords: st0298, stpm2cif, survival analysis, competing risks, cumulative incidence, cause-specific hazard

1 Introduction

In survival analysis, if interest lies in the true probability of death from a particular cause, then it is important to appropriately account for competing risks. Competing risks occur when patients are at risk of more than one mutually exclusive event, such as death from different causes (Putter, Fiocco, and Geskus 2007). The occurrence of a competing event may prevent the event of interest from ever occurring. It therefore seems logical to conduct an analysis that considers these competing risks. The two main measures of interest for analyses of this type are the cause-specific hazard and the cumulative incidence function. The cause-specific hazard is the instantaneous risk of dying from a specific cause given that the patient is still alive at a particular time. The cumulative incidence function is the proportion of patients who have experienced a particular event at a certain time in the follow-up period. Several methods are already available to estimate this; however, it is not always clear which approach should be used.

In this article, we explain how to fit flexible parametric models using the stpm2 command by estimating the cause-specific hazard for each cause of interest in a competing-risks situation. The stpm2cif command is a postestimation command used to estimate the cumulative incidence function for up to 10 competing causes along with confidence intervals, cause-specific hazards, and two other useful measures.
2 Methods

If a patient is at risk from \( K \) different causes, the cause-specific hazard, \( h_k(t) \), is the risk of failure at time \( t \) given that no failure from cause \( k \) or any of the \( K-1 \) other causes has occurred. In a proportional hazards model, \( h_k(t) \) is

\[
h_k(t | Z) = h_{k,0}(t) \exp \left( \beta_k^T Z \right)
\]  

(1)

where \( h_{k,0}(t) \) is the baseline cause-specific hazard for cause \( k \), and \( \beta_k \) is the vector of parameters for covariates \( Z \). The cumulative incidence function, \( C_k(t) \), can be derived from the cause-specific hazards through the equation

\[
C_k(t) = \int_0^t h_k(u | Z) \prod_{k=1}^K S_k(u) du
\]  

(2)

where \( \prod_{k=1}^K S_k(u) du = \exp \left( - \int_0^t \sum_{k=1}^K h_k \right) \) is the overall survival function (Prentice et al. 1978).

Several programs are currently available in Stata that can compute the cumulative incidence function. The command stcompet calculates the function by using the Kaplan–Meier estimator of the overall survival function (Coviello and Boggess 2004). It therefore does not allow for the incorporation of covariate effects. A follow-on to stcompet is stcompadj, which fits the cumulative incidence function based on the Cox model or the flexible parametric regression model (Coviello 2009). However, it only allows one competing event, and because the regression models are built into the command internally, it does not allow users to specify their own options with stcox or stpm2. Finally, Fine and Gray’s (1999) proportional subhazards model can be fit using stcrreg.

The flexible parametric model was first proposed by Royston and Parmar in 2002. The approach uses restricted cubic spline functions to model the baseline log cumulative hazard. It has the advantage over other well-known models such as the Cox model because it produces smooth predictions and can be extended to incorporate complex time-dependent effects, again through the use of restricted cubic splines. The Stata implementation of the model using stpm2 is described in detail elsewhere (Royston and Parmar 2002; Lambert and Royston 2009). Both the cause-specific hazard (1) and the overall survival function can be obtained from the flexible parametric model to give the integrand in (2). This can be done by fitting separate models for each of the \( k \) causes, but this will not allow for shared parameters. It is possible to fit one model for all \( k \) causes simultaneously by stacking the data so that each individual patient has \( k \) rows of data, one for each of the \( k \) causes. Table 1 illustrates how the data should look once they have been stacked (in the table, CVD stands for cardiovascular disease). Each patient can fail from one of three causes. Patient 1 is at risk from all three causes for 10 years but does not experience any of them and so is censored. Patient 2 is at risk from all three causes for eight years but then experiences a cardiovascular event. By expanding
the dataset, one can allow for covariate effects to be shared across the causes, although it is possible to include covariates that vary for each cause.

<table>
<thead>
<tr>
<th>ID</th>
<th>Age</th>
<th>Time</th>
<th>Cause</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>10</td>
<td>Cancer</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>10</td>
<td>CVD</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>10</td>
<td>Other</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>70</td>
<td>8</td>
<td>Cancer</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>70</td>
<td>8</td>
<td>CVD</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>70</td>
<td>8</td>
<td>Other</td>
<td>0</td>
</tr>
</tbody>
</table>

3 Syntax

```
stsmt2cif newvarlist, cause1(varname # [varname # ...]) cause2(varname #
[varname # ...]) cause3(varname # [varname # ...]) ...
cause10(varname # [varname # ...]) obs(#) ci mint(#) maxt(#)
timetype(newvar) hazard contmort conthaz
```

The names specified in `newvarlist` coincide with the order of the causes inputted in the options.

3.1 Options

- `cause1(varname # [varname # ...])` ... `cause10(varname # [varname # ...])` request that the covariates specified by the listed `varname` be set to # when predicting the cumulative incidence functions for each cause. `cause1()` and `cause2()` are required.
- `obs(#)` specifies the number of observations (of time) to predict. The default is `obs(1000)`. Observations are evenly spread between the minimum and maximum values of follow-up time.
- `ci` calculates a 95% confidence interval for the cumulative incidence function and stores the confidence limits in `CIF_newvar_lci` and `CIF_newvar_uci`.
- `mint(#)` specifies the minimum value of follow-up time. The default is set as the minimum event time from `stset`.
- `maxt(#)` specifies the maximum value of follow-up time. The default is set as the maximum event time from `stset`.
S. R. Hinchliffe and P. C. Lambert

`timename(newvar)` specifies the time variable generated during predictions for the cumulative incidence function. The default is `timename(newt)`. This is the variable for time that needs to be used when plotting curves for the cumulative incidence function and the cause-specific hazard function.

`hazard` predicts the cause-specific hazard function for each cause.

`contmort` predicts the relative contribution to total mortality.

`conthaz` predicts the relative contribution to hazard.

4 Example

Data were used on 506 patients with prostate cancer who were randomly allocated to treatment with diethylstilbestrol. The data have been used previously to illustrate the command `stcompet` (Coviello and Boggess 2004). Patients are classified as alive or having died from one of three causes: cancer (the event of interest), cardiovascular disease (CVD), or other causes. To use `stpm2cif`, the user must first expand the dataset:

```plaintext
. use prostatecancer
. expand 3
   (1012 observations created)
. by id, sort: generate cause= _n
. generate cancer = cause==1
. generate cvd = cause==2
. generate other = cause==3
. generate treatcancer = treatment*cancer
. generate treatcvd = treatment*cvd
. generate treatother = treatment*other
. generate event = (cause==status)
```

The data have been expanded so that each patient has three rows of data, one for each cause as shown in table 1. Three indicator variables have been created for each of the three competing causes and interactions between `treatment`. Three causes have also been generated. The indicator variable `event` defines whether a patient has died and the cause of death. We now need to `stset` the data and run `stpm2`.
Competing risks

```
. stset time, failure(event)
  failure event: event != 0 & event < .
  obs. time interval: (0, time]
  exit on or before: failure

  1518 total obs.
  0 exclusions

  1518 obs. remaining, representing
  356 failures in single record/single failure data
  54898.8 total analysis time at risk, at risk from t = 0
  earliest observed entry t = 0
  last observed exit t = 76

. stpm2 cancer cvd other treatcancer treatcvd treatother, scale(hazard)
   > rcsbaseoff dftvc(3) nocons tvc(cancer cvd other) eform no log

Log likelihood = -1150.4866  Number of obs = 1518

| exp(b) | Std. Err. | z     | P>|z|  |  [95% Conf. Interval] |
|--------|-----------|------|------|-------------------|
| xb     |           |      |      |                   |
| cancer | 0.2363179 | 0.0275697 | -12.37 | 0.000 | 0.188015 | 0.2970303 |
| cvd    | 0.1801668 | 0.0238668 | -12.94 | 0.000 | 0.1389876 | 0.2335983 |
| other  | 0.1008464 | 0.0238668 | -12.94 | 0.000 | 0.070895 | 0.1335983 |
| treatcancer | 0.6722196 | 0.1096964 | -12.76 | 0.000 | 0.4882109 | 0.9285819 |
| treatcvd | 1.188189  | 0.0139015 | -10.70 | 0.000 | 1.053815 | 1.324565 |
| treatother | 0.6345498 | 0.0238668 | -12.94 | 0.000 | 0.4382109 | 0.9285819 |
| _rcs_cancer1 | 3.501847 | 0.44435 | 7.88 | 0.000 | 2.637088 | 4.540625 |
| _rcs_cancer2 | 0.8842712 | 0.0742915 | -11.80 | 0.000 | 0.7600191 | 1.048584 |
| _rcs_cancer3 | 1.046436 | 0.0371625 | -11.80 | 0.000 | 0.9760756 | 1.121868 |
| _rcs_cvd1 | 2.841936 | 0.2619063 | 11.13 | 0.000 | 2.372299 | 3.405465 |
| _rcs_cvd2 | 0.8772848 | 0.0498866 | -17.70 | 0.000 | 0.7847607 | 0.980176 |
| _rcs_cvd3 | 1.008804 | 0.052009 | 11.13 | 0.000 | 0.942175 | 1.075465 |
| _rcs_other1 | 2.751505 | 0.3563037 | 7.72 | 0.000 | 2.134738 | 3.458467 |
| _rcs_other2 | 0.7962094 | 0.0558599 | -13.94 | 0.000 | 0.6893208 | 0.913576 |
| _rcs_other3 | 0.9614597 | 0.0512891 | -14.94 | 0.000 | 0.8660117 | 1.067428 |
```

By including the three cause indicators (`cancer`, `cvd`, and `other`) as both main effects and time-dependent effects (using the `tvc()` option), we have fit a stratified model with three separate baselines, one for each cause. For this reason, we have used the `rcsbaseoff` option together with the `nocons` option, which excludes the baseline hazard from the model. The interactions between treatment and the three causes have also been included in the model. This estimates a different treatment effect for each of the three causes. The hazard ratios (95% confidence intervals) for the treatment effect are 0.67 [0.49, 0.93], 1.19 [0.85, 1.66], and 0.63 [0.38, 1.06] for cancer, CVD, and other causes, respectively.

Now that we have run `stpm2`, we can run the new postestimation command `stpm2cif` to obtain the cumulative incidence function for each cause. Because we have two groups of patients, treated and untreated, we must run the command twice. This will give separate cumulative incidence functions for the treated and the untreated groups and for each of the three causes.
The `cause1()` to `cause3()` options give the linear predictor for each of the three causes for which we want a prediction. The commands have generated six new variables containing the cumulative incidence functions. The untreated group members are denoted with a 0 at the end of the variable name, and the treated group members are denoted with a 1. These labels come from the input into `newvarlist` in the above command line. The six cumulative incidence functions are therefore labeled `CIF_cancer0`, `CIF_cvd0`, `CIF_other0`, `CIF_cancer1`, `CIF_cvd1`, and `CIF_other1`. Each of these variables has a corresponding high and low confidence bound, for example, `CIF_cancer0_lci` and `CIF_cancer1_uci`. These were created because the `ci` option was specified. The `maxt()` option has been specified to restrict the predictions for the cumulative incidence function to a maximum follow-up time of 60 months; this was done for illustrative purposes only.

By specifying the `hazard` option, we have generated cause-specific hazards that correspond with each of the cumulative incidence functions. These are labeled as `h_cancer0`, `h_cvd0`, `h_other0`, `h_cancer1`, `h_cvd1`, and `h_other1`. The options `contmort` and `conthaz` are the two additional measures mentioned previously. The `contmort` option produces what we have named the “relative contribution to the total mortality”. This is essentially the cumulative incidence function for each specific cause divided by the sum of all the cumulative incidence functions. It can be interpreted as the probability that you will die from a particular cause given that you have died by time $t$. The `conthaz` option produces what we have named the “relative contribution to the overall hazard”. This is similar to the last measure in that it is the cause-specific hazard for a particular cause divided by the sum of all the cause-specific hazards. It can be interpreted as the probability that you will die from a particular cause given that you die at time $t$. 
If we plot the cumulative incidence functions for each cause against time, we can achieve plots as shown in figure 1.

![Cumulative incidence of cancer, CVD, and other causes of death in treated and untreated patients with prostate cancer](image)

Figure 1. Cumulative incidence of cancer, CVD, and other causes of death in treated and untreated patients with prostate cancer.

The plots in figure 1 give the actual probabilities of dying from each cause, taking into account the competing causes. The treated group have a lower probability of dying from cancer or other causes compared with the untreated group, but have a higher probability of dying from CVD.

The model fit above is relatively simple because it only considers treatment as a predictor for the three causes of death. Age is an important factor when fitting the probability of death, so we shall now consider a model including age as a continuous variable with a time-dependent effect. Although the effect of age will most likely differ between the three causes of death, for demonstrative purposes, we will assume that the effect of age can be shared across all three causes. This is one of the main advantages of stacking the data as shown previously. The *stpm2* command can be rerun to include age in both the variable list and the `tvc()` option. The three cause indicators (*cancer*, *cvd*, and *other*) remain as time-dependent effects with 3 degrees of freedom to maintain the stratified model with three separate baselines. Age is now included as a time-dependent effect with only 1 degree of freedom.
As before, we can now use the \texttt{stpm2cif} command to obtain the cumulative incidence functions for cancer, CVD, and other causes. This time, we want to predict for ages 65 and 75 in both of the two treatment groups, so we will need to run the command four times.

\begin{verbatim}
   . stpm2cif age65cancer0 age65cvd0 age65other0, cause1(cancer 1 age 65) cause2(cvd 1 age 65) cause3(other 1 age 65) ci hazard contmortal cont haz maxt(60)
   . stpm2cif age65cancer1 age65cvd1 age65other1, cause1(cancer 1 treatcancer 1 age 65) cause2(cvd 1 treatcvd 1 age 65) cause3(other 1 treatother 1 age 65) ci hazard contmortal cont haz maxt(60)
   . stpm2cif age75cancer0 age75cvd0 age75other0, cause1(cancer 1 age 75) cause2(cvd 1 age 75) cause3(other 1 age 75) ci hazard contmortal cont haz maxt(60)
   . stpm2cif age75cancer1 age75cvd1 age75other1, cause1(cancer 1 treatcancer 1 age 75) cause2(cvd 1 treatcvd 1 age 75) cause3(other 1 treatother 1 age 75) ci hazard contmortal cont haz maxt(60)
\end{verbatim}

The \texttt{stpm2cif} commands have generated 12 new variables for the cumulative incidence functions, labeled \texttt{CIF.age65cancer0}, \texttt{CIF.age65cvd0}, \texttt{CIF.age65other0}, \texttt{CIF.age65cancer1}, \texttt{CIF.age65cvd1}, \texttt{CIF.age65other1}, \texttt{CIF.age75cancer0}, \texttt{CIF.age75cvd0}, \texttt{CIF.age75other0}, \texttt{CIF.age75cancer1}, \texttt{CIF.age75cvd1}, and \texttt{CIF.age75other1}. A 65 next to \texttt{age} represents the prediction for those 65 years old; a 75 represents a prediction for those 75 years old.
Rather than plotting the cumulative incidence function as a line for each cause separately as we did previously, we display them by stacking them on top of each other. This produces a graph as shown in figure 2. To do this, we need to generate new variables that sum up the cumulative incidence functions. This is done for each of the two treatment groups and two ages. The code shown below is for the 65-year-olds in the treatment group only.

```stata
. generate age65treat1 = CIF_age65cancer1
   (518 missing values generated)
. generate age65treat2 = age65treat1+CIF_age65cvd1
   (518 missing values generated)
. generate age65treat3 = age65treat2+CIF_age65other1
   (518 missing values generated)
. twoway (area age65treat3 _newt, sort fintensity(100))
  > (area age65treat2 _newt, sort fintensity(100))
  > (area age65treat1 _newt, sort fintensity(100)), ylabel(0(0.2)1, angle(0)
  > format(%3.1f)) ytitle("") xtitle(""")
  > legend(order(3 "Cancer" 2 "CVD" 1 "Other") rows(1) size(small))
  > title("Treated") plotregion(margin(zero)) scheme(sj)
  > saving(treatedage65, replace)
```

Figure 2. Stacked cumulative incidence of cancer, CVD, and other causes of death for those aged 65 and 75 in treated and untreated patients with prostate cancer
The results in figure 2 allow us to visualize the total probability of dying in both the treated and the untreated groups for those aged 65 and 75 and allow us to see how this is broken down by the specific causes. As expected, the total probability of death is higher for the oldest age in both treatment groups. The distribution of deaths across the three causes in each treatment group is roughly the same for both ages. Again we see that although the treatment reduces the total probability of death, it actually increases the probability of death from CVD.

Using a similar process to the one used above to obtain the stacked cumulative incidence plots, we can also produce stacked plots of the relative contribution to the total mortality and the relative contribution to the hazard. These graphs are shown in figures 3 and 4.

Figure 3. Relative contribution to the total mortality for those aged 65 and 75 in treated and untreated patients with prostate cancer

Figure 2 shows the relative contribution to the total mortality for those aged 65 and 75 in the two treatment groups. If we focus on the 65-year-olds in the treated group, the plot shows us that given a patient 65 years old is going to die by 40 months if treated, then the probability of dying from cancer is 0.39, the probability of dying from CVD is 0.48, and the probability of dying from other causes is 0.13. However, if the patient is untreated, then the same probabilities are 0.49, 0.34, and 0.17, respectively.
Figure 4 shows the relative contribution to the overall hazard for those aged 65 and 75 in the two treatment groups. Again, if we focus on the 65-year-olds in the treated group, the plot shows us that given a patient 65 years old is going to die at 40 months if treated, then the probability of dying from cancer is 0.39, the probability of dying from CVD is 0.45, and the probability of dying from other causes is 0.16. However, if the patient is untreated, then the same probabilities are 0.48, 0.32, and 0.20, respectively.

5 Conclusion

The new command \texttt{stpm2cif} provides an extension to the command \texttt{stpm2} to enable users to estimate the cumulative incidence function through the flexible parametric function. We hope that it will be a useful tool in medical research.

6 References


**About the authors**

Sally Hinchliffe is a PhD student at the University of Leicester, UK. She is currently working on developing a methodology for application in competing risks.

Paul Lambert is a reader in medical statistics at the University of Leicester, UK. His main interest is in the development and application of methods in population-based cancer research.