Survival Analysis with Time-Dependent Covariates: A Practical Example

October 28, 2016
SAS Health Users Group
Maria Eberg
Outline

• Why use time-dependent covariates?
• Things to consider in definition of time-dependent covariates
• Counting process type of data input
• Example of definition and analysis of data with multiple time-dependent covariates
When should one use time-dependent covariates?

Solution for “immortal time-bias”.

• Immortal time refers to a period of follow-up during which, by design, death or the study outcome cannot occur.

Important covariates may change over time in studies with long-term follow-up.
Before you start coding...

1. What functional form of the time-dependent covariate makes sense:
   - Cumulative, current, ever-use
   - Time-lag (depends on the biological effect)

2. Exogeneity

   Covariates are external or exogenous if they are determined by factors outside the system or the individual under study.

   Air pollution – exogenous (external) covariate
   Blood pressure, disease complications – endogenous (internal) covariates
Counting process type of input

Multiple records are created for each subject, one record for each distinct pattern of the time-dependent measurements.

Caution:
- No nested or overlapping intervals;
- Intervals of zero length do not contribute to the analyses.
Practical example – study design

Index event – prostate cancer diagnosis

Main exposure – statin use
• 1-year lag to avoid reverse causality
• Multiple definitions:
  – Ever-never,
  – cumulative days of use,
  – cumulative dose
• Cumulative days of use or cumulative dose taken – a summary of exposure history

Outcomes – death due to prostate cancer, all-cause mortality

Covariates:
• Fixed at baseline:
  - Demographic and lifestyle (age, sex, smoking status, BMI, alcohol abuse)
  - Comorbidities (MI, stroke/TIA, peripheral artery disease, chronic kidney disease, previous cancer)
  - Indicators of disease severity (Gleason score, prostate cancer treatments)
  - Medication use (including pre-diagnostic use of statins)
• Time-dependent:
  Prostate specific antigen (PSA) testing count
Creating long data

data byday;
    set original;
    do istart = t0 to (end_fu - 1);
        iend = istart + 1;
        output;
    end;
run;

<table>
<thead>
<tr>
<th>ID</th>
<th>DIAG_DATE</th>
<th>T0</th>
<th>END_FU</th>
<th>EVENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01JAN2001</td>
<td>01JAN2002</td>
<td>31MAR2010</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>15APR2004</td>
<td>15APR2005</td>
<td>09OCT2007</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>27SEP2003</td>
<td>27SEP2004</td>
<td>10JUL2009</td>
<td>2</td>
</tr>
</tbody>
</table>
## Creating long data (2)

<table>
<thead>
<tr>
<th>ID</th>
<th>DIAG_DATE</th>
<th>T0</th>
<th>END_FU</th>
<th>EVENT</th>
<th>ISTART</th>
<th>IEND</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01JAN2001</td>
<td>01JAN2002</td>
<td>31MAR2010</td>
<td>0</td>
<td>01JAN2002</td>
<td>02JAN2002</td>
</tr>
<tr>
<td>1</td>
<td>01JAN2001</td>
<td>01JAN2002</td>
<td>31MAR2010</td>
<td>0</td>
<td>02JAN2002</td>
<td>03JAN2002</td>
</tr>
<tr>
<td>1</td>
<td>01JAN2001</td>
<td>01JAN2002</td>
<td>31MAR2010</td>
<td>0</td>
<td>03JAN2002</td>
<td>04JAN2002</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>01JAN2001</td>
<td>01JAN2002</td>
<td>31MAR2010</td>
<td>0</td>
<td>30MAR2010</td>
<td>31MAR2010</td>
</tr>
<tr>
<td>2</td>
<td>15APR2004</td>
<td>15APR2005</td>
<td>09OCT2007</td>
<td>1</td>
<td>16APR2005</td>
<td>17APR2005</td>
</tr>
<tr>
<td>2</td>
<td>15APR2004</td>
<td>15APR2005</td>
<td>09OCT2007</td>
<td>1</td>
<td>17APR2005</td>
<td>18APR2005</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Getting covariate information

<table>
<thead>
<tr>
<th>ID</th>
<th>RX_DATE</th>
<th>DURATION</th>
<th>DOSE</th>
<th>END_FU</th>
<th>Exposed_1st</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01JUL2003</td>
<td>28</td>
<td>1</td>
<td>31MAR2010</td>
<td>01JUL2004</td>
</tr>
<tr>
<td>1</td>
<td>15SEP2004</td>
<td>14</td>
<td>0.5</td>
<td>31MAR2010</td>
<td>01JUL2004</td>
</tr>
<tr>
<td>1</td>
<td>22SEP2004</td>
<td>28</td>
<td>1</td>
<td>31MAR2010</td>
<td>01JUL2004</td>
</tr>
</tbody>
</table>

data statin_byday;
  set statin_rx;
  do istart = rx_date to (rx_date + duration);
    iend = istart + 1;
    statin_i = 1;
    output;
  end;
run;

proc sort data = statin_byday nodupkey;
  by id istart iend;
run;
Defining cumulative variables

data TD_combined;
    merge byday (in=a) statin_byday(where = (istart >= exposed_1st));
    by id istart iend;
    if a;
        retain statin_01 cumdur cumdose;
        if first.id then do;
            statin_01 = 0;
            cumdur = 0;
            cumdose = 0;
        end;
        if istart = exposed_1st then do;
            statin_01 = 1;
            cumdur = basedur;
            cumdose = basedose;
        end;
        if statin_i = 1 then do;
            cumdur = cumdur + 1;
            cumdose = cumdose + dose;
        end;
    end;
run;
### Dataset with cumulative variables

<table>
<thead>
<tr>
<th>ID</th>
<th>T0</th>
<th>END_FU</th>
<th>EVENT</th>
<th>ISTART</th>
<th>IEND</th>
<th>STATIN_I</th>
<th>STATIN_01</th>
<th>CUMDUR</th>
<th>CUMDOSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td>01JAN2002</td>
<td>31MAR2010</td>
<td>0</td>
<td>01JAN2002</td>
<td>02JAN2002</td>
<td>0</td>
<td>No Use</td>
<td>No Use</td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>01JAN2002</td>
<td>31MAR2010</td>
<td>0</td>
<td>02JAN2002</td>
<td>03JAN2002</td>
<td>0</td>
<td>No Use</td>
<td>No Use</td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>....</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>01JAN2002</td>
<td>31MAR2010</td>
<td>0</td>
<td>01JUL2004</td>
<td>02JUL2004</td>
<td>0</td>
<td>1</td>
<td>&lt; 1 year</td>
<td>&lt; 365 DDD</td>
</tr>
<tr>
<td>1001</td>
<td>01JAN2002</td>
<td>31MAR2010</td>
<td>0</td>
<td>02JUL2004</td>
<td>03JUL2004</td>
<td>0</td>
<td>1</td>
<td>&lt; 1 year</td>
<td>&lt; 365 DDD</td>
</tr>
<tr>
<td>1001</td>
<td>....</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>01JAN2002</td>
<td>31MAR2010</td>
<td>0</td>
<td>15SEP2004</td>
<td>16SEP2004</td>
<td>1</td>
<td>1</td>
<td>&lt; 1 year</td>
<td>&lt; 365 DDD</td>
</tr>
<tr>
<td>1001</td>
<td>01JAN2002</td>
<td>31MAR2010</td>
<td>0</td>
<td>16SEP2004</td>
<td>17SEP2004</td>
<td>1</td>
<td>1</td>
<td>&lt; 1 year</td>
<td>&lt; 365 DDD</td>
</tr>
<tr>
<td>1001</td>
<td>....</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>01JAN2002</td>
<td>31MAR2010</td>
<td>0</td>
<td>30MAR2010</td>
<td>31MAR2010</td>
<td>0</td>
<td>1</td>
<td>&lt; 1 year</td>
<td>&lt; 365 DDD</td>
</tr>
</tbody>
</table>

Cumulative duration and dose variables were formatted as:

- **No use** - patient is unexposed during the follow-up
- **< 1 year** – cumulative exposure is less than 365 days or defined daily doses (DDD)
- 1 – 2 years – cumulative exposure is 365 – 730 days or DDDs
- 2 – 3 years – cumulative exposure is 731 – 1095 days or DDDs
- 3+ years – cumulative exposure is over 1095 days or DDDs
Combining time intervals

```sql
proc sql;
create table FINAL as
select id, event, statin_01, cumdur, cumdose, psa_count,
     min(istart) as start format date9.,
     max(iend) as end format date9.,
     (calculated start) - t0 as time1,
     (calculated end) - t0 as time2
from TD_combined
group by id, event, statin_01, cumdur, cumdose, psa_count
order by id, time1, time2;
quit;
```

Within PROC SQL we:
1) Combined daily episodes into informative intervals when changes occur;
2) Assigned the same time origin for all the patients.

Last step is to assign time-dependent event variable to be:
- 0 for all intervals prior to the last one;
- the value of event variable for the last interval.
## Final dataset and fitting PROC PHREG

<table>
<thead>
<tr>
<th>ID</th>
<th>EVENT</th>
<th>STATIN_01</th>
<th>CUMDUR</th>
<th>CUMDOSE</th>
<th>START</th>
<th>END</th>
<th>TIME1</th>
<th>TIME2</th>
<th>EVENT_TD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>No Use</td>
<td>No use</td>
<td>01JAN2002</td>
<td>01JUL2004</td>
<td>0</td>
<td>912</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>&lt; 1 year</td>
<td>&lt; 1 year</td>
<td>01JUL2004</td>
<td>31MAR2010</td>
<td>912</td>
<td>3011</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>No Use</td>
<td>No use</td>
<td>15APR2005</td>
<td>09OCT2007</td>
<td>0</td>
<td>907</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>No Use</td>
<td>No use</td>
<td>27SEP2004</td>
<td>15SEP2005</td>
<td>0</td>
<td>353</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>&lt; 1 year</td>
<td>&lt; 1 year</td>
<td>15SEP2005</td>
<td>01NOV2006</td>
<td>353</td>
<td>765</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1 – 2 years</td>
<td>1 – 2 years</td>
<td>01NOV2006</td>
<td>08AUG2008</td>
<td>765</td>
<td>1411</td>
<td>2</td>
</tr>
</tbody>
</table>

### Crude model with binary statin variable:

```sql
proc phreg data = FINAL;
   class statin_01/ref = first;
   model (time1, time2) * event_TD (0, 2) = statin_01/ties = EFRON rl;
run;
```
Take Home Messages

• Use of time-dependent vs time-fixed covariates offers a solution to “immortal time” bias and allows one to update information on covariates that vary over time.

• However, covariates must be carefully constructed to ensure interpretability.

• Counting process type of input may be more preferable in case of multiple time-dependent covariates BUT need to ensure:
  - time intervals do not overlap;
  - there are no intervals of zero length.
THANK YOU!

Questions?
Key References


