A novel approach to competing risks analysis using case-base sampling

Maxime Turgeon
April 27th, 2016

McGill University
Department of Epidemiology, Biostatistics, and Occupational Health
This project is joint work with:

- Sahir Bhatnagar
- Olli Saarela (U. Toronto)
- Jim Hanley
Introduction
Motivation

• In epidemiological studies of time-to-event data, a quantity of interest to the clinician and the patient is the absolute risk of an event, e.g. 5-year risk of developing cancer.

• In some settings, the analysis is complicated by the presence of competing events (e.g. complications due to bone-marrow transplant in a study of acute leukemia recurrence).

• A proper estimation of absolute risks needs to take these competing events into account.
Motivation

• In epidemiological studies of time-to-event data, a quantity of interest to the clinician and the patient is the absolute risk of an event, e.g. 5-year risk of developing cancer.
Motivation

- In epidemiological studies of time-to-event data, a quantity of interest to the clinician and the patient is the absolute risk of an event, e.g. 5-year risk of developing cancer.
- In some settings, the analysis is complicated by the presence of competing events (e.g. complications due to bone-marrow transplant in a study of acute leukemia recurrence).
Motivation

- In epidemiological studies of time-to-event data, a quantity of interest to the clinician and the patient is the absolute risk of an event, e.g. 5-year risk of developing cancer.
- In some settings, the analysis is complicated by the presence of competing events (e.g. complications due to bone-marrow transplant in a study of acute leukemia recurrence).
- A proper estimation of absolute risks needs to take these competing events into account.
Current methods

To compute absolute risks, we need to model the cause-specific hazards. Therefore, the Fine-Gray model is not appropriate. A common alternative is the Cox proportional hazards model. However, this model leads to a two-step procedure for estimating the hazard function.
Current methods

- To compute absolute risks, we need to model the cause-specific hazards.
Current methods

- To compute absolute risks, we need to model the cause-specific hazards.
  - Therefore, the Fine-Gray model is not appropriate.
Current methods

- To compute absolute risks, we need to model the *cause-specific hazards*.
  - Therefore, the Fine-Gray model is not appropriate.
- A common alternative is the Cox proportional hazards model.
Current methods

- To compute absolute risks, we need to model the *cause-specific hazards*.
  - Therefore, the Fine-Gray model is not appropriate.
- A common alternative is the Cox proportional hazards model.
  - However, this model leads to a two-step procedure for estimating the hazard function.
Summary

We propose a simple approach to modeling directly the cause-specific hazards using (smooth) parametric families. Our approach relies on Hanley & Miettinen's case base sampling method [1]. Our approach allows for a symmetric treatment of all time variables. Finally, it also allows for variable selection. This method is currently available as an R package: http://sahirbhatnagar.com/casebase/
• We propose a **simple** approach to modeling **directly** the cause-specific hazards using (smooth) parametric families.

Our approach relies on Hanley & Miettinen's case base sampling method [1].

Our approach allows for a **symmetric** treatment of all time variables.

Finally, it also allows for **variable selection**. This method is currently available as an R package: http://sahirbhatnagar.com/casebase/
• We propose a **simple** approach to modeling **directly** the cause-specific hazards using (smooth) parametric families.
  • Our approach relies on Hanley & Miettinen’s **case base sampling** method [1].
Summary

- We propose a **simple** approach to modeling **directly** the cause-specific hazards using (smooth) parametric families.
  - Our approach relies on Hanley & Miettinen’s **case base sampling** method [1].
- Our approach allows for a **symmetric** treatment of all time variables.

This method is currently available as an R package: [http://sahirbhatnagar.com/casebase/](http://sahirbhatnagar.com/casebase/)
Summary

- We propose a **simple** approach to modeling **directly** the cause-specific hazards using (smooth) parametric families.
  - Our approach relies on Hanley & Miettinen’s **case base sampling** method [1].
- Our approach allows for a **symmetric** treatment of all time variables.
- Finally, it also allows for **variable selection**.

This method is currently available as an R package: [http://sahirbhatnagar.com/casebase/](http://sahirbhatnagar.com/casebase/)
We propose a **simple** approach to modeling **directly** the cause-specific hazards using (smooth) parametric families.

- Our approach relies on Hanley & Miettinen’s **case base sampling** method [1].

Our approach allows for a **symmetric** treatment of all time variables.

Finally, it also allows for **variable selection**.

This method is currently available as an R package:

http://sahirbhatnagar.com/casebase/
Case-base sampling
Follow-up time (months)

Population

- Red: Relapse
- Blue: Competing event
- Black: Base series
Case-base sampling

- "Paradigm shift": the unit of analysis is a person-moment.
- Case-base sampling reduces the model fitting to a familiar multinomial regression.
- The sampling process is taken into account using an offset term.
- By sampling a large base series, the information loss eventually becomes negligible.
- This framework can easily be used with time-varying covariates (e.g. time-varying exposure).
Case-base sampling

- “Paradigm shift”: the unit of analysis is a *person-moment*. 
Case-base sampling

- “Paradigm shift”: the unit of analysis is a *person-moment*.
- Case-base sampling reduces the model fitting to a familiar multinomial regression.
Case-base sampling

- “Paradigm shift”: the unit of analysis is a *person-moment*.
- Case-base sampling reduces the model fitting to a familiar multinomial regression.
  - The sampling process is taken into account using an offset term.
Case-base sampling

- “Paradigm shift”: the unit of analysis is a *person-moment*.
- Case-base sampling reduces the model fitting to a familiar multinomial regression.
  - The sampling process is taken into account using an offset term.
- By sampling a large base series, the information loss eventually becomes negligible.
Case-base sampling

- “Paradigm shift”: the unit of analysis is a person-moment.
- Case-base sampling reduces the model fitting to a familiar multinomial regression.
  - The sampling process is taken into account using an offset term.
- By sampling a large base series, the information loss eventually becomes negligible.
- This framework can easily be used with time-varying covariates (e.g. time-varying exposure).
We can fit any model of the following form:

\[
\log \lambda(t; \alpha, \beta) = g(t; \alpha) + \beta X.
\]

Different choices of the function \(g\) lead to familiar parametric families:

- **Exponential**: \(g\) is constant.
- **Gompertz**: \(g(t; \alpha) = \alpha t\).
- **Weibull**: \(g(t; \alpha) = \alpha \log t\).
We can fit any model of the following form:

$$\log \lambda(t; \alpha, \beta) = g(t; \alpha) + \beta X.$$
We can fit any model of the following form:

$$\log \lambda(t; \alpha, \beta) = g(t; \alpha) + \beta X.$$ 

Different choices of the function $g$ leads to familiar parametric families:
We can fit any model of the following form:

\[
\log \lambda(t; \alpha, \beta) = g(t; \alpha) + \beta X.
\]

Different choices of the function \( g \) leads to familiar parametric families:

- **Exponential**: \( g \) is constant.
- **Gompertz**: \( g(t; \alpha) = \alpha t \).
- **Weibull**: \( g(t; \alpha) = \alpha \log t \).
We can fit any model of the following form:

$$\log \lambda(t; \alpha, \beta) = g(t; \alpha) + \beta X.$$ 

Different choices of the function $g$ leads to familiar parametric families:

- **Exponential**: $g$ is constant.
- **Gompertz**: $g(t; \alpha) = \alpha t$. 
We can fit any model of the following form:

$$\log \lambda(t; \alpha, \beta) = g(t; \alpha) + \beta X.$$  

Different choices of the function $g$ leads to familiar parametric families:

- **Exponential**: $g$ is constant.
- **Gompertz**: $g(t; \alpha) = \alpha t$.
- **Weibull**: $g(t; \alpha) = \alpha \log t$. 
Data analysis
<table>
<thead>
<tr>
<th>Variable description</th>
<th>Statistical summary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td>M=Male (100)</td>
</tr>
<tr>
<td></td>
<td>F=Female (77)</td>
</tr>
<tr>
<td><strong>Disease</strong></td>
<td>ALL (73)</td>
</tr>
<tr>
<td></td>
<td>AML (104)</td>
</tr>
<tr>
<td><strong>Phase</strong></td>
<td>CR1 (47)</td>
</tr>
<tr>
<td></td>
<td>CR2 (45)</td>
</tr>
<tr>
<td></td>
<td>CR3 (12)</td>
</tr>
<tr>
<td></td>
<td>Relapse (73)</td>
</tr>
<tr>
<td><strong>Type of transplant</strong></td>
<td>BM+PB (21)</td>
</tr>
<tr>
<td></td>
<td>PB (156)</td>
</tr>
<tr>
<td><strong>Age of patient (years)</strong></td>
<td>4–62</td>
</tr>
<tr>
<td></td>
<td>30.47 (13.04)</td>
</tr>
<tr>
<td><strong>Failure time (months)</strong></td>
<td>0.13–131.77</td>
</tr>
<tr>
<td></td>
<td>20.28 (30.78)</td>
</tr>
<tr>
<td><strong>Status indicator</strong></td>
<td>0=censored (46)</td>
</tr>
<tr>
<td></td>
<td>1=relapse (56)</td>
</tr>
<tr>
<td></td>
<td>2=competing event (75)</td>
</tr>
</tbody>
</table>
Absolute risk for female patient, median age, in relapse at transplant (stem cells from peripheral blood).
<table>
<thead>
<tr>
<th>Variable</th>
<th>Hazard ratio 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.68 (0.39, 1.20)</td>
</tr>
<tr>
<td>Disease</td>
<td>0.51 (0.28, 0.92)</td>
</tr>
<tr>
<td>Phase CR2</td>
<td>1.18 (0.47, 2.96)</td>
</tr>
<tr>
<td>Phase CR3</td>
<td>1.51 (0.39, 5.86)</td>
</tr>
<tr>
<td>Phase Relapse</td>
<td>4.38 (2.01, 9.54)</td>
</tr>
<tr>
<td>Source</td>
<td>1.37 (0.45, 4.23)</td>
</tr>
<tr>
<td>Age</td>
<td>0.99 (0.97, 1.02)</td>
</tr>
<tr>
<td>Variable</td>
<td>Hazard ratio</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Sex</td>
<td>0.68</td>
</tr>
<tr>
<td>Disease</td>
<td>0.51</td>
</tr>
<tr>
<td>Phase CR2</td>
<td>1.18</td>
</tr>
<tr>
<td>Phase CR3</td>
<td>1.51</td>
</tr>
<tr>
<td>Phase Relapse</td>
<td>4.38</td>
</tr>
<tr>
<td>Source</td>
<td>1.37</td>
</tr>
<tr>
<td>Age</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Discussion
• We proposed a simple and flexible way of directly modeling the hazard function.
• This leads to smooth estimates of the absolute risks.
• We are explicitly modeling time, and we cannot model the effect of time on the hazard function.
• We can test the significance of covariates, in a similar way to traditional competing risks approaches.
• We proposed a simple and flexible way of directly modeling the hazard function.
• We proposed a simple and flexible way of directly modeling the hazard function.
  • This leads to smooth estimates of the absolute risks.
Discussion

• We proposed a simple and flexible way of directly modeling the hazard function.
  • This leads to smooth estimates of the absolute risks.
• We are explicitly modeling time, and we cant therefore how to model the effect of time on the hazard function.
• We proposed a simple and flexible way of directly modeling the hazard function.
  • This leads to smooth estimates of the absolute risks.
• We are explicitly modeling time, and we cant therefore how to model the effect of time on the hazard function.
• We can test the significance of covariates, in a similar way to traditional competing risks approaches.
J. A. Hanley and O. S. Miettinen.  
**Fitting smooth-in-time prognostic risk functions via logistic regression.**  
*The International Journal of Biostatistics, 5*(1), 2009.

O. Saarela.  
**A case-base sampling method for estimating recurrent event intensities.**  
*Lifetime data analysis, pages 1–17, 2015.

O. Saarela and J. A. Hanley.  
**Case-base methods for studying vaccination safety.**  
L. Scrucca, A. Santucci, and F. Aversa.

Regression modeling of competing risk using R: an in depth guide for clinicians.

Questions or comments?

For more details, visit
http://sahirbhatnagar.com/casebase/