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

Acknowledgements

This project is joint work with:

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

Introduction

Motivation

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

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

Summary

- We propose a **simple** approach to modeling **directly** the cause-specific hazards using (smooth) parametric families.

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].

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.

Summary

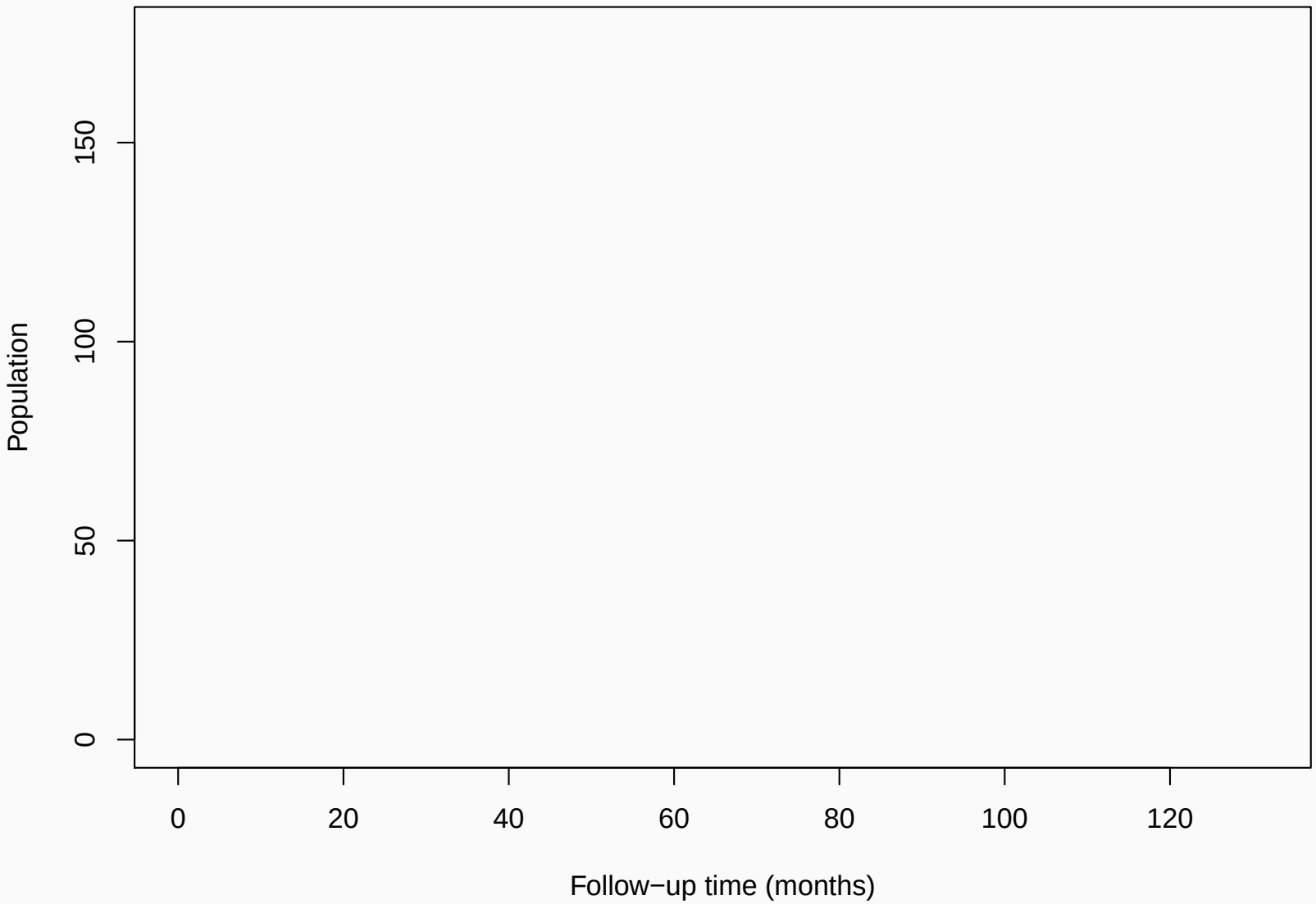
- We propose a **simple** approach to modeling **directly** the cause-specific hazards using (smooth) parametric families.
    - Our approach relies on Hanley & Miettinen's **case base sampling** method [1].
- Our approach allows for a **symmetric** treatment of all time variables.
- Finally, it also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

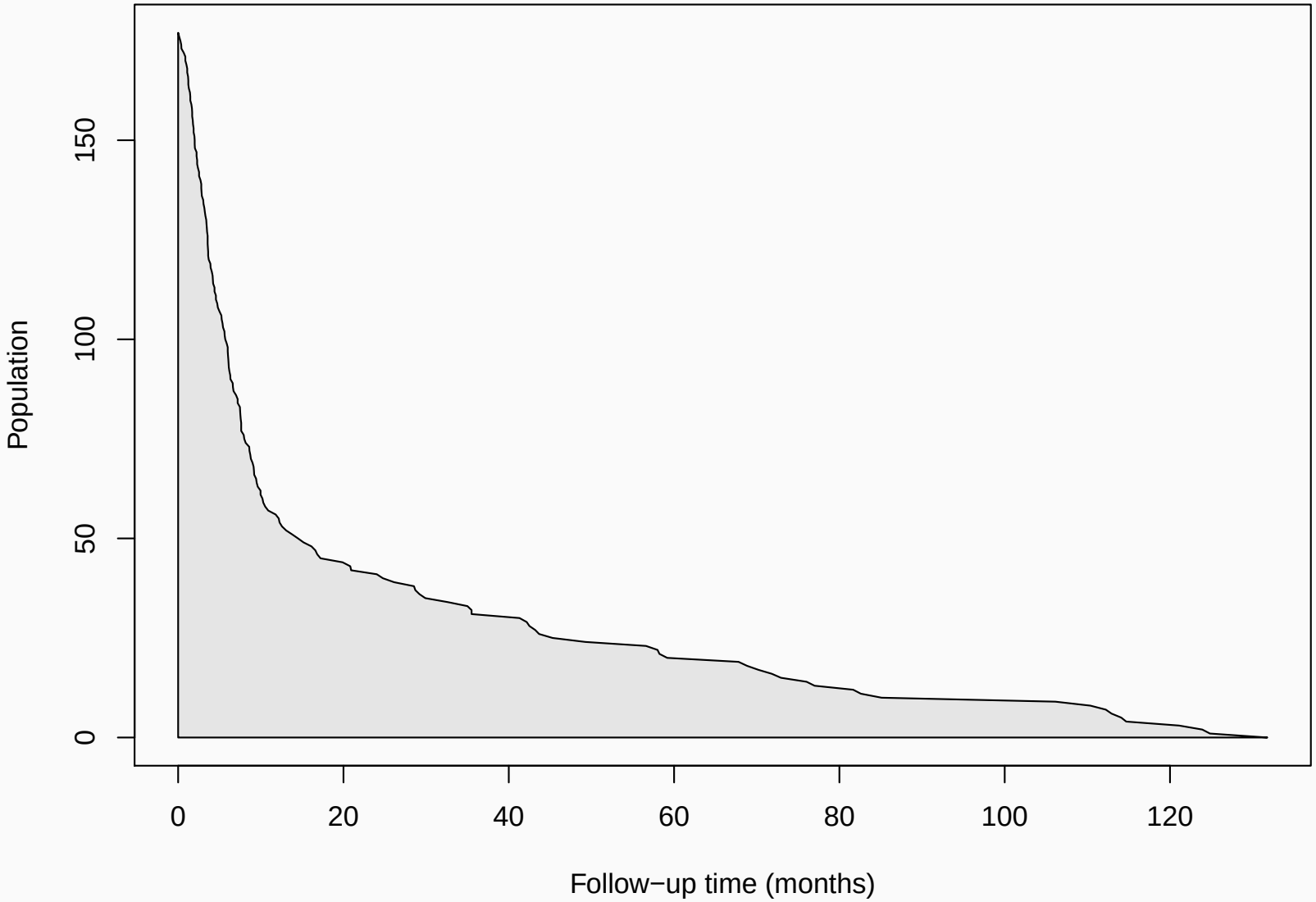
Summary

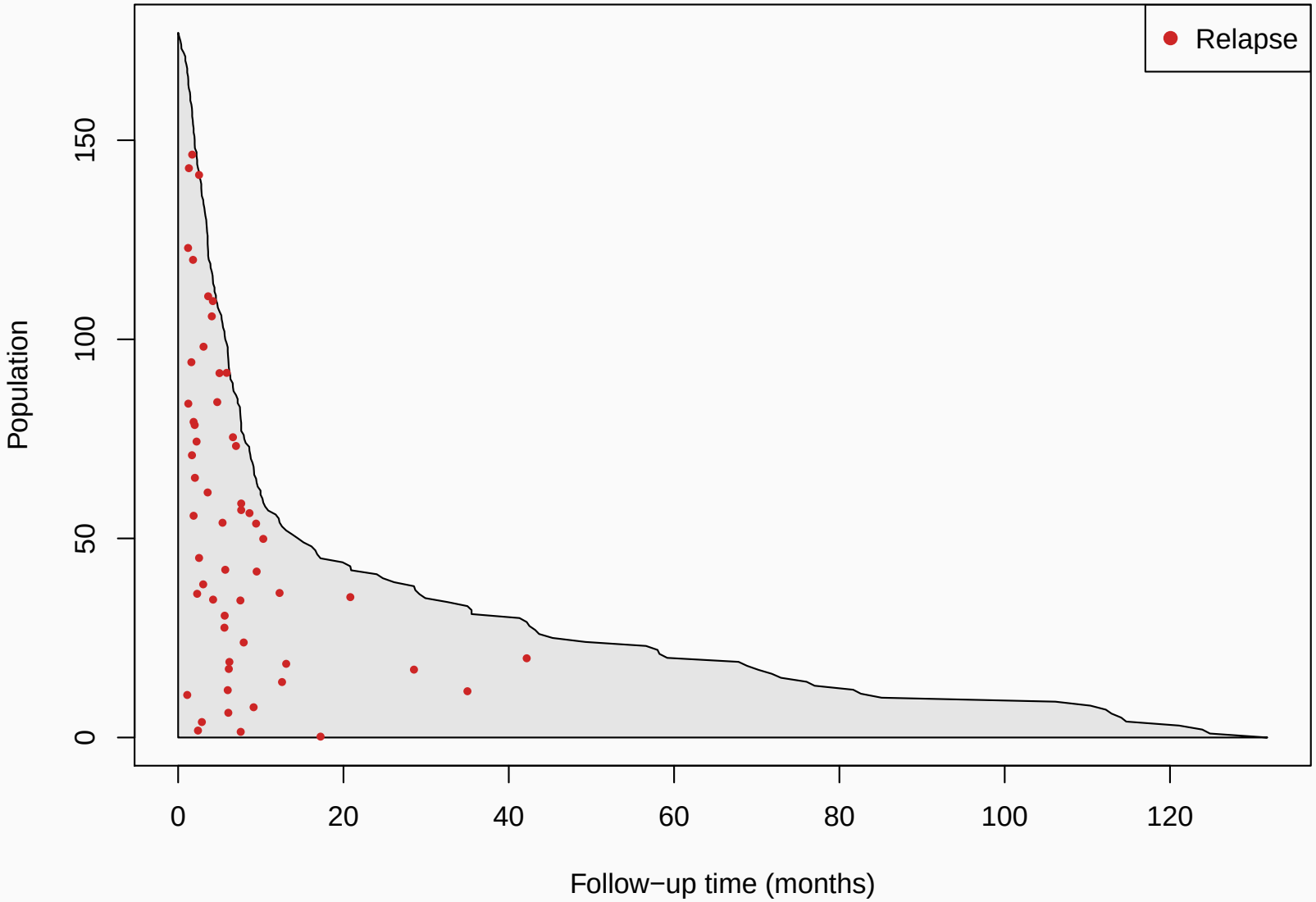
- We propose a **simple** approach to modeling **directly** the cause-specific hazards using (smooth) parametric families.
    - Our approach relies on Hanley & Miettinen's **case base sampling** method [1].
- Our approach allows for a **symmetric** treatment of all time variables.
- Finally, it also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

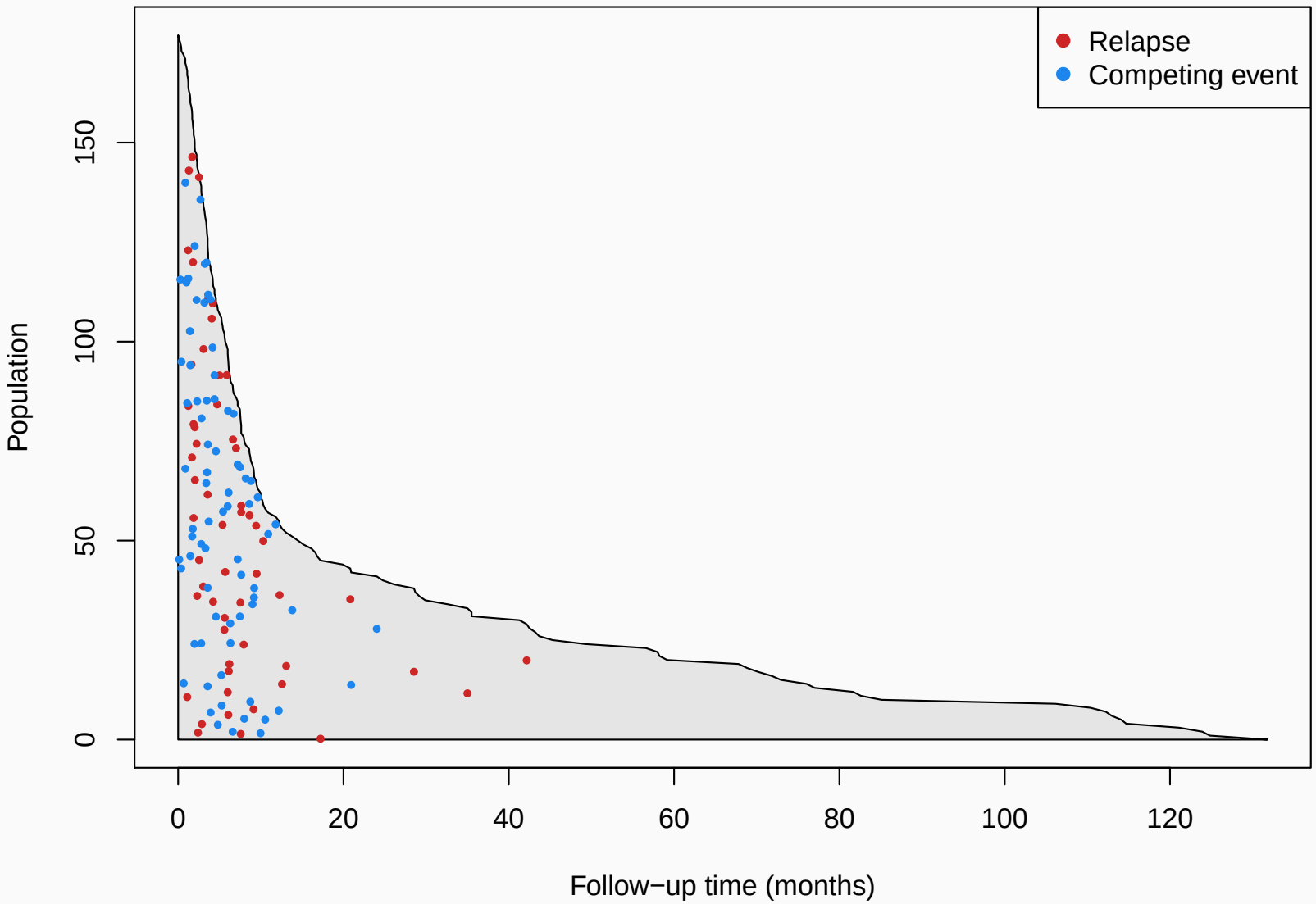
This method is currently available as an R package:
http://sahirbhatnagar.com/casebase/

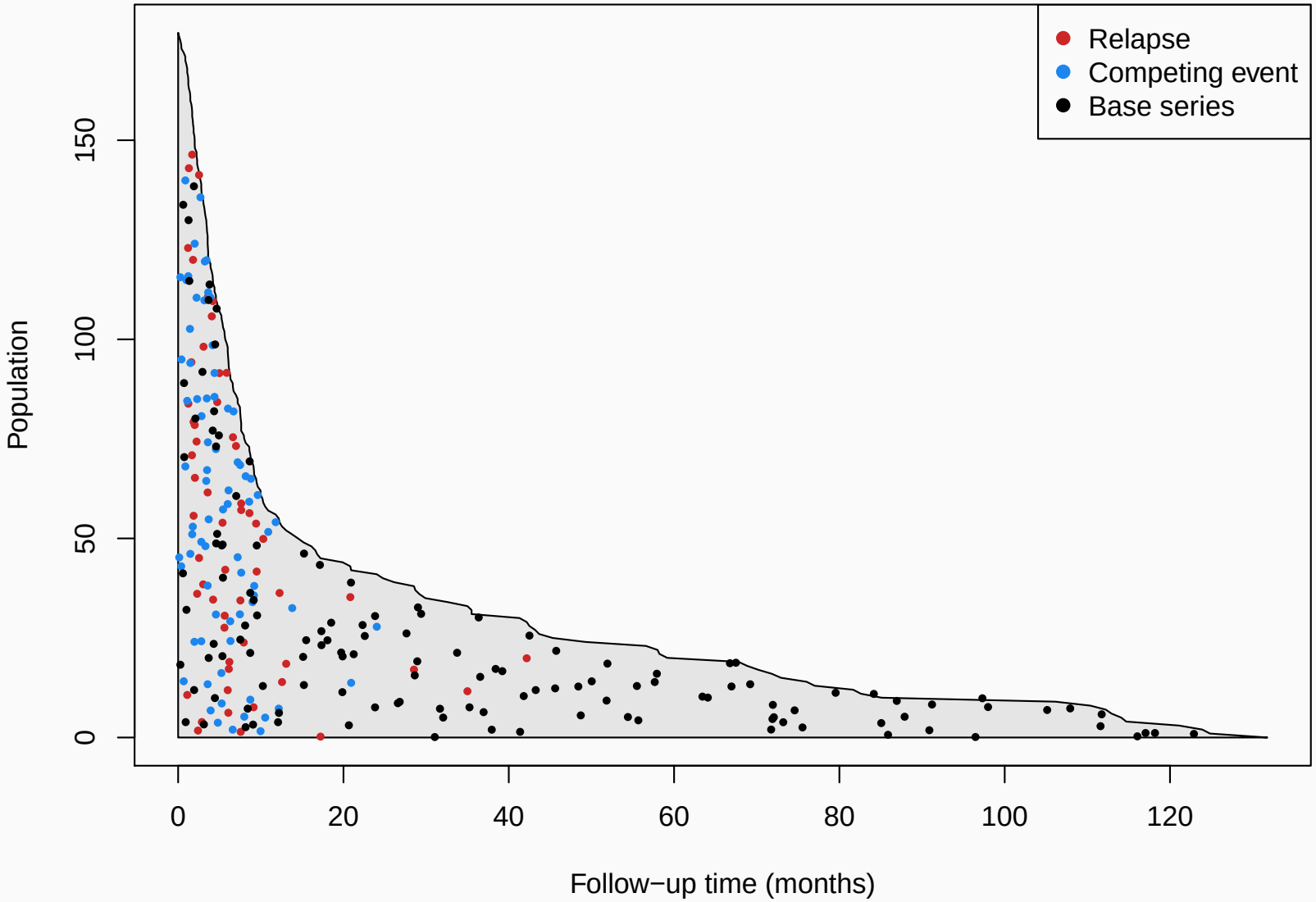
Case-base sampling


Population
0
50
100
150
0
20
40
60
80
100
120
Follow−up time (months)


Population
0
50
100
150
0
20
40
60
80
100
120
Follow−up time (months)


Relapse
Population
0
50
100
150
0
20
40
60
80
100
120
Follow−up time (months)


Relapse
Competing event
Population
0
50
100
150
0
20
40
60
80
100
120
Follow−up time (months)


Relapse
Competing event
Base series
Population
0
50
100
150
0
20
40
60
80
100
120
Follow−up time (months)

Case-base sampling

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).

Parametric families

Parametric families

We can fit any model of the following form:

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

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:

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.

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: $g(t; \alpha) = \alpha t$.

Parametric families

We can fit any model of the following form:

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

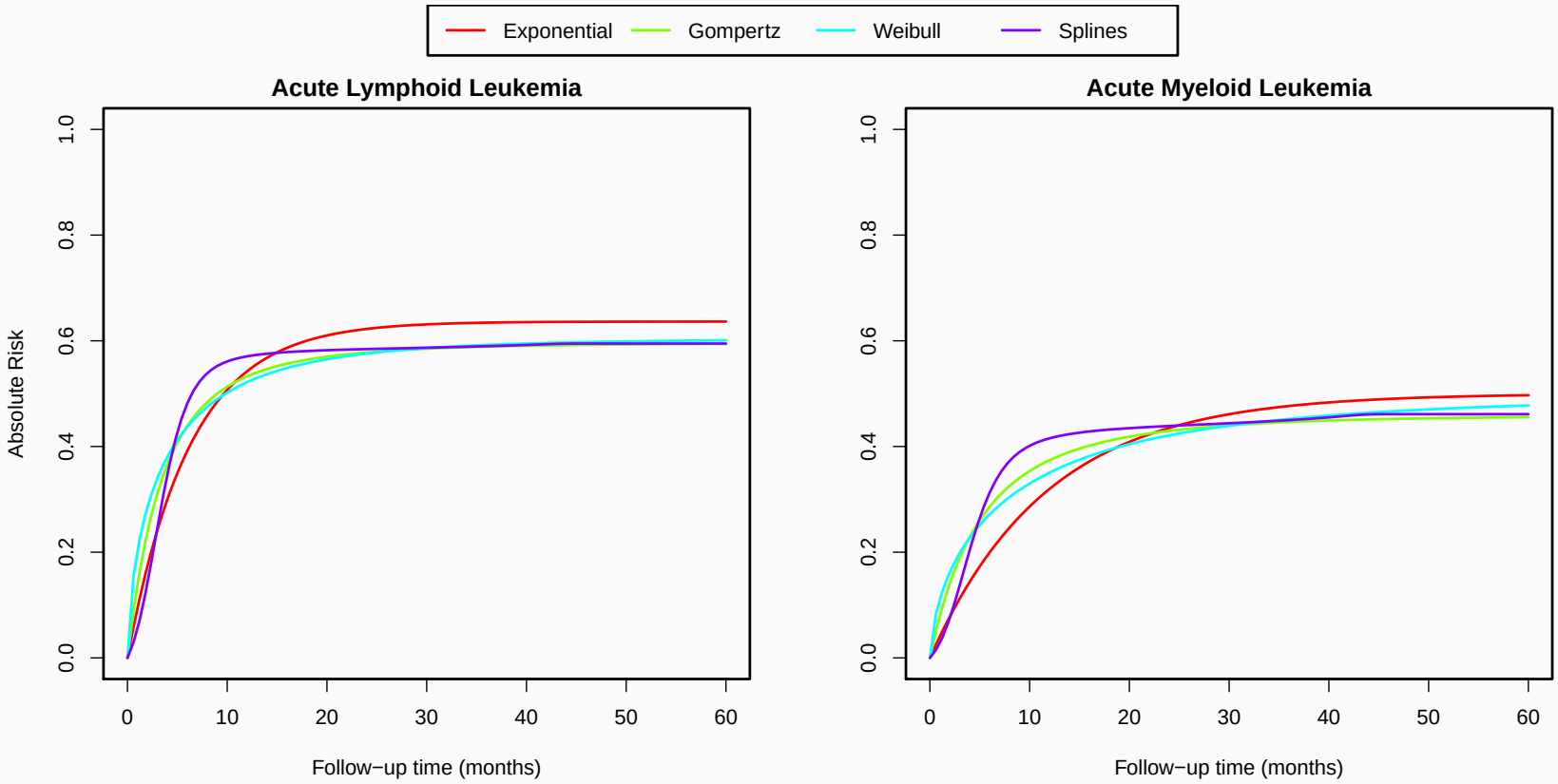
Different choices of the function g leads to familiar parametric families:

- Exponential: g is constant.
- Gompertz: $g(t; \alpha) = \alpha t$.
- Weibull: $g(t; \alpha) = \alpha \log t$.

Data analysis

Data

Variable description	Statistical summary
Sex	M=Male (100)
	F=Female (77)
Disease	ALL (73)
	AML (104)
Phase	CR1 (47)
	CR2 (45)
	CR3 (12)
	Relapse (73)
Type of transplant	BM+PB (21)
	PB (156)
Age of patient (years)	4–62
	30.47 (13.04)
Failure time (months)	0.13–131.77
	20.28 (30.78)
Status indicator	0=censored (46)
	1=relapse (56)
	2=competing event (75)



Absolute risk for female patient, median age, in relapse at transplant (stem cells from peripheral blood).

Model fit

Model fit

Variable	Hazard ratio	95% CI
Sex	0.68	(0.39, 1.20)
Disease	0.51	(0.28, 0.92)
Phase CR2	1.18	(0.47, 2.96)
Phase CR3	1.51	(0.39, 5.86)
Phase Relapse	4.38	(2.01, 9.54)
Source	1.37	(0.45, 4.23)
Age	0.99	(0.97, 1.02)

Discussion

Discussion

Discussion

- We proposed a simple and flexible way of directly modeling the hazard function.

Discussion

- 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 explicitely modeling time, and we cant therefore how to model the effect of time on the hazard function.

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 explicitely 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.

References I

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.
Biometrics, 71(1):42–52, 2015.

References II

L. Scrucca, A. Santucci, and F. Aversa.
Regression modeling of competing risk using R: an in depth guide for clinicians.
Bone marrow transplantation, 45(9):1388–1395, 2010.

Questions or comments?

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