Performance of Variance Estimators of The Cumulative Incidence Function for Left-Truncated Competing Risks Data

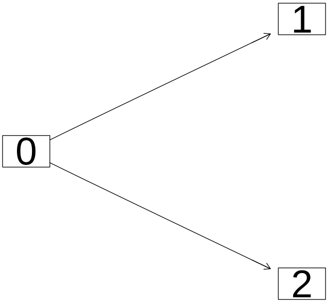
Arthur Allignol[1,2,*] Martin Schumacher[2]
Reinhard Meister[3] Jan Beyersmann[1,2]

[1]Freiburg Center for Data Analysis and Modeling, University of Freiburg
[2]Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg
[3]Department of Mathematics, University of Applied Science,
Technische Fachhochschule Berlin
[*]arthur.allignol@fdm.uni-freiburg.de

DFG Forschergruppe FOR 534

UNIVERSITÄTS KLINIKUM FREIBURG

Competing Risks Multistate Model



- Competing risks process: $X_{t \geq 0} \in \{0, 1, 2\}$
- Failure time: $T = \inf(t \in [0, \infty\} \,|\, X_t \neq 0)$
- Time until the first event T, and type of this first event X_T

Competing Risks Multistate Model

- Cause-specific hazard (CSH)

$$\alpha_{0j}(t)dt = \mathrm{P}(T \in dt, X_T = j | T \geq t), \, j = 1, 2$$

- Cumulative incidence function (CIF)

$$\mathrm{P}(T \leq t, X_T = j) = \int_0^t \mathrm{P}(T \geq u-)\alpha_{0j}(u)du, \, j = 1, 2,$$

which is estimated by the Aalen-Johansen estimator in the presence of right-censoring and left-truncation

Variance Estimation

- Two major variance estimators
  - Greenwood
  - Aalen
- Derived using Martingale and counting processes techniques for the general Aalen-Johansen estimator of transition probabilities in time-inhomogeneous Markov multistate models (Andersen et al., 1993)

Previous Findings

- Klein (1991) compared the analogues of the Greenwood and Aalen variance estimators for the variance of the Kaplan-Meier estimator
  - Both estimators were found to be negatively biased
  - Greenwood estimator had the smallest absolute bias

Previous Findings

- Klein (1991) compared the analogues of the Greenwood and Aalen variance estimators for the variance of the Kaplan-Meier estimator
  - Both estimators were found to be negatively biased
  - Greenwood estimator had the smallest absolute bias

- Estimating the variance of the CIF in small samples has been investigated recently in the presence of right-censoring
  - Some estimators performed poorly

    $\Rightarrow$ Critics of Martingale-based estimators

Previous Findings

- Klein (1991) compared the analogues of the Greenwood and Aalen variance estimators for the variance of the Kaplan-Meier estimator
  - Both estimators were found to be negatively biased
  - Greenwood estimator had the smallest absolute bias
- Estimating the variance of the CIF in small samples has been investigated recently in the presence of right-censoring
  - Some estimators performed poorly

    $\Rightarrow$ Critics of Martingale-based estimators
- The preferred estimator in these simulation studies is in fact equal to the Greenwood variance estimator (Allignol et al., 2009b)

Previous Findings

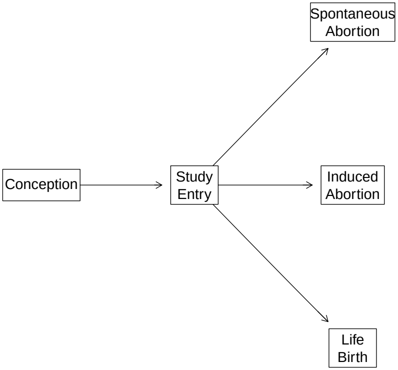
- Klein (1991) compared the analogues of the Greenwood and Aalen variance estimators for the variance of the Kaplan-Meier estimator
  - Both estimators were found to be negatively biased
  - Greenwood estimator had the smallest absolute bias
- Estimating the variance of the CIF in small samples has been investigated recently in the presence of right-censoring
  - Some estimators performed poorly
    $\Rightarrow$ Critics of Martingale-based estimators
- The preferred estimator in these simulation studies is in fact equal to the Greenwood variance estimator (Allignol et al., 2009b)
- Objectives:
  Extend these simulation studies to the left-truncated competing risks data situation
  - With left-truncation, risk sets are extremely small at the beginning of the time interval

Data Example (Meister and Schaefer, 2008)

- Study on spontaneous abortion in pregnancies exposed to coumarin derivatives

Data Example (Meister and Schaefer, 2008)

- Study on spontaneous abortion in pregnancies exposed to coumarin derivatives



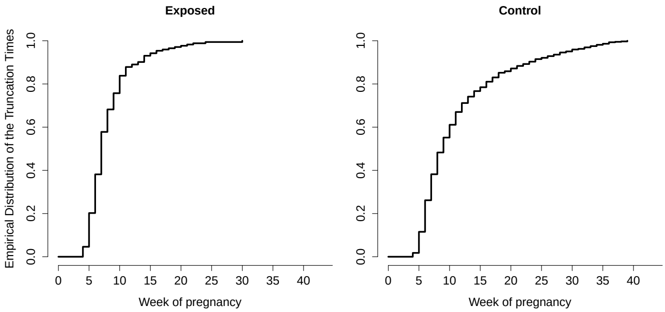
- Spontaneous abortion is the event of interest
- 2 competing events:
  - Induced Abortion
  - Life Birth
- Data subject to left-truncation as women usually enter the study several weeks after conception

Data Example

- Informations on 1186 pregnant women
  - 173 exposed to coumarin derivatives
  - 1013 controls

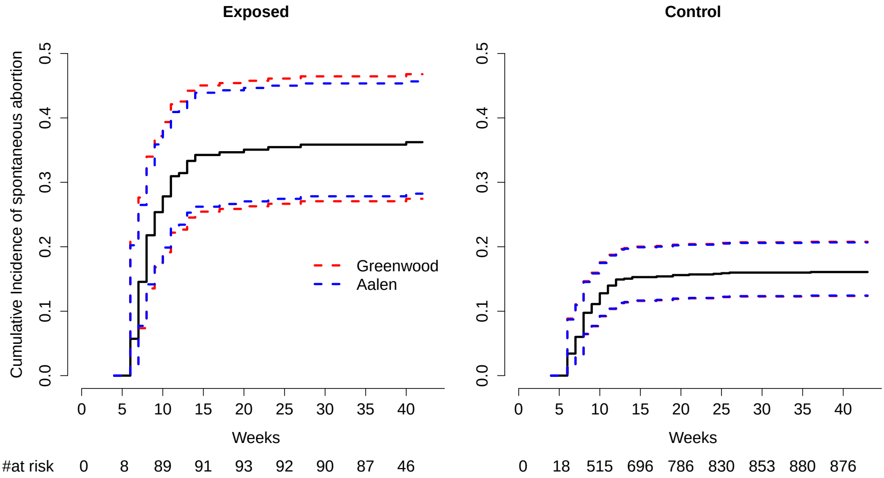
Data Example

- Informations on 1186 pregnant women
  - 173 exposed to coumarin derivatives
  - 1013 controls



- Independence of left-truncation has been investigated (Allignol et al., 2009b)

Data Example



Simulation Study

- Simulation of competing risks data with 2 risks, event of type 1 being the one of interest
- Specification of the cause-specific hazards motivated by the data example
  - Constant hazard $\alpha_{01}(t) = 0.015$
  - Weibull-type hazard $\alpha_{02}(t) = 10.5/39.5 \cdot (t/39.5)^{9.5}$
- Simulation algorithm as in Beyersmann et al. (2009)

Simulation Study

- Simulation of competing risks data with 2 risks, event of type 1 being the one of interest
- Specification of the cause-specific hazards motivated by the data example
  - Constant hazard $\alpha_{01}(t) = 0.015$
  - Weibull-type hazard $\alpha_{02}(t) = 10.5/39.5 \cdot (t/39.5)^{9.5}$
- Simulation algorithm as in Beyersmann et al. (2009)
- Gamma distributed truncation times L independent of (T, X_T) with parameters chosen to give light, medium and heavy truncation

Simulation Study

- Simulation of competing risks data with 2 risks, event of type 1 being the one of interest
- Specification of the cause-specific hazards motivated by the data example
  - Constant hazard $\alpha_{01}(t) = 0.015$
  - Weibull-type hazard $\alpha_{02}(t) = 10.5/39.5 \cdot (t/39.5)^{9.5}$
- Simulation algorithm as in Beyersmann et al. (2009)
- Gamma distributed truncation times L independent of (T, X_T) with parameters chosen to give light, medium and heavy truncation
- We simulated $N \in \{100, 200, 300\}$ individuals. Only those with $L < T$ entered the sample
  - Light left-truncation: $0.85 \cdot N$ on average
  - Medium left-truncation $0.80 \cdot N$ on average
  - Heavy left-truncation $0.45 \cdot N$ on average
- 10 000 simulation studies for each scenario

Simulation Study

- Results are presented in terms of:
  - Relative bias
$$((\bar{\hat{V}} - V)/V)$$
  - Root mean squared error
$$\sqrt{(\bar{\hat{V}} - V)^2 + \mathrm{SE}(\hat{V})^2}$$

V is the empirical variance
$\bar{\hat{V}}$ is the average of the estimated variance $\hat{V}$

Simulation Study

- Results are presented in terms of:
  - Relative bias

$$((\bar{\hat{V}} - V)/V)$$

  - Root mean squared error

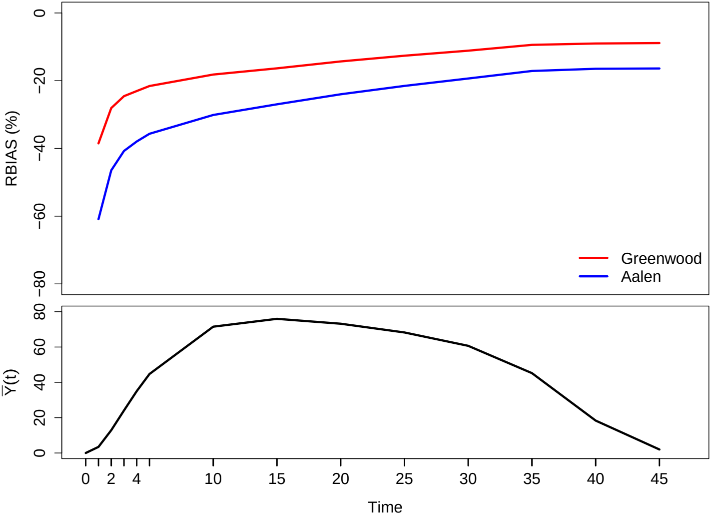
$$\sqrt{(\bar{\hat{V}} - V)^2 + \mathrm{SE}(\hat{V})^2}$$

V is the empirical variance
$\bar{\hat{V}}$ is the average of the estimated variance $\hat{V}$

- Empty inner risk sets might occur with left-truncated data
  - Risk set $Y(t) = 0$ for some t after the first study entry time and before the last event time
  - Estimators rendered useless as they remain constant thereafter
  - Removed of the study, in line with Keiding and Gill (1990)

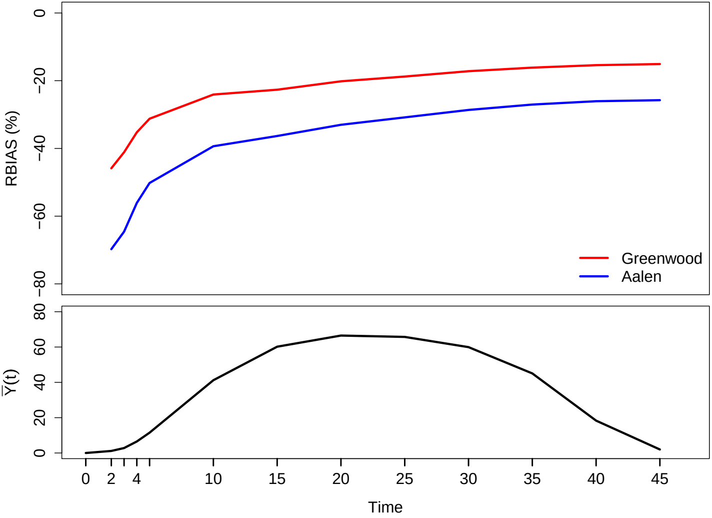
Light Left-Truncation

100 Individuals



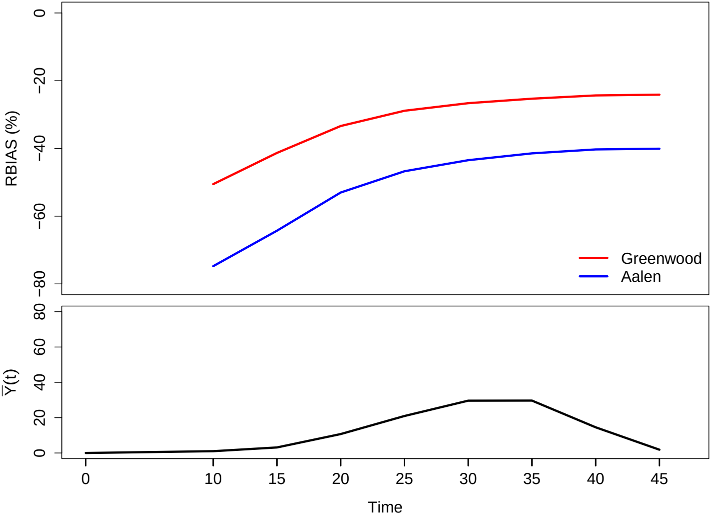
Medium Left-Truncation

100 Individuals



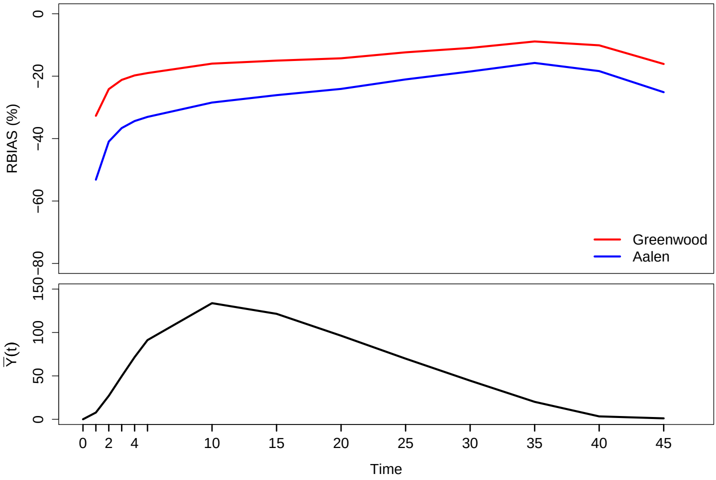
Heavy Left-Truncation

100 Individuals



Light Left-Truncation, Heavy Right-Censoring

200 Individuals



Summary

- The Greenwood estimator was found to perform well with left-truncated data
- It is algebraically identical to the one that performed the best in right-censored competing risks data
- The Greenwood estimator also available for more complex multistate models
- Implemented in the R-package **etm** (Allignol et al., 2009a)

Summary

- The Greenwood estimator was found to perform well with left-truncated data
- It is algebraically identical to the one that performed the best in right-censored competing risks data
- The Greenwood estimator also available for more complex multistate models
- Implemented in the R-package **etm** (Allignol et al., 2009a)
- Outlook:
  - Potential ways to repair the small sample bias
  - Contribution of truncation and censoring in the 'overall' bias

Summary

- The Greenwood estimator was found to perform well with left-truncated data
- It is algebraically identical to the one that performed the best in right-censored competing risks data
- The Greenwood estimator also available for more complex multistate models
- Implemented in the R-package **etm** (Allignol et al., 2009a)
- Outlook:
  - Potential ways to repair the small sample bias
  - Contribution of truncation and censoring in the 'overall' bias

$\Rightarrow$ We recommend the use of the Greenwood estimator both for right-censored and left-truncated competing risks data

Bibliography

Allignol, A., Schumacher, M., and Beyersmann, J. (2009a). Empirical transition matrix of multistate models: The etm package. *Submitted*.

Allignol, A., Schumacher, M., and Beyersmann, J. (2009b). A note on variance estimation of the Aalen-Johansen estimator of the cumulative incidence function in competing risks, with a view towards left-truncated data. *Submitted*.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes.* Springer Series in Statistics. New York, NY: Springer.

Beyersmann, J., Latouche, A., Buchholz, A., and Schumacher, M. (2009). Simulating competing risks data in survival analysis. *Statistics in Medicine*, 28:956–971.

Keiding, N. and Gill, R. D. (1990). Random truncation models and Markov processes. *The Annals of Statistics*, 18:582–602.

Klein, J. P. (1991). Small sample moments of some estimators of the variance of the Kaplan–Meier and Nelson–Aalen estimators. *Scandinavian Journal of Statistics*, 18:333–340.

Meister, R. and Schaefer, C. (2008). Statistical methods for estimating the probability of spontaneous abortion in observational studies–analyzing pregnancies exposed to coumarin derivatives. *Reproductive Toxicology*, 26(1):31–35.