The Kaplan-Meier theater

Thomas Alexander Gerds

Research Report 14/06

Department of Biostatistics
University of Copenhagen

THE KAPLAN-MEIER THEATER

THOMAS ALEXANDER GERDS

1. Abstract

The Kaplan-Meier theater is an interactive teaching lesson which lasts about one hour. It is designed for a class of at least 16 students. The aim is to base an intuitive understanding of how the Kaplan-Meier method deals with right censored data on an unforgettable experience. The idea is to illustrate Efron's re-distribution to the right algorithm based on data collected during the class. The students will learn why naive summaries of right censored survival data can lead to wrong conclusions. The theater also explains the main assumptions underlying the Kaplan-Meier method and why it fails in situations with competing risks.

2. Introduction

In this article I present the concept for an interactive teaching lesson about one of the most important statistics of medical and epidemiological research: the Kaplan-Meier estimator. The idea occurred to me in the early in morning of the third day of a three day course on statistics for medical researchers. On the very same day the Kaplan-Meier theater had its premiere. This is to say that the lesson does not require a lot of preparation and that it is useful to ease teaching of a complex statistical topic to students without a strong background mathematics. However, the material is also useful for students of (bio-)statistics who usually otherwise only learn about the product-limit formula of the Kaplan-Meier estimate. Over the years I have received highly positive feedback from students with all kinds of backgrounds. In addition of a class with at least 16 students, only few items are needed to perform the lesson (see Appendix A).

Date: October 9, 2014.

3. On the Titanic

After some introductory slides about the different aims and the main parameters of survival analysis, the next LaTeX beamer slide (see Appendix B) announces the theater. In the following I am the teacher and text in *italics* refers to what I would say during the class. I start by asking: *Are you ready?* People are looking curious and some a little scared, but usually everyone is nodding agreement. The original and most successful form of the story to be told is the following.

We are all on the Titanic, and the Titanic is going down. Once under water, one has to hold the breath. But for how long can you do this? Every second person in the room will participate as an actor in the theater. The nearest not participating neighbors will act as timekeepers and count the number of seconds the participants can hold their breath. For this exceptional purpose the idea of equipping every civilized person with a mobile phone is very useful. However, we will not all start simultaneously. Because the Titanic is sinking slowly, the participants touch the water for the first time at different time-points. This is to mimic the situation of a clinical trial where patients meeting the inclusion criteria are not enrolled on the same day but within an accrual period which can last several years. The accrual period of our theater will last about 50 seconds. As soon as each participant has found a timekeeper, I will start the trial. For this I have this vintage sports stop watch. When I knock on the first table at which the first participant is sitting, the trial has started. At this sign the first participant starts holding the breath. About five seconds later, I knock on the next table, hereby enrolling the second participant. In this way I move through the classroom until all participants are enrolled and have started to hold their breath. This is the end of the accrual period. The participants continue to try to hold their breath until I say "stop". This is the end of the follow-up period of the trial. So the individual follow-up period stops for all participants at the same time point. We can imagine that the person who was recording the data on the Titanic had to save herself at some critical time, and thus only data are available up to this time. In the case of a clinical trial the stop-time *is usually the time of statistical analysis.*

Then we will collect the data. A participant who was not able to hold the breath until the stop-time is said to have "drowned". For those who have "drowned" we will need the number of seconds between enrollment and the moment when they cannot hold their breath. A participant who was still able to hold the breath at the stop-time is said to have "survived". For those who "survived" we will need the number of seconds until the end of the trial. Once everyone is set and ready, we start to let the Titanic sink.

Numbers of participants between 8 and 12 work well. The aim is to have a mixture of participants who experience an event within the study period ("drowned") and participants who do not ("survived"). The event time is right censored for those who "survived". In a given class of students the outcome will depend on the athletic condition and personal ambition of the students and is generally hard to forecast. However, the results are influenced by the enrollment process. If the last person enrolled looks strong, then this person will likely contribute with a small right censored event time. Useful results are expected, and were achieved in previous performances of the theater, when the trial is stopped about 90 seconds after the first person was enrolled, that is about 40 seconds after the end of the accrual period. To be sure that there are at least some events one can arrange with the participants that they raise an arm as soon as they cannot hold the breath anymore.

4. When simple summary statistics fail

Participants who "drowned" write the number of seconds with at most one decimal in large red colored letters on a large (letter size/din A4) piece of paper. Participants who "survived" write the number of seconds in large green colored letters. All participants take their piece of paper and come to the front.

Ideally, the "participants" can be lined up in front of the rest of the class, such that the teacher (you) can still write on a broadly visible part of the board.

The first task is that you sort the line of participants from left to right in increasing order of the number of seconds.

This action is sketched in Figure 1. Once the participants have ordered, the first question goes to the people who are still sitting (timekeepers and other non-participants). Here are proposals for the first question:

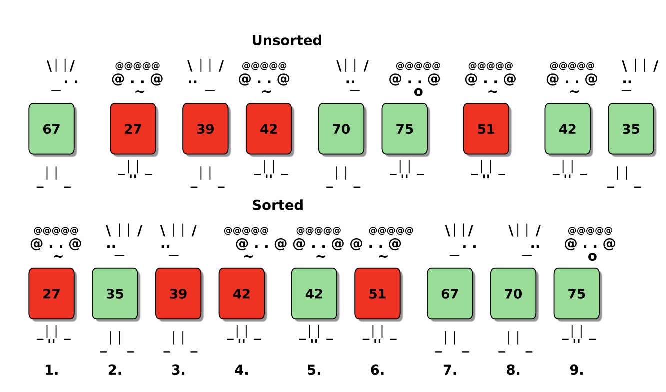

FIGURE 1. Comic of the results of a Kaplan-Meier theater. The upper panel illustrates the situation when the participants have lined up in front of the board. The lower panel shows the same participants when they have ordered themselves according to the number of seconds. It is a convention of survival analysis that in case of a tie between an event time and a censored time, the event time comes first. That is why in the lower panel the person with the red 42 seconds is standing to the left of the person with the green 42 seconds.

Based on the data that you see,

- *what is the probability to survive 40 seconds?*
- *what is the median survival time?*
- *what is the mean survival time?*

The aim of the first question is to obtain a wrong answer. Thus, in order to increase the likelihood of receiving a wrong answer to the first of the proposed questions there should be at least one of the participants lost to follow-up before 40 seconds. Note that the answer to the mean survival time is always wrong when there is at least one green number. If really no one is willing to give a wrong answer, one can anyway proceed by asking:

Why would it be wrong to estimate the probability to survive 40 by the relative frequency of participants who have more than 40 seconds on their piece of paper?

After some reflection someone in the class will discover the problem, namely that for the participants who have a green number of seconds which is smaller than 40, it is unknown when the event happened. The teacher will then introduce that the Kaplan-Meier method solves this problem.

5. The algorithm in action

The following description of how Efron's re-distribution to the right algorithm works is tailored to the data shown in Figure 1. On the board we draft a table with the following columns: *Time*, *Number at-risk*, *Number of events*, *Number lost to followup*, *Survival probability (%)*. Then we fill the first row with the data at time zero. See Table 1 for the final version which is obtained at the end of the theater.

5.1. **Participant number 1.**

We address the first participant, i.e., the person standing most to the left in the lower panel of Figure 1:

You have "drowned" after 27 seconds. At the time where you have "drowned" all other 8 participants were still alive and still in the study. In a sample of 9 participants each participant weighs 1/9. The value 1/9 is the probability mass assigned to each participant. Thus, at 27 seconds the Kaplan-Meier estimate of survival drops from 100% by 1/9 and takes on the value 88.9%. Please take your piece of paper and sit-down.

These statements can as well be obtained by communicating with the audience. No matter how they are obtained, they lead us to write the second line of the table (see Table 1).

5.2. **Participant number 2.**

Now we address the second participant:

You "survived" 35 seconds, and then we lost track of what happened to you. The Kaplan-Meier method assumes that you "drowned" at a later time point (there are no *competing risks). The Kaplan-Meier method assumes that your survival probabilities after 35 seconds are equal to the survival probabilities of the remaining seven participants standing to your right. Under these assumptions it is consistent that your contribution to the Kaplan-Meier estimate occurs at time points later than 35 seconds. This is done by distributing your probability mass to the right. Note that all the remaining seven participants standing on your right hand side have a time greater than 35 seconds. You tear your piece of paper into seven equally large pieces and then you give one piece to each of the remaining seven participants.*

In order to aid the computations at later stages of the algorithm, the participant can write the fractional weight of the distributed paper pieces on the back side of each of the 7 pieces of paper, i.e., $1/9 \times 1/7$. However, it is good training for the memory and it saves time not to do so. Appendix C provides some R-code which provides the details of the computations for a data set.

When you have done this you may sit down. The Kaplan-Meier estimate does not change at this time, and the following data are written into the third line of our table: 35, 8, 0, 1, 88.9%.

5.3. **Participant number 3.** *Participant number 3 has "drowned" after 39 seconds. Thus, the value of the Kaplan-Meier estimate is reduced by 1/9 at 39 seconds. However, it may be that participant number 2 has also "drowned" in the period between 35 seconds (were we lost track of his fate) and 39 seconds. To take this possibility into account the Kaplan-Meier estimate drops not just by 1/9, which is the weight of participant number 3, but in addition drops by the weight of the piece of paper that participant number 3 has received from participant number 2. Thus, at 39 seconds the Kaplan-Meier estimate of survival is reduced by the weight corresponding to the total amount of paper participant number 3 has in his hands:*

$$\underbrace{1/9}_{\text{own contribution}} \quad + \quad \underbrace{1/9 \times 1/7}_{\text{contr. from participant no. 2}}$$

5.4. **Participants number 4 and 5.** *The action for participant number 4 is similar to that just performed for participant number 3. The contribution of participant number 4 leads to a drop in the estimated survival probability at 42 seconds of* $1/9+1/9\times1/7$*. You (participant number 4) may take the two pieces of papers with you and sit down.*

Before we can write the fourth line of the table we need to deal with participant number 5 because she has the same number of seconds on her piece of paper as participant number 4. This is called a "tie". In case of a "tie" the convention of survival analysis is that events occur before loss of followup.

Participant number 5 has "survived" 42 seconds. In order to account for that she will drown at a later time-point the Kaplan-Meier method distributes the probability mass of participant number 5 in equal pieces to the remaining four participant who are standing on her right hand side. Importantly, you (participant number 5) need to distribute both you own paper (which weighs 1/9) and also the piece of paper that you received from participant number 2 (which weighs $1/9\times1/7$*). When you are done, you can sit down and finally we can write the fourth line of the table.*

In order to facilitate the computations, participant number 5 can write the weight of the distributed paper pieces on the back sides of the papers that she distributes to the remaining four participants. Thus, she would write $1/9\times1/4$ on each of the 4 pieces into which she divided her own paper, and $1/9\times1/7\times1/4$ on each of the 4 pieces into which she divided the piece of paper that she received from participant number 2.

5.5. **Participant number 6.** *You have "drowned" after 51 seconds. Now, the Kaplan-Meier estimate takes into account that maybe participants number 2 and 5 have drowned until 51 seconds just this was not observed. Thus the Kaplan-Meier estimate drops by the total weight of the paper participant number 6 is holding in her hands. As soon as you (or someone else) has counted the total weight of the four pieces of paper, you may sit down.*

For the convenience of the reader, here is the computation:

$$\underbrace{1/9}_{\text{own contribution}} + \underbrace{1/9 \times 1/7}_{\text{contr. from part. 2}} + \underbrace{1/9 \times 1/4}_{\text{contr. from part. 5}} + \underbrace{1/9 \times 1/7 \times 1/4}_{\text{contr. from part. 5 via part. 2}}$$

5.6. **Participants number 7, 8 and 9.** *Participant number 7 has "survived" for 67 seconds. In order to account for that you will "drown" at a later time-point you give half of the pieces of paper to each of the remaining 2 participants. Then, you sit down.*

Participant number 8 has "survived" for 70 seconds. In order to account for that you will "drown" at a later time-point you give all the pieces of paper to the remaining participant. Then, you sit down.

Participant number 8 has "survived" for 75 seconds. Since, no one is left on the right all we can do is stop the Kaplan-Meier estimate at 75 seconds. You take all the pieces of paper with you. The weight of these papers reflect the value of the Kaplan-Meier estimate at the latest time-point where it is defined. Note that if the last participant would have "drowned" then the Kaplan-Meier estimate would have dropped all the way down to the value zero.

6. The Kaplan-Meier plot

After the action of the theater, I repeat the assumptions of the Kaplan-Meier method (see next section), and then I would usually hold a coffee break. This allows the participants to catch up and copy the data from the board. The data from the specific Kaplan-Meier theater outlined in the previous section are summarized in Table 1. Based on the data I draw the Kaplan-Meier graph on the board (see Figure 2). While doing this I discuss each step of the line, and mark the line at the time points at which subjects were lost to follow-up (right censored). It is useful to pronounce that and repeat why the step size of the Kaplan-Meier graph is increasing over time. Finally, I ask the students to look at the figure and read off the answers to the introductory questions:

- *What is the probability to survive 40 seconds?* **Answer**: *76.2%.*

TABLE 1. Kaplan-Meier analysis of the data shown in Figure 1.

Time	Number at-risk	Number of events	Number lost to followup	Survival probability (%)
0	9	0	0	100
27	9	1	0	88.9
35	8	0	1	88.9
39	7	1	0	76.2
42	6	1	1	63.5
51	4	1	0	47.6
67	3	0	1	47.6
70	2	0	1	47.6
75	1	0	1	47.6

- *What is the median survival time?* **Answer**: *51 seconds.*
- *What is the mean survival time?* **Answer**: *not estimable.*

7. Discussion

The Kaplan-Meier method (Kaplan and Meier, 1958) makes three important assumptions. The first is that everyone in the population will experience the event of interest. This assumption justifies the re-distribution of probability mass to the right algorithm which is due to Efron (1967). This assumption implies that there are no competing risks. For example, in the context of the story supporting the Kaplan-Meier theater the assumption rules out the possibility that a lifeboat comes to save "drowning" people. In the more relevant context of a medical study on the occurrence of an event, say the development of cardiovascular disease, the Kaplan-Meier assumption rules out the possibility that the patients can die in a state which is free of cardiovascular diseases. It should be clear for anyone who witnessed a Kaplan-Meier theater that the Kaplan-Meier method is biased if it is applied naively in the presence of competing risks. A subject who experiences a competing risk will not experience the event at a later time point. E.g., a participant who is saved by a lifeboat will not "drown" and a patient who died from cancer can not suffer a cardiovascular disease after death. Thus, if one would re-distribute the probability mass to the right of a subject who experienced a competing risk then this would lead to a systematically to low survival probability

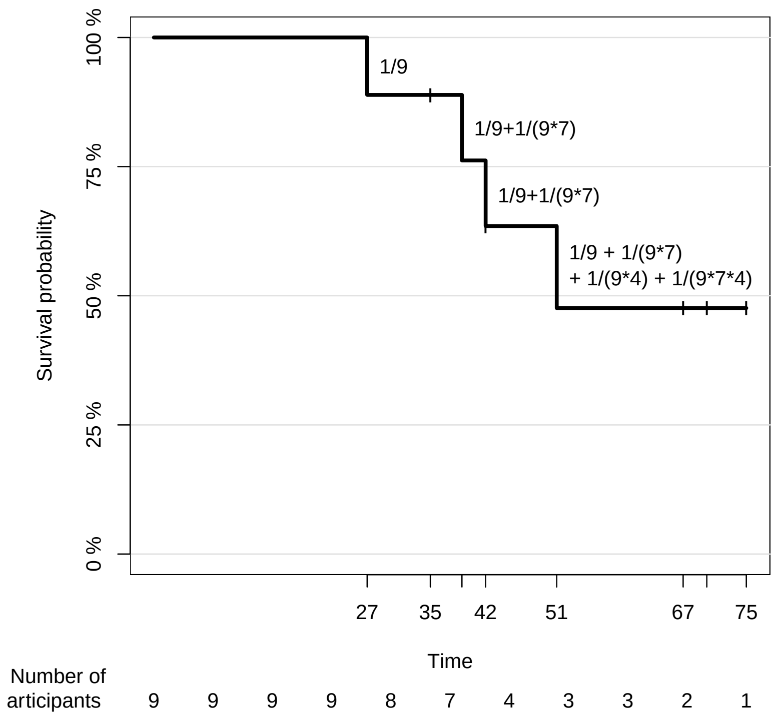

FIGURE 2. Kaplan-Meier graph based on the data from the comic of the theater shown in Figure 1.

at later time points. Unfortunately, this mistake has been made many times (see e.g. Southern et al., 2006).

The second assumption is that it makes sense to estimate the average (marginal) survival distribution in the population. This assumption is reflected during the theater where subjects who were lost to followup distributed the same amount of the probability mass to all subjects standing to the right, irrespectively of the gender, lung volume or a Viking gene of these subjects.

The third assumption is that there is no information contained in the fact and time that a person is lost to followup about the future of this person. Indeed, the idea of distributing the probability mass is strongly relying on that a subject who is event-free at the end of followup has the same future survival chances than the remaining subjects standing to the right. This assumption is often called *independent censoring* and the censoring mechanism which determines the probability that a subject is lost to followup is called *non-informative* (Andersen et al., 1993). This assumption can be relaxed, so that the censoring mechanism is allowed to depend on the covariates. In the context of the Kaplan-Meier theater this would mean that women with a high lung volume who are lost to followup are distributing their probability mass only to the remaining women who also have a high lung volume. A corresponding extension of the Efron's re-distribution to the right algorithm is described in Malani (1995).

References

P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer, New York, 1993.

B. Efron. The two-sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 4, pages 831–853, 1967.

E. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

H. M. Malani. A modification of the redistribution to the right algorithm using disease markers. 82:515–526, 1995.

D. A. Southern, P. D. Faris, R. Brant, P. D. Galbraith, C. M. Norris, M. L. Knudtson, and W. A. Ghali. Kaplan–meier methods yielded misleading results in competing risk scenarios. *Journal of clinical epidemiology*, 59(10):1110–1114, 2006.

8. Appendix A

The following items are needed to perform the Kaplan-Meier theater:

- A stopwatch (or another time tracking device)
- A bunch of red and green text markers.
- About 10 pieces of paper.
- About 10 mobile phones (or 10 other time tracking devices).
- A white or black board.

9. Appendix B

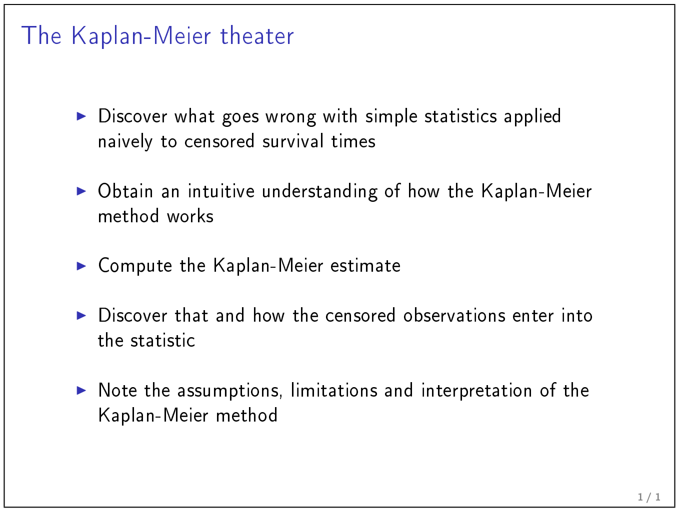

Figure 3. Beamer slide introducing the Kaplan-Meier theater

10. Appendix C: R code

```
source("http://staff.pubhealth.ku.dk/~tag/download/redist.R")
f <-redist(time=c(27,35,39,42,42,51,67,70,75),
      status=c(1,0,1,1,0,1,0,0,0))
```

```
Kaplan-Meier estimate via re-distribution to the right algorithm:


Subject 1:
---------------------------
Survival before = 100%
Event at time = 27
Contribution to Kaplan-Meier estimate:


                 fractions value
own contribution 1/9       0.1111
                 sum       0.1111


Survival after = 100% - (1/9)
```

```
              = 100% - 11.11% = 88.89%


Subject 2:
---------------------------
Survival before = 88.89%
No event until time = 35
Re-distribute mass 0.11 to remaining 7 subjects
Survival after = 88.89%


Subject 3:
---------------------------
Survival before = 88.89%
Event at time = 39
Contribution to Kaplan-Meier estimate:

                 fractions value
own contribution 1/9       0.11111
from subject 2   1/9*1/7   0.01587
                 sum       0.1270

Survival after = 88.89% - (1/9 + 1/9*1/7)
              = 88.89% - 12.7% = 76.19%


Subject 4:
---------------------------
Survival before = 76.19%
Event at time = 42
Contribution to Kaplan-Meier estimate:

                 fractions value
own contribution 1/9       0.11111
from subject 2   1/9*1/7   0.01587
                 sum       0.1270

Survival after = 76.19% - (1/9 + 1/9*1/7)
```

```
              = 76.19% - 12.7% = 63.49%


Subject 5:

---------------------------

Survival before = 63.49%

No event until time = 42

Re-distribute mass 0.13 to remaining 4 subjects

Survival after = 63.49%


Subject 6:

---------------------------

Survival before = 63.49%

Event at time = 51

Contribution to Kaplan-Meier estimate:


                 fractions   value

own contribution 1/9         0.111111

from subject 2   1/9*1/7     0.015873

                 1/9*1/4     0.027778

from subject 5   1/9*1/7*1/4 0.003968

                 sum         0.1587


Survival after = 63.49% - (1/9 + 1/9*1/7 + 1/9*1/4 + 1/9*1/7*1/4)

              = 63.49% - 15.87% = 47.62%


Subject 7:

---------------------------

Survival before = 47.62%

No event until time = 67

Re-distribute mass 0.16 to remaining 2 subjects

Survival after = 47.62%


Subject 8:

---------------------------

Survival before = 47.62%
```

```
No event until time = 70

Re-distribute mass 0.24 to remaining 1 subject

Survival after = 47.62%


Subject 9:

---------------------------

Survival before = 47.62%

Last subject lost to follow-up event free at time = 75

Survival after = 47.62%


Summary table:
```

	time	n.risk	n.event	n.lost	surv
1	0	9	0	0	100.00000
2	27	9	1	0	88.88889
3	35	8	0	1	88.88889
4	39	7	1	0	76.19048
5	42	6	1	1	63.49206
6	51	4	1	0	47.61905
7	67	3	0	1	47.61905
8	70	2	0	1	47.61905
9	75	1	0	1	47.61905

Department of Biostatistics, University of Copenhagen, Østerfarimagsgade 5b, 1014 Copenhagen, Denmark

E-mail address: tag@biostat.ku.dk

Research Reports available from Department of Biostatistics

http://www.pubhealth.ku.dk/bs/publikationer

Department of Biostatistics
University of Copenhagen
Øster Farimagsgade 5
P.O. Box 2099
1014 Copenhagen K
Denmark

12/01 Keiding, N., Hansen, O.K.H., Sørensen, D.N. & Slama, R. The current duration approach to estimating time to pregnancy.

12/02 Andersen, P.K. A note on the decomposition of number of life years lost according to causes of death.

12/03 Martinussen, T. & Vansteelandt, S. A note on collapsibility and confounding bias in Cox and Aalen regression models.

12/04 Martinussen, T. & Pipper, C.B. Estimation of Odds of Concordance based on the Aalen additive model.

12/05 Martinussen, T. & Pipper, C.B. Estimation of Causal Odds of Concordance based on the Aalen additive model.

12/06 Binder, N., Gerds, T.A. & Andersen, P.K. Pseudo-observations for competing risks with covariate dependent censoring.

12/07 Keiding, N. & Clayton, D. Standardization and control for confounding in observational studies: a historical perspective.

12/08 Hilden, J. & Gerds, T.A. Evaluating the impact of novel biomarkers: Do not rely on IDI and NRI.

12/09 Gerster, M., Madsen, M. & Andersen, P.K. Matched Survival Data in a Co-twin Control Design.

12/10 Scheike, T.H., Holst, K.K. & Hjelmborg, J.B. Estimating heritability for cause specific mortality based on twin studies.

12/11 Jensen, A.K.G., Gerds, T.A., Weeke, P., Torp-Pedersen C. & Andersen, P.K. A note on the validity of the case-time-control design for autocorrelated exposure histories.

13/01 Forman, J.L. & Sørensen M. A new approach to multi-modal diffusions with applications to protein folding.

13/02 Mogensen, U.B., Hansen, M., Bjerrum, J.T., Coskun, M., Nielsen, O.H., Olsen, J. & Gerds, T.A. Microarray based classification of inflammatory bowel disease: A comparison of modelling tools and classification scales.

13/03 Wolbers, M., Koller, M.T., Witteman, J.C.M. & Gerds, T.A. Concordance for prognostic models with competing risks.

13/04 Andersen, P.K., Canudas-Romo, V. & Keiding, N. Cause-specific measures of life lost.

13/05 Keiding, N. Event history analysis.

13/06 Kreiner, S. & Nielsen, T. Item analysis in DIGRAM 3.04 - Part I: Guided tours.

13/07 Hansen, S., Andersen, P.K. & Parner, E. Events per variable for risk differences and relative risks using pseudo-observations.

13/08 Rosthøj, S., Henderson, R. & Barrett, J.K. Determination of optimal dynamic treatment strategies from incomplete data structures.

13/09 Gerds, T.A., Andersen, P.K. & Kattan M.W. Calibration plots for risk prediction models in the presence of competing risks.

13/10 Scheike, T.H., Holst, K.K. & Hjelmborg J.B. Estimating twin concordance for bivariate competing risks twin data.

13/11 Olsbjerg, M. & Christensen K.B. SAS macro for marginal maximum likelihood estimation in longitudinal polytomous Rasch models.

13/12 Touraine, C., Gerds, T.A. & Joly, Pierre. The SmoothHazard package for R: Fitting regression models to interval-censored observations of illness-death models.

14/01 Ambrogi, F. & Andersen, P.K. Predicting Smooth Survival Curves through Pseudo-Values.

14/02 Olsbjerg, M. & Christensen, K.B. Modeling local dependence in longitudinal IRT models.

14/03 Olsbjerg, M. & Christensen, K.B. LIRT: SAS macros for longitudinal IRT models.

14/04 Jacobsen, M. & Martinussen, T. A note on the large sample properties of estimators based on generalized linear models for correlated pseudo-observations.

14/05 De Neve, J. & Gerds, T. A note on the interpretation of the Cox regression model.

14/06 Gerds, TA. The Kaplan-Meier theater.