

# **Biostatistics III: Survival analysis for epidemiologists**

Anna Johansson

Department of Medical Epidemiology and Biostatistics (MEB)

Karolinska Institutet

Stockholm, Sweden

<http://www.biostat3.net/>

Karolinska Institutet

Feb 3–12, 2025

<https://doctoralcourses.application.ki.se/fubasextern/info?kurs=C8F3142>

## Friendly faces

- Primary teachers:  
Anna Johansson (anna.johansson@ki.se),  
Elisavet (Betty) Syriopoulou (elisavet.syriopoulou@ki.se)
- Course director: Anna Johansson (anna.johansson@ki.se)
- Course administrator: Gunilla Nilsson Roos (gunilla.nilsson.roos@ki.se)
- Computer lab tutors: Therese Andersson, Jiayao Lei, Elisavet Syriopoulou, Anna Johansson.

# Teaching format

- Generally, the morning session 09:00-12:00 comprises 2 hours of lectures followed by a 1 hour computer lab, while the afternoon session 13:00-16:00 comprises a 1 hour lecture and 2 hours of computer lab.
- The computer labs are important for your learning and active participation will help you achieve the learning outcomes.
- We encourage you to work in pairs during the computer labs. We also encourage you to discuss with the teaching assistants.
- We have constructed exercises and provided solutions to most exercises. We will suggest appropriate exercises for each afternoon.
- There are a lot of exercises, don't worry if you don't have time to finish all of them. We will let you know which exercises that are the key exercises.

- Course participants have a wide range of backgrounds and diverse interests. It is hoped that the lab sessions will provide time for you to study or ask questions about topics of special interest.
- The lecture notes are very comprehensive, and contain topics which we will only cover briefly. Some slides will not be covered during the lectures, but are included for completeness.
- On days 2, 3, 4 and 5 there will be a lab review in the morning, where key concepts from the previous day will be reiterated and further developed.
- Review questions will also be sent to you by email after each day, where you can self-assess your own learning. Responses are anonymous.

## Group work

- Group work on day 4 and 5, each group will be assigned an article to read and discuss.
- You can see which group you belong to in the list of participants.
- The article(s) can be found on the course web site.
- Read your allocated article in advance, and discuss the article in your group on day 4 of the course.
- Prepare a short presentation (10 minutes) that you will give to the other groups on Day 5.
- More information will be given later during the course.

# Computer labs

- We will be using Stata.
- We have created a zip-file, containing all files needed for the labs, that can be downloaded from the course website.
- Further details will come on working with the labs.
- We encourage you to work in pairs/small groups and discuss the computer exercises.

## For SAS users

- Although we will be using Stata during the course, we have made the data sets available in SAS format and you are welcome to attempt the exercises using SAS.
- We have written brief notes on how the methods described during the course can be implemented in SAS.
- We have provided SAS code for reproducing some of the key exercises.
- Please keep in mind that if you choose to work with SAS we expect you to bring your own laptop (with SAS installed) and to be familiar with SAS so that the teaching assistants can focus on helping with the statistical concepts rather than statistical programming.

## For R users

- Moreover, you are welcome to attempt the exercises using R.
- There is an R version of this course offered in the autumn each year. The material from that course is available online.
- Please keep in mind that if you choose to work with R we expect you to bring your own laptop (with R installed) and to be familiar with R so that the teaching assistants can focus on helping with the statistical concepts rather than statistical programming.



## Textbook

- We have not assigned any compulsory texts since experience has shown that course participants have widely varying preferences.
- We will provide extensive course notes and many participants do not find a great need for additional texts. A large number of textbooks are available and we suggest students interested in additional reading identify a textbook at a technical level suitable for them. Many general textbooks in medical statistics contain a chapter on survival analysis.
- Very few books are targeted at epidemiologists (e.g., you won't find Poisson regression mentioned in many books).
- One suggestion is 'Statistical Methods in Cancer Research: Volume II - The Design and Analysis of Cohort Studies' by Breslow and Day [2] although it is rather advanced. [full text available free from <https://publications.iarc.fr/Book-And-Report-Series/Iarc-Scientific-Publications/Statistical-Methods-In-Cancer-Research-Volume-II-The-Design-And-Anal>]

- 'An Introduction to Survival Analysis Using Stata' [4] is highly recommended for Stata users. Many parts, however, assume a solid grasp of mathematical statistics.
- The SAS 'books by user' [3, 1] are recommended for SAS users.

## Learning outcomes

- The course plan is the formal document upon which the course (and the examination) is based (see <https://doctoralcourses.application.ki.se/fubasextern/info?kurs=C8F3142>).
- KI uses outcome-based learning. The learning outcomes are listed in the course plan and reproduced below.

After successfully completing this course you should be able to:

- Propose a suitable statistical model for assessing a specific research hypothesis using data from a cohort study, fit the model using standard statistical software, evaluate the fit of the model and interpret the results.
- Explain the similarities and differences between Cox regression and Poisson regression.
- Discuss the concept of timescales in statistical models for time-to-event data, control for different timescales using standard statistical software, and argue for an appropriate timescale for a given research hypothesis.
- Discuss the concept of confounding in epidemiological studies and control/adjust for confounding using statistical models.

- Apply and interpret appropriate statistical models for studying effect modification and be able to reparameterise a statistical model to estimate appropriate contrasts.
- Critically evaluate the methodological aspects (design and analysis) of a scientific article reporting a cohort study.

## Examination

The course grade is based solely on a take-home examination. The exam is the only mandatory part of the course.

- The examination is individual-based: **you are not allowed to cooperate with anyone**, although you are encouraged to consult the available literature. The teachers will use Urkund in order to assess potential plagiarism.
- The examination will be available at 12:00 on Wednesday 12 February 2025 from <https://biostat3.net/download/?dir=stata>.
- The **examination is due by 17:00 on Wednesday 19 February 2025**.
- Students who do not obtain a passing grade in the first examination will be offered a second possibility of examination within 2 months of the final day of the course.
- Do not write answers by hand: please use Word,  $\text{\LaTeX}$  or a similar format for your examination report.
- Motivate all answers and show all calculations in your examination report, but write as brief an answer as possible without loss of clarity. Define any notation that you use for equations. The examination report should be written in English.

- You will not be expected to write computer code, but you will be expected to interpret output from Stata.
- Email the examination report containing the answers **as a pdf file** to `gunilla.nilsson.roos@ki.se`. **Write your name in the email, but do not write your name in the document containing the answers.**

## References

- [1] Paul D. Allison. *Survival Analysis Using the SAS System: A Practical Guide*. Cary, NC: SAS Institute Inc., 1996.
- [2] N. E. Breslow and N. E. Day. *Statistical Methods in Cancer Research: Volume II - The Design and Analysis of Cohort Studies*. IARC Scientific Publications No. 82. Lyon: IARC, 1987.
- [3] Alan Cantor. *SAS Survival Analysis Techniques for Medical Research*. BBU Press, second edition, 2003.
- [4] M. A. Cleves, W. W. Gould, R. G. Gutierrez, and Y. Marchenko. *An Introduction to Survival Analysis Using Stata*. Stata Press, third edition, 2010.