# Practical aspects of large register studies: Multiple states and multiple timescales: The diabetes and cancer study.

**Bendix Carstensen**
Senior Statistician, Steno Diabetes Center, Denmark
bxc@steno.dk www.biostat.ku.dk/~bxc

This talk will in detail describe the data sources and statistical and computational methods underlying a study of cancer occurrence among Danish diabetes patients.

**Background and aims:** A large number of studies of cancer occurrence among diabetes patients have been conducted. We used a register-based follow-up of the entire Danish population in the period 1995–2007 to determine how cancer incidence rates vary diabetes duration and duration of insulin therapy.

**Methods:** We linked the Danish National Diabetes Register and the Danish Cancer Register and followed diabetes patients for the occurrence of cancer. For persons diagnosed with diabetes after 1995 we have reliable data on DM duration and the time since first insulin use. For these patients we modelled the effect of DM duration and insulin duration on the cancer occurrence rates, and describe the cancer incidence rate ratio relative to the non-diabetic part of the population.

I will describe how we prepared analysis tables from the register data, and how we estimated rates and rate-ratios for transitions to cancer and death from the well and diabetic parts of the population respectively, and how we plan to extend this using more detailed drug purchase information.

Furthermore, I will discuss probability statements such as "the cumulative risk of diabetes patient getting cancer within the next 10 years", and what assumptions are needed and desired to attach meaning to them. Subsequently I will discuss how to use these assumptions to compute the cumulative risk based on a multistate model for the follow-up data.

**Results:** For all cancers combined we saw a clear decrease in RR by DM duration stabilizing at about 1 after 2–3years. A similar decrease was seen by duration of insulin therapy, however stabilizing at a RR about 1.2.

The differences between groups (Well, DM, DM/Ins) look considerably less spectacular when using cumulative risks than when using rate-ratios to describe them.

**Conclusion:** This study is the largest of diabetes and cancer incidence so far, and the only one to model the duration effects of both diabetes and insulin treatment. With register information on the entire population it is possible to model transition rates between several states (Well, DM, DM/Ins, Cancer, Dead,...) and how they depend on several timescales (age, calendar time, duration,...). With a parametric representation of the results it is fairly straight-forward to compute relevant probabilities, taking all competing risks into account, but with multiple timescales used to describe the rates, it requires some book-keeping skills to get it right.

## References

B Carstensen, JK Kristensen, P Ottosen, and K Borch-Johnsen: The Danish National Diabetes Register: Trends in incidence, prevalence and mortality. Diabetologia, 51: 2187-2196, 2008.

L Wiederoff, G Gridley et al.: Cancer incidence in a population based cohort of patients hospitalized with Diabetes Mellitus in Denmark. JNCI, 89: 1360-1365, 1997.

SS Coughlin, EE Calle, et al.: Diabetes Mellitus as a predictor of cancer mortality in a large cohort of US adults. Am J Epidemiol 2004, 159: 1160-1167

## About the speaker

Bendix Carstensen has a master's degree in Mathematical Statistics from the University of Copenhagen. He has worked as a research statistician with the Danish Cancer Registry in 1984–95, and as a senior statistician at the Danish Zoonosis Centre 1995–99 and since 1999 at Steno Diabetes Center, where he is with the epidemiology group. His main area of work is in register based follow-up studies. He has an affiliation as external lecturer at the department of Biostatistics at the University of Copenhagen, where he teaches on courses for medical PhD-students. He gives courses internationally both in the area of follow-up studies (medical demography) and on the analysis of clinical method comparison studies. He is the maintainer of the Epi package for R, and the author and maintainer of the (yet unpublished) MethComp package for R.

Further details available at his web-site www.biostat.ku.dk/~bxc.