# Penalised competing risks regression

*Federico Ambrogi*[1], *Thomas Scheike*[2]

[1]federico.ambrogi@unimi.it, Department of Clinical Sciences and Community Health, University of Milan

[2]thsc@sund.ku.dk, Department of Biostatistics, University of Copenhagen

High dimensional data analysis is an important topic in many research fields. For example, biomedical research generates increasing amount of data to characterise patients bio-profiles (e.g. from genomic high-throughput assay, imaging, physiological measurements, laboratory tests, patient monitoring, etc.). Variable selection is a long-established problem in statistical research and is every day more and more important. In the last decades many forms of penalised regression have been developed, as a modern form of variable selection, to cope with high and ultra-high dimensional settings. The increasing complexity in the characterisation of patients bio-profiles, is added to the complexity related to the prolonged follow-up of patients with the registration of the occurrence of possible adverse events, that may offer useful insight in disease dynamic and in identifying subset of patients with worse prognosis and better response to the therapy. Although in the last years the number of contributions for coping with high and ultra-high dimensional data in standard survival analysis have increased, the research regarding competing risks is less developed. The aim of this work is to consider how to do penalized regression when considering the crude cumulative incidence. The direct binomial regression model developed by Scheike, Zhang and Gerds (Biometrika, 2008) is reformulated in a penalised framework to possibly fit a sparse regression model. The proposed approach is easily implementable using existing high performance software to do either ridge, or lasso or elastic net penalization. Results from simulation studies are presented together with an application to genomic data when the endpoint is progression free survival.

**Keywords:** Competing risks, penalized estimating equations, high dimensional data