Summary

Family studies examining disease occurrence in related individuals typically aim at determining whether the disease of interest aggregates in families or not. That is, whether the risk of disease is greater in individuals with an affected relative compared to that of the general population. In addition to having the potential to offer insight into disease etiology and pathogenesis, studies of familial aggregation of diseases can be used to assess familial risk, which is highly relevant for preventive purposes.

This PhD project was initiated by the Danish Cancer Society Research Center motivated by a desire for studying familial aggregation of cancer utilising the information available in the Danish registers. The extent and quality of the information that can be obtained from the registers enable elaborate statistical modelling, recognising the register based family data as multivariate competing risks data.

The work presented in this thesis aims at contributing to the statistical methodology for analysis of family studies in the form of multivariate competing risks data. Focus is on methods for analysis of the cumulative incidence of disease including assessment of within-family dependence. Two methods are presented. The first method is an adaptation of an existing method from the twin setting to the case-control-family setting. It provides a useful and simple measure of the cumulative incidence of disease and the within-family dependence at a given age. The second method is a novel method. It is more dynamic than the first method and models the cumulative incidence of disease in a given time period while allowing for within-family dependence with regard to both risk and timing. That is, the method allows for family heterogeneity with regard to risk level (in terms of absolute risk) and age distribution. Accordingly, the second method fully utilises the information available in the data while the first method does not. Yet, the second method imposes a fairly elaborate structure on the data and is more computationally intensive. As a consequence of the computational intensity involved in parameter estimation, it is recommended to analyse a case-cohort sample of families. Both of the presented methods are assessed using simulation studies and applied in analysis of Danish register-based family data on different cancer types.