# Selection models for the efficient design of family studies

**Yujie Zhong**

*School of Statistics and Management, Shanghai University of Finance and Economics*
*E-mail: zhong.yujie@mail.shufe.edu.cn*

**Abstract:** Family studies on the genetic basis for disease typically recruit a random sample individuals from a disease registry with their respective relatives to obtain genetic and phenotypic data. Individuals in the disease registry, however, may be viewed as comprising a phase I sample in a two-phase study. We propose the use of selection models to exploit this phase I data and more efficiently identify a phase II sample of families for study. Copula models are adopted characterize the within family dependence in disease onset times but the likelihood is constructed based on current status data in order to rely on the physician diagnosis when family members are examined. Empirical studies are conducted to demonstrate the efficiency gains that can be realized over simple random or balanced sampling schemes and to study the effect of misspecifying the design parameters.