# Design and Analysis for Two-Phase studies with Survival Data

Xu Cao

Preprint no. 2024-05

**Abstract**

Large cohort studies under simple random sampling could be prohibitive to conduct with a limited budget for epidemiological studies seeking to relate a failure time to some exposure variables that are expensive to obtain. In this case, two-phase studies are desirable. In this dissertation, Chapter 2 discusses the semiparametric inference for a two-phase failure-time-auxiliary-dependent sampling (FADS) design that allows the probability of obtaining the expensive exposures to depend on both the failure time and cheaply available auxiliary variables.

In the studies of time-to-event outcomes, it often happens that a fraction of subjects will never experience the event of interest, and these subjects are said to be cured. When survival data include a fraction of long-term survivors, censored observations encompass both uncured individuals and cured individuals. Consequently, the cure status is unknown, and survival data comprise a mixture of cured and uncured individuals that can't be distinguished beforehand. Chapter 3 of this dissertation considers the generalized case-cohort design for studies with a cure fraction. We propose a two-step estimation procedure under the semiparametric transformation mixture cure models. We show that the proposed estimator is consistent and asymptotically normal, regardless of whether the working model is correctly specified or not.

Department of Mathematics and Statistics, UNC Charlotte, Charlotte, NC 28223