

# Efficient linear regression for time-to-event data under length-biased sampling

Michael R. Kosorok<sup>1\*</sup>, Rajat Mukherjee<sup>2</sup>, and Jason P. Fine<sup>3</sup>

<sup>1</sup>University of North Carolina-Chapel Hill

<sup>2</sup>Public Health Foundation of India

<sup>3</sup>University of Wisconsin-Madison

*E-Mail:* kosorok@unc.edu

*Abstract:* In this talk, we develop both efficient and inefficient methods of inference for accelerated failure time regression under length-biased sampling and right censoring. This kind of data arises in both prevalent cohort and current duration study designs. An HIV sero-prevalence study can have a prevalent cohort design if the status but not the time of HIV infection is known at a fixed cross-sectional sampling time, after which patients are followed until they acquire AIDS or the study is censored. Certain time-to-pregnancy studies have a current duration design if time of initiation of unprotected intercourse is known but time to successful pregnancy is censored at a fixed cross-sectional sampling time. In both cases, there exists a bias in the sampling of the event time (the time between the initiating and terminal events) because the probability of observing the patient is proportional to the magnitude of the event time. This biased sampling structure induces a dependence between the covariates and the event time. We obtain the surprising result that this dependency can actually be ignored without decreasing the efficiency of the regression parameter estimator. We apply the proposed procedures to data from a French Fecundity study and present several simulation results for illustration.