Statistica Sinica **36** (2026), 1-19
doi:https://doi.org/10.5705/ss.202024.0099

PROBIT TIME-TO-EVENT REGRESSION FOR MISCLASSIFIED GROUP TESTING DATA

Lijun Fang[1], Tao Hu[1], Shuwei Li*[2], Lianming Wang[3], Christopher S. McMahan[4] and Joshua M. Tebbs[3]

[1]*Capital Normal University,* [2]*Guangzhou University,* [3]*University of South Carolina and* [4]*Clemson University*

Abstract: Group testing has been used extensively to reduce screening costs in epidemiological studies involving low-prevalence diseases. This testing strategy involves combining specimens (e.g., blood, urine, or swabs) from several individuals to form a pool and then testing the pooled specimen for infection. When the endpoint of interest is a time-to-event outcome, for example, the time until infection or disease, and pools are measured only once, the resulting data are called group-tested current status data (Petito and Jewell, 2016). In this paper, we propose a new type of regression analysis for these data using a semiparametric probit model, an alternative to the proportional hazards model in survival analysis. A sieve maximum likelihood estimation approach is developed that approximates the model's nonparametric nuisance function by using logarithmic monotone splines, and an efficient expectation-maximization algorithm is proposed. Asymptotic properties of the resulting estimators are investigated by using empirical process techniques and sieve estimation theory. Numerical results from simulation studies suggest our estimation methods perform nominally, even when pools are possibly misclassified due to assay error, and can outperform individual testing when the number of assays (tests) is fixed. We illustrate our work by estimating a time-to-event regression model for Chlamydial infection using group testing data from a large public health laboratory in Iowa.

Key words and phrases: Current status data, EM algorithm, maximum likelihood estimation, pooled testing, sieve estimation.

1. Introduction

Group testing was originally proposed by Dorfman (1943) to screen members of the United States military for syphilis during World War II. This strategy works by collecting a biological specimen (e.g., blood, urine, or swab) from different individuals and pooling the specimens together. The pooled specimen is then tested for infection or disease. If a pooled specimen tests negatively, then all individuals in the pool are declared to be negative at the expense of a single test. If a pooled specimen tests positively, individuals within it can be retested one at a time or in some other predetermined manner. When the

*Corresponding author. E-mail: seslishuw@gzhu.edu.cn