

## Editorial

Mei-Cheng Wang<sup>1</sup> · Chiung-Yu Huang<sup>1</sup>

Published online: 16 January 2017

© Springer Science+Business Media New York 2017

Various outcome-dependent sampling (ODS) designs have been employed under different settings for data collection. These various designs share a common feature - the sampling mechanism depends on the outcome variable or vector, therefore induces bias in the observed data. Significant research areas of outcome-dependent sampling include case-control design with extensions to continuous or failure time outcomes, truncation and sampling-bias models, and non-ignorable missing data. In these models or studies, the outcome-dependent sampling is adopted either by design or because of practical constraints or needs for data collection. In these situations, standard statistical approaches are typically inappropriate as they do not account for sampling bias properly and are likely to result in biased analysis and interpretation. In statistical literature there exists a collection of models and methods that account for the outcome-dependent sampling in the analytical steps, though these approaches (or even the sampling problems) may or may not be recognized by practitioners and data analysts.

This special issue invited statisticians with expertise in the field to either review the existing statistical models/methods or address new outcome-dependent sampling problems. Ding et al. (2017) provided a comprehensive review of the recent progress in ODS with univariate or multivariate failure time outcomes. The authors reviewed the case-cohort design (original, generalized, stratified and modified), the general failure-time ODS design, as well as other biased-sampling designs, such as length-biased sampling and interval sampling, with failure time data. Estimation procedures for

---

✉ Mei-Cheng Wang  
mcwang@jhu.edu

Chiung-Yu Huang  
cyhuang@jhu.edu

<sup>1</sup> Johns Hopkins University, Baltimore, MD, USA

commonly used models (proportional hazards model, additive rates model, transformation models, etc.) under different designs were presented and their extensions to multivariate failure times were discussed.

ODS design can be viewed as a special case of the general two-phase sampling scheme. Fu and Gilbert (2017) considered joint modeling of biomarker trajectories and time to event, where longitudinal measurements of biomarkers are obtained under a two-phase sampling design, that is, biomarkers are assessed only on a selected group of individuals and the selection probability may be covariate and/or outcome-dependent. Guan and Qin (2017) proposed empirical likelihood-based methods for estimating the parameters in the selection probability model in the presence of nonignorable missing data. The resulting empirical weights on complete cases are also used to estimate the mean of the response.

It is well known that survival data collected in prevalent cohort studies are subject to left-truncation. As a result, the sampling scheme favors individuals who survive longer and thus is outcome dependent. Vakulenko-Lagun, Mandel, and Goldberg (2017) considered nonparametric estimation of bivariate sojourn time distribution under left truncation and right censoring. Huang and Chen (2017) considered analysis of bivariate sojourn times when the first sojourn time is unobservable. In the special case where the incidence of disease follows a Poisson process, the probability of a survival time being sampled is proportional to its length and thus is length-based. Shen, Ning and Qin (2017) provided a comprehensive review of recent development in nonparametric and semiparametric methods for length-biased data subject to right censoring. For maximum likelihood estimation, Chan (2017) considered alternative numerical methods to accelerate the convergence of Vardi's EM algorithm (Vardi 1989).

All the papers were carefully and thoughtfully written, and are definitely worth reading.