

# VARIANCE INFLATION DUE TO CENSORING IN SURVIVAL PROBABILITY ESTIMATES

Andreas Gustavsson<sup>1</sup>, Szilárd Nemes<sup>2</sup>, Zaid Taib<sup>3</sup>

<sup>1</sup>BioPharma Early Biometrics & Statistical Innovation, Data Science & AI,  
BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden  
andreas.gustavsson@astrazeneca.com

<sup>2</sup>BioPharma Early Biometrics & Statistical Innovation, Data Science & AI,  
BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

<sup>3</sup>Atlasbiostat, Gothenburg, Sweden

## Abstract

The main attractiveness of the Kaplan-Meier curve is the simple but elegant way of handling incomplete data, a common feature of survival analysis. This partial loss of information, due to censoring, ought to inflate the variance of the estimated probabilities. Cantor (2001) was among the first to study the issue of variance inflation of Kaplan-Meier curves under censoring. Before that, functionals of the Kaplan-Meier curve, namely the restricted mean survival time or life expectancy had received a lot of interest, regarding the issue of the variance under censoring.

In this presentation, we present an analytical expression for the variance inflation due to incomplete observations. This estimator allows the estimation of the expected variance under different assumptions for survival and censoring times. The main advantage of the proposed formula is the relative ease of use. For the exponential distribution, there exists a closed form solution. For other distributions, or combination of distributions, it can be applied based on available computational methods. Further properties of the proposed estimator are illustrated using simulation studies and its applicability is highlighted based on published data.

Furthermore, we attempt to establish a direct connection between variance inflation and sample size calculations, a subject of particular interest in planning of clinical trials. We believe that our estimator is useful in that area as it allows making a quick assessment of the excess variability caused by censoring and loss to follow-up.