

Comparison of Statistical Approaches for Interval-Censored Data: Analysis of data from an HIV-negative MSM Cohort

*Inês Carvalho*¹, *Leandro Duarte*², *Carla Moreira*^{2,4,5}, *Luís Machado*², *Ana Paula Amorim*²,
*Miguel Rocha*³, *Paula Meireles*^{4,5}

¹b13232@math.uminho.pt, Centro de Matemática da Universidade do Minho, Universidade do Minho,
4800-058 Guimarães, Portugal

² Centro de Matemática da Universidade do Minho, Universidade do Minho, 4800-058 Guimarães,
Portugal

³ Grupo de Activistas em Tratamentos, Lisboa, Portugal

⁴ EPIUnit - Instituto de Saúde Pública, Universidade do Porto, Rua das Taipas, n° 135, 4050-600 Porto,
Portugal

⁵ Laboratório para a Investigação Integrativa e Translacional em Saúde Populacional (ITR),
Universidade do Porto, Rua das Taipas, n° 135, 4050-600 Porto, Portugal

Survival analysis differentiates from other domains in statistics as the data can be censored, meaning that the event during a follow-up period, such as death or occurrence of a disease, is incomplete. In standard survival analysis, the event of interest might be observed exactly or is right-censored, that is, individuals will be event-free throughout follow-up, but the event may occur after the last observation time. In other situations, however, the times of the events of interest are known only to have occurred within a time interval from the last examination without the event to the first examination after the event has occurred. These data are known to be interval-censored. Although interval-censored data requires specific methods, those developed for right-censored data are still standard practice. A common ad-hoc approach for dealing with interval-censored data is to assume that the event has occurred at the end, midpoint or beginning of each interval and then apply the usual methods for standard time-to-event data. Nevertheless, this approach can lead to misguided inferences and, particularly, underestimate the standard errors of the estimated parameters. The goal is to compare these approaches with those that do not assume such imputation and that are based on the maximum likelihood estimation (MLE), the expectation-maximization (EM) algorithm, and the Turnbull estimator. All approaches will be compared using data from the Lisbon Cohort of HIV-negative of Men Who Have Sex With Men (MSM).

Keywords: Estimation of Survival, Interval Censoring, Survival Analysis.