



SADDLEPOINT MOSAICS™ FACT SHEET

SaddlePoint-Mosaics is a multi-core software pipeline providing a Bayesian multi-risk survival analysis for data from heterogeneous medical and epidemiological cohorts and those under the influence of competing risks.

1. General description

Version	SaddlePoint-Mosaics 1.1.0
Platforms	most versions of Windows Unix/Linux and MacOS (tested on Windows 7, 8, 10, Ubuntu and MacOS Sierra)
Data	n samples representing d covariates (x_1, \dots, x_d) and corresponding time-to-event survival outcomes (t, r) (event time and type)
Deliverables	automated regression for candidate multi-risk latent class models optimal model determination using Bayesian model selection insight into cohort heterogeneity and any differences between sub-groups insight into presence or absence of heterogeneity-induced informative censoring risk- and class-specific hazard ratio p-value and 95% confidence intervals estimation retrospective class allocation of samples extensive results and dataset summary reporting (automated)

2. Detailed functionality

Operation	multi-threading with user controlled maximum number of threads command-line user interface protocol for adjusting regression due to informative missingness of covariates.
Data types	time-to-event survival data (with multiple risks and or censoring) discrete ordinal outcomes data (following mapping to time-to-event representation)
Dataset summary report generator	automated report generation for user-friendly overview of cohort including: tables of descriptive statistics of covariates and outcomes covariates distributions Pearson correlations across covariates Pearson correlations between covariates and time-to-event covariate-conditioned Kaplan-Meier curves
Data preprocessing	SaddlePoint-compatible dataset creation from comma-separated files: automated metadata generation user-selected covariate inclusion/exclusion user-selected removal of patients with missing data user-selected inclusion of auxiliary missing covariate data variables (allows regression to account for informative missingness) automated binary expansion of categorical covariates with user-selected baseline



for regression (to avoid multi-collinearity issues).
linear normalization of covariates (to zero average and unit variance)

Regression pipelines

latent class models with different level of heterogeneity
model parameters include class weightings covariate associations and frailties
model baseline hazard rates estimated using a spline construction
Bayesian priors on all estimated parameters
automated regression for all candidate models
automated Bayesian model selection following regression stage
models ranked according to relative probability given cohort's survival data
Aikake and Bayesian Information Criterion scores are also calculated
retrospective class-assignment probabilities based on individual's survival data
automated generation of stratified class-specific datasets (useful for further analysis)

Outputs

dataset summary reporting (automated) for extensive overview of cohort
candidate latent class model ranking details:

Bayesian model selection for optimal model to describe cohort without overfitting.
model summary reports for insight into cohort heterogeneity and informative censoring:
risk- and class-specific hazard ratio p-value and 95% confidence intervals estimation
risk- and class-specific baseline hazard rate plots
risk- and class-specific crude and marginal survival curves

Pearson correlations between retrospective class assignment probability and covariate.
retrospectively assigned class-specific datasets (for subsequent analyses)