Survival-Trajectory Phenotypes from Individualized Survival Curves with unsurv

A methodological note on unsupervised phenotyping from predicted survival curves.

Imad EL BADISY

2026-05-23

Table of contents

Survival-trajectory phenotyping clusters patients by the shape of their predicted survival curves. The input is not the baseline covariate matrix X, but a survival probability matrix

$$S = \left[S_i(t_j)\right]_{i=1,\ldots,n;\; j=1,\ldots,q},$$

where each row is an individualized predicted curve. A distance between two patients can be defined on the curve scale, for example an L_2 distance

$$d(i, i') = \left\{ \sum_{j=1}^{q} w_j \left[S_i(t_j) - S_{i'}(t_j)\right]^2 \right\}^{1/2},$$

with optional time weights w_j. Clustering on S groups patients by predicted survival behavior, not by isolated covariate values. This is useful when several different covariate profiles imply similar risk trajectories.[1]

The resulting clusters are descriptive phenotypes. They should be interpreted as summaries of model-predicted prognosis, not as causal treatment effects.

[1] `unsurv` package documentation, CRAN: https://CRAN.R-project.org/package=unsurv.

Example

The `unsurv` package implements unsupervised clustering of individualized survival curves.[2]

```
library(unsurv)

set.seed(1)
times <- seq(0, 36, by = 3)
hazards <- rep(c(0.02, 0.05, 0.09), each = 10)

S <- outer(
  hazards,
  times,
  function(h, t) exp(-h * t)
)

fit <- unsurv(S, times, K = 3, seed = 1)
fit
```

```
unsurv (PAM) fit
  K:3
  distance:L2 silhouette_mean:0.974
  n:30 Q:13
```

The fitted object stores the cluster assignment, medoid curves, the selected distance, and a silhouette summary.

```
table(fit$clusters)
```

```
 1  2  3
10 10 10
```

```
round(fit$medoids[, c(1, 5, 9, 13)], 3)
```

```
     [,1]  [,2]  [,3]  [,4]
[1,]    1 0.787 0.619 0.487
[2,]    1 0.549 0.301 0.165
[3,]    1 0.340 0.115 0.039
```

[2] `unsurv` R-universe documentation: https://ielbadisy.r-universe.dev/unsurv.

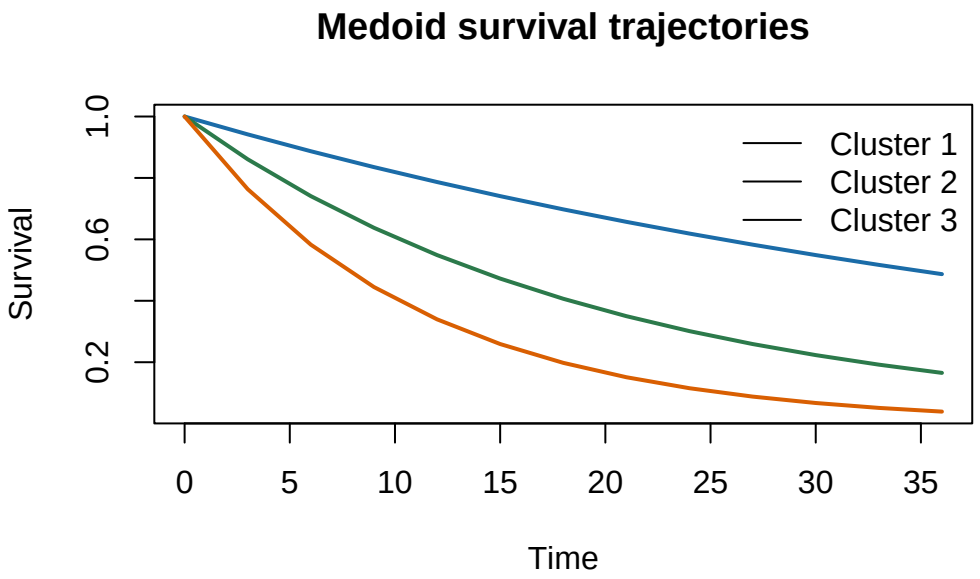
```
fit$silhouette_mean
```

```
[1] 0.9744392
```

Plots

The medoid plot displays the representative survival trajectory for each cluster.

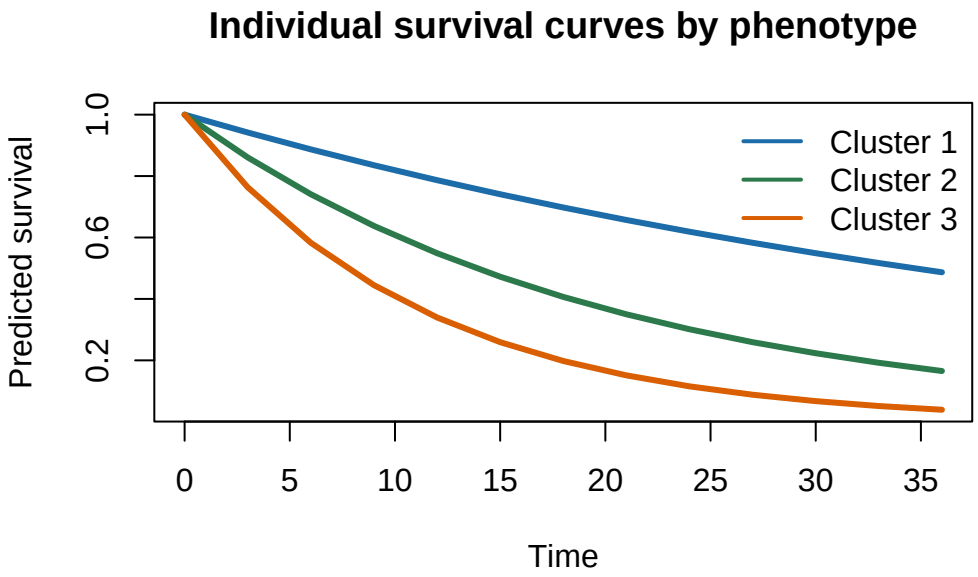
```
plot(
  fit,
  main = "Medoid survival trajectories",
  col = c("#1B6CA8", "#2C7A4B", "#D95F02"),
  lwd = 2
)
```



A second plot overlays all individualized survival curves, colored by the assigned phenotype. This shows the within-cluster spread around each medoid.

```
cluster_cols <- c("#1B6CA8", "#2C7A4B", "#D95F02")
matplot(
  times,
  t(S),
  type = "l",
  lty = 1,
```

```
  col = adjustcolor(cluster_cols[fit$clusters], alpha.f = 0.45),
  xlab = "Time",
  ylab = "Predicted survival",
  main = "Individual survival curves by phenotype"
)
lines(times, fit$medoids[1, ], col = cluster_cols[1], lwd = 3)
lines(times, fit$medoids[2, ], col = cluster_cols[2], lwd = 3)
lines(times, fit$medoids[3, ], col = cluster_cols[3], lwd = 3)
legend(
  "topright",
  legend = paste("Cluster", 1:3),
  col = cluster_cols,
  lty = 1,
  lwd = 2,
  bty = "n"
)
```



The three medoid rows summarize low-, intermediate-, and high-risk survival trajectories at 0, 12, 24, and 36 months. In this simulated example, the clusters are balanced because the data were generated from three known exponential survival profiles. The mean silhouette is high, indicating strong separation between trajectory groups.

Interpretation

A practical reporting template is:

- Cluster 1: slow decline in predicted survival, high expected event-free time.
- Cluster 2: intermediate decline, moderate expected event-free time.
- Cluster 3: rapid decline, low expected event-free time.

After clustering, each phenotype should be described with survival probabilities at clinically meaningful horizons, RMST summaries, and baseline covariate distributions. This keeps the phenotype anchored to survival behavior while making it interpretable for clinical collaborators.[3]

[3]Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.