ICLR Conference
TS4H Workshop
23 | 27 JULY

A Novel Methodological Framework for the Analysis of Health Trajectories and Survival Outcomes in Heart Failure Patients

Juliette Murris[1], Tristan Amadei[2], Tristan Kirscher[2], Antoine Klein[2], Anne-Isabelle Tropeano[3] & Sandrine Katsahian[1,3]

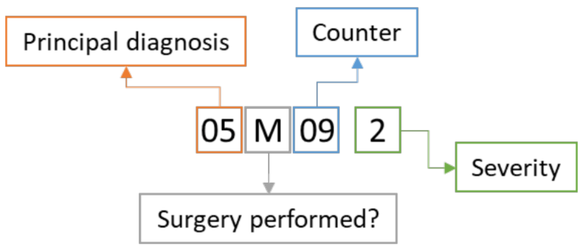
[1] HeKA, Inserm, Inria, Université Paris Cité, Pierre Fabre R&D [2] ENSAE, IP de Paris, [3] CIC-1418, HEGP, AP-HP, Paris, France

CONTEXT & OBJECTIVES

Motivating example

- Heart failure (HF) is common amongst **elderly** patients and associated with a **high mortality rate** *[Farré 2017]*
- Chronic HF is often accompanied with **repeated hospitalizations** and is the condition with **highest 30-days re-hospitalization rate** *[Constantinou 2021]*

Figure 1. ICD-10 architecture



Available data

- The **EGB** (*Echantillon Généraliste des Bénéficiaires*) is a random sample representative of the **French health insurance databases** and provides **in-hospital electronic health records**
- **International Classification of Disease** (10th edition (ICD-10)) is used to establish **primary and associated diagnoses** of hospitalizations (Figure 1)

Study objectives

- Identify frequent care sequences in HF patients in France
- Investigate associations with mortality

METHODOLOGICAL FRAMEWORK

Identify similarities from care pathways using clustering

Input – Care pathways (examples in Table 1), excluding factors like gender and age

Methodology

- Defining the appropriate distance metric to quantify the distance between two patients' care pathways, based on Levenshtein distance with weighted components of the ICD-10 codes. For two ICD-10 codes A and B:

Table 1. Three patients and their care pathways

Patient	hospit_course
P0	05K051,05M042,05M16T,05M09T,05M092,05C191,…
P6	02C05J,02C05J,05M093,04M132,05C222,23M103
P8	11M041,06C194

$$D_{ICD10}(A,B) = \omega_1 * lev(A_{0:2}, B_{0:2}) + \omega_2 * lev(A[2], B[2]) + \omega_3 * lev(A_{3:5}, B_{3:5}) + \omega_4 * lev(A[5], B[5])$$

→ We then compare the i^{th} ICD-10 code of a patient with the $(i-1)^{th}$, i^{th} and $(i+1)^{th}$ ICD-10 codes of another patient, compute the distances and keep the minimum to get the distance between two patient sequences.

- K-medoids algorithm used to group data points into k clusters. Two hyperparameters require settings and are under constraint:
  - Ω the weights of the distance metric, with $0 \leq \omega_4 \leq \omega_3 \leq \omega_2 \leq \omega_1 \leq 100$,
  - $k \in [2,20]$ the number of clusters.

Extract frequent care pathway patterns using sequential pattern mining

Input – Subset of items (ICD-10 code), event sequence (ordered list of ICD-10 codes)

Methodology – PrefixSpan algorithm was retained and uses the concept of "prefixes" to efficiently search for frequent patterns in a sequence database *[Pei 2001]*

Clinicians' interest in survival analysis to predict the survival probability of a specific HF patient

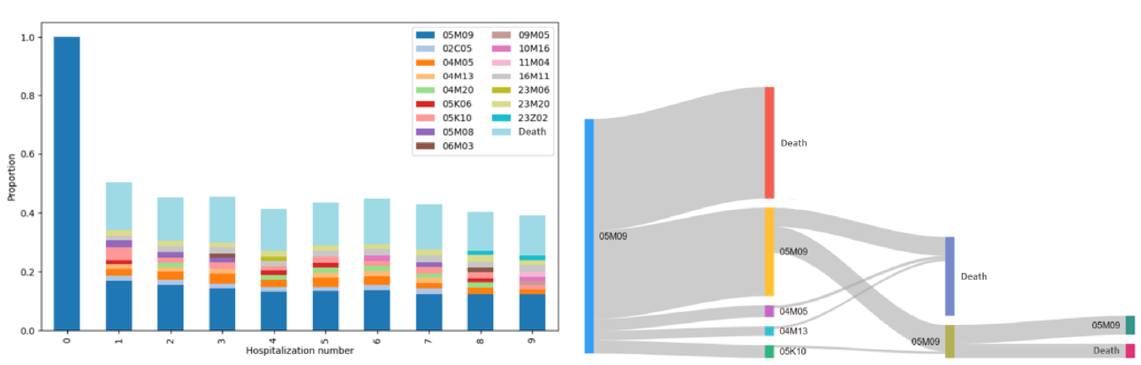
Input – Survival time and status (dead or alive), adjusted covariates

Methodology – Random survival forests and survival gradient boosting algorithm (hyperparameter tuning conducted*) [Ishwaran 2008, Hothorn 2006]*, and evaluation using AIC and concordance index *[Harrell 1982]*

Outputs

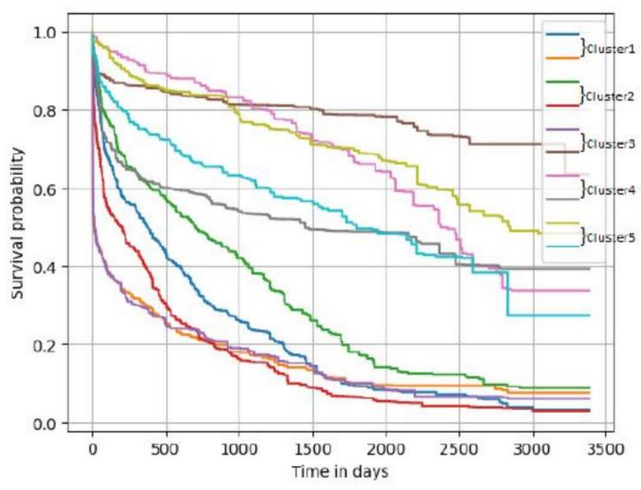
- The ten most frequent ICD-10 codes collectively accounted for approximately 50% of trajectories, indicating significant similarity in the care sequences (Figure 2)
- Common sequences leading to death are '05M09' for HF hospitalization, '04M05' for pleurisy, and '04M13' for pulmonary edema and respiratory distress
- Different survival trajectories from the 5 cluster s obtained (Figure 3)
- Aging and prolonged hospital stays are also impactful risk factors

Figure 2. Key figures for Cluster 1



(a) Frequent ICD-10 codes (b) Sankey diagram

Figure 3. Survival predictions in clusters with best and worst scenarios



DISCUSSION & CONCLUSION

- Our approach is **simple** and easily **accessible**
- While our focus has been on the HF patients, our approach is **adaptable** and **can be extended** to other clinical problematics and populations

Perspectives

- Include **up-to-date methodologies with NLP and embedding techniques** to extract event more relevant information from ICD-10 codes
- Include **risky patterns** identified as covariates in survival models

BIBLIOGRAPHY

Constantinou P, Pelletier-Fleury N, Olié V, Gastaldi-Ménager C, Juillère Y, and Tuppin P. Patient stratification for risk of readmission due to heart failure by using nationwide administrative data. Journal of Cardiac Failure, 27(3):266–276, 2021.

Farré N, Vela E, Clèries M, Bustins M, Cainzos-Achirica M, Enjuanes C, Moliner P, Ruiz S, Verdu-Rotellar JM, and Comin-Colet J. Real-world heart failure epidemiology and outcome: A population-based analysis of 88,195 patients. PLOS ONE, 12(2):e0172745, February 2017. ISSN 1932-6203. doi: 10.1371/journal.pone. 0172745. URL https://dx.plos.org/10.1371/journal.pone.0172745.

Harrell F, Califf R, Pryor D, Lee K, and Rosati R. Evaluating the yield of medical tests. Jama, 247(18):2543–2546, 1982.

Hothorn T, Buhlmann P, Dudoit S, Molinaro A, and Van Der Laan M. Survival ensembles. Biostatistics, 7(3):355–373, 2006.

Ishwaran H, Kogalur U, Blackstone E, and Lauer M. Random survival forests. The Annals of Applied Statistics, 2(3):841–860, September 2008. ISSN 1932-6157, 1941-7330. doi: 10.1214/08-AOAS169. URL https://projecteuclid.org/journals/annals-of-applied-statistics/volume-2/issue-3/Random-survival-forests/10.1214/08-AOAS169.full.

Pei J, Han J, Mortazavi-Asl B, Pinto H, Chen Q, Dayal U, and Hsu m. PrefixSpan,:mining sequential patterns efficiently by prefix-projected pattern growth. In Proceedings 17th International Conference on Data Engineering, pp. 215–224, Heidelberg, Germany,2001. IEEE Comput. Soc. ISBN 978-0-7695-1001-9. doi: 10.1109/ICDE.2001.914830. URL http://ieeexplore.ieee.org/document/914830/.

Pinaire J, Azé J, Bringay S, and Landais P. Patient healthcare trajectory, an essential monitoring tool: a systematic review. Health information science and systems, 5:1–18, 2017.

https://github.com/Kirscher/TextMining_Parcours_de_soin