

Meta-TAK¹: A Scalable Double-Clustering Method for Treatment Sequences Visualization in Breast Cancer from Claim data



Context

This work focuses on the study of **treatment sequences**, particularly the extraction of patterns from non-clinical claim databases through clustering. For this purpose, the **TAK² algorithm** was proposed and demonstrate its usefulness. However, the **scalability** of the TAK algorithm regarding the number of patients is an issue, the method being impossible to use in practice for thousands of patients. For this purpose, an **extension of the TAK algorithm** is presented in the following. Referred to as Meta-TAK, this method appears to be **robust and computationally efficient**.

An example using HER2-positive early breast cancer (eBC) extracted from the PMSI (2,477 patients) illustrates the **quantitative and qualitative performances of the method**.

Method

TAK algorithm

1. **Model** each patient treatment sequence as a timeline
2. Algorithmically **arrange** timelines through a similarity criterion (using an Agglomerative Clustering configured with the Hamming distance and the Ward³ linkage method)
3. Use final ordering to produce an image

Arranging timelines: Highly combinatorial problem

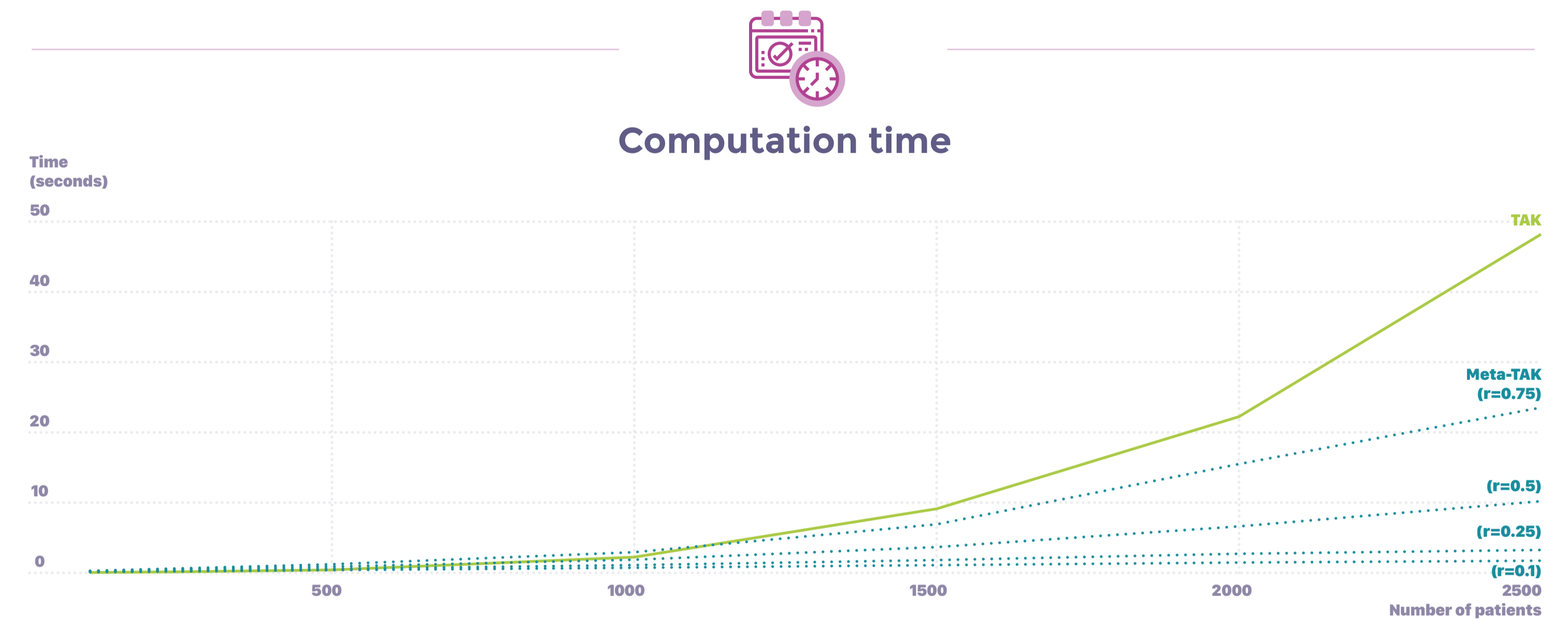
Evolution of the TAK: Meta-TAK algorithm



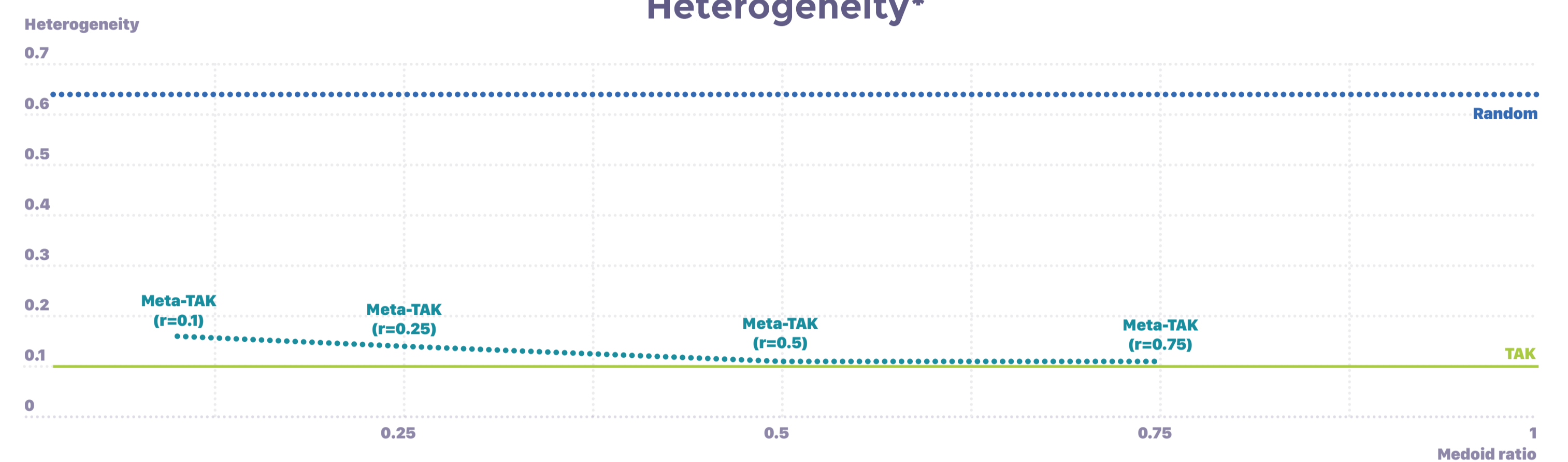
Results

Performance

The Meta-TAK is a great alternative to the TAK for large cohorts of patients (thousands of patients)



Heterogeneity*

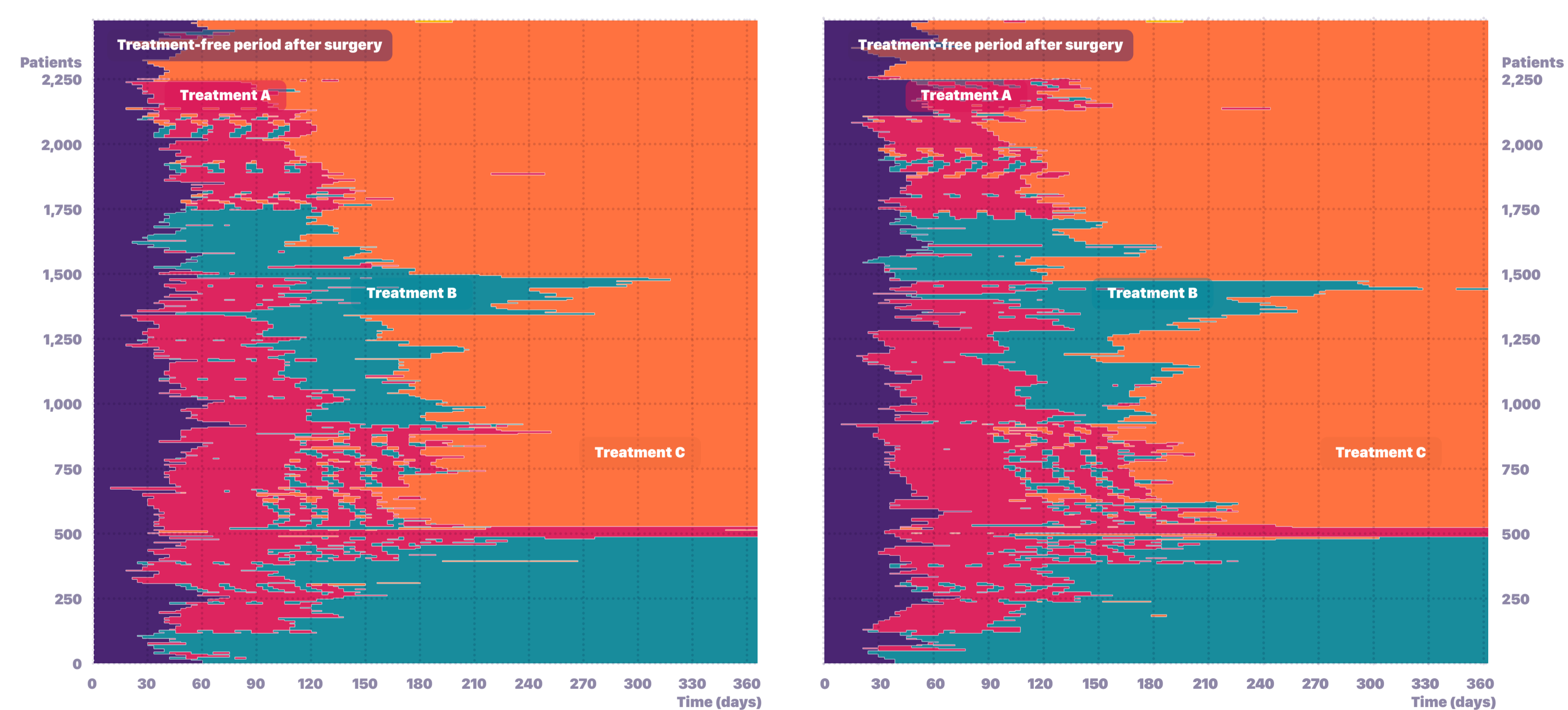


*The heterogeneity is defined as the mean value of the information entropy⁵ computed on a 3x3 grid of pixels, all over the image. The lower bound of this measure is 0, for a uniform image.

Medical interpretation

The Meta-TAK maintains a very close heterogeneity while reducing the computing time (-79% on the example)

TAK	Meta-TAK (r=0.5)
48s. Computation time	10s. Computation time
0.15 Heterogeneity	0.16 Heterogeneity



Conclusion

The Meta-TAK paves the way to scale up the clustering method to large cohorts of tens of thousands of patients while maintaining a low heterogeneity, which makes clinical interpretations still trustworthy.