Tristan Trébaol
26.08.2021

Full report & code documentation

Bayesian inference with Expectation Maximisation for the characterisation of antibiotic treatment recovery in Cystic Fibrosis (CF)

Objective

Improve hospital-based monitoring of CF *by asking clinical questions related to recovery:*

- How does a recovery after an antibiotic treatment look like?
- What is the typical response to treatment?
- Are there different types of recoveries?

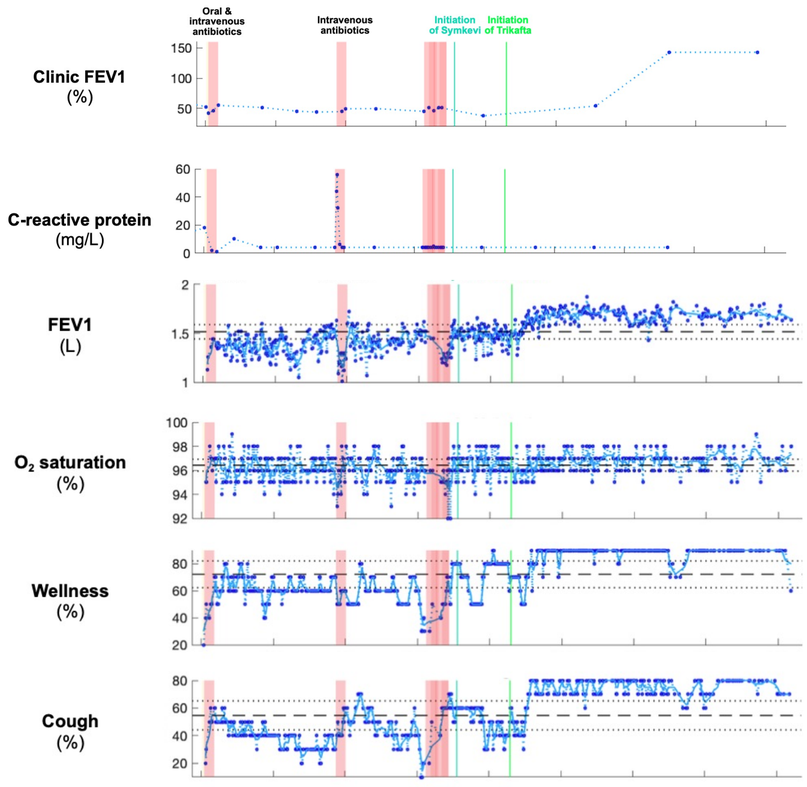
and answering them with Machine Learning

1. CF patient's physiological data

Physiological data from 258 individuals with CF is recorded for 2 years on a daily basis, as part of a home monitoring study with 2 hospitals in the UK. The dataset contains 500k measurements in 17 features, from which 8 are studied (wellness, cough, FEV1, FEF2575, O2 saturation, pulse rate, temperature, minutes asleep). Bimonthly clinical data for those patients is provided by the hospitals.

Clinical data:
Bimonthly recordings

Home monitoring data:
Measurements ~3 times per week

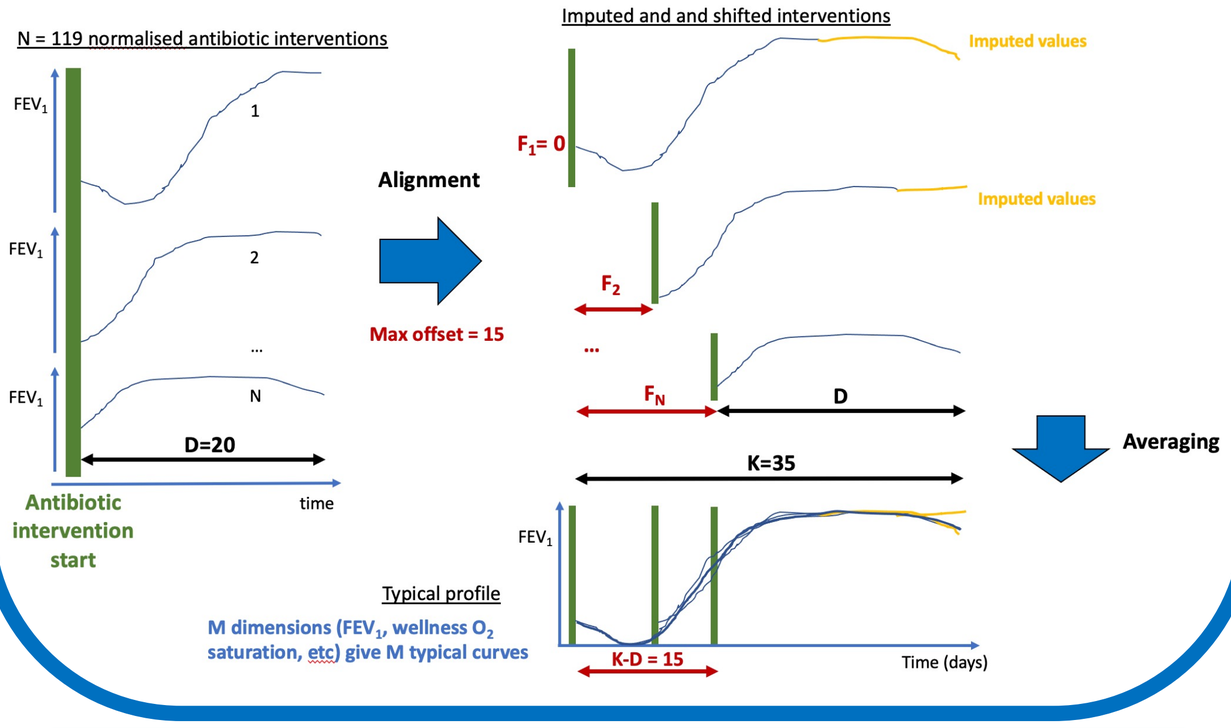


2. The alignment model

A model is built to align the time series of patient's physiological data at recovery start, thereby drawing the typical profile of a recovery.

Model assumptions:

1. For each measure, the recorded values for the period immediately following treatment are a noisy version of a single typical profile.
2. he amount a measured value deviates from this profile is controlled by the position on the profile and is independent from one day to the next.
3. The treatment start can happen anytime between day 1 and the maximum allowed offset (K-D).

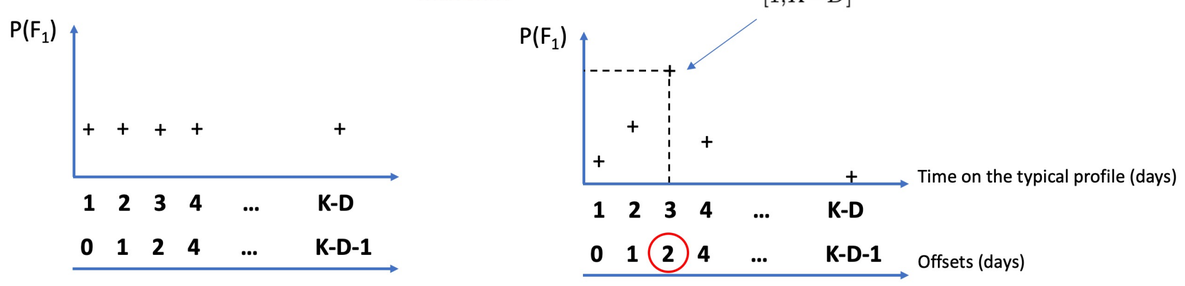


3. Probabilistic generative algorithm

The curve alignment is performed using a Bayesian inference approach with convergence through Expectation Maximisation. The information below shows the mathematical derivation for the single class (see report for multiple class inference).

Initialisation: $P(F_N)$ is uniformly initialised over the offsets' span [0; K-D-1]

End-state: $\mathcal{F} : [1;N] \longrightarrow [0;K-D-1], n \longmapsto F_n, F_n = \underset{[1;K-D]}{\arg\max}(P(F_n))) - 1$ (no shift: $K-D-1$; max shift: $[1;K-D]$)



Expectation-step: **Update $P(F_n)$**

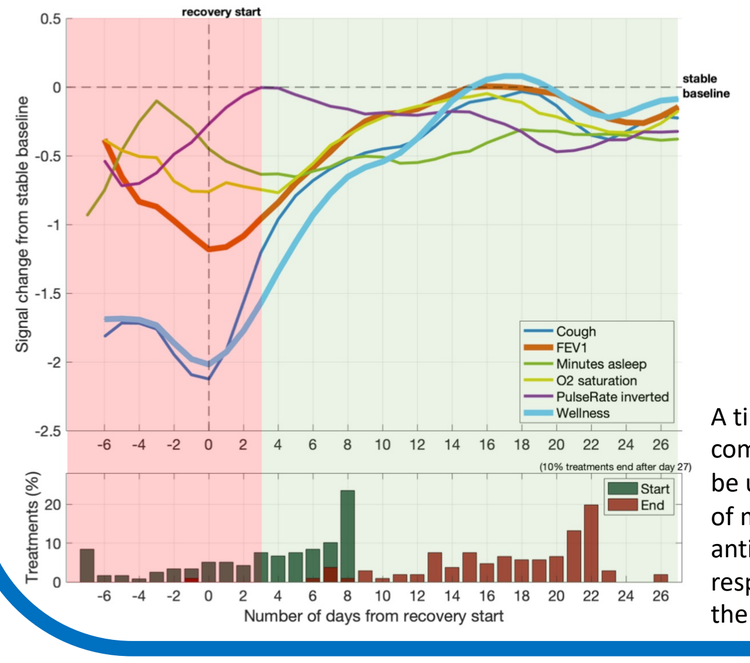
$$\forall n \in N, Obj_n = \ln(P(F_n|V_n,Z,S)) = -\frac{1}{2}\cdot\sum_{m=1}^{M}\sum_{d=1}^{D}\left[A+\left(\frac{V_{n,m,d}-Z_{m,d+F_n}+B}{S_{m,d+F_n}}\right)^2\right]$$

Maximisation-step: **Averaging**

$$\forall k \in [1;K], m \in [1;M], \begin{cases} Z'_{m,k} = \frac{\sum_{n=1}^{N}\sum_{f=0}^{K-D-1} V_{n,m,k-f}\cdot P(f)\cdot W(k,f)}{\sum_{f=0}^{K-D-1} P(f)\cdot W(k,f)} \\ S'_{m,k} = \frac{\left(\sum_{n=1}^{N}\sum_{f=0}^{K-D-1} V_{n,m,k-f}\cdot P(f)\cdot W(k,f)\right)^2}{\sum_{f=0}^{K-D-1} P(f)\cdot W(k,f)} - Z'^2_{m,k} \end{cases}$$

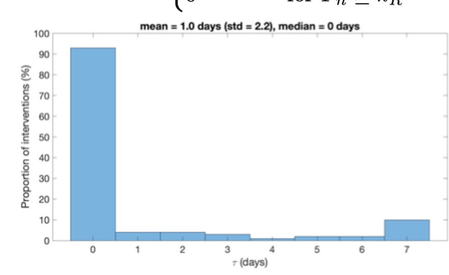
5. Typical recovery profile

A characteristic profile of the changes in physiology and symptoms during a recovery was generated using the algorithm. The profile allows to define an accurate recovery start date, which provides a label to explore the time to response, and the quality of the recovery. It also revealed that health bio-markers typically respond sharply to treatment and recover fully back to stable baseline. 40% of recoveries started from a full decline (red zone in figure). After a recovery paroxysm, there is a call-back with stabilisation nearby the stable baseline. More importantly, no clear decline can be observed in a typical recovery.



Recovery start - offset

$$\forall n \in N, \tau = \begin{cases} k_R - F_n & \text{for } F_n < k_R \\ 0 & \text{for } F_n \geq k_R \end{cases}$$



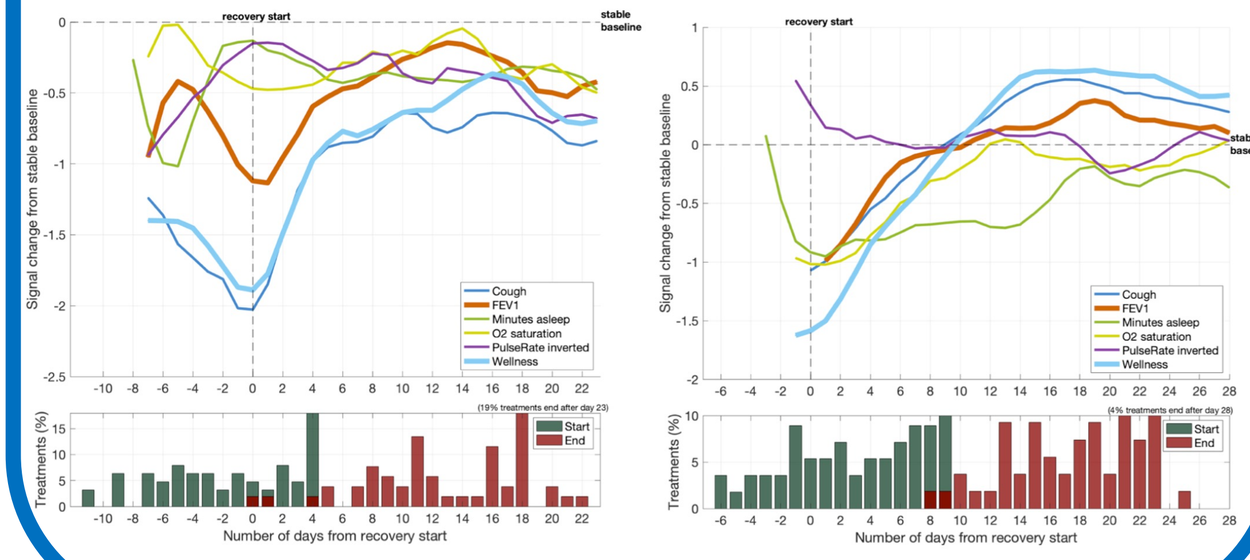
A time to treatment response was computed with the learned offset. This can be used in the future to analyse the impact of multiple competing features, including antibiotic choice, on the time to treatment response and eventually give the patient the antibiotic that minimises this time.

6. Multiple recoveries inference

The probabilistic inference algorithm was also used to infer the two most typical types of recoveries. The two-fold partition of the samples was balanced: class 1 contained 53% of the data records and 47% for class 2. Based on the following results, this indicates that half of the recoveries were successful and half were only partial recoveries, thereby unsuccessful.

Class 1: 53% of antibiotic interventions

Class 2: 47% of antibiotic interventions



Conclusion

Actionable information for the clinician:

1. Patients with higher amount of treatments are more likely to experience successful recoveries. Already known, hence validates the ML approach
2. Prognosis of recovery quality: A high increase in subjective parameters (cough and wellness) not followed by physiological signals (FEV1, O2 saturation) can indicate the beginning of an unsuccessful recovery.

Future work:

1. Analyse the impact of multiple competing features, including antibiotic choice, patient microbiology on the quality of recovery.
2. Infer long-term outcomes of combined treatments and therapies (*in particular CFTR modulators*).

7. Definition of a recovery

A recovery is a process of change in the patient's health status following an antibiotic treatment, closely linked to the preceding acute pulmonary exacerbation. It lasts from the treatment start until the day where a recovery label can be assigned with sufficient certitude using the related graph: