A novel personalized stepwise dynamic predictive algorithm in Chronic Lymphocytic Leukemia

Theodoros Moysiadis[1,*], Dimitris Koparanis[1], Konstantinos Liapis[1], Maria Ganopoulou[2], George Vrachiolias[1], Ioannis Katakis[3], Chronis Moyssiadis[4], Ioannis Vizirianakis[5], Lefteris Angelis[2], Konstantinos Fokianos[6], Ioannis Kotsianidis[1]

[1]Department of Hematology, Democritus University of Thrace Medical School, Alexandroupolis, Greece; [2]School of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece; [3]Department of Computer Science, School of Sciences and Engineering, University of Nicosia, Nicosia, Cyprus; [4]School of Mathematics, Aristotle University of Thessaloniki, Thessaloniki, Greece; [5]School of Pharmacy, Aristotle University of Thessaloniki, Thessaloniki, Greece; [6]Department of Mathematics & Statistics, University of Cyprus, Nicosia, Cyprus

Abstract: Personalized prediction is an ideal perspective in Chronic Lymphocytic Leukemia (CLL), yet presently an unattainable goal. This study proposes a novel personalized stepwise dynamic predictive algorithm (PSDPA) for the time-to-first-treatment (TTFT) for the individual patient.

PSDPA involves a Score computation at each time-point (patient visit), exploiting the time dependent nature of the patient's follow-up records, and forming a continuous time series (TS). Higher Score reflects more intense disease deterioration compared to lower values. PSDPA involves (i) develop a pool of treated/control patients, and compute their Score-TS from diagnosis until TTFT/last follow-up, respectively, (ii) for a "new" patient, not having received treatment at time-point *n*, compute the Score until *n*, and detect the group of patients to which is mostly similar to, and (iii) employ binary TS modeling to this group, and predict, at *n*, the TTFT for the "new" patient.

PSDPA was applied on 14 treated and 6 control patients. The average time between time-points was approximately 3 months, and were assumed to be equally distanced. Each of the 14 treated patients, assumed the role of a "new" patient and was compared to all other 19 patients. The *n* was set to 13. Prediction was performed in 8/14 cases with 7/8 predictions differing less than 4 time-points compared to the observed time. The results provide evidence that PSDPA could be employed to promote personalized prediction in CLL.

Aim: The personalized stepwise dynamic predictive algorithm (**PSDPA**) aims to:

- ❖ exploit the information within the time-dependent follow-up of an individual patient (followed-up prospectively, and considered as the "new" patient), and represented by a customized Score time series
- ❖ compare it to corresponding Scores of reference patients (already retrospectively analyzed)
- ❖ dynamically predict for the individual "new" patient (starting from a specific time point) the time required until first treatment

The **PSDPA** can flexibly integrate additional biological information (cytogenetic, genomic, immunophenotypic, immunogenetic, epigenetic, transcriptomic, etc.)

The PSDPA

(i)
- Develop a "reference" pool of patients (retrospective follow-up data available)
- N_T treated patients, N_C controls
- Compute the Score TS, $S_i(t)$ (N_T: until 1st treatment, N_C: last follow-up)

(ii)
- For each treated patient, estimate an appropriate threshold $T_i, i = 1, 2, \ldots, N_T$, to transform the continuous TS, $S_i(t)$, into a binary TS, $BS_i(t)$

(iii)
- Consider a "new" patient. Appropriately select a time point *n* to initiate performing personalized predictions for TTFT (condition: patient has not yet received treatment at *n*)

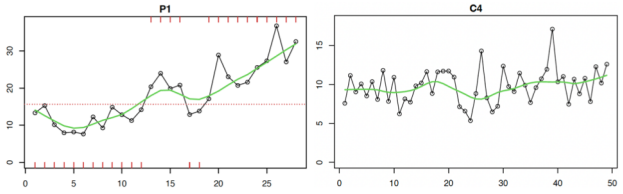
(iv)
- Compute $S_{new}(t)$, and compare it with the corresponding first *n* time points in each of the $S_i(t)$ in the reference pool, using Dynamic Time Warping alignment
- Biological information may be optionally utilized within this step

(v)
- Determine the group with the "mostly similar" patients to the "new" patient. Use appropriate criteria to assess it, and, if the criteria are satisfied, extract information that represents the mean group-behavior until TTFT. If the criteria are not satisfied, go back to (iii) and use *n+1*

(vi)
- Assess the conditions at time point *n*, and, if satisfied, proceed with the personalized prediction of the TTFT for the "new" patient (estimate the time points until the TTFT). If the conditions are not satisfied, go back to (iii) and use *n+1*

Application: 20 CLL patients (14 treated, and 6 controls)

Figure 1: The $S_i(t)$ (in black) is displayed for two patients, one treated (left), and one control (right), along with a smooth spline (in green). The treated patient's plot includes a horizontal dotted line, representing the estimated binary threshold (15.64), and the corresponding vertical segments on the top and bottom of the plot (in red), which correspond to the, $BS_i(t)$, with values one and zero, respectively. Time point 0 represents the time of diagnosis for both patients. The last time point corresponds to the time point before the initiation of first treatment for the treated patient, and the last follow-up for the control.



Funding: The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "2nd Call for H.F.R.I. Research Projects to support Post-Doctoral Researchers" (Project Number: 553).

Figure 2: The pairwise distances between the continuous Score TSs of all patients at *n*=15, are displayed after they were divided with the maximum distance.

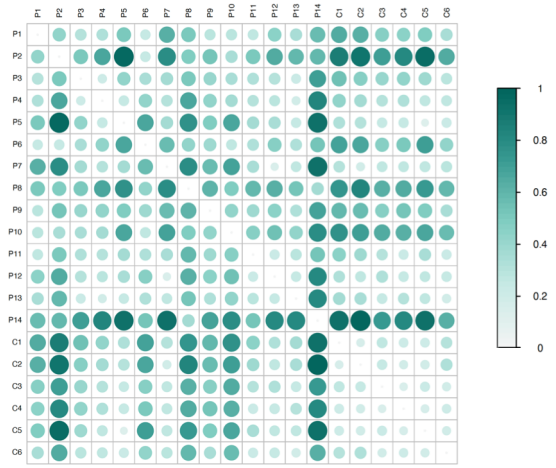
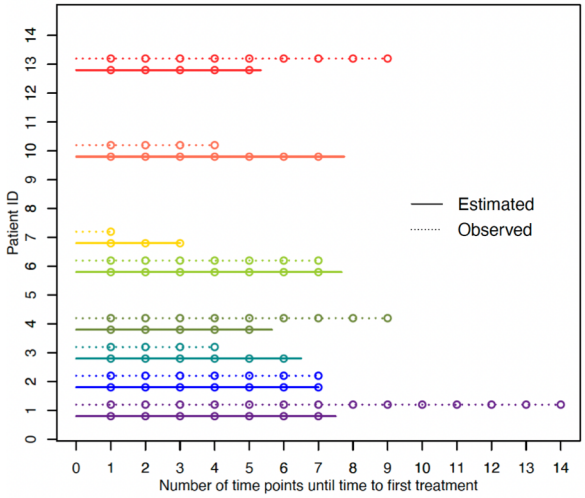


Figure 3: The personalized predictions of the number of time points required until the time to first treatment ("estimated") are provided for the 14 treated patients along with the corresponding "observed" numbers. The number of time points is calculated based on the time point that prediction is actually made (*n*, *n+1*, *n+2*, etc.)



Conclusion: The PSDPA can utilize the time-dependent information from the clinical follow-up of the individual patient, and, optionally, standard time-independent information reflecting his/her biological background. Using a specific time point n (after diagnosis) as a starting point, it can provide (when appropriate) in a dynamic fashion a personalized prediction for the time required until first treatment for the individual patient of interest.

The PSDPA could be used to build personalized approaches, essential to accommodate the clinical heterogeneity in CLL, and could potentially be employed as a supplemental tool to support decision making in clinical practice.