A novel stepwise dynamic algorithm to personalize prediction: Preliminary results from 20 Chronic Lymphocytic Leukemia patients

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

[1]Department of Hematology, University Hospital of Alexandroupolis, Democritus University of Thrace Medical School, Dragana Area, 681 00, Alexandroupolis, Greece; [2]School of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; [3]School of Mathematics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; [4]School of Pharmacy, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; [5]Department of Mathematics & Statistics, University of Cyprus, 2109 Nicosia, Cyprus

Abstract: Personalized prediction is an ideal perspective in Chronic Lymphocytic Leukemia (CLL), yet presently an unattainable goal. An ongoing effort has taken place in the recent years to develop refined prognostic models and stratify patients in risk groups based on information at the time of diagnosis and/or time of first treatment. Still this approach cannot accommodate time-dependent information for the individual patient, failing to address the considerable clinical heterogeneity, still observed within the resultant risk groups. This study aims to contribute towards personalized prediction in CLL by proposing a novel personalized stepwise predictive algorithm for disease progression for the individual patient.

The algorithm involves a Score computation to exploit the time series nature of the individual patient's follow-up records. At each time-point (patient visit) a record-based Score value is computed, forming a discrete Score time series (TS). Higher Score values represent more intense disease deterioration compared to lower values. The stepwise algorithm involves the following steps: (i) Develop a reference pool of retrospectively analyzed patients. For each treated/control (not having received treatment) patient, compute the Score-TS at all time-points from diagnosis until time-to-first-treatment (TTFT)/last follow-up, respectively, (ii) For a "new" patient prospectively analyzed and not having yet received treatment at time-point n (when prediction for the TTFT will be initiated), compute his/her Score TS from diagnosis until time-point n, compare it with the patients' pool (using an appropriate TS distance measure), and determine the group of patients to which is mostly similar to, (iii) Apply binary time series modeling to this "similar" patient group, and utilize the results to predict for the "new" patient at time-point n, the time until first treatment. This could be generalized for time to next treatment, and overall survival.

In a short application 14 treated and 6 untreated patients (controls) were retrospectively analyzed and constituted the reference pool [step (i)]. The average time between the Score time-points was approximately 3 months (imputation was performed when needed, and time-points were assumed to be equally distanced). Then, each of the 14 treated patients, separately assumed the role of a "new" patient and was compared to all other 19 patients (step (ii)). The value of n was initially set to 13 when appropriate amount of information will be available. The results have shown that treated patients are more likely to be mostly similar to other treated patients than to controls, and control patients were similar to each other and more dissimilar to the treated ones. Prediction was performed in 8/14 cases with 6/8 predictions being close to the observed time. In 6/14 cases no prediction was made (algorithm conditions were not satisfied).

The proposed predictive algorithm may take into account the whole information from the follow-up of the patient. Since many parameter aspects are involved (distance measure, similarity distance threshold, etc.), an extensive study is required to assess the sensitivity of the algorithm to different parameter settings. The proposed approach could dynamically refine the stratification of a patient in a time dependent manner and promote personalized prediction in CLL.

Aim: The personalized predictive algorithm aims to:

- ❖ exploit the time-dependent information of an individual patient, (followed-up prospectively, and considered as the "new" patient),
- ❖ compare it to 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 algorithm can integrate biological information (cytogenetic, genomic, immunophenotypic, immunogenetic, epigenetic, transcriptomic, etc.).

The personalized stepwise predictive algorithm

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

(ii)
- For each treated patient, estimate an appropriate threshold $T_i, i = 1, 2, \dots, N_T$, 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, when relevant (patient has not yet received treatment).

(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 (DTW) 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.

(vi)
- Assess the conditions at time point *n*, and, if appropriate, proceed with the personalized prediction of the TTFT for the "new" patient.

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.

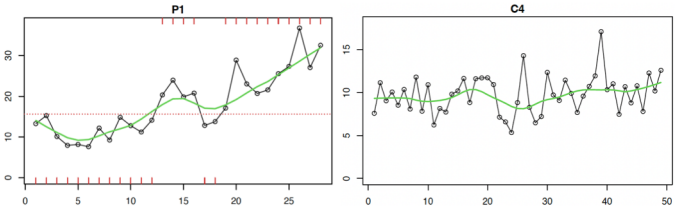


Figure 2: The pairwise distances between the continuous Score TSs of all patients at *n*=15, are displayed after they were divided with the maximum distance.

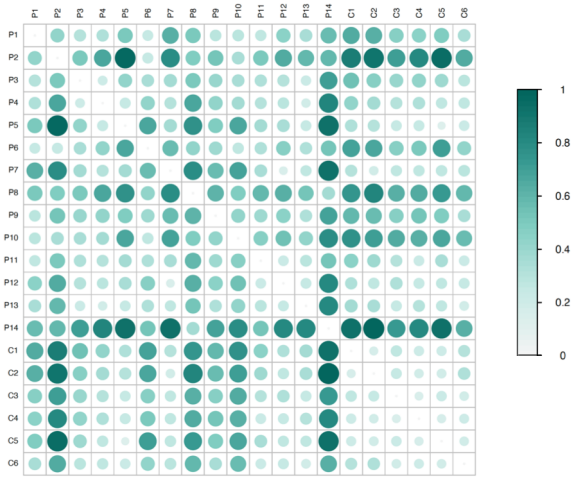
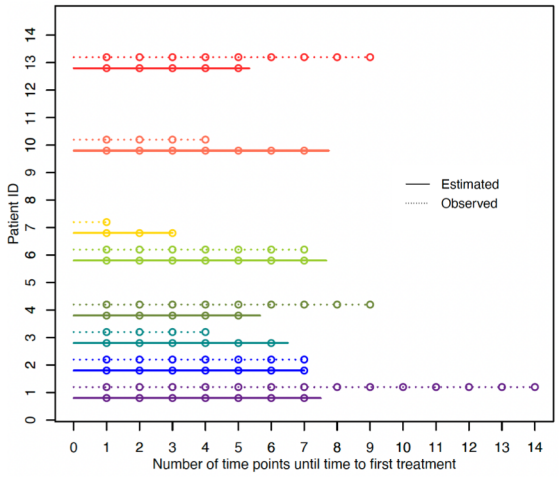


Figure 3: The personalized predictions of the number of time points required until the time to first treatment ("estimated") are provided for the 14 treated patients along with the corresponding "observed" numbers.



Conclusion: The personalized predictive algorithm 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 the clinical practice.

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).