Predicting imminent disease progression in advanced colorectal cancer by a machine-learning algorithm

Yuri Kogan[1], Moran Elishmereni[1], Marina Kleiman[1], Shmuel Shannon[1], Larisa Aptekar[1], Eldad Taub[1], Hovav Nechushtan[2], Zvia Agur[1]

[1] Optimata Ltd. Israel, [2] Hadassah Medical Center, Israel

Abstract #645

Background

In advanced cancers, predicting disease progression before its clinical manifestation enables an earlier switch to the next treatment line prior to deterioration in the patient's state, potentially improving survival. Yet, at present, serum tumor markers such as carcinoembryonic antigen (CEA) are poor indicators even of the current tumor state, and certainly cannot be used for forecasting future outcomes such as progression.

We developed a machine-learning algorithm alerting to approaching disease progression in patients with Colorectal Cancer (CRC), using longitudinal tumor marker input.

Methods & Results

CRC patient data under standard 1st line treatment: (1) **Clinical study data** (denoted FL-4/6, FL-Pan; projectdatasphere.org); (2) **Real-world evidence** from Hadassah Medical Center (denoted HMC). Study-eligible patients contained sufficient tumor assessment and marker evaluations.

Data type	Clinical studies		Hospital-registry
Dataset (acronym)	FL-4/6	FL-Pan	HMC
Study (NCT identifiers)	NCT00272051 NCT00305188	NCT00364013	NCT02577627
Number of patients			
Total patients	756	935	132
Study-eligible patients	489	729	92
Tumor assessments			
Progression events	175	251	32
Non-progression events	741	1818	142
Follow-up time (until progression)			
Median, months	5.6	6.7	6.3
Number of tumor marker measurements			
CEA	1047	2400	313
CA19.9	-	-	224

Tab. 1. Data characteristics.

Can Tumor Marker Values Directly Predict Progression?

Basic prediction potential of the tumor markers CEA and CA19.9 was examined by the standard method of **receiver operating characteristic (ROC) analysis** on marker values monitored in-treatment.

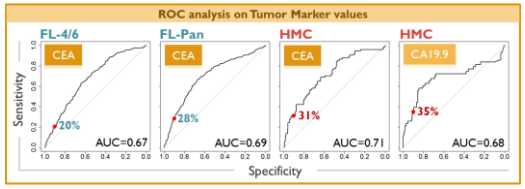


Fig. 1. Tumor markers - weak signals for progression

With current methods, **tumor markers carry weak signals** and are not useful to signal approaching progression (Fig. 1)

Algorithm Development

Modeling approach: machine-learning using R.
Training: the Random Forest algorithm trained using dynamic tumor marker features as input, and tumor assessment records (by RECIST 1.1) as output.
Testing: leave-one-out cross-validation on the same dataset or separate testing on another dataset.

Tab. 2. Algorithm performance metrics in clinical study datasets.

	Cross-Validation		Testing	
Training dataset	FL-4/6	FL-Pan	FL-4/6	FL-Pan
Testing dataset	FL-4/6	FL-Pan	FL-Pan	FL-4/6
Sensitivity	57%	52%	68%	51%
Specificity	88%	90%	77%	93%
Pos. Predictive Value	64%	54%	40%	72%
Neg. Predictive Value	84%	90%	92%	83%
Accuracy	79%	83%	76%	81%
Cohen's Kappa	0.46	0.43	0.35	0.48

The algorithm signals imminent progression at **high specificity and sensitivity**, and can prompt a timely switch to next line therapy (Tab. 2, Tab. 3)

Algorithm Validation

Tab. 3. Algorithm testing on CRC patients in the Hadassah hospital registry.

	Independent Testing	
Training dataset	FL-4/6	FL-Pan
Testing dataset	HMC	HMC
Sensitivity	70%	58%
Specificity	81%	88%
Pos. Predictive Value	53%	61%
Neg. Predictive Value	90%	87%
Accuracy	78%	81%
Cohen's Kappa	0.46	0.47

The algorithm uses **simple, accessible, low-cost markers, and enhances their value** for predicting progression

Multiple-Marker Algorithm

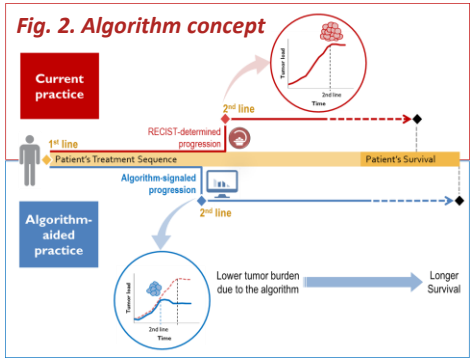
Tab. 4. Algorithm performance metrics in CRC patients using CEA and CA19.9 data.

The algorithm combines **multiple tumor markers** to produce an **even stronger progression signal**– Tab. 4

	Cross-Validation
Training/Testing dataset	HMC (CEA+CA19.9)
Sensitivity	58%
Specificity	92%
Pos. Predictive Value	65%
Neg. Predictive Value	90%
Accuracy	86%
Cohen's Kappa	0.53

A New Paradigm for Algorithm-aided Clinical Practice

Fig. 2. Algorithm concept



Conclusions and Implications

- By use of machine-learning, we created a **new algorithm that amplifies weak signals from tumor markers** monitored during treatment, **to produce a strong alert of disease progression just before the tumor surges** (Fig. 2).
- The algorithm-amplified ability of CEA to predict progression in CRC complements our recent findings in non-small cell lung cancer, where CEA integrated with 4 other markers provides 91% specificity and 66% sensitivity in predicting progression, surpassing the low capacity of each separate marker.[1]
- Similarly, adding more markers is expected to further boost the prediction capacity of the current algorithm for CRC.
- By **individually timing the therapy switch before disease deterioration**, the algorithm can **enhance the efficacy of 2nd line drugs**, thereby **extending the progression-free survival and overall survival rates** in cancer patients (Fig. 2).
- The paradigm of **algorithm-aided improvement of cancer treatment** can also be applied to further lines of therapy (e.g. 3rd line drugs) and additional indications.

Reference: [1] Kogan et al. J Clin Oncol 36, 2018 (suppl; abstr e21190).

Website: www.optimata.com
Correspondence: yurikogan@optimata.com