PREDICTING SOLID-TUMOR BEHAVIOUR OVER TIME WITH EXPLAINABLE GENAI FOR PRECISION ONCOLOGY

Author: Anna Tidstam, PhD

FPN:295P

Read more

WWW.AILEENHEALTH.COM

Background

Treatment decisions in breast cancer require individualized risk-benefit analysis, as recommended by ESMO guidelines. Current treatment decisions rely on evidence from large-scale clinical trials, which provide population-level average efficacy outcomes. These population-level data, though informative, does not provide patient-level information on how an individual may respond to a treatment, given their characteristics, which can delay adoption of emerging therapies.

Prediction of pCR with and without therapy-specific modelling

Prior models predicting pathological complete response (pCR) achieved good performance (AUC = 0.81). However, these models commonly included therapy information as a confounding factor, meaning treatment influenced outcomes but was not explicitly modeled. As a result, predictions from these models cannot be directly interpreted for therapy-specific outcomes. To estimate how a patient would respond to a specific therapy—a counterfactual scenario—therapy-specific modeling is required, where outcomes are predicted assuming a defined treatment, independent of the therapy actually received.

Methods

Factual (observable) treatment scenarios are compared with counterfactual (non-observable) treatment scenarios, using a generative approach with novel counterfactual metrics and novel validation methods. It is explainable because it generates therapy-specific, visual comparisons.

Generative approach for non-observable treatment scenarios

Generative models for image-based disease trajectories could predict tumor progression and treatment response. This generative approach has potential in personalised medicine, enabling visualisation of hypothetical follow-up images under different treatment scenarios, i.e. non-observable (counterfactual) treatment scenarios. By generating counterfactual image sequences, clinicians gain a visual aid to support decision-making and assess the efficacy of interventions. Models can be conditioned on clinical metadata, such as patient demographics or pathology status, to simulate patient-specific outcomes. As we are preparing generative models for counterfactual modelling, we need to find ways to evaluate them.

Counterfactual Δ-metrics for multi-treatment predictions

Patient-level counterfactuals are evaluated with therapy-specific Δ-metrics, e.g., ΔpCR, representing the predicted difference in pathological complete response between the actual (observable) and an alternative (non-observable) treatment scenario. There are many more endpoints that can be modelled with Δ-metrics, e.g. ΔAE for adverse effects. Here we are however considering ΔpCR.

Counterfactual validation

The retrospective validation of the actual (observable) treatment scenario is relatively straight-forward, as there is a ground truth. For the counterfactual treatment (non-observable) treatment scenario, novel methods are needed.

1. Blind Treatment Assignment Testing (BTAT)

A novel patient-level evaluation that captures the ability to correctly detect treatment assignments by generating both factual and counterfactual treatment scenarios. The identification is done by comparing with the ground truth. This method is primarily suitable during training of Cancer4D as a patient-level evaluation of its performance.

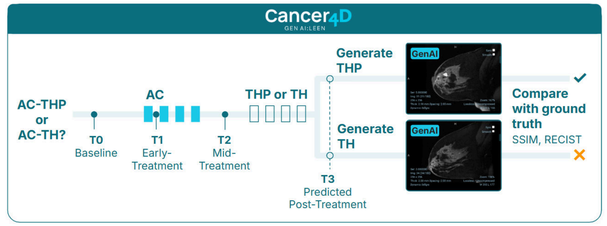

Fig 1: Blind Treatment Assignment Test, where the task is to identify which treatment that was assigned by comparing generated outcomes with the ground truth

2. Non-Observable RCT Benchmarking (NORB)

This benchmarking method compares counterfactual predictions with published trial results to assess consistency with known treatment efficacy differences.

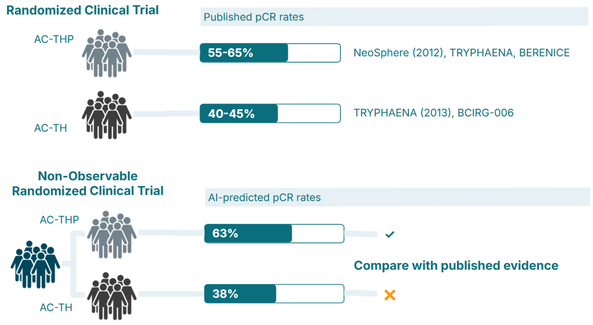

Fig 2: Non-Observable RCT Benchmarking by comparing predicted treatment outcomes with published clinical trial results.

3. Patient stratification with counterfactual ΔpCR

The third method we are proposing is the treatment recommendations that may be derived from the counterfactual metrics, in our case ΔpCR.

Pioneering Δ-Metrics for Oncologist's Treatment Decisions

Patient	*Switch*	*Switching Rationale*	*Cost €*	*Age*	*Frailty Index*	*ΔpCR*	*Tumor Δcm³*	*Tumor Δ%*	*ΔTol*
P0001	TH→THP	pCR achieved only with THP	+20k	55	0	Yes	15 (13, 16)	-25%	-10%
P0002	TH→THP	Significant tumor shrinkage	+20k	45	1	-	10 (8, 12)	-10%	-5%
P0003	THP→TH	Less side-effects	-20k	65	2	-	0 (0, 6)	0%	15%

Fig 3: Patient stratification with counterfactual ΔpCR

We are applying logistic regression modelling with he ISPY2 dataset in order to show the evaluation principles and metrics of this third method.

Results

The ISPY2 dataset contains 985 patients with sequential MRI imaging, 385 with detailed tumour data, distributed over 13 treatment arms. This forces us to create two treatment groups for feasibility demonstration based on A: higher pCR (n = 364, pCR = 34.1%) and B: lower pCR (n = 364, pCR = 21.5%). Both treatment groups are for HER2-negative subtypes.

Treatment A (group of treatments with higher pCR):
Paclitaxel/ABT-888/Carboplatin, Paclitaxel/Neratinib, Paclitaxel/AMG-386, Paclitaxel/Pembrolizumab, Paclitaxel/MK-2206

Treatment B (group of treatments with lower pCR):
Paclitaxel, Paclitaxel/Ganetespib, Paclitaxel /Ganitumab

We created two logistic regression models for groups A and B, respectively. 11 patients were recommended based on ΔpCR to switch from A to B (see Table 1), and 15 patients who should switch from B to A, (Table 2). This is aligned with the higher efficacy in group A. The tables contain longest diameter measurements corresponding to the timepoint T0, T1, T2 and T3 from Figure 1.

The patients that are strongly recommended to switch regimens are patients who had a negative pCR, but were predicted to have a positive pCR. For example, patient 320538 has no radiological visible tumor already at L2, but still has a negative pCR, see Table 1. The same is true for patient 790722, who also had no radiologically visible tumor are L2, see Table 2. This was then true for both recommendations, from group A to B, and B to A. The interpretation should be that these patients are predicted to achieve pCR with both treatment A and B, and the regimen switch is triggered since the pCR was not achieved with the factual treatment.

Table 1: 15 patients recommended to switch from treatment B to A (A has higher pCR rates).

Patient ID	pCR	ΔpCR	Arm (B→A recommendation)	HR	Age	L0 [cm]	L1 [cm]	L2 [cm]	L3 [cm]
320538	0	0,80	Paclitaxel	0	57	1,5	1,5	0	0
507198	0	0,78	Paclitaxel	0	44	3,9	3,8	0,5	0
454808	0	0,77	Paclitaxel + Ganitumab	0	56	2,6	0,7	0,6	0
115987	0	0,72	Paclitaxel + Ganitumab	0	42	4,0	2,0	1,7	0
233191	0	0,71	Paclitaxel	0	42	10,1	9,5	3,0	2,4
780272	0	0,69	Paclitaxel + Ganitumab	0	54	4,9	3,8	2,7	0
571276	0	0,67	Paclitaxel + Ganetespib	0	48	4,3	1,6	1,5	1,4
100899	0	0,67	Paclitaxel + Ganitumab	0	53	2,4	2,0	1,0	0,8
934906	0	0,66	Paclitaxel	0	32	7,9	5,8	3,8	2,9
421829	0	0,66	Paclitaxel	0	46	4,0	2,7	1,5	1,6
104268	0	0,66	Paclitaxel	0	50	3,5	3,0	1,6	1,3
875089	0	0,66	Paclitaxel + Ganitumab	0	42	3,5	2,1	2,0	0
559263	0	0,64	Paclitaxel	0	64	7,1	1,3	0,7	5,1
955035	0	0,61	Paclitaxel + Ganitumab	0	46	7,6	5,9	3,6	2,8
411950	0	0,60	Paclitaxel + Ganitumab	1	61	3,1	2,8	0	0

Table 2: 11 patients recommended to switch from treatment A to B (B has lower pCR rates).

Patient ID	pCR	ΔpCR	Arm (A→B recommendation)	HR	Age	L0 [cm]	L1 [cm]	L2 [cm]	L3 [cm]
790722	0	0,80	Paclitaxel + ABT 888 + Carboplatin	0	64	3,7	3,5	0	0
291515	0	0,79	Paclitaxel + ABT 888 + Carboplatin	0	44	4,0	3,1	0	0
652480	0	0,74	Paclitaxel + ABT 888 + Carboplatin	0	36	2,4	1,9	1,2	0
118307	0	0,70	Paclitaxel + AMG 386	0	44	3,9	2,9	1,2	0,6
411684	0	0,70	Paclitaxel + ABT 888 + Carboplatin	0	39	4,9	2,9	1,5	0,8
943633	0	0,68	Paclitaxel + ABT 888 + Carboplatin	0	50	2,8	2,3	1,5	0
286984	0	0,63	Paclitaxel + MK-2206	1	50	4,3	3,5	0	0
637566	0	0,62	Paclitaxel + MK-2206	0	68	3,7	3,6	2,1	0
228192	0	0,62	Paclitaxel + AMG 386	0	56	3,9	3,1	2,2	0
630758	0	0,61	Paclitaxel + ABT 888 + Carboplatin	0	57	5,2	3,0	1,7	0,5
591180	0	0,60	Paclitaxel + Neratinib	0	57	4,7	4,1	1,9	2,3

Conclusions

Of the three proposed methods for validating counterfactual predictions, only the patient stratification with ΔpCR could be calculated with the ISPY2 dataset. Future work requires larger patient cohorts with well-informed treatments and evidence of efficacy in order to validate a full image-based Cancer4D implementation for treatment-specific counterfactuals.

Corresponding author

Author: Anna Tidstam, Aileen Health, Salzburg, Austria
E-Mail: anna@aileenhealth.com

2025 ESMO AI & DIGITAL ONCOLOGY
Annual Congress
BERLIN GERMANY
12-14 NOVEMBER 2025

ai:leen HEALTH