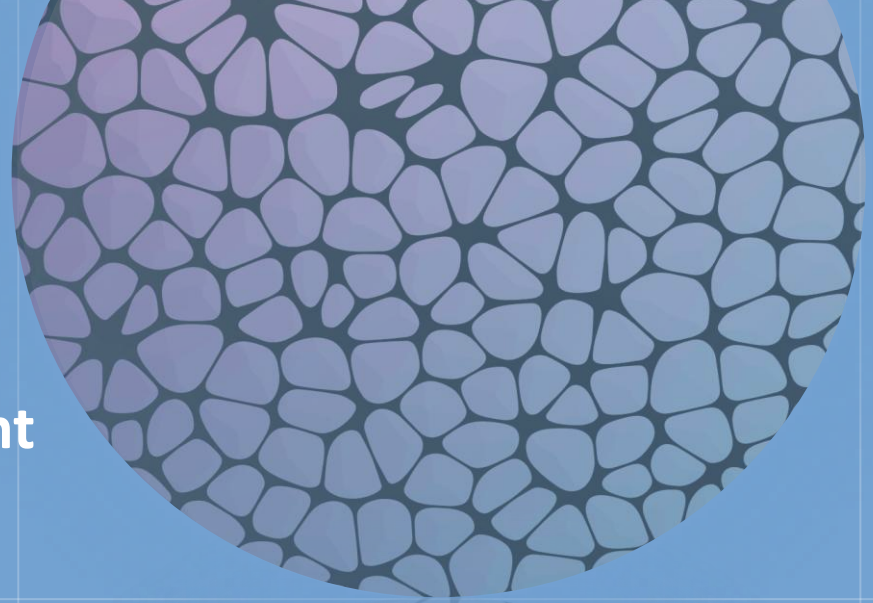
POLITECNICO MILANO 1863 | DIPARTIMENTO DI MATEMATICA

Digital Twins in Computational Oncology: Probabilistic Formulation of Personalized Risk Assessment

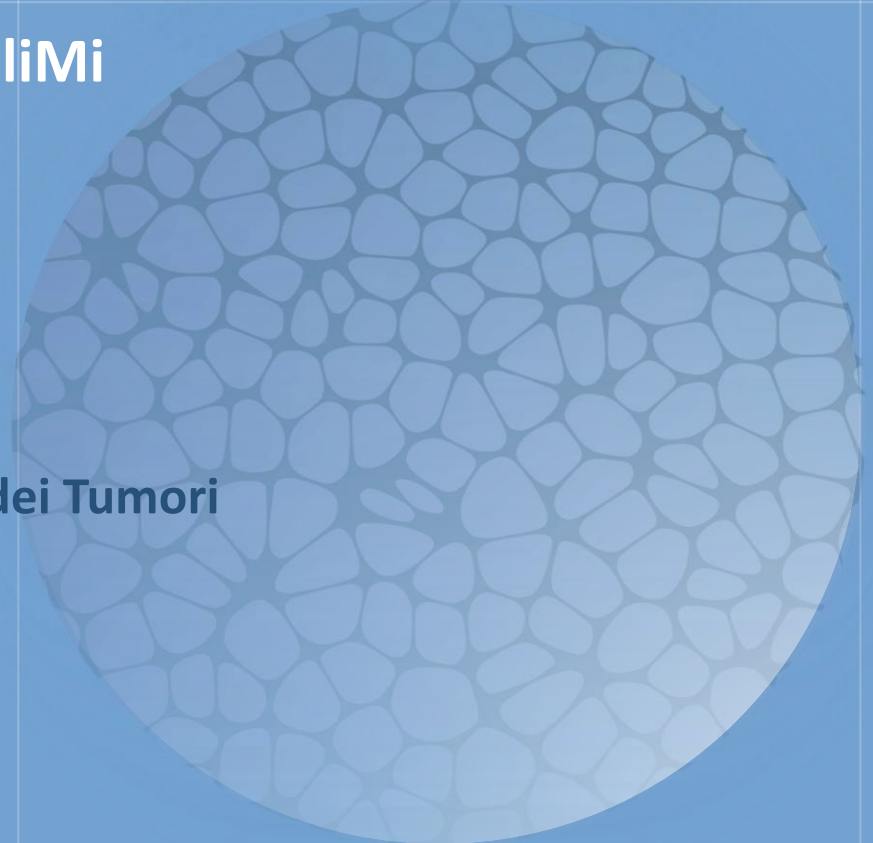


Francesca Arceci - MOX, Department of Mathematics, PoliMi

Joint work with

Prof. Paolo Zunino, Dr. Piermario Vitullo, MOX, DMAT, PoliMi

CMON Lab, Data Science Unit, Fondazione IRCCS Istituto Nazionale dei Tumori



20-24.11.2025 | Workshop

"The Mathematics of Scientific Machine Learning and Digital Twins"

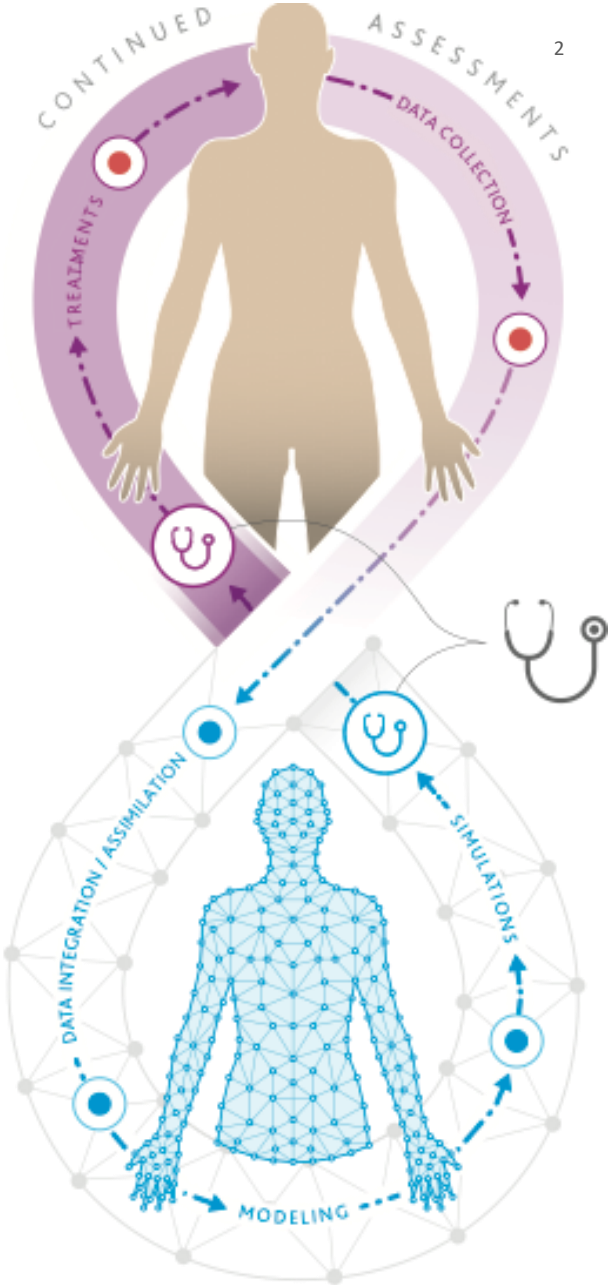
Digital Twin in Oncology

Radiation therapy (RT) is commonly used after surgery for various types of cancer, including **breast cancer.**

Despite protection and safety measures, radiation therapy can expose the lung and heart to radiation, **raising the risk of severe cardiac and pulmonary diseases.**

The project TETRIS proposes to explore the opportunities and challenges of applying **digital twins (DTs)** in RT safety.

What is a digital twin?



tetris

Risk assessment *Tools* for severe side *Effects* after *breas*T *R*adiotherapy: radiation safety through biological extended models and *d*igital *twin*S

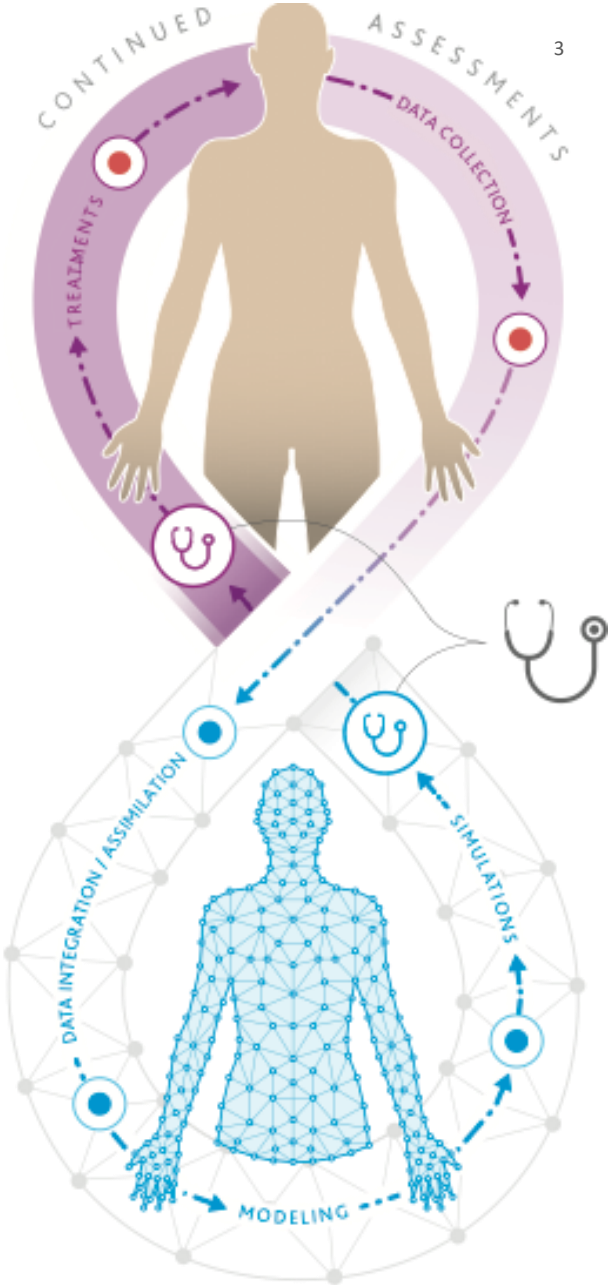
EU Grant Agreement n. 101166699

This information is from *Foundational Research Gaps and Future Directions for Digital Twins* available online at www.nationalacademies.org/digital-twins.

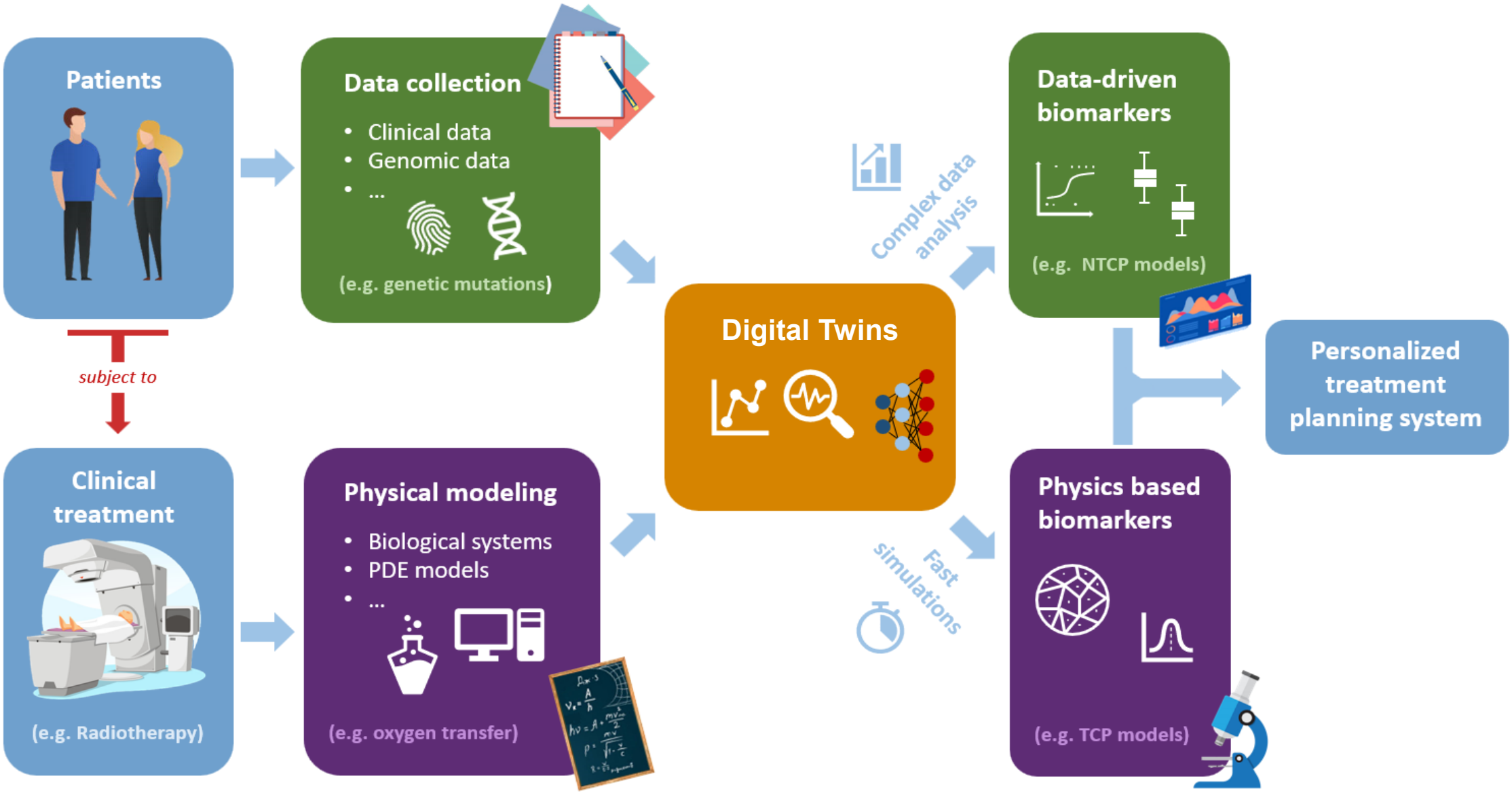
Digital Twin in Oncology

What is a digital twin?

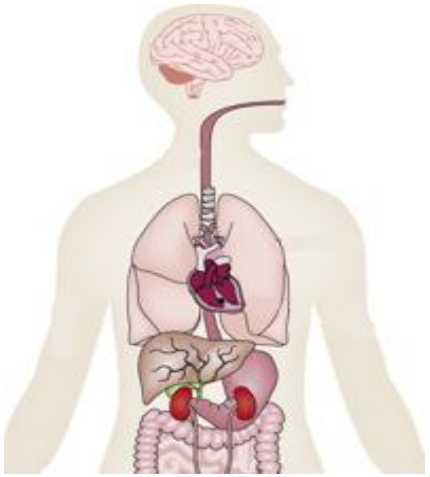
- A digital twin is a set of **virtual information constructs** that mimics **structure, context, and behaviour** of a physical system.
- A digital twin is **dynamically updated** with real-time **data** from its physical twin.
- A digital twin has **predictive capabilities.**
- It informs **decision-making that realize value**
- The **bidirectional interaction** between the virtual and the physical is central to the digital twin.



Digital Twins for Computational Oncology

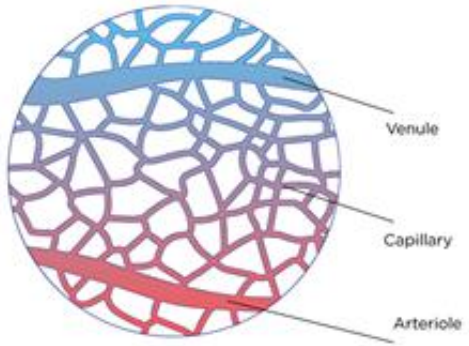


Personalized Radiotherapy



Personalize the treatment at the patient level to **maximize** the chances of **tumor eradication** and **prevent late toxicity.**

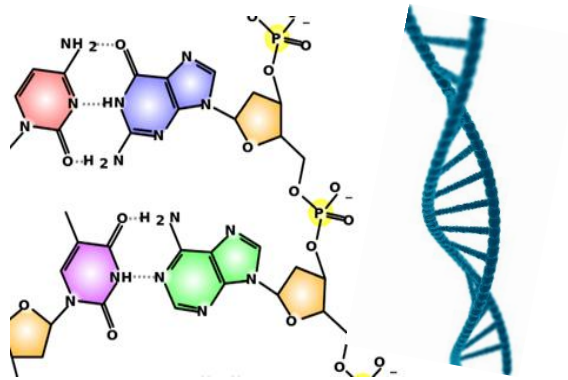
Physics driven



- How does the **tumoral micro-environment** affect radiotherapy?
- Can we provide **robust** and **efficient** numerical simulations?

Personalized Treatment Planning

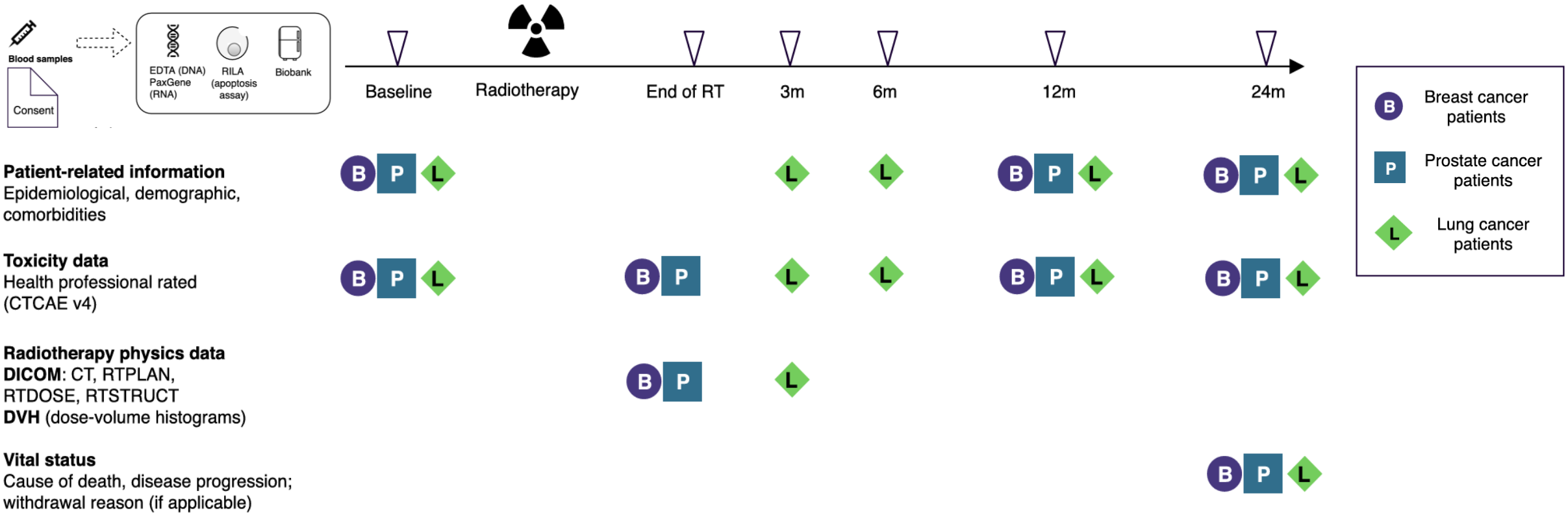
Data driven



- Can we quantify the risk?
- How do **genetic** traits influence radiosensitivity?

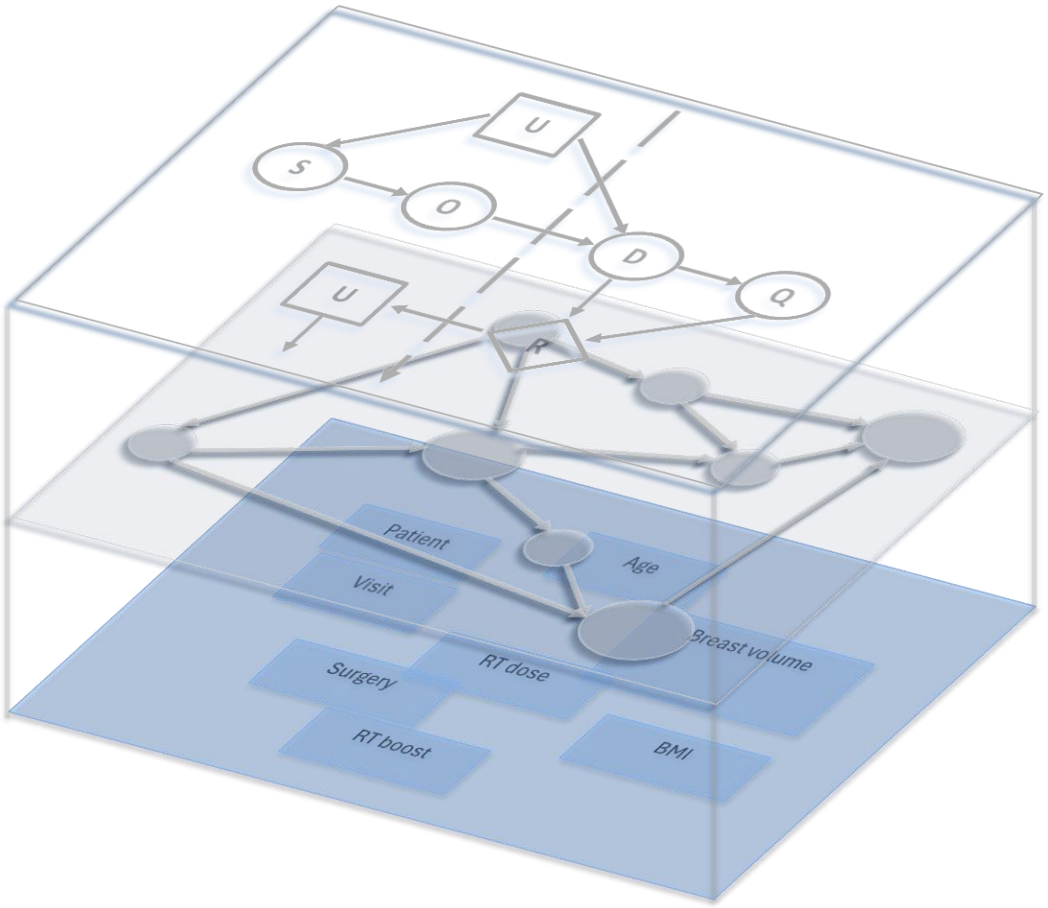
The REQUITE dataset

REQUITE is a large, international, multi-center cohort comprising over 5000 patients treated for breast, prostate, or lung cancer (*Seibold, Petra et al. Radiotherapy and Oncology, Volume 138, 59 – 67*).



RADprecise REQUITE

Towards the the DT for breast cancer patients



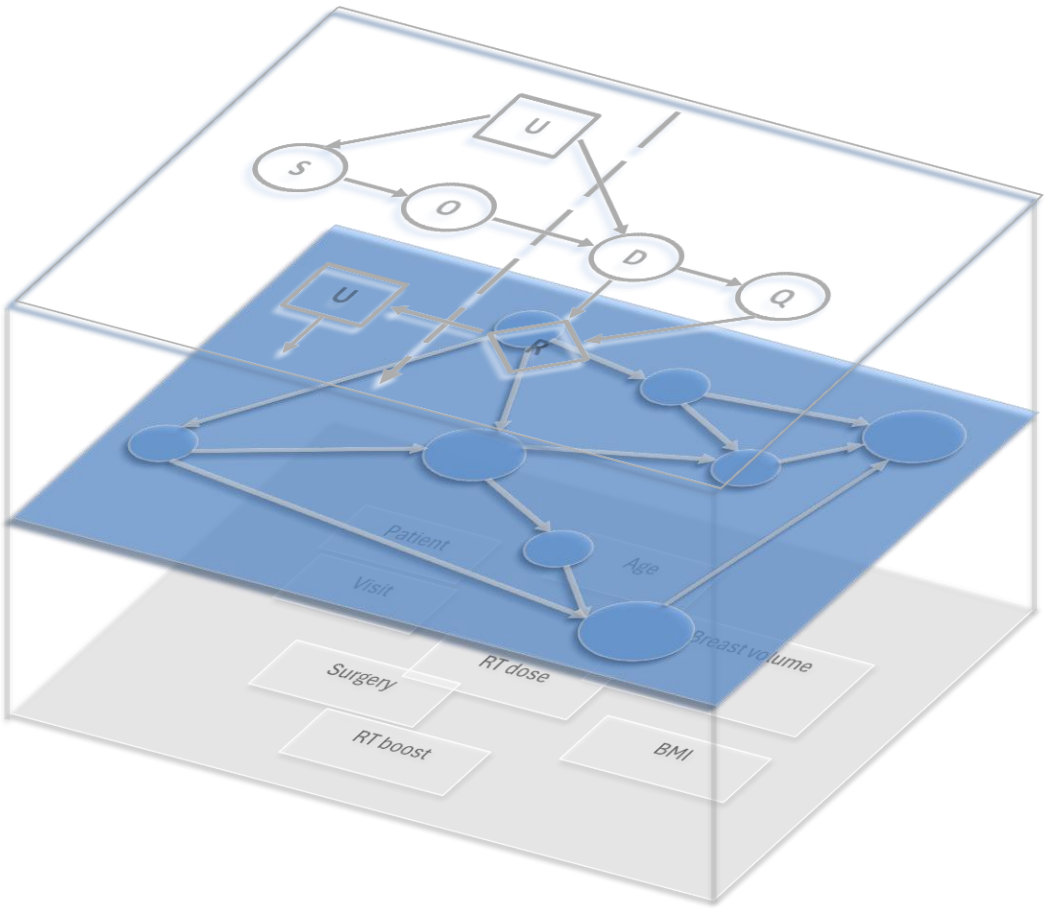
Computational level

Causal level

Data level

Seibold, Petra et al., *REQUITE: A prospective multicentre cohort study of patients undergoing radiotherapy for breast, lung or prostate cancer,* Radiotherapy and Oncology, Volume 138, 59 – 67 *(2019)*

Workflow for breast cancer patients



Computational level

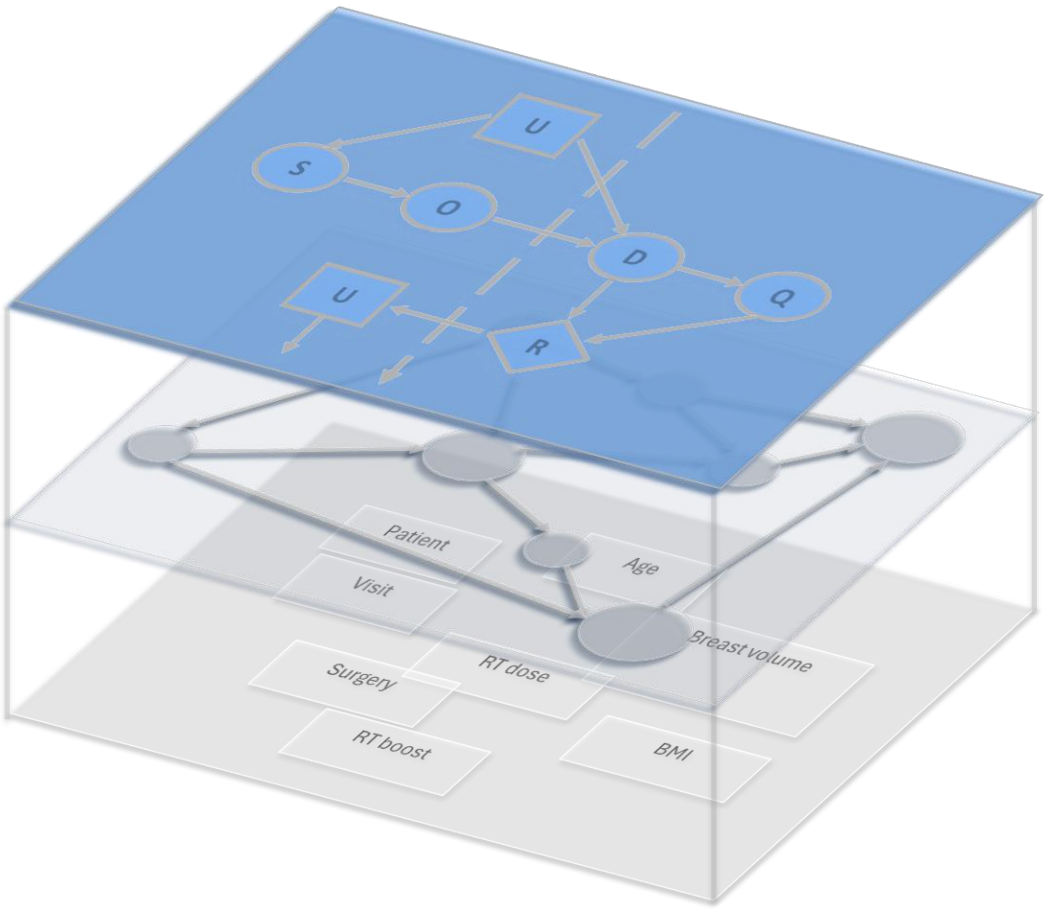
Causal level

Data level

van Amsterdam WAC et al., *Causal Inference in Oncology: Why, What, How and When,* Clininical Oncology, Volume 38 *(2025)*

B. Dionisi Ferrara & T. Rancati, CMON Lab, Istituto Nazionale dei Tumori

Workflow for breast cancer patients



Computational level

Causal level

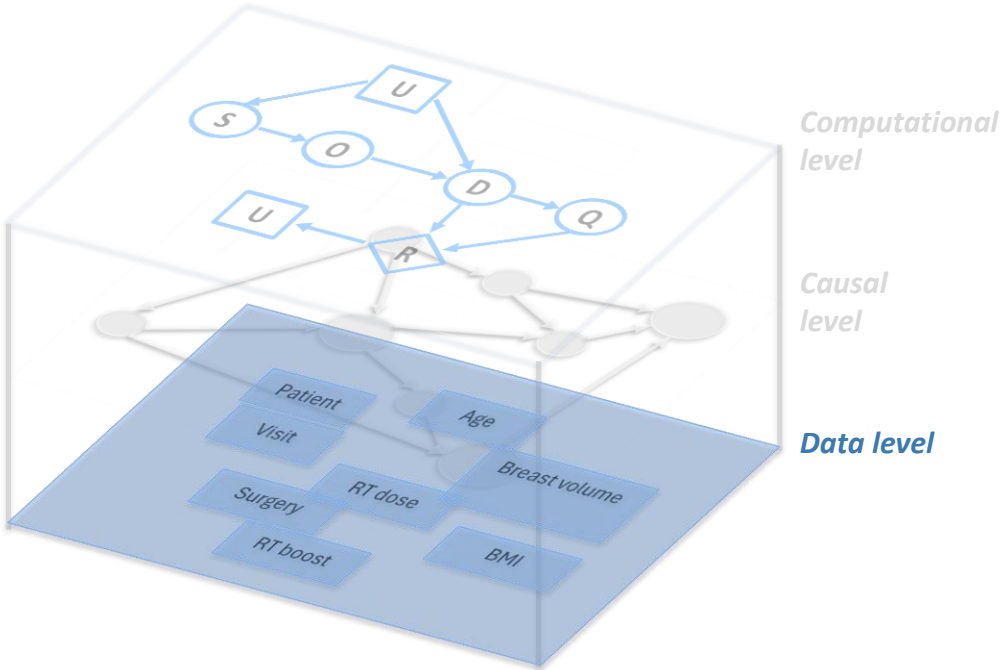
Data level

Willcox K.E. et al., *A probabilistic graphical model foundation for enabling predictive digital twins at scale,* Nature Computational Science, 1 (5), pp. 337 – 347 *(2021)*

Breast cancer patients: data level

The breast dataset consists in 2057 patients.

Data available up to 24 months after RT.



Patient information:

- ***Clinical data:*** *Age, BMI, comorbidities*
- ***Genomic data:*** *SNP genotypes, polygenic risk scores*
- ***Imaging data:*** *DICOM-RT*

B2a Breast patient factors baseline 2025-05-16 18-28-34.xlsx

Breast toxicity data:

- *Nipple retraction*
- *Skin induration (fibrosis)*
- *Erythema*
- *Pneumonitis*

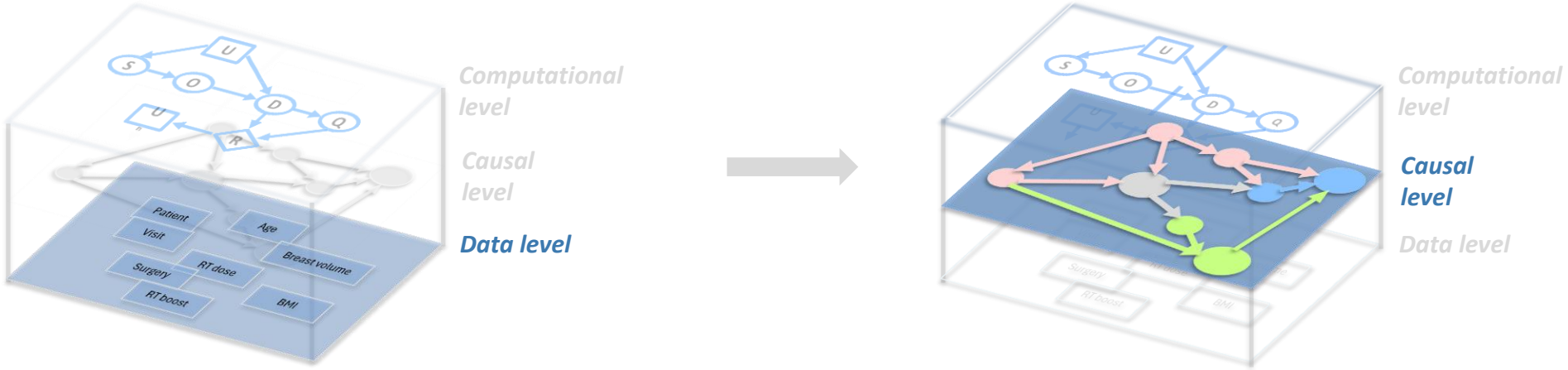
B5 Health professional 2025-05-16 18-31.xlsx

Breast clinical and treatment information:

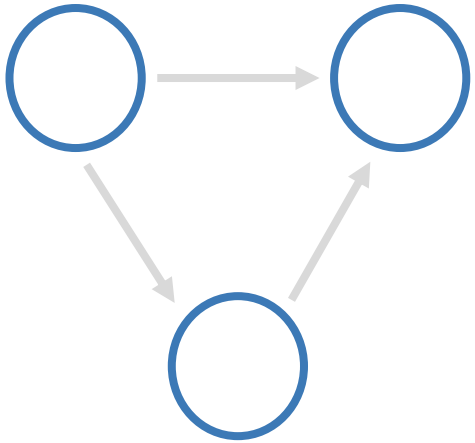
- *RT prescribed dose, fraction dose*
- *Breast skin dose, heart mean dose, lung mean dose*
- *RT boost, photon boost, neoadjuvant*

B3 Breast clinical and treatment data form 2025-05-16 18-30-19.xlsx

Breast cancer patients: causal level



- **Predictive models** not necessarily focus on the relationships among variables
- **Causal inference models** to uncover the mechanisms behind observed phenomena

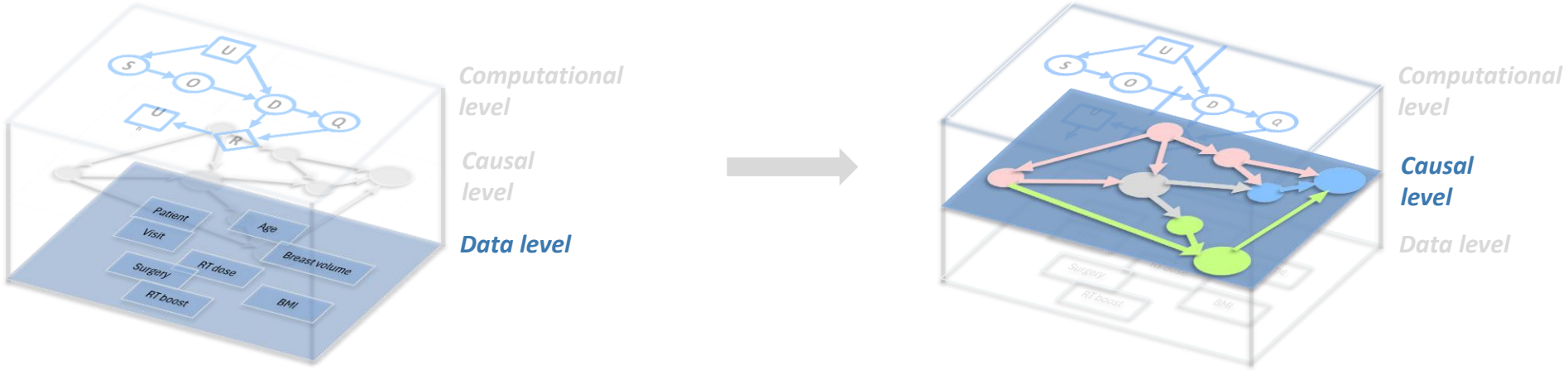


Directed Acyclic Graphs

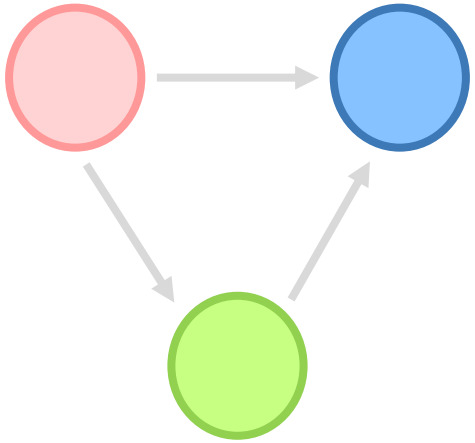
- **Nodes:** *selected variables*
- **Arrows:** *relationships from cause to effect*

B. Dionisi Ferrara & T. Rancati, CMON Lab,Istituto Nazionale dei Tumori

Breast cancer patients: causal level



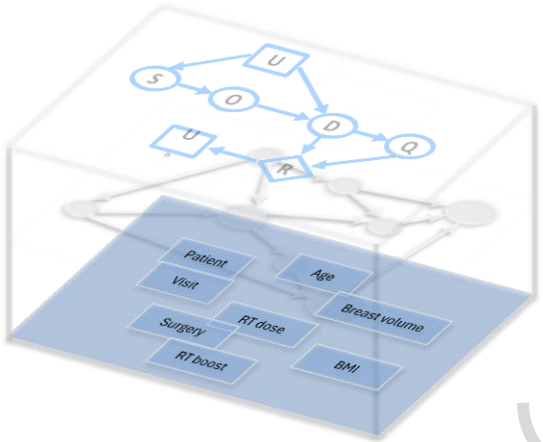
- **DAG** wants to explore the **direct effect** of the **exposure** on the **outcome**
- **Confounding variables affect both** the exposure and the outcome



- **Exposure variable**
- **Outcome variable**
- **Confounding variable**

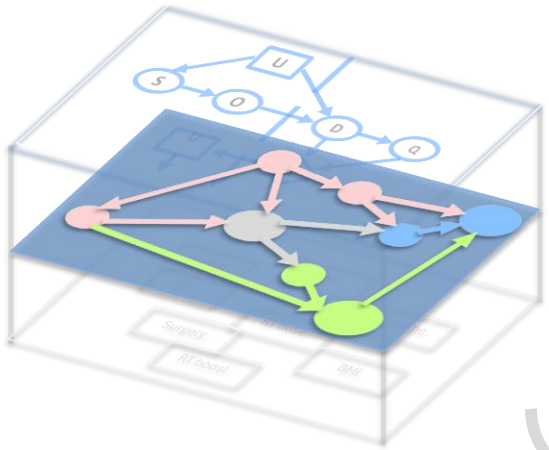
B. Dionisi Ferrara & T. Rancati, CMON Lab,Istituto Nazionale dei Tumori

Causal model



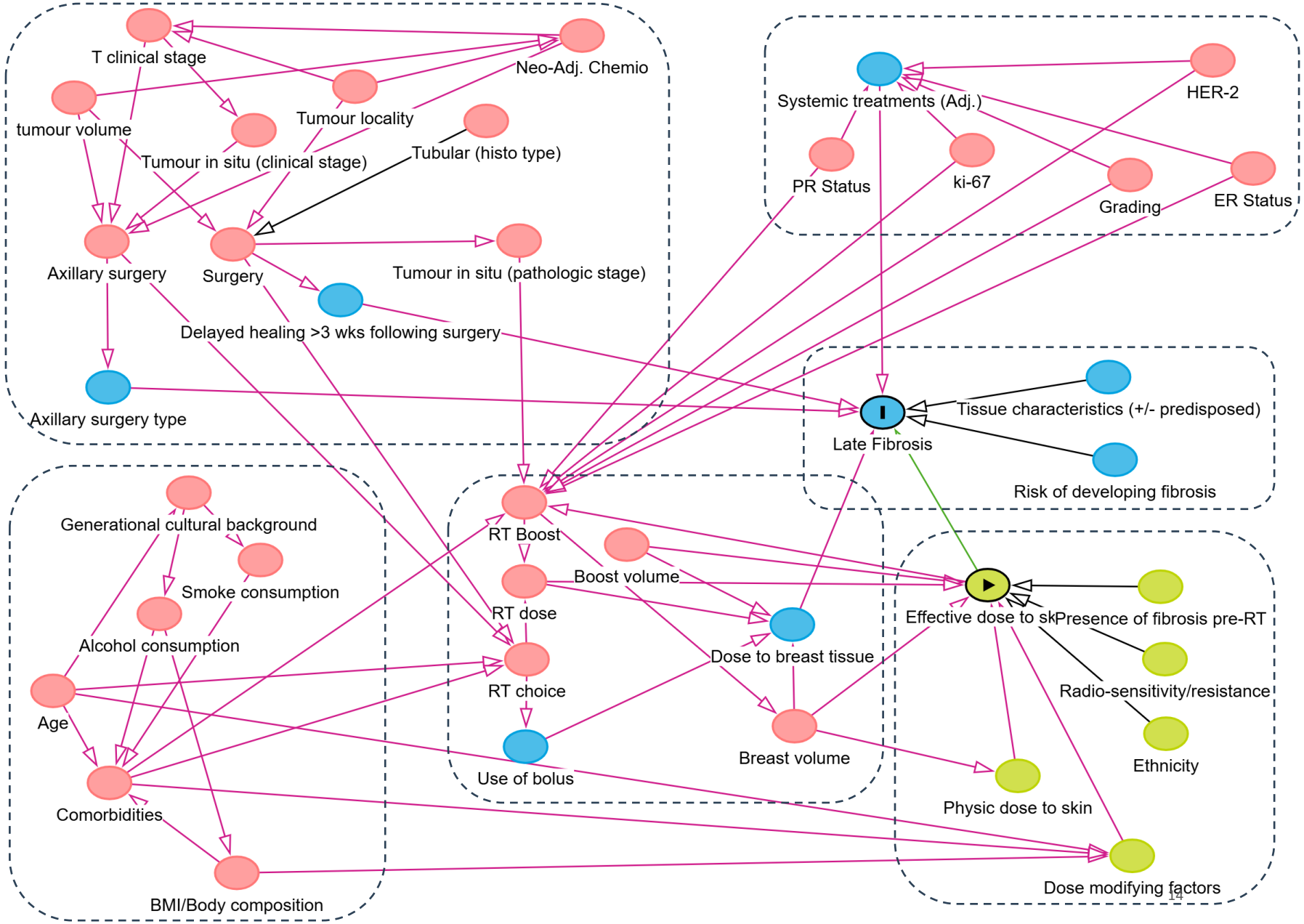
Data level

Causal model



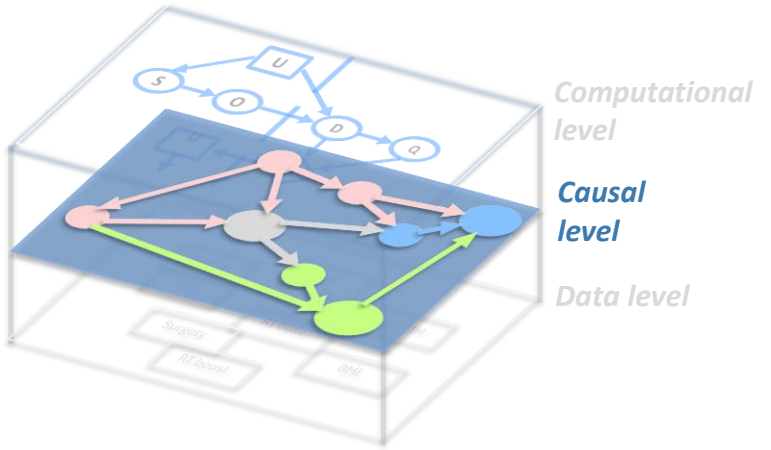
Causal level

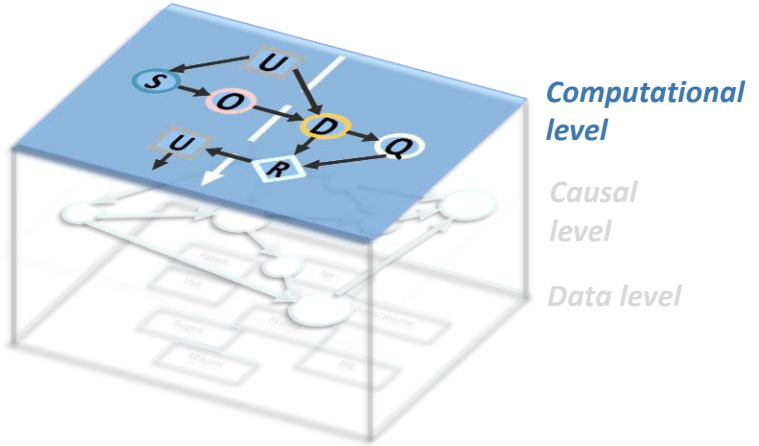
B. Dionisi Ferrara & T. Rancati, CMON Lab,Istituto Nazionale dei Tumori



- exposure
- outcome
- ancestor of exposure
- ancestor of outcome
- ancestor of exposure *and* outcome

Breast cancer patients: computational level





Observed variables $O^{pat}, O^{gen}, O^{tum}, O^{rad}$	Control inputs/actions U	Quantities of interest E	Outcome R
▪ ***Clinical data:*** *Age, BMI, comorbidities,…* ▪ ***Genomic data:*** *SNP genotypes, polygenic risk scores.* ▪ ***Imaging data:*** *DICOM-RT.*	▪ *RT prescribed dose* ▪ *RT boost* ▪ *RT choice*	▪ *Effective skin radiation dose*	▪ *Late fibrosis*

Digital Twin: PGM Overview

Semantic categories

S_t **Physical State:** State of the patient or of the disease.

O_t **Observational Data:** Available information describing the state of the patient.

U_t **Control Inputs:** Therapy decisions that influence the patient condition.

D_t **Digital State:** Configuration of the computational models comprising the digital twin.

Q_t **Quantities of Interest:** Quantities for monitoring the patient, estimated via model outputs.

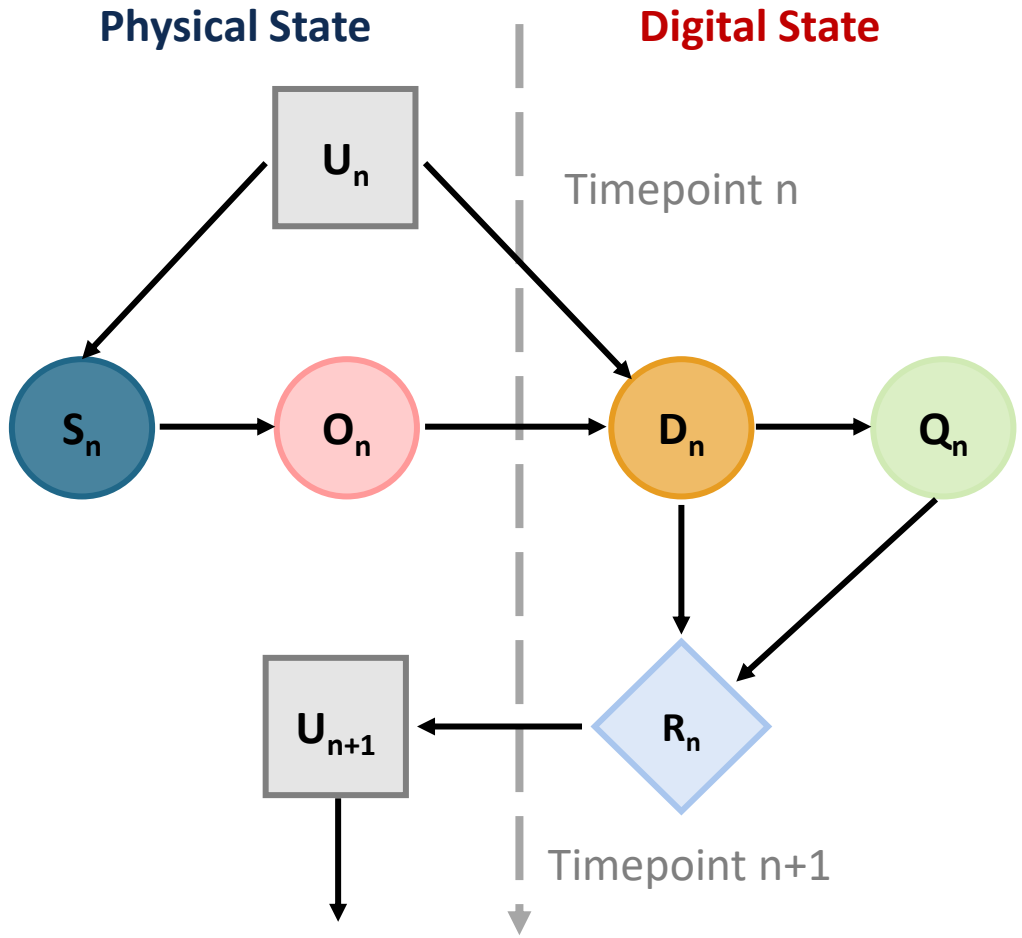
R_t **Reward:** Quantifies overall performance of the patient-twin system

National Academies publication "Foundational Research Gaps and Future Directions for Digital Twins,"

Unified representation of the asset-twin system

Mathematical/ Computational/ Statistical model

Probabilistic Graphical Models



Kapteyn M.G., Pretorius J.V.R., Willcox K.E. A probabilistic graphical model foundation for enabling predictive digital twins at scale (2021) Nature Computational Science, 1 (5), pp. 337 - 347

Quasi-static PGM: first attempt

We design a **quasi-static decision network** that exploits patient-specific personalized treatment

- **Observed variables**

$$O^{pat}, O^{gen}, O^{tum}, O^{rad}$$

- **Control inputs**

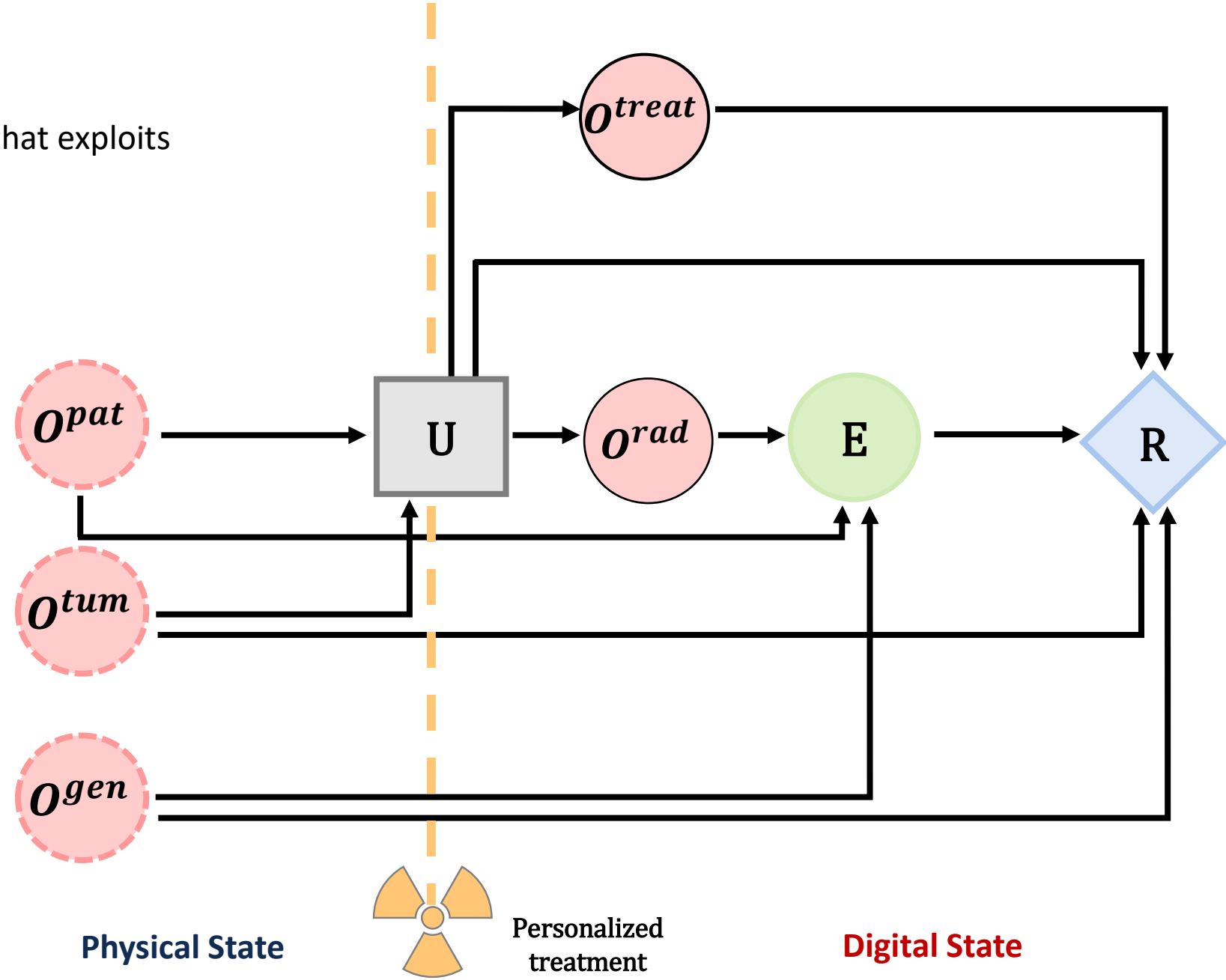
$$U$$

- **Quantity of interest**

$$E$$

- **Outcome**

$$R$$



Probabilistic approach to risk assessment

We propose a probabilistic decomposition of **late toxicity** risk into the following components.

Observed variables	Control inputs/actions	Quantities of interest	Outcome
$O^{pat}, O^{gen}, O^{tum}, O^{rad}$	U	E	$R = 0,1$

$$O^i = O^{pat}, O^{gen}, O^{tum}, O^{rad}$$

$$\begin{aligned} Toxicity &= \mathbb{P}(R=1|E,U,O^i=o^i) \\ &= \mathbb{P}(R=1|E)[\mathbb{P}(E|U)\cdot\mathbb{P}(U|o^i)+\mathbb{P}(E|o^i)] \\ &+ \mathbb{P}(R=1|U)\cdot\mathbb{P}(U|o^i) \\ &+ \mathbb{P}(R=1|o^i) \end{aligned}$$

Probabilistic modeling framework

$$
\begin{aligned}
Toxicity &= \mathbb{P}(R=1|E,U,O^i=o^i) \qquad O^i = O^{pat}, O^{gen}, O^{tum}, O^{rad} \\
&= \mathbb{P}(R=1|E)[\mathbb{P}(E|U)\cdot\mathbb{P}(U|o^i)+\mathbb{P}(E|o^i)] \\
&+ \mathbb{P}(R=1|U)\cdot\mathbb{P}(U|o^i) \\
&+ \mathbb{P}(R=1|o^i)
\end{aligned}
$$

Using the **Beta-Bernoulli** framework we have:

Prior

Beta-Bernoulli for toxicity outcome

$$R|E,U,O^i \sim \beta(a_0(E,U,O^i), b_0(E,U,O^i))$$

- *Retrospective*
- *Population informed*

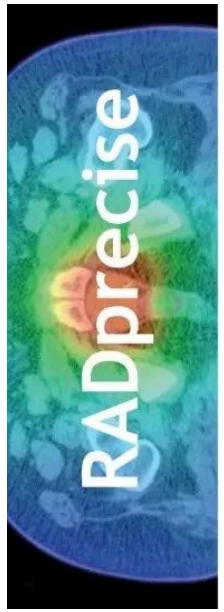
Posterior

Beta-Bernoulli for toxicity outcome

$$R|E,U,O^i \sim \beta(a_0 + \mathbb{I}[R=1], b_0 + \mathbb{I}[R=0])$$

- *Retrospective*
- *Population informed*

Training Data Leveraging Multi-Center Cohorts



REQUITE

- **Clinical covariates** (*B2a Breast patient factors baseline 2025-05-16 18-28-34.xlsx*)

$$O^{pat} = \mathbb{I}[\texttt{age} \geq 60 \texttt{ or } (\texttt{diabetes+hypertension+history of heart desease}) > 0]$$

- **Dosimetric covariates** (*B3 Breast clinical and treatment data form 2025-05-16 18-30-19.xlsx*)

$$E = \begin{cases} Low, & \text{if } \text{dose}_{\text{heart}} < 4\text{Gy and } \text{dose}_{\text{lung}} < 10\text{Gy} \\ Medium, & \text{otherwise} \\ High, & \text{if } \text{dose}_{\text{heart}} \geq 5\text{Gy and } \text{dose}_{\text{lung}} \geq 12.5\text{Gy} \end{cases}$$

- **Standardized toxicity outcomes (CTCAE v4.0)** (*B5 Health professional 2025-05-16 18-31.xlsx*)

$$R = \mathbb{I}[e \in \mathcal{T}_{breast} \subset \mathbb{R}^{13} : CTCAE(e) \geq 2]$$

- **Synthetic (preliminary) PRS score (proportional to to toxicity risk + noise)**

$$PRS = \omega R + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

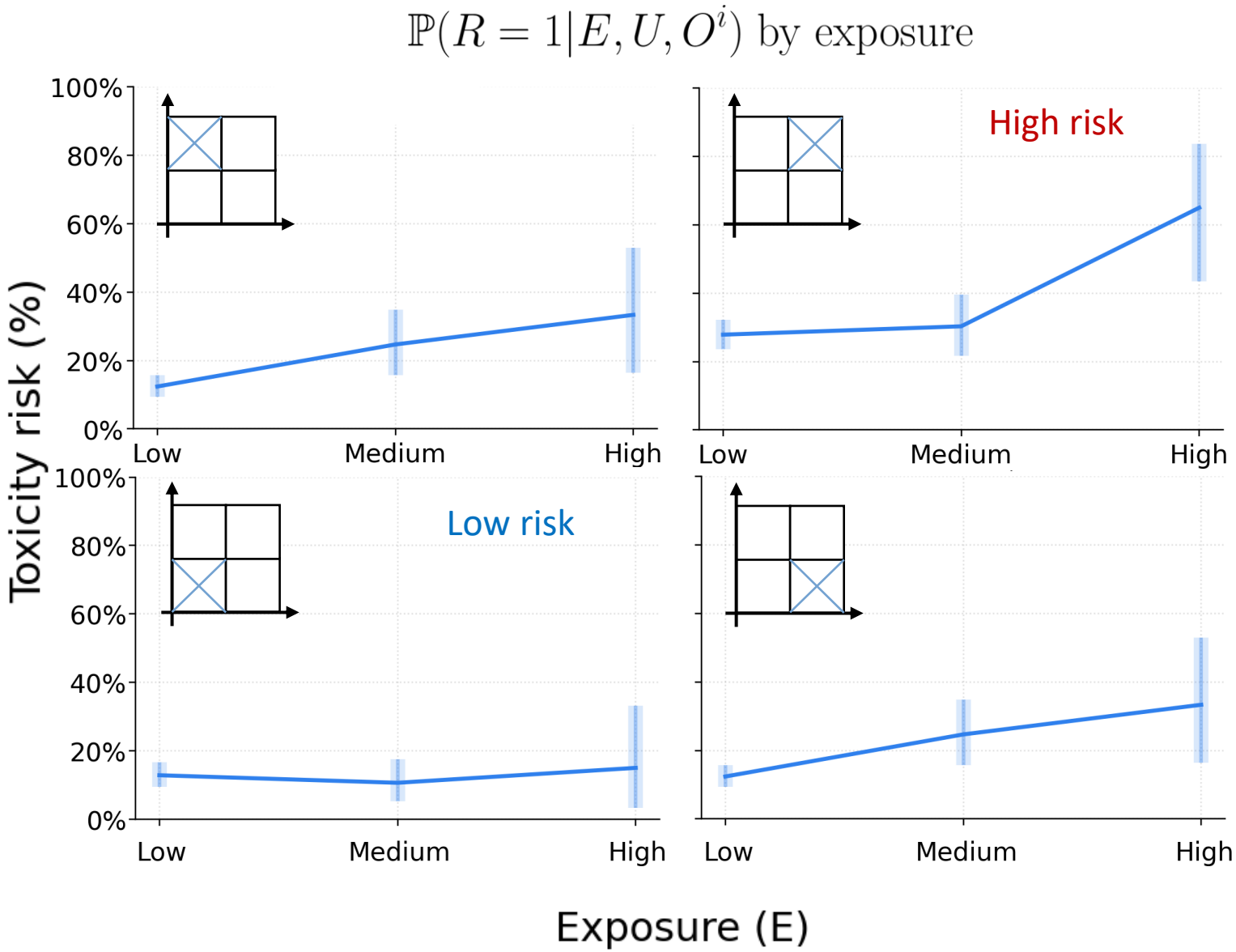
$$O^{gen} = \mathbb{I}[PRS > \overline{PRS}]$$

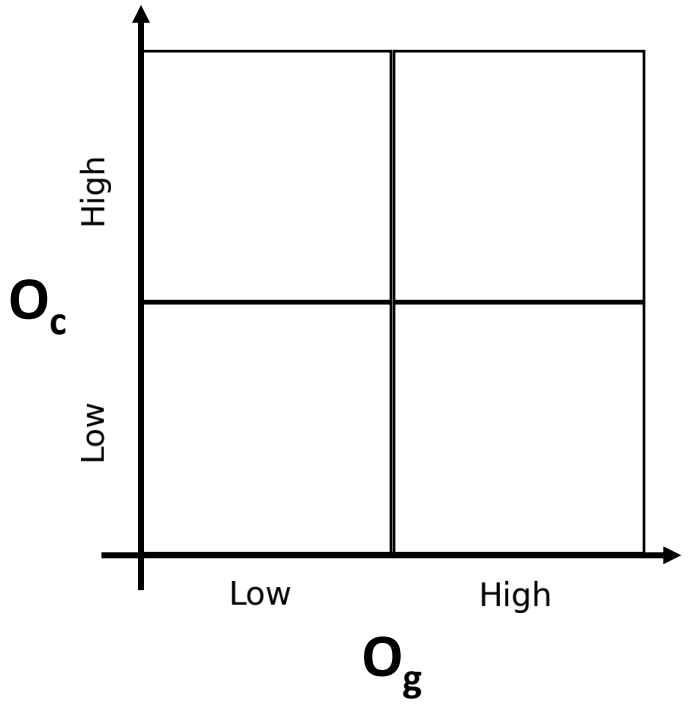
Results: toxicity table

Mean Toxicity Risk with 95% Credible Intervals

Exposure	Clinical Risk	Genomic Risk	n	Toxicity Risk (95% CI)
Low	Low	Low	325	12.8% (9.4–16.7)
Low	Low	High	287	15.9% (11.9–20.3)
Low	High	Low	434	12.4% (9.5–15.6)
Low	High	High	418	27.9% (23.7–32.2)
Medium	Low	Low	92	10.6% (5.3–17.6)
Medium	Low	High	100	26.5% (18.4–35.4)
Medium	High	Low	75	24.7% (15.8–34.8)
Medium	High	High	97	30.3% (21.7–39.7)
High	Low	Low	18	15.0% (3.4–33.1)
High	Low	High	23	36.0% (18.8–55.3)
High	High	Low	22	33.3% (16.4–52.9)
High	High	High	18	65.0% (43.5–83.7)

Results: toxicity plots



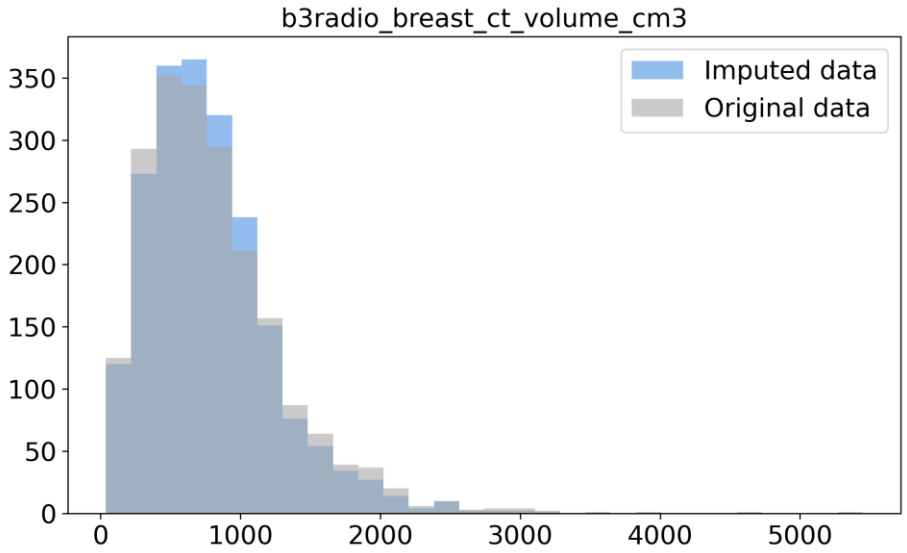


Next Steps in the Digital Twin Model

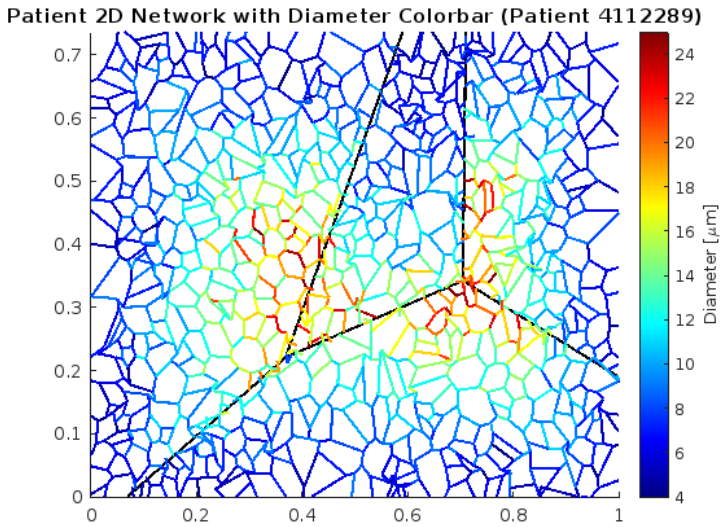
- **Real-time refinement** via Bayesian updates for **prior/posterior** distribution
- **Update** from quasi-static to **dynamic Digital Twin:**

$$R \rightarrow R_t$$

- **Imputation of** missing data with **Gaussian Processes**



- **Integrate the interaction** with **TCP** physical models for the **personalized microvasculature damage**



Acknowledgments to many collaborators

Paolo Zunino and Piermario Vitullo

Politecnico di Milano

Benedetta Dionisi Ferrara and Tiziana Rancati

Istituto Nazionale Tumori

Fondazione IRCCS
Istituto Nazionale dei Tumori