Practical Identifiability Failure in Physics-Constrained Cancer Models: The Therapeutic Controllability Index

Per Magnus Swedenborg[1,]

[1] DNAI Biotech Correspondence: per.swedenborg@dnai.bio | ORCID: 0009-0008-2750-9735

Abstract

Background: Physics-constrained machine learning aims to ground predictions in biological laws, theoretically enabling counterfactual treatment simulation. However, inferring dynamic drug sensitivity parameters () from static survival endpoints presents a severe inverse problem. We investigate whether pharmacological parameters are structurally or practically identifiable when longitudinal tumor volume data is absent.

Methods: We audited the identifiability of a production cancer digital twin (DNAI; VAE v5.10 + Neural ODE) trained on 9,393 patients from The Cancer Genome Atlas (TCGA). We formalized the identifiability limit via a didactic counterexample, demonstrating that under a threshold-crossing observation model with constant treatment, the parameter space collapses to a non-unique manifold. To address this, we developed the **Therapeutic Controllability Index (TCI)**, a composite diagnostic metric quantifying decision sensitivity via posterior and input perturbation analysis. We further applied a propensity-score-based Inverse Probability Weighting (IPW) pipeline to evaluate the stability of the inferred immune parameter .

Results: We report a practical identifiability failure: drug sensitivity () is under-identified, utilizing only a fraction of its constrained $[0, 1]$ range and exhibiting a gradient sensitivity ratio of 0.0017 relative to proliferation (). Consequently, the model defaults to a prognostic risk scorer driven by intrinsic biology, largely insensitive to treatment choice. However, the derived TCI successfully stratifies decision stability, distinguishing Therapeutically Controllable from Biologically Determined cohorts with clear separation in survival curves. IPW-adjusted analysis of the immune parameter removes naive treatment benefits in the pooled cohort, highlighting that as an outcome-trained latent variableinduces collider bias if used for stratification without rigorous deconfounding.

Conclusion: In the absence of longitudinal tumor volume data and treatment switching, physics-constrained models prioritize biological drivers over pharmacological effects. The TCI does not restore the identifiability of , but rather reframes this limitation into a clinical decision tool, identifying patients for whom standard therapy is mathematically insensitive to intervention.

---

1. Introduction

The integration of multi-omics data into Digital Twin frameworks promises to revolutionize precision oncology by simulating patient-specific responses to varying treatment schedules [1, 2]. A leading paradigm is **physics-constrained deep learning**, where neural networks map static patient data (genomics, histology) to the parameters of Ordinary Differential Equations (ODEs) governing tumor growth and inhibition [3, 12].

However, a fundamental tension exists between the richness of the model and the sparsity of

clinical training data. While the forward simulation (parameters survival) is well-posed, the inverse problem (survival parameters) is often ill-posed. In systems biology, this is known as **practical non-identifiability** [4]: multiple parameter combinations (e.g., high growth/high kill vs. low growth/low kill) can explain the same time-to-event outcome.

We posit that under **survival-only supervision**defined here as time-to-event labels without longitudinal tumor burden measurements and with coarse treatment encodingthe digital twin functions effectively as a *prognostic* risk model (predicting outcome regardless of treatment) rather than a *predictive* simulator (predicting response to specific treatments). Specifically, the loss function constrains only the *effective net growth rate*, creating a stiff parameter spectrum [9] where intrinsic proliferation () explains outcomes while the drug sensitivity parameter () collapses.

Our contribution proceeds in four steps: First, we formalize the structural unidentifiability of pharmacological parameters under static supervision (Section 3.2). Second, we empirically demonstrate this parameter collapse in a production model trained on 9,393 patients (Section 4.1). Third, we introduce the **Therapeutic Controllability Index (TCI)**, a metric that quantifies intervention headroomthe extent to which a patients outcome is mathematically steerable by treatment given their intrinsic biological aggression (Section 3.3). Finally, we audit the inferred immune parameter () using causal inference techniques to distinguish biological signal from confounding-by-indication (Section 3.4).

2. Related Work

Mechanistic Tumor Modeling: Classical mathematical oncology utilizes ODEs to model tumor growth inhibition [5, 6]. These models are typically calibrated on dense longitudinal data (e.g., RECIST measurements) [7]. Our work addresses the challenge of applying these laws in the data desert of retrospective clinical genomics (e.g., TCGA), where only static baselines and single endpoints are available. Recent perspectives emphasize the gap between digital twin theory and data reality [1, 2].

Identifiability in Systems Biology: Structural identifiability refers to the theoretical uniqueness of parameters given perfect data [14, 15]. Raue et al. [4] distinguish this from *practical identifiability*, which concerns recoverability given noisy or sparse data. While profile likelihood methods [19] are standard for low-dimensional systems, we extend this analysis to deep learning-based parameter inference, aligning with the concept of sloppy parameter spectra [9].

Causal Inference & Uncertainty: Estimating Conditional Average Treatment Effects (CATE) from observational data is a core ML focus [10, 11]. We leverage Monte Carlo Dropout [20] for approximate Bayesian inference to estimate parameter uncertainty. Furthermore, we incorporate Inverse Probability Weighting (IPW) methods [13, 21] to correct for treatment assignment bias, specifically addressing the confounding-by-indication common in immunotherapy cohorts.

3. Methods

3.1. The DNAI Architecture

The system utilizes a dual-paradigm architecture comprising a frozen Foundation Model and a trainable Hypernetwork.

Foundation Model (VAE v5.10): A hierarchical Variational Autoencoder compresses multi-omics data (RNA, DNA, CNV, Methylation) into a 328-dimensional latent space (z).

Hypernetwork & Emulator: A specialist Hypernetwork maps z and Whole Slide Imaging (WSI) embeddings to ODE parameters. To accelerate training and enable gradient-based optimization, we utilize the analytical solution (SimpleEmulator) of the underlying Lotka-Volterra dynamics in the log-linear growth regime. The emulator equation is:

$$x(t) = (\ _{eff}(t)\ )\ t + x_0$$

Where: * $x(t) = \log V(t)$: Log tumor burden relative to detection threshold. * $[0, 0.3]$: Intrinsic tumor growth rate (day^1). * > 0: Aggregate immune pressure (inferred from baseline covariates, not observed). * x_0: Initial log tumor volume (inferred parameter).

In plain English: under constant treatment, the log-tumor burden evolves approximately linearly with a slope equal to the intrinsic growth rate minus the net kill rate.

Treatment Encoding: The Hypernetwork outputs a vector R^D representing sensitivity to $D = 5$ drug classes (Chemotherapy, Immunotherapy, Targeted, Hormone, Radiation).

$$_{eff}(t) = \sum_{d=1}^{D} {}_d\ u_d(t)$$

Here, $u_d(t)$ $\{0, 1\}$ is the binary treatment status for class d. In TCGA, $u_d(t)$ is usually time-constant within a patient (binary assignment), limiting the excitation of dynamics required to disentangle parameters.

3.2. Structural Unidentifiability Analysis

We formally characterize the identifiability limit. Structural identifiability requires the map $T()$ to be injective under the observation operator T.

Proposition 1 (Didactic Counterexample): Consider the system state $x(t;)$ governed by $x = u$ with initial condition x_0. The observable output is the time-to-event $T = \inf\{t \quad x(t) \quad x_{lethal}\}$. **Assumptions:** 1. Treatment $u(t)$ 1 is constant. 2. Initial volume x_0 and lethal burden x_{lethal} are known constants. 3. Parameters $= [, ,]^T$ are constrained to a hypercube C R^3.

Claim: Under these assumptions, is structurally unidentifiable.

Proof: The time-to-event is given analytically by:

$$T() = \frac{x_{lethal}\ x_0}{}$$

Let $_{eff} =$ be the effective net growth rate. The observation T uniquely determines $_{eff}$ (assuming $x_{lethal} > x_0$ are known), defining a plane in parameter space:

$$= C$$

where $C = (x_{lethal}\ x_0)/T_{obs}$. The set of valid solutions is the intersection of this plane with the constraint hypercube C. For any admissible C where the intersection is not a single vertex, the level set contains a continuum of solutions. Thus, there exist infinitely many distinct vectors $_1, _2$ C such that $T(_1) = T(_2)$. Specifically, one can increase intrinsic growth and drug sensitivity simultaneously while maintaining the same net growth rate $_{eff}$.

Summary: A single event time identifies only the net rate , not the individual components.

Remark 1 (Real-World Implications): In the actual DNAI training scenario, x_0 is unknown (inferred) and data is right-censored. This strictly *enlarges* the unidentifiable manifold described in Proposition 1. If the parameters are unidentifiable under the idealized assumptions of known x_0 and observed event times, they remain unidentifiable in the noisier real-world setting. Given this limit, we should not interpret as a patient-specific drug response parameter in TCGA-like settings. Next, we introduce TCI to manage this uncertainty.

3.3. Therapeutic Controllability Index (TCI)

To reframe the identifiability collapse as a clinical feature, we define TCI. This composite score $[0, 1]$ estimates the susceptibility of the tumor trajectory to treatment modulation.

Design Intent: We use TCI as an abstention/triage signal: low TCI indicates recommendations are mathematically weakly determined. It aggregates four dimensions: 1. **Mechanistic Headroom (C_{dose}):** Can the drug theoretically overcome the tumors growth rate? 2. **Counterfactual Spread (C_{auth}):** Does changing the treatment actually change the predicted outcome? 3. **Decision Stability (C_{stab}):** Is the best treatment consistent across Monte Carlo parameter samples? 4. **Local Robustness (C_{rob}):** Is the prediction stable under small perturbations of the latent input?

Algorithm 1: TCI Computation Input: Patient latent z, Hypernetwork H, MC samples $M = 1{,}000$. **Output:** TCI $[0, 1]$.

1. **Generate Posterior:** Sample parameters $_m$ $H(z)$ for $m = 1..M$ using MC Dropout [20] (dropout active, BatchNorm frozen).
2. **Compute Components:**
   - **Dose Control (C_{dose}):**

$$C_{dose} = \frac{1}{M} \sum_{m=1}^{M} \text{clip}\left(\frac{\max(_m)\ D_{max}}{_m}, 0, 1\right)$$

     Note: $D_{max} = 1.0$ represents maximum tolerated dose intensity.
   - **Treatment Authority (C_{auth}):** Let $R(u_d)$ be the predicted 3-year mortality risk if treated with drug class d U.

$$C_{auth} = \text{sigmoid}\left(\text{StdDev}_{dU}[E_m[R(u_d)]] \text{ Œ } 10\right)$$

     Where $U = \{$Chemo, IO, Targeted, Hormone, Radiation$\}$.
   - **Decision Stability (C_{stab}):** Let $u_m = \arg\min_{uU} R_m(u)$.

$$C_{stab} = \frac{1}{M} \sum_{m=1}^{M} I(u_m = \text{Mode}(u))$$

   - **Robustness (C_{rob}):** Add noise $N(0, 0.1)$ to standardized z.

$$C_{rob} = \exp\left(\frac{|E_m[R(z)]\ E_m[R(z+)]|}{E_m[R(z)] + 1e\ 6}\right)$$

3. **Aggregate:**

$$TCI = 0.25C_{dose} + 0.30C_{auth} + 0.25C_{stab} + 0.20C_{rob}$$

3.4. Causal Inference Pipeline

Even if is not identifiable, one might still be tempted to interpret other learned mechanistic parameters (e.g.,). We therefore audit with IPW to isolate true effects from confounding-by-indication.

Assumptions & Diagnostics: * **Unconfoundedness:** We assume treatment assignment is independent of potential outcomes given observed covariates (X). * **Positivity:** We require $0 < P(T = 1|X) < 1$. * **Balance Check:** We evaluate covariate balance using Standardized Mean Differences (SMD), targeting SMD < 0.1 post-weighting [16].

Procedure: 1. **Propensity Estimation:** We fit a Logistic Regression model (class_weight=balanced) to estimate $P(T = 1|X)$ where T is Immunotherapy (IO) assignment and X includes immune_score, log_TMB, purity, proliferation_score, aneuploidy, and cancer type. 2. **Weight Calculation:** Propensity scores $e(X)$ are clipped to $[0.05, 0.95]$. Stabilized weights w_i are computed following [13, 21]:

$$w_i = \frac{P(T = 1)}{e_{clip}(X_i)} \; I(T_i = 1) + \frac{P(T = 0)}{1 \; e_{clip}(X_i)} \; I(T_i = 0)$$

3. **Estimand:** We report the **Restricted Mean Survival Time (RMST)** difference at 5 years [18]. 4. **Interaction Analysis:** To test if is a predictive biomarker, we stratify patients by quartiles and compute the IPW-adjusted RMST difference (IO vs. Non-IO) within each stratum.

4. Experiments & Results

Dataset: TCGA Pan-Cancer Atlas ($N = 9{,}393$, 33 cancer types). **Model:** Trained on a random 85/15 split (seed=42; 7,985 training, 1,408 validation).

4.1. The Practical Identifiability Failure

Our audit confirmed the theoretical prediction: collapses under survival-only supervision.

Parameter Statistics (Full Cohort, N=9,393):

Parameter	Mean	Std	Range
(growth)	0.0313	0.0157	[0.0013, 0.1346]
(drug sensitivity)	7.0 Œ 10^5	9.2 Œ 10^5	[0.0, 0.00143]
(immune kill)	0.0082	0.00144	[0.0041, 0.0140]

- **Sensitivity Hierarchy:** The loss function is dominated by proliferation. $S_{/} = 0.0017$, indicating the model is ~588x more sensitive to than .
- **Range Collapse:** While is constrained to $[0, 1]$, predicted values clustered in $[0.0, 0.00143]$.
- **Interpretation:** This empirical range collapse is the practical signature of the non-identifiable manifold described in Proposition 1.

4.2. TCI Stratification and Incremental Value

Despite the collapse of the raw parameter, the composite TCI metric successfully recovered clinical signal. We stratified patients ($N = 9{,}393$) based on their TCI scores into **Controllable** (TCI 0.6), **Moderate**, and **Determined** (TCI < 0.3).

Table 1: Survival Characteristics by TCI Group ($N = 9{,}393$)

Group	N (%)	Median OS (Days)	Event Rate	Mean TCI
Controllable (0.6)	1,774 (18.9%)	993	10.3%	0.632
Moderate	6,957 (74.1%)	699	32.2%	0.454
Determined (< 0.3)	662 (7.0%)	458.5	64.5%	0.276

TCI C-index = 0.735. Controllable patients receiving treatment: median OS 1,153 days vs Determined patients receiving treatment: 489 days ($P = 4.3$ Œ 10^{55}, Mann-Whitney U).

Addressing the Tautology: A key concern is whether TCI is simply a proxy for prognostic risk. 1. **Correlation:** TCI correlates with overall survival ($r = 0.274$), reflecting the biological reality that aggressive tumors () are harder to control. 2. **Distinct Dimension:** However, TCI captures a distinct *decision-sensitivity* dimension. Controllable patients have treatment authority scores of 0.847 (high spread across treatment classes), compared to 0.091 for Determined patients. This confirms TCI identifies patients where the model sees a meaningful difference between interventions, not just patients who live longer.

4.3. Case Study: Deconfounding the Immune Signal

We investigated the immune kill parameter to determine if it captured biological signal or treatment assignment bias.

1. Propensity & Balance: The logistic regression model achieved an AUC of 0.933 in predicting IO assignment, confirming strong selection bias. IPW weighting successfully reduced the maximum Standardized Mean Difference (SMD) across covariates to below the 0.1 diagnostic threshold [16].

2. Effect Modification (Interaction Analysis): We stratified patients by quartiles and computed the IPW-adjusted RMST difference (IO vs. Control).

Table 2: IPW-Adjusted Immunotherapy Benefit by Stratum ($N_{IO} = 170$)

Stratum	N_{IO} / $N_{control}$	Naive IO Benefit (Days)	IPW-Adjusted IO Benefit (Days)	Interpretation
Low	90 / 4,606	+514	**+313**	Benefit
High	80 / 4,617	+613	**231**	Harm
All	170 / 9,223	+539.5	26	Null

Interaction magnitude: 544 days (low high). Naive benefit is entirely explained by confounding-by-indication (propensity AUC = 0.933).

Conclusion: The pattern is compatible with the hypothesis of effect modification. Patients with *low* intrinsic immune pressure () appear to derive the most benefit from checkpoint blockade, while patients with *high* show no benefit or potential harm. However, we caution that is a learned latent parameter trained on outcomes; stratifying on it carries a risk of collider bias if absorbs unobserved treatment effects.

5. Discussion: Guidelines for Digital Twin Builders

Our findings challenge the assumption that physics constraints alone enable mechanistic inference from sparse clinical data. We offer three guidelines for the community:

1. **The Constraint Trap:** Do not assume that because parameters are bounded (e.g., $[0, 1]$), they are identifiable. As Proposition 1 shows, constraints merely bound the solution manifold; they do not collapse it to a point.
2. **Minimum Viable Supervision:** To learn valid pharmacological parameters, survival data is insufficient. Models must incorporate **longitudinal tumor burden** (at least two timepoints) and **treatment switching** (on/off periods) to break the symmetry between intrinsic growth and drug sensitivity.
3. **Fail Gracefully with TCI:** When identifiability fails, do not report raw parameters as personalized drug sensitivity. Instead, compute a stability metric like TCI. If TCI is low, the model is effectively stating: The outcome is biologically determined; treatment choice is mathematically irrelevant.

6. Limitations

1. **Retrospective Nature:** This study relies on retrospective TCGA data; TCI has not been validated in a prospective trial.
2. **Positivity Violation:** The limited overlap in propensity scores restricts the precision of causal estimates in the tails of the distribution.
3. **Treatment Proxy:** The binary treatment encoding in the ODE is a simplification; it does not capture dose intensity or combination schedules.

7. Conclusion

We report that drug sensitivity is practically unidentifiable in physics-constrained cancer models trained on survival data alone. However, this failure mode is informative. The Therapeutic Controllability Index extracts value from this limitation, stratifying patients by their potential to benefit from treatment. Future work will focus on integrating longitudinal imaging (RECIST) and PDX data to restore the identifiability of pharmacological parameters.

Ethics Statement

This study exclusively utilized de-identified, publicly available retrospective data from The Cancer Genome Atlas (TCGA). All data were accessed in accordance with the TCGA data use agreement. No patient contact, intervention, or collection of new human biological material was performed. As all data were previously collected, de-identified, and publicly released under institutional review, additional IRB approval was not required per the Common Rule (45 CFR 46.104(d)(4)).

Author Contributions (CRediT)

P.M.S.: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing Original Draft, Visualization.

Competing Interests

P.M.S. is the founder of DNAI Biotech and has a financial interest in the commercialization of the DNAI platform. P.M.S. is the inventor on provisional patent applications related to the DNAI platform.

---

8. References

[1] Stahlberg, E. A., et al. (2022). Exploring the future of cancer digital twins. *Nature Cancer*, 3, 533535. [2] Hernandez-Boussard, T., et al. (2021). The future of digital twins in precision medicine. *Nature Medicine*, 27, 20152016. [3] Chen, R. T. Q., et al. (2018). Neural Ordinary Differential Equations. *NeurIPS*. [4] Raue, A., et al. (2009). Structural and practical identifiability analysis. *Bioinformatics*. [5] Simeoni, M., et al. (2004). Predictive pharmacokinetic-pharmacodynamic modeling. *Cancer Research*. [6] Ribba, B., et al. (2012). A tumor growth inhibition model for low-grade glioma. *Clin Cancer Res*. [7] Benzekry, S., et al. (2014). Classical Mathematical Models for Description of Tumor Growth. *PLoS Comp Bio*. [8] Villaverde, A. F. (2019). Observability and structural identifiability. *Complexity*. [9] Gutenkunst, R. N., et al. (2007). Universally sloppy parameter sensitivities. *PLoS Comp Bio*. [10] Shalit, U., et al. (2017). Estimating individual treatment effect. *ICML*. [11] Wager, S., & Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects. *JASA*. [12] Raissi, M., et al. (2019). Physics-informed neural networks. *J Comp Phys*. [13] Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. [14] Bellman, R., & Åström, K. J. (1970). On structural identifiability. *Mathematical Biosciences*. [15] Ljung, L., & Glad, T. (1994). On global identifiability for arbitrary model parametrizations. *Automatica*. [16] Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*. [17] Walter, E., & Pronzato, L. (1997). *Identification of Parametric Models from Experimental Data*. Springer. [18] Royston, P., & Parmar, M. K. (2013). Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Medical Research Methodology*. [19] Raue, A., et al. (2013). Lessons learned from quantitative dynamical modeling in systems biology. *PLoS One*. [20] Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *ICML*. [21] Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*.