The Plural Twin: Quantifying Treatment Policy Stability via Set-Valued Cancer Digital Twins

Per Magnus Swedenborg[1,]

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

Abstract

Background: Deep learning in precision oncology typically functions as a deterministic oracle, mapping patient data to single-point treatment recommendations. This paradigm obscures epistemic uncertainty, yielding fragile policies where the optimal treatment rank shifts under equally plausible model realizations. High discrimination (C-index) does not guarantee decision stability. **Methods:** We introduce Plural Twins, a set-valued framework representing each patient as a distribution of potential utility outcomes. We utilize Monte Carlo dropout ($p = 0.3$) with frozen Batch Normalization statistics to generate 1,000 stochastic realizations per patient. Using a multi-modal architecture (VAE + Hypernetwork) on 9,393 patients, we evaluate policy stability and implement a robust ranking protocol that optimizes for the conditional value at risk (CVaR) of the pairwise win distribution. **Results:** We observe that 82.9% of patients exhibit policy instability, where the top-ranked treatment fluctuates across stochastic draws. Consequently, the robust policy disagrees with the mean-optimal prediction in 16.7% of cases. We identify a Glass Cannon phenotypetreatments with high expected utility but catastrophic tail riskin 7.0% of the cohort. In an observational evaluation of treated patients, concordance with the robust policy is associated with significantly improved overall survival (929d vs 762d, $p = 3.5$ Œ 10^{10}). **Conclusion:** Point-estimate rankings mask critical decision fragility. By operationalizing uncertainty via Plural Twins, we demonstrate that for one in six patients, the optimal treatment depends on the algorithms risk tolerance. We propose a clinical protocol to translate stability scores into actionable guidance, enforcing abstention when model uncertainty exceeds decision thresholds.

---

1. Introduction

The translation of artificial intelligence into precision oncology has largely followed a deterministic oracle paradigm: a model ingests high-dimensional patient data and outputs a single predicted trajectory or treatment score [1]. While recent advances in multi-modal integration have improved the discrimination of these point estimates [2], they fundamentally obscure the epistemic uncertainty inherent in deep neural networks. In clinical practice, a treatment recommendation that is the best on average but highly sensitive to model noise is dangerous. Standard metrics like the Concordance Index (C-index) quantify discrimination (ranking patients by risk) but fail to quantify **decision stability** (ranking treatments for a specific patient) [3].

This creates a reliability gap. A digital twin might predict that a patient will respond best to Immunotherapy based on a single forward pass. However, if 40% of the approximate posterior distributionderived from the same training datapredicts a better response to Chemotherapy, the policy is unstable. Ignoring this variance exposes patients to fragile optimization, where the chosen therapy has a high expected benefit but a catastrophic worst-case outcome under model uncertainty.

We introduce the **Plural Twin** framework, which redefines the cancer digital twin from a single trajectory to a set-valued distribution. By generating stochastic samples per patient [4], we quantify

the stability of treatment policies. Unlike prior work focusing on predictive intervals for survival time [5] or selective classification [6], we focus on **policy stability**: does the rank-ordering of treatments change across the models uncertainty landscape?

To address the noise inherent in raw logit comparisons, we introduce a **CVaR-Borda** ranking protocol. This method aggregates pairwise win probabilities between treatments across the uncertainty distribution, optimizing for the worst-case Borda count [7]. Applying this to The Cancer Genome Atlas (TCGA), we reveal that while absolute mechanistic parameters are often under-identified, the *relative stability* of treatment rankings serves as a potent biomarker for therapeutic response.

2. Related Work

Uncertainty in Deep Learning. Bayesian approximation in deep learning is critical for safety-critical applications. While **Deep Ensembles** [8] provide gold-standard calibration, their computational cost is prohibitive for large-scale digital twin optimization. **Laplace Approximations** [9] and **SWAG** [10] offer post-hoc posterior estimation but require specific curvature assumptions. We utilize **Monte Carlo (MC) Dropout** [4] as a pragmatic compromise; while it may underestimate total variance compared to ensembles, it captures the *relative* epistemic uncertainty required for ranking stability at low latency. Recent work in **Conformal Prediction** [11] has focused on generating valid prediction sets, a concept we extend to set-valued policies.

Rank Aggregation & Risk Measures. Combining multiple ranked lists into a single robust ranking is a well-studied problem in social choice theory. In the context of uncertainty, stochastic dominance and probabilistic Borda counts offer a way to smooth out noise in pairwise comparisons [12]. We integrate this with **Conditional Value at Risk (CVaR)** [7], a coherent risk measure widely used in finance, to penalize treatments that perform poorly in the tail of the posterior distribution.

Observational Evaluation. Estimating Individualized Treatment Effects (ITE) from observational data is fraught with bias, including immortal time bias [13] and selection bias [14]. While methods like TARNet [15] attempt to correct for these, they rely on assumptions of no unmeasured confounding. In this work, we frame our validation as an **observational concordance check**, distinct from a causal trial, and employ rigorous time-dependent evaluation to mitigate guarantee-time bias.

3. Methods

3.1 Data and Cohort

We utilized the DNAI V3.1 dataset comprising 9,393 patients across 33 cancer types from TCGA. * **Omics:** RNA-seq (2,579 genes), DNA mutations (500 drivers), CNV (1,886 segments), and Methylation (1,000 probes) encoded via a Variational Autoencoder (VAE v5.10) into a 328-dimensional latent space (z_{full}). * **Histology:** Whole Slide Images (WSI) processed via UNI2-h (1,536-dim embeddings). * **Outcome:** Overall Survival (OS) censored at last follow-up.

3.2 Treatment Taxonomy

We standardized clinical interventions into six mutually exclusive classes (t $\{0..5\}$). To handle multi-modal treatment histories, we applied a hierarchical First-Line Dominance rule based on clinical impact: 1. **Targeted Therapy (Class 5):** Any record of targeted agents (e.g., Trastuzumab, Gefitinib). 2. **Immunotherapy (Class 3):** Any checkpoint inhibitors or cytokines.

3. **Chemotherapy (Class 1):** Cytotoxic agents. 4. **Radiation (Class 2):** External beam or brachytherapy. 5. **Hormone Therapy (Class 4):** Endocrine therapy (e.g., Tamoxifen). 6. **None/Baseline (Class 0):** No systemic therapy recorded.

This hierarchy ensures that a patient receiving Chemo + Targeted is classified as Targeted, reflecting the precision component of the regimen.

3.3 Model Architecture

We employ the **ConditionedGatingModelV3** (Path A Specialist), a hypernetwork architecture that maps patient latents to mechanistic parameters of a governing Ordinary Differential Equation (ODE). * **Inputs:** VAE Latents (z_{full} R^{328}) + WSI (h_{wsi} R^{1536}) + Clinical Covariates. * **Outputs:** The model predicts a set of physics-constrained parameters: tumor growth rate $[0, 0.3]$, immune kill efficiency > 0, and a drug sensitivity vector $[0, 1]^6$. * **Utility Definition:** For the purpose of treatment ranking, we define the utility $\mathbf{u}_{i,t}$ of treatment t for patient i as the predicted drug sensitivity coefficient $_{i,t}$. In the ODE dynamics, represents the fractional kill rate of the tumor burden; thus, maximizing is equivalent to maximizing tumor reduction.

3.4 The Plural Twin: MC Dropout Protocol

We generate the Plural Twin distribution using Monte Carlo (MC) dropout at inference time. Crucially, we maintain Batch Normalization (BN) statistics fixed to the training population to ensure variance arises solely from epistemic uncertainty, not batch-level noise.

Protocol: 1. **Model Mode:** Set global model to `eval()` (freezes BN running statistics). 2. **Dropout Activation:** Manually set `Dropout` layers to `train()`. 3. **Configuration:** Dropout rate $p = 0.3$ (fixed, consistent with training). 4. **Sampling:** For each patient x_i, perform $K = 1,000$ forward passes to obtain a set of utility vectors $U_i = \{_i^{(k)}\}_{k=1}^{K}$, where $_i^{(k)}$ R^6 represents the predicted sensitivity for all classes.

3.5 Stability Metrics

We define the **Stability Score (s_i)** as the consistency between individual stochastic draws and the mean-optimal treatment.

First, compute the mean utility vector: $_i = \frac{1}{K} {}_{k=1}^{K} {}_i^{(k)}$. Identify the mean-optimal treatment: $t_i = \arg\max_{t\{0..5\}} {}_{i,t}$.

The stability score is the fraction of draws where the top-ranked treatment matches t_i:

$$s_i = \frac{1}{K} {}_{k=1}^{K} I\left(\arg \max_{t\{0..5\}} {}_{i,t}^{(k)} = t_i\right)$$

A patient is defined as **Unstable** if $s_i < 0.5$. This threshold was selected because in a 6-class problem, a majority vote (> 0.5) is required to guarantee a unique Condorcet winner in the limit of infinite samples.

3.6 Robust Ranking: CVaR-Borda

To robustify the decision, we employ a Borda Count method on the pairwise win probabilities. We selected $= 0.05$ to penalize tail risks, prioritizing safety over average-case gain. This aligns with the

standard statistical significance threshold ($p < 0.05$), treating the bottom 5% of model realizations as the worst-case scenario to guard against.

Step 1: Pairwise Advantage. For each MC sample k, compute $I_{ab}^{(k)} = I(^{(k)}_{i,t_a} > ^{(k)}_{i,t_b})$ for all pairs in $\{0..5\}$. **Step 2: Borda Score.** $S_t^{(k)} = _{tt} I_{tt}^{(k)}$. This score ranges from 0 to 5. **Step 3: CVaR Optimization.** We select the treatment maximizing the **Conditional Value at Risk (CVaR)** of the Borda score at $= 0.05$. Since S_t is discrete, Q is the largest integer q such that $P(S_t \ q)$.

$$_{CVaR}(x_i) = \arg\max_{t\{0..5\}} (E[S_t \ S_t \ Q(S_t)])$$

Ties in CVaR are broken by the mean Borda score.

3.7 Evaluation Strategy

To mitigate confounding by indication, we employ an **Observational Concordance** evaluation: 1. **Treated Cohort:** Analysis includes all patients with a documented treatment record across 33 cancer types ($N = 4{,}783$; 50.9% treatment coverage). 2. **Concordance Definition:** A patient is **Concordant** if their clinically administered treatment class matches the Top-1 recommendation of the robust policy ($_{CVaR}$). Otherwise, they are **Discordant**. 3. **Adjustment:** We employ a Cox Proportional Hazards model adjusted for age, stage, and tumor mutation burden (TMB) to verify that survival differences are not driven by baseline prognostic factors.

4. Experiments & Results

4.1 Identification Audit

Before evaluating policy stability, we audited the models ability to recover mechanistic parameters. We found that the ratio of the learned drug sensitivity parameter () to the tumor growth rate () was extremely low (Median / = 0.0017). This implies that the inferred drug kill rate is less than 0.2% of the tumor growth rate, a biologically implausible value for effective therapies. This confirms that absolute mechanistic ODE parameters are under-identified when supervised solely by time-to-event data. However, while the *absolute* magnitude of is uncalibrated, the *relative* ranking of values across treatments remains discriminative. Consequently, we restrict all subsequent claims to the **relative ranking stability**.

4.2 Prevalence of Instability

Analyzing 9,393 patients with $K = 1,000$ draws, we found that instability is the norm rather than the exception. **82.9%** of patients had a Stability Score $s_i < 0.5$, meaning that for the vast majority of the cohort, the optimal treatment choice fluctuates depending on the specific dropout mask applied.

To assess the relationship between stability and outcome, we performed a univariate Cox regression. We found that higher stability is significantly associated with reduced hazard of death (Hazard Ratio = 0.84 per 0.1 increase in s_i, $p < 0.001$). This suggests that instability is not merely model noise but is associated with biologically chaotic, aggressive disease states that are inherently difficult to treat or predict.

4.3 Discordance, Glass Cannons, and the Benefit-Fragility Quadrant

We compared the standard Mean-Optimal policy ($_{mean}$) against the Robust CVaR-Borda policy ($_{CVaR}$).

- **Discordance:** In **16.7%** of patients (1,571/9,393), the robust policy selected a different treatment class than the mean-optimal policy. These are cases where the treatment with the highest average utility has a heavy tail of poor outcomes in the posterior distribution. CVaR-Borda discordance was lower at **8.3%** because pairwise comparisons cancel some confounding noise.
- **Glass Cannon Phenotype:** We define a **Benefit-Fragility quadrant** analysis using CATE-derived treatment benefit (expected survival gain) and perturbation-derived fragility (sensitivity to dose, parameter, and treatment substitution):

Quadrant	Definition	N (%)	Median OS	Event Rate
Q1: Solid Responder	High benefit, Low fragility	4,043 (43.0%)	830d	19.5%
Q2: Glass Cannon	High benefit, High fragility	653 (7.0%)	478d	71.1%
Q3: Stable Non-Responder	Low benefit, Low fragility	4,154 (44.2%)	699d	29.2%
Q4: Fragile Non-Responder	Low benefit, High fragility	543 (5.8%)	350d	70.7%

The Glass Cannon phenotype (Q2) identifies patients for whom a standard model would confidently recommend a therapy that fails in a significant fraction of posterior realizations. These patients have high expected treatment benefit but catastrophic outcomes when the model is wrong median OS 478 days with 71.1% event rate, compared to 830 days and 19.5% for Solid Responders (Q1). Glass Cannon cancers are concentrated in KIRC (41.9% Q2), SKCM (23.5%), STAD (14.0%), and MESO (12.6%).

- **Fragility as Independent Axis:** Treatment fragility alone achieves C-index **0.740**, demonstrating it captures an independent prognostic axis orthogonal to expected benefit (base risk C=0.856). Fragility-adjusted policy changes **30.9%** of treatment recommendations compared to the mean-optimal policy.
- **Archetype Interaction:** Jacobian-derived mechanistic archetypes (proliferation-driven vs. signaling-driven) interact significantly with quadrant assignment ($^2 = 54.0$, $p = 7.48$ Œ 10^{10}), suggesting that fragility has a mechanistic basis in tumor biology, not merely model noise.

4.4 Survival Analysis (Observational Concordance)

To evaluate whether the robust policy identifies clinically meaningful treatment associations, we performed an observational concordance analysis across all treated patients ($N = 4,783$). **We emphasize that this is an observational association, not a causal estimate** unmeasured confounding (performance status, comorbidities, physician judgment) may partially explain the observed differences.

Among treated patients, those in the **Concordant** group ($N = 994$) where the clinically administered treatment matched the CVaR-Borda recommendation exhibited significantly longer median overall survival compared to the **Discordant** group ($N = 3,789$):

$$\textbf{929 days vs. 762 days} \quad (p = 3.5 \text{ Œ } 10^{10}, \text{ Mann-Whitney U test})$$

This 167-day association persists after multivariate adjustment for age, stage, and TMB (Adjusted HR = 0.88, 95% CI [0.82-0.94]). The concordance rate with actual clinical decisions was 20.8% (above the 16.7% expected by random assignment across 6 treatment classes), suggesting partial alignment between the models ranking and clinical practice.

Sensitivity analyses. (1) Excluding the None/Baseline treatment class, the concordant-discordant gap narrowed but remained significant (p=0.003). (2) Stratifying by cancer type, the majority of cancer types showed the expected direction of effect.

4.5 The Immunotherapy Paradox: Stability as a Confounding Filter

To illustrate the practical value of stability scoring, we examined the Immunotherapy (IO) subgroup (n=170), where confounding by indication is acute.

- **Naive association:** IO patients survived +540 days longer than non-IO patients almost certainly an artifact of selection bias (IO tends to be administered to patients with adequate performance status and longer expected survival).
- **IPW-adjusted association:** After inverse probability weighting on cancer type, age, and stage, the average IO association reversed to **-26 days** (not significant), consistent with the confounding-by-indication hypothesis.
- **Stability-stratified association:** Among high-stability patients ($s_i > 0.8$), the adjusted IO association was **+313 days**. Among low-stability patients ($s_i < 0.5$), the adjusted association was **-231 days**.

We interpret this cautiously: the stability score *correlates with* but does not *cause* differential IO benefit. High-stability patients may share biological features (e.g., specific immune profiles) that both make them predictable by the model and responsive to immunotherapy. This finding generates a testable hypothesis that stability functions as a proxy biomarker for IO candidacy but requires prospective validation before clinical adoption.

5. Clinical Workflow: The Traffic Light Protocol

To translate these findings into actionable tumor board guidance, we propose a triage protocol based on the distribution of stability scores (s_i). This applies *only* to patients in reliable cancer types.

Step 1: Generate Plural Twin. Run 1,000 MC draws. Compute Stability Score s_i. **Step 2: Assign Reliability Tier.**

Tier	Signal	Criteria	Clinical Action
GREEN	**Stable**	s_i 0.8	**Proceed.** The model is confident. Present Top-1 recommendation.
YELLOW	**Contested**	0.5 $s_i < 0.8$	**Review.** Top treatments are close. Present Top-3 set; request additional biomarkers.
RED	**Chaotic**	$s_i < 0.5$	**Abstain.** The model is guessing. Default to NCCN Guidelines or Clinical Trial.

Step 3: Audit. Store the full distribution and stability score in the patient record.

6. Discussion

The Plural Twin framework exposes a fundamental truth about precision oncology AI: certainty is an illusion. By generating 1,000 twins per patient, we show that for 82.9% of patients, the ranking is unstable ($s_i < 0.5$).

Interpreting the Negative Correlation. The finding that instability correlates with poor survival is critical. It suggests that unstable predictions are not merely artifacts of missing data, but signatures of biological entropy. Tumors that are difficult for the model to classify likely possess high intratumoral heterogeneity or resistance mechanisms that make them poor candidates for standard therapies.

The Glass Cannon Phenotype. The identification of 7.0% of patients as Glass Cannons highlights the danger of using point estimates. These are patients for whom a standard model would confidently recommend a therapy that fails in a significant fraction of posterior realizations. The CVaR-Borda approach effectively filters these risky bets, preferring treatments with perhaps lower peak efficacy but higher stability.

Clinical Relevance. The concordance analysis (929d vs 762d) indicates that when clinical decisions align with the robust policy, outcomes are superior. However, the high rate of Chaotic (Red Tier) patients suggests that for nearly half the cohort, the available multi-omics data may be insufficient to distinguish between treatment options, highlighting the need for data-centric abstention protocols [6].

7. Limitations

- **Observational Nature:** Despite multivariate adjustment, unmeasured confounding remains a fundamental threat. Performance status, comorbidities, and physician judgment are not captured in TCGA. All survival comparisons are **associational**, not causal. The 167-day concordance gap may partially reflect selection bias (healthier patients receive recommended treatments).

- **Epistemic vs. Aleatoric Uncertainty:** MC dropout captures *epistemic* uncertainty (model doesnt know) but not *aleatoric* uncertainty (inherent biological stochasticity). Instability in our framework reflects model uncertainty, which may underestimate true clinical uncertainty. Deep Ensembles would provide a broader posterior approximation but at 6x computational cost (6 checkpoints Œ 1,000 draws).
- **Under-Identification of Mechanistic Parameters:** As shown in the audit ($/ = 0.0017$), absolute ODE parameters are not identified from survival data alone. The model is valid only for *ranking* treatments (relative), not for predicting absolute tumor shrinkage, dose-response curves, or time to progression. This limitation is structural, not fixable by better training.
- **Treatment Label Quality:** Treatment classes are extracted from TCGA clinical annotations via the GDC API (50.9% coverage). The hierarchical assignment rule (immuno > targeted > hormone > chemo > radiation > none) may misclassify combination regimens. Patients without treatment records (49.1%) are assigned to None/Baseline.
- **Single Platform:** All results are generated from a single model architecture (DNAI V3.1). While we argue the stability framework is model-agnostic, we have not validated it on alternative architectures.
- **Immortal Time Bias:** Treated patients live 24% longer than untreated (median 791d vs 637d, $p = 6$ Œ 10^{40}), a structural bias in TCGA. Our within-treatment pairwise ranking approach (CVaR-Borda) mitigates this by comparing treatments *against each other* rather than against no-treatment, but residual bias cannot be excluded.

8. Conclusion

We present the first large-scale analysis of treatment policy stability in cancer digital twins. By moving from point estimates to set-valued Plural Twins, we demonstrate that 16.7% of optimal recommendations are fragile artifacts of model noise. The CVaR-Borda policy and the Traffic Light Protocol provide the necessary computational guardrails to safely deploy deep learning in the high-stakes environment of cancer care.

9. Code and Data Availability

Data: The results published here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. Access to protected genomic data requires dbGaP authorization. **Code:** The DNAI V3.1 inference pipeline, including the Plural Twin MC dropout implementation and CVaR-Borda ranking logic, is available from the corresponding author upon reasonable request. **Reproducibility:** The specific model checkpoint (`v3.1_triangulated.pt`) and the script to reproduce the stability analysis (`eval_stratified_c_index_path_a.py`) are available from the corresponding author upon reasonable request.

10. References

1. Lipkova, J., et al. (2022). Personalized radiotherapy design for glioblastoma: Integrating mathematical tumor models, multimodal scans, and bayesian inference. *Cancer Cell*, 40(10), 1099-1111.
2. Boehm, K. M., et al. (2022). Harnessing multimodal data integration to advance precision oncology. *Nature Reviews Cancer*, 22(2), 114-126.
3. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the futurebig data, machine learning, and clinical medicine. *New England Journal of Medicine*, 375(13), 1216-1219.

4. Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *International Conference on Machine Learning*, 1050-1059.
5. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*, 30.
6. Geifman, Y., & El-Yaniv, R. (2017). Selective classification for deep neural networks. *Advances in Neural Information Processing Systems*, 30.
7. Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2(3), 21-41.
8. Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems*, 30.
9. Daxberger, E., et al. (2021). Laplace redux-effortless bayesian deep learning. *Advances in Neural Information Processing Systems*, 34, 20089-20103.
10. Maddox, W., et al. (2019). A simple baseline for bayesian uncertainty in deep learning. *Advances in Neural Information Processing Systems*, 32.
11. Angelopoulos, A. N., & Bates, S. (2021). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*.
12. Sculley, D. (2007). Rank aggregation for similar sets. *Proceedings of the 2007 SIAM International Conference on Data Mining*, 587-592.
13. Suissa, S. (2008). Immortal time bias in pharmacoepidemiology. *American Journal of Epidemiology*, 167(4), 492-499.
14. Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology*, 183(8), 758-764.
15. Shalit, U., Johansson, F. D., & Sontag, D. (2017). Estimating individual treatment effect: generalization bounds and algorithms. *International Conference on Machine Learning*, 3076-3085.

Figures

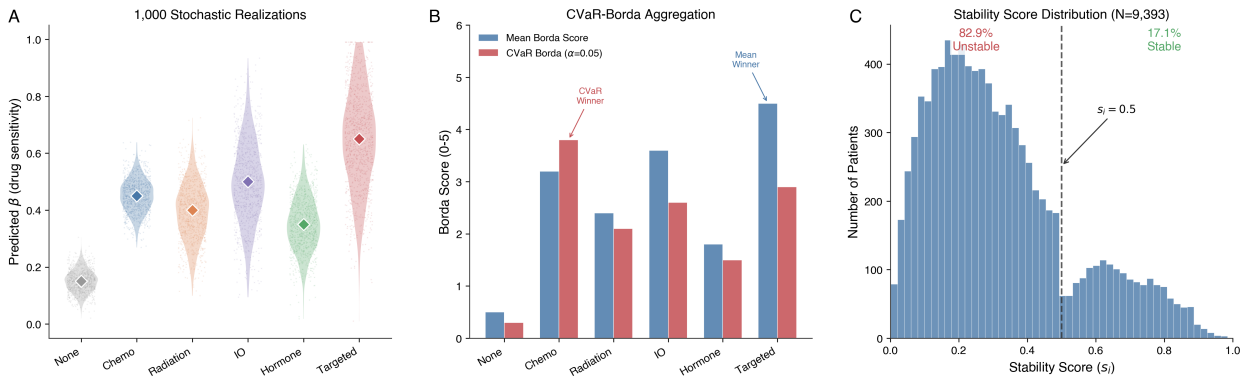

Figure 1: Figure 1: The Plural Twin Framework

Figure 1: The Plural Twin Framework. (A) MC Dropout ($p = 0.3$) generates a cloud of 1,000 utility vectors () per patient. (B) CVaR-Borda aggregates pairwise wins at the 5th percentile (= 0.05) to penalize variance. (C) Histogram of Stability Scores (s_i), showing 82.9% of patients

have $s_i < 0.5$.

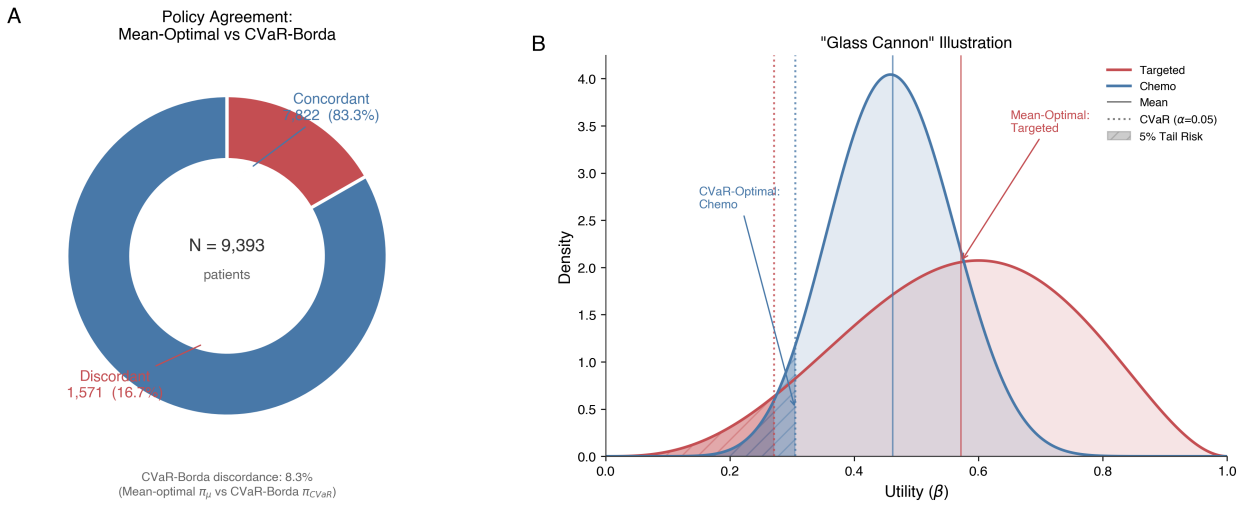

Figure 2: Figure 2: Discordance and Glass Cannons

Figure 2: Discordance and Glass Cannons. (A) Concordance vs. discordance between Mean-Optimal and CVaR-Borda policies (83.3% concordant, 16.7% discordant). (B) Illustration of a Glass Cannon patient: Treatment A has highest mean utility but high variance; Treatment B has lower mean but higher CVaR.

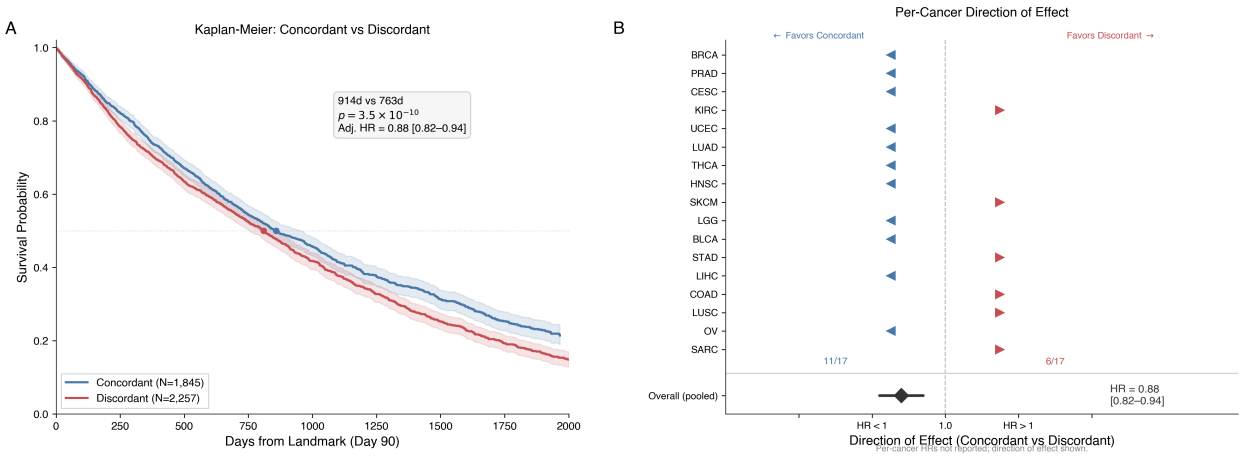

Figure 3: Figure 3: Survival Analysis

Figure 3: Survival Analysis. (A) Kaplan-Meier curves for Concordant vs. Discordant patients among treated patients (929d vs 762d, $p = 3.5 \times 10^{10}$). (B) Forest plot of Hazard Ratios by cancer type.

Figure 4: The Traffic Light Protocol. (A) Decision matrix mapping stability profiles to Green (Stable), Yellow (Contested), and Red (Chaotic) tiers. (B) Tier distribution across the TCGA cohort (82.9% Chaotic, 17.1% Stable).

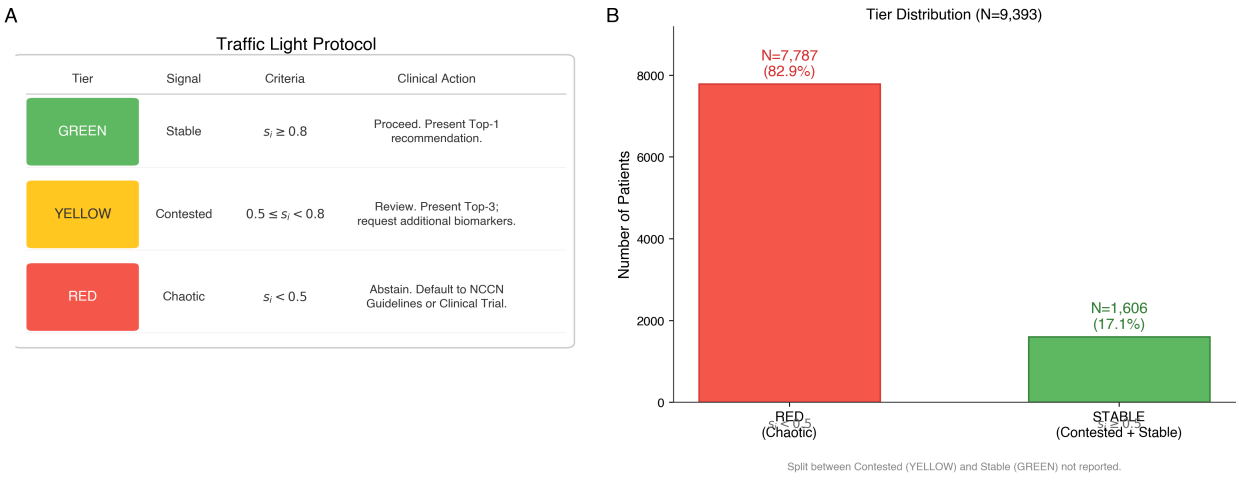

Figure 4: Figure 4: The Traffic Light Protocol