

# SocultPaperV6

2024-05-19

## Introduction

### What is adaptive cancer therapy

Worldwide, cancer is the cause of 1 out of 6 deaths. An estimated 90% of these cancer deaths are due to the development of drug resistance (Bukowski, Kciuk, and Kontek 2020). While initial cancer treatment usually shows a positive response in tumor burden, drug resistance develops over time. To highlight the inefficiency of traditional approaches, (Staňková et al. 2019) models cancer treatment as a game-theoretic contest between a physician and a tumor. In this model, the physician’s move in each round is to apply a certain treatment, and the tumor adapts. While it is a “Stackelberg game”—a game where one player is the leader (the physician) and another player is a follower (the tumor)—this asymmetry is rarely exploited. Instead of using their advantage to steer the evolutionary pressures placed on tumors, physicians let the tumor adapt not only to the current round of the game but also to future rounds. Thus, the advantage of leading the game is lost. The authors analogize the current practice:

*“Consider cancer treatment as a rock-paper-scissors game in which almost all cells within the cancer play, for example, “paper.” It is clearly advantageous for the treating physician to play “scissors.” Yet, if the physician only plays “scissors,” the cancer cells can evolve to the unbeatable resistance strategy of “rock.””* (Staňková et al. 2019)

*Adaptive Therapy* is an approach to cancer treatment based on controlling the intra-tumoral evolutionary dynamics by leveraging the fitness costs that cancer cells incur when evolving mechanisms to resist drugs. For example, it has been shown that tumor cells can mutate to increase the expression of the PGP membrane pump, which uses ATP to expel drugs from the cell. While this makes the cell more resistant to treatment, it also comes at a metabolic cost. (Gatenby et al. 2009) found that PGP activity was responsible for approximately 50% of cell metabolism. As a result, if resistant and non-resistant cells are competing for space and resources, drug-sensitive cells should, over time, outcompete resistant cells. Adaptive Therapy utilizes this Darwinian competition to make cancer fight itself. The allure of adaptive therapy is not to eradicate cancer but to make it a controllable chronic disease.

Initial results from a pilot clinical trial applying adaptive therapy for a group of *metastatic castrate-resistant prostate cancer* (mCRPC) patients are promising, showing both lower cumulative dosages and longer survival compared to a similar group of patients receiving standard care. The trial has been utilizing a range-bounded heuristic to manage the tumor burden: if the blood marker *Prostate-specific Antigen* (PSA), a proxy for tumor burden, increases back to pre-trial levels, treatment is applied until PSA drops to 50% of pre-trial levels (Zhang et al. 2017). The trial is expected to run until December 2024 (Center and Research Institute 2024).

### Desired qualities of models

The clinical trial reported by (Zhang et al. 2017) only utilized a single drug. A group of researchers has recently surveyed the use of mathematical modeling in the field of Adaptive Therapy, identifying several needed developments. For example, they argue that modelling approaches will need the capacity to simulate

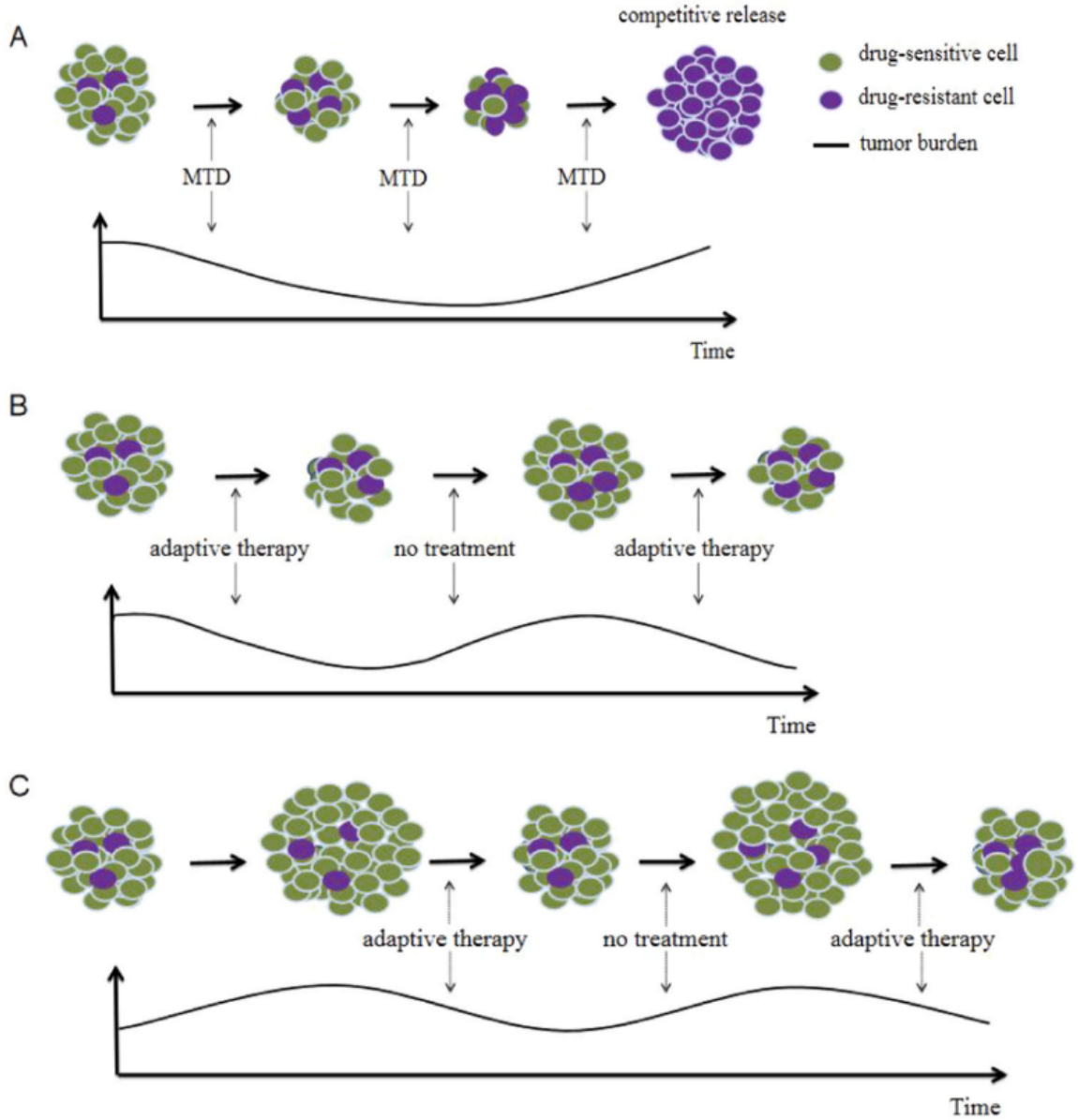


Figure 1: Figure 1: From Zhang et al. 2023. Panel A shows a typical "Maximal Tolerable Dose" treatment protocol. While the initial response in tumor burden is promising, a competitive release of resistant cells ensures. Panel B shows Adaptive Therapy approach with a small tumor burden. Panel C shows adaptive therapy with a high tumor burden, which has been theorized to further increase the suppression of resistance cells through darwinian competition.

multidrug therapies. Additionally, they argue that it is unlikely that any single treatment approach will be able to delay the emergence of resistant cells, while lowering tumor burden and minimizing toxicity. Given that patients likely differ in their ability to tolerate tumor burden, drug toxicity, and the evolutionary dynamics of their cancers, therapies will have make informed trade-off between these objectives for each specific patient. Additionally, more complicated modelling of tumor dynamics is required. The authors highlight several biological mechanisms, such as how competition dynamics change depending on how close a tumor is to carrying capacity. Likewise, high tumor burdens might also come with other costs, such as an increased risk of new metastasis. Models will therefore need the flexibility to specify complex causal systems.

Predicting individual patient treatment responses would greatly enhance adaptive therapy since dose modulation could be individualized further, and evolutionary dynamics could be controlled with more precision. A key issue going forward is rethinking how data on patients is collected. It will be crucial to collect data not only to detect progression but also to inform future treatment decisions. For example, a risk of using a set testing frequency, as done in the pilot clinical trial, is treatment *overshooting*, since drops in PSA weren’t detected early enough. This could have limited future treatment opportunities. The authors also stress the importance of careful analysis of the uncertainties involved in the treatment suggestions by any model (West et al. 2023).

## Partially Observable Markov Decision Process

*Partially Observable Markov Decision Processes* (POMDPs) is a class of controller models that model an Markovian process. A markovian process is a discrete time and state environment, the next step of the system only depends on the current step. Crucially, partial observability is a key facet of these models and refers to the fact that these models don’t directly observe the actual environment or Markovian process but instead only potentially noisy signals emitted by environment (Åström 1969). This allows these models to differentiate an observed signal from what it “believes” about the environment and use a single reward function to trade off uncertainty for achieving a certain goal state (Kaelbling, Littman, and Cassandra 1998) while yielding Bayes optimal beliefs. For example, if discretized, a POMDP could understand a PSA reading as a noisy signal of the actual tumor state and thus try to control the tumor state rather than the PSA reading. While POMDPs are typically difficult to solve analytically, various approximate approaches exist.

## Using Active Inference to solve POMDPs

One approximate solution is born out of the neuroscience paradigm of *Active Inference*. By minimizing two objective functions—*Variational Free Energy* (VFE), a measure of a generative model and past sensory inputs, and *Expected Free Energy* (EFE), a score of future courses of action against a set of preferences. Active Inference has been used to model psychopathology (Da Costa et al. 2020) but also applied to control scenarios such as the mountain-car problem (Friston, Daunizeau, and Kiebel 2009) and, albeit augmented with deep-learning, robotics control (Çatal et al. 2020). Active Inference implementations of the POMDP scheme have been implemented in MATLAB and recently in Python with the python package *pymdp* [(Heins et al. 2022a) and (Heins et al. 2022b)].

Typically, Active Inference POMDPs are technically a joint probability:

$$p(o, s, u; \phi)$$

where  $o$  are observations,  $s$  are hidden states,  $u$  are “control states” (states that an agent can influence), and  $\phi$  are the hyperparameters of the model, such as  $\alpha$  typically used as ‘inverse temperature’ i.e. how deterministically the model selects actions. By conditioning on certain observations, the POMDP is solvable by various approximation schemes for the optimal posterior beliefs at certain time point. This yields a posterior beliefs about the current hidden states are, but also what the optimal course of action is for a given set of preferences. For example, a POMDP could model the joint probability of observing a certain amount of tumor burden in a patient, given some hidden states, such as a tumor burden, a degree of treatment

resistance, and whether a certain treatment was applied. Computability is the only limit to how far into the future these POMDPs can be solved. Due to considering time and states discrete, this joint probability can be made tractable by factorizing the joint probability into the following categorical probability distributions:

- A likelihood model *A*. Usually modeled as a set of arrays where each array corresponds to an observation modality and describes how observations map to a particular state. For example, one modality could be PSA readings, and its likelihood array describes how likely different test results are under different tumor burdens. It can be modeled as a perfect signal of the tumor burden, i.e., the same tumor burden always gives the same test result, but it could also be implemented as a noisy signal. If multiple states are modeled, each likelihood modality is an array with a dimension corresponding to the number of possible observations of that modality, and for each hidden state, another dimension is added with the same length as the number of possible hidden states. In this way, every combination of states can be mapped to an observation.
- A transition model *B*. It describes the probability of a state transitioning to another at any particular timestep. This also encodes how actions are expected to influence transition probabilities. It could describe how a tumor is likely to evolve from one timestep to the next depending on whether treatment is being applied. The transition model is usually coded as a collection of three-dimensional arrays, a first dimension for the next state, a dimension for the current state, and a third dimension with the length of each action that would influence transition probabilities of the state. Modeling the transition probabilities would depend on whether treatment is being applied or not.
- A prior over preferred states *C*. A particularity of these models is that utility is specified as preferences for states rather than rewards. Thus, the agent isn’t seeking to maximize cumulative future rewards but to seek states it has a preference for. However, this doesn’t stop an agent from modeling reward states, as preferences can still be implemented over these reward states.
- A prior over initial states *D*. Finally, a POMDP is solved by implementing a generative model that approximates the joint probability. This can be achieved by computing the approximate posterior beliefs over the observations and hidden states, given observations so far. While many techniques exist, Active Inference implementations are based on free-energy minimization (FEM) techniques. FEM tries to maximize model evidence of the POMDP model by using model parameters as expectations to create a predictive model that can explain what is observed. Given that these models are equipped to reason about how future policies and plans can maximize the likelihood of future observations matching preferences, this approximate belief model can also be used to evaluate possible policies against a generative model by FEM and select actions based on the evaluation. A simplified version of this process can be described by “perception” and “planning” steps.

The POMDP can be solved at any time step by minimizing two measures, *Variational Free Energy* (VFE) and *Expected Free Energy* (EFE), to yield the optimal beliefs about current states given observations and the best course of action. VFE is an upper bound on *surprisal* and is approximated to infer the optimal posterior of current states given some observations, the likelihood model, and the prior over either initial states if no earlier observations have been made. Otherwise, the preceding posterior over states is used as the current prior (Smith, Friston, and Whyte 2022). VFE scores how well the model’s representation aligns with a set of observations, and by using different approximate posteriors over states, choosing the approximate posterior over states which minimizes VFE will approximate Bayesian belief updating (Da Costa et al. 2020). EFE instead scores *policies* — plans about future actions — on their expected outcomes. EFE is used to construct a posterior distribution over what policy is most preferable given *C*, the set of prior preferences over observations.

## The aim of this paper

The treatment decisions used by (Zhang et al. 2017) were based on a, albeit well-informed, heuristic about how resistance dynamics likely would change based when treatment was applied or withdrawn. This paper will

investigate whether the Active Inference implementation of POMDPs could be viable in explicitly modeling the degree of treatment resistance, with the goal in mind of controlling this long term. Beyond the fitting the desiderata reported by (West et al. 2023), the use of the Active Inference paradigm for the present project is further motivated by the following considerations:

- Real-time data is likely sparse in clinical settings. Extracting the maximum amount of information from each data point will therefore be essential. Using models that perform optimal Bayesian inference would therefore be preferable.
- Short-term exploitation of tumor vulnerability is essential for keeping the patient alive, but for long-term success, controlling resistance dynamics will be crucial. A therapy plan will have to balance keeping the patient alive now against limiting its future options for treatment. Using a measure such as Expected Free Energy could be instrumental in striking this balance.
- Resistance dynamics are not directly measurable, but should be inferable through observing how the tumor burden responds to treatment. Choosing whether to treat would therefore not only be a consideration of the tumor level but also about how much it informs us about the resistance dynamics.
- The information gain from testing can be quantified and used to prioritize testing for periods of greater uncertainty.

An interesting corollary to these motivations is that POMDPs could suggest some counterintuitive treatment decisions. For example, a model could correctly estimate that one patient has a low tumor burden and suggest not applying treatment, while proposing treatment for another patient with exactly the same burden. This is a feature, not a bug. If the model has less certainty about the resistance dynamics of the second patient, applying treatment could be the optimal move simply to gain information about the resistance dynamics.

## Analysis

### Simulated environment

To investigate the appropriateness of the Active Inference implementation of the POMDP scheme, multiple simulations were run. To easily comply with the discrete state and time implementation typical of the Active Inference literature, discrete states and time were used. In each simulation run, a simulated clinical setting of a maximum of 200 timesteps was created. In each timestep, a POMDP has to keep a virtual cancer patient alive by deciding when to apply a treatment based on a signal of the tumor burden. This environment was inspired by the use of PSA testing and Abiraterone treatment in the pilot clinical trial (Zhang et al. 2017). In the simulations, a patient’s *tumor state* determines whether they survive to the next timestep. The simulation ends if the patient does not survive. Simulation runs always begin with the tumor state at 0, but it increases with a fixed risk at each timestep. If the tumor state reaches 5, the virtual patient “dies” and the simulation ends.

Applying treatment can reduce the tumor state and thus prevent the patient from dying. However, whether a treatment successfully reduces the tumor state depends on an underlying *resistance state*, which also begins at state 0 out of 5. When the resistance is low, the chance of treatment succeeding is high. But for each round of treatment application, the resistance state has a fixed risk of increasing. This renders treatment less and less effective. The resistance state can decrease if treatment is withdrawn, and the probability of the resistance state decreasing depends on the tumor level. At high tumor levels, the resistance state is most likely to decrease. This is based on the work of (Hansen and Read 2020), who suggests that larger tumor sizes should generate more competitive pressure on resistant cells.

All in all, this means that a model has to balance the tumor state not growing out of control, which would “kill the patient,” while not painting itself into a corner by applying treatment too frequently.

## Modifications to POMDP scheme

Typically for Active Inference implementations of POMDPs, different types of hidden states such as resistance and tumor states are not modeled to affect each other directly. Instead, the interactions of combinations of different states are modeled to lead to different observations. This means that their interactions would be coded in the  $A$ -array, which contains the likelihood mappings.

However, this setup was deemed incompatible with the clinical setting attempted emulated. Therefore, custom changes to the implementation in `pymdp` were made to allow a POMDP to infer causally upstream “occluded” nodes, which don’t themselves produce observations (E.g the resistance state). Instead the state of the occluded node will have to be inferred, through its downstream effects on partially observable state (the tumor state).

Specifically, modifications were made to the structure of the generative model’s transition beliefs. Usually, three-dimensional “B-tensors” describe expected transition probabilities for a hidden state. This would mean a B-tensor would hold one dimension for the current state, one dimension for the next state, and a third for every allowed action. These B-tensors were given another dimension that corresponds to the state of another type of hidden state. Concretely, this meant that the tumor state factor had a fourth dimension corresponding to tumor transition probabilities for each resistance state. The expected transition probabilities could then be estimated by matrix multiplication of the B-tensor containing the tumor state transition probabilities and a vector of expected resistance levels at a given time point. One can imagine that instead of having only one B-tensor for the transition probabilities of the tumor state, multiple B-tensors of tumor transition rates were specified. The transition probabilities of each tensor then corresponded to a specific resistance level. This modification to the generative model necessitates taking an average of all these tensors weighted by the expected probabilities for each resistance state. This amounts to calculating the expected tumor state conditional on a set of beliefs about the resistance state and whether treatment was being applied. This modification strategy was repeated for the resistance level since decreases in the resistance state depended on the tumor state. While these changes rendered much of the functionality of `pymdp` immediately unusable, this strategy should be able to model arbitrarily complex dependencies between states. Hopefully, a robust implementation can be designed without overhauling `pymdp`. It should also be noted that this modification means that it is no longer meaningless in which order beliefs about states are evaluated. Instead, this must be specified. For example, for the present simulations, the effect of resistance level on treatment efficacy was evaluated first since this was evaluated first at each step of the simulation.

## Exploratory Simulation

An exploratory simulation was conducted, where a single run of POMDP controlling treatments and testing was performed to investigate the feasibility of using the modified POMDP scheme. This model could choose whether to test and treat. It would always observe a perfect signal of whether the patient is alive, whether the model is testing, and whether treatment is being applied. If the model decided to test, it would receive a signal of the tumor state. Its prior preferences were heavily against observing a dead patient, somewhat against treating, and slightly against testing. This was done to simulate a cost to both treating and testing. This means that the model will have to balance the utility gained and the “cost” of all these actions while considering the potential information gain of each choice. The model was further handicapped by adding noise to the tumor signal. This means that the resistance state, which must be inferred through the tumor signal, is doubly obfuscated. However, the likelihood mapping accounted for the expected noise in the tumor signal, and it perfectly knew the transition probabilities of the environment. The model was given uniform priors over initial states, meaning it had no knowledge at the beginning of each run. The model evaluated the same set of predetermined policies considering 6 timesteps into the future. This was done to ease computation by limiting the search space. The policies consisted of two blocks of either testing or treating for three timesteps in a row for each combination of applying testing or not at the timestep.

## Learning underlying hidden state

The simulation run managed to keep the patient alive for 68 timesteps and chose to test on 55 of these (see Fig. 1).

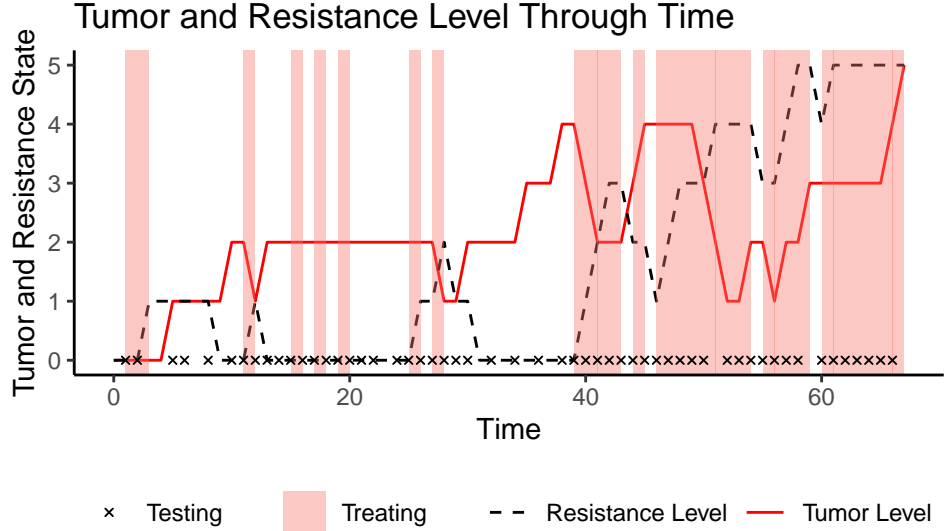


Figure 2: Figure 1. Shows a simulation run for exploratory analysis.

The model seems to be somewhat sensibly applying treatment. Generally, it refrains from applying treatment when the tumor state is low. Interestingly, it applies treatment at the first timestep, despite the tumor state and resistance state being at their lowest possible values. Considering that the model has uniform priors about the resistance state and tumor level, this means that it is maximally uninformed about both. Treating and testing would therefore immediately provide information about both. Qualitatively, the model also seems to prefer testing when treatment is being applied. This suggests that it finds testing to be more worthwhile while also treating. Considering that testing during treatment not only provides information about the tumor state but also the resistance state, this behavior is useful.

A plot of the model's beliefs about the resistance state at each time-point (see Fig. 2) also shows that its beliefs generally follow the development of the actual resistance level.

All in all, the modification to POMPD structure seems to allow the model to infer an the underlying state of causally upstream state factor through a noised signal form of a downstream node. This suggests that the model could be used to infer the resistance state from the tumor state. However, the model's behavior is not yet optimal.

## Expected trajectories

For each policy, the model evaluates the expected trajectory of the hidden states (see Fig. 3). These evaluations provide insights into the model's uncertainty and what it expects to happen under the best policy at each time point.

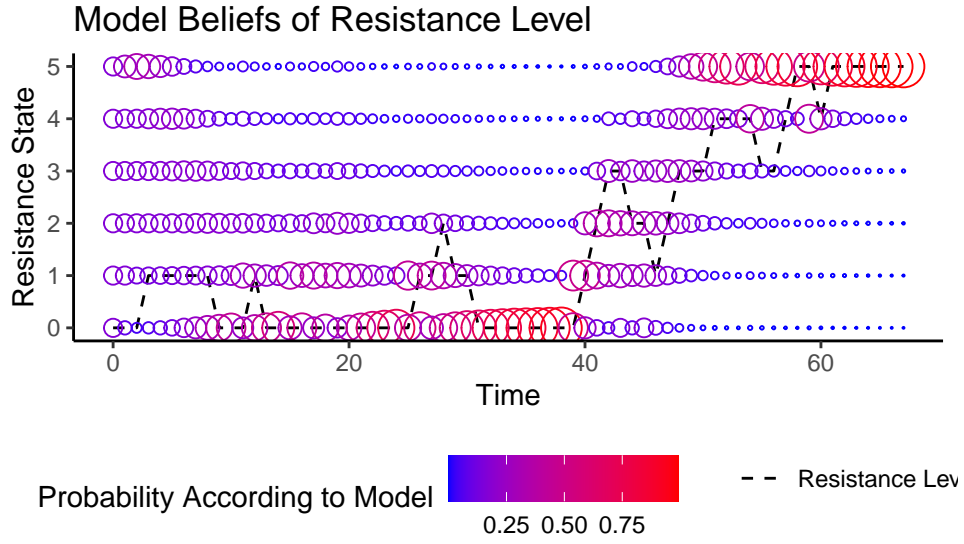
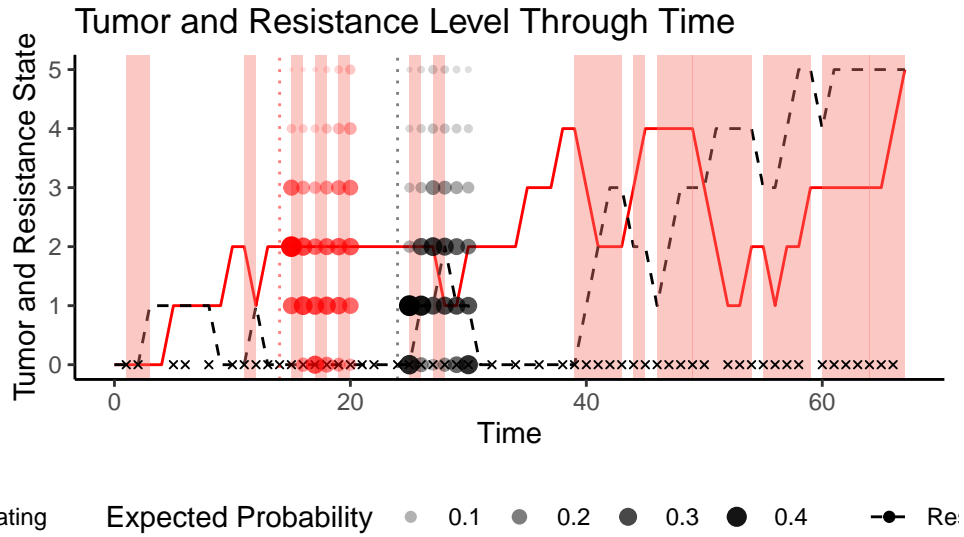
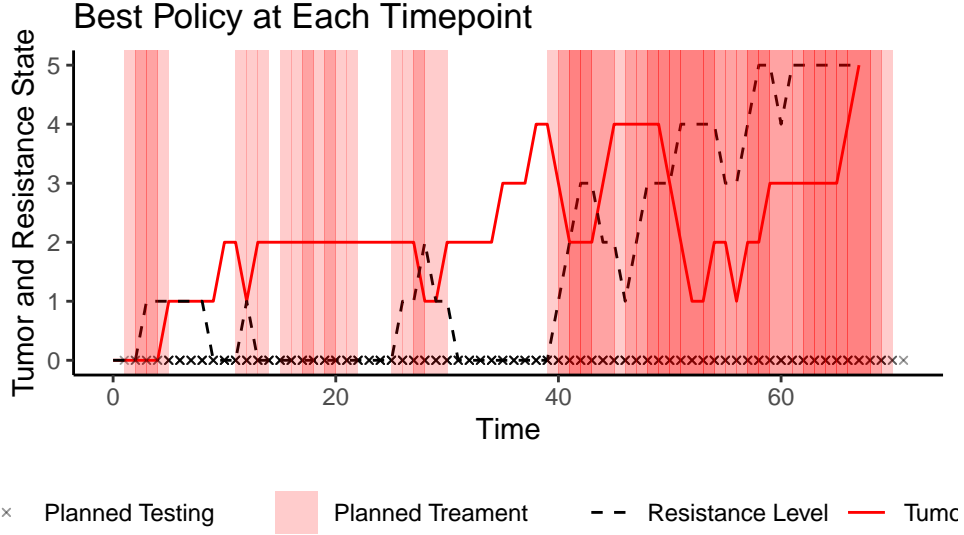


Figure 3: Figure 2. Plot of the models belief about the reistance level at every timepoint

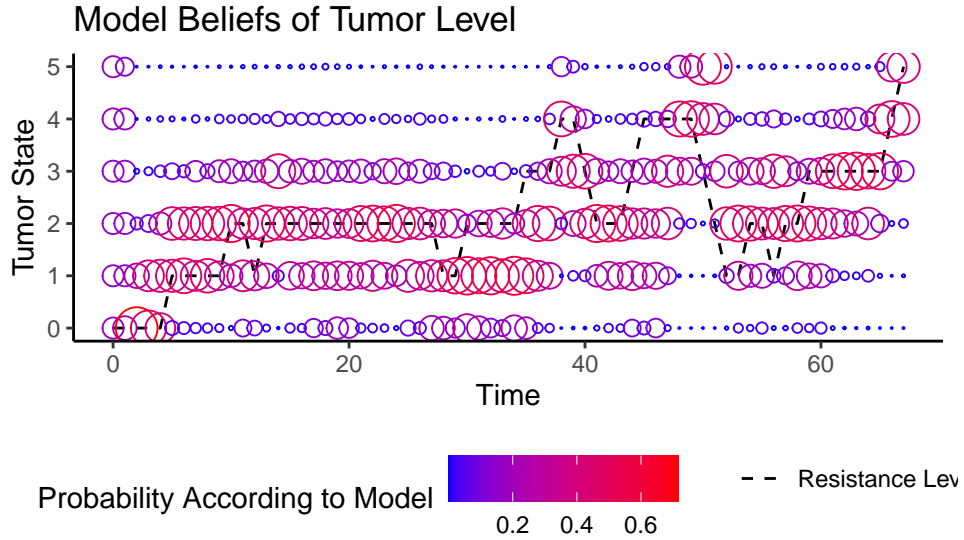


The amount of “flip-flopping” on decisions can also be investigated by examining what policy the model finds the most promising at each time step (see Fig. 4).

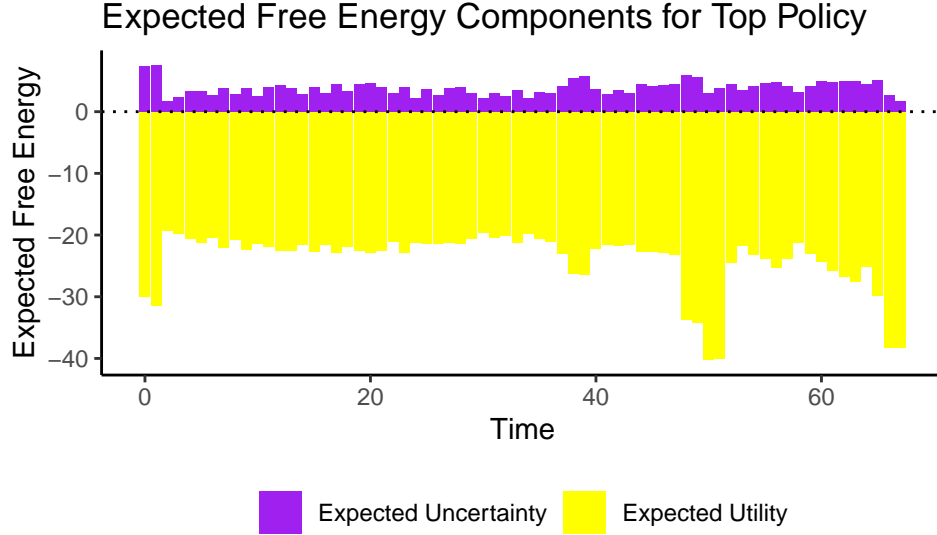




Curiously, there is a long stretch, approximately from timesteps 30 to 39, where the model never considers treating to be the optimal course of action, even though the resistance level is low. The model also holds accurate beliefs about the resistance level at this timepoint (see Fig. 2). However, this is likely due to a combination of multiple factors. First, there is a cost to treating, which means that the model generally prefers not to treat. It also underestimates the tumor burden at this timepoint (see Fig. 5) and gets “caught off guard” by a sudden rise in the tumor level. Since there is a cost associated with testing, it is also hesitant to apply tests during this period. Because it only considers six timesteps into the future, it likely doesn’t consider the long-term consequences of not reducing the tumor level.



The exact “decision making” for each policy can be evaluated. Each policy is evaluated for its EFE, which is a combination of the expected utility penalized for expected uncertainty (see Fig. 6).



The free energy components reveal several interesting decision-making points. At first, the model, given its uniform priors, is very uncertain about outcomes, and decreasing uncertainty weighs heavily in its decision making. This uncertainty also means that the model believes it is more likely to be in a risky situation, and reducing this risk is a priority. During the earlier identified pivotal period between timesteps 30-39, there is a concerning lack of increase in the weight of the utility component. This could mean that the model is too concerned with avoiding both treating and testing. It is only from timestep 40 onwards that the model realizes the severity of the situation, i.e., the rising tumor level (see Fig. 5). This analysis warrants strategies to increase the model’s ability. Changing the prior preferences to make testing and treatment “cheaper” in terms of utility might help. However, simply treating and testing more would translate to actual costs, both financial and to patient well-being. While it is important that the prior preferences are tuned to costs in a clinical setting, it would probably be more fruitful to give the model longer policies to consider. This would hopefully yield models that are better tuned to the risks of not detecting and controlling tumor levels at low values.

## Performance of POMPDs against a Range-Bounded Strategy

Several modified Active Inference POMPDs were tested against a range-bounded strategy to benchmark the performance of POMPDs against a strategy inspired by the treatment protocol used in the pilot clinical trial (Zhang et al. 2017). The POMPDs were modified so that they could only observe the tumor state, receiving a noiseless signal of this at every timestep. They were also only given prior preferences for observations of tumor states. These preferences were uniform over all states except for the highest tumor state, for which the probability was set to 0. Three POMPDs were tested with different lengths of policy horizons:

- A short horizon model that could consider a block of treating or not treating for the next three steps.
- A medium horizon model that could consider two blocks of treating or not treating for three steps, yielding a total horizon of 6 timesteps.
- A long horizon model that could consider four blocks of treating or not treating for three steps, resulting in a total horizon of 12 steps.

These three models were benchmarked on four different “tumor aggressivity settings” against a range-bounded strategy inspired by the pilot clinical trial. The tested growth rates were 0.1, 0.3, 0.5, and 0.7 probability of increasing the tumor state at each timestep. Each of these four bouts consisted of 100 simulation runs for each model. For each single run, the environment was predetermined by generating 200 outcomes at each combination of tumor state, resistance state, and treatment state. This was done to ensure comparability at the level of each run.

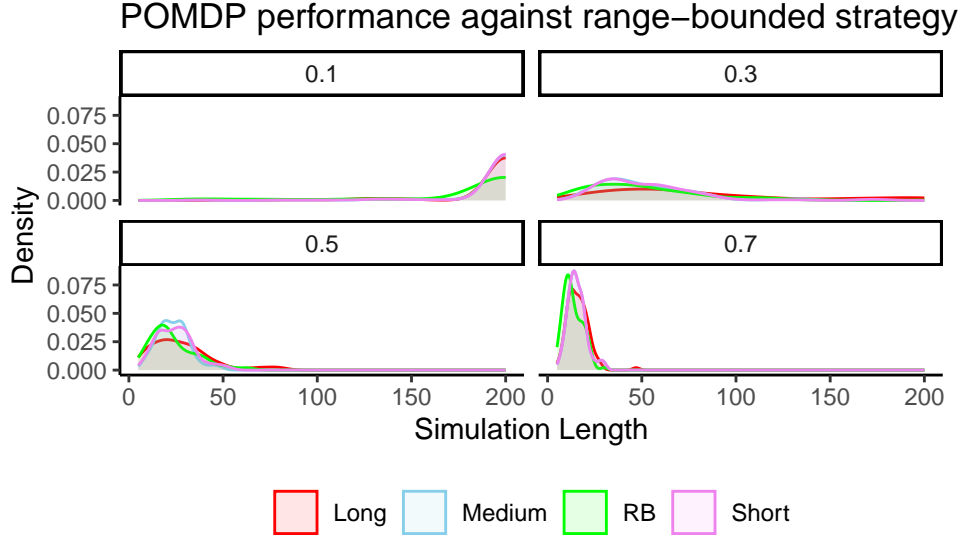


Figure 4: Figure. 7. The observed simulation lengths for each model type and tumor risk.

At the lowest tumor growth rate, nearly all the runs reached the maximal length (see Fig. 7). The longest policy strategy also appeared to perform slightly better than the other models. Since the difficulty of managing the tumor burden changed, contrasts were computed between the long policy POMPD and the range-bounded strategy. Because the stochasticity of each run was predetermined, the percentage difference from switching from an RB-strategy to a POMPD can be computed for each run (see Fig. 8).

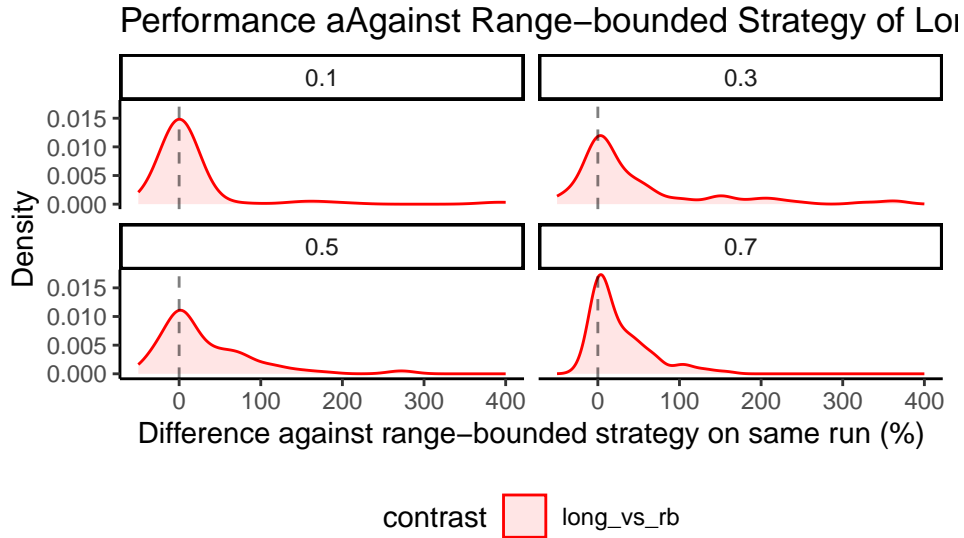


Figure 5: Figure 8. Shows the change in percent of between the range-bound approach and the longest horizon POMPD for tumor aggressivity settings

In the 100 runs with a 0.1 risk of tumor growth, the long horizon POMPD seems to perform similarly to the RB strategy. However, as the tumor growth risk increases, the POMPD model increasingly shows large percentage improvements compared to the RB strategy. This could be because, as the tumor dynamics speed up, the need for proactive action becomes more important, favoring models that attempt to also control the underlying resistance state.

# Discussion

## What This Project Found

This project tested a simple modification to how POMDPs are typically implemented in Active Inference to investigate whether POMDPs could learn and control a hidden state that could only be inferred based on its consequences on other states. This was done in a simulated environment designed to mimic the dynamics of adaptive therapy. The modification was successful, allowing the POMDP to model an underlying resistance state that controlled the efficacy of treatment, despite the underlying state not producing any observable signals itself. POMDPs were further tested against a range-bounded treatment strategy and were found to outperform the range-bounded strategy. The difference in performance increased as the aggressiveness of tumor growth increased and as the future time-horizon that the POMDPs considered at each time step lengthened. This finding underscores the potential of using the Active Inference paradigm to plan adaptive therapy, and particularly demonstrates that, under simplified circumstances, POMDPs could be a useful choice for implementing Active Inference for real-time treatment decision-making.

Additionally, this paper demonstrates that the Active Inference implementation of POMDPs shows several desirable qualities for adaptive therapy POMDPs:

- The models are flexible. Computation and our ability to inform likelihood and transition dynamics are the only limits to the combinations of hidden states, actions, and outcome modalities. This paper implemented a POMDP capable of combining multiple decisions, such as treating and testing, but more complex generative models could include multiple testing and treatment actions. As (West et al. 2023) highlight, it is important to incorporate biological factors into modeling decisions.
- Another quality is the capacity to balance expected information gain against expected utility. The certainty of the model is incorporated directly in selecting when to apply treatments and when to test. This suggests that there is more information to be gained from testing when treatment is being applied, as it provides a view into the underlying resistance dynamics. Immediate testing seems to suggest that the models are more likely to apply testing when also applying treatment. Another interesting corollary is that a POMDP could believe that two patients have exactly the same tumor burden but suggest treatment for one and not the other. This is a feature, not a bug. If the model is certain about an underlying state (e.g., resistance in the models in this paper), there is less gain in information from treating. On the other hand, if the model is uncertain about the resistance state, it might be worthwhile to treat simply to gain information. This is appropriate behavior, as it allows the model to better its chances of keeping the patient alive in the future.
- Another quality desired (West et al. 2023) is rigorous uncertainty. Since the POMDP’s certainty about current and future states is easily extractable, this criterion is arguably met. Additionally, its “reasoning,” i.e., free energy estimates, can be examined for each action.

## Future Research

While the results are initially promising, the simple nature of the simulation prohibits drawing any strong conclusions about whether Active Inference POMDPs could eventually be a good model choice in a clinical setting.

To further investigate whether POMDPs could apply to a clinical setting, multiple problems will need to be solved. For example, in the present study, the environment featured only discrete values. Since biomarkers and dosage intensities will likely be continuous values, some combination of binning continuous values to discrete values or adapting the model structure to work with continuous data is necessary. Deep learning has been used to construct likelihood mappings and transition probabilities in POMDPs from continuous data (Çatal et al. 2020). Another key issue is determining a way to inform likelihood and transition probabilities in a fashion that would be viable in a clinical setting. The present study simply used the actual transition

probabilities of the simulated environment and, depending on the simulation run, also used noiseless likelihood mappings. Possibly, simulations of cancer dynamics, such as those by (Zhang et al. 2017), could be used to inform transition probabilities, and a clinical model could potentially readjust to patient data. The Active Inference implementation of POMDPs has a developed literature on learning probabilities from data and even on how to incorporate the information gain of learning transition and likelihood probabilities into decision-making ((Smith, Friston, and Whyte 2022) and Da Costa et al. (2020)).

Another key issue is constructing a more rigorous benchmarking system for proposed models. In the present paper, the ranges of the compared range-bounded model were selected fairly arbitrarily. It produced convincing results during initial testing and did manage to control tumor levels for a substantial amount of time. However, a systematic method of comparing models is necessary, especially considering the difficulty of real-world tests. No matter the choice of modeling framework, building a benchmarking suite is crucial for the adaptive therapy discipline. Given the difficulty of real-world testing, we must maximize the information that can be extracted from simulation work. (West et al. 2023) has produced a detailed qualitative account of needed developments in mathematical models. Translating this account to a set of simulation environments would be extremely useful. A number of useful simulation scripts probably already exist. For example, (Zhang et al. 2017) released their simulations in Matlab. Wrapping existing simulations into environments that can easily exchange actions and observations could be facilitated by using an API like the one used in Gymnasium (formerly OpenAI Gym) (Towers et al. 2024) to minimize friction for non-oncology researchers. If experts in adaptive therapy were to predetermine a set of benchmarks, it would also greatly ease the burden on outsiders to identify useful contributions, allowing them more time to implement these contributions.

If an easy-to-use benchmarking suite existed, POMPDs could also be “reverse-engineered” from more complex models. If a black-box model, like a neural network, can be shown to successfully control treatment application in more complicated simulations of cancer dynamics, this would presumably not be too difficult considering that POMPDs have often been fitted to human decision-making in computational psychiatry. The same techniques could potentially allow us to translate the decision-making of a black-box model into the structure of a POMPD, thus combining the performance of the black-box model with the transparency of POMPD structure.

- Åström, K. J. 1969. “Optimal Control of Markov Processes with Incomplete State-Information II. The Convexity of the Lossfunction.” *Journal of Mathematical Analysis and Applications* 26 (2): 403–6. [https://doi.org/10.1016/0022-247X\(69\)90163-2](https://doi.org/10.1016/0022-247X(69)90163-2).
- Bukowski, Karol, Mateusz Kciuk, and Renata Kontek. 2020. “Mechanisms of Multidrug Resistance in Cancer Chemotherapy.” *International Journal of Molecular Sciences* 21 (9): 3233. <https://doi.org/10.3390/ijms21093233>.
- Çatal, Ozan, Samuel Wauthier, Cedric De Boom, Tim Verbelen, and Bart Dhoedt. 2020. “Learning Generative State Space Models for Active Inference.” *Frontiers in Computational Neuroscience* 14 (November). <https://doi.org/10.3389/fncom.2020.574372>.
- Center, H. Lee Moffitt Cancer, and Research Institute. 2024. “A Pilot Study of Adaptive Abiraterone Therapy for Metastatic Castration Resistant Prostate Cancer.” <https://clinicaltrials.gov/study/NCT02415621>.
- Da Costa, Lancelot, Thomas Parr, Noor Sajid, Sebastijan Veselic, Victorita Neacsu, and Karl Friston. 2020. “Active Inference on Discrete State-Spaces: A Synthesis.” *Journal of Mathematical Psychology* 99 (December): 102447. <https://doi.org/10.1016/j.jmp.2020.102447>.
- Friston, Karl J., Jean Daunizeau, and Stefan J. Kiebel. 2009. “Reinforcement Learning or Active Inference?” *PLoS ONE* 4 (7): e6421. <https://doi.org/10.1371/journal.pone.0006421>.
- Gatenby, Robert A., Ariosto S. Silva, Robert J. Gillies, and B. Roy Frieden. 2009. “Adaptive Therapy.” *Cancer Research* 69 (11): 4894–4903. <https://doi.org/10.1158/0008-5472.CAN-08-3658>.
- Hansen, Elsa, and Andrew F. Read. 2020. “Modifying Adaptive Therapy to Enhance Competitive Suppression.” *Cancers* 12 (12): 3556. <https://doi.org/10.3390/cancers12123556>.
- Heins, Conor, Beren Millidge, Daphne Demekas, Brennan Klein, Karl Friston, Iain D. Couzin, and Alexander Tschantz. 2022a. “Pymdp: A Python Library for Active Inference Indiscrete State Spaces.” *Journal of Open Source Software* 7 (73): 4098. <https://doi.org/10.21105/joss.04098>.
- Heins, Conor, Beren Millidge, Daphne Demekas, Brennan Klein, Karl Friston, Iain Couzin, and Alexander Tschantz. 2022b. “Pymdp: A Python Library for Active Inference in Discrete State Spaces.” *Journal of*

- Open Source Software* 7 (73): 4098. <https://doi.org/10.21105/joss.04098>.
- Kaelbling, Leslie Pack, Michael L. Littman, and Anthony R. Cassandra. 1998. “Planning and Acting in Partially Observable Stochastic Domains.” *Artificial Intelligence* 101 (1): 99–134. [https://doi.org/10.1016/S0004-3702\(98\)00023-X](https://doi.org/10.1016/S0004-3702(98)00023-X).
- Smith, Ryan, Karl J. Friston, and Christopher J. Whyte. 2022. “A Step-by-Step Tutorial on Active Inference and Its Application to Empirical Data.” *Journal of Mathematical Psychology* 107 (April): 102632. <https://doi.org/10.1016/j.jmp.2021.102632>.
- Staňková, Kateřina, Joel S. Brown, William S. Dalton, and Robert A. Gatenby. 2019. “Optimizing Cancer Treatment Using Game Theory.” *JAMA Oncology* 5 (1): 96–103. <https://doi.org/10.1001/jamaoncol.2018.3395>.
- Towers, Mark, Jordan K Terry, Ariel Kwiatkowski, John U. Balis, Gianluca Cola, Tristan Deleu, Manuel Goulão, et al. 2024. *Gymnasium*. Zenodo. <https://doi.org/10.5281/zenodo.10655021>.
- West, Jeffrey, Fred Adler, Jill Gallaher, Maximilian Strobl, Renee Brady-Nicholls, Joel Brown, Mark Roberson-Tessi, et al. 2023. “A Survey of Open Questions in Adaptive Therapy: Bridging Mathematics and Clinical Translation.” Edited by Richard M White. *eLife* 12 (March): e84263. <https://doi.org/10.7554/eLife.84263>.
- Zhang, Jingsong, Jessica J. Cunningham, Joel S. Brown, and Robert A. Gatenby. 2017. “Integrating Evolutionary Dynamics into Treatment of Metastatic Castrate-Resistant Prostate Cancer.” *Nature Communications* 8 (1): 1816. <https://doi.org/10.1038/s41467-017-01968-5>.