PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Gluzman M, Scott JG, Vladimirsky A. 2020 Optimizing adaptive cancer therapy: dynamic programming and evolutionary game theory. *Proc. R. Soc. B* **287**: 20192454.
http://dx.doi.org/10.1098/rspb.2019.2454

Received: 31 October 2019
Accepted: 26 March 2020

Subject Category:
Evolution

Subject Areas:
health and disease and epidemiology, systems biology, evolution

Keywords:
adaptive therapy, optimal treatment policy, evolutionary game theory, Hamilton–Jacobi–Bellman equation, tumour heterogeneity

Author for correspondence:
Alexander Vladimirsky
e-mail: vladimirsky@cornell.edu

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.4931604.

THE ROYAL SOCIETY PUBLISHING

Optimizing adaptive cancer therapy: dynamic programming and evolutionary game theory

Mark Gluzman[1], Jacob G. Scott[2] and Alexander Vladimirsky[3]

[1]Center for Applied Mathematics, Cornell University, Ithaca, NY, USA
[2]Department of Translational Hematology and Oncology Research, Cleveland Clinic, Cleveland, OH, USA
[3]Department of Mathematics and Center for Applied Mathematics, Cornell University, 561 Malott Hall, Ithaca, NY 14853-4201, USA

MG, 0000-0002-8404-4232; JGS, 0000-0003-2971-7673; AV, 0000-0003-4284-4546

Recent clinical trials have shown that adaptive drug therapies can be more efficient than a standard cancer treatment based on a continuous use of maximum tolerated doses (MTD). The adaptive therapy paradigm is not based on a preset schedule; instead, the doses are administered based on the current state of tumour. But the adaptive treatment policies examined so far have been largely ad hoc. We propose a method for systematically optimizing adaptive policies based on an evolutionary game theory model of cancer dynamics. Given a set of treatment objectives, we use the framework of dynamic programming to find the optimal treatment strategies. In particular, we optimize the total drug usage and time to recovery by solving a Hamilton–Jacobi–Bellman equation. We compare MTD-based treatment strategy with optimal adaptive treatment policies and show that the latter can significantly decrease the total amount of drugs prescribed while also increasing the fraction of initial tumour states from which the recovery is possible. We conclude that the use of optimal control theory to improve adaptive policies is a promising concept in cancer treatment and should be integrated into clinical trial design.

1. Background

Intratumoural heterogeneity is increasingly recognized as a cause of metastasis, progression and resistance to therapy [1]. While genetic instability, a hallmark of malignancy [2], can result in this heterogeneity, it is being increasingly understood that eco-evolutionary factors, like selection and clonal interference, can also drive and maintain it [3,4].

While sequencing technologies have enabled increasingly in-depth quantitative understanding of the genetic heterogeneity, relatively little experimental work has sought to directly quantify the eco-evolutionary interactions involved. As more studies come to light showing the efficacy of treatments based on eco-evolutionary trial designs, this lack of quantification is coming into focus.

In line with standard, cell-autonomous growth-based theories, conventional chemotherapy is given to patients at the *maximum tolerated doses* (MTD): the highest doses that most patients can safely tolerate. Although the MTD-based therapy offers advantages in survival compared to no therapy, cures remain elusive, and side effects can be severe. In addition to the toxicity, it is known that relapse is nearly inevitable due to the emergence of therapeutic resistance: a process driven by Darwinian evolutionary dynamics in which the MTD-based chemotherapy kills off the chemotherapy-sensitive cells, and chemo-refractory cells eventually dominate in the tumour. While it is unknown whether these resistant cells are present before therapy or acquire resistance mutations during therapy, it is the process of variation and selection under standard therapy that drives the inevitable failure in the patient.

Metronomic chemotherapy has been proposed as a possible alternative to the MTD strategy [5–7]. Metronomic chemotherapy is given in an on/off fashion at

frequent time intervals according to a set periodic schedule. The idea is to give less overall medication, allowing therapy-sensitive cells to regrow, keeping the tumour sensitive to the therapy, while at the same time mitigating toxicity. This decrease in toxicity has further been postulated to prevent tumour angiogenesis and stimulate resensitization of anticancer immune protection. Frustratingly, the results of clinical trials of metronomic chemotherapy have been 'variable' [8] and some of them have *not* demonstrated significant efficacy [9–11] compared with standard therapy. In recent clinical trials, drug dosage intensity was optimized for a *preset schedule*, with a positive effect on toxicity [12]. However, the problem of finding an appropriate schedule has not been addressed, even though models suggest that timing is equally important [13,14].

Based on the hypothesis that disease dynamics depend on the evolution of tumour heterogeneity as modulated by competition between subtypes, the idea of using *adaptive therapy* (AT) has been proposed [15]. AT is much like metronomic therapy, with an important difference: the doses of therapy are administered according to the current *state* of tumour growth and its anticipated evolutionary changes (*trajectory*). These can be estimated using direct (e.g. taking biopsies) or indirect (e.g. antigen testing, mathematical modelling) methods. Therefore, unlike the MTD-based or metronomic protocols, AT *does not have a preset schedule*. The adaptive adjustment of doses and timing also prolongs the period until a tumour becomes chemotherapy-resistant. Recently, the adaptive strategies have shown promise in pre-clinical trials of breast cancer [16] and a phase 2 clinical trial in metastatic castrate-resistant prostate cancer [17].

These two recent successes [16,17] in AT have been based on mathematical modelling of tumour evolution under therapy using a dynamical systems approach based on evolutionary game theory (EGT) [18,19]. This formalism explicitly considers interactions between sub-populations and models their fitness in frequency-dependent terms.[1] EGT has been used to theoretically consider many scenarios in cancer before, including therapy scheduling and timing in prostate cancer [13,17,22–24]; the use of tumour microenvironment targeting therapy in glioblastoma [25]; the trade-off between healthy tissue and cancer in multiple myeloma [26,27]; and drug resistance in general [28–31]. These theoretical studies, combined with the recent empiric realizations, suggest significant opportunities to improve therapy by using this evolutionarily enlightened approach. Nevertheless, therapeutic decisions in general practice are currently *not* based on this knowledge and continue to use the MTD paradigm.

Even if we make very strong assumptions that an oncologist has perfect information about the current state of a tumour and a faithful mathematical model that can predict its trajectory, it is not clear how he/she should adjust the schedule and doses. Based on a stage of the disease and patient's needs, the therapy can have different final goals: maximization of patient's life duration, ensuring the best possible quality of life, decreasing probability of new metastases appearing, decreasing time/cost of the treatment, etc. Unfortunately, an oncologist can usually only focus on one or two of these goals, having some reasonable constraints on the secondary parameters. Thus, an important step toward optimizing AT is to define an *objective* of the therapy and 'translate' it into mathematical language. The next step is to *quantify* how good each particular strategy is with respect to that chosen objective. Optimizing this objective is a mathematical goal which can be addressed by the tools of optimal control theory.

Optimal control theory, a branch of mathematics typically used in engineering, can be also applied to a wide class of problems arising in oncology [32]. The first such application was due to Swan & Vincent [33], who found the optimal treatment strategy for multiple myeloma with the objective to minimize the total amount of drugs by using the Pontryagin minimum principle (PMP) [34]. Since then, others have used the PMP in different cancer treatment problems: a chemotherapy optimization under evolving drug resistance [35–36], optimal scheduling of a vessel disruptive agent [38], MAPK inhibitors [39] input in cancer treatment, minimizing the amount of drugs prescribed in tumour-immune model [40], finding a compromise between drug toxicity and tumour repression for the myeloma bone disease [41], and many others.

While these approaches have offered benefits in their ability to formally optimize problems written as dynamical systems, the PMP method has several limitations. First, PMP yields only a necessary condition for an optimum, and any *locally* optimal trajectory of the control system satisfies PMP. Local optimality means that the trajectory is optimal when comparing it with its small perturbations, but there may well be a different trajectory that is even better (*globally* optimal, compared with *all* possible trajectories). Secondly, PMP provides a time-dependent (open-loop) control: given an initial state, the method provides an optimal treatment strategy as a function of time—therefore a treating oncologist has to follow it regardless of the changing state of the tumour. However, if the underlying model has been perturbed or includes some noise (like a tumour acquiring mutations, say), the control cannot adapt to these unexpected changes.

A significantly different perspective on control theory is based on the notion of feedback (closed-loop) controls. Using the Hamilton–Jacobi–Bellman (HJB) equations, one can obtain controls that depend on the current state of the dynamical system (current distribution of sub-populations of cancer cells) rather than only the current time [42]. In this case, the treating oncologist's decisions can be adjusted if something unexpected has happened with the trajectory. Moreover, the HJB equations guarantee that the resulting treatment feedback strategies are globally optimal. Despite these advantages of the HJB, there are only a few treatment optimization studies [43,44] which use this feedback control paradigm.

Here, we apply the HJB approach to compute optimal treatment strategies for a model of lung cancer proposed by Kaznatcheev *et al*. [20]. In that paper, the authors introduce an evolutionary game (system of replicator-type equations) that models the dynamics of three sub-populations of tumour cells. The article highlights the importance of a good scheduling in the polyclonal regime, when the game has cyclic dynamics. The article has an example of two different scheduling strategies with the same set of initial parameters that lead the system to opposite outcomes: putative recovery, versus putative death of a patient. While several qualitatively different treatment schedules are presented, optimal therapy is not discussed. Given the growing interest in EGT in clinical applications [17] and recent work connecting these models using direct *in vitro* parametrization [45], we believe the optimization of therapies based on such models will become increasingly important and the HJB-based approach will be used far more often in the future.

Box 1. Mathematical model of cancer sub-population evolution from Kaznatcheev *et al.* [20].

Transformation/reduction to a 2D system:
(x_G, x_D, x_V) for GLY, DEF and VOP, respectively.

Note: $x_G + x_D + x_V = 1$.

Evolution dynamics in reduced coordinates with control on therapy intensity:

Subpopulation proportions:

$$\begin{cases} q = \frac{x_V}{x_V + x_D}, \\ p = x_G, \end{cases} \quad \text{or} \quad \begin{cases} x_D = (1-q)(1-p), \\ x_G = p, \\ x_V = (1-p)q. \end{cases} \tag{2.1}$$

$$\begin{cases} \dot{q}(t) = q(t)\left(1 - q(t)\right)\left(\frac{b_v}{n+1}\sum_{k=0}^{n} p^k(t) - c\right), \\ \dot{p}(t) = p(t)\left(1 - p(t)\right)\left(\frac{b_a}{n+1} - (b_v - c)q(t) - d(t)\right); \\ q(0) = q_0,\ p(0) = p_0. \end{cases} \tag{2.2}$$

Control and parameters:

— $d : \mathbb{R}_+ \to [0, d_{\max}]$, time-dependent intensity of GLY-targeting therapy;

— b_a, the benefit per unit of acidification;

— b_v, the benefit from the oxygen per unit of vascularization;

— c, the cost of production VEGF;

— n, the number of cells in the interaction group.

Conditions for homogeneous regime:

$$\frac{b_a}{n+1} < b_v - c < cn. \tag{2.3}$$

Process *terminates* as soon as either

$$\begin{cases} p(t) < r_b, & \text{if therapy succeeds;} \\ p(t) > 1 - f_b, & \text{if therapy fails.} \end{cases}$$

Terminal set:

$$\Delta = \left\{(q, p) \in [0, 1] \times [0, 1] : p < r_b \text{ or } p > 1 - f_b\right\}. \tag{2.4}$$

2. Methods

We illustrate our approach on a model of cancer evolution that has been proposed in Kaznatcheev *et al.* [20] and summarized in box 1 with the key steps of derivation included in electronic supplementary material, Section 1S. This model considers interactions between three different sub-populations of cancer cells playing a modified version of the public goods social dilemma: glycoltyic cells (GLY), vascular overproducers (VOP) and cells called defectors (DEF) which use both strategies to 'cheat' on the others. While this model is highly simplified, the qualitative aspects of these cell types have been well documented in tumours. GLY cells are anaerobic and produce lactic acid, these are classic 'Warburg shifted' cells, and have been implicated in the acid-mediated invasion hypothesis [46]. These cells are largely responsible for local progression and metastasis, and while it has been suggested they can be treated with buffer therapy [47], this has yet to be shown to be clinically actionable; therefore, adaptive dynamics remain a viable approach to control this population. VOP cells spend extra energy to produce HIF-1α, a signalling molecule which drives further production of Vascular Endothelial Growth Factor (VEGF—a protein that works to improve the vasculature), which ultimately benefits both VOP and DEF cells since they require oxygen for aerobic respiration. For this reason, the VOP cells could be targeted by drugs like bevacizumab, which is an anti-body against VEGF. DEF cells, the 'cheaters', have analogues in many cancers, and have no cell-autonomous advantages, but instead depend on the ecological interactions with other types for survival. These cells are best targeted by modifying the ecology; e.g. via AT.

Based on the replicator model from EGT [18,19], transformed into equation (2.2) in box 1, the evolution of the tumour is described by tracking the changing *proportions* of GLY, VOP and DEF cells in the full population. The patient is viewed as recovered when the GLY proportion falls below some low threshold r_b. (Below this *recovery barrier*, the validity of replicator-based model is harder to justify and we assume that the GLY cells are essentially extinct.) Conversely, we assume that GLY cells *suppress* other tumour cells and a patient dies if the total proportion of aerobic cells (VOP and DEF sub-populations combined) falls below some low threshold (a *failure barrier*) f_b.

For a range of parameter values (2.3), this model predicts a heterogeneous regime[2] in cancer evolution with coexistent and oscillating proportions of GLY, VOP and DEF. Without any treatment, these sub-populations follow cyclic dynamics and a patient never recovers (figure 1*a*).

Following section 4.1. in [20], we consider a cell-type-targeting therapy that preferentially penalizes the fitness of GLY cells; see formula (2.2) in box 1. But we emphasize that the same approach can be used to target any cell-type and optimize therapy for any chosen end-condition. During the treatment, a doctor defines the timing of therapy and its intensity. This time-dependent intensity $d(t)$ can vary between 0 (no therapy) and $d_{\max} > 0$ (the MTD). Two extreme cases ($d(t) = 0$ versus $d(t) = d_{\max}$ for all times t) are illustrated in figure 1*a*,*b*, respectively.[3] In the latter case, GLY cells become extinct and the patient recovers quickly; however, it is natural to ask whether this treatment strategy is optimal in some sense (e.g. would the recovery be much delayed if the patient received therapy less often or at a lower intensity?). In the following section, we will show that the MTD-based treatment can result in an avoidably high cumulative amount of drugs (figure 2*c*) or might even fail to achieve a recovery in situations where AT-based treatment would have succeeded (figure 4), much like the early results from Zhang *et al.* in metastatic prostate cancer [17].

For the Kaznatcheev *et al.* tumour model (2.2), a natural objective function to minimize is the total amount of therapy administered over the course of treatment (which in this case

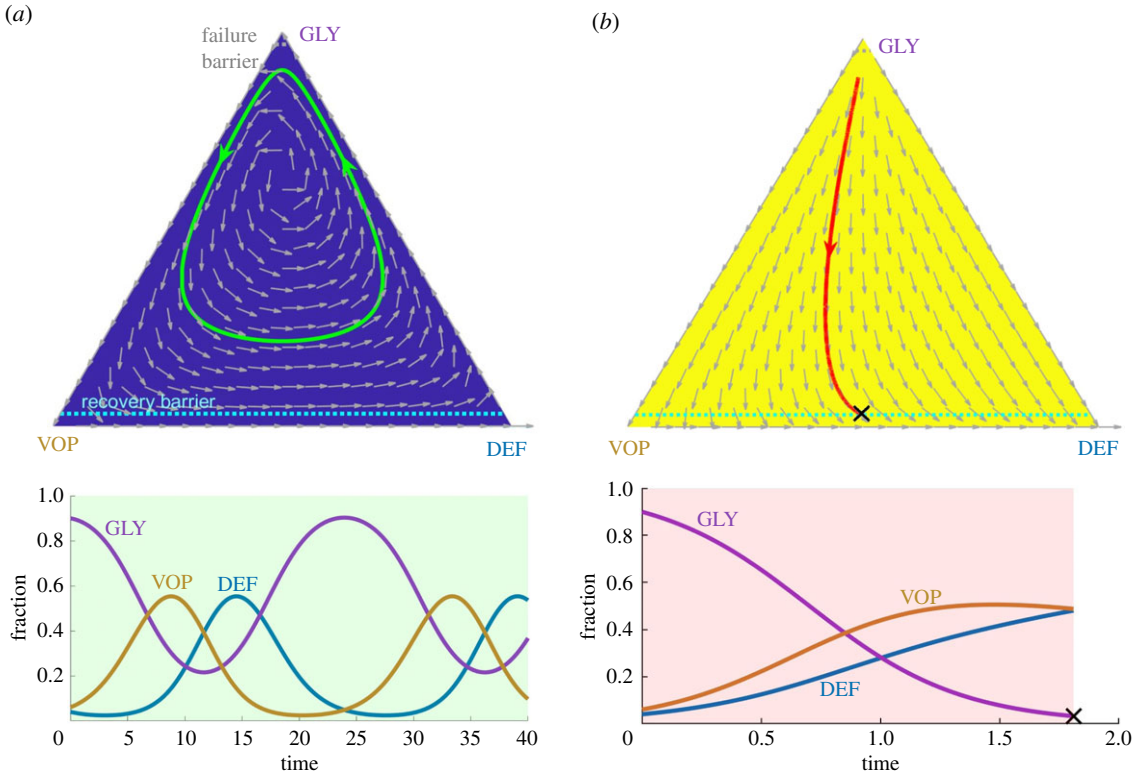
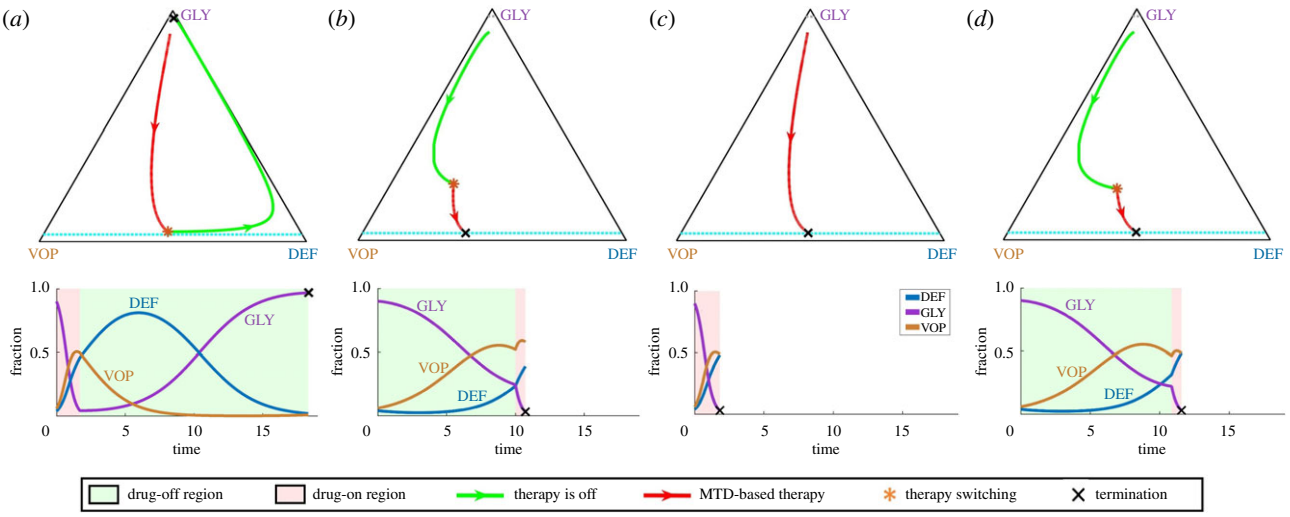


Figure 1. A comparison of two possible constant treatment scenarios starting from an initial state $(x_D, x_G, x_V) = (0.04, 0.9, 0.06)$: (*a*) without any therapy; (*b*) with the MTD-based therapy. Top row: phase portraits of corresponding vector fields (shown by *grey arrows*) on a GLY–VOP–DEF triangle with illustrative trajectories. *Blue background* and *green reference trajectory*—no therapy at all. *Yellow background* and *red reference trajectory*—MTD-based therapy at all times. *Dash light blue* and *grey lines* separate the recovery zone (bottom) and the failure zone (top), respectively. *Black cross*—termination due to crossing the recovery barrier. Bottom row: evolution of sub-populations with respect to time based on the reference trajectories above. *Green time range*—no therapy. *Pink time range*—MTD-based therapy. Note the different scaling of the time axis. (Online version in colour.)



subfigure	policy	total time	time till switching (denoted by *)	overall cost
(*a*)	'treat immediately but not long enough'	18.32	1.7	$+\infty$
(*b*)	'start treating after 10 time steps'	10.73	10	2.29
(*c*)	'MTD-based'	1.81	—	5.45
(*d*)	'optimal'	11.56	10.86	2.25

Figure 2. Importance of optimal scheduling: ensuring recovery and decreasing the cost of treatment. Tumour evolution under four different treatment strategies for the same initial state. A seemingly reasonable treatment strategy may not lead to a recovery; see (*a*). Even if a patient eventually recovers (as in (*b*,*c*)), the overall cost of treatment can be reduced by pursuing a provably optimal policy (*d*). Top row: tumour evolutionary trajectories under different strategies on a GLY-VOP-DEF triangle. *Green part of a trajectory*—no therapy is used. *Red part of a trajectory*—MTD-based (standard) therapy. The moment of switching from one regime to another is denoted by (red *). *Dash light blue* and *grey lines* separate the recovery zone and the failure zone, respectively. *Black cross*—termination due to crossing the failure or recovery barrier. Bottom row: evolution of sub-populations with respect to time based on the reference trajectories above. *Green time range*—no therapy. *Pink time range*—MTD-based therapy. (Online version in colour.)

Box 2. Objective function.

Total treatment (terminal) time: $T(q_0,\ p_0, d(\cdot)) = \min\{t \in \mathbb{R}_+ | (q(t), p(t)) \in \Delta, q(0) = q_0, p(0) = p_0\}.$ (3.1)

If the system never gets to the terminal set, we assume that $T(q_0, p_0, d(\cdot)) = +\infty$.

Terminal cost function is $g(q, p)$: $\Delta \to \{0, +\infty\}$ such that $g(q,p) = \begin{cases} +\infty, & \text{if } p > 1 - f_b, \\ 0, & \text{otherwise.} \end{cases}$ (3.2)

Treatment cost (objective) function: $J(q_0, p_0, d(\cdot)) = \int_0^T (d(s) + \sigma) ds + g(q(T), p(T)),$ (3.3)

where $T := T(q_0, p_0, d(\cdot))$ is the terminal time. J is finite if the system (2.2) terminates at the recovery barrier.

Value function $u(q_0, p_0) = \inf_{d(\cdot)} J(q_0, p_0, d(\cdot))$ can be found by solving HJB PDE:

$$\min_{d \in [0, d_{\max}]} \left\{ \nabla u(q, p) \cdot \begin{pmatrix} \dot{q}(q, p, d) \\ \dot{p}(q, p, d) \end{pmatrix} + d + \sigma \right\} = 0, (q, p) \in ([0, 1] \times [0, 1]) \setminus \Delta. \quad (3.4)$$

The boundary conditions of HJB equation: $\begin{cases} u(q,p) = 0, & \text{if } p < r_b; \\ u(q,p) = +\infty, & \text{if } p > 1 - f_b. \end{cases}$ (3.5)

Once u and ∇u are computed, they can be used to obtain the optimal control *in feedback form:* $d^* = d(q, p)$.

could be a surrogate for both toxicity and cost). This can be quantified as $D = \int_0^T \mathrm{d}(t)\,\mathrm{d}t$, where the total time of treatment T is dependent on the initial cancer subpopulation fractions and on our chosen therapy policy $d(\cdot)$. However, this objective is problematic for two reasons. First, the minimum of D is clearly attained without any therapy (taking $d(t) \equiv 0$ implies $D = 0$ and $T = +\infty$), even though the dynamics become cyclic and the recovery is never achieved. Second, if we constrain our minimization to only those $d(\cdot)$ that lead to recovery, an optimal treatment policy does not exist. Instead, there is a sequence of treatment policies that lead to successively smaller d-values but with an unbounded increase in corresponding treatment times T. The idea of such policies is simple: travel along the therapy-free trajectories of figure 1*a* for most of the time, but use short bursts of therapy only when the drugs are most effective. To approach the optimally small d, one would need to use shorter and shorter bursts, resulting in policies that are hard to implement in practice and would require unrealistically long treatment times T, but would yield a situation like a chronic disease, where while the tumour is never cured, it is always controlled.

In order to get a meaningful optimal policy we will penalize the treatment time by a *time penalty* $\sigma > 0$. The total time spent on the treatment, including time between the doses, is an important factor by itself. Much longer treatment time results in additional costs and lower quality of life for a patient. Therefore, our objective is to minimize the sum of a *therapy cost* D and a *treatment time cost* σT, while guaranteeing eventual recovery. For every choice of $\sigma > 0$, the resulting treatment policies are thus *Pareto-optimal* with respect to D and T.

Our method computes the value function, u, which is defined for every starting tumour state as the minimum of $(\int_0^T \mathrm{d}(t)\,\mathrm{d}t + \sigma T)$ over the set of treatment policies that lead to recovery. (If recovery is not possible then $u = +\infty$.) Any policy $d(\cdot)$ that realizes this minimum is called optimal. Due to the structure of this optimization problem, one can show that optimal treatment policies are *bang–bang*: at any given time t, they either administer drugs at MTD-rate ($d(t) = d_{\max}$) or administer no drugs at all ($d(t) = 0$); see electronic supplementary material, §2S. For such policies, the objective function becomes a weighted sum of the total *therapy time* $\widetilde{T}$ (when $d(t) \equiv d_{\max}$) and the total treatment time T, with $d_{\max}$ and σ as the corresponding weights. Moreover, this allows for a simple visual representation of any such policy: splitting the full state space into two parts (MTD dynamics versus no-therapy dynamics, shown in yellow and blue, respectively, in all of our figures) and simplifies application of therapy into something familiar to clinicians: on or off.

3. Results

(a) Quantifying the benefits of optimal treatment strategies

We now focus on an optimal control problem summarized in box 2 and based on the example considered in the *therapeutic implications* section of Kaznatcheev *et al.* [20]. All model parameter values ($d_{\max} = 3$, $b_a = 2.5$, $b_v = 2$, $c = 1$, $n = 4$) correspond to those in [20], except for the recovery and failure barriers[4] $r_b = f_b = 10^{-1.5}$. We also use $\sigma = 0.01$ to incorporate the time-penalty absent in the original model. In figure 2, we compare the treatment cost (3.3) and treatment time (3.1) of trajectories corresponding to four different treatment strategies starting from the same initial configuration $(x_D, x_G, x_V) = (0.04, 0.9, 0.06)$. The first two strategies are similar to those modelled in Kaznatcheev *et al.* [20]: (a) is an example of a bad policy that may cause a failure by stopping the therapy prematurely, while (b) is a good policy based on ad-hoc adjustment of the start time for the therapy. We also illustrate the standard of care 'MTD-based' policy (c). Even though both (b) and (c) lead to recovery, neither of these is optimal (with the MTD-based approach resulting in an excessive amount of drugs, captured by the higher cost). The policy minimizing our objective function is found by solving the HJB equation and illustrated in (d).

The corresponding therapy on/off regions and the resulting vector field are shown in figure 3*a*. The zoomed version shows that trajectories can be prevented from crossing the failure barrier by using the MTDs just before crossing. In fact, a *chattering control* (with intermittent and sufficiently frequent use of MTDs) would be sufficient to guarantee this.

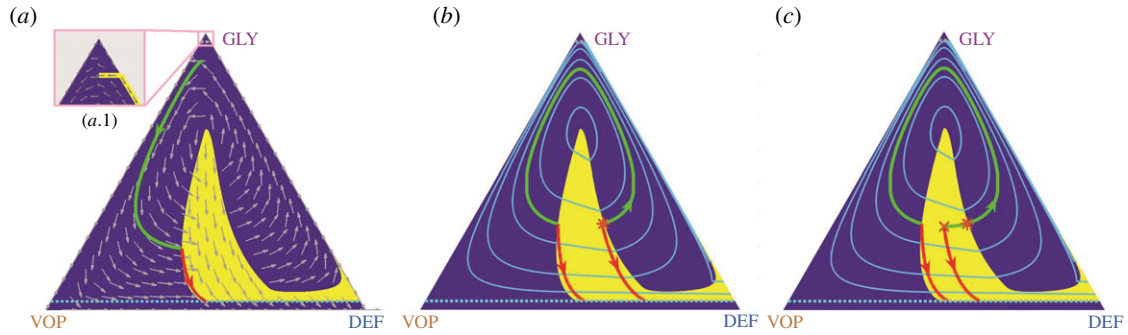


Figure 3. Optimal control in feedback form, the value function, and the pitfalls of PMP. (*a*) A phase portrait of the optimal system dynamics. The vector field is shown by *grey arrows* over the optimal drugs-off (*blue background*) and drugs-on (*yellow background*) regions. A sample optimal trajectory (in green and red) corresponds to the initial state from figure 2. (*b*) Computation of the value function u (whose level curves are shown by *light blue lines*) is used to determine the optimal drugs-on and drugs-off regions (shown in yellow and dark blue, respectively). Optimal trajectories are not unique for initial states on the shockline (where the level curves of u are not smooth). Two such optimal trajectories are shown starting from an asterisk (*). The green-red trajectory takes longer to reach the recovery, but uses less drugs than the red (start-drugs-right-away) trajectory. The cumulative cost is the same for both of them. (*c*) For initial conditions off the shocklines of u, there can still be multiple *locally optimal* trajectories. We show an example of two such trajectories starting from a cross marker (×). The risk of applying the PMP method is that it might yield either of them, but only the red (start-drugs-right-away) is globally optimal. (Online version in colour.)

The level sets of u in figure 3*b* show that value functions need not be smooth. Since the gradient of u is used to determine the optimal course of action (therapy on or off), there can actually be more than one optimal policy for initial states on a *shockline* (where the gradient is undefined). We show an example of such trajectories (solid green and red lines in figure 3*b*, both yielding the same cost of 2.764) for an initial point denoted by (*). Non-smoothness of the value function often poses a challenge for methods based on PMP [34] even if the initial state *is not* on a shockline. For example, perturbing the initial state to a nearby one, denoted by a cross (×) in figure 3*c*, one sees two *locally optimal* trajectories and PMP might yield either of these depending on the initial guess. The green one is, however, inferior to the *globally* optimal red trajectory, which can be always recovered by solving the HJB equation.

The trade-offs between our two optimization objectives (the total administered drugs versus the time to recovery) are further examined in electronic supplementary material, §S5.

(b) 'Incurable' states and periodic trajectories under MTD treatment

One might think that, despite being sub-optimal, an aggressive MTD-based strategy is at least always fully reliable and the resulting trajectories are guaranteed to reach the recovery zone from every initial configuration, as shown in figure 2*c*. Indeed, if $b_a/(n+1) \le d_{\max}$, the MDT-based policy ($d(t) \equiv d_{\max}$) guarantees that $\dot{p}$ is always negative; see equation (2.2). But with $b_a/(n+1) > d_{\max}$ the recovery might not be attained with the constant use of MTDs (even if some other treatment policies are successful!).

Consider, for example, the following set of parameters: $b_a = 4$, $b_v = 2$, $c = 1$, $n = 4$; $r_b = f_b = 10^{-1.5}$; $d_{\max} = 0.3$, $\sigma = 0.03$ and an initial state $(x_D, x_G, x_V) = (0.02, 0.8, 0.18)$ denoted by (*) in figure 4. Under these parameters, the MTD-based therapy has a periodic trajectory[5] (figure 4*b*). Since the treatment time is infinite, the cost (3.3) of such a policy is $+\infty$. (In reality, this would lead to the emergence of drug resistance and eventual failure, but this biological situation is not modelled in Kaznatcheev *et al*. [20].)

We can see that neither of two extreme strategies ('no-drugs-at-all' in figure 4*a* and the MTD-based 'drugs-all-the-time' in figure 4*b*) can bring the trajectory to the recovery zone. However, their adaptive combination can still achieve the objective. We show a trajectory corresponding to the optimal policy in figure 4*c*. With a larger failure zone (e.g. $r_b = f_b = 10^{-1}$), a previously successful MTD-based treatment might even result in death (figure 4*d*), while the adaptive strategy still leads to recovery (figure 4*e*).

For a fixed treatment policy, we define its corresponding 'incurable' area to be a set of states starting from which it is impossible to cross the recovery barrier. For example in figure 4, the incurable area of the MTD-based policy includes the state denoted by (*) (when $r_b = f_b = 10^{-1.5}$ or $r_b = f_b = 10^{-1}$). However, this state is *not* in the (dramatically smaller) incurable area of the adaptive/optimal policy (figure 5). Of course, the 'incurable areas' are also highly dependent on model parameters. In electronic supplementary material, §S6, we show that they can grow due to an increase in the MTD rate $d_{\max}$ or a decrease in vascularization benefits b_v.

Starting from any incurable configuration, one could similarly pose a different control problem of *maximizing* the time until crossing the failure barrier. While we do not address it here, we note that the HJB approach would be quite suitable to find optimal treatment policies for this problem as well.

4. Discussion

By now, it is widely accepted that cancer is an evolutionary process, and that variation and selection drive the emergence of drug resistance. While this new knowledge is driving cancer research forward, it has largely not yet affected the practice, with the majority of clinical protocols relying on MTD-based approaches, which invariably fail in the setting of most metastatic disease. This is changing with the advent of AT—therapeutic strategies specifically designed with a changing regimen prescribed: one that adapts to an evolving tumour [48].

To optimally design AT protocols, the underling dynamics of the tumour growth and treatment response must be known. While the methods of learning these dynamics are still in their infancy, the last decade has seen

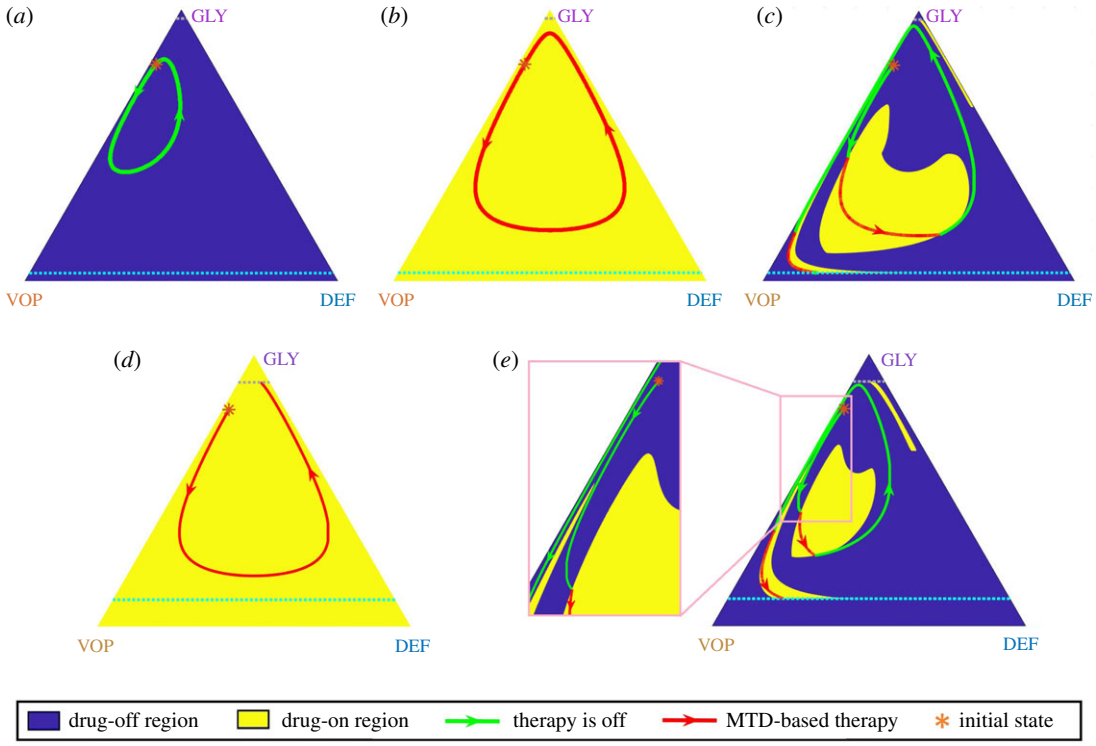


Figure 4. MTD-based policy versus the optimal (adaptive) policy when the MTD rate is low ($d_{max} < b_a/(n+1)$). Top row: trajectories under both (*a*) 'no therapy' policy and (*b*) the MTD-based policy are cyclic and cannot cross either the recovery or the failure barrier from the initial state denoted by (*). Nevertheless, the adaptive/optimal switching leads to a full recovery (*c*). Bottom row: With a larger failure zone, an MTD-based policy leads to patient's putative death (*d*) even though it is still possible to cross the recovery barrier under an adaptive/optimal policy (*e*). (Online version in colour.)

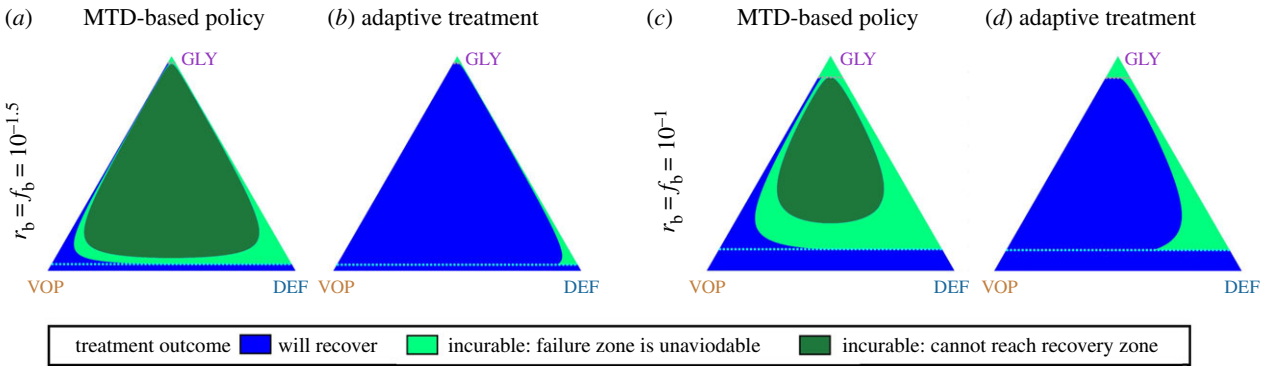


Figure 5. (*a–d*) Comparison of the 'incurable' area for the MTD-based policy versus the adaptive policy. Unlike the MTD-based treatment, adaptive policies lead to a patient recover from most initial states even with a small d_{max}. *The blue colour* indicates initial states from which the recovery is achievable. *The green colours* represent initial states from which the recovery is impossible (i.e. the 'incurable' area): *the light green*—trajectories eventually get into the failure zone; *the dark green*—trajectories are cyclic and cannot cross either the recovery or failure barriers. (Online version in colour.)

a flurry of activity using EGT and other evolutionary models for a range of prescribed dynamics. These works have shown qualitative changes in tumour behaviour in a range of treatment scenarios, and importantly, demonstrated that the sequence [20,49–51] and timing [13,20] of therapy can drastically change the outcomes. As we come closer to the reality of evolutionarily designed therapeutic trials in the mainstream, it is important to develop methods that do not just improve outcomes but aim to formally optimize them.

However, before using any optimization tools, one needs to choose a specific quantifiable criterion for comparing the outcomes. Once that criterion is selected and the underlying mathematical model is sufficiently accurate, the best treatment strategy can be found by the techniques of optimal control theory. In this paper, we show how this can be done for one particular heterogeneous cancer model previously described in Kaznatcheev *et al.* [20]. We show that the optimal treatment policy can have multiple regimes: always on $d^*(\cdot) \equiv d_{max}$, always off $d^*(\cdot) \equiv 0$, and involving several contiguous treatment periods. For the latter, the challenge is to accurately approximate the on/off 'switching curves' in the state space. We show that the definition of optimal treatment policy is heavily dependent on a parameter σ describing the relative importance of minimizing the total amount of drugs versus the total time to recovery. We further show that, for some parameter regimes, there are 'incurable regions' in the state space—the starting configurations that will not lead to a recovery regardless of the chosen treatment strategy—suggesting an alternative therapy (or goal) should be considered. Moreover, for some starting configurations, the 'always on' treatment might not lead to a recovery even if it is achievable with some on/off hybrid strategies.

Just like any other model, the approach in Kaznatcheev *et al*. [20] is based on simplifying assumptions (e.g. only the subpopulation fractions are important, and no novel types can arise), which limit its practical applicability. But our message is broader, and we use this specific model primarily to illustrate the general optimization approach applicable to more detailed cancer evolution equations or even in data-driven/equation-free framework.

Of the two main approaches of optimal control theory, the PMP has been much more widely used in cancer treatment research up until now. By contrast, our approach here is based on dynamic programming and the numerical methods for HJB equations. The higher computational cost of these methods is balanced by several important practical considerations. First, they yield a policy in feedback-form and are thus more robust to modelling/measurement errors. Second, they always return the globally optimal treatment strategies and avoid some of the pitfalls well-known for the PMP-based methods (e.g. figure 3*b*). With the advent of efficient numerical methods, we posit that the HJB equations will be soon playing a larger role in treatment optimization.

There are several obvious directions for extending our approach. First, the ability to optimize outcomes for a *range* of criteria will open new avenues to quantify physician/patient discussions (on trade-offs and personalized therapy) that have previously been only qualitative. The computation of optimal policies for different values of σ (covered in the electronic supplementary material) can be viewed as a small step in this direction. But there are also many other possible optimization criteria of practical interest. Methods for approximating *all* Pareto-optimal policies are available [52] but are usually more computationally challenging. For probabilistic cancer evolution models, one can also choose between optimizing different characteristics of the same random quantity (e.g. minimize the average time-to-recovery versus maximizing the probability of recovery in the next year). Finally, one can also use the choice of criterion to promote *robustness* by systematically treating possible measurement/modelling errors as perturbations chosen by some adversarial player. Such 'games-against-nature', as described recently by Stanková *et al*. [48], can be similarly treated by solving Hamilton–Jacobi–Isaacs equations.

While we believe that our optimization approach will have a major role in design of future clinical trials of AT, the presented version is not yet sufficiently practical. One important limitation is our assumed full knowledge of the system state: the exact subpopulation fractions are needed at every point in time to decide whether to administer the drugs. In practice, one can periodically obtain an approximation of these quantities (e.g. based on a repeat biopsy), but most of the time the decisions must be made based on some less invasive measurements (e.g. based on PSA-levels [17] or on circulating cell-free DNA [53]). A rigorous treatment of such *partially observable controlled processes* [54] would require a much more detailed model of 'tumour state' uncertainties. As an intermediate step, we would recommend validation of such methods *in vitro* and *in vivo*, using measured, instead of prescribed games. This is a topic of active interest to our group [45,55], and would be prudent to address before any clinical trials.

Data accessibility. This article has no additional data.

Competing interests. We declare we have no competing interests.

Funding. M.G. acknowledges the support from the Institute for Data and Decision Analytics at The Chinese University of Hong Kong, Shenzhen during his visit when this research was completed. J.G.S. would like to thank the NIH Case Comprehensive Cancer Center support grant no. P30CA043703 and the Calabresi Clinical Oncology Research Program, National Cancer Institute Award number K12CA076917. A.V. would like to thank the Simons Foundation for its fellowship support and the National Science Foundation (award DMS-1738010) for supporting development of numerical methods for Hamilton–Jacobi equations. A part of this work was performed during a sabbatical leave at Princeton/ORFE, and A.V. is grateful to ORFE Department for its hospitality.

Endnotes

[1]Using frequency dependence only (not strictly tumour size) is a strong assumption in cancers which typically are growing. But while the assumption of 'constant population' is standard in EGT, a mapping to exponential growth has been also considered without loss of generality [20,21].

[2]Other tumour regimes (fully angiogenic and glycolyctic) also exist outside of this parameter range. They are less interesting from the point of view of treatment strategies, but we still consider them for the sake of completeness in electronic supplementary material, §S7.

[3]Model parameter values and initial states of trajectories are specified for all figures in electronic supplementary material, §S4.

[4]This change in parameter values is meant to decrease the computational cost of our numerical approach (see electronic supplementary material, §S3). The original $r_b = f_b = 10^{-4}$ from Kaznatcheev *et al*. [20] would require computations on a much finer mesh.

[5]This is easy to prove by redefining the parameter $b_a := b_a - (n+1)\,d_{\max} > 0$ and reducing the MTD-based case to the periodic behaviour of the uncontrolled system in the heterogeneous parameter regime; i.e. equation (2.2) with $d(t) \equiv 0$.

References

1. Marusyk A, Polyak K. 2010 Tumor heterogeneity: causes and consequences. *Biochim. Biophys. Acta Rev. Cancer* **1805**, 105–117. (doi:10.1016/j.bbcan.2009.11.002)
2. Hanahan D, Weinberg RA. 2000 The hallmarks of cancer. *Cell* **100**, 57–70. (doi:10.1016/S0092-8674(00)81683-9)
3. Marusyk A, Tabassum DP, Altrock PM, Almendro V, Michor F, Polyak K. 2014 Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature* **514**, 54–58. (doi:10.1038/nature13556)
4. Scott J, Marusyk A. 2017 Somatic clonal evolution: a selection-centric perspective. *Biochim. Biophys. Acta Rev. Cancer* **1867**, 139–150. (doi:10.1016/j.bbcan.2017.01.006)
5. Browder T, Butterfield CE, Kräling BM, Shi B, Marshall B, O'Reilly MS, Folkman J. 2000 Antiangiogenic scheduling of chemotherapy improves efficacy against experimental drug-resistant cancer. *Cancer Res.* **60**, 1878–1886.
6. Hanahan D, Bergers G, Bergsland E. 2000 Less is more, regularly: metronomic dosing of cytotoxic drugs can target tumor angiogenesis in mice. *J. Clinical Invest.* **105**, 1045–1047. (doi:10.1172/JCI9872)
7. Klement G, Baruchel S, Rak J, Man S, Clark K, Hicklin DJ, Bohlen P, Kerbel RS. 2000 Continuous low-dose therapy with vinblastine and VEGF receptor-2 antibody induces sustained tumor regression without overt toxicity. *J. Clinical Invest.* **105**, R15–R24. (doi:10.1172/JCI8829)
8. Pasquier E, Kavallaris M, André N. 2010 Metronomic chemotherapy: new rationale for new directions. *Nat. Rev. Clinical Oncol.* **7**, 455–465. (doi:10.1038/nrclinonc.2010.82)

9. Kesari S *et al.* 2007 Phase II study of metronomic chemotherapy for recurrent malignant gliomas in adults. *Neuro Oncol.* **9**, 354–363. (doi:10.1215/15228517-2007-006)
10. Senerchia AA *et al.* 2017 Results of a randomized, prospective clinical trial evaluating metronomic chemotherapy in nonmetastatic patients with high-grade, operable osteosarcomas of the extremities: a report from the Latin American Group of Osteosarcoma Treatment. *Cancer* **123**, 1003–1010. (doi:10.1002/cncr.30411)
11. Steinbild S *et al.* 2007 Metronomic antiangiogenic therapy with capecitabine and celecoxib in advanced tumor patients: results of a phase II study. *Oncol. Res. Treat.* **30**, 629–635. (doi:10.1159/000110580)
12. Hénin E, Meille C, Barbolosi D, You B, Guitton J, Iliadis A, Freyer G. 2016 Revisiting dosing regimen using PK/PD modeling: the MODEL1 phase I/II trial of docetaxel plus epirubicin in metastatic breast cancer patients. *Breast Cancer Res. Treat.* **156**, 331–341. (doi:10.1007/s10549-016-3760-9)
13. Basanta D, Scott JG, Fishman MN, Ayala G, Hayward SW, Anderson ARA. 2012 Investigating prostate cancer tumour-stroma interactions: clinical and biological insights from an evolutionary game. *Br. J. Cancer* **106**, 174–181. (doi:10.1038/bjc.2011.517)
14. Chen C-S, Doloff JC, Waxman DJ. 2014 Intermittent metronomic drug schedule is essential for activating antitumor innate immunity and tumor Xenograft regression. *Neoplasia* **16**, 84–96. (doi:10.1593/neo.131910)
15. Gatenby RA, Silva AS, Gillies RJ, Frieden BR. 2009 Adaptive therapy. *Cancer Res.* **69**, 4894–4903. (doi:10.1158/0008-5472.CAN-08-3658)
16. Enriquez-Navas PM *et al.* 2016 Exploiting evolutionary principles to prolong tumor control in preclinical models of breast cancer. *Sci. Transl. Med.* **8**, 327ra24. (doi:10.1126/scitranslmed.aad7842)
17. Zhang J, Cunningham JJ, Brown JS, Gatenby RA. 2017 Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nat. Commun.* **8**, 1816. (doi:10.1038/s41467-017-01968-5)
18. Hofbauer J, Sigmund K. 1998 *Evolutionary games and population dynamics*. Cambridge, UK: Cambridge University Press.
19. Smith JM. 1982 *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.
20. Kaznatcheev A, Vander Velde R, Scott JG, Basanta D. 2017 Cancer treatment scheduling and dynamic heterogeneity in social dilemmas of tumour acidity and vasculature. *Br. J. Cancer* **116**, 785–792. (doi:10.1038/bjc.2017.5)
21. Melbinger A, Cremer J, Frey E. 2010 Evolutionary game theory in growing populations. *Phys. Rev. Lett.* **105**, 178101. (doi:10.1103/PhysRevLett.105.178101)
22. Cunningham JJ, Brown JS, Gatenby RA, Staňková K. 2018 Optimal control to develop therapeutic strategies for metastatic castrate resistant prostate cancer. *J. Theor. Biol.* **459**, 67–78. (doi:10.1016/j.jtbi.2018.09.022)
23. You L, Brown JS, Thuijsman F, Cunningham JJ, Gatenby RA, Zhang J, Stanková K. 2017 Spatial vs. non-spatial eco-evolutionary dynamics in a tumor growth model. *J. Theor. Biol.* **435**, 78–97. (doi:10.1016/j.jtbi.2017.08.022)
24. West J, You L, Zhang J, Gatenby RA, Brown JS, Newton PK, Anderson ARA. In press. Towards multidrug adaptive therapy. *Cancer Res.* (doi:10.1158/0008-5472.CAN-19-2669)
25. Basanta D, Scott JG, Rockne R, Swanson KR, Anderson ARA. 2011 The role of IDH1 mutated tumour cells in secondary glioblastomas: an evolutionary game theoretical view. *Phys. Biol.* **8**, 015016. (doi:10.1088/1478-3975/8/1/015016)
26. Dingli D, Offord C, Myers R, Peng KW, Carr TW, Josic K, Russell SJ, Bajzer Z. 2009 Dynamics of multiple myeloma tumor therapy with a recombinant measles virus. *Cancer Gene Ther.* **16**, 873–882. (doi:10.1038/cgt.2009.40)
27. Wu A, Liao D, Tlsty TD, Sturm JC, Austin RH. 2014 Game theory in the death galaxy: interaction of cancer and stromal cells in tumour microenvironment. *Interface Focus* **4**, 20140028. (doi:10.1098/rsfs.2014.0028)
28. Komarova NL, Wodarz D. 2005 Drug resistance in cancer: principles of emergence and prevention. *Proc. Natl Acad. Sci. USA* **102**, 9714–9719. (doi:10.1073/pnas.0501870102)
29. Orlando PA, Gatenby RA, Brown JS. 2012 Cancer treatment as a game: integrating evolutionary game theory into the optimal control of chemotherapy. *Phys. Biol.* **9**, 065007. (doi:10.1088/1478-3975/9/6/065007)
30. West J, Hasnain Z, Mason J, Newton PK. 2016 The prisoner's dilemma as a cancer model. *Convergent Sci. Phys. Oncol.* **2**, 035002. (doi:10.1088/2057-1739/2/3/035002)
31. Yoon N *et al*. 2018 Optimal therapy scheduling based on a pair of collaterally sensitive drugs. *Bull. Math. Biol.* **80**, 1776–1809. (doi:10.1007/s11538-018-0434-2)
32. Schättler H, Ledzewicz U 2015 *Optimal control for mathematical models of cancer therapies*, vol. 42. Interdisciplinary Applied Mathematics. New York, NY: Springer.
33. Swan GW, Vincent TL. 1977 Optimal control analysis in the chemotherapy of IgG multiple myeloma. *Bull. Math. Biol.* **39**, 317–337. (doi:10.1016/S0092-8240(77)80070-0)
34. Pontryagin L, Boltyanskii V, Gamkrelidze R, Mishchenko E. 1962 *The mathematical theory of optimal processes*. New York, NY: John Wiley & Sons.
35. Carrère C. 2017 Optimization of an *in vitro* chemotherapy to avoid resistant tumours. *J. Theor. Biol.* **413**, 24–33. (doi:10.1016/j.jtbi.2016.11.009)
36. Schattler H, Ledzewicz U. 2006 Drug resistance in cancer chemotherapy as an optimal control problem. *Discrete Continuous Dyn. Syst. Series B* **6**, 129–150. (doi:10.3934/dcdsb.2006.6.129)
37. Wang S, Schättler H. 2016 Optimal control of a mathematical model for cancer chemotherapy under tumor heterogeneity. *Math. Biosci. Eng.* **13**, 1223–1240. (doi:10.3934/mbe.2016040)
38. D'Onofrio A, Ledzewicz U, Maurer H, Schättler H. 2009 On optimal delivery of combination therapy for tumors. *Math. Biosci.* **222**, 13–26. (doi:10.1016/j.mbs.2009.08.004)
39. Su Y, Jia C, Chen Y. 2016 Optimal control model of tumor treatment with oncolytic virus and MEK inhibitor. *BioMed Res. Int.* **2016**, 5621313.
40. Ledzewicz U, Naghnaeian M, Schättler H. 2012 Optimal response to chemotherapy for a mathematical model of tumor-immune dynamics. *J. Math. Biol.* **64**, 557–577. (doi:10.1007/s00285-011-0424-6)
41. Lemos JM, Caiado DV, Coelho R, Vinga S. 2016 Optimal and receding horizon control of tumor growth in myeloma bone disease. *Biomed. Signal Process. Control* **24**, 128–134. (doi:10.1016/j.bspc.2015.10.004)
42. Bardi M, Dolcetta I. 1997 *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman Equations*. Basel, Switzerland: Birkhäuser.
43. Lorz A, Lorenzi T, Hochberg ME, Clairambault J, Perthame B. 2013 Populational adaptive evolution, chemotherapeutic resistance and multiple anti-cancer therapies. *ESAIM: Math. Modell. Numer. Anal.* **47**, 377–399. (doi:10.1051/m2an/2012031)
44. Nowakowski A, Popa A. 2013 A dynamic programming approach for approximate optimal control for cancer therapy. *J. Optim. Theory Appl.* **156**, 365–379. (doi:10.1007/s10957-012-0137-z)
45. Kaznatcheev A, Peacock J, Basanta D, Marusyk A, Scott JG. 2019 Fibroblasts and alectinib switch the evolutionary games played by non-small cell lung cancer. *Nature Ecol. Evol.* **3**, 450–456. (doi:10.1038/s41559-018-0768-z)
46. Gatenby RA, Gawlinski ET, Gmitro AF, Kaylor B, Gillies RJ. 2006 Acid-mediated tumor invasion: a multidisciplinary study. *Cancer Res.* **66**, 5216–5223. (doi:10.1158/0008-5472.CAN-05-4193)
47. Silva AS, Yunes JA, Gillies RJ, Gatenby RA. 2009 The potential role of systemic buffers in reducing intratumoral extracellular ph and acid-mediated invasion. *Cancer Res.* **69**, 2677–2684. (doi:10.1158/0008-5472.CAN-08-2394)
48. Stanková K, Brown JS, Dalton WS, Gatenby RA. 2019 Optimizing cancer treatment using game theory. *JAMA Oncol.* **5**, 96–103. (doi:10.1001/jamaoncol.2018.3395)
49. Maltas J, Wood KB. 2019 Pervasive and diverse collateral sensitivity profiles inform optimal strategies to limit antibiotic resistance. *PLoS Biol.* **17**, e3000515. (doi:10.1371/journal.pbio.3000515)
50. Nichol D, Jeavons P, Fletcher AG, Bonomo RA, Maini PK, Paul JL, Gatenby RA, Anderson AR, Scott JG. 2015 Steering evolution with sequential therapy to prevent the emergence of bacterial antibiotic resistance. *PLoS Comput. Biol.* **11**, e1004493. (doi:10.1371/journal.pcbi.1004493)

51. Nichol D *et al.* 2019 Antibiotic collateral sensitivity is contingent on the repeatability of evolution. *Nat. Commun.* **10**, 334. (doi:10.1038/s41467-018-08098-6)
52. Kumar A, Vladimirsky A. 2010 An efficient method for multiobjective optimal control and optimal control subject to integral constraints. *J. Comput. Math.* **28**, 517–551. (doi:10.4208/jcm.1003-m2809)
53. Khan KH *et al.* 2018 Longitudinal liquid biopsy and mathematical modeling of clonal evolution forecast time to treatment failure in the Prospect-C Phase II Colorectal Cancer Clinical Trial. *Cancer Discov.* **8**, 1270–1285. (doi:10.1158/2159-8290.CD-17-0891)
54. Davis MHA, Varaiya P. 1973 Dynamic programming conditions for partially observable stochastic systems. *SIAM J. Control* **11**, 226–261. (doi:10.1137/0311020)
55. Kaznatcheev A. 2018 Effective games and spatial structure. *PNAS* **115**, E1709. (doi:10.1073/pnas.1719031115)