MDPI

*Article*

# Practical Understanding of Cancer Model Identifiability in Clinical Applications

Tin Phan [1,2,*] , Justin Bennett [2,3] and Taylor Patten [2,4]

1 Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM 87544, USA
2 School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85281, USA
3 Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218, USA
4 Arizona College of Osteopathic Medicine, Midwestern University, Glendale, AZ 85308, USA
* Correspondence: ttphan@lanl.gov

**Abstract:** Mathematical models are a core component in the foundation of cancer theory and have been developed as clinical tools in precision medicine. Modeling studies for clinical applications often assume an individual's characteristics can be represented as parameters in a model and are used to explain, predict, and optimize treatment outcomes. However, this approach relies on the identifiability of the underlying mathematical models. In this study, we build on the framework of an observing-system simulation experiment to study the identifiability of several models of cancer growth, focusing on the prognostic parameters of each model. Our results demonstrate that the frequency of data collection, the types of data, such as cancer proxy, and the accuracy of measurements all play crucial roles in determining the identifiability of the model. We also found that highly accurate data can allow for reasonably accurate estimates of some parameters, which may be the key to achieving model identifiability in practice. As more complex models required more data for identification, our results support the idea of using models with a clear mechanism that tracks disease progression in clinical settings. For such a model, the subset of model parameters associated with disease progression naturally minimizes the required data for model identifiability.

## 1. Introduction

Mathematical models serve an important role in the development of cancer theory and provide a framework to integrate and understand clinical data [1–9]. The attractiveness of mathematical models in clinical application comes from their ability to predict possible outcomes of a hypothetical treatment scenario based off of a set of mechanisms, which contrasts the black-box approaches in machine learning. Mathematical modelers assume that the relevant characteristics of an individual for treatment design can be represented by a set of parameters [10–13]. The influence of these parameters is then expressed in functional responses, which dictate the rate of each reaction within a preset model structure. The particular forms of the functional response and model structure are often borrowed from classic ecology and population studies and tested on the data of cohorts of patients, which make them a shared canvas for all patients [14]. On the other hand, the set of parameters that distinguishes the treatment outcome is unique to each individual [15–17]. If this unique set of parameters can be determined for a particular patient, then the model is presumed to be useful in predicting the most appropriate treatment for that patient. Thus, the concept of mathematical modeling coincides with the central idea behind precision medicine, where treatment should be formulated from the characteristics of each individual [18,19].

Mathematical models can be used to fit clinical observations and make real-time predictions of treatment outcomes. Given a sufficient amount of data for a patient and an appropriate model, one can use various statistical techniques to estimate the patient-specific set of parameters. Yet, "sufficient" is a quantity determined by model complexity. Comprehensive models, such as those found in system biology studies, are more biologically realistic, but the extra layers of complexity often hinder attempts to estimate the patient-specific set of parameters. On the flip side, simpler models with fewer parameters may not be capable of fully capturing the set of possible clinical outcomes qualitatively and quantitatively. Yet, simple models can still sometimes be too complex relative to the available data. Let us consider a thought experiment using a simple differential equation describing the birth and death processes of a population of cancer cells $x$:

$$x' = \beta x - \delta x = (\beta - \delta)x = mx, \tag{1}$$

where $\beta$ and $\delta$ are the per capita birth and death rates, respectively. We define $m = \beta - \delta$. This is perhaps the simplest growth equation that can be used as a building block for a more complex model. One example of the use of this simple model is by Claret et al. [20]. In clinical settings, estimates of cancer growth are available over time, either by direct methods, such as imaging or indirect estimates from tumor proxy [21–23]. The clinical data would then come in pairs of $(t_n, x_n)$, where $x_n$ is the cancer population measured at time $t_n$. We can assume that the model fits the data perfectly in this thought experiment. Yet, without prior knowledge of $\beta$ or $\delta$, there would be no way to determine a unique value for either parameters in this very simple model. We can only find an infinite number of pairs of $\beta$ and $\delta$ that give the same value for $m$. Note that this is related, but distinct from uncertainty or sensitivity analysis, where the uncertainty in each parameter is constrained based on the error in the parameter estimation.

The above thought experiment is a toy example meant to demonstrate the concept of model unidentifiability, which heuristically implies the inability to estimate the unique patient-specific set of parameters from the data. In other words, if the model is not identifiable (relative to the available data), then it is possible to find many sets of parameters that fit the data equally well statistically. To see why model identifiability represents a great obstacle in realizing the potential of mathematical models in precision medicine, we look at a specific example by Wu et al. [24]. This example comes from a modeling study using data from a clinical trial testing intermittent androgen deprivation therapy for treatment of metastatic prostate cancer [25]. The model used in the study is a mechanistic model developed from the classic Droop cell quota model and tested against several clinical datasets [11,15,26,27]. In fact, most of the model parameters can be determined directly from information in literature [13,28], yet the model is still unidentifiable with respect to the available clinical data. Figure 1 shows that five distinct sets of parameters demonstratively capture the data well in the fitting portion, yet only one provides accurate forecast. This example illustrates why potential issues of model unidentifiability should be addressed completely prior to clinical application of mathematical models.
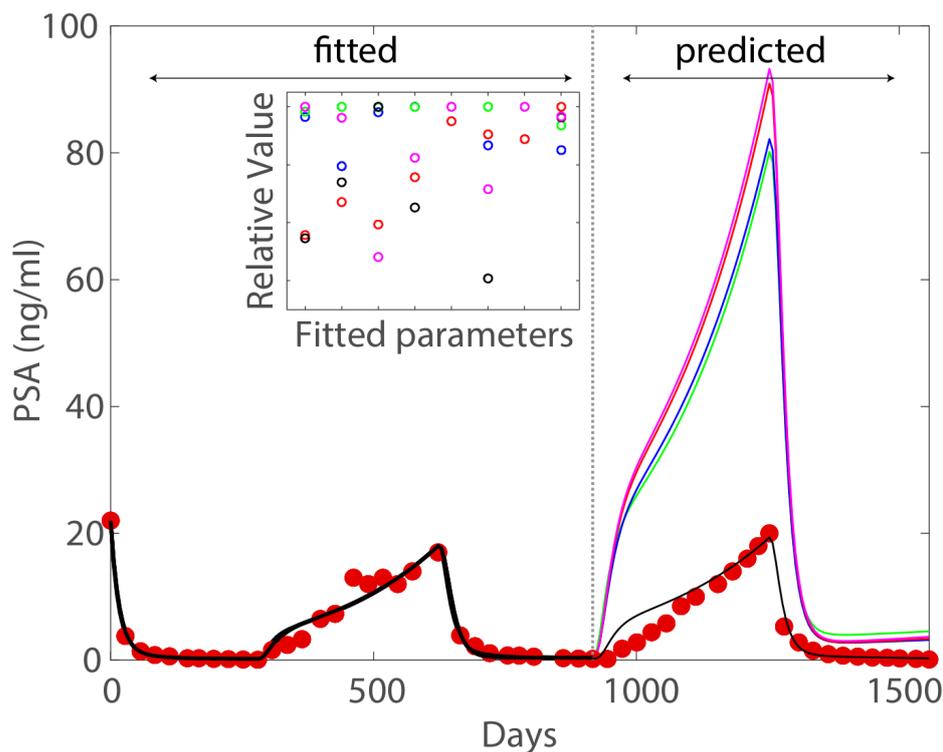
**Figure 1.** Figure adapted from Wu et al. [24] with permission distributed under a Creative Commons Attribution (CC BY) license. The color of the fitted parameters corresponds to the forecast trajectory of the same color. In the fitting portion, five different sets of parameters produce nearly indistinguishable good fits to the data. However, in the forecasting portion, only one set provides accurate forecasting.

## 2. Materials and Methods

### 2.1. Structural and Practical Identifiability

Depend on the types of model identifiability, there are various examples and techniques to address the issues of model identifiability [29–40]. Here, we offer our perspective on this issue. Consider a dynamical system of the form:

$$x'(t) = f(t, x(t), u(t), \theta), \tag{2}$$

$$y(t) = h(x(t), u(t), \theta), \tag{3}$$

where $x(t) \in \mathbb{R}^m$ represents the state variables, $y(t) \in \mathbb{R}^d$ represents the measurable output (e.g., the data), $u(t) \in \mathbb{R}^p$ represents the (control) input vectors, for example the administered drug, and $\theta \in \mathbb{R}^q$ represent the set of constant parameters. Note that while $\theta$ can contain time-varying parameters, for a biological system, this complexity can usually be avoided by using a functional representation of the time-varying parameters or explicitly modeling the underlying processes that drive the temporal changes. Thus, for simplicity, we take $\theta$ to contain only constant parameters. The general definition of model identifiability follows [29,31].

**Definition 1** (Model identifiability). *The dynamic system given by Equations (2) and (3) is identifiable if $\theta$ can be uniquely determined from the given input $u(t)$ and measurable output $y(t)$. If a system is not identifiable, then it is unidentifiable. Furthermore, if $y(t)$ does not contain error, then the identifiability of the system is referred to as structural identifiability. Otherwise, it is referred to as practical identifiability.*

The first example that we gave is an instance of structural unidentifiability and the second is a case of practical unidentifiability. We expand on what it means for a dynamical

model to be identifiable with respect to a measurable output. For this next example, we will look at a one dimensional logistic equation

$$x' = rx\left(1 - \frac{x}{K}\right). \tag{4}$$

Here, $r$ and $K$ are the intrinsic rate of growth and carrying capacity, respectively, for the population denoted by $x$. Let $(r_1, K_1)$ and $(r_2, K_2)$ be two sets of parameters and assume perfect measurement $y(t) = x(t)$. We will also assume that $x(0)$ is known. Then, if $(r_1, K_1)$ and $(r_2, K_2)$ results in the same models dynamics, then

$$x'(t, r_1, K_1) = x'(t, r_2, K_2) \iff r_1 x(t) - \frac{r_1}{K_1}x(t)^2 = r_2 x(t) - \frac{r_2}{K_2}x(t)^2. \tag{5}$$

For the condition above to hold for all $x(t)$, we must have $r_1 = r_2$ and $\frac{r_1}{K_1} = \frac{r_2}{K_2}$, which implies $(r_1, K_1) = (r_2, K_2)$ for almost all $x(0)$, except for possibly a set of measure zero. For instance, if $x(0) = 0$ or $K$, then the system is at its steady state, making the aforementioned comparison obsolete. *To put it simply, if a model is identifiable, then two sets of parameters that give the same model dynamics must be identical.* This means that an identifiable model does not have potential issue in Figure 1. Instead, the uncertainty in model forecast is solely dependent on the uncertainty in the data.

We remark that the above system does not contain a control term; however, for biological systems, if the identifiability of the system without the control can be studied, then adding the control term afterward usually does not change its structural identifiability, see the example given by Eisenberg and Jain [31]. What we showed here is an example of a direct test of model identifiability, originally used by Denis-Vidal and Joly-Blanchard [41]. While the direct test method is often not used in practice, it serves as an intuitive description for model identifiability from a dynamical system perspective.

To test for model structural identifiability, software built on differential algebra theory, such as DAISY, is the gold standard [42]. However, structural identifiability does not guarantee practical identifiability, which is necessary for clinical application. The practical identifiability of a model should be studied with data, in particular using the Fisher information matrix or profile likelihood [24,31]. In these scenarios, the available data dictates the formulation of the model. However, because mathematical models are often developed independent of the data collection, modelers often must sacrifice certain realistic aspects of the model to keep it identifiable relative to the available data. Reversely, if one first builds a set of candidate models and finds out the required data for accurate model identification, then it may be possible to obtain these data during the collection process. We consider the latter case an ideal scenario, where mathematical modelers and clinicians can collaborate effectively.

### 2.2. Observing-System Simulation Experiment via Monte Carlo Method

In the ideal scenario mentioned above, Monte Carlo simulation experiment is our tool of choice to obtain the information on the data required for a model to be identifiable. First, we introduce the statistical model [43]:

$$\boldsymbol{y}(t_i) = \boldsymbol{h}(\boldsymbol{x}(t_i), \boldsymbol{u}(t_i), \boldsymbol{\theta}) + \boldsymbol{\varepsilon}(t_i). \tag{6}$$

where $\boldsymbol{h}(\boldsymbol{x}(t_i), \boldsymbol{u}(t_i), \boldsymbol{\theta})$ is the measurement, $\boldsymbol{\theta}$ is the vector of parameters estimated from $\{\boldsymbol{y}(t_i)\}_{i=1}^N$ observations at time $\{t_1, \ldots, t_N\}$. Assuming no model error, then the general form of the measurement error is

$$\varepsilon_i = h_i(\boldsymbol{x}(t_i), \boldsymbol{u}(t_i), \boldsymbol{\theta})^f \epsilon_i, \tag{7}$$

with $f \geq 0$, $\epsilon_i$ taken to be independent and identically distributed random variables with mean 0 and variance $\sigma_0^2$. For biological application, it is reasonable to expect the

measurement error to be proportional to the measurement itself, so we fix $f = 1$, giving us a relative error. The steps of the Monte Carlo simulation method follow [29,44].

1. Determine the appropriate set of true parameters $\theta_0$ for the simulation.
2. Numerically solve the ODE model to obtain the measurements at desired time points.
3. Generate M sets of simulated data from the statistical model (6) and (7) with a Gaussian error structure and a chosen standard deviation $\sigma_0\%$ around mean 0.
4. Fit the model to each of the M simulated data sets to obtain the parameter estimates $\tilde{\theta}_i, i = 1, \ldots, M$. Here, we take M to be 200 sets.
5. Calculate the average relative estimation error (ARE) for each element of $\theta$ as

$$ARE(\theta_0^{(k)}) = 100\% \frac{1}{M} \sum_{j=1}^{M} \frac{\left| \theta_0^{(k)} - \tilde{\theta}_i^{(k)} \right|}{|\theta_0^{(k)}|}, \tag{8}$$

where $\theta_0^{(k)}$ and $\tilde{\theta}_i^{(k)}$ are the k-th element of $\theta_0$ and $\tilde{\theta}_i$, respectively.
6. Repeats steps 2 through 5 with increasing $\sigma_0 = 0, 5, 15, 25\%$.

A model is practically identifiable if the *ARE* is less than the variance ($\sigma > 0\%$), meaning we want the error in the parameter estimation to be less than the error in the data. When the variance $\sigma$ is 0% and *ARE* is sufficiently close to 0%, then the model is considered to be structurally identifiable. We borrow the idea from the observing-system simulation experiment where we will use different hypothetical sets of data to test whether they can help us identify key parameters [45,46]. By continually restricting the amount of information we have from the data, we can approximate the threshold of information required for model identification.

The MC simulation approach is not error-free. One such limitation comes from choosing the initial guesses. For example, if we start our initial guess close enough to the true set of parameters, then the effect of the error $\sigma$ may be limited. However, if we have our guesses far away from the true set of parameters, then the numerical optimization may become trapped in some local minimum away from the true estimate or the parameters may not be sensitive enough to be estimable. One can start with random samples of initial guesses to have a better chance at reaching the true estimates. However, this sampling approach does not inherently deal with the issue of the insensitive parameters. Here, we pick the initial guesses randomly within 50% of the true value. In order to rule out any parameters that are not sensitive with respect to the tolerance of the numerical optimization schemes, we carry out the MC approach for each individual parameter with error-free data ($\sigma_0 = 0\%$). Any parameters that cannot be refitted within reasonable ARE will be eliminated (or become fixed) from the pool of free parameters. The remaining parameters are deem to be sensitive enough for the numerical scheme. As we amp up the tolerance of the numerical scheme, eventually we should be able to fit all parameters when no error is present in the data.

### 2.3. Two Mathematical Models for Prostate Cancer

Many mathematical models for prostate cancer have been developed in the past two decades [1,18,47] with many recent studies focused on immuno- and chemo-treatments of prostate cancer [48–56]. Here, we divert from this trend and instead use two simpler mechanistic models to demonstrate the concept of model identifiability in practice. Both models contain a clear prognostic parameter that keeps track of cancer progression, which greatly simplifies their structure. Using these models, we aim to show that even if the model itself may not be identifiable, having a model-based prognostic parameter allows modelers to focus the resource to identify these key parameters. This would be helpful in practical settings due to limitation in data acquisition.

*A cancer stem cell model.* Cancer stem cells propel cancer's therapeutic resistance and are thought to be a primary factor in the initiation and progression of prostate cancer [57–60]. Utilizing the mathematical model below, in conjunction with the stem

cell hypothesis, could provide a better understanding of prostate cancer's acquisition of castration resistant cells and their heterogeneity within a mass. Prostate cancer stem cells are thought to express little to no androgen receptors, giving them the ability to multiply their population without a hormone requirement [61]. Resistance is achieved with cancer stem cells' ability to thrive in the absence of androgen, which provides a means for cancer to continue to evolve during and after intervention with intermittent androgen deprivation therapy [12,17,62].

Prostate cancer stem cells continue to rapidly divide after treatment, either asymmetrically to form differentiated cells or symmetrically to form additional stem cells. The production of differentiated cells results in negative feedback of the production of stem cells. However, unlike stem cells, differentiated cells are affected negatively by androgen deprivation therapy. The ability to withstand androgen deprivation is just one of the many contributing factors that give rise to the renewal of stem cells. For instance, mitochondrial fission factor expression plays a role in the evolution and multiplicity of prostate cancer stem cells [63].

Here, we use a novel model built upon this concept for prostate cancer from the studies by Brady-Nicholls et al. [12,17,62]. The model consider three compartments, the cancer stem cells ($S$), the differentiated cancer cell ($D$), and the PSA byproduct ($P$). While it is simpler in structure, the model has shown promises in its applicability.

$$\frac{dS}{dt} = \underbrace{\frac{S}{S+D}p_s\lambda S}_{growth}, \tag{9}$$

$$\frac{dD}{dt} = \underbrace{\left(1 - \frac{S}{S+D}p_s\right)\lambda S}_{growth} - \underbrace{\alpha T_x D}_{death}, \tag{10}$$

$$\frac{dP}{dt} = \underbrace{\rho D}_{production\ by\ cancer\ cell} - \underbrace{\psi P}_{clearance}. \tag{11}$$

The cancer stem cell population $S$ is assumed to divide at a rate $\lambda$ to produce either one stem cell and one cancer cell with probability $p_s$, or two cancer cells. This division has a negative feedback from the differentiated cancer cells, which takes the form $\frac{S}{S+D}$. The cancer cell is killed by the drug at a constant rate $\alpha$, where $T_x$ denotes the application of the drug. PSA is produced by cancer cells at a rate $\rho$, which is cleared from the blood stream at a rate $\psi$.

Since the drug applications for these model, $u$ and $T_x$, are known input. For simplicity, we can treat them as constant. Since they are known, their variation in time should not affect the identification of the other factors. Additionally, in practice, the drug application would be fixed for a certain period of time depending on the specific treatment. We take the following parameter values as the true values for our study: $T_x = 0.5$ (dimensionless), $p_s = 0.03$ (dimensionless), $\lambda = ln(2)$ day$^{-1}$, $\alpha = 0.05$ day$^{-1}$, $\rho = 1.87 \times 10^{-4}$ µg L$^{-1}$day$^{-1}$, and $\psi = 0.085$ day$^{-1}$ [62].

*A cell quota cancer model.* Prostate cancer cells require androgen for growth, which is why the effect of androgen is regularly incorporated into prostate cancer model [1,64,65]. However, the quantitative connection between androgen and prostate cancer growth is not well characterized, leading to various functional forms used for this purpose.

Here, we use a cancer model that integrates the effect of androgen based on a stoichiometric modeling framework [15,66,67]. The model was developed in a series of studies that highlight the importance of androgen dynamics in prostate cancer growth [11,13,16,26–28,64,68,69]. In this model, cancer independence to androgen is modeled as a variable explicitly and can be used as an indicator of cancer growth. Meade et al. later expanded on this idea to build a more biologically realistic model of cancer growth for

predicting treatment failure [16]. Despite its simplicity, the model is founded on established biological principle and can capture and predict the dynamics of cancer progression.

$$\frac{dx}{dt} = \underbrace{\mu_m \left(1 - \frac{q}{Q}\right) x}_{\text{growth}} - \underbrace{\left(\nu(t)\frac{R}{Q+R} + \delta x\right) x}_{\text{death}}, \tag{12}$$

$$\frac{dv}{dt} = - \underbrace{dv}_{\text{rate of gaining androgen independence}}, \tag{13}$$

$$\frac{dQ}{dt} = \underbrace{(\gamma_1 u + \gamma_2)(Q_m - Q)}_{\text{androgen influx to cells}} - \underbrace{\mu_m(Q - q)}_{\text{uptake}}, \tag{14}$$

$$\frac{dP}{dt} = \underbrace{bQ}_{\text{baseline production}} + \underbrace{\sigma x Q}_{\text{production by cancer cells}} - \underbrace{\epsilon P}_{\text{PSA clearance}}. \tag{15}$$

The cancer population, denoted by $x$, grows based on the Droop cell-quota model. The death rates are contributed by an androgen dependent term, $\nu(t)\frac{R}{Q+R}x$, and a density dependent term, $\delta x^2$. Here, $\nu(t)$ is the maximal androgen dependent death rate for the cancer. The authors assume that the cancer cells lose their dependence on androgen at a rate $-dv$, which can be interpreted as the "rate of gaining androgen independence". With this interpretation, under androgen deprivation therapy, the treatment would gradually become ineffective. $Q$ and $P$ are the intracellular androgen level and serum PSA, respectively. The dynamics of $Q$ is governed by an influx of serum androgen and the uptake of cancer cells. $\gamma_1$ and $\gamma_2$ represent the rates at which androgen is being produced by the testes and the adrenal gland, respectively, with the drug application denoted by $u$. $P$ is assumed to be produced as a baseline by normal cells, but mainly by cancer cells, and is cleared from the blood stream at a constant rate. We take the following parameter values as the true values for our study: $u = 0.5$ (dimensionless), $\mu_m = 0.009$ day$^{-1}$, $q = 0.4$ nmol day$^{-1}$, $R = 3$ nmol L$^{-1}$, $\delta = 45$ L$^{-1}$day$^{-1}$, $d = 0.0001$ day$^{-1}$, $\gamma_1 = 0.08$ day$^{-1}$, $\gamma_2 = 0.004$ day$^{-1}$, $Q_m = 30$ nmol L$^{-1}$, $b = 0.0001$ µg nmol$^{-1}$day$^{-1}$, $\sigma = 0.001$ µg nmol$^{-1}$L$^{-1}$day$^{-1}$, and $\epsilon = 0.1$ day$^{-1}$ [13,15].

### 2.4. Parameter Optimization

When the data are of a single type, we use the standard root mean squared error (RMSE), for example, the cancer stem cell model with PSA data. When the data composed of multiple types of data, for example with PSA and androgen, we weigh the error contribution from each source equally. Any variation from this fitting procedure will be mentioned on a case-by-case basis. Finally, we use the built-in function *lsqnonlin* (MATLAB) for our optimization.

## 3. Results

### 3.1. The Identifiability of Two Prostate Cancer Models

First, we study the identifiability of the model given the measurements that are usually available directly for parameter estimation. In the case of the cancer stem cell model, this measurement is taken to be PSA. In the case of the cell quota model, the measurements are PSA and androgen. We also note that the spacing between the synthetic data points is kept constant in this section.

*The cancer stem cell model.* An example of the synthetic data and fitting for the cancer stem cell model is presented in Figure 2. Table 1 shows the results from the sensitivity test for each individual parameter for the cancer stem cell model. Out of the seven parameters and initials, $p_s$ and $D(0)$ are not sensitive enough to be identifiable for larger measurement error. On the other hand, $\lambda, \alpha, \rho, \psi$, and $S(0)$ appear to be sufficiently sensitive. Thus, we fix $p_s$ and $D(0)$.
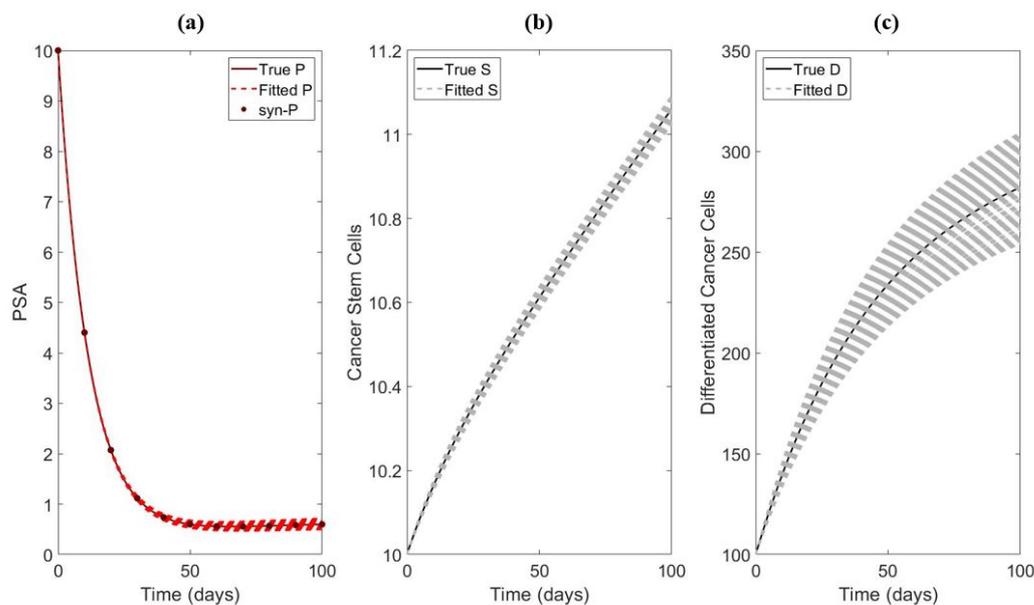
**Figure 2.** An example of data fitting with only PSA synthetic data. In this example, the parameters being fitted are $\lambda$ and $\rho$ with $\sigma_0 = 0$. (**a**) Fitting of PSA. (**b**) Simulation of the S using the best fitted parameters. (**c**) Simulation of D using the best fitted parameters.

**Table 1.** ARE% calculated for Brady et al. model with respect to $\sigma$. The test is carried out individually for each parameter and initial. 200 samples are used. The results indicate that five ($\lambda, \alpha, \rho, \psi, S(0)$) out of seven components are sufficiently sensitive to to the numerical optimization.

| $\sigma_0$ | 0% | 5% | 15% | 25% |
|:---:|:---:|:---:|:---:|:---:|
| $p_s$ | 0.028 | 37.72 | 92.89 | 130.77 |
| $\lambda$ | 0.01 | 2.17 | 6.51 | 10.85 |
| $\alpha$ | 0.01 | 2.94 | 8.89 | 15.20 |
| $\rho$ | 0.01 | 2.50 | 7.49 | 12.48 |
| $\psi$ | 0.01 | 2.12 | 6.61 | 11.89 |
| $S(0)$ | 0.01 | 2.18 | 6.54 | 10.90 |
| $D(0)$ | 0.20 | 35.95 | 87.01 | 121.86 |

Next, we carry out the MC scheme for all of the remaining parameters at once, namely $\lambda, \alpha, \rho, \psi, S(0)$. The results in Table 2 show that only $\psi$ is practically identifiable. To see why the other parameters are not identifiable with only PSA data alone, we fix $\psi$ and test the identifiability of all 2-combinations of the remaining parameters (e.g., fit two parameters at a time while fixing the rest). We find that none of the 2-combinations are practically identifiable. Since each parameter being tested is sensitive enough to be identifiable by themselves, this indicates the existence of an unknown relationship among the remaining four parameters (e.g., $\lambda, \alpha, \rho, S(0)$). To demonstrate this point, we plot the estimated values of these parameters in 2-combinations and show that an approximate relationship between these parameters can be obtained by a simple regression Figure 3. Without error ($\sigma_0 = 0\%$), the relationships between $\lambda, \alpha$, and $\rho$ are evident by performing the 2-combination test.

In Brady et al. [62], the authors find that the prostate cancer stem cell renewal rate $p_s$ is a good indicator of resistance timing. However, our analysis shows that in order to utilize $p_s$ to make clinical predictions, modelers must have a solid grasp on the values of all other parameters and a good understanding of the appropriate value for $p_s$ in the cancer stem cell model. This agrees with the approach taken in Brady et al. to obtain model identifiability for $p_s$ [17,62].

**Table 2.** ARE% calculated for Brady et al. model with respect to $\sigma$ using the MCMC method. The test is carried out for all five parameters and initial together. In total, 200 samples are used. The results indicate that when fitting together, only $\psi$ is practically identifiable.

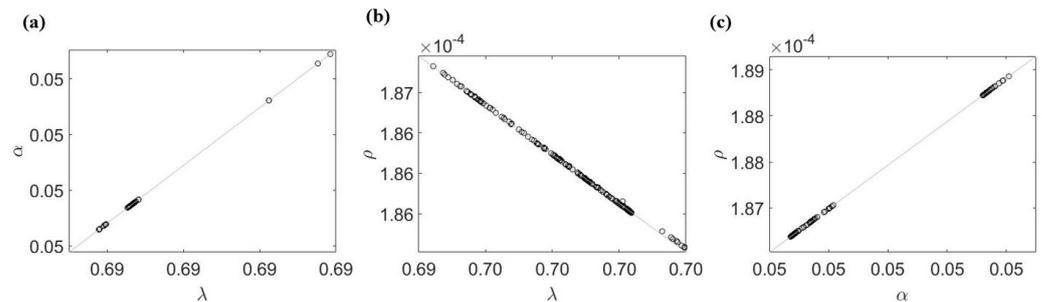| $\sigma_0$ | 0% | 5% | 15% | 25% |
|---|---|---|---|---|
| $\lambda$ | 54.79 | 246.64 | 243.88 | 220.03 |
| $\alpha$ | 52.51 | 131.37 | 123.11 | 131.75 |
| $\rho$ | 50.59 | 130.77 | 200.50 | 206.93 |
| $\psi$ | 1.12 | 6.36 | 13.98 | 21.11 |
| $S(0)$ | 13.66 | 37.48 | 70.88 | 101.84 |



**Figure 3.** Parameter relation obtained from the 2-combination parameter test for the case of $\sigma_0 = 0$. (**a**) $\alpha$ and $\lambda$ are positively correlated. (**b**) $\rho$ and $\lambda$ are negatively correlated. (**c**) $\alpha$ and $\rho$ are positively correlated. If the parameters are linearly correlated, (**a**,**b**) would imply that $\alpha$ and $\rho$ are negatively correlated; however, this is not the case in (**c**). This suggests all three parameters are involved in a non-linear relationship.

*The cell quota cancer model.* An example of the synthetic data and fitting for the cancer stem cell model is presented in Figure 4 Similarly, we carry out these tests with respect to the 13 parameters and initials for the cell quota cancer model. Table 3 shows that out of these, only $\mu$, $Q_m$, and $\epsilon$ are sufficiently sensitive when each parameter is fitted individually with PSA and androgen data.

**Table 3.** ARE% calculated for BK 1 model with respect to $\sigma$ using the MCMC method. The test is carried out individually for each parameter and initials. In total, 200 samples are used. The results indicate that 3 ($\mu$, $Q_m$, $\epsilon$) out of 13 components are sufficiently sensitive to continue with the procedure. Note that $\delta$ and $\sigma$ appear to be insensitive, so we exclude them.

| $\sigma_0$ | 0% | 5% | 15% | 25% |
|---|---|---|---|---|
| $\mu$ | 0.00 | 4.48 | 13.49 | 22.68 |
| $q$ | $5.78 \times 10^{-5}$ | 86.18 | 94.87 | 97.21 |
| $R$ | 0.51 | 68.42 | 68.05 | 67.87 |
| $\delta$ | 5.02 | 5.02 | 5.02 | 5.02 |
| $d$ | 11.70 | 66.07 | 65.74 | 65.91 |
| $\gamma_1$ | $2.39 \times 10^{-4}$ | 10.89 | 32.76 | 51.04 |
| $\gamma_2$ | 0.00 | 21.73 | 57.28 | 73.29 |
| $Q_m$ | $2.28 \times 10^{-5}$ | 2.27 | 6.81 | 11.35 |
| $b$ | 5.51 | 32.40 | 46.65 | 50.32 |
| $\sigma$ | 1.05 | 14.94 | 14.96 | 15.97 |
| $\epsilon$ | $1.82 \times 10^{-4}$ | 2.34 | 7.33 | 13.24 |
| $x(0)$ | 0.57 | 59.84 | 60.14 | 59.75 |
| $v(0)$ | 0.46 | 78.75 | 78.57 | 78.29 |

When fitting the three sensitive parameters together, we find that only $\epsilon$ is practically identifiable, see Table 4. To see why the remaining two parameters ($\mu$ and $Q_m$) are not identifiable, we fix $\epsilon$ and study the correlation between these two parameters. Here, we find a similar linear relationship between estimates for $\mu$ and $Q_m$ with or without error in the data, see Figure 5. This hidden correlation interferes with the estimability of these two parameters.

In Baez and Kuang, the parameter $d$ (or variable $\nu(t)$) is created to keep track of the development of cancer resistance. However, our analysis indicates a similar issue to the cancer stem cell model where the relevant parameter is not identifiable with the available data. If we want to have an accurate estimate of $d$, we must have a strong grasp on the values of all other variables in the model and a good guess for an appropriate value for $d$.
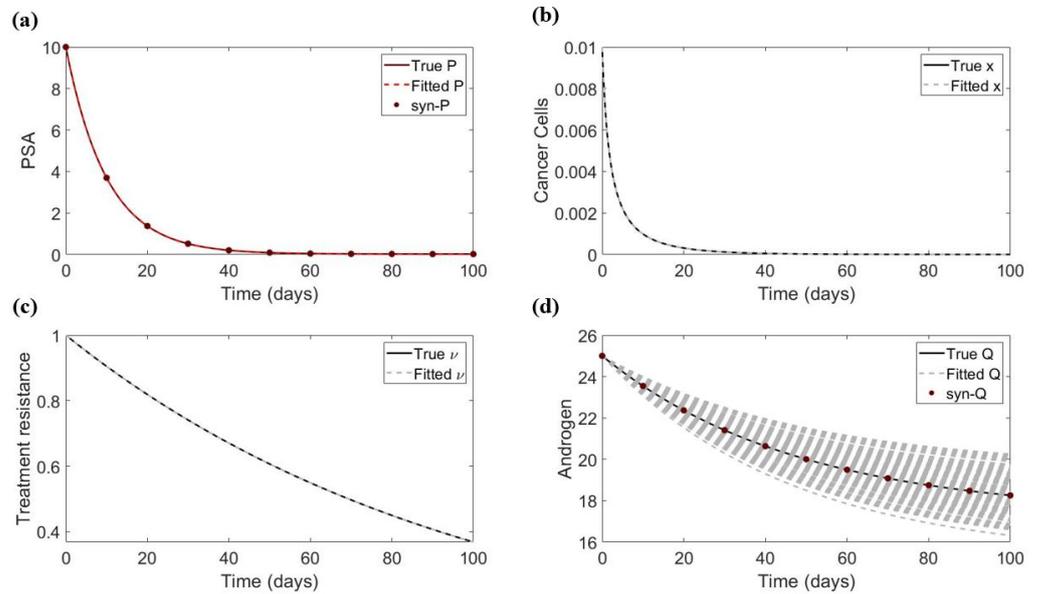


**Figure 4.** An example of data fitting with PSA and androgen synthetic data. In this example, the parameters being fitted are $\mu$ and $Q_m$ with $\sigma_0 = 0$. (**a**) Fitting of PSA. (**b**) Simulation of the cancer population using the best fitted parameters. (**c**) Simulation of the parameter $\nu$ (associated with treatment resistance) using the best fitted parameters. (**d**) Fitting of androgen.
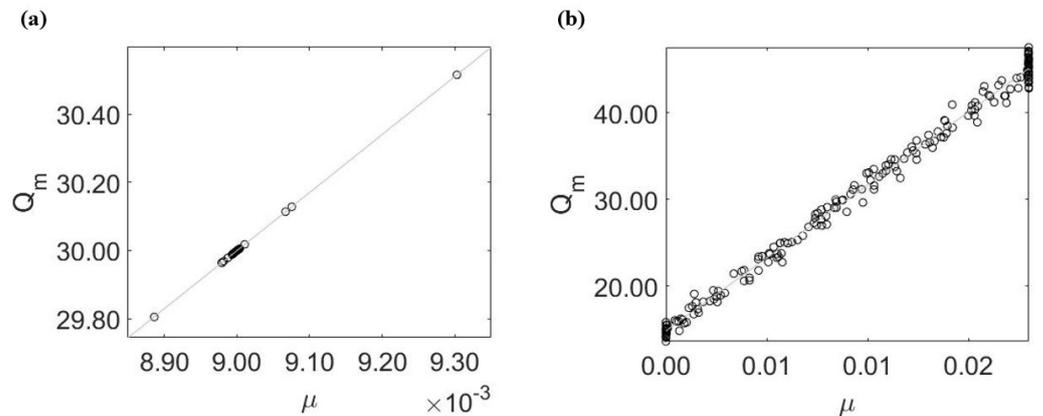


**Figure 5.** Parameter relation obtained from the 2-combination parameter test. (**a**) When $\sigma_0 = 0$, $\mu$ and $Q_m$ are slightly correlated with the best fit values concentrated around $\mu = 9.00 \times 10^{-3}$ and $Q_m = 30$. (**b**) When $\sigma_0 = 5\%$, the relationship between $\mu$ and $Q_m$ is much clearer, leading to a larger ARE%.

**Table 4.** ARE% calculated for BK 1 pop model with respect to $\sigma$ using the MCMC method. The test is carried out for all five parameters and initials together. In total, 200 samples are used. The results indicate that when fitting all four parameters together, only $\epsilon$ is practically identifiable. It is worth pointing out that when the measurement error is 0, the ARE% of all four parameters are very close to 0, indicating structural identifiability.

| $\sigma_0$ | 0% | 5% | 15% | 25% |
|---|---|---|---|---|
| $\mu$ | 0.02 | 64.85 | 87.60 | 92.13 |
| $Q_m$ | 0.01 | 32.77 | 43.91 | 45.80 |
| $\epsilon$ | 0.00 | 2.34 | 7.32 | 13.23 |

*3.2. Observing-System Simulation Experiment—Identifiability of Treatment Resistance Parameter*

Now, we turn our focus to answering the question: what amount of data are necessary to determine the key model-based prognostic parameter? To address this question, we synthesize candidate sets of data that vary in the type and frequency of data collected. Then, we attempt to study the identifiability of these key parameters using each synthetic dataset.

*The identifiability of $p_s$ in the cancer stem cell model.* Recall that $p_s$ is not identifiable even when being estimated by itself with only PSA data (Table 1). Thus, we will carry out simulation to determine the amount of data required to sufficiently characterize the $p_s$, which is used to predict treatment success and failure in Brady et al. [17,62]. For the experiment, we assume all parameters (except for $p_s$) can be obtained from other means, which means they are fixed to the values used to make the synthetic dataset for these simulation experiments.

Table 5 summarizes the main results of the experiments to determine the identifiability of $p_s$. The frequency of data points appears to be the most influential factor to identify $p_s$, which is followed by the inclusion of the measurement of cancer stem cells. On the other hand, measurement of the cancer population, optimization tolerance, and (linear) weight contribution from different sources of error have less of an impact on the identifiability of $p_s$. Interestingly, if a measurement of PSA can be taken roughly every 5 h, then we could accurately determine $p_s$ as well (given that it is the only parameter we need to estimate).

While the results in Table 5 suggest the frequency of measurements plays a key role, when fitting all parameters together with pseudo-continuous measurement of PSA and cancer stem cells, the model remains to be unidentifiable (see Table 6). However, if the measurements are very precise ($\sigma_0 \approx 0\%$), estimated values of each parameter are within acceptable ranges that may still be useful in making predictions (less than 10% difference from the true value).

*The identifiability of $d$ in the cell quota cancer model.* Similarly, recall that $d$ is not identifiable in the cell quota cancer model even when being estimated by itself with both androgen and PSA data (Table 3). Hence, we carry out simulation to determine the amount of data required to identify $p_s$. As before, we fix all other parameters for these simulation experiments.

Table 7 summarizes the main results of the experiments. We reach a similar conclusion that the frequency of data points appears to be the most influential factor in identifying $d$, which is followed by the inclusion of the measurement of cancer cells. However, with larger error margins, the measurement of cancer cells loses its effectiveness, which is problematic due to the fact that accurate estimations of cancer populations are difficult in practice. Meanwhile, all other factors have a negligible effect on the estimation of $d$. Finally, if a measurement of cancer cells, androgen, and PSA can be taken every 24 h, then we could accurately determine $d$. Interestingly, if a measurement of the cancer population is not available, but we can obtain pseudo-continuous data for $Q$ and $P$, then it is possible to determine the value of $d$ within a reasonable range.

**Table 5.** The identification of $p_s$ in the cancer stem cell model. The test is carried out for $p_s$ (all other parameters and initials are fixed to their true values). Baseline frequency (data) indicates a measurement is taken every 10 days. Pseudo-continuous data indicates a measurement is taken roughly every 2.4 h. Increased optimization tolerance refers to one fold increase in the function tolerance and optimality tolerance of the optimization function. Weight ($\omega$) comes from the minimization objective. $\omega > 0.5$ means higher weight is given to the error in $P$ and $\omega < 0.5$ means higher weight is given to the error in $S$, which is given by $\omega \times RMSE^P + (1 - \omega) \times RMSE^S$. Asterisk ($*$) indicates practical identifiability.

| $\sigma_0$ | 0% | 5% | 15% | 25% |
|---|---|---|---|---|
| Fitting using P (baseline frequency) | 0.03 | 37.72 | 92.89 | 130.77 |
| Fitting using S and P (baseline frequency) | 0.00 | 19.45 | 55.43 | 81.93 |
| Fitting using D and P (baseline frequency) | 0.00 | 30.33 | 79.48 | 113.78 |
| Fitting using S, D, and P (baseline frequency) | 0.00 | 29.47 | 77.71 | 111.34 |
| Fitting using S and P (baseline frequency $\times$ 2) | 0.00 | 13.38 | 39.08 | 60.22 |
| Fitting using S and P (baseline frequency $\times$ 10) | 0.00 | 6.15 | 18.45 | 30.60 |
| Fitting using S and P (pseudo-continuous data) ($*$) | 0.00 | 1.86 | 5.57 | 9.29 |
| Fitting using P (pseudo-continuous data) ($*$) | 0.00 | 3.69 | 11.05 | 18.42 |
| Fitting using P (baseline frequency $\times$ 50) | 0.00 | 5.53 | 16.58 | 27.62 |
| Fitting using S and P (increased optimization tolerance) | 0.00 | 19.45 | 55.43 | 81.93 |
| Fitting using S and P (weight = 0.4) | 0.00 | 19.43 | 55.38 | 81.87 |
| Fitting using S and P (weight = 0.25) | 0.00 | 19.41 | 55.34 | 81.82 |

**Table 6.** ARE% calculated for the cancer stem cell model with respect to $\sigma$. The experiment is carried out using pseudo-continuous data of $S$ and $P$.

| $\sigma_0$ | 0% | 5% | 15% | 25% |
|---|---|---|---|---|
| $p_s$ | 1.97 | 34.68 | 97.35 | 107.28 |
| $\lambda$ | 6.46 | 90.26 | 154.45 | 167.19 |
| $\alpha$ | 2.20 | 30.27 | 53.20 | 72.75 |
| $\rho$ | 5.07 | 37.08 | 60.01 | 70.64 |
| $\psi$ | 0.06 | 0.86 | 1.95 | 3.77 |
| $S(0)$ | 0.02 | 0.43 | 0.93 | 1.44 |
| $D(0)$ | 4.83 | 11.26 | 17.68 | 34.10 |

**Table 7.** The identification of $d$ in cancer cell quota model. The test is carried out for $p_s$ (all other parameters and initials are fixed to their true values). Baseline frequency (data) indicates a measurement is taken every 10 days. Pseudo-continuous data indicates a measurement is taken roughly every 2.4 h. Increased optimization tolerance refers to a one-fold increase in the function tolerance and optimality tolerance of the optimization function fmincon (MATLAB). Weight$_1 = \omega_1$ and weight$_2 = \omega_2$ come from the minimization objective, which is $\omega_1 \times RMSE^P + \omega_2 \times RMSE^Q + (1 - \omega_1 - \omega_2) \times RMSE^X$. Asterisk ($*$) indicates practical identifiability. ($**$) indicates $d$ is identifiable at very high error.

| $\sigma_0$ | 0% | 5% | 15% | 25% |
|---|---|---|---|---|
| Fitting using Q and P (baseline frequency) | 11.70 | 66.07 | 65.74 | 65.91 |
| Fitting using X, Q, and P (baseline frequency) | 0.09 | 24.12 | 50.96 | 60.37 |
| Fitting using X, Q, and P (baseline frequency $\times$ 10) | 0.10 | 6.34 | 19.73 | 27.75 |
| Fitting using X, Q, and P (baseline frequency $\times$ 50) ($*$) | 0.10 | 3.90 | 8.23 | 12.29 |
| Fitting using X, Q, and P (pseudo continuous data) ($*$) | 0.08 | 2.78 | 6.78 | 10.79 |
| Fitting using Q and P (pseudo continuous data) ($**$) | 3.67 | 10.44 | 10.69 | 10.75 |
| Fitting using X, Q, and P (increase optimization tolerance) | 0.09 | 29.04 | 62.88 | 74.00 |
| Fitting using X, Q, and P (weight$_1$ = 0.4, weight$_2$ = 0.33) | 0.11 | 25.33 | 46.66 | 52.43 |
| Fitting using X, Q, and P (weight$_1$ = 0.25, weight$_2$ = 0.25) | 0.09 | 19.90 | 43.54 | 52.83 |
| Fitting using X, Q, and P (weight$_1$ = 0.4, weight$_2$ = 0.2) | 0.10 | 23.33 | 46.81 | 54.39 |
| Fitting using X, Q, and P (weight$_1$ = 0.25, weight$_2$ = 0.5) | 0.10 | 21.26 | 41.25 | 45.78 |

As before, if none of the other parameters are known, then the identifiability of the model is not possible even with pseudo-continuous data of the cancer population, androgen, and PSA (see Table 8). Yet, if those measurements can be taken very precisely ($\sigma_0 \approx 0\%$), then the parameters can still be estimated within reasonable accuracy for application.

**Table 8.** ARE% calculated for the cell quota cancer model with respect to $\sigma$. The experiment is carried out using pseudo-continuous data of $x$, $Q$, and $P$.

| $\sigma_0$ | **0%** | **5%** | **15%** | **25%** |
|---|---|---|---|---|
| $\mu$ | 6.55 | 5.41 | 11.41 | 19.38 |
| $q$ | 0.75 | 2.53 | 1.53 | 1.69 |
| $R$ | 4.18 | 3.90 | 4.53 | 5.46 |
| $\delta$ | 5.17 | 4.48 | 4.97 | 4.90 |
| $d$ | 0.04 | 12.78 | 14.30 | 18.69 |
| $\gamma_1$ | 5.83 | 11.79 | 14.98 | 23.74 |
| $\gamma_2$ | 3.36 | 15.88 | 11.63 | 14.00 |
| $Q_m$ | 4.81 | 2.66 | 3.71 | 5.56 |
| $b$ | 0.44 | 9.57 | 27.39 | 42.59 |
| $\sigma$ | 0.03 | 26.09 | 34.05 | 39.14 |
| $\epsilon$ | 0.01 | 0.34 | 1.02 | 1.68 |
| $x(0)$ | 0.37 | 1.65 | 4.30 | 6.84 |
| $v(0)$ | 1.71 | 3.97 | 4.37 | 5.25 |

## 4. Discussion

Mathematical models not only contribute to the foundation of cancer theory but can also be integrated to provide a better prognostic tool for clinicians in clinical settings. For example, one may apply mathematical models to better understand cancer progression dynamics and make predictions of treatment outcomes based on a patient's characteristics [18]. Yet, the issue of practical identifiability remains a major obstacle to realizing the clinical potential of mathematical models. In this study, we explore the issue of model identifiability from a clinical perspective. First, we study the general identifiability property of the model. Then, we narrow down the parameter that can be used to predict treatment outcomes and look for the appropriate set of data for its identification using Monte Carlo simulation. Our results provide insights into the type of data acquisition that can enable future incorporation of mathematical models into clinical applications.

*The frequency of data collection plays a major role in model identifiability.* It is well known that increasing the number of measurements increases the chances of obtaining true estimates of model parameters assuming that the measurements are perfect and the model is structurally identifiable [70]. Here, we demonstrate in both examples that increasing the frequency of measurements, even in the presence of Gaussian noise, can increase model identifiability. However, simply increasing the number of data points will not overcome the issue of structural identification. The dataset should cover multiple temporal regions of cancer growth, so that the model can be tested more comprehensively to prove its usefulness. This finding suggests the development of devices or procedures to obtain measurements, such as PSA and androgen, on a regular basis can help to accurately identify the values of prognostic parameters.

*Cancer population data can help reduce the uncertainty in model identifiability.* We demonstrate that the inclusion of cancer population measurements (or stem cells) can increase the model identifiability. In practice, this can be completed by using imaging data or indirectly measuring circulating cancer cells [71–75]. On the other hand, for models that incorporate mechanisms for cancer growth using androgen, androgen data seem to be necessary for model identifiability. Unfortunately, these measurements are not widely adopted, making it difficult to integrate these ideas effectively. Furthermore, we carried out the same computational experiments on several cancer models with multiple subpopulations (not shown). The results suggest a measurement that helps to distinguish different cancer

subpopulations (e.g., the frequency of each cancer subpopulation) may be necessary to obtain model identification.

*Highly accurate data may be the key to addressing model identifiability in practice.* Perhaps the most intriguing finding is that with very accurate data, the prognostic parameters and some other model parameters are reasonably identifiable. Unlike the frequency of measurements, the accuracy of measurements does not require additional compliance from the patients. With continual advances in the techniques and equipment to measure the relevant biomarkers, high-accuracy data may be the key to obtaining model identifiability in practice. We also note that certain biological markers, such as androgen, vary significantly throughout the day and with diets [64], so better clinical protocols may need to be implemented to obtain more accurate measurements.

So far, we have only discussed the applicability of model identifiability in terms of key prognostic parameters. However, one can analytically derive treatment outcomes based on a combination of a set of model parameters with mathematical analysis. This can provide a deeper understanding of key factors that drive the progression of cancer and may even shed light on novel treatments. However, to apply analytical results in practice, one needs to assess the interconnection between the parameters and how they may change based on external factors. If how these parameters change over time during treatment can be assessed, one can then use the analytical condition to determine the treatment outcome directly. Nevertheless, the issue of model identifiability remains a crucial component for this approach to work.

Another aspect of model identifiability is the statistical method used for parameter estimation. Most approaches in literature use an individual fitting, which limits the amount of data used for parameter estimation for each individual data. An alternative approach is to use population fitting with mixed effects. This should not be confused with pulling individual data and fit to the average. Instead, this approach assumes that for each parameter, its value varies per individual, but follows some distributions for the whole population. Thus, we can utilize the data of all patients simultaneously to fit the model. A software often used to implement this approach is Monolix [76]. Examples of this approach can be found in within-host viral dynamics literature [77,78]; however, it has yet to gain traction in mathematical and computational oncology literature.

In summary, we find that incorporating frequent data measurements, different types of data (especially those related to the cancer population), and high-accuracy measurements will increase the likelihood of practical identification of prognostic parameters. As more complex models contain more parameters, making it a more difficult task to obtain complete model identification, our results advocate for the use and development of models with a mechanism that tracks disease progression. By incorporating such a mechanism, a subset of model parameters (associated with the mechanism) naturally becomes the focus of model identifiability. This reduces the issue of model unidentifiability and provides a means for making predictions regarding the outcome of treatment. There are several major limitations to our studies, such as the assumption of a perfect model (no model error). These issues can perhaps be accounted for by continually improving the model development or by using a data assimilation approach, such as the Kalman filter [79,80]. We also do not employ patient-specific data for our simulation study. These can be explored in future studies.

**Author Contributions:** All authors conceptualized the study. T.P. (Tin Phan) wrote the code for model analysis, which was tested by J.B. and T.P. (Taylor Patten) All authors contributed to the analysis equally. T.P. (Tin Phan) and T.P. (Taylor Patten) drafted the initial draft. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All codes will be made available upon request.

**Acknowledgments:** We would like to acknowledge various discussions with Eric J. Kostelich, Yang Kuang, and Rebecca Everett.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviation is used in this manuscript:

MC     Monte Carlo

## References

1.  Kuang, Y.; Nagy, J.D.; Eikenberry, S.E. *Introduction to Mathematical Oncology*; CRC Press: Boca Raton, FL, USA, 2016; Volume 59.
2.  Wodarz, D.; Komarova, N. *Dynamics of Cancer: Mathematical Foundations of Oncology*; World Scientific: River Edge, NJ, USA, 2014.
3.  Bull, J.A.; Byrne, H.M. The hallmarks of mathematical oncology. *Proc. IEEE* **2022**, *110*, 523–540. [CrossRef]
4.  Anderson, A.R.; Quaranta, V. Integrative mathematical oncology. *Nat. Rev. Cancer* **2008**, *8*, 227–234. [CrossRef] [PubMed]
5.  Zhang, X.P.; Liu, F.; Wang, W. Two-phase dynamics of p53 in the DNA damage response. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 8990–8995. [CrossRef] [PubMed]
6.  Gupta, S.; Silveira, D.A.; Mombach, J.C.M. Towards DNA-damage induced autophagy: A Boolean model of p53-induced cell fate mechanisms. *DNA Repair* **2020**, *96*, 102971. [CrossRef]
7.  Altrock, P.M.; Liu, L.L.; Michor, F. The mathematics of cancer: Integrating quantitative models. *Nat. Rev. Cancer* **2015**, *15*, 730–745. [CrossRef]
8.  Jones, H.M.; Mayawala, K.; Poulin, P. Dose selection based on physiologically based pharmacokinetic (PBPK) approaches. *AAPS J.* **2013**, *15*, 377–387. [CrossRef]
9.  Bartelink, I.; van de Stadt, E.; Leeuwerik, A.; Thijssen, V.; Hupsel, J.; van den Nieuwendijk, J.; Bahce, I.; Yaqub, M.; Hendrikse, N. Physiologically based pharmacokinetic (PBPK) modeling to predict PET image quality of three generations EGFR TKI in advanced-stage NSCLC patients. *Pharmaceuticals* **2022**, *15*, 796. [CrossRef]
10. Hirata, Y.; Bruchovsky, N.; Aihara, K. Development of a mathematical model that predicts the outcome of hormone therapy for prostate cancer. *J. Theor. Biol.* **2010**, *264*, 517–527. [CrossRef]
11. Portz, T.; Kuang, Y.; Nagy, J.D. A clinical data validated mathematical model of prostate cancer growth under intermittent androgen suppression therapy. *Aip Adv.* **2012**, *2*, 011002. [CrossRef]
12. Brady-Nicholls, R.; Enderling, H. Range-Bounded Adaptive Therapy in Metastatic Prostate Cancer. *Cancers* **2022**, *14*, 5319. [CrossRef]
13. Phan, T.; Nguyen, K.; Sharma, P.; Kuang, Y. The impact of intermittent androgen suppression therapy in prostate cancer modeling. *Appl. Sci.* **2018**, *9*, 36. [CrossRef]
14. Yin, A.; Moes, D.J.A.; van Hasselt, J.G.; Swen, J.J.; Guchelaar, H.J. A review of mathematical models for tumor dynamics and treatment resistance evolution of solid tumors. *CPT Pharmacomet. Syst. Pharmacol.* **2019**, *8*, 720–737. [CrossRef]
15. Baez, J.; Kuang, Y. Mathematical models of androgen resistance in prostate cancer patients under intermittent androgen suppression therapy. *Appl. Sci.* **2016**, *6*, 352. [CrossRef]
16. Meade, W.; Weber, A.; Phan, T.; Hampston, E.; Resa, L.F.; Nagy, J.; Kuang, Y. High Accuracy Indicators of Androgen Suppression Therapy Failure for Prostate Cancer—A Modeling Study. *Cancers* **2022**, *14*, 4033. [CrossRef]
17. Brady-Nicholls, R.; Zhang, J.; Zhang, T.; Wang, A.Z.; Butler, R.; Gatenby, R.A.; Enderling, H. Predicting patient-specific response to adaptive therapy in metastatic castration-resistant prostate cancer using prostate-specific antigen dynamics. *Neoplasia* **2021**, *23*, 851–858. [CrossRef]
18. Phan, T.; Crook, S.M.; Bryce, A.H.; Maley, C.C.; Kostelich, E.J.; Kuang, Y. Mathematical modeling of prostate cancer and clinical application. *Appl. Sci.* **2020**, *10*, 2721. [CrossRef]
19. West, J.; Adler, F.; Gallaher, J.; Strobl, M.; Brady-Nicholls, R.; Brown, J.S.; Robertson-Tessi, M.; Kim, E.; Noble, R.; Viossat, Y.; et al. A survey of open questions in adaptive therapy: Bridging mathematics and clinical translation. *arXiv* **2022**, arXiv:2210.12062.
20. Claret, L.; Girard, P.; Hoff, P.M.; Van Cutsem, E.; Zuideveld, K.P.; Jorga, K.; Fagerberg, J.; Bruno, R. Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *J. Clin. Oncol.* **2009**, *27*, 4103–4108. [CrossRef]
21. He, C.; Bayakhmetov, S.; Harris, D.; Kuang, Y.; Wang, X. A Predictive Reaction-Diffusion Based Model of E. coliColony Growth Control. *IEEE Control Syst. Lett.* **2020**, *5*, 1952–1957. [CrossRef]
22. Han, L.; Eikenberry, S.; He, C.; Johnson, L.; Preul, M.; Kostelich, E.; Kuang, Y. Patient-specific parameter estimates of glioblastoma multiforme growth dynamics from a model with explicit birth and death rates. *Math. Biosci. Eng. MBE* **2019**, *16*, 5307–5323. [CrossRef]
23. Rutter, E.M.; Stepien, T.L.; Anderies, B.J.; Plasencia, J.D.; Woolf, E.C.; Scheck, A.C.; Turner, G.H.; Liu, Q.; Frakes, D.; Kodibagkar, V.; et al. Mathematical analysis of glioma growth in a murine model. *Sci. Rep.* **2017**, *7*, 1–16. [CrossRef] [PubMed]

24. Wu, Z.; Phan, T.; Baez, J.; Kuang, Y.; Kostelich, E.J. Predictability and identifiability assessment of models for prostate cancer under androgen suppression therapy. *Math. Biosci. Eng. MBE* **2019**, *16*, 3512–3536. [CrossRef] [PubMed]

25. Bruchovsky, N.; Klotz, L.; Crook, J.; Malone, S.; Ludgate, C.; Morris, W.J.; Gleave, M.E.; Goldenberg, S.L. Final results of the Canadian prospective phase II trial of intermittent androgen suppression for men in biochemical recurrence after radiotherapy for locally advanced prostate cancer: Clinical parameters. *Cancer* **2006**, *107*, 389–395. [CrossRef] [PubMed]

26. Morken, J.D.; Packer, A.; Everett, R.A.; Nagy, J.D.; Kuang, Y. Mechanisms of Resistance to Intermittent Androgen Deprivation in Patients with Prostate Cancer Identified by a Novel Computational MethodIdentifying CRPC Mechanisms in Individual Patients. *Cancer Res.* **2014**, *74*, 3673–3683. [CrossRef]

27. Bennett, J.; Hu, X.; Gund, K.; Liu, J.; Porter, A. Clinical data validated mathematical model for intermittent abiraterone response in castration-resistant prostate cancer patients. *SIAM Undergrad Res Online* **2021**, *14*, 58–77. [CrossRef]

28. Phan, T.; Changhan, H.; Martinez, A.; Kuang, Y. Dynamics and implications of models for intermittent androgen suppression therapy. *Math. Biosci. Eng* **2019**, *16*, 187–204. [CrossRef]

29. Miao, H.; Xia, X.; Perelson, A.S.; Wu, H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev.* **2011**, *53*, 3–39. [CrossRef]

30. Murphy, H.; Jaafari, H.; Dobrovolny, H.M. Differences in predictions of ODE models of tumor growth: A cautionary example. *BMC Cancer* **2016**, *16*, 1–10. [CrossRef]

31. Eisenberg, M.C.; Jain, H.V. A confidence building exercise in data and identifiability: Modeling cancer chemotherapy as a case study. *J. Theor. Biol.* **2017**, *431*, 63–78. [CrossRef]

32. Nguyen, K.; Li, K.; Flores, K.; Tomaras, G.; Dennison, S.M.; McCarthy, J. Parameter estimation and identifiability analysis for a bivalent analyte model of monoclonal antibody-antigen binding. *bioRxiv* **2022**. [CrossRef]

33. Evangelou, N.; Wichrowski, N.J.; Kevrekidis, G.A.; Dietrich, F.; Kooshkbaghi, M.; McFann, S.; Kevrekidis, I.G. On the parameter combinations that matter and on those that do not: data-driven studies of parameter (non) identifiability. *PNAS Nexus* **2022**, *1*, pgac154. [CrossRef]

34. Renardy, M.; Kirschner, D.; Eisenberg, M. Structural identifiability analysis of age-structured PDE epidemic models. *J. Math. Biol.* **2022**, *84*, 1–30. [CrossRef]

35. Tsamandouras, N.; Rostami-Hodjegan, A.; Aarons, L. Combining the 'bottom up' and 'top down' approaches in pharmacokinetic modelling: Fitting PBPK models to observed clinical data. *Br. J. Clin. Pharmacol.* **2015**, *79*, 48–55. [CrossRef]

36. Tuncer, N.; Timsina, A.; Nuno, M.; Chowell, G.; Martcheva, M. Parameter identifiability and optimal control of an SARS-CoV-2 model early in the pandemic. *J. Biol. Dyn.* **2022**, *16*, 412–438. [CrossRef]

37. Laubmeier, A.N.; Cazelles, B.; Cuddington, K.; Erickson, K.D.; Fortin, M.J.; Ogle, K.; Wikle, C.K.; Zhu, K.; Zipkin, E.F. Ecological dynamics: Integrating empirical, statistical, and analytical methods. *Trends Ecol. Evol.* **2020**, *35*, 1090–1099. [CrossRef]

38. Phan, T.; He, C.; Loladze, I.; Prater, C.; Elser, J.; Kuang, Y. Dynamics and growth rate implications of ribosomes and mRNAs interaction in E. coli. *Heliyon* **2022**, *8*, e09820. [CrossRef]

39. Jain HV, Norton K-A, P.B.; TL, J. SMoRe ParS: A novel methodology for bridging modeling modalities and experimental data applied to 3D vascular tumor growth. *Front. Mol. Biosci.* **2022**, *9*, 1056461. [CrossRef]

40. Phan, T.; Brozak, S.; Pell, B.; Gitter, A.; Xiao, A.; Mena, K.D.; Kuang, Y.; Wu, F. A simple SEIR-V model to estimate COVID-19 prevalence and predict SARS-CoV-2 transmission using wastewater-based surveillance data. *Sci. Total Environ.* **2023**, *857*, 159326. [CrossRef]

41. Denis-Vidal, L.; Joly-Blanchard, G. An easy to check criterion for (un) indentifiability of uncontrolled systems and its applications. *IEEE Trans. Autom. Control* **2000**, *45*, 768–771. [CrossRef]

42. Bellu, G.; Saccomani, M.P.; Audoly, S.; D'Angiò, L. DAISY: A new software tool to test global identifiability of biological and physiological systems. *Comput. Methods Programs Biomed.* **2007**, *88*, 52–61. [CrossRef]

43. Banks, H.T.; Hu, S.; Thompson, W.C. *Modeling and Inverse Problems in the Presence of Uncertainty*; CRC Press: Boca Raton, FL, USA, 2014.

44. Tuncer, N.; Gulbudak, H.; Cannataro, V.L.; Martcheva, M. Structural and practical identifiability issues of immuno-epidemiological vector–host models with application to rift valley fever. *Bull. Math. Biol.* **2016**, *78*, 1796–1827. [CrossRef] [PubMed]

45. Errico, R.M.; Yang, R.; Privé, N.C.; Tai, K.S.; Todling, R.; Sienkiewicz, M.E.; Guo, J. Development and validation of observing-system simulation experiments at NASA's Global Modeling and Assimilation Office. *Q. J. R. Meteorol. Soc.* **2013**, *139*, 1162–1178. [CrossRef]

46. Durazo, J.; Kostelich, E.; Mahalov, A.; Tang, W. Observing system experiments with an ionospheric electrodynamics model. *Phys. Scr.* **2016**, *91*, 044001. [CrossRef]

47. Pasetto, S.; Enderling, H.; Gatenby, R.; Brady-Nicholls, R. Intermittent Hormone Therapy Models Analysis and Bayesian Model Comparison for Prostate Cancer. *Bull. Math. Biol.* **2022**, *84*, 1–36. [CrossRef] [PubMed]

48. Coletti, R.; Leonardelli, L.; Parolo, S.; Marchetti, L. A QSP model of prostate cancer immunotherapy to identify effective combination therapies. *Sci. Rep.* **2020**, *10*, 1–18. [CrossRef]

49. Coletti, R.; Pugliese, A.; Lunardi, A.; Caffo, O.; Marchetti, L. A Model-Based Framework to Identify Optimal Administration Protocols for Immunotherapies in Castration-Resistance Prostate Cancer. *Cancers* **2021**, *14*, 135. [CrossRef]

50. Jain, H.V.; Sorribes, I.C.; Handelman, S.K.; Barnaby, J.; Jackson, T.L. Standing variations modeling captures inter-individual heterogeneity in a deterministic model of prostate cancer response to combination therapy. *Cancers* **2021**, *13*, 1872. [CrossRef]

51. Valle, P.A.; Coria, L.N.; Carballo, K.D. Chemoimmunotherapy for the treatment of prostate cancer: Insights from mathematical modelling. *Appl. Math. Model.* **2021**, *90*, 682–702. [CrossRef]
52. Coletti, R.; Pugliese, A.; Marchetti, L. Modeling the effect of immunotherapies on human castration-resistant prostate cancer. *J. Theor. Biol.* **2021**, *509*, 110500. [CrossRef]
53. Barnaby, J.P.; Sorribes, I.C.; Jain, H.V. Relating prostate-specific antigen leakage with vascular tumor growth in a mathematical model of prostate cancer response to androgen deprivation. *Comput. Syst. Oncol.* **2021**, *1*, e1014. [CrossRef]
54. Barnaby, J.; Jain, H.V. Combining Androgen Deprivation and Immunotherapy in Prostate Cancer Treatment: A Mechanistic Approach. *Appl. Sci.* **2022**, *12*, 6954. [CrossRef]
55. Siewe, N.; Friedman, A. Combination therapy for mCRPC with immune checkpoint inhibitors, ADT and vaccine: A mathematical model. *PLoS ONE* **2022**, *17*, e0262453. [CrossRef]
56. Foryś, U.; Nahshony, A.; Elishmereni, M. Mathematical model of hormone sensitive prostate cancer treatment using leuprolide: A small step towards personalization. *PLoS ONE* **2022**, *17*, e0263648. [CrossRef]
57. Hynes, P.G.; Kelly, K. Prostate cancer stem cells: The case for model systems. *J. Carcinog.* **2012**, *11*, 1.
58. Mei, W.; Lin, X.; Kapoor, A.; Gu, Y.; Zhao, K.; Tang, D. The contributions of prostate cancer stem cells in prostate cancer initiation and metastasis. *Cancers* **2019**, *11*, 434. [CrossRef]
59. Enderling, H.; Hlatky, L.; Hahnfeldt, P. Cancer stem cells: A minor cancer subpopulation that redefines global cancer features. *Front. Oncol.* **2013**, *3*, 76. [CrossRef]
60. Enderling, H.; Hahnfeldt, P. Cancer stem cells in solid tumors: Is 'evading apoptosis'a hallmark of cancer? *Prog. Biophys. Mol. Biol.* **2011**, *106*, 391–399. [CrossRef]
61. Vummidi Giridhar, P.; Williams, K.; VonHandorf, A.P.; Deford, P.L.; Kasper, S. Constant Degradation of the Androgen Receptor by MDM2 Conserves Prostate Cancer Stem Cell Integrity Constant MDM2-Mediated AR Degradation Conserves CSC Stemness. *Cancer Res.* **2019**, *79*, 1124–1137. [CrossRef]
62. Brady-Nicholls, R.; Nagy, J.D.; Gerke, T.A.; Zhang, T.; Wang, A.Z.; Zhang, J.; Gatenby, R.A.; Enderling, H. Prostate-specific antigen dynamics predict individual responses to intermittent androgen deprivation. *Nat. Commun.* **2020**, *11*, 1750. [CrossRef]
63. Civenni, G.; Bosotti, R.; Timpanaro, A.; Vazquez, R.; Merulla, J.; Pandit, S.; Rossi, S.; Albino, D.; Allegrini, S.; Mitra, A.; et al. Epigenetic control of mitochondrial fission enables self-renewal of stem-like tumor cells in human prostate cancer. *Cell Metab.* **2019**, *30*, 303–318. [CrossRef]
64. Reckell, T.; Nguyen, K.; Phan, T.; Crook, S.; Kostelich, E.J.; Kuang, Y. Modeling the synergistic properties of drugs in hormonal treatment for prostate cancer. *J. Theor. Biol.* **2021**, *514*, 110570. [CrossRef] [PubMed]
65. Padmanabhan, R.; Meskin, N.; Moustafa, A.E.A. Hormone Therapy Models. In *Mathematical Models of Cancer and Different Therapies*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 135–156.
66. Droop, M.R. Vitamin B12 and marine ecology. IV. The kinetics of uptake, growth and inhibition in Monochrysis lutheri. *J. Mar. Biol. Assoc. United Kingd.* **1968**, *48*, 689–733. [CrossRef]
67. Loladze, I.; Kuang, Y.; Elser, J.J. Stoichiometry in producer–grazer systems: Linking energy flow with element cycling. *Bull. Math. Biol.* **2000**, *62*, 1137–1162. [CrossRef] [PubMed]
68. Eikenberry, S.E.; Nagy, J.D.; Kuang, Y. The evolutionary impact of androgen levels on prostate cancer in a multi-scale mathematical model. *Biol. Direct* **2010**, *5*, 1–28. [CrossRef] [PubMed]
69. Everett, R.; Packer, A.; Kuang, Y. Can mathematical models predict the outcomes of prostate cancer patients undergoing intermittent androgen deprivation therapy? *Biophys. Rev. Lett.* **2014**, *9*, 173–191. [CrossRef]
70. Sontag, E.D. For differential equations with r parameters, 2r+ 1 experiments are enough for identification. *J. Nonlinear Sci.* **2002**, *12*, 553–583. [CrossRef]
71. De Bono, J.S.; Scher, H.I.; Montgomery, R.B.; Parker, C.; Miller, M.C.; Tissing, H.; Doyle, G.V.; Terstappen, L.W.; Pienta, K.J.; Raghavan, D. Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clin. Cancer Res.* **2008**, *14*, 6302–6309. [CrossRef]
72. Azad, A.A.; Volik, S.V.; Wyatt, A.W.; Haegert, A.; Le Bihan, S.; Bell, R.H.; Anderson, S.A.; McConeghy, B.; Shukin, R.; Bazov, J.; et al. Androgen Receptor Gene Aberrations in Circulating Cell-Free DNA: Biomarkers of Therapeutic Resistance in Castration-Resistant Prostate CancerAR Gene Aberrations in Circulating Cell-Free DNA. *Clin. Cancer Res.* **2015**, *21*, 2315–2324. [CrossRef]
73. Wyatt, A.W.; Azad, A.A.; Volik, S.V.; Annala, M.; Beja, K.; McConeghy, B.; Haegert, A.; Warner, E.W.; Mo, F.; Brahmbhatt, S.; et al. Genomic alterations in cell-free DNA and enzalutamide resistance in castration-resistant prostate cancer. *JAMA Oncol.* **2016**, *2*, 1598–1606. [CrossRef]
74. Kohli, M.; Tan, W.; Zheng, T.; Wang, A.; Montesinos, C.; Wong, C.; Du, P.; Jia, S.; Yadav, S.; Horvath, L.G.; et al. Clinical and genomic insights into circulating tumor DNA-based alterations across the spectrum of metastatic hormone-sensitive and castrate-resistant prostate cancer. *EBioMedicine* **2020**, *54*, 102728. [CrossRef]
75. Ionescu, F.; Zhang, J.; Wang, L. Clinical Applications of Liquid Biopsy in Prostate Cancer: From Screening to Predictive Biomarker. *Cancers* **2022**, *14*, 1728. [CrossRef]
76. Lavielle, M.; Mentré, F. Estimation of population pharmacokinetic parameters of saquinavir in HIV patients with the MONOLIX software. *J. Pharmacokinet. Pharmacodyn.* **2007**, *34*, 229–249. [CrossRef]
77. Ke, R.; Zitzmann, C.; Ho, D.D.; Ribeiro, R.M.; Perelson, A.S. In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2111477118. [CrossRef]

78. Gonçalves, A.; Bertrand, J.; Ke, R.; Comets, E.; De Lamballerie, X.; Malvy, D.; Pizzorno, A.; Terrier, O.; Rosa Calatrava, M.; Mentré, F.; et al. Timing of antiviral treatment initiation is critical to reduce SARS-CoV-2 viral load. *CPT Pharmacomet. Syst. Pharmacol.* **2020**, *9*, 509–514. [CrossRef]

79. Durazo, J.; Kostelich, E.J.; Mahalov, A. Data assimilation for ionospheric space-weather forecasting in the presence of model bias. *Front. Appl. Math. Stat.* **2021**, *7*, 679477. [CrossRef]

80. Hunt, B.R.; Kostelich, E.J.; Szunyogh, I. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Phys. D Nonlinear Phenom.* **2007**, *230*, 112–126. [CrossRef]