Review

System identification and parameter estimation in mathematical medicine: examples demonstrated for prostate cancer

Yoshito Hirata[1,2], Kai Morino[2], Taiji Suzuki[3], Qian Guo[4], Hiroshi Fukuhara[5] and Kazuyuki Aihara[1,2,*]

[1] Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan
[2] Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan
[3] Department of Mathematical and Computing Sciences, Graduate School of Information Science and Engineering, Tokyo Institute of Technology, Tokyo 152-8552, Japan
[4] Department of Mathematics, Shanghai Normal University, Shanghai 200234, China
[5] Department of Urology, Graduate School of Medicine, The University of Tokyo, Tokyo 113-8655, Japan
* Correspondence: aihara@sat.t.utokyo.ac.jp

We review our studies on how to identify the most appropriate models of diseases, and how to determine their parameters in a quantitative manner given a short time series of biomarkers, using intermittent androgen deprivation therapy of prostate cancer as an example. Recently, it has become possible to estimate the specific parameters of individual patients within a reasonable time by employing the information concerning other previous patients as a prior. We discuss the importance of using multiple mathematical methods simultaneously to achieve a solid diagnosis and prognosis in the future practice of personalized medicine.

Keywords: mathematical medicine; dynamical model; parameter estimation

INTRODUCTION

Developments in measurement techniques have enabled us to model diseases quantitatively and mathematically. For example, several mathematical models [1–5] have been proposed for intermittent androgen deprivation therapy for prostate cancer [6–10]. When we employ such a mathematical dynamical model in a clinical setting, there are certain problems that commonly appear, even if the target diseases are different. For example, we need to estimate a set of parameters characterizing each patient from a short time series of biomarkers [11,12]. Further problems include how to optimize a treatment schedule given such a personalized set of parameters [13–15].

In this review, we focus on the first problem of mathematical medicine, i.e., how to estimate a set of parameters for quantitatively characterizing each patient. Because the mathematical modeling of intermittent androgen deprivation therapy for prostate cancer with the sensitive biomarker Prostate-Specific Antigen (PSA) is well developed, we employ it as an example.

PHYSIOLOGY OF PROSTATE CANCER

Androgen deprivation therapy (ADT) is often used to treat prostate cancer. ADT suppresses the growth of prostate cancer by lowering the androgen level in patients. Although ADT often shows remarkable effects in tumor regression initially, prostate cancer acquires the ability to grow without androgen in long-term ADT treatment, resulting in eventual relapse.

Intermittent ADT [6–10] has been introduced in order to overcome this problem and sustain the hormonal sensitivity in prostate cancer. Intermittent ADT is an approach that is based on the premise that cycles of androgen deprivation followed by re-exposure may delay "androgen independence", reduce treatment morbidity, and improve quality of life. In intermittent ADT, patients continue ADT until the PSA levels decline sufficiently, and then cease ADT. After the PSA levels elevate again to upper threshold values, ADT is resumed. From a mathematical viewpoint, one problem concerning intermittent ADT is that the protocol of intermittent ADT is

dependent on the decisions of individual medical doctors, and it is not clear whether the applied protocol is optimal or not for each patient.

In order to personalize intermittent ADT, several mathematical models of intermittent ADT have been proposed [1–5]. Among these models, Ideta *et al.* [1] were the first to mention the mathematical possibility that intermittent ADT can prevent the relapse of prostate cancer if the population of androgen independent cancer cells shrinks after the hormonal therapy is stopped. However, this model is unable to explain the bi-phasic decline of PSA in periods that the hormonal therapy is carried out. In order to describe the bi-phasic decline during on-treatment periods, Hirata *et al.* [3] have proposed a three-dimensional piecewise-linear model. This model describes the dynamics of prostate cancer effectively and quantitatively, and it will be introduced mathematically in the next section.

CONSIDERED MODEL

In general, a dynamical system describing disease dynamics can be written in the following form:

$$\frac{dx}{dt} = f_m(x), \qquad (1)$$

where x describes a state of the target disease, m corresponds to a regimen of treatment, and f_m describes the dynamics of the disease given a regimen m of treatment.

Such a system encompasses intermittent ADT of prostate cancer [6–10]. Here, we describe the model of Hirata *et al.* [3]. The model describes the dynamical behavior of prostate cancer, and the dynamics has been comprehensively studied [11,12,14–22]. To define the model, we first define the following three variables: x_1 corresponds to the non-dimensionalized number of androgen dependent cancer cells, and x_2 and x_3 to the non-dimensionalized numbers of androgen independent cancer cells. When the hormonal therapy is active ($m = 1$), the cancer population specified by x_1 may change to that specified by x_2 or x_3, and the cancer population specified by x_2 may change to that specified by x_3 (see Figure 1). When the hormonal therapy is suspended ($m = 0$), androgen independent cells, expressed by x_2, may change back to androgen dependent cells, expressed by x_1. However, androgen independent cells expressed by x_3, cannot change back to cancer population specified by either x_1 or x_2 (see Figure 1). That is, the androgen independent cells expressed by x_2 are cells generated through reversible changes, and the androgen independent cells expressed by x_3 are cells generated through irreversible changes like mutation. Mathematically, the dynamics among these variables is written as

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} w_{1,1}^1 & 0 & 0 \\ w_{2,1}^1 & w_{2,2}^1 & 0 \\ w_{3,1}^1 & w_{3,2}^1 & w_{3,3}^1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \qquad (2)$$

while the hormonal therapy is active ($m = 1$), and

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} w_{1,1}^0 & w_{1,2}^0 & 0 \\ 0 & w_{2,2}^0 & 0 \\ 0 & 0 & w_{3,3}^0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \qquad (3)$$

while the hormonal therapy is inactive ($m = 0$). The level of PSA is now assumed to be $p(t) = x_1(t) + x_2(t) + x_3(t)$, for simplicity. The unit of t is one day.
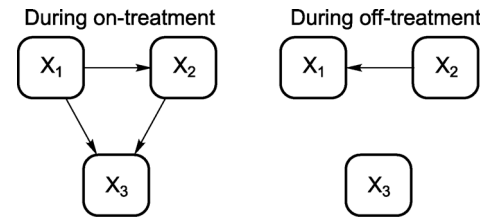
Figure 1. Possible state transitions among cancer cells. The symbols x_1, x_2, and x_3 represent the non-dimensionalized numbers of androgen dependent cancer cells, androgen independent cancer cells generated through reversible changes, and androgen independent cancer cells generated through irreversible changes, respectively [3]. The left panel shows possible state transitions when the hormonal therapy is active, and the right panel shows a possible state transition when the hormonal therapy is inactive.

In the paper of Hirata *et al.* [3], a method of fitting with constraints was also proposed to achieve physiological validity of the model. In particular, the method focused on the non-negativities of x_1, x_2, and x_3; moderate amounts of changes within a day; and the reproducibility of relapse when the hormonal therapy is continued. To achieve the non-negativity of x_1, x_2, and x_3, we assume that $w_{i,j}^m \geqslant 0$ when $i \neq j$ and $m \in \{0,1\}$. To ensure moderate amounts of changes within one day, we assume that $-0.2 \leqslant w_{i,i}^m \leqslant 0.2$, $w_{i,j}^m \leqslant 0.1$ when $i \neq j$, and $-0.2 \leqslant \sum_i w_{i,j}^m \leqslant 0.2$ for $j \in \{1,2,3\}$ and $m \in \{0,1\}$. For the reproducibility of relapse, we assume that $w_{3,3}^1 \geqslant 0$, $\overline{p}(360) \leqslant 2$, and $\overline{p}(360 \times 5) \geqslant 10$, where $\overline{p}(t)$ is the level of PSA at time t in the case that we keep the hormonal therapy active ($m = 1$) for all time. We use the penalty method to enforce these conditions. If we would like to realize $x \geqslant 0$, then we minimize the following quantity $h(x)$:

$$h(x) = \begin{cases} 100(1-x), & x < 0, \\ 0, & otherwise \end{cases}. \qquad (4)$$

The function $h(x) = 0$ when $x \geqslant 0$ is satisfied. Thus, by combining this with the fitting error represented by the L_1 norm, the cost function to determine the parameters and initial conditions $q = (w, x(0))$ can be written as follows:

$$\min_{q}\{\Sigma_{i=1}^{I}|o_i - p(t_i)| + C(q)\}, \qquad (5)$$

where $\{(t_i, o_i)|i = 1, 2, ..., I\}$ are the observed PSA values o_i at t_i, and

$$C(q) = \sum_{m, i \neq j} h(w_{i,j}^m) + \sum_{i,m} h(w_{i,i}^m + 0.2)$$

$$+ \sum_{i,m} h(-w_{i,i}^m + 0.2) + \sum_{m, i \neq j} h(0.1 - w_{i,j}^m)$$

$$+ \sum_{m, i \neq j} h(\sum_i w_{i,j}^m + 0.2) + \sum_{m, i \neq j} h(-\sum_i w_{i,j}^m + 0.2)$$

$$+ h(w_{3,3}^1) + h(2 - \overline{p}(360)) + h(\overline{p}(360 \times 5) - 10). \qquad (6)$$

The last two terms of Equation (6) correspond to constraints realizing the assumptions that the PSA value will once go down and reach the value less than 2 ng/ml one year after the start of the hormonal therapy, while it will become greater than 10 ng/ml after 5 years if we continue the hormonal therapy. These assumptions came from general common observations made by urologists [3].

A summary of recent methods [11,12,17,19,20] for the system identification and parameter estimation is presented in Table 1. The methods are roughly classified into three classes: parametric estimation, semi-parametric estimation, and non-parametric estimation. In the following sections, we introduce these methods one at a time.

PARAMETRIC ESTIMATION

Two parameter-estimation methods have been proposed for estimating a set of parameters for dynamical diseases. The first method is to use cross entropy [19], and the second method is to use the Bayesian theorem with the data of past patients [12].

The method of cross entropy [19] involves choosing a set of parameters by recursively minimizing the Kullback-Leibler divergence [23] between the previous truncated distribution and the current distribution without the truncation such that the cost function is sufficiently small. The method of cross entropy can be formulated as follows. First, we define a cost function $F(q)$ as

$$F(q) = \sqrt{\sum_i (o_i - p(t_i))^2} + C(q). \qquad (7)$$

At each step of iteration, we randomly draw N sets of parameters $q_1, q_2, ..., q_N$ such that each ith parameter value follows the Gaussian distribution with mean $u_i^{(k-1)}$ and standard deviation $\sigma_i^{2^{(k-1)}}$, where k represents the iteration number. Then, we obtain the r-quantile $\gamma^{(k)}$ among $\{F(q_n)|n = 1, 2, ..., N\}$. After obtaining the optimizer $q* = \underset{q \in \{q_n|n=1, 2, ..., N\}}{\text{argmin}} F(q)$, we move each

Table 1. Summary of methods.

Methods	Fitting with constraints [3]	Bootstrapping [17]	Cross entropy [19]	Variational Bayes [20]	Gaussian process [20]	Temporal expert advice [11]	Bayes with past patients' prior [12]
Parametric	○	×	○	×	×	×	○
Semi-parametric	×	○	×	×	×	○	×
Non-parametric	×	×	×	○	○	×	×
Parameter distribution	×	○	○	○	○	×	×
Prediction interval	×	○	○	○	○	○	×
Point prediction	○	×	×	○	○	○	○
Using constraints	○	○	○	×	×	○	○
Reproducibility of relapse under continuous ADT	○	○	○	×	×	○	○
Use of past other patients' data	×	×	×	×	×	○	○
Necessary length of cycles in terms of intermittent ADT	2.5	2.5	2.5	2	2	$\leqslant 0.5$	$\leqslant 1.5$
Computational efficiency	○	×	×	◎	◎	△	○

In this table, ○ means that the corresponding property is satisfied, and × means that the corresponding property is not satisfied. The symbol ◎ means that the corresponding property is strongly satisfied, and △ means that the corresponding property is satisfied in part.

q_n slightly closer to the optimizer q^*, namely $\hat{q}_n = \delta q^* + (1-\delta)q_n$. Then, we choose mean $\hat{\mu}_i^{(k)}$ and deviation $\hat{\sigma}_i^{2(k)}$ such that we can minimize the Kullback-Leibler divergence between the previous truncated distribution and the current distribution. Equivalently, we maximize the following quantity in terms of the parameters $\left\{\hat{\mu}_i^{(k)}\right\}$ and $\left\{\hat{\sigma}_i^{2(k)}\right\}$ of the current distribution, expressed by the product of univariate Gaussian distributions:

$$\frac{1}{N}\sum_{n=1}^{N}\left\{I_{\left\{F(\hat{q}_n)<\gamma^{(k)}\right\}} + \varsigma I_{\left\{F(\hat{q}_n)\geqslant\gamma^{(k)}\right\}}\right\}\ln f_k(\hat{q}_n), \quad (8)$$

Where ς is a weight, and f_k follows the product of univariate Gaussian distributions of mean $\hat{\mu}_i^{(k)}$ and deviation $\hat{\sigma}_i^{2(k)}$. By differentiating Equation (8) in terms of $\hat{\mu}_i^{(k)}$ and $\hat{\sigma}_i^{2(k)}$, we obtain

$$\hat{\mu}_i^{(k)} = \frac{\sum_{n=1}^{N}\left\{I_{\left\{F(\hat{q}_n)<\gamma^{(k)}\right\}} + \varsigma I_{\left\{F(\hat{q}_n)\geqslant\gamma^{(k)}\right\}}\right\}\hat{q}_{ni}}{\sum_{n=1}^{N}\left\{I_{\left\{F(\hat{q}_n)<\gamma^{(k)}\right\}} + \varsigma I_{\left\{F(\hat{q}_n)\geqslant\gamma^{(k)}\right\}}\right\}}, \quad (9)$$

and

$$\hat{\sigma}_i^{2(k)} = \frac{\sum_{n=1}^{N}\left\{I_{\left\{F(\hat{q}_n)<\gamma^{(k)}\right\}} + \varsigma I_{\left\{F(\hat{q}_n)\geqslant\gamma^{(k)}\right\}}\right\}\left(\hat{q}_{ni}-\hat{\mu}_i^{k}\right)^2}{\sum_{n=1}^{N}\left\{I_{\left\{F(\hat{q}_n)<\gamma^{(k)}\right\}} + \varsigma I_{\left\{F(\hat{q}_n)\geqslant\gamma^{(k)}\right\}}\right\}}. \quad (10)$$

Then, we adjust $\mu_i^{(k-1)}$ and $\sigma_i^{2(k-1)}$ towards $\hat{\mu}_i^{(k)}$ and $\hat{\sigma}_i^{2(k)}$, as

$$\mu_i^{(k)} = \alpha\hat{\mu}_i^{(k)} + (1-\alpha)\mu_i^{(k-1)}, \quad (11)$$

$$\sigma_i^{2(k)} = \frac{\beta}{k}\hat{\sigma}_i^{2(k)} + \left(1-\frac{\beta}{k}\right)\sigma_i^{2(k-1)}, \quad (12)$$

and complete a set of routines for an iteration. Here, we set $\alpha\in(0.5, 1)$ and $\beta\in(0.8, 1)$. We stop the iterations when either the cost function reaches our intended lower bound or the number of iterations reaches a certain number.

This method was demonstrated by Guo et al. [19]. The parameter values can be estimated using the first two and half cycles of intermittent ADT. The prediction intervals for the PSA values were also estimated.

The second method is based on the Bayesian theorem, with a prior that is constructed based on the data of past patients [12]. Suppose that we maximize the posterior probability $Q(q|o)$ of a set q of parameters, given a dataset o. Using the Bayesian theorem, the posterior probability can be written as follows:

$$Q(q|o) \propto Q(o|q)Q(q). \quad (13)$$

By taking the logarithm, Equation (13) can be rewritten as

$$\ln Q(q|o) = \ln Q(o|q) + \ln Q(q) + V, \quad (14)$$

where V is constant. We now model the prior distribution $Q(q)$ by using the multivariate Gaussian distribution estimated from past patients. Letting $\overline{q}$ and W denote the mean and the covariance matrix, and enforcing the constraints of Equation (6) for the physiological appropriatenesss, the problem of estimating a set of parameters for the current patient can be written as

$$\min_q\left\{\sum_{i=1}^{I}\frac{1}{2\sigma^2}(o_i-p(t_i))^2+(q-\overline{q})^TW^{-1}(q-\overline{q})+C(q)\right\}, \quad (15)$$

where σ is the standard deviation for the observations of the PSA values. In the paper of Hirata et al. [12], we employed the entire datasets of 36 other patients, who were not used later in evaluating the fitting accuracy for constructing the prior distribution.

In the paper of Hirata et al. [12], it was shown that the estimated parameter values with the first one and half cycles exhibited correlations with those obtained using the entire datasets.

SEMI-PARAMETRIC ESTIMATION

By a semi-parametric method, we mean a method that combines the parametric model described in Equations (2) and (3) with a non-parametric approach such that we can quantify the uncertainty in terms of the estimation of the parameters in Equations (2) and (3). Two methods have been proposed for estimating a set of parameters by using the semi-parametric estimation. The first method is that of bootstrapping [17]. The second method is called the temporal expert advice method [11].

The bootstrapping method [17] is quite simple, and takes into account the uncertainty owing to a few measurements of PSA. We resample a given dataset with replacements 100 times. For each resampled dataset, we obtain a set of parameters resulting from Equation (5), and represent the uncertainty as the distribution of the 100 estimated sets of parameters. The problem with the bootstrapping method is its huge computational cost.

In the temporal expert advice method [11], we prepare experts by using the sets of parameters for past patients and fitting only the initial conditions with the data of the target patient. Then, the initial value obtained using the set of parameters for each past patient provides a prediction for the future. Thus, we use the first few time points to evaluate whether the past patients have similar dynamical disease behavior as the current patient or not. Then, depending on the similarities, we either take the weighted average over the predictions or we take an ensemble of

predictions to obtain the prediction intervals. This method has been tested in Morino *et al.* [11] on the datasets of patients who had previously undergone the first radical treatments and were waiting for the start of their next additional treatment.

NON-PARAMETRIC ESTIMATION

There have also been two non-parametric-estimation methods proposed in the context of the mathematical modeling of prostate cancer under intermittent ADT. The first method is the variational Bayes [20] method, and the second is the Gaussian process [20] method.

In the variational Bayes method [20], we use a piecewise affine model. Consider a phase space X that includes $(2m+1)$ — dimensional vectors $x_t = [y_{t-1}, y_{t-2}, ..., y_{t-m}, u_t, u_{t-1}, ..., u_{t-m}]$, where y_t corresponds to the logarithm of the PSA value at time t, and u_t to the treatment option at time t. Suppose that a partition $X_i = \{z \in R^{2m+1} | H_i z \leqslant K_i, I_i z < G_i\}$ of the phase space X is given, satisfying $X = U_i X_i$. Then, a piecewise affine model to be used in this case is defined as

$$f(x_t) = w_i^T \begin{bmatrix} x_t \\ 1 \end{bmatrix}, \qquad (16)$$

for $x_t \in X_i$. We would like to approximate y_t by $f(x_t)$. We denote $D = \{(x_t, y_t)\}$.

When we apply the Gaussian mixture for modeling x_t, it can be written as

$$Q(x|\Psi, M) = \sum_i \varphi_i N(x|\rho_i, S_i^{-1}), \qquad (17)$$

Where φ_i denotes the weight of the ith Gaussian distribution, N represents the Gaussian distribution, and ρ_i and S_i^{-1} denote the mean and the covariance matrix for the ith Gaussian distribution. We also use a latent variable z_i^n, where if the nth data point belongs to the ith component, then $z_i^n = 1$, otherwise $z_i^n = 0$. Then, each x_t is modeled as

$$Q(x_t|z_i^n = 1, \Psi) = N(x_n|\rho_i, S_i^{-1}), \qquad (18)$$

and each y_t is modeled as

$$Q(y_t|x_t, z_i^n = 1, \Theta) = N(y_t|f(x_t), \beta_i^{-1}). \qquad (19)$$

Therefore, the total probability that we have data D and latent variables $Z = \{z_i^n\}$ for the Gaussian mixture M, a model of input Ψ, and a model of output Θ, is given by

$$Q(D, Z|M, \Psi, \Theta) = \Pi_n \Pi_i \{\varphi_i N(x_t|\rho_i, S_i^{-1}) N(y_t|f(x_t), \beta_i^{-1})\}. \qquad (20)$$

The estimates of the parameters are obtained as the mean

with respect to the posterior distribution constructed by the Bayes theorem. However, the computation of the posterior is a difficult task in the Gaussian mixture model. Thus, in the variational Bayes method we approximate the posterior as a computationally tractable form and obtain the parameters M, Ψ, and Θ. This is the variational Bayes method for non-parametric estimation.

The Gaussian process method [20] can be summarized as follows. Suppose that $\widetilde{f}(x)$ obeys the zero-mean Gaussian distribution. That is, for any finite input points $(x_1, x_2, ..., x_k)$, the joint distribution of $(\widetilde{f}(x_1), \widetilde{f}(x_2), ..., \widetilde{f}(x_k))$ is zero-mean multivariate Gaussian. Then, we can define a covariance function as

$$k(x, x') = E_{\widetilde{f} \sim GP}[\widetilde{f}(x)\widetilde{f}(x')], \qquad (21)$$

where $E_{\widetilde{f} \sim GP}$ represents the operation of taking the expectation in terms of the Gaussian process $\widetilde{f}$. Let $f = [\widetilde{f}(x_1), \widetilde{f}(x_2), ..., \widetilde{f}(x_n)]^T$, and $f^* = \widetilde{f}(x)$. Then, we can define the joint distribution $Q([f^T f^*]^T)$ as

$$N\left([f^T f^*]^T | 0, \begin{bmatrix} K_{11} & K_{12} \\ K_{12}^T & K_{22} \end{bmatrix}\right), \qquad (22)$$

where we define

$$K_{11} = (k(x_n, x_{n'}))_{n=1, n'=1}^{N,N}, \qquad (23)$$

$$K_{12} = [k(x_1, x), k(x_2, x), ..., k(x_n, x)]^T, \qquad (24)$$

and

$$K_{22} = k(x, x). \qquad (25)$$

We employ the distribution introduced above as a prior distribution of $[f^T f^*]$. We also model the likelihood of f as a Gaussian distribution, as follows:

$$Q(y|f, \sigma^2 I). \qquad (26)$$

Because the posterior distribution of f^* also follows the Gaussian distribution, the mean $\overline{f^*}$ of f^* becomes

$$\overline{f^*} = K_{12}^T (K_{11} + \sigma^2 I)^{-1} y. \qquad (27)$$

Writing the mean in a different way, we have

$$\overline{f^*} = \Sigma_n k(x, x_n)\alpha_n, \qquad (28)$$

where $\alpha = (K_{11} + \sigma^2 I)^{-1} y$. The variance of f^* is given by

$$K_{22} - K_{12}^T (K_{11} + \sigma^2 I)^{-1} K_{12}. \qquad (29)$$

Suzuki and Aihara [20] discussed the fact that the variational Bayes method is slightly more effective for the long-term predictions, while the Gaussian process is better for short-term predictions.

DISCUSSION

Here, we return to Table 1. The methods can be discussed from three different aspects. The first is whether we can provide the parameter distribution and prediction interval. The methods of the bootstrapping [17], the cross entropy [19], the variational Bayes [20], and the Gaussian process [20] can provide the parameter distribution, and thus the prediction interval as well. The method of the temporal expert advice [11] can provide the prediction interval by weighting predictions according to each expert with Gaussian distributions, although this method cannot provide the parameter distribution itself.

The second aspect is whether we use constraints for the parameters. By using constraints for the parameters, we can reproduce relapse under continuous ADT. Although the parametric and semi-parametric estimations can be combined with constraints, and thus reproduce relapse of prostate cancer under continuous ADT, this is not the case for non-parametric estimation.

The third aspect is whether we can use the data of past patients simultaneously with the proposed methods. The methods of the temporal expert advice [11] and the Bayes method with the prior of past patients [12] can be combined with the information of past patients, while the other methods currently cannot. This property is directly linked with how short the time series of the current patient can be. Namely, if we use the data of past patients, we can use a shorter time series to estimate the set of parameters for the current patient.

An important point in the actual clinical setting is whether we can provide multiple indices to make diagnosis and prognosis. Because the methods for parameter estimation discussed here can be used in conjunction with other mathematical models, and other mathematical models for intermittent androgen deprivation therapy of prostate cancer have already been proposed [1,2,4,5], the use of multiple methods with various models will provide useful information. This is a future direction that we should explore in order to make further progress in mathematical medicine.

REFERENCES

1. Ideta, A. M., Tanaka, G., Takeuchi, T. and Aihara, K. (2008) A mathematical model of intermittent androgen suppression for prostate cancer. J. Nonlinear Sci., 18, 593–614

2. Shimada, T. and Aihara, K. (2008) A nonlinear model with competition between prostate tumor cells and its application to intermittent androgen suppression therapy of prostate cancer. Math. Biosci., 214, 134–139

3. Hirata, Y., Bruchovsky, N. and Aihara, K. (2010) Development of a mathematical model that predicts the outcome of hormone therapy for prostate cancer. J. Theor. Biol., 264, 517–527

4. Jain, H. V., Clinton, S. K., Bhinder, A. and Friedman, A. (2011) Mathematical modeling of prostate cancer progression in response to androgen ablation therapy. Proc. Natl. Acad. Sci. USA, 108, 19701–19706

5. Portz, T., Kuang, Y. and Nagy, J. D. (2012) A clinical data validated mathematical model of prostate cancer growth under intermittent androgen suppression therapy. AIP Adv., 2, 011002.

6. Akakura, K., Bruchovsky, N., Goldenberg, S. L., Rennie, P. S., Buckley, A. R. and Sullivan, L. D. (1993) Effects of intermittent androgen suppression on androgen-dependent tumors. Apoptosis and serum prostate-specific antigen. Cancer, 71, 2782–2790

7. Bruchovsky, N., Klotz, L., Crook, J., Malone, S., Ludgate, C., Morris, W. J., Gleave, M. E. and Goldenberg, S. L. (2006) 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, 107, 389–395

8. Bruchovsky, N., Klotz, L., Crook, J. and Goldenberg, S. L. (2007) Locally advanced prostate cancer—biochemical results from a prospective phase II study of intermittent androgen suppression for men with evidence of prostate-specific antigen recurrence after radiotherapy. Cancer, 109, 858–867

9. Crook, J. M., O'Callaghan, C. J., Duncan, G., Dearnaley, D. P., Higano, C. S., Horwitz, E. M., Frymire, E., Malone, S., Chin, J., Nabid, A., *et al.* (2012) Intermittent androgen suppression for rising PSA level after radiotherapy. N. Engl. J. Med., 367, 895–903

10. Hussain, M., Tangen, C. M., Berry, D. L., Higano, C. S., Crawford, E. D., Liu, G., Wilding, G., Prescott, S., Kanaga Sundaram, S., Small, E. J., *et al.* (2013) Intermittent versus continuous androgen deprivation in prostate cancer. N. Engl. J. Med., 368, 1314–1325

11. Morino, K., Hirata, Y., Tomioka, R., Kashima, H., Yamanishi, K., Hayashi, N., Egawa, S. and Aihara, K. (2015) Predicting disease progression from short biomarker series using expert advice algorithm. Sci. Rep., 5, 8953

12. Hirata, Y., Morino, K., Akakura, K., Higano, C. S., Bruchovsky, N., Gambol, T., Hall, S., Tanaka, G. and Aihara, K. (2015) Intermittent androgen suppression: Estimating parameters for individual patients based on initial PSA data in response to androgen deprivation therapy. PLoS One, 10, e0130372

13. Suzuki, T., Bruchovsky, N. and Aihara, K. (2010) Piecewise affine systems modelling for optimizing hormone therapy of prostate cancer. Philos. Trans. A Math. Phys. Eng. Sci., 368, 5045–5059

14. Hirata, Y., di Bernardo, M., Bruchovsky, N. and Aihara, K. (2010) Hybrid optimal scheduling for intermittent androgen suppression of prostate cancer. Chaos, 20, 045125

15. Hirata, Y., Azuma, S. and Aihara, K. (2014) Model predictive control for optimally scheduling intermittent androgen suppression of prostate cancer. Methods, 67, 278–281

16. Tanaka, G., Hirata, Y., Goldenberg, S. L., Bruchovsky, N. and Aihara, K. (2010) Mathematical modelling of prostate cancer growth and its application to hormone therapy. Philos. Trans. A Math. Phys. Eng. Sci., 368, 5029–5044

17. Kuramae, H., Hirata, Y., Bruchovsky, N., Aihara, K. and Suzuki, H. (2011) Nonlinear systems identification by combining regression with bootstrap resampling. Chaos, 21, 043121

18. Hirata, Y., Akakura, K., Higano, C. S., Bruchovsky, N. and Aihara, K. (2012) Quantitative mathematical modeling of PSA dynamics of prostate cancer patients treated with intermittent androgen suppression. J. Mol. Cell Biol., 4, 127–132

19. Guo, Q., Lu, Z., Hirata, Y. and Aihara, K. (2013) Parameter estimation and optimal scheduling algorithm for a mathematical model of intermittent androgen suppression therapy for prostate cancer. Chaos, 23, 043125

20. Suzuki, T. and Aihara, K. (2013) Nonlinear system identification for prostate cancer and optimality of intermittent androgen suppression therapy. Math. Biosci., 245, 40–48

21. Suzuki, Y., Sakai, D., Nomura, T., Hirata, Y. and Aihara, K. (2014) A new protocol for intermittent androgen suppression therapy of prostate cancer with unstable saddle-point dynamics. J. Theor. Biol., 350, 1–16

22. Tao, Y., Guo, Q. and Aihara, K. (2015) A partial differential equation model and its reduction to an ordinary differential equation model for prostate tumor growth under intermitten thormone therapy. J. Math. Biol., 69, 817–838

23. Cover, T. M. and Thomas, J. A. (1991) Elements of Information Theory. New York: Wiley Interscience Publication