Time-variant tumor growth trajectory models for *in silico* randomized controlled trials for patients with early-stage non-small cell lung cancer in optimizing stereotactic body radiation therapy

Authors

Kazuki MITSUSHIMA[1], Hidetaka ARIMURA[*2], Yuko SHIRAKAWA[3], Takumi KODAMA[1], Tadamasa YOSHITAKE[3]

[1] Department of Health Sciences, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
[2] Department of Health Sciences, Faculty of Medical Sciences, Kyushu University, Fukuoka, Japan
[3] Departments of Radiology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

***Corresponding author:**

Hidetaka Arimura, Ph.D.

Division of Medical Quantum Science, Department of Health Sciences, Faculty of Medical Sciences,

Kyushu University, 3-1-1, Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

Tell: +81-92-642-6719

E-mail: arimura.hidetaka.616@m.kyushu-u.ac.jp

Declarations

-Funding:

This study was supported by grants from JSPS KAKENHI (JP23KJ1733 and JP24K10840).

-Conflicts of interest/Competing interests (include appropriate disclosures):

Not applicable.

-Availability of data and material (data transparency):

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

-Code availability (software application or custom code):

This study used following applications;

・MATLAB, 2018a, https://jp.mathworks.com/products/matlab.html

・Python, 3.10.5., https://www.python.org/

Preprocessing, model construction, and optimization in this study were conducted using in-house functions. The code used in this study is not publicly available but can be obtained from the

corresponding author upon reasonable request.

-Authors' contributions (optional - please review the submission guidelines from the journal whether statements are mandatory):

Conceptualization: Kazuki Mitsushima, Hidetaka Arimura

Data curation: Kazuki Mitsushima, Hidetaka Arimura, Yuko Shirakawa, Tadamasa Yoshitake

Formal analysis: Kazuki Mitsushima, Hidetaka Arimura, Yuko Shirakawa, Takumi Kodama

Funding acquisition: Hidetaka Arimura, Takumi Kodama

Investigation: Kazuki Mitsushima, Hidetaka Arimura, Takumi Kodama

Methodology: Kazuki Mitsushima, Hidetaka Arimura, Takumi Kodama

Project administration: Hidetaka Arimura, Yuko Shirakawa

Resources: Kazuki Mitsushima, Hidetaka Arimura, Yuko Shirakawa, Tadamasa Yoshitake

Software: Kazuki Mitsushima, Takumi Kodama

Supervision: Hidetaka Arimura, Yuko Shirakawa

Validation: Kazuki Mitsushima, Hidetaka Arimura, Yuko Shirakawa, Takumi Kodama

Visualization: Kazuki Mitsushima, Hidetaka Arimura

Writing – original draft: Kazuki Mitsushima, Hidetaka Arimura

Writing – review & editing: Kazuki Mitsushima, Hidetaka Arimura, Yuko Shirakawa, Takumi Kodama, Tadamasa Yoshitake

-Ethics approval (include appropriate approvals or waivers):

This retrospective study was performed with ethical approval from the Institutional Review Boards of Kyushu University Hospital (approval number:22307-00).

-Consent to participate (include appropriate statements):

Given the retrospective study design, informed consent was obtained through an opt-out procedure, whereby patients were provided with the clear information and the opportunity to withdraw their data from the study.

-Consent for publication (include appropriate statements):

Given the retrospective study design, informed consent was obtained through an opt-out procedure, whereby patients were provided with the clear information and the opportunity to withdraw their data from the study.

Abstract

Purpose: Applying new treatments to real patients to verify therapeutic efficacy may induce various risks, such as critical adverse events. Additionally, there are ethical and financial issues in real-world randomized controlled trials (RCTs). This study aimed to develop mathematical models of time-variant tumor growth trajectories (TGTs) for *in silico* RCTs targeting patients with stage I non-small cell lung cancer (NSCLC) to optimize stereotactic body radiation therapy (SBRT).

Methods: The basic idea of the *in silico* RCT was to evaluate the endpoint of progression-free survival (PFS) curves for the two regimens derived from TGTs for virtual patient data produced via mathematical models. TGT models with a relative number of tumor cells were proposed by integrating the Bertalanffy-Pütter (BP) model and linear quadratic model into tumor growth models. To validate the proposed models, we performed three RCTs, 30 Gy/1 fraction (Fr) versus 60 Gy/3 Fr, 48 Gy/4 Fr versus 75 Gy/25 Fr, and 34 Gy/1 Fr versus 48 Gy/4 Fr.

Results: The three *in silico* RCTs showed no statistically significant differences in PFS curves, which was similar to the results of three previous studies.

Conclusions: The proposed mathematical models could be leveraged for *in silico* RCTs to optimize SBRT.

Keywords: lung cancer, tumor growth trajectory, *In silico* simulations, radiation therapy, differential equations

Time-variant tumor growth trajectory models for *in silico* randomized controlled trials for patients with early-stage non-small cell lung cancer in optimizing stereotactic body radiation therapy

Abstract

Purpose: Applying new treatments to real patients to verify therapeutic efficacy may induce various risks, such as critical adverse events. Additionally, there are ethical and financial issues in real-world randomized controlled trials (RCTs). This study aimed to develop mathematical models of time-variant tumor growth trajectories (TGTs) for *in silico* RCTs targeting patients with stage I non-small cell lung cancer (NSCLC) to optimize stereotactic body radiation therapy (SBRT).

Methods: The basic idea of the *in silico* RCT was to evaluate the endpoint of progression-free survival (PFS) curves for the two regimens derived from TGTs for virtual patient data produced via mathematical models. TGT models with a relative number of tumor cells were proposed by integrating the Bertalanffy-Pütter (BP) model and linear quadratic model into tumor growth models. To validate the proposed models, we performed three RCTs, 30 Gy/1 fraction (Fr) versus 60 Gy/3 Fr, 48 Gy/4 Fr versus 75 Gy/25 Fr, and 34 Gy/1 Fr versus 48 Gy/4 Fr.

Results: The three *in silico* RCTs showed no statistically significant differences in PFS curves, which was similar to the results of three previous studies.

Conclusions: The proposed mathematical models could be leveraged for *in silico* RCTs to optimize SBRT.

Keywords: lung cancer, tumor growth trajectory, *In silico* simulations, radiation therapy, differential equations

1. Introduction

Lung cancer, the second most frequently diagnosed cancer worldwide [1], is categorized into two primary types: small-cell lung cancer and non-small cell lung cancer (NSCLC). NSCLC accounts for approximately 85% of all lung cancer cases [2]. Currently, stereotactic body radiation therapy (SBRT) is recommended for patients who are inoperable or refuse surgery, especially elderly people in an aging society [3, 4]. Optimizing the fractionated dose and number of fractions in SBRT is crucial for improving tumor control rates and reducing complication rates in normal tissues of patients with NSCLC [5, 6]. For this purpose, clinical trials such as RTOG trials for SBRT have been conducted to determine the optimum conditions (e.g., fractional dose and number of fractions) [7-9]. However, many institutions must be involved in multi-center research and require long periods of time, and many research grants (financial resources) are required for real-world clinical trials [8, 9], especially randomized controlled trials (RCTs) [7, 10].

RCTs are considered the gold standard for generating high-quality clinical evidence [11], but personnel costs for patient enrollment, randomization, follow-up, data management, monitoring, and statistical analysis, should be secured [12]. In multi-center trials, additional efforts and expenses are needed for coordination and data harmonization across institutions. For

instance, industry-sponsored trials have reported average per-participant costs of approximately $36,500 [12]. In addition, applying new treatments to real patients to verify therapeutic efficacy may induce various risks, such as critical adverse events. These safety concerns, combined with ethical challenges specific to real-world RCTs, make their execution even more demanding [10].

In silico RCTs have emerged as a promising alternative to traditional clinical trials due to their cost-effectiveness, scalability, and ethical advantages [13, 14]. These trials employ computer simulations with virtual patients to evaluate the efficacy and safety of medical interventions. They eliminate the need for enrolling human participants, thereby reducing the risk of adverse events and the burden of ethical and regulatory approvals. Moreover, *in silico* RCTs significantly reduce time and cost compared to conventional trials. The VICTRE study demonstrated that an *in silico* trial could be completed in approximately one-third the time and at a fraction of the cost of a comparable human trial, without compromising data quality [13]. Besides, *in silico* trials offer flexibility in design and allow for rapid hypothesis testing across diverse patient populations, which is particularly valuable for rare diseases or personalized medicine [14]. Therefore, virtual or *in silico* RCTs are the key to overcoming these issues. However, there are currently no mathematical models available for performing *in silico* RCTs in patients with stage I NSCLC to determine the optimum regimens of SBRT.

Therefore, this study aimed to develop mathematical models of time-variant tumor growth trajectories (TGTs) for *in silico* RCTs targeting patients with stage I NSCLC to optimize SBRT. TGT models can help us calculate progression-free survival (PFS) curves for *in silico* RCTs. The TGT models of this study focus on *in silico* RCTs for patients with stage I NSCLC for optimization of fractionated dose, the number of fractions, irradiation interval in SBRT.

2. Materials and methods

Figure 1 illustrates the overall workflow of this study. The tumor contours were delineated by radiation oncologists for calculation of tumor volumes. The number of tumor cells in each tumor was calculated as a reference value by multiplying the tumor volume by tumor cell density. TGT models (BP-based and Gompertz-based models) were constructed by optimizing the parameters of the models. The TGT models were evaluated with mean absolute percentage error (MAPE) and Spearman's correlation coefficient (SCC). *In silico* RCTs were simulated according to actual RCTs and a retrospective study with different fractionated doses and numbers of fractions.

Figure 1 Overall workflow of this study (A) and tumor growth trajectory (TGT) models

for the three types of tumor cells (B).

2.1 Patient data

This retrospective study was performed with ethical approval from the Institutional Review Boards of our university hospital (approval number:22307-00). Given the retrospective study design, informed consent was obtained through an opt-out procedure, whereby patients were provided with the clear information and the opportunity to withdraw their data from the study. From 2004 to 2016, 17 patients were selected from 131 patients with stage I NSCLC who underwent SBRT according to the two criteria: (1) the availability of pretreatment planning CT images and at least two follow-up CT images, where lung tumors were clearly visible; (2) the SBRT for the patients was the first irradiation administered to the lung field. 4D CT imaging was not employed for the patients. The contours of the largest primary gross tumor volume (GTV) on pretreatment planning CT images were delineated as the regions of interest (ROI) in a treatment planning system (Eclipse version 15.1, Varian Medical System Inc., Palo Alto, CA, USA) based on a consensus between two radiation oncologists with 18 and 23 years of clinical experiences, respectively. GTV contours on follow-up CT images were delineated by a medical physicist with 21 years of experiences and confirmed by the radiation oncologist (23 years of clinical

experiences). These 17 patients were then divided into seven patients for the training dataset and ten patients for the test dataset. Table S1 (Supplementary Tables) presents the demographic and clinical characteristics of the patients included in this study. All 17 patients received a prescribed total dose of 48 Gy in four fractions. Table S2 (Supplementary Tables) provides information on age, sex, lung cancer stage classification, histological classification, tumor appearance, recurrence patterns, and the number of follow-up CT images included in this study. The CT images used in this study were acquired using the following CT systems (Aquilion and Xvigor Laudator, Toshiba, Japan; Mx8000, Brilliance iCT and Brilliance 64, Philips, Netherlands; SOMATOM Volume Zoom, Siemens, Germany). From the 17 patients, 64 ROIs of primary tumors were extracted from 17 planning CT images and 47 follow-up CT images to calculate the reference number of tumor cells. The ROI with in-plane pixel sizes of 0.390 - 0.977 mm and slice thicknesses of 2.0 - 3.0 mm for the planning CT images were converted to binary images with an isovoxel size of 0.977 mm using a shape-based interpolation method [15]. The ROI with in-plane pixel sizes of 0.390 - 0.625 and slice thicknesses of 2.0 - 5.0 mm for the follow-up CT images were transformed in the same way as the planning CT images. This isotropic voxelization was performed using MATLAB 2018a.

2.2 Reference numbers of tumor cells obtained from CT images

The number of tumor cells in each tumor was calculated as a reference value by multiplying the tumor volume by tumor cell density. Tumor volume was calculated from the contours of the tumor region on the slices of a three-dimensional CT image. The tumor lesions on the CT images were displayed with a window width of 400 and level of 50 under mediastinal settings, and with a window width of 1500 and level of -600 under lung settings [16]. Finally, the reference total number of tumor cells, $N_{Ref}(t)$, was calculated by:

$$N_{Ref}(t) = n_V V d, \quad (1)$$

where n_V is the number of voxels in the ROI, V is the volume per voxel (cm^3), and d is the tumor cell density ($/cm^3$). Tumor cell densities of 8.56×10^7 ($/cm^3$), 1.08×10^8 ($/cm^3$), and 9.68×10^7 ($/cm^3$) were applied to patients with LUAD, LUSC, and other histological types, respectively, as identified from bronchoscopic biopsy [17]. The reference numbers of tumor cells, N_{Ref}, were normalized by the number of tumor cells obtained at the time of treatment planning CT acquisition for each patient. The relative numbers of tumor cells at all time points were utilized for constructing the TGT model.

2.3 TGT models

We developed two TGT models based on the Bertalanffy-Pütter (BP) and Gompertz models by integrating the linear quadratic model [18] into both models. Figure S1 (Supplementary Figures) illustrates the overall workflow of the TGT generation. The lower panel of Figure 1B illustrates the conceptual diagram for tumor growth, consisting of three types of cells and tumor cells killed during SBRT. We hypothesized that tumor cells in patients with NSCLC are composed of three populations: sensitive cells, resistant cells, and persister cells to radiation, denoted by subscripts S, R, and P, respectively [19]. The resistant tumor cells are intrinsically resistant to radiation therapy, and they are capable of continuing the proliferation, which would be independent of radiation exposure [20]. Since we assume that the resistant tumor cells could have the property of resistance as well as response to radiation, the increasing and decreasing rates (u_R and v_R) were set in Eqs. (3) and (6) of Supplementary Document. Persister tumor cells could stay at the dormant state, but they may subsequently resume the proliferation by transitioning to resistant cells with an alteration probability due to irradiation [20]. Thus, the number of tumor cells $N_{Total}(t)$ at a time of t in NSCLC patients is calculated as the sum of the sensitive cells $N_S(t)$, resistant cells $N_R(t)$, and persister cells $N_P(t)$:

$$N_{Total}(t) = N_S(t) + N_R(t) + N_P(t). \quad (2)$$

The initial numbers of tumor cells at the time of treatment planning CT image, $N_S(initial)$,

$N_R(initial)$, and $N_P(initial)$, were defined as follows:

$$N_S(initial) = (1 - a - b)N_{Total}(initial), \quad (3)$$

$$N_R(initial) = aN_{Total}(initial), \quad (4)$$

and

$$N_P(initial) = bN_{Total}(initial), \quad (5)$$

where a and b represent the initial proportions of resistant and persister cells, respectively. "$Sensitive\ cell\ ratio\ (SCR)$" represents the predicted proportion of N_S within N_{Total} and was calculated using the following equation:

$$SCR = (1 - a - b). \quad (6)$$

In the TGT model, $N_{Total}(t)$, $N_S(t)$, $N_R(t)$, and $N_P(t)$ were always expressed as relative tumor cell counts, obtained by dividing by their respective initial values $N_{Total}(initial)$, $N_S(initial)$, $N_R(initial)$, and $N_P(initial)$. The details of the BP-based- and Gompertz-based models are described in Supplementary Document.

2.4 Optimization of parameters for prediction models

Twelve parameters in the BP-based models and nine parameters in the Gompertz-based models were optimized using a simulated annealing method [21]. Table 1 summarizes the parameter descriptions of the BP-based and Gompertz-based models, $(u_S, v_S, u_R, v_R, u_P, v_P, \mu, a, b, n, \alpha, \alpha\beta)$ in BP-based models and $(\lambda_S, \lambda_R, \lambda_P, \mu, a, b, n, \alpha, \alpha\beta)$ in Gompertz-based models. The simulated annealing algorithm can search for a global minimum of the energy [the objective function of Eq. (7) in this study] based on a stochastic exploration of the multidimensional space of parameters by sampling neighboring candidates of parameter sets. The transition from a current candidate to the next candidate was determined by the Metropolis procedure [21], where a probabilistic acceptance rule that permits transitions to worse solutions with a probability governed by the Boltzmann distribution. The acceptance probability decreases as the temperature is gradually lowered according to a predefined annealing schedule. This mechanism facilitates the escape from local minima by allowing uphill moves, thereby enhancing the algorithm's ability to converge toward the global minimum of the objective function.

The MAPE in the relative total number of tumor cells between the reference numbers at all CT examinations and the numbers predicted with TGT models at the same time points was minimized for each patient as an objective function for the optimization of the parameters in the simulated annealing method. The MAPE is defined as:

$$MAPE(\%) = \frac{1}{n}\sum_{i=1}^{n}\frac{|N_{Total}(i) - N_{Ref}(i)|}{N_{Ref}(i)} \times 100, \quad (7)$$

where n is the total number of CT examinations, i is the CT examination number, $N_{Total}(i)$ is the relative total number of tumor cells predicted by the TGT models for each patient at a CT examination i, and $N_{Ref}(i)$ is the reference value representing the relative total number of tumor cells obtained from the pretreatment or follow-up CT images at the CT examination i. The parameters were initially set using randomly generated values within an arbitrary range. The ranges of 12 parameters (minimum and maximum values) for the test patients were obtained from the training patients (Table S1). The initial temperature, minimum temperature, cooling rate, and number of iterations for each case in the simulated annealing method were 10 000, 0.0001, 0.9, and 100, respectively. This process was performed using Python 3.10.5 on a Windows 10 operating system. The following libraries were used: NumPy (v1.23.1), pandas (v1.5.3), SciPy (v1.9.3), scikit-learn (v1.2.1), matplotlib (v3.6.2), seaborn (v0.12.2), OpenCV (v4.7.0.68), SimpleITK (v2.1.1.2), lifelines (v0.28.0), pydicom (v3.0.1), and pyradiomics (v3.0.1).

Table 1 Parameter descriptions of Bertalanffy-Pütter-based and Gompertz-based models.

2.5 Evaluation of TGT models

The TGT models were evaluated with MAPE and SCC. SCC evaluates the correlation between the reference numbers of tumor cells and the predicted numbers in the TGT for each patient. The metrics of the BP-based models were compared with those of the Gompertz-based models for training patients. The best models that achieved the lowest average MAPE and highest SCC were applied to the test patients for the evaluation of robustness. Finally, *in silico* simulations of RCT were performed using the best models.

2.6 *In silico* RCT

Three *in silico* RCTs were simulated according to two actual RCTs [5, 22] and a retrospective study comparing two regimens [6] with different fractionated doses and numbers of fractions. The basic idea of the *in silico* RCT was to evaluate the endpoint of PFS curves for the two regimens, which were derived by calculating the time to the event in TGTs of the number of tumor cells with virtual patient data produced via the mathematical models. Figure 2 illustrates

an overall workflow of *in silico* RCTs used in this study. We conducted three RCTs, RCT I: 30 Gy/1 Fr (n=65) versus 60 Gy/3 Fr (n=94) [5], RCT II: 48 Gy/4 Fr (n=56) versus 75 Gy/25 Fr (n=103) [6], RCT III: 34 Gy/1 Fr (n=39) versus 48 Gy/4 Fr (n=45) [22]. These trials were selected from RCTs targeting SBRT, in which Kaplan-Meier (KM) curves of PFS were employed to evaluate clinical outcomes. The inclusion criteria for selecting these RCTs were the availability of KM data of PFS, SBRT treatment regimens, the number of participants more than 30 (because 50 patients were employed in this study), and sufficient reports of endpoints necessary for TGT model validation. Table S3 (Supplementary Tables) shows details of the irradiation schedules used in the *in silico* RCTs. As for the regimen of 75 Gy/25, a cycle of 5 consecutive days of irradiation followed by 2 days of rest was assumed to be repeated. It was assumed that other irradiations were performed daily, because there were no statistically significant differences in PFS between daily and every other day ($p > 0.05$).

The 10 test patients were randomly assigned to Group A and Group B with five patients for each group. Twelve parameters of the TGT models were randomly generated nine times for each test patient within the ranges defined by the minimum and maximum values for the optimized parameters obtained from the training patients. Forty-five TGTs were constructed from 45 virtual patient data produced from 5 test patient. In total, each group had 50 patient data, which was composed of 5 real patient data and 45 virtual patient data.

PFSs for three *in silico* RCTs were obtained for the 50 test patient data by calculating the time to progression (TTP) from TGTs derived from the BP-based model. Progression was defined as a 20% increase from the minimum total number of tumor cells predicted by the model after irradiation according to the response evaluation criteria in solid tumors (RECIST) guideline [23]. The day to reach the threshold of 20% in a TGT was defined as TTP, but before five years. If the TTPs were longer than five years, the patients were considered censored. Statistically significant differences between the two regimens were evaluated using the p-value of the log-rank test, which was less than 0.05, indicating statistical significance. This process was performed using Python 3.10.5.

Figure 2 Overall workflow of *in silico* randomized controlled trials (RCTs) used in this study.

3. Results

Figure 3 shows the TGTs predicted by BP-based models for four patients (srt009, srt010, srt016, and srt031) in the training patients. For reference, Figure S2 shows the TGTs predicted by the Gompertz-based models for the same four patients as shown in Fig. 3. The reference tumor

cell counts derived from CT images are represented by blue circles. The total tumor cell count predicted by the tumor growth model, which is the sum of three types of cells (sensitive tumor cells [red line], resistant tumor cells [yellow line], and persister tumor cells [blue line]), is shown as a green line. The case ID is shown in the upper left corner of each graph, while the MAPE, SCC, and the p-value of the SCC are indicated in the upper right corner. These figures are further discussed in the following section.

Table 2 shows the MAPE and SCC in the relative total number of tumor cells between the reference values obtained from CT images at each time point and the numbers of TGTs predicted by the BP-based and Gompertz-based models at the same time points for training patients. In the BP-based models, the MAPE and SCC values were 5.51% and 0.947, respectively. In contrast, the corresponding values for the Gompertz-based models were 26.5% and 0.733, respectively. Tables S4 and S5 (Supplementary Tables) show optimized parameters and SCR for BP-based and Gompertz-based model, respectively. Therefore, BP-based models have greater potential to predict TGTs in tumor cells than Gompertz-based models.

Figure 3 Time variant tumor growth trajectories (TGTs) derived from the Bertalanffy-Pütter-based models for the training patients. 'Predicted' tumor cells were estimated by TGT models.

Table 2 Mean absolute percentage errors (MAPEs) and Spearman's correlation coefficients (SCCs) in the relative total number of tumor cells between the reference numbers at all CT examinations and numbers of TGTs predicted with Bertalanffy-Pütter- and Gompertz-based models at the same time points as the CT scans for the training patients.

Table 3 shows the MAPE and SCC in the relative total number of tumor cells between the reference numbers at all CT examinations and numbers of TGTs predicted by the BP-based models at the same time point for test patients. The MAPE and SCC values in the test patients were 2.89% and 0.960, respectively, demonstrating excellent performance with MAPE values below 5%. Table S6 (Supplement Tables) shows optimized parameters and SCR for BP-based model for test patients. Figure 4 shows the TGTs of four patients included in the test patients (srt067, srt068, srt070, and srt071) as predicted by the BP-based models. This figure will be discussed further in the next section. Figure S3 (Supplementary Figures) shows the remaining six test patients (srt058, srt063, srt081, srt090, srt105, and srt129) as predicted by the BP-based models.

Figure 4 Time variant tumor growth trajectories (TGTs) derived from the Bertalanffy-Pütter-based models for the test patients. 'Predicted' tumor cells were estimated by TGT models.

Table 3 Mean absolute percentage errors (MAPEs) and Spearman's correlation coefficients (SCCs) in the relative total number of tumor cells between the reference numbers at all CT examinations and numbers of TGTs predicted with Bertalanffy-Pütter-based models at the same time points as the CT scans for the test patients.

Figure 5 shows the PFS for three *in silico* RCTs derived from the BP-based model for the test patients. We performed three RCTs: 30 Gy/1 fraction (Fr) versus 60 Gy/3 Fr, 48 Gy/4 Fr versus 75 Gy/25 Fr, and 34 Gy/1 Fr versus 48 Gy/4 Fr. The three *in silico* RCTs showed no statistically significant differences in PFS curves [30 Gy/1 Fr versus 60 Gy/3 Fr (p-value = 0.3863), 48 Gy/4 Fr versus 75 Gy/25 Fr (p-value = 0.6159), 34 Gy/1 Fr versus 48 Gy/4 Fr (p-value = 0.4042)], which were similar to three past studies [5, 6, 22].

Figure 5 Progression-free survivals for *in silico* randomized controlled trials derived from the Bertalanffy-Pütter-based model for the test patients.

4. Discussion

In summary, the MAPE and SCC in the BP-based models for training patients were 5.51% and 0.947, respectively, which were superior to those of the Gompertz-based models. Furthermore, the BP-based model demonstrated a MAPE of 2.89% and an SCC of 0.960 for the test patients. The three *in silico* RCTs indicated no statistically significant differences in PFS curves derived from TGT models, which were similar to past studies.

The scientific significance of our findings was to have shown that PFS curves can be predicted from time-variant tumor growth trajectories of three types of sensitive, persister, and resistant tumor cells obtained from the proposed mathematical models. The clinical implication of the proposed models was that *in silico* RCTs can be performed by changing irradiation parameters (fractionated dose, number of fractions, irradiation interval) prior to clinical trials of new regimens in the real world to search for optimum irradiation conditions. Once irradiation parameters are narrowed down a few feasible candidates in *in silico* RCTs, real world RCTs can be effectively conducted based the candidates. As validation tests, we performed three *in silico*

RCTs according to actual SBRT regiments [5, 6, 22] with the endpoint of PFS curves. As the results, we found that there were no significant differences between simulated PFS curves, which were similar to the results of three previous real world data. However, we should perform various other *in silico* RCTs including multicenter RCTs, IMRT or VMAT and chemotherapy regimens. In addition, we should resolve several limitations such as the increase of the number of training patients, performing external tests using patient data from different institutions, and inclusion of other irradiation parameters (e.g. dose rate [24]) (see the last paragraph of this section).

From Figs. 3 and 4, the behaviors of total numbers of tumor cells (references) occurred during the radiation therapy can be described by predicting those of the three types of tumor cells for each patient. The number of sensitive tumor cells rapidly decreased with treatment and follow-up time in all patients (Figs. 3, 4, S2, and S3). As for the patient of srt010 (left bottom, Fig. 3), the total number of tumor cells increased with time after approximately one month. We can explain the phenomena as follows: persister tumor cells slightly increased and transitioned to resistant cells with an alteration probability due to irradiation, and then resistant tumor cells rapidly increased with the time while keeping the proliferation even after the treatment. In the patient of srt071 (right bottom, Fig. 4), the number of resistant tumor cells decreased because of decreasing rate, but persister tumor cells increased with the time. Consequently, the total number of tumor cells increased with the time after approximately six months.

There was a correlation between SCR and radiation sensitivity across different subtypes. Table S7 (Supplementary Tables) shows the α values, α/β ratios, and SCRs obtained from the BP-based and Gompertz-based models for all patients with two different histological tumor types. The table indicates that all patients with LUSC had a higher average value of SCR than those with LUAC. Previous studies have shown that LUSC is more radiosensitive than LUAC [25]. According to the report of Tai A et al., the optimized parameters included in the regression model for analyzing tumor control probability after SBRT for NSCLC were in an agreement with single-institution data, with the extracted radiobiological values being α = 0.010 ± 0.001 Gy^{-1} and α/β = 21.5 ± 1.0 Gy. On the other hand, the values extracted from multi-institution data were α = 0.072 ± 0.006 Gy^{-1} and α/β = 15.9 ± 1.0 Gy [26]. Based on the results in Table S7, although the α values differed from Tai A et al.'s findings, the α/β values derived from both models in the training patients were similar to those for the single-institution data by Tai A et al.

In future studies, we will integrate our TGT models for radiation therapy and drug effects [19] into new models for the combination of radiation therapy with conventional chemotherapy, molecularly targeted drug therapy [19]], and immunotherapy [27-29]. We should revise the tumor types (sensitive, resistant, and persister tumor cells) by considering the interactions of tumor cells with drugs.

This study has several limitations. First, the number of patients chosen for this study was limited and the data were provided from a single institution. Therefore, more patients should be collected from different institutions to evaluate the robustness of the prediction models. Second, we did not perform external tests using patient data from different institutions for multicenter RCTs, which should be done in future works. Third, we did not evaluate the tumor volumes during SBRT using cone-beam CT. If we deployed the volumes at four fractions, we may develop more accurate models to be performed in a future work. Fourth, only the tumor region with the largest diameter was targeted for each patient. According to the RECIST guideline [23], the entire tumor burden including all tumors must be considered when calculating the number of tumor cells. Fifth, we did not consider other irradiation parameters such as dose rate [24]. Finally, we did not take into account new irradiation regimens such as SBRT with and without intensity modulated radiotherapy (IMRT) or volumetric-modulated arc therapy (VMAT) [30]. Therefore, we will explore novel TGT models by including other irradiation parameters and the effects of treatment approaches in our TGT models for future works.

5. Conclusions

We have developed TGT models for *in silico* RCTs in optimizing SBRT. In three *in silico* RCTs,

there were no statistically significant differences in PFS curves derived from TGT models, which were similar to results of past studies. The proposed mathematical models could be leveraged for *in silico* RCTs to optimize SBRT.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209-249. doi: 10.3322/caac.21660.
2. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clin Proc. 2008 May;83(5):584-94. doi: 10.4065/83.5.584.
3. Borghetti P, Costantino G, Santoro V, Mataj E, Singh N, Vitali P, Greco D, Volpi G, Sepulcri M, Guida C, Tomasi C, Buglione M, Nardone V. Artificial Intelligence-suggested Predictive Model of Survival in Patients Treated With Stereotactic Radiotherapy for Early Lung Cancer. In Vivo. 2024 May-Jun;38(3):1359-1366. doi: 10.21873/invivo.13576.
4. Dalwadi SM, Szeja SS, Bernicker EH, Butler EB, Teh BS, Farach AM. Practice Patterns

and Outcomes in Elderly Stage I Non-Small-cell Lung Cancer: A 2004 to 2012 SEER Analysis. Clin Lung Cancer. 2018 Mar;19(2):e269-e276. doi: 10.1016/j.cllc.2017.11.004.

5. Ma SJ, Serra LM, Syed YA, Hermann GM, Gomez-Suescun JA, Singh AK. Comparison of Single- and Three-fraction Schedules of Stereotactic Body Radiation Therapy for Peripheral Early-stage Non-Small-cell Lung Cancer. Clin Lung Cancer. 2018 Mar;19(2):e235-e240. doi: 10.1016/j.cllc.2017.10.010.

6. Karasawa K, Hayakawa S, Machitori Y, Shibata Y, Ogawa H, Ito K, Shimizuguchi T, Kawamoto T, Nihei K. Accelerated Hypofractionated Radiotherapy Versus Stereotactic Body Radiotherapy for the Treatment of Stage I Nonsmall Cell Lung Cancer-A Single Institution Experience With Long-Term Follow-Up. Technol Cancer Res Treat. 2018 Jan 1;17:1533033818806318. doi: 10.1177/1533033818806318.

7. Videtic GM, Hu C, Singh AK, Chang JY, Parker W, Olivier KR, Schild SE, Komaki R, Urbanic JJ, Timmerman RD, Choy H. A Randomized Phase 2 Study Comparing 2 Stereotactic Body Radiation Therapy Schedules for Medically Inoperable Patients With Stage I Peripheral Non-Small Cell Lung Cancer: NRG Oncology RTOG 0915 (NCCTG N0927). Int J Radiat Oncol Biol Phys. 2015 Nov 15;93(4):757-64. doi: 10.1016/j.ijrobp.2015.07.2260.

8. Timmerman RD, Paulus R, Pass HI, Gore EM, Edelman MJ, Galvin J, Straube WL, Nedzi

LA, McGarry RC, Robinson CG, Schiff PB, Chang G, Loo BW Jr, Bradley JD, Choy H. Stereotactic Body Radiation Therapy for Operable Early-Stage Lung Cancer: Findings From the NRG Oncology RTOG 0618 Trial. JAMA Oncol. 2018 Sep 1;4(9):1263-1266. https://doi.org/10.1001/jamaoncol.2018.1251

9. Bezjak A, Paulus R, Gaspar LE, Timmerman RD, Straube WL, Ryan WF, Garces YI, Pu AT, Singh AK, Videtic GM, McGarry RC, Iyengar P, Pantarotto JR, Urbanic JJ, Sun AY, Daly ME, Grills IS, Sperduto P, Normolle DP, Bradley JD, Choy H. Safety and Efficacy of a Five-Fraction Stereotactic Body Radiotherapy Schedule for Centrally Located Non-Small-Cell Lung Cancer: NRG Oncology/RTOG 0813 Trial. J Clin Oncol. 2019 May 20;37(15):1316-1325.

10. Goldstein CE, Weijer C, Brehaut JC, Fergusson DA, Grimshaw JM, Horn AR, Taljaard M. Ethical issues in pragmatic randomized controlled trials: a review of the recent literature identifies gaps in ethical argumentation. BMC Med Ethics. 2018 Feb 27;19(1):14.

11. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. BJOG. 2018 Dec;125(13):1716. doi: 10.1111/1471-0528.15199.

12. Lovegreen, O., Riggs, D., Staten, M.A. et al. Financial management of large, multi-center trials in a challenging funding milieu. Trials **19**, 267 (2018).

13. Badano, A. In silico imaging clinical trials: cheaper, faster, better, safer, and more scalable. *Trials* **22**, 64 (2021).

14. Wedlund L, Kvedar J. Simulated trials: in silico approach adds depth and nuance to the RCT gold-standard. NPJ Digit Med. 2021 Aug 11;4(1):121. doi: 10.1038/s41746-021-00492-7.

15. Gabor T. Herman, Jingsheng Zheng, and Carolyn A. Bucholtz. 1992. Shape-Based Interpolation. IEEE Comput. Graph. Appl. 12, 3 (May 1992), 69–79. doi: 10.1109/38.135915.

16. Ikushima K, Arimura H, Jin Z, Yabu-Uchi H, Kuwazuru J, Shioyama Y, Sasaki T, Honda H, Sasaki M. Computer-assisted framework for machine-learning-based delineation of GTV regions on datasets of planning CT and PET/CT images. J Radiat Res. 2017 Jan;58(1):123-134. doi: 10.1093/jrr/rrw082.

17. Katsurada N, Tachihara M, Jimbo N, Yamamoto M, Yoshioka J, Mimura C, Satoh H, Furukawa K, Otoshi T, Kiriu T, Yasuda Y, Tanaka T, Nagano T, Nishimura Y. Yield of tumor samples with a large guide-sheath in endobronchial ultrasound transbronchial biopsy for non-small cell lung cancer: A prospective study. PLoS One. 2021 Oct 29;16(10):e0259236. doi: 10.1371/journal.pone.0259236.

18. Geng C, Paganetti H, Grassberger C. Prediction of Treatment Response for Combined Chemo- and Radiation Therapy for Non-Small Cell Lung Cancer Patients Using a Bio-

Mathematical Model. Sci Rep. 2017 Oct 19;7(1):13542. doi: 10.1038/s41598-017-13646-z.

19. Grassberger C, McClatchy D 3rd, Geng C, Kamran SC, Fintelmann F, Maruvka YE, Piotrowska Z, Willers H, Sequist LV, Hata AN, Paganetti H. Patient-Specific Tumor Growth Trajectories Determine Persistent and Resistant Cancer Cell Populations during Treatment with Targeted Therapies. Cancer Res. 2019 Jul 15;79(14):3776-3788. doi: 10.1158/0008-5472.CAN-18-3652.

20. Zhao Y, Song Y, Zhao R, Zhao M, Huang Q. Gene Panel of Persister Cells as a Prognostic Indicator for Tumor Repopulation After Radiation. Front Oncol. 2020 Nov 20;10:607727. doi: 10.3389/fonc.2020.607727.

21. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by Simulated Annealing. Science 1983:220(4598);671-680. DOI:10.1126/science.220.4598.671

22. Videtic GM, Paulus R, Singh AK, Chang JY, Parker W, Olivier KR, Timmerman RD, Komaki RR, Urbanic JJ, Stephans KL, Yom SS, Robinson CG, Belani CP, Iyengar P, Ajlouni MI, Gopaul DD, Gomez Suescun JB, McGarry RC, Choy H, Bradley JD. Long-term Follow-up on NRG Oncology RTOG 0915 (NCCTG N0927): A Randomized Phase 2 Study Comparing 2 Stereotactic Body Radiation Therapy Schedules for Medically Inoperable Patients With Stage I Peripheral Non-Small Cell Lung Cancer. Int J Radiat Oncol Biol Phys. 2019 Apr 1;103(5):1077-1084. doi: 10.1016/j.ijrobp.2018.11.051.

23. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009 Jan;45(2):228-47. doi: 10.1016/j.ejca.2008.10.026.

24. Nakano H, Shiinoki T, Tanabe S, Utsunomiya S, Takizawa T, Kaidu M, Nishio T, Ishikawa H. Mathematical model combined with microdosimetric kinetic model for tumor volume calculation in stereotactic body radiation therapy. Sci Rep. 2023 Jul 6;13(1):10981. doi: 10.1038/s41598-023-38232-4.

25. Murad TM, August C. Radiation-induced atypia. A review. Diagn Cytopathol. 1985 Apr-Jun;1(2):137-52. doi: 10.1002/dc.2840010210.

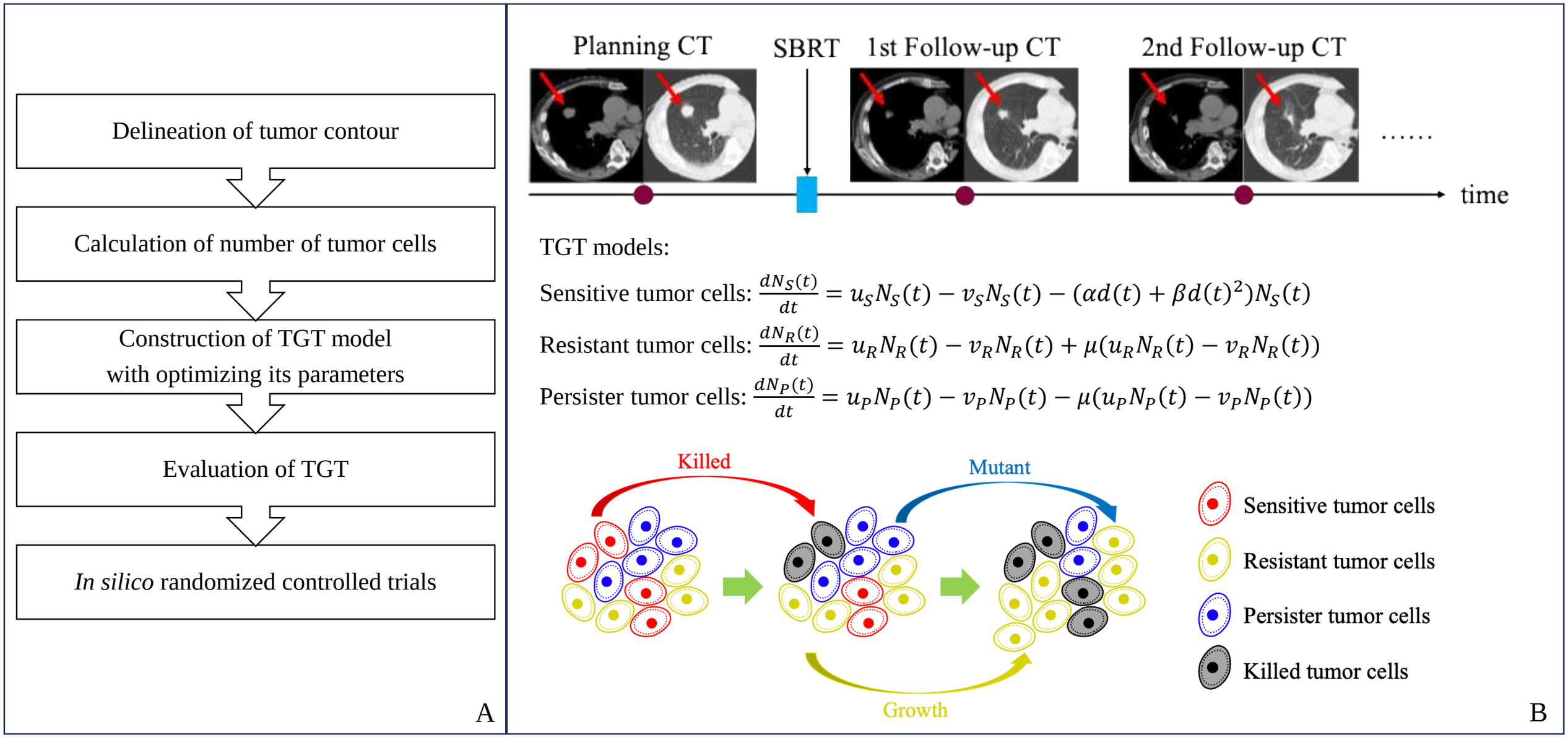
26. Tai A, Liu F, Gore E, Li XA. An analysis of tumor control probability of stereotactic body radiation therapy for lung cancer with a regrowth model. Phys Med Biol. 2016 May 21;61(10):3903-13. doi: 10.1088/0031-9155/61/10/3903.

27. Jenner AL, Cassidy T, Belaid K, Bourgeois-Daigneault MC, Craig M. In silico trials predict that combination strategies for enhancing vesicular stomatitis oncolytic virus are determined by tumor aggressivity. J Immunother Cancer. 2021 Feb;9(2):e001387. doi: 10.1136/jitc-2020-001387. Erratum in: J Immunother Cancer. 2021 Oct;9(10):e001387corr1. doi: 10.1136/jitc-2020-001387corr1.

28. Sung W, Hong TS, Poznansky MC, Paganetti H, Grassberger C. Mathematical Modeling to Simulate the Effect of Adding Radiation Therapy to Immunotherapy and Application to Hepatocellular Carcinoma. Int J Radiat Oncol Biol Phys. 2022 Mar 15;112(4):1055-1062. doi: 10.1016/j.ijrobp.2021.11.008.

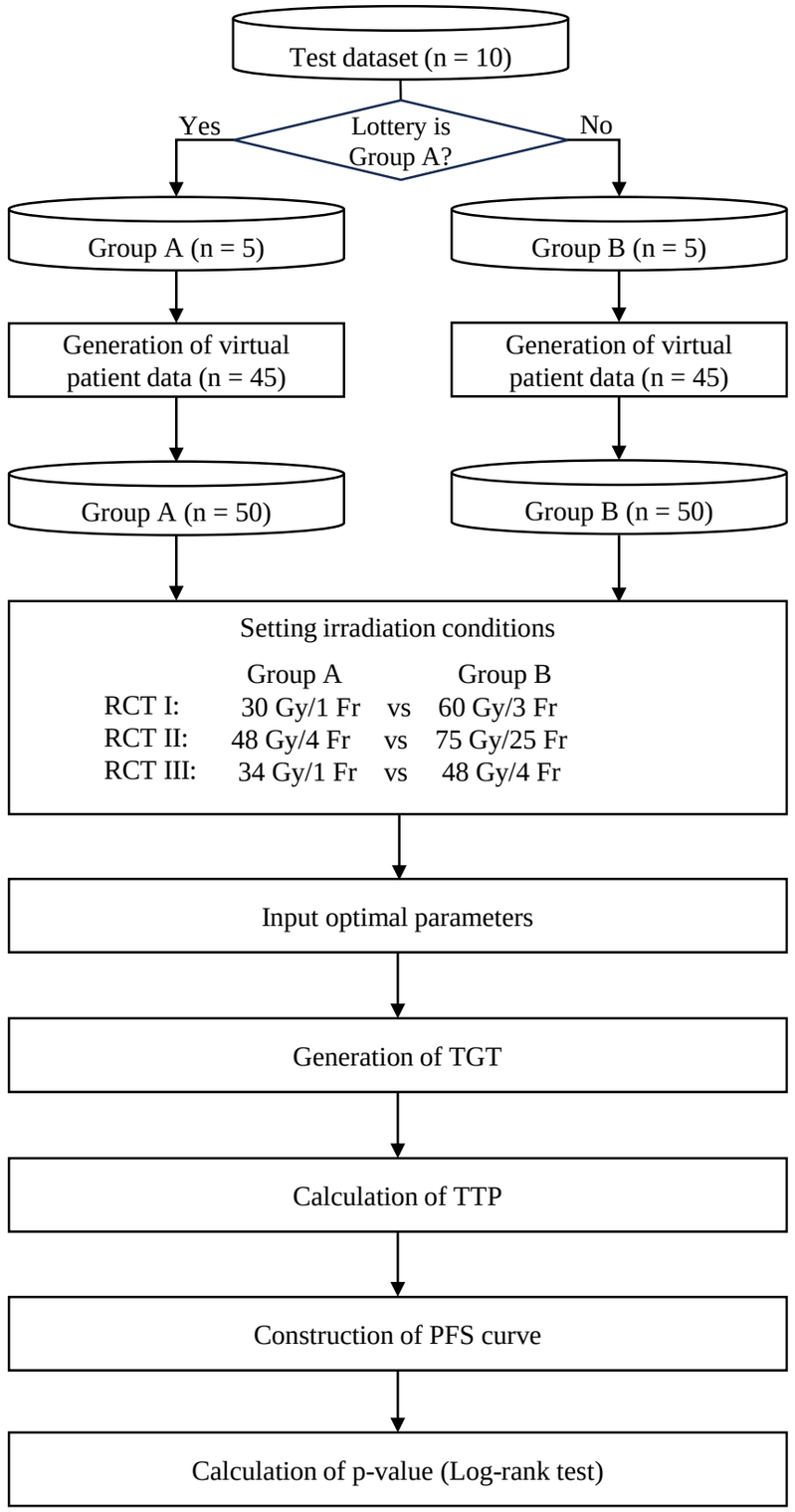
29. Creemers JHA, Ankan A, Roes KCB, Schröder G, Mehra N, Figdor CG, de Vries IJM, Textor J. In silico cancer immunotherapy trials uncover the consequences of therapy-specific response patterns for clinical trial design and outcome. Nat Commun. 2023 Apr 24;14(1):2348. doi: 10.1038/s41467-023-37933-8.

30. Wei Z, Peng X, He L, Wang J, Liu Z, Xiao J. Treatment plan comparison of volumetric-modulated arc therapy to intensity-modulated radiotherapy in lung stereotactic body radiotherapy using either 6- or 10-MV photon energies. J Appl Clin Med Phys. 2022 Aug;23(8):e13714. doi: 10.1002/acm2.13714.



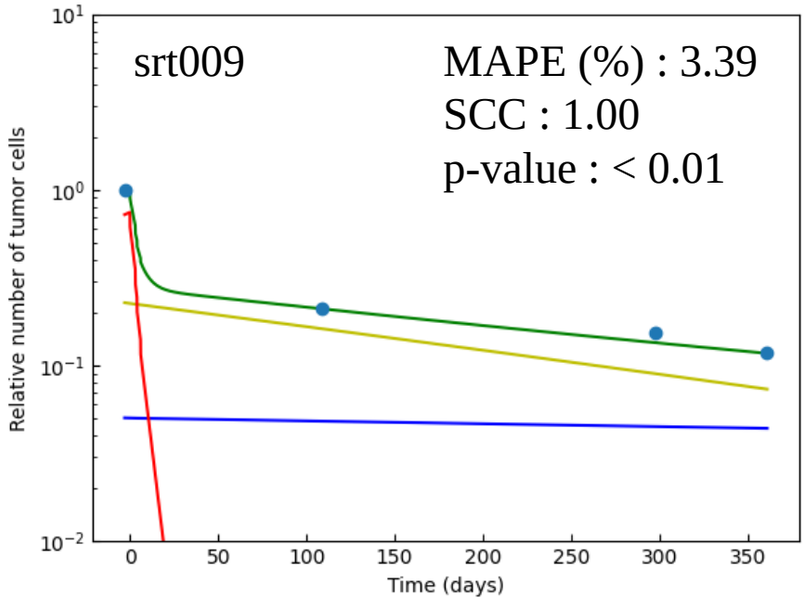
SBRT: Stereotactic body radiotherapy, CT: Computed tomography.

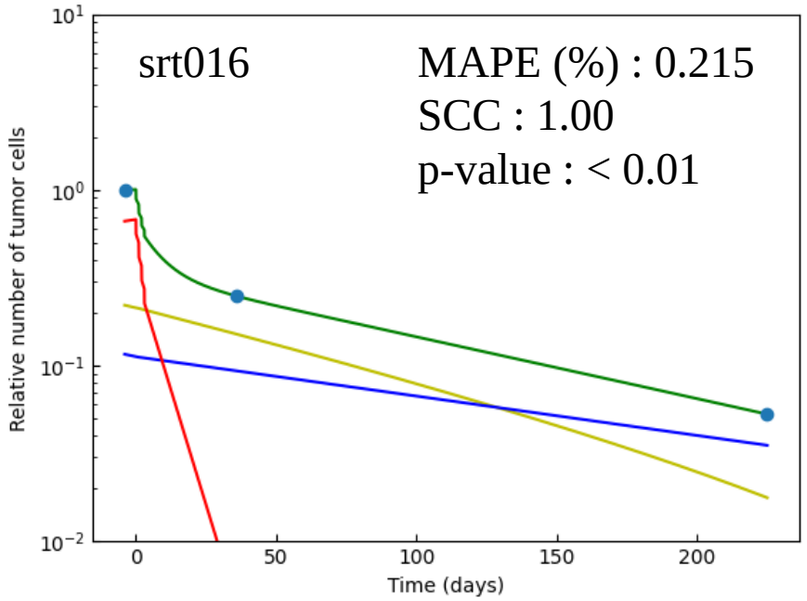
Figure 1 Overall workflow of this study (A) and tumor growth trajectory (TGT) models for three types of tumor cells (B).

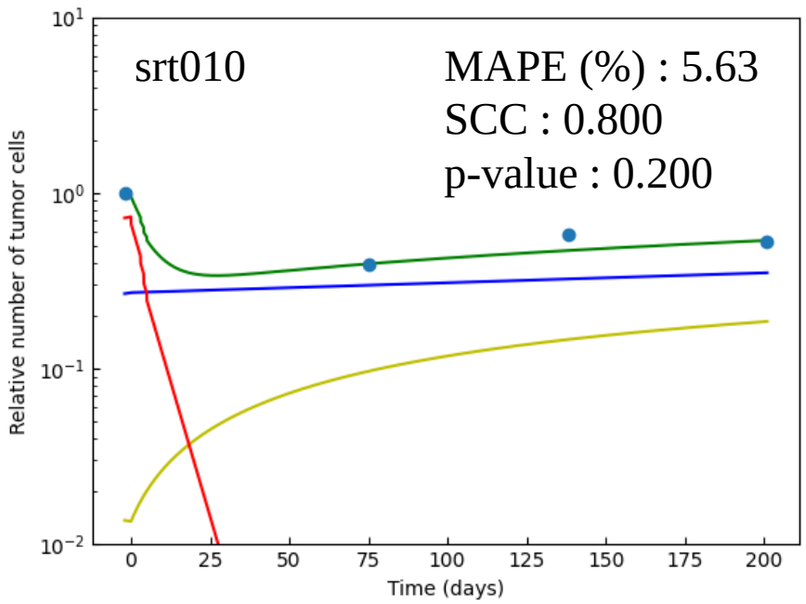


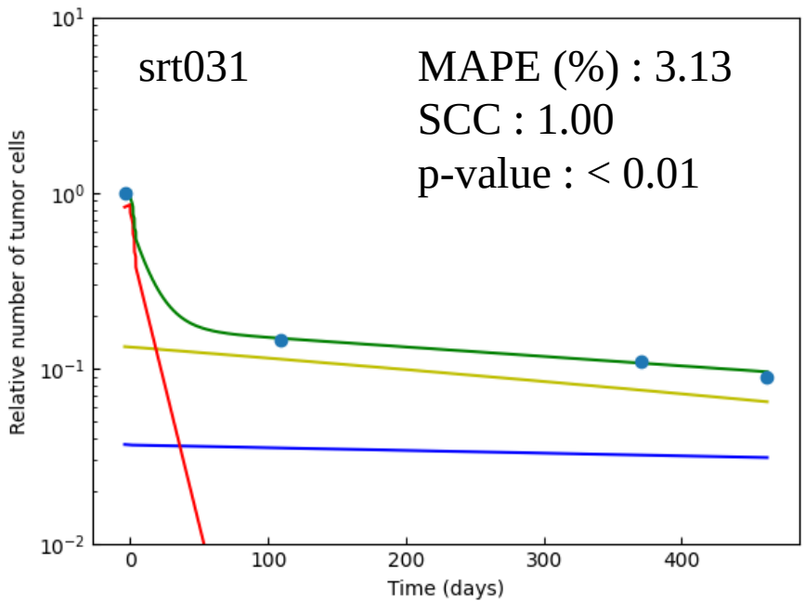
Fr: Fraction(s), TGT: Tumor growth trajectory, TTP: Time to progression, PFS: progression-free survival

Figure 2 Overall workflow of *in silico* randomized controlled trials (RCTs) used in this study.





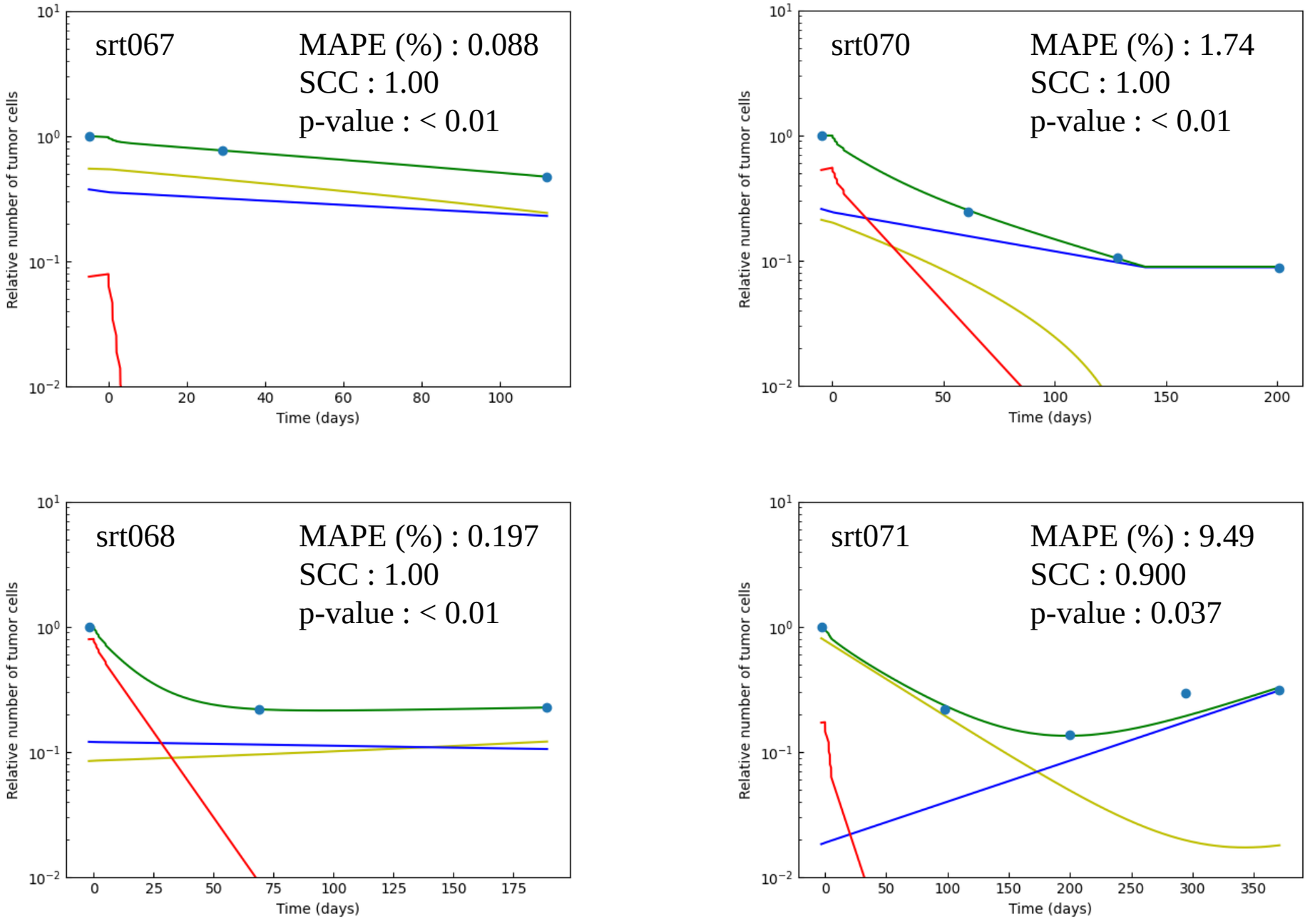




Predicted sensitive tumor cells: N_S
Predicted resistant tumor cells: N_R
Predicted persister tumor cells: N_P
Predicted total tumor cells : N_{Total}
Reference total number of tumor cells from CT image : N_{Ref}

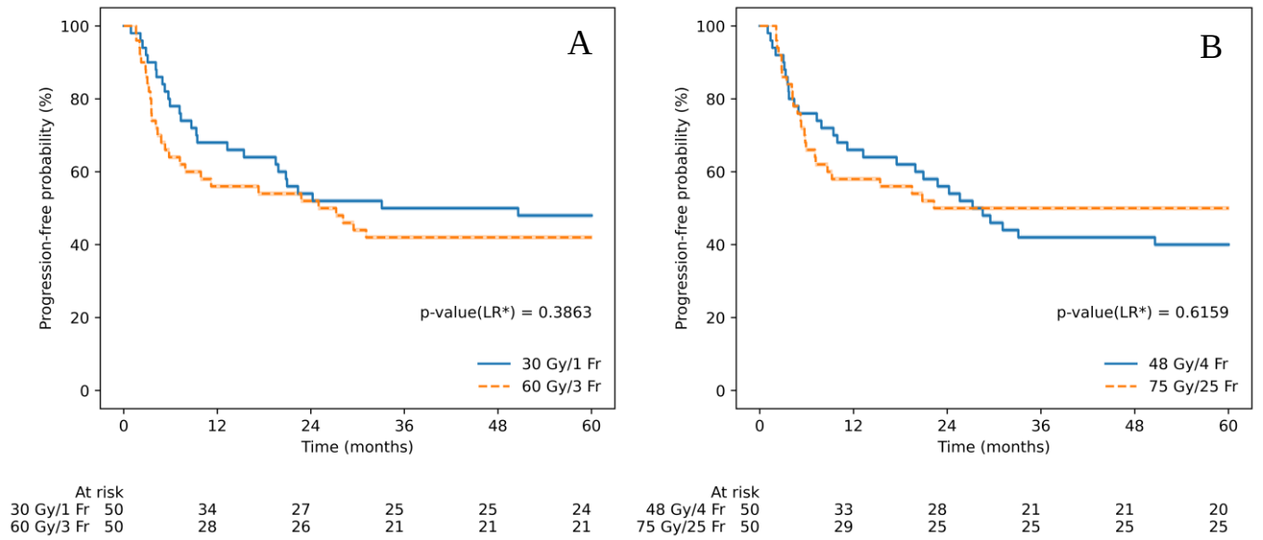
MAPE: Mean absolute percentage errors, SCC: Spearman's correlation coefficient.

Figure 3 Time variant tumor growth trajectories (TGTs) derived from the Bertalanffy-Pütter-based models for the training patients. 'Predicted' tumor cells were estimated by TGT models.



MAPE: Mean absolute percentage errors, SCC: Spearman's correlation coefficient.

Figure 4 Time variant tumor growth trajectories (TGTs) derived from the Bertalanffy-Pütter-based models for the test patients. 'Predicted' tumor cells were estimated by TGT models.



(A) 30 Gy/1 Fr vs 60 Gy/3 Fr

(B) 48 Gy/4 Fr vs 75 Gy/25 Fr

(C) 34 Gy/1 Fr vs 48 Gy/4 Fr

Fr: Fraction(s), LR: Log-rank test.

Figure 5 Progression-free survivals for *in silico* randomized controlled trials derived from the Bertalanffy-Pütter-based model for the test patients.

Table 1 Parameter descriptions of Bertalanffy-Pütter-based and Gompertz-based models.

Parameter	Description
N_{Total}	Total number of tumor cells
$N_S,\ N_R,\ N_P$	Numbers of sensitive, resistant, and persister tumor cells
a, b	Initial ratio of resistant and persister tumor cells
$u_S,\ u_R,\ u_P$	Increasing rate of sensitive, resistant, and persister tumor cells
v_S, v_R, v_P	Decreasing rate of sensitive, resistant, and persister tumor cells
$\lambda_S, \lambda_R, \lambda_P$	Growth rate of sensitive, resistant, and persister tumor cells
K	Carrying capacity
n	Cell loss parameter
μ	Mutant probability from persister cells to resistant cells
α	Growth rate of sensitive, resistant, and persister tumor cells
$\alpha\beta$	$\alpha\beta$-ratio
d	Fraction dose (Gy)

Table 2 Mean absolute percentage errors (MAPEs) and Spearman's correlation coefficients (SCCs) in the relative total number of tumor cells between the reference numbers at all CT examinations and numbers of TGTs predicted with Bertalanffy-Pütter- and Gompertz-based models at the same time points as the CT scans for the training patients.

	BP-based model			Gompertz-based model		
Patients	MAPE (%)	SCC	p-value (SCC)	MAPE (%)	SCC	p-value (SCC)
srt009	3.39	1.00	< 0.01	8.39	0.800	0.200
srt010	5.63	0.800	0.200	5.79	0.800	0.200
srt016	0.215	1.00	< 0.01	3.45	1.00	< 0.01
srt031	3.13	1.00	< 0.01	16.6	0.200	0.800
srt033	22.4	0.800	0.042	137	0.829	0.042
srt034	0.169	1.00	< 0.01	11.7	0.500	0.667
srt061	3.61	1.00	< 0.01	3.11	1.00	< 0.01
Mean	5.51	0.947	-	26.5	0.733	-

Table 3 Mean absolute percentage errors (MAPEs) and Spearman's correlation coefficients (SCCs) in the relative total number of tumor cells between the reference numbers at all CT examinations and numbers of TGTs predicted with Bertalanffy-Pütter-based models at the same time points as the CT scans for the test patients.

Patients	MAPE (%)	SCC	p-value (SCC)
srt058	0.860	1.00	< 0.01
srt063	10.1	0.821	0.089
srt067	0.088	1.00	< 0.01
srt068	0.197	1.00	< 0.01
srt070	1.74	1.00	< 0.01
srt071	9.49	0.900	0.037
srt081	0.284	0.866	0.333
srt090	0.224	1.00	< 0.01
srt105	5.81	1.00	< 0.01
srt129	0.114	1.00	< 0.01
Mean	2.89	0.960	-

Supplementary Figure_Revised

Supplementary Table_Revised clean