Classification of tolerable/intolerable mucosal toxicity of head-and-neck radiotherapy schedules with a biomathematical model of cell dynamics

Juan Pardo-Montero[a)]
Group of Medical Physics and Biomathematics, Instituto de Investigación Sanitaria de Santiago (IDIS), Santiago de Compostela, Spain
Department of Medical Physics, Complexo Hospitalario Universitario de Santiago de Compostela, Spain

Martín Parga-Pazos[*]
Group of Medical Physics and Biomathematics, Instituto de Investigación Sanitaria de Santiago (IDIS), Santiago de Compostela, Spain

John D. Fenwick
Department of Molecular and Clinical Cancer Medicine, Institute of Translational Medicine, University of Liverpool, Liverpool, UK
Department of Physics, Clatterbridge Cancer Centre, Clatterbridge Road, Wirral, UK

(Received 9 October 2020; revised 7 February 2021; accepted for publication 1 March 2021; published 9 July 2021)

Purpose: The purpose of this study is to present a biomathematical model based on the dynamics of cell populations to predict the tolerability/intolerability of mucosal toxicity in head-and-neck radiotherapy.

Methods and Materials: Our model is based on the dynamics of proliferative and functional cell populations in irradiated mucosa, and incorporates the three As: Accelerated proliferation, loss of Asymmetric proliferation, and Abortive divisions. The model consists of a set of delay differential equations, and tolerability is based on the depletion of functional cells during treatment. We calculate the sensitivity (sen) and specificity (spe) of the model in a dataset of 108 radiotherapy schedules, and compare the results with those obtained with three phenomenological classification models, two based on a biologically effective dose (BED) function describing the tolerability boundary (*Fowler* and *Fenwick*) and one based on an equivalent dose in 2 Gy fractions (EQD_2) boundary (*Strigari*). We also perform a machine learning-like cross-validation of all the models, splitting the database in two, one for training and one for validation.

Results: When fitting our model to the whole dataset, we obtain predictive values (sen + spe) up to 1.824. The predictive value of our model is very similar to that of the phenomenological models of *Fowler* (1.785), *Fenwick* (1.806), and *Strigari* (1.774). When performing a $k = 2$ cross-validation, the specificity and sensitivity in the validation dataset decrease for all models, from ~1.82 to ~1.55–1.63. For *Fowler*, the worsening is higher, down to 1.49.

Conclusions: Our model has proved useful to predict the tolerability/intolerability of a dataset of 108 schedules. As the model is more mechanistic than other available models, it could prove helpful when designing unconventional dose fractionations, schedules not covered by datasets to which phenomenological models of toxicity have been fitted. © *2021 The Authors. Medical Physics published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine* [https://doi.org/10.1002/mp.14834]

Key words: classification, head-and-neck, linear-quadratic model, mucositis, radiotherapy

1. INTRODUCTION

Many clinical trials have tested nonconventional schedules to improve tumor control probability in head-and-neck cancer, including acceleration, dose escalation, or gaps.[1] Intolerable mucosal toxicity is a limiting factor in dose escalation and/or treatment shortening in head-and-neck cancer.[2–4] Finding the boundary separating tolerable and intolerable schedules is of particular interest, especially when aiming at designing novel unconventional fractionation schedules.

Mathematical models have been employed to tackle this problem. Kaanders et al.[5] initially explored this issue, and by plotting the total dose vs the treatment time, they found that intolerable schedules tend to deliver larger doses over shorter times. Fowler et al.[6] used the linear-quadratic model to investigate the tolerable/intolerable boundary, and proposed a boundary between tolerable and intolerable schedules based on the biologically effective dose (BED). Fenwick et al.[7] developed that model using delay differential equations to obtain a novel BED vs treatment time formula that resulted in better separation of tolerable and intolerable schedules. Strigari et al.[8] tackled this problem by employing the LKB formulation of normal tissue complication probability (NTCP). With their model, they achieved values of sensitivity/specificity similar to Ref. [7] while at the same time predicting the level of severe mucositis NTCP of those schedules. More recently, machine learning techniques have been used to predict patientwise the probability of severe mucosal toxicity.[9–11]

All models mentioned above are phenomenological. Mechanistic models can add to our understanding of the processes involved in response to radiotherapy, and may prove especially useful when designing unconventional dose fractionations. In this work, we present a more mechanistic approach to the problem, based on the three A's of accelerated proliferation:[12] loss of asymmetry, accelerated proliferation, and abortive divisions. We rely on the compartment model of cell population dynamics in mucosa presented in Ref. [13], which has been extended to multiple fractions, including incomplete repair between fractions, and we link tolerability/intolerability of a given schedule to the numbers of cells in the mucosa. Systems of differential equations have previously been used to model dynamics of cell populations in normal tissues, including, for example, toxicity[14] and cancer induction,[15] tumor reoxygenation,[16] or response to molecular radiotherapy.[17,18] We validate our model in an extended dataset of tolerability/intolerability of head-and-neck schedules, building upon those provided in Refs. [7,8], and adding some additional schedules.

2. METHODS AND MATERIALS

2.A. Model of dynamics of cell populations

The model presented in Ref. [13] for single fraction irradiation has been extended to multiple fractions. We refer the reader to that publication for a detailed presentation and study of the model's description of cell dynamics in preclinical data of cell. We initially assume two compartments of cells: proliferative cells (SC) and fully differentiated, nonproliferative, functional cells (FC). Proliferation is controlled by the fraction of proliferating cells, p; the division time, τ; and the asymmetry factor AF (the fraction of SCs arising from a cell division, in equilibrium, each division will produce one SC and one FC, and $AF = 0.5$). Irradiation will generate a new compartment of cells, containing SCs that have been lethally damaged by radiation and are doomed. Cells in this compartment still carry some proliferative capacity (abortive divisions), but will eventually die and disappear with a death rate, γ_A. FCs are nonproliferative and insensitive to radiation, but are lost due to natural turnover (rate μ). A graphical sketch of the model is presented in Fig. 1.

The set of delay differential equations describing the evolution of these three populations of cells during a radiotherapy treatment is as follows:

$$\frac{dS(t)}{dt} = \frac{2p(t-\tau)}{\tau} S(t-\tau) e^{-\gamma_S \tau} [1 - AF(t-\tau)] - \frac{p(t)}{\tau} S(t) - (1 - SF(t)) S(t) \delta(t - \{t_D\}) \quad (1)$$

$$\frac{dS_A(t)}{dt} = \frac{2p(t-\tau)}{\tau} S_A(t-\tau) e^{-\gamma_A \tau} [1 - AF(t-\tau)] - \frac{p(t)}{\tau} S_A(t) + (1 - SF(t)) S(t) \delta(t - \{t_D\}) \quad (2)$$

$$\frac{dF(t)}{dt} = \frac{2p(t-\tau) AF(t-\tau)}{\tau} [e^{-\gamma_S \tau} S(t-\tau) + e^{-\gamma_A \tau} S_A(t-\tau)] - \mu F(t) \quad (3)$$

γ_A and γ_S are the death rate of abortive and undamaged stem cells, respectively. The term $SF(t)$ is the surviving fraction of irradiated cells, and accounts for the transfer of cells from the healthy compartment, S, to the doomed compartment, S_A, when a radiation fraction is applied. $d(t)$ is the applied dose at time t, which is zero when t is not in $\{t_D\}$, the vector of irradiation times. The delivery of each fraction is considered to be instantaneous, hence the formal introduction of the Dirac delta function in the equations. If time between fractions is low, as can happen in schedules with multiple fractions per day, incomplete repair of sublethal damage may need to be considered. Therefore, the surviving fraction has been modeled with the linear-quadratic (LQ) model with incomplete repair correction.[19,20] The surviving fraction of cells following the k-th radiation fraction is given by:

$$\log SF_k = -\alpha d_k - \beta d_k^2 - 2\beta d_k \sum_{p=1}^{k-1} d_p \prod_{q=p}^{k-1} \theta_q \quad (4)$$

$$\theta_q = \exp(-\nu \Delta t_q); \Delta t_q = t_{q+1} - t_q \quad (5)$$

α and β are the LQ parameters characterizing cell radiosensitivity, and ν is the repair rate of sublethal damage.

AF and p describe the asymmetry factor and proliferative fraction via:

$$AF(t) = \frac{S(t) + S_A(t)}{S_0} \quad (6)$$

$$p(t) = max\left(p_0, 1 - (1 - p_0)\left[\frac{F(t)}{F_0}\right]\right) \quad (7)$$

where S_0 and F_0 are numbers of proliferative and functional cells in the steady state. According to this functional form, when an external perturbation (radiation) kills a substantial fraction of proliferative cells, the remaining proliferative cells will adapt and try to recover by decreasing AF (loss of asymmetry). This shortage of proliferative cells will lead to a shortage of functional cells because tissue turnover cannot be compensated. Loss of functional cells will trigger an increase in the fraction of proliferative cells undergoing division (accelerated proliferation). In Eq. (7), p_0 is the steady proliferation fraction, which can be determined from the steady-state condition ($dF/dt = 0$) to be $p_0 = \mu\tau F_0/(S_0\, exp(-\gamma_S \tau))$.

2.B. Intolerable/Tolerable mucositis model

We formulate simple binary criteria that assign a tolerable/intolerable classification to a given schedule according to the population of functional cells. Our tolerability criteria rely on the requirement that functional cells maintain the functionality of the tissue, and their elimination may trigger toxicity, while proliferating cells contribute to maintain intended cell densities. While toxicity is a complex phenomenon, which

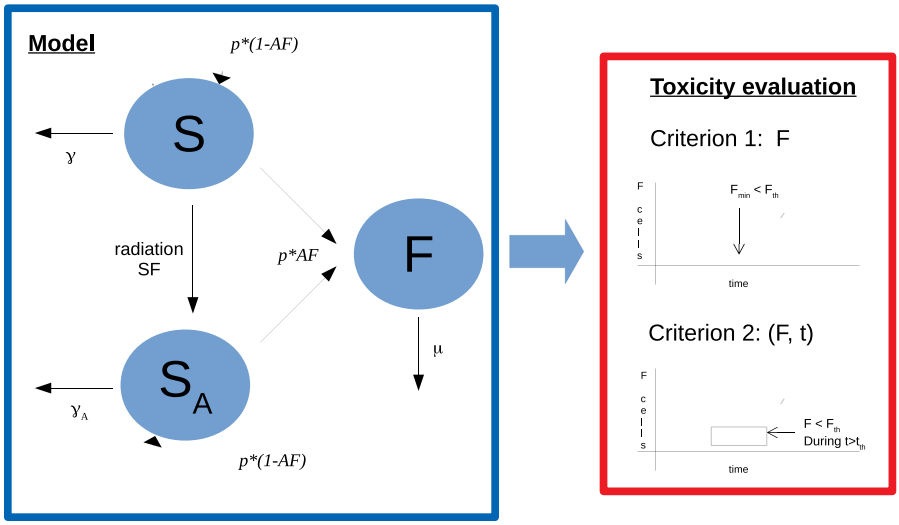

FIG. 1. Sketch of the model of cell dynamics of stem cells (S), abortive cells (S_A), and functional cells (F) (left), and the criteria for toxicity evaluation (right). [Color figure can be viewed at wileyonlinelibrary.com]

depends on more factors than the number of cells, in rapidly turning over tissues, the number of cells is bound to play an important role.[21–23]

We have investigated two simple criteria (Fig. 1): First, we consider a schedule to be intolerable if the number of functional cells falls below a given threshold (F_{th}), relative to the steady-state population of cells (*Criterion 1*):

$$\text{Criterion}_1 = \begin{cases} \text{intolerable, if } F \le F_{th} \\ \text{tolerable, otherwise} \end{cases} \quad (8)$$

Second, we consider that if the number of functional cells remains below the threshold longer than a time t_{th}, the schedule will be intolerable (*Criterion 2*):

$$\text{Criterion}_2 = \begin{cases} \text{intolerable, if } F \le F_{th} \text{for a length time } t \ge t_{th} \\ \text{tolerable, otherwise} \end{cases} \quad (9)$$

2.C. Comparison with other classification models

Our model has been compared with three phenomenological models presented in Refs. [6–8], which we will refer to as *Fowler*, *Fenwick*, and *Strigari*. These models are based on a BED boundary between tolerable/intolerable schedules, and they are presented in detail in an appendix.

2.D. Experimental dataset

We have tested our model on an extended dataset cataloging the tolerability/intolerability of radiation mucositis induced by different head-and-neck radiotherapy schedules, which builds upon schedules provided in Refs. [6–8] while also adding some further schedules. As in our model, we need not only total dose, dose per fraction, and treatment time but also the time of delivery of each fraction, we have reviewed each original article to find the necessary data. If times of dose delivery could easily be inferred from available data, such schedules were incorporated to the dataset, otherwise they were left out. In many cases, elements of the radiation schedule such as duration, gaps, and doses varied among the studied population, and we considered median values if reported. For accelerated schedules delivering several doses per day, a separation of 6 h between doses given in the same day was considered when not explicitly stated in the reference.

Some very low dose, tolerable schedules used in Refs. [6–8] were left out of the analysis, because in these cases, tolerability is not an issue. These schedules typically correspond to the first part of gap treatments, when toxicity is evaluated and treatment continued if acceptable (schedules 77–84 in table A.1.1 in Ref. [7]). Nonetheless, they are included in the dataset available as Supplementary Materials in Openoffice spreadsheet format in order to provide the most complete picture.

In some cases, the tolerability of a schedule is disputable. There are four such schedules in our dataset (table A.1.2. in Ref. [7]). These schedules are reported as "disputable" in our dataset, but for fitting purposes, we have considered K11 ("Program is not doable outside the experimental protocol") and F31 (" ... may necessitate lowering the dose ... ") as intolerable, and K41/F8 ("Acute toxicity does not represent a major obstacle") and F28 ("Toxicity is high but manageable") as tolerable.

The dataset contains data for 108 radiotherapy schedules (plus eight additional tolerable schedules that correspond to the first part of gap treatments that have not been used in the analysis, as discussed above), of which 15 were deemed intolerable (13.9%), a fraction similar to that reported in Ref. [7] (84 schedules, 13 intolerable, 15.5%), and Ref. [8] (107 schedules).

2.E. Numerical implementation, optimization, and parameters

The set of differential equations is coded in Matlab (The Mathworks Inc.) and solved by employing an Euler method,[24]

with a time step of 1 h. The classification models presented in Refs. [6–8] were also implemented. A simulated annealing algorithm[13,25] was implemented for the optimization problem (model fitting to the experimental dataset). This methodology seems appropriate for the problem under study, where gradient-descent methods may be suboptimal due to local minima/maxima, and closed-form solutions do not exist (or are unknown). The objective function, C, to be maximized is defined as the sum of sensitivity (sen) and specificity (spe),

$$C = sen + spe \tag{10}$$

where the sensitivity is defined as the number of true positives (TP) over positive cases (P), and the specificity as the number of true negatives (TN) over negative cases (N):

$$\begin{aligned} sen &= \frac{TP}{P} \\ spe &= \frac{TN}{N} \end{aligned} \tag{11}$$

Positives are defined as intolerable schedules, and TP as schedules correctly classified as intolerable, while negatives are defined as tolerable schedules, and TN as schedules correctly classified as tolerable. We use such an objective function in order to give equal weight to tolerable and intolerable schedules, even if the number of intolerable schedules in the dataset is much lower than the number of tolerable ones. Other optimization criteria, like minimizing the number of misclassifications (either tolerable or intolerable), would not be convenient as they would undervalue the importance of correct classification of intolerable schedules.

We have fixed the values of $\alpha/\beta = 10$ Gy (as typically done for mucositis[12]), and the incomplete repair parameter, $T_{1/2} = 3.2$ h as in Ref. [7,26] ($\nu = \log(2)/T_{1/2} = 0.2166$ h^{-1}). The BED model of Fowler has three free-fitting parameters, αT_d, T_k, and BED_0, while the BED model of Fenwick and TD50-model have two free parameters each, BED_0 and T_c, and $TD_{50}(0)$ and Δ, respectively. These free parameters have been fitted to maximize the cost function C defined in the previous section. The number of free parameters in our model is larger, six to seven parameters, namely, α, τ, γ_A (we take $\gamma_S = 0$), μ, the fraction of functional cells, f_F (the percentage of proliferating cells is $f_S = 1 - f_F$), and F_{th} and t_{th} (the parameters of the tolerability model). In order to reduce the number of free parameters in our model, we have fixed $\alpha = 0.1$ Gy^{-1} and $\tau = 48$ h. These parameters have not been as frequently studied experimentally or used in models, especially in clinical studies. Our approach has been to set round values within ranges reported in the literature. For the proliferation time τ, Dörr and colleagues have reported values ranging from ~1 day to ~4 days ([12,27] and references therein). We have picked a value in this range, 2 days. This value would also approximately match the proliferation needed to counter a turnover time of 4 days (in the range reported for those tissues) if ~50% of the cells are proliferative and $p_0 = 0.5$. Regarding the value of α, analysis of animal studies[13,27] shows very low values of α (~0.02–0.05 Gy^{-1}), yet values typically reported in clonogenic studies tend to be higher, ~0.2–0.3 Gy^{-1}. Therefore, we have opted for selecting a relatively low value for α, yet not as low as some reported values.

2.F. Model validation, parameter stability, and cross-validation

First, we have investigated the performance of our model (with two classification criteria) and models reported in Refs. [6–8] (Eqs. A1–A4) on the whole dataset. Several optimizations were run in order to avoid convergence to local maxima. Once best-fitting parameters were obtained, the stability of the solution around those parameters was evaluated, in order to provide an estimation of the uncertainties of best-fitting parameters. Random perturbations were applied to the set of best-fitting parameters, the value of the objective function evaluated, and the combination of parameters saved if the value of the objective function did not change (until reaching 100 combinations). The dispersion of parameters was then analyzed.

Secondly, we investigated the cross-validation of all the models studied: the performance of the model, with optimal parameters obtained in a *training dataset*, on an independent *validation dataset*. In order to do so, and due to the lack of several large datasets, we have used the methodology of *stratified k-fold* cross-validation[28]: Our dataset has been split in two datasets ($k = 2$), training and validation; each schedule is randomly included in one or another dataset (50% probability); model parameters are optimized in the training dataset; and the model with those same parameters is used on the validation dataset, and the sensitivity and specificity are calculated; this procedure is repeated 20 times (with different training/validation datasets) to check the consistency of the results. Due to the relatively low number of intolerable schedules, the generation of the training/validation datasets was stratified, imposing a constraint of at least five intolerable schedules in every training/validation dataset.

3. RESULTS

3.A. Parameter optimization and validation: whole dataset

We have investigated the performance of the models under study on the whole dataset. In Table I, we present sets of best-fitting parameters (that maximize the cost function), as well as achieved values of sensitivity and specificity, for every single model.

We have observed through multiple optimizations that our models present maxima clustered around two different sets of parameters for *Criterion 1* and *Criterion 2*: The first configuration corresponds to rapidly disappearing abortive cells (γ_A ~ 0.04 h^{-1}, corresponding to a half-life of ~17 h) and very low values of F_{th} to trigger intolerability (~2×10^{-3}); the second cluster is associated with longer lived abortive cells (γ_A ~ 0.007 h^{-1}, corresponding to a half-life ~100 h) and much higher values of F_{th} (~0.6). The turnover rate of functional cells μ is in the 0.005–0.011 h^{-1} range, which

TABLE I. Optimal parameters, and sensitivity and specificity values when applying our model (*classification criterion 1* and *criterion 2*) and *Fenwick*, *Strigari*, and *Fowler* models to the whole dataset.

Optimal parameters						
	Criterion 1	Criterion 2		Fenwick Ref. [7]	Strigari Ref. [8]	Fowler Ref. [6]
α	0.1 Gy^{-1}*	0.1 Gy^{-1}*	BED_0 / $TD_{50}(0)$	64.25	48.67	67.38
τ	48 h*	48 h*	T_c / Δ	45.60 d	84.44 d	19.44 d
$T_{1/2}$	3.2 h*	3.2 h*	$T_{1/2}$	3.2 h*	3.2 h*	3.2 h*
α/β	10 Gy*	10 Gy*	α/β	10 Gy*	10 Gy	10 Gy*
γ_A	0.044 h^{-1}	0.006 h^{-1}	αT_d	–	–	0.677 d Gy^{-1}
p_0	0.685	0.200	–	–	–	–
f_S	0.437	0.417	–	–	–	–
F_{th}	0.002	0.682	–	–	–	–
t_{th}	–	246 h	–	–	–	–

Sensitivity and Specificity					
	Criterion 1	Criterion 2	Fenwick Ref. [7]	Strigari Ref. [8]	Fowler Ref. [6]
sen	0.933	0.867	1.000	1.000	1.000
spe	0.839	0.957	0.806	0.774	0.785
C	1.772	1.824	1.806	1.774	1.785

The symbol * indicates that the parameter value has been fixed and it is not part of the optimization. These sets of parameters are those used in Figs. 2–4.

yields half-lives for such cells of ~60–140 h, within the range of typically reported 4 days.[21]

Classification *criterion 1* can reach sensitivity and specificity values of 93.3% and 83.9%, respectively (C = 1.772), while *criterion 2*, which incorporates an extra free parameter can improve the value of the cost function to C = 1.824.

In Fig. 2, we show the evolution of the population of cells during treatment for three representative types of schedules in our database, a conventional 35 × 2 Gy-7 weeks schedule, an accelerated 68 × 1.2 Gy-2 fracs/day schedule, and an accelerated + gap schedule. The depletion during treatment and subsequent repopulation of the proliferative and functional compartments can be qualitatively appreciated, together with the appearance and elimination of doomed cells after some abortive divisions. In Fig. 3, we present histograms of the populations of functional cells, split in tolerable and intolerable schedules (classification *criterion 1*).

The predictive value of our models is similar to that of models presented in Refs. [6–8] (Table I). When using the *Fowler* model (Eq. A1) to analyze the whole dataset, we obtain sensitivity and specificity values of 100% and 78.5% (C = 1.785); with *Fenwick* model (Eq. A2), we obtain sensitivity and specificity values of 100% and 80.6% (C = 1.806); and with the *Strigari* model (Eq. A4), 100.0% and 77.4% (C = 1.774), respectively. In Fig. 4, we show the tolerability/intolerability boundary obtained with *Fowler*, *Fenwick*, and *Strigari* models.

In Fig. 5, we present a study of the uncertainties of the best-fitting parameters and stability of the solution under perturbations of parameters, as described in Section 2.6. The figure shows correlations between different parameters, with different values leading to the same value of the objective function. In general, the maxima of the objective function found with the optimization algorithm (reported in Table I) are stable under relative perturbations of ~0.2–1% for *Fowler*, *Fenwick,* and *Strigari*, ~1–2% for *classification criterion 1*. The best solution obtained with *criterion 2*, which is the best solution among all models, is extremely sensitive to perturbations of parameters, and variations <0.02% for any parameter other than t_{th} lead to a worsening of the solution.

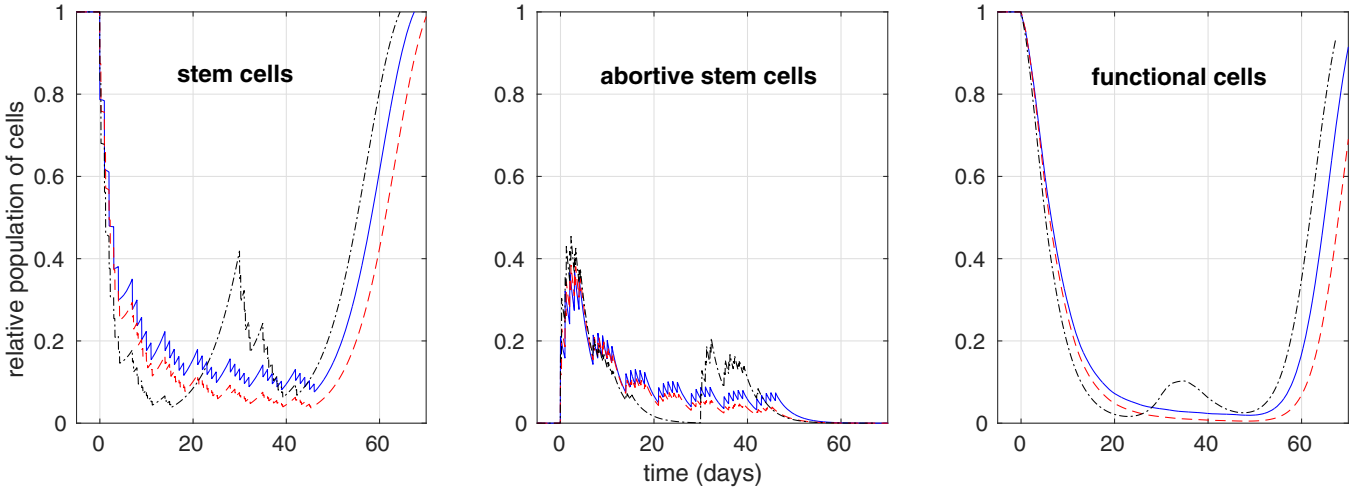

FIG. 2. Dynamics of populations of cells vs time for three different treatments in our dataset: a conventional 35 × 2 Gy-7 weeks treatment (solid blue line), an accelerated 68 × 1.2 Gy-2 fracs/day treatment (dashed red line), and an accelerated + gap treatment, 42 × 1.6 Gy-2 fracs/day-15 days gap after 12 fractions (dash-dot black line). The three panels show healthy proliferative cells (left), doomed proliferative cells (center), and functional cells (right). These results have been obtained with parameters presented in Table I (*criterion 1*). [Color figure can be viewed at wileyonlinelibrary.com]

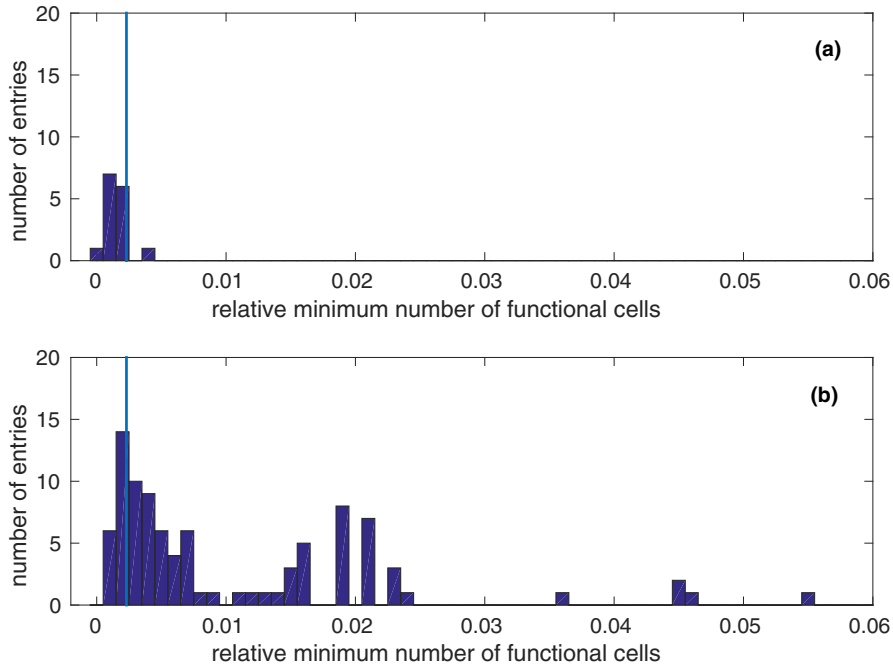

FIG. 3. Histogram of the relative minimum number of functional cells for each of the 108 schedules in our dataset, split in intolerables (a) and tolerables (b). The thick lines in both panels represent the threshold F_{th} of tolerability. These results have been obtained with model parameters presented in Table I (*criterion 1*). [Color figure can be viewed at wileyonlinelibrary.com]

3.B. Cross-validation

When we performed a cross-calibration of by splitting the dataset in two sets, one for training and one for validation, the specificity and sensitivity in the validation dataset were lower than in the training dataset for all models, and this was particularly the case for *Fowler* model. In Table II, we report mean values and standard deviation for 20 cross-validations (with different randomly generated training/validation datasets).

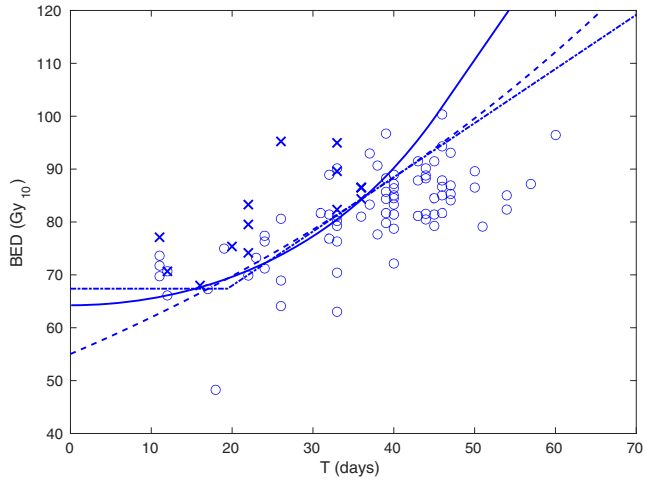

FIG. 4. Biologically effective dose (BED) vs treatment time for each schedule in the dataset. Intolerable schedules are represented with crosses, and tolerable schedules with circles. Lines show the optimal boundaries between tolerable and intolerable schedules obtained with *Fowler* model (dash line), *Fenwick* model (solid line), and *Strigari* model (dash-dot line). Notice that EQD_2 values from Strigari's fit have been converted to BED values. [Color figure can be viewed at wileyonlinelibrary.com]

Average sensitivity and specificity values are 95–99% and 83–87%, respectively, for the training dataset, falling to 72–84% and 79–83% for the validation dataset for all models except Fowler's, with large standard variations (up to 20%) for sensitivity in the validation dataset. C-values (sen + spe) decrease from ~1.82 for the training datasets to ~1.55–1.63 for the validation datasets. The performance of *Fowler* model in the validation datasets is notably less good, with sensitivity and specificity values around 68% and 81%, respectively, and C ~ 1.49.

4. DISCUSSION

Mucosal toxicity is a limiting factor in dose escalation and/or treatment shortening in head-and-neck cancer. Predicting the tolerability or intolerability of a given radiotherapy schedule can help the design of novel radiotherapy treatments, aiming at maximizing tumor control while keeping toxicity under control. Different mathematical models have been introduced to tackle this issue. In this work, we present a novel model, based on population dynamics of cells, which is more mechanistic than other purely phenomenological models. Our model has been used to classify tolerable/intolerable schedules in a dataset comprising 108 schedules, and its performance has been compared with three classification models based on BED and EQD2, developed by Fowler et al.,[6] Fenwick et al.,[7] and Strigari et al.[8]

Remarkably, our models can match the predictive power of classical phenomenological models. After an optimization aiming at maximizing the sum of sensitivity and specificity on the whole dataset, we have obtained similar values of

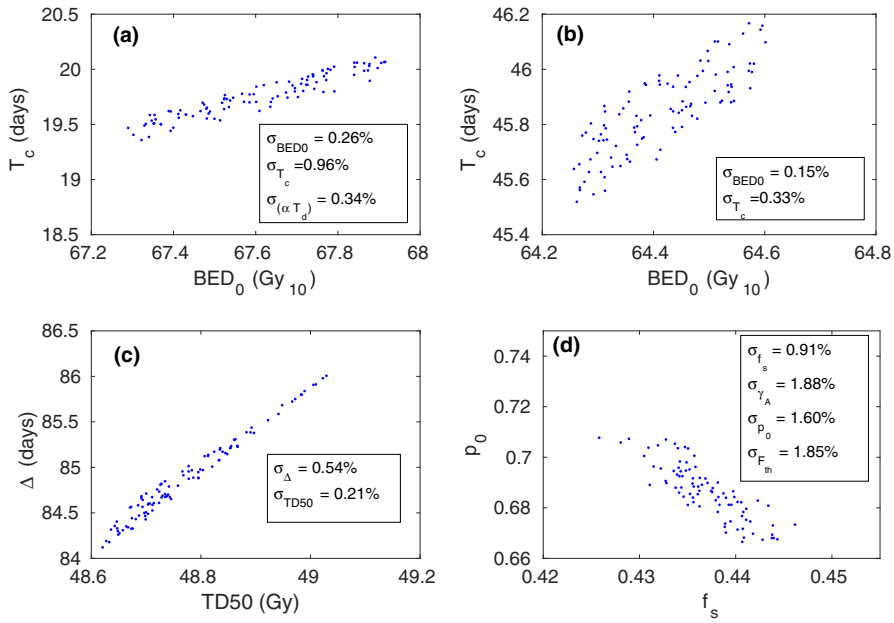

FIG. 5. Plots of different combinations of parameters (perturbations around values reported in Table I) that lead to the same value of the objective function for *Fowler* (a), *Fenwick* (b), *Strigari* (c), and *Criterion 1* (d). We also show relative standard deviations, σ, for each parameter. Notice that only two parameters are shown in each panel, but *Fowler* and *Criterion 1* have more degrees of freedom. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE II. Results of the cross-validation: The model is trained with the training dataset, aiming at maximizing C = sensitivity +specificity, and then applied to the validation dataset. Results are presented for 20 runs by reporting means and standard deviations. *Criterion 1* and *Criterion 2* refer to the two different classification criteria presented in this work, while *Fowler*, *Fenwick*, and *Strigari* refer to models presented in Refs. [6–8].

	Training dataset			Validation dataset		
	Specificity	Sensitivity	C	Specificity	Sensitivity	C
Criterion 1	0.857 ± 0.059	0.961 ± 0.063	1.818 ± 0.078	0.825 ± 0.064	0.798 ± 0.135	1.622 ± 0.108
Criterion 2	0.871 ± 0.070	0.950 ± 0.076	1.821 ± 0.078	0.830 ± 0.083	0.723 ± 0.142	1.553 ± 0.137
Fenwick	0.827 ± 0.064	0.993 ± 0.032	1.819 ± 0.056	0.806 ± 0.057	0.822 ± 0.209	1.628 ± 0.187
Strigari	0.828 ± 0.057	0.988 ± 0.036	1.817 ± 0.064	0.787 ± 0.069	0.836 ± 0.190	1.621 ± 0.168
Fowler	0.859 ± 0.067	0.959 ± 0.074	1.818 ± 0.072	0.814 ± 0.072	0.668 ± 0.230	1.492 ± 0.206

sensitivity and specificity for our models (sen = 93.3%, spe = 83.9%, *criterion 1*; sen = 93.3%, spe = 88.2%, *criterion 2*) and the other three models under study (sen = 100.0% and spe = 77.4%, *Fowler*; sen = 100.0% and spe = 80.6%, *Fenwick*; and sen = 100.0% and spe = 77.4%, *Strigari*). The values of the objective function, C = sen+spe, are very similar for the five models studied: 1.772 (*criterion 1*), 1.824 (*criterion 2*), 1.785 (*Fowler*), 1.806 (*Fenwick*), and 1.774 (*Strigari*).

Sensitivity and specificity are always overrated when the model is evaluated on the dataset used to optimize its performance. But, after performing a cross-validation of our model, in which the model is trained on half of the dataset (on average), and validated on the other half, we have found reasonable values of sensitivity and specificity on the validation dataset (around 80% and 82% for *criterion 1*). This demonstrates that the good performance of the model is not due to overfitting in the (training) dataset, as predictive power is still good in a different dataset, and shows that the model is reliable for the classification of tolerable/intolerable schedules. Interestingly, results are somewhat worse with *criterion 2* in the validation dataset, which may be caused by the extra free parameter causing overfitting in the training dataset. The same cross-validation procedure was applied to *Strigari*, *Fenwick*, and *Fowler* models. The first two models also maintain fair predictive power when applied to the validation dataset. However, *Fowler* model had limited success in correctly classifying intolerable schedules in the validation dataset (sen ~ 68%). This limitation may result from both the rather small number of intolerable schedules in the dataset, which can cause the training and validation cohorts to contain very different information about intolerable schedules, and the greater degree of flexibility of the *Fowler* model, which has three separate parameters controlling the baseline tolerable BED, the rate at which this increases with schedule duration and the critical duration beyond which the increase begins, compared to only two parameters in the models of *Fenwick* and *Strigari*.

Like the models to which it has been compared, our model aims at separating populationwise tolerable and intolerable

schedules. While this is important, for example, when designing new fractionations, certainly the optimal approach would be to develop methods that would allow to predict the risk of intolerable toxicity of each individual patient, based on their dose distribution, via NTCP approaches or patient-based machine-learning methodologies. However, our model potentially permits some individualization by linking model parameters such as radiosensitivity and proliferation capacity, cell turnover to genetic profiling.[29,30]

A limitation of our model arises from its compartment model nature: The model does not include spatial dimensions, and therefore cannot directly account for dose–volume or dose–area effects, including repopulation by cell migration from undamaged tissue. This limitation does not allow to incorporate into the model complex individualized dose distributions in mucosa, as can be done in some NTCP-based methods.[9,10] Spatial dimensions could be added to the model, transforming it in a spatiotemporal model. Such model would be a more accurate representation of the mucosa, but also rather more complex. With the good results obtained with the current method, it is doubtful that such spatiotemporal models would improve obtained sensitivity/specificity, but it is an issue worth investigating, as a model of that kind would allow to simulate the effect of dose distribution heterogeneities on the response of the mucosa. Certainly, the added complexity of such a model would probably only be worthwhile if it was used to fit data including spatial information.

Our model relies on the underlying description of fractionation effects provided by the LQ model. The validity of the LQ model has been questioned, especially when delivering large radiation doses, and some alternatives have been suggested.[31,32] Those models could be easily implemented in our model, but given that highly hypofractionated schedules are rarely used for head-and-neck radiotherapy, using the LQ model seems to be well reasonable.

An interesting approach would be to develop and apply a similar model to chemo-irradiation treatments, which have been investigated by Meade et al.[33] using the models presented in Refs. [6–8]. In our model, chemotherapy could be introduced through a new term of cell death in the proliferative compartment, thereby limiting repopulation of both the proliferative and functional compartments.

In general, the success of phenomenological models (in particular *Fenwick* and *Strigari* models) in predicting the tolerability of different radiotherapy schedules in head-and-neck cancer (including cross-calibration), and their simplicity, should make them preferable over complex mechanistic models in the clinic. However, mechanistic models can still usefully add to our understanding of toxicity, and may assist with the design of unconventional dose fractionations.

5. CONCLUSIONS

Our model has proved useful to predict the tolerability/intolerability different radiotherapy schedules on a dataset of 108 schedules, achieving sensitivity/specificity values similar to those obtained using other models (*Fowler*, *Fenwick,* and *Strigari*).

Mechanistic models can clarify our understanding of the processes involved in the response to radiotherapy, and our approach may prove helpful when designing unconventional dose fractionations delivering differing doses per week/day, the tolerability of such schedules not being described by the datasets to which phenomenological models of toxicity have been fitted. However, *Fenwick* and *Strigari* models can classify tolerable/intolerable schedules as well as the more mechanistically based models presented in this work, and their relative simplicity may facilitate their use in the clinic.

ACKNOWLEDGMENTS

This project was funded by Instituto de Salud Carlos III (ISCIII) through research grants PI17/01428 and DTS17/00123 (FEDER co-funding). J.P-M. is supported by ISCIII through a Miguel Servet II grant (CPII17/00028, FEDER co-funding).

CONFLICT OF INTEREST

The authors report no conflict of interest.

APPENDIX A

CLASSIFICATION MODELS OF FOWLER, FENWICK, AND STRIGARI

A.1. Fowler

In Ref. [6], the boundary between tolerable/intolerable radiotherapy schedules was modeled with a simple hockey stick BED function:

$$BED_{boundary}(T) = BED_0 + \log(2)\frac{(T - T_k)}{\alpha T_p} \quad \text{(A1)}$$

where T_k is the accelerated proliferation kick-off time, and T_p is the doubling time of stem cells in the mucosa.

A.2. Fenwick

In Ref. [7], the boundary between tolerable/intolerable radiotherapy schedules was modeled with a piecewise function:

$$BED_{boundary}(T) = \begin{cases} BED_0\left[\frac{T}{T_c}\right]/\sin\left(\frac{T}{T_c}\right), \text{ if } T \le T_c \\ BED_0\left[\frac{T}{T_c}\right], \text{ if } T > T_c \end{cases} \quad \text{(A2)}$$

In the above equations, T is the treatment time (in days), including the first day of treatment, and the argument of the sine term is expressed in radians. If the BED of a given schedule is above the boundary, it is classified as intolerable,

while if it is below the boundary, it is classified as tolerable. The BED of each schedule is computed including incomplete repair, as in Ref. [7]:

$$BED = \sum_{i=1}^{n} d_i + \frac{\beta}{\alpha}\sum_{i=1}^{n} d_i^2 + \frac{2\beta}{\alpha}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} d_i d_j \times 0.5^{\sum_{k=i}^{j-1}\Delta t_k/T_{1/2}} \tag{A3}$$

Here, Δt_k is the time between fraction k and $k + 1$, and $T_{1/2}$ is the repair half-time.

A.3. Strigari

In Ref. [8], the authors used a TD_{50} boundary, in units of EQD_2, equivalent dose in 2 Gy fractions.

$$TD_{50}(T) = TD_{50}(0)\exp((T-5)/\Delta) \tag{A4}$$

where T is again the treatment time, and $TD_{50}(0)$ and Δ are fitting parameters. If the EQD_2 of a given schedule is below the boundary, it is classified as tolerable; otherwise, it is intolerable. The EQD_2 is computed from the BED as $EQD_2 = BED/(1 + 2/(\alpha/\beta))$, and also includes the effect of incomplete repair.

DATA AVAILABILITY STATEMENT

The data used for this study are provided in a supplementary spreadsheet. The code containing the implementation of the models and optimization methods is available from the corresponding author upon reasonable request.

[#]Now at the Basque Center for Applied Mathematics (BCAM) and CIC-bioGUNE

[a)]Author to whom correspondence should be addressed. Electronic mail: juan.pardo.montero@sergas.es; Telephone: +34981955604.

REFERENCES

1. Ang KK. Altered fractionation trials in head and neck cancer. *Semin Radiat Oncol*. 1998;8:230–236.
2. Trotti A, Bellm LA, Epstein JB, et al. Mucositis incidence, severity and associated outcomes in patients with head and neck cancer receiving radiotherapy with or without chemotherapy: A systematic literature review. *Radiother Oncol*. 2003;66:253–262.
3. Sroussi HY, Epstein JB, Bensadoun R-J, et al. Common oral complications of head and neck cancer radiation therapy: Mucositis, infections, saliva change, fibrosis, sensory dysfunctions, dental caries, periodontal disease, and osteoradionecrosis. *Cancer Med*. 2017;6:2918–2931.
4. Vera-Llonch M, Oster G, Hagiwara M, Sonis S. Oral mucositis in patients undergoing radiation treatment for head and neck carcinoma. *Cancer*. 2006;106:329–336.
5. Kaanders JHAM, van der Kogel AJ, Ang KK. Altered fractionation: Limited by mucosal reactions? *Radiother Oncol*. 1998;50:247–260.
6. Fowler JF, Harari PM, Leborgne F, Leborgne JH. Acute radiation reactions in oral and pharyngeal mucosa: Tolerable levels in altered fractionation schedules. *Radiother Oncol*. 2003;69:161–168.
7. Fenwick JD, Lawrence GP, Malik Z, Nahum AE, Mayles WP. Early mucosal reactions during and after head-and-neck radiotherapy: Dependence of treatment tolerance on radiation dose and schedule duration. *Int J Radiat Oncol Biol Phys*. 2008;71:625–634.
8. Strigari L, Pedicini P, D'Andrea M, et al. A new model for predicting acute mucosal toxicity in head-and-neck cancer patients undergoing radiotherapy with altered schedules. *Int J Radiat Oncol Biol Phys*. 2012;83:e697–702.
9. Dean JA, Wong KH, Welsh LC, et al. Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy. *Radiother Oncol*. 2016;120:21–27.
10. Dean JA, Welsh LC, Wong KH, et al. Normal tissue complication probability (NTCP) modelling of severe acute mucositis using a novel oral mucosal surface organ at risk. *Clin Oncol*. 2017;29:263–273.
11. Dean JA, Wong KH, Gay H, et al. Functional data analysis applied to modeling of severe acute mucositis and dysphagia resulting from head and neck radiation therapy. *Int J Radiat Oncol Biol Phys*. 2016;96:820–831.
12. Dörr W. Three A's of repopulation during fractionated irradiation of squamous epithelia: Asymmetry loss, acceleration of stem-cell divisions and Abortive divisions. *Int J Radiat Biol*. 1997;72:635–643.
13. Parga-Pazos M, López Pouso O, Fenwick JD, Pardo-Montero J. A mathematical model of dynamics of cell populations in squamous epithelium after irradiation. *Int J Rad Biol*. 2020;96:1165–1172.
14. Fenwick JD. Delay differential equations and the dose-time dependence of early radiotherapy reactions. *Med Phys*. 2006;33:3526–3540.
15. Schneider U. Mechanistic model of radiation-induced cancer after fractionated radiotherapy using the linear-quadratic formula. *Med Phys*. 2009;36:1138–1143.
16. Kuperman VY, Lubich LM. Effect of reoxygenation on hypofractionated radiotherapy of prostate cancer. *Med Phys*. 2020;47:5383–5391.
17. Neira S, Gago-Arias A, Guiu-Souto A, Pardo-Montero J. A kinetic model of continuous radiation damage to populations of cells: Comparison to the LQ model and application to molecular radiotherapy. *Phys Med Biol*. 2020;65:245015.
18. Jeremic MZ, Matovic MD, Krstic DZ, Pantovic SB, Nikezic DR. A five-compartment biokinetic model for 90 Y-DOTATOC therapy. *Med Phys*. 2018;45:5577–5585.
19. Fowler JF. The linear-quadratic formula and progress in fractionated radiotherapy. *Br J Radiol*. 1989;62:679–694.
20. Wakisaka Y, Yagi M, Sumida I, Takashina M, Ogawa K, Koizumi K. Impact of time-related factors on biologically accurate radiotherapy treatment planning. *Radiat Oncol*. 2018;13:30.
21. Gruber S, Dörr W. Tissue reactions to ionizing radiation - Oral mucosa. *Mutat Res*. 2016;770:292–298.
22. Villa A, Sonis ST. Mucositis: Pathobiology and management. *Curr Opin Oncol*. 2015;27:159–164.
23. Rutkowska E, Baker C, Nahum A. Mechanistic simulation of normal-tissue damage in radiotherapy - implications for dose-volume analyses. *Phys Med Biol*. 2010;55:2121–2136.
24. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes: The Art of Scientific Computing*, Cambridge: Cambridge University Press; 2007.
25. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by Simulated Annealing. *Science*. 1983;220:671–680.
26. Bentzen SM, Ruifrok AC, Thames HD. Repair capacity and kinetics for human mucosa and epithelial tumors in the head and neck: Clinical data on the effect of changing the time interval between multiple fractions per day in radiotherapy. *Radiother Oncol*. 1996;38:89–101.
27. Dörr W, Obeyesekere MN. A mathematical model for cell density and proliferation in squamous epithelium after single-dose irradiation. *Int J Radiat Biol*. 2001;77:497–505.
28. Burman P. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*. 1989;76:503–514.
29. Fachal L, Gómez-Caamaño A, Barnett GC, et al. A three-stage genome-wide association study identifies a susceptibility locus for late radiotherapy toxicity at 2q24.1. *Nat Genet*. 2014;46:891–894.
30. Barnett GC, Thompson D, Fachal L, et al. A genome wide association study (GWAS) providing evidence of an association between common genetic variants and late radiotherapy toxicity. *Radiother Oncol*. 2014;111:178–185.

31. Guerrero M, Li XA. Extending the linear- quadratic model for large fraction doses pertinent to stereotactic radiotherapy. *Phys Med Biol*. 2004;49:4825–4835.
32. Wang JZ, Huang Z, Lo SS, Yuh WTC, Mayr NA. A generalized linear-quadratic model for radiosurgery, stereotactic body radiation therapy, and high-dose rate brachytherapy. *Sci Transl Med*. 2010;2:1–7.
33. Meade S, Sanghera P, Glaholm J, Hartley A. Models of acute mucosal tolerance to radiotherapy alone applied to synchronous chemoradiation schedules in head and neck cancer. *Tumor Biology*. 2014;35:2017–2023.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1: Dataset containing detailed information of the schedules analyzed in this study.

WILEY

First Clinical Use of SunSCAN 3D Webinar

Initial Impressions and Workflow Improvements

December 13, 2022 - 12:00 PM EST

AMERICAN ASSOCIATION of PHYSICISTS IN MEDICINE

Sponsored by SUN NUCLEAR