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

Juan Pardo-Montero[1,2,*], Martín Parga-Pazos[1,+], John D. Fenwick[3,4]

1. Group of Medical Physics and Biomathematics, Instituto de Investigación Sanitaria de Santiago (IDIS), Santiago de Compostela, Spain.

2. Department of Medical Physics, Complexo Hospitalario Universitario de Santiago de Compostela, Spain.

3. Department of Molecular and Clinical Cancer Medicine, Institute of Translational Medicine, University of Liverpool, Liverpool, United Kingdom.

4. Department of Physics, Clatterbridge Cancer Centre, Clatterbridge Road, Wirral, United Kingdom.

+ Now at the Basque Center for Applied Mathematics (BCAM) and CIC-bioGUNE

*Corresponding author: Juan Pardo-Montero, Instituto de Investigación Sanitaria de Santiago, Servizo de Radiofísica e Protección Radiolóxica, Hospital Clínico Universitario de Santiago, Trav. Choupana s/n, 15706, Santiago de Compostela; Phone: +34981955604; E-mail: juan.pardo.montero@sergas.es

Conflict of interest
The authors report no conflict of interest

Running title - Mucosal toxicity in H&N radiotherapy

Abstract

Purpose: 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 cells populations in irradiated mucosa, and incorporates the three A's: Accelerated proliferation, loss of Asymmetric proliferation, Abortive divisions. The model consist 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 that 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.

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

1. Introduction

Many clinical trials have tested non-conventional 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 at.* [5] initially explored this issue, and by plotting the total dose versus 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 [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 patient-wise the probability of severe mucosal mucosal toxicity [9, 10, 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 [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 cell populations 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 [7] and [8], and adding some additional schedules.

3

2. Methods and Materials

105

2.1. Model of dynamics of cell populations

The model presented in [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 110 description of cell dynamics in preclinical data of cell. We initially assume two compartments of cells: proliferative cells (SC), and fully differentiated, non-proliferative, 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 115 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 non-proliferative and insensitive to radiation, but are lost due to natural turnover (rate μ). A graphical sketch of the model is presented in Figure 1.

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

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

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

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

γ_A and γ_S are the death rate of abortive and undamaged stem cells, respectively. The term *SF(t)* is the 130 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)* 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 135 with multiple fractions per day, incomplete repair of sub-lethal damage may need to be considered.

4

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:

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

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

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

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

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

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

where S_0 and F_0 are numbers of proliferative and functional cells 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 $p_0 = \mu\tau F_0/(S_0 \exp(-\gamma_S \tau))$.

2.2. Intolerable/Tolerable mucositis model

We formulate simple binary criteria that assign a tolerable/intolerable classification to a given

5

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 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 (Figure 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*):

$$\text{Criterion}_1 = \begin{cases} \text{intolerable}, \text{if } F \leq F_{th} \\ \text{tolerable}, \text{otherwise} \end{cases} \tag{8}$$

Secondly, 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*):

$$\text{Criterion}_2 = \begin{cases} \text{intolerable}, \text{if } F \leq F_{th} \text{ for a length time } t \geq t_{th} \\ \text{tolerable}, \quad \text{otherwise} \end{cases} \tag{9}$$

2.3. Comparison with other classification models

Our model has been compared with three phenomenological models presented in references [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 the detail in an Appendix.

2.4. 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 [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 [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 to 84 in Table A1.1 in [7]). Nonetheless, they are included in a 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 8 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 [7] (84 schedules, 13 intolerable, 15.5%), and [8] (107 schedules).

2.5. 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 hour. The classification models presented in [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),

$$C = sen + spe \qquad (14)$$

235

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):

$$sen = \frac{TP}{P}$$
$$spe = \frac{TN}{N} \qquad (15)$$

240

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

245   than the number of tolerable ones. Other optimization criteria, like minimizing the number of miss-classifications (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
250   repair parameter, $T_{1/2} = 3.2$ h as in [7, 26] ($v = \log(2)/T_{1/2} = 0.2166$ 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, 6-7 parameters, namely, α, τ, γ_A (we take $\gamma_S = 0$), μ, the fraction
255   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
260   values ranging from ~1 day to ~4 days ([27, 12] 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 ([27, 13]) show very low values of

8

α (~0.02-0.05 Gy^{-1}), yet values typically reported in clonogenic studies tend to be higher, ~0.2-0.3

265 Gy^{-1}. Therefore, we have opted for selecting a relatively low value for α, yet not as low as some reported values.

2.6. Model validation, parameter stability, and cross-validation

270

First, we have investigated the performance of our model (with two classification criteria) and models reported in references [6-8] (equations A1-4) 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

275 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.

280 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%

285 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 5 intolerable schedules in

290 every training/validation dataset.

3. Results

295 **3.1. Parameter optimization and validation: whole dataset**

We have investigated the performance of the models under study on the whole dataset. In Table 1

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.

300

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 ($\gamma_A \sim 0.04$ 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-

305 lived abortive cells ($\gamma_A \sim 0.007$ 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 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%,

310 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 Figure 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

315 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 Figure 3 we present histograms of the populations of functional cells, split in tolerable and intolerable schedules (classification *criterion 1*).

320

The predictive value of our models is similar to that of models presented in [6-8] (Table 1). When using the *Fowler* model (equation A1) to analyze the whole dataset, we obtain sensitivity and specificity values of 100% and 78.5% (C=1.785); with *Fenwick* model (equation A2) we obtain sensitivity and specificity values of 100.% and 80.6% (C=1.806); and with the *Strigari* model

325 (equation A4), 100.0% and 77.4% (C=1.774), respectively. In Figure 4 we show the tolerability/intolerability boundary obtained with *Fowler*, *Fenwick*, and *Strigari* models.

In Figure 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

330 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

10

algorithm (reported in Table 1) are stable under relative perturbations of ~0.2-1% for *Fowler, Fenwick* and *Strigari*, ~1-2% for *classification criterion 1.* On the other hand, 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.

3.2. Cross-validation

When we performed a cross-calibration of our model by splitting the dataset in two sets, one for training one for validation, for all models the specificity and sensitivity of classification of tolerability in the validation dataset were lower than in the training dataset, and this was particularly the case for *Fowler* model. In Table 2 we report mean values and standard deviation for 20 cross-validations (with different randomly generated training/validation datasets).

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], 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 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% 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*).

On the other hand, 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 population-wise 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, viaNTCP approaches or patient-based machine-learning methodologies. However, our model

400   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,
405   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 spatio-temporal 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
410   that such spatio-temporal 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.

415   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.
420

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 [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
425   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
430   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

435

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*).

440 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

445 mechanistically based models presented in this work, and their relative simplicity may facilitate their use in the clinic.

455 ## Appendix: Classification models of Fowler, Fenwick, and Strigari

Fowler:

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

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

460
(A1)

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

465 *Fenwick:*

In [7], the boundary between tolerable/intolerable radiotherapy schedules was modeled with a

piecewise function:

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

(A2)

470

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 [7]:

475

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

(A3)

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

480   *Strigari:*

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

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

(A4)

485   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.

490

Data availability

The data used for this study is 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.

495

References

1. Ang KK. Altered fractionation trials in head and neck cancer. Semin Radiat Oncol 1998;8:230-236. DOI: 10.1016/s1053-4296(98)80020-9

2. Trotti A, Bellm LA, Epstein JB, Frame D, Fuchs HJ, Gwede CK, 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(3):253-262.. DOI: 10.1016/s0167-8140(02)00404-8

3. Sroussi HY, Epstein JB, Bensadoun RJ, Saunders DP, Lalla RV, Migliorati CA, 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(12):2918-2931. DOI: 10.1002/cam4.1221

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(2):329-336.. DOI: 10.1002/cncr.21622

5. Kaanders JHAM, van der Kogel AJ, Ang KK. Altered fractionation: limited by mucosal reactions?. Radiother Oncol 1998;50:247-260. DOI: 10.1016/s0167-8140(99)00012-2

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(2):161-168. DOI: 10.1016/s0167-8140(03)00231-7

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(2):625-634. DOI: 10.1016/j.ijrobp.2008.02.008

8. Strigari L, Pedicini P, D'Andrea M, Pinnarò P, Marucci L, Giordano C, 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(5):e697-702. DOI: 10.1016/j.ijrobp.2012.02.004

9. Dean JA, Wong KH, Welsh LC, Jones AB, Schick U, Newbold KL, 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(1):21-27. DOI: 10.1016/j.radonc.2016.05.015

535

10. Dean JA, Welsh LC, Wong KH, Aleksic A, Dunne E, Islam MR, 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(4):263-273. DOI: 10.1016/j.clon.2016.12.001

540 11. Dean JA, Wong KH, Gay H, Welsh LC, Jones AB, Schick U, Oh JH, Apte A, Newbold KL, Bhide SA, Harrington KJ, Deasy JO, Nutting CM, Gulliford SL. 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(4):820-831.

12. Dörr W. Three A's of repopulation during fractionated irradiation of squamous epithelia:
545 Asymmetry loss, Acceleration of stem-cell divisions and Abortive divisions. Int J Radiat Biol. 1997;72(6):635-643. DOI: 10.1080/095530097142780

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-
550 1172. DOI: 10.1080/09553002.2020.1787540

14. Fenwick JD. Delay differential equations and the dose-time dependence of early radiotherapy reactions. Med Phys. 2006;33(9):3526-3540. DOI: 10.1118/1.2241995

15. Schneider U. Mechanistic model of radiation-induced cancer after fractionated radiotherapy using the linear-quadratic formula. Med Phys 2009;36(4):1138-43. DOI: 10.1118/1.3089792

16. Kuperman VY, Lubich LM. Effect of reoxygenation on hypofractionated radiotherapy of
555 prostate cancer. Med Phys 2020;47(10):5383-5391. DOI: 10.1002/mp.14343.

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(24):245015. DOI: 10.1088/1361-6560/aba21d

560

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(12):5577-5585. DOI: 10.1002/mp.13229.

19. Fowler JF. The linear-quadratic formula and progress in fractionated radiotherapy. Br J Radiol. 565   1989;62(740):679-694. DOI: 10.1259/0007-1285-62-740-679

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. DOI: 10.1186/s13014-018-0973-6

570

21. Gruber S, Dörr W. Tissue reactions to ionizing radiation - Oral mucosa. Mutat Res 2016;770:292-298. DOI: 10.1016/j.mrrev.2016.07.007

22. Villa A, Sonis ST. Mucositis: pathobiology and management. Curr Opin Oncol. 2015;27(3):159- 575   164. DOI: 10.1097/CCO.0000000000000180

23. Rutkowska E, Baker C, Nahum A. Mechanistic simulation of normal-tissue damage in radiotherapy - implications for dose-volume analyses. Phys Med Biol 2010;55:2121-36. DOI: 10.1088/0031-9155/55/8/001

580

24. Press WH, Teukolsky SA, Vetterling WT, and Flannery BP. Numerical Recipes: The Art of Scientific Computing, Cambridge University Press 2007. DOI: 10.1145/1874391.187410

25. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by Simulated Annealing. Science 1983;220:671-680. DOI: 10.1126/science.220.4598.671 585

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 Feb;38(2):89-101. 590   DOI: 10.1016/0167-8140(95)01689-9.

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(4):497-505. DOI:

10.1080/09553000010022391

595

28. Burman P. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. Biometrika. 1989;76(3):503–514 DOI: 10.2307/2336116

29. Fachal L, Gómez-Caamaño A, Barnett GC, Peleteiro P, Carballo AM, Calvo-Crespo P, et al. A

600   three-stage genome-wide association study identifies a susceptibility locus for late radiotherapy toxicity at 2q24.1. Nat Genet 2014;46:891-894. DOI: 10.1038/ng.3020

30. Barnett GC, Thompson D, Fachal L, Kerns S, Talbot C, Elliott RM, A genome wide association study (GWAS) providing evidence of an association between common genetic variants and late

605   radiotherapy toxicity. Radiother Oncol 2014;111:178-185. DOI: 10.1016/j.radonc.2014.02.012

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. DOI: 10.1088/0031-9155/49/20/012

610   32. Wang JZ, Huang Z, Lo SS, Yuh WTC, and 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. DOI: 10.1126/scitranslmed.3000864

33. Meade S, Sanghera P, Glaholm J, Hartley A. Models of acute mucosal tolerance to radiotherapy

615   alone applied to synchronous chemoradiation schedules in head and neck cancer. Tumor Biology 2014;35:2017-2023. DOI: 10.1007/s13277-013-1267-0

Figure captions

620

Figure 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).

Figure 2: Dynamics of populations of cells versus time for three different treatments in our dataset:

625   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 1 (*criterion 1*).

630

Figure 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 1 (*criterion 1*).

635

Figure 4: BED versus 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

640   have been converted to BED values.

Figure 5: Plots of different combinations of parameters (perturbations around values reported in Table 1) 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

645   that only two parameters are shown in each panel, but *Fowler* and *Criterion 1* have more degrees of freedom.