Mathematical Medicine and Biology (2016) **33**, 211–252
doi:10.1093/imammb/dqv015
Advance Access publication on May 15, 2015

Optimal fractionation in radiotherapy with multiple normal tissues

Fatemeh Saberian* and Archis Ghate
Industrial and Systems Engineering, University of Washington, Box 352650, Seattle, WA 98195, USA
*Corresponding author. Email: saberian@uw.edu

and

Minsun Kim
Radiation Oncology, University of Washington, Box 356043, Seattle, WA 98195, USA

[Received on 3 August 2014; revised on 7 December 2014; accepted on 9 April 2015]

The goal in radiotherapy is to maximize the biological effect (BE) of radiation on the tumour while limiting its toxic effects on healthy anatomies. Treatment is administered over several sessions to give the normal tissue time to recover as it has better damage-repair capabilities than tumour cells. This is termed fractionation. A key problem in radiotherapy involves finding an optimal number of treatment sessions (fractions) and the corresponding dosing schedule. A major limitation of existing mathematically rigorous work on this problem is that it includes only a single normal tissue. Since essentially no anatomical region of interest includes only one normal tissue, these models may incorrectly identify the optimal number of fractions and the corresponding dosing schedule. We present a formulation of the optimal fractionation problem that includes multiple normal tissues. Our model can tackle any combination of maximum dose, mean dose and dose-volume type constraints for serial and parallel normal tissues as this is characteristic of most treatment protocols. We also allow for a spatially heterogeneous dose distribution within each normal tissue. Furthermore, we do not *a priori* assume that the doses are invariant across fractions. Finally, our model uses a spatially optimized treatment plan as input and hence can be seamlessly combined with any treatment planning system. Our formulation is a mixed-integer, non-convex, quadratically constrained quadratic programming problem. In order to simplify this computationally challenging problem without loss of optimality, we establish sufficient conditions under which equal-dosage or single-dosage fractionation is optimal. Based on the prevalent estimates of tumour and normal tissue model parameters, these conditions are expected to hold in many types of commonly studied tumours, such as those similar to head-and-neck and prostate cancers. This motivates a simple reformulation of our problem that leads to a closed-form formula for the dose per fraction. We then establish that the tumour-BE is quasiconcave in the number of fractions; this ultimately helps in identifying the optimal number of fractions. We perform extensive numerical experiments using 10 head-and-neck and prostate test cases to uncover several clinically relevant insights.

Keywords: intensity modulated radiation therapy; quadratically constrained quadratic programming; linear-quadratic cell survival model.

1. Introduction

In external beam radiation therapy for cancer, ionizing radiation targeted towards the cancerous region also passes through nearby normal tissue and thus damages both. The goal, therefore, is to maximize tumour damage while ensuring that the healthy anatomies can safely tolerate the resulting toxic effects of radiation. In intensity-modulated radiation therapy (IMRT) (Webb, 2010), this is attempted by following a two-pronged approach that includes a spatial and a temporal component.

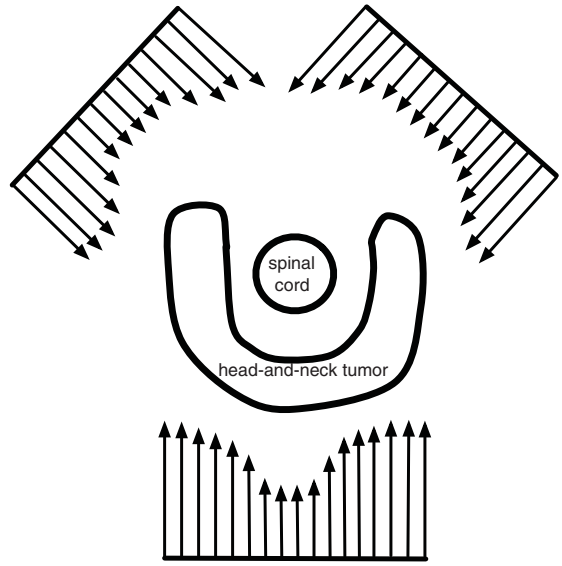

FIG. 1. A schematic showing intensity modulation to spare the spinal cord from excessive damage during head-and-neck cancer treatment. The length of each arrow in the radiation field is proportional to the intensity of the corresponding beamlet. In particular, lower intensity is assigned to beamlets that pass directly through the spinal cord. The schematic shows that radiation fields from three different directions are used to avoid areas of low radiation dose in the tumour.

The first, spatial component of this approach in essence is to prescribe a uniform or non-uniform high dose to the cancerous region and put upper limits on the dose delivered to the healthy anatomies. The intensity profile of the radiation field, which is also called the fluence-map, is then optimized so that the delivered dose is as close as possible to this tumour-conforming prescription. This is called fluence-map optimization. Figure 1 shows how the fluence-map typically relates to the relative anatomy of the tumour and the nearby healthy tissue. The fluence-map optimization problem has been studied extensively over the last two decades. Increasingly, sophisticated dose prescriptions have been proposed and efficient optimization algorithms to find the corresponding intensity profiles have been designed, tested clinically and incorporated into commercial treatment planning systems (Burman *et al.*, 1997; Shepard *et al.*, 1999; Eisbruch, 2002; Langer *et al.*, 2003; Romeijn *et al.*, 2006; Ehrgott *et al.*, 2008).

The second, temporal, component of the two-pronged approach involves breaking the total planned dose into several equal-dose fractions that are administered in multiple, well-separated treatment sessions (called fractions) spread over many weeks. This is called fractionation and it is designed mainly to give healthy anatomies sufficient time to recover between consecutive treatment sessions as healthy cells have better damage-repair capabilities than tumour cells (Withers, 1985; Hall & Giaccia, 2005). The key challenge on this temporal side of the equal-dosage fractionation problem is to find the optimal number of fractions and the corresponding dose in different fractions. Roughly speaking, using several sessions with a small dose in each session reduces the damage done to the healthy anatomies. This may enable the treatment planner to administer a much larger total dose to the tumour as compared to using only a few sessions with a higher dose per session. However, especially in fast proliferating tumours, long treatment courses may adversely affect the likelihood of adequate tumour control. Several alternative number of fractions have been clinically studied over more than a hundred years (Horiot *et al.*, 1992, 1997; Rockwell, 1998; Fu *et al.*, 2000; Garden, 2001; Ahamad *et al.*, 2005; Trotti *et al.*, 2005;

Ho *et al.*, 2009; Marzi *et al.*, 2009; Arcangeli *et al.*, 2010; Kader *et al.*, 2011). Many of these studies were conducted before the IMRT technology became widely available in the late 1990s. It is generally believed that the optimal number of fractions depends on the relative difference between the radiation dose–response of the tumour and that of the healthy anatomies, and also on the anatomy of the tumour and the healthy tissues relative to the radiation field. Currently, the most widely accepted mathematical framework for radiation dose–response is the linear-quadratic (LQ) model (Hall & Giaccia, 2005). Using this model, some researchers have mathematically explored the effect of model parameters on the optimal number of fractions. We review these studies next.

2. Limitations of existing work

To the best of our knowledge, the use of the LQ model to determine the optimal number of fractions started in the 1990s with Fowler's work (Fowler, 1990; Fowler & Ritter, 1995) and the investigation in Jones *et al.* (1995). Specifically, Fowler showed in a series of papers using insightful calculations that some treatment courses are too short and some treatment courses are too long; there is a range in between these two extremes that should include the optimal number of treatment sessions (Fowler, 1990, 2001, 2007, 2008; Fowler & Ritter, 1995). These papers consider a tumour and a single normal tissue both receiving dose d in each of the N fractions. From an IMRT perspective, the assumption that the tumour and the normal tissue receive a homogeneous and identical dose essentially ignores the spatial side of the problem. The key idea that repeatedly occurs in all papers in this area is to choose the number of fractions N and the corresponding dose d per fraction that is delivered to the tumour to maximize the biological effect (BE) on the tumour for a fixed biologically effective dose (BED) delivered to the normal tissue (Hall & Giaccia, 2005). The papers by Fowler and co-authors were arithmetic in nature in that they calculated and plotted the BE on tumour as a function of N to reach their conclusions (without rigorously proving the concavity structure of the tumour-BE as a function of N).

Jones *et al.* (1995) also considered a single normal tissue and proposed a model that was conceptually similar to Fowler's. They employed calculus to maximize the tumour-BE by setting its derivative with respect to N to equal zero. This led to a quadratic equation for the optimal dose per fraction. The positive solution of this equation was then used in turn to derive a formula for the optimal number of fractions. Thus, to the best of our knowledge, Jones *et al.* obtained the first analytical results in this area. However, they did not include a time-lag before tumour proliferation begins after treatment initiation (see Fowler, 2008 for a discussion of the importance of including this time-lag). Jones and co-authors later improved their earlier model by including such a time-lag (Armpilia *et al.*, 2004). However, these two papers by Jones and co-authors ignored the constraint that N must be an integer and did not rigorously prove that setting the derivative with respect to N to zero was a sufficient condition for global optimality in their model (i.e. they did not investigate concavity properties of tumour-BE by computing its second derivative).

Yang & Xing (2005) also studied a model similar to Fowler's but they utilized an extension of the basic LQ model to incorporate redistribution and reoxygenation of tumour cells (Brenner *et al.*, 1995) and included two normal tissues. They applied the simulated annealing heuristic to optimize the number of fractions without any guarantee of optimality. Their broad conclusions were similar to Fowler's. Bertuzzi *et al.* (2013) recently considered a model with two normal tissues and used Karush–Kuhn–Tucker (KKT) conditions to characterize the structure of optimal solutions, but the number of fractions was fixed in their model.

Mizuta *et al.* (2012) recently considered a slight generalization of Fowler's calculations, where the tumour and the single normal tissue received doses d and sd in each fraction, respectively, but they

did not consider tumour proliferation. Here s is called the sparing factor and it attempts to model the realistic possibility that the tumour and the normal tissue do not receive identical doses. They reached the intuitive conclusion that the optimal number of fractions also depends on the sparing factor s in addition to the α/β ratios of the tumour and the normal tissue in the LQ model (here α and β are parameters of the LQ model; see Section 4 for details). In fact, a similar observation was made much earlier by Jones *et al.* (1995).

Unkelbach *et al.* (2013) recently generalized the result in Mizuta *et al.* They also considered a single normal tissue but explicitly allowed for a spatially heterogeneous normal tissue dose distribution. Consequently, the analysis in Unkelbach *et al.* is a notable exception to all previous efforts towards optimizing the number of fractions. In particular, they used voxel-dependent sparing factors for each normal tissue voxel. The use of such sparing factors is justified because optimization models in radiotherapy routinely model dose as a linear transformation of the fluence-map. See Jeraj and Keall (1999), Siebers *et al.* (2001), Spirou & Chui (1998), Tian *et al.* (2013) and Webb & Oldham (1996) and Section 4 for more details about how this transformation is defined using a dose-deposition coefficient matrix. Unkelbach *et al.* derived an effective sparing factor, $s_{\text{effective}}$, which was a function only of the voxel-dependent sparing factors. The specific functional form of this effective sparing factor depends on whether the normal tissue constraint is a maximum dose constraint or a mean dose constraint. Their key conclusion was that a single fraction is optimal if $(\alpha/\beta)_{\text{tumour}} \leqslant (\alpha/\beta)_{\text{normal tissue}}/s_{\text{effective}}$ and an infinite number of fractions is optimal otherwise. A similar conclusion was also reached independently by Keller *et al.* (2012). However, and crucially, Unkelbach *et al.* and Keller *et al.* did not incorporate tumour proliferation and this was the reason for the optimality of an infinite number of fractions.

Bortfeld *et al.* (2013) extended the model in Unkelbach *et al.* by including tumour proliferation, but they did not include a time-delay before proliferation starts. Specifically, their focus was on modelling accelerated tumour re-population whereby the tumour exhibits faster growth towards the end of the treatment course. They used dynamic programming to demonstrate how dose per fraction increases over the treatment course under accelerated re-population; this was the main contribution of their paper. As an aside, they rigorously proved the correctness of a previously known closed-form formula (as e.g. in Jones *et al.*, 1995 and Armpilia *et al.*, 2004) for the optimal number of fractions under constant re-population. However, as in Unkelbach *et al.*, Bortfeld *et al.* also considered only a single normal tissue.

In fact, the inability to simultaneously include two or more normal tissues is a major practical limitation of all work in this area that derives a provably optimal number of fractions. Cancer sites always include at least two normal tissues—unspecified normal tissue and an organ at risk such as the spinal cord. In fact, in most cases, several serial and parallel normal tissues play a role in limiting the tumour-dose.

Finally, the existing works do not explicitly describe precisely how their optimal fractionation models could be implemented in practice. Specifically, it is not clear exactly how sparing factors are to be calculated and how the resulting temporal dosing schedule is to be combined with a spatially optimized fluence-map. In short, these works do not integrate their models with state-of-the-art IMRT methods. In this paper, we are able to simultaneously overcome these limitations, and yet rigorously derive optimal dosing schedules as summarized next.

3. Our contributions

We present a mathematical formulation of the optimal fractionation problem based on the LQ model using sparing factors. This formulation allows for doses to vary across fractions. It incorporates (i) any combination of several maximum dose, mean dose and dose-volume type constraints on multiple normal

tissues that receive spatially heterogeneous doses and (ii) tumour proliferation with a time-lag. Finally, since the starting point for our model is a spatially optimized fluence-map (say from any treatment planning system) it is implementable after appropriate clinical validation in the future without requiring any additional hardware or technological developments. The generality of our model introduces several challenges. In this section, we highlight our contributions in the context of these challenges.

- *Optimality of equal-dosage fractionation*: Although equal-dosage fractionation is ubiquitous in practice and in the existing literature, it is not known whether this strategy is in fact optimal in the case of multiple normal tissues. Consequently, our formulation *does not a priori* assume equal-dosage fractionation. As a result, our initial formulation is a mixed-integer, *non-convex*, quadratically constrained quadratic program (QCQP); such problems belong to the class NP-hard (Luo *et al.*, 2010). Fortunately, we are able to prove two simple sufficient conditions under which equal-dosage fractionation *is* optimal (see Propositions 5.1 and 5.3). Based on the prevalent estimates of the parameters of the LQ model, we expect these conditions to hold in many commonly studied tumours such as head-and-neck and prostate. We thus devise an equal-dosage reformulation of our problem and study it in detail in all subsequent sections.

- *Switching of the dose-limiting normal tissue*: When the problem formulation includes a single normal tissue, that normal tissue, by definition, is what limits tumour-dose. This facilitates the derivation of a simple closed-form formula for the optimal number of fractions and the corresponding dose per fraction. However, in the case of multiple normal tissues, we show that the dose-limiting normal tissue may depend on the number of fractions (see, e.g. Fig. 3(a), which demonstrates that the limiting normal tissue switches from the unspecified normal tissue to the right parotid). Furthermore, it is not evident if and when such a switch occurs as this is a function of many model parameters especially when the formulation includes several maximum dose, mean dose and dose-volume type constraints to mimic realistic anatomies and treatment protocols, and allows for heterogeneous dose distributions.

- *Non-monotonic behavior of tumour-BE*: The tumour-BE *without proliferation* is monotonic in the number of fractions in optimal fractionation problems that include a single normal tissue. We show that this is no longer true with multiple normal tissues; see, e.g. Fig. 3(b), which illustrates that the tumour-BE without proliferation can first increase and then decrease as the number of fractions is increased. In particular, *even without proliferation*, when $(\alpha/\beta)_{\text{tumour}} > (\alpha/\beta)_{\text{normal tissue}}/s_{\text{effective}}$, *a finite number of fractions may be optimal* in the case of multiple normal tissues (contrast this with the aforementioned conclusion in Unkelbach *et al.* and Keller *et al.* that with a single normal tissue, an infinite number of fractions is optimal). More generally, when a proliferation term is included, the tumour-BE is either decreasing or strictly concave in the number of fractions in the case of a single normal tissue (see Appendix B). Again, this is at the heart of the derivation of a simple closed-form formula for the optimal number of fractions and the corresponding dose per fraction. In the case of multiple normal tissues, the behavior of tumour-BE (with or without proliferation) as a function of the number of fractions is complicated owing to the possible switching of the limiting normal tissue and this has not been investigated in the existing literature.

- *Piecewise convexity of the limiting dose per fraction and quasiconcavity of tumour-BE*: We perform a rigorous mathematical analysis of our model to completely characterize how the limiting dose and tumour-BE behave as functions of the number of fractions (see Lemmas 6.1 and Theorem 6.3). Specifically, in (48), we obtain a closed-form formula for the limiting dose and show that it is a strictly decreasing, piecewise strictly convex function of the number of fractions (see Lemma 6.1

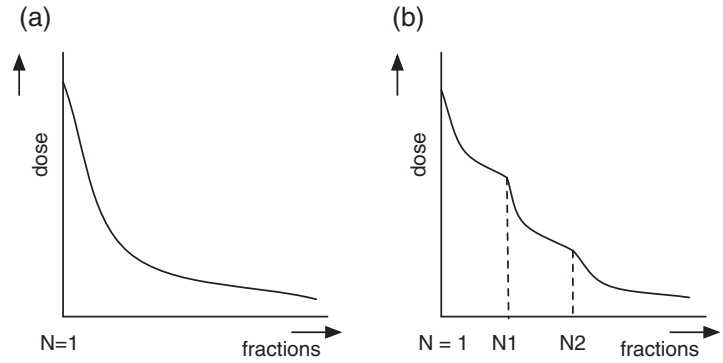

FIG. 2. A schematic illustration of the structure of the limiting dose $d^*(N)$: (a) for a single normal tissue, the limiting dose is strictly decreasing and strictly convex; (b) with multiple normal tissues, it is strictly decreasing and piecewise strictly convex; the kinks at fractions $N1$ and $N2$ represent switching of the limiting normal tissues. Also see Fig. 3(a).

and Fig. 2). This leads to a closed-form formula for the corresponding tumour-BE. Furthermore, we establish in Theorem 6.3 that this tumour-BE eventually decreases and that it is a quasiconcave (Boyd & Vandenberghe, 2004) function of the number of fractions.

- *Simple calculation procedure for problem solution*: We propose a simple calculation procedure rooted in the above quasiconcavity property to obtain an optimal number of fractions and the corresponding dose per fraction. Specifically, we calculate the tumour-BE using our closed-form formula for dose per fraction sequentially for $N = 1, 2, \ldots$ until the tumour-BE 'turns around' to correctly identify the optimal number of fractions (see Theorem 6.3 and Corollary 6.4). We note here that in the case of multiple normal tissues, it is not possible to find the optimal number of fractions in closed-form by setting a derivative to equal zero unlike the single normal tissue case as in Jones *et al.*, Bortfeld *et al.* and our Appendix B.
- *Uncovering clinically relevant insights through numerical experiments*: In Section 7, we perform extensive numerical experiments and sensitivity analyses using five head-and-neck cancer cases and five prostate cancer cases to gain further insight into the optimal fractionation problem. Some examples of such insights are listed here.
  - *Mismatch between predictions of our model with multiple normal tissues and existing models with a single normal tissue*: We show in Section 7.3 that the optimal number of fractions in our model with multiple normal tissues cannot in general be found simply by choosing from the set of optimal number of fractions obtained from separately solving single normal tissue problems one by one. This point is especially important as all existing analytical literature considers only a single normal tissue, and furthermore, the difference between the optimal number of fractions from our model and that obtained from a single normal tissue model can be rather large. Since essentially all cancer anatomies of interest include multiple normal tissues in practice, this observation provides motivation for using our model as opposed to existing single normal tissue models.
  - *Sensitivity to tumour doubling time*: We demonstrate in Section 7.4 that the optimal number of fractions increases as the tumour doubling time increases—i.e. a longer treatment schedule is optimal for slower growing tumours. We demonstrate that the sensitivity of the optimal number

of fractions to the tumour doubling time is higher when the alpha over beta ratio of the tumour is larger as compared to that of a normal tissue.

– *Effect of time-lag before proliferation starts*: We show that when the sufficient condition in Proposition 5.1 is met, it is not optimal to administer less than $1 + T_{\text{lag}}$ fractions where T_{lag} is the lag in days before tumour proliferation begins (see Remark 5.2). When the tumour doubling time is smaller than a threshold determined by T_{lag} and other model parameters, it is optimal to administer exactly $1 + T_{\text{lag}}$ fractions. When the doubling time is larger than this threshold, the optimal number of fractions is larger than $1 + T_{\text{lag}}$ but (otherwise) no longer sensitive to T_{lag}. Our numerical results in Section 7.5 further illustrate this behavior for our head-and-neck cases.

The rest of this paper is organized as follows. In the next section, we describe our problem formulation in detail. In Section 5, we prove sufficient conditions for equal-dosage fractionation. Then in Section 6, we present an equal-dosage reformulation of our original model. We then present its mathematical analysis, and propose a simple solution procedure. We present numerical results in Section 7. Finally, in Section 8, we conclude by discussing opportunities for extensions of our work here. Proofs of all results are provided in Appendix A. A previously known closed-form formula for the optimal number of fractions in the case of a single normal tissue is derived in Appendix B for the sake of completeness and notational consistency as we will need this formula in Section 7.3 when its results are compared with our multiple normal tissue model.

4. Problem formulation

Our overall methodology is to choose the number of fractions to maximize the BE of the average tumour-dose subject to constraints on normal tissue doses. All components of this methodology are described next.

4.1 *Expression for the objective function*

Let N be the number of fractions. This is the key decision variable in our optimization model. We assume that a single fraction is administered every day. As we will describe in Section 8, our model can be easily extended to tackle other fractionation schedules such as the one that administers treatment on weekdays only or the one that administers treatment twice daily on weekdays. Let T_{double} be the doubling time for the tumour (Fowler, 1990; Fowler & Ritter, 1995; Jones *et al.*, 1995) and let T_{lag} be the time-lag after which tumour proliferation starts after treatment initiation (Fowler, 2008); both these are measured in days. Finally, let α_0 and β_0 denote the parameters of the LQ model for the tumour.

In optimization models in IMRT, various anatomical regions of interest in the patient's body are discretized into equal volume cubes called voxels. Let n denote the number of such voxels in the tumour. We index these voxels by $i = 1, 2, \ldots, n$. Similarly, the radiation field is also discretized into small segments called beamlets. Let k be the number of such beamlets. We use $u \in \Re^k_+$ to denote a generic k-dimensional beamlet intensity vector (fluence-map). Let A be the $n \times k$, non-negative, tumour-dose deposition coefficient matrix. Let A_i denote its ith row. This row corresponds to the ith tumour-voxel. Then, according to the linear dose deposition model, the dose delivered to the ith voxel in the tumour equals $A_i u$ if fluence-map u is used.

We assume that a nominal fluence-map is available to the treatment planner. For example, this could be obtained by solving any of the standard fluence-map optimization models as is routinely done in commercial treatment planning systems available in hospitals. We denote this fluence-map by u_{nominal}

and it serves as an input and a starting point for our model. An important benefit of using such a fluence-map as input is that we are assured that the spatial properties of the treatment plan such as its dose-volume histograms are clinically acceptable. Let d_{nominal} denote the average tumour-dose over all tumour-voxels delivered by this nominal fluence-map. Consider any dose $d \geqslant 0$. According to the linear dose deposition model, a fluence-map that delivers an average tumour-dose of d can be constructed by scaling u_{nominal} to $u_{\text{nominal}} d/d_{\text{nominal}}$.

Now suppose that average tumour-doses $\vec{d} = (d_1, d_2, \ldots, d_N)$ are delivered over the N fractions; $\vec{d}$ is the second decision variable in our problem. According to the LQ model, the BE of delivering average tumour-doses $\vec{d}$ over N fractions equals

$$\alpha_0 \sum_{t=1}^{N} d_t + \beta_0 \sum_{t=1}^{N} d_t^2 - \tau(N). \tag{1}$$

Here, we have used the notation

$$\tau(N) = \frac{[(N-1) - T_{\text{lag}}]^+ \ln 2}{T_{\text{double}}}. \tag{2}$$

The expression $[(N-1) - T_{\text{lag}}]^+$, which is defined as $\max((N-1) - T_{\text{lag}}, 0)$, models the possibility that tumour proliferation does not start until T_{lag} days after treatment begins (Fowler, 2008). The appearance of $N-1$ instead of N in the above expression may seem surprising at first; however, note that the time that *elapses* between the first and the Nth fractions is actually $N-1$ days and hence this is the time over which the tumour could proliferate. The biological significance of the BE formula in (1) is that, according to the LQ model, different dosing schedules with identical BE are assumed to inflict the same damage on the tumour when measured in terms of the proportion of surviving cells. Note that $\tau(N) = 0$ for all N such that $N \leqslant 1 + T_{\text{lag}}$, and that $\tau(N)$ is strictly increasing and linear for all N such that $N \geqslant 1 + T_{\text{lag}}$. In particular, $\tau(N)$ is non-decreasing for all $N \geqslant 1$; it is also convex for all $N \geqslant 1$. These properties of $\tau(N)$ will be used in our mathematical analysis. We next describe the constraints in our model.

4.2 *Normal tissue constraints*

Let $O_1, O_2, \ldots, O_M$ denote the M different normal tissues under consideration. Let $\mathcal{M} = \{1, \ldots, M\}$ be the set of indices of these normal tissues. For $m \in \mathcal{M}$, let n_m denote the number of voxels in O_m. These voxels are indexed by $j = 1, 2, \ldots, n_m$. Let $\mathcal{N}_m$ denote the set $\{1, 2, \ldots, n_m\}$ of these voxels. All normal tissue voxels are assumed to have equal volume as is standard practice in IMRT. Let A^m be the $n_m \times k$, non-negative, dose-deposition coefficient matrix for O_m. Let A_j^m be the jth row of this matrix; this is the row that corresponds to the jth voxel in O_m. That is, according to the linear dose deposition model, $A_j^m u$ is the dose delivered to the jth voxel in O_m if fluence-map u is used. For each $m \in \mathcal{M}$ and $j \in \mathcal{N}_m$, let s_j^m denote the sparing factor for the jth voxel in normal tissue O_m. This sparing factor is given by

$$s_j^m = \frac{A_j^m u_{\text{nominal}}}{d_{\text{nominal}}}. \tag{3}$$

Then, according to the linear dose deposition model, the scaled fluence-map $u_{\text{nominal}} d_t / d_{\text{nominal}}$ delivers a dose of

$$\frac{A_j^m u_{\text{nominal}} d_t}{d_{\text{nominal}}} = s_j^m d_t \tag{4}$$

to voxel j in normal tissue O_m. Let α_m and β_m denote the parameters of the LQ model for normal tissue O_m, and let $\rho_m = 1/(\alpha_m/\beta_m)$ denote the inverse of the alpha over beta ratio for this normal tissue. Then, the BED corresponding to the doses $s_j^m\vec{d}$ that are delivered to voxel j in O_m over N fractions is given by

$$\sum_{t=1}^{N}(s_j^m d_t) + \rho_m \sum_{t=1}^{N}(s_j^m d_t)^2. \tag{5}$$

The above BED formula will be used in our formulation to incorporate the three most common types of constraints on normal tissue that are used in practice—maximum dose constraints for serial normal tissue, and mean dose or dose-volume type constraints for parallel normal tissue. We describe each of these in detail in the next three sections.

4.2.1 *Serial normal tissue.* Let $\mathcal{M}_1 \subseteq \mathcal{M}$ be the set of indices of serial normal tissues; these are the normal tissues whose function is hampered even when a small portion suffers damage (Marks *et al.*, 2010). Examples include spinal cord and brainstem. Suppose for any $m \in \mathcal{M}_1$ that a total dose $D_{\max}^m$ is known to be tolerated by each voxel in O_m if administered in N_{conv}^m equal-dose fractions. The BED corresponding to this schedule equals

$$\text{BED}_{\max}^m = D_{\max}^m(1 + \rho_m(D_{\max}^m/N_{\text{conv}}^m)). \tag{6}$$

Then, doses $s_j^m\vec{d}$ over N fractions can be tolerated by the jth voxel in normal tissue O_m if

$$\sum_{t=1}^{N}(s_j^m d_t) + \rho_m \sum_{t=1}^{N}(s_j^m d_t)^2 \leqslant \text{BED}_{\max}^m, \quad \forall j \in \mathcal{N}_m. \tag{7}$$

Thus, for each $m \in \mathcal{M}_1$, our problem formulation will include constraints (7).

4.2.2 *Mean dose constraints for parallel normal tissue.* Let $\mathcal{M}_2 \subseteq \mathcal{M}$ be the set of indices of parallel normal tissues with mean dose constraints; these are the normal tissues where a sufficiently small portion can be damaged without affecting the organ function (Marks *et al.*, 2010). Examples include parotid glands, heart, liver, kidneys and lungs. Suppose for any $m \in \mathcal{M}_2$ that mean dose D_{mean}^m is known to be tolerated by O_m if administered in N_{conv}^m equal-dose fractions. The BED corresponding to this mean dose is given by

$$\text{BED}_{\text{mean}}^m = D_{\text{mean}}^m(1 + \rho_m(D_{\text{mean}}^m/N_{\text{conv}}^m)). \tag{8}$$

Then, doses $s_j^m\vec{d}$ in over N fractions can be tolerated by O_m if

$$\sum_{t=1}^{N}\sum_{j=1}^{n_m}(s_j^m d_t + \rho_m(s_j^m d_t)^2) \leqslant n_m \text{BED}_{\text{mean}}^m. \tag{9}$$

As in Unkelbach *et al.* (2013), the quadratic constraint (9) can be simplified as

$$p_m \sum_{t=1}^{N} d_t + \rho_m q_m \sum_{t=1}^{N} d_t^2 \leqslant n_m \text{BED}_{\text{mean}}^m, \tag{10}$$

where we have used the notation

$$p_m \triangleq \sum_{j=1}^{n_m} s_j^m \quad \text{and} \quad q_m \triangleq \sum_{j=1}^{n_m} (s_j^m)^2. \tag{11}$$

We also define the effective sparing factor

$$s_{\text{mean}}^m \triangleq q_m / p_m, \tag{12}$$

and let

$$B_{\text{mean}}^m \triangleq q_m n_m \text{BED}_{\text{mean}}^m / (p_m)^2. \tag{13}$$

Then, multiplying both sides by the non-negative number $q_m/(p_m)^2$, inequality (10) can be rewritten as

$$\sum_{t=1}^{N} s_{\text{mean}}^m d_t + \rho_m \sum_{t=1}^{N} (s_{\text{mean}}^m d_t)^2 \leqslant B_{\text{mean}}^m. \tag{14}$$

Thus, for each $m \in \mathcal{M}_2$, our optimization model will include constraint (14).

4.2.3 *Dose-volume constraints.* Let $\mathcal{M}_3 \subseteq \mathcal{M}$ be the set of indices of normal tissues with dose-volume constraints. Examples include rectum, bladder and lung. In particular, suppose for any $m \in \mathcal{M}_3$ that no more than a volume fraction ϕ_m of normal tissue O_m can receive a dose more than D_{dv}^m if administered in N_{conv}^m fractions. The BED of total dose D_{dv}^m administered in N_{conv}^m equal-dose fractions is given by

$$\text{BED}_{\text{dv}}^m = D_{\text{dv}}^m (1 + \rho_m (D_{\text{dv}}^m / N_{\text{conv}}^m)). \tag{15}$$

Since all voxels in O_m are assumed to have equal volume, the volume fraction is the same as the voxel fraction. For each $m \in \mathcal{M}_3$ and for each voxel $j \in \mathcal{N}_m$, we thus define binary-valued functions $f_j(\vec{d}, N)$ such that

$$f_j(\vec{d}, N) = \begin{cases} 1 & \text{if } \displaystyle\sum_{t=1}^{N} (s_j^m d_t) + \rho_m \sum_{t=1}^{N} (s_j^m d_t)^2 > \text{BED}_{\text{dv}}^m, \\ 0 & \text{if } \displaystyle\sum_{t=1}^{N} (s_j^m d_t) + \rho_m \sum_{t=1}^{N} (s_j^m d_t)^2 \leqslant \text{BED}_{\text{dv}}^m. \end{cases} \tag{16}$$

We use the integer K_m to denote $\lfloor n_m \phi_m \rfloor$. Then the dose-volume constraints are written in our formulation as

$$\sum_{j=1}^{n_m} f_j(\vec{d}, N) \leqslant K_m, \quad \text{for } m \in \mathcal{M}_3. \tag{17}$$

Here, for simplicity of notation, we have assumed that there is at most one dose-volume constraint for each normal tissue. This simplifying assumption can be easily relaxed to allow for multiple dose-volume constraints on one or more of the normal tissues. In fact, we do this in our computational experiments for prostate cancer in Section 7 (see Table 4). For each $m \in \mathcal{M}_3$, our formulation will include constraint (17). We emphasize here that, to the best of our knowledge, despite the prevalence

of dose-volume constraints in practice, they have thus far not been incorporated into existing optimal fractionation models.

4.2.4 *A formulation of the optimal fractionation problem.* Putting together the various components discussed in the above sections, we formulate our optimal fractionation problem as

$$\text{(OPTFRAC1)} \quad \max_{N,\vec{d}} \quad \alpha_0 \sum_{t=1}^{N} d_t + \beta_0 \sum_{t=1}^{N} d_t^2 - \tau(N), \tag{18}$$

$$\text{subject to} \quad \sum_{t=1}^{N} (s_j^m d_t) + \rho_m \sum_{t=1}^{N} (s_j^m d_t)^2 \leqslant \text{BED}_{\max}^m,\ j \in \mathcal{N}_m,\ m \in \mathcal{M}_1, \tag{19}$$

$$\sum_{t=1}^{N} s_{\text{mean}}^m d_t + \rho_m \sum_{t=1}^{N} (s_{\text{mean}}^m d_t)^2 \leqslant B_{\text{mean}}^m,\ m \in \mathcal{M}_2, \tag{20}$$

$$\sum_{j=1}^{n_m} f_j(\vec{d}, N) \leqslant K_m,\ \text{for } m \in \mathcal{M}_3, \tag{21}$$

$$\vec{d} \geqslant 0, \tag{22}$$

$$N \geqslant 1,\ \text{integer}. \tag{23}$$

Note that for each $m \in \mathcal{M}_1$, constraint (19) is most restrictive for a voxel $j \in \mathcal{N}_m$ that has the largest sparing factor. Therefore, as in Unkelbach *et al.* (2013), we define the effective sparing factor for each $m \in \mathcal{M}_1$ as

$$s_{\max}^m \triangleq \max_{j \in \mathcal{N}_m} s_j^m. \tag{24}$$

Then, for each $m \in \mathcal{M}_1$, the group of constraints (19) is equivalent to the single constraint

$$\sum_{t=1}^{N} (s_{\max}^m d_t) + \rho_m \sum_{t=1}^{N} (s_{\max}^m d_t)^2 \leqslant \text{BED}_{\max}^m. \tag{25}$$

Now consider constraint (21) for any $m \in \mathcal{M}_3$. This dose-volume constraint requires that $\sum_{t=1}^{N} (s_j^m d_t) + \rho_m \sum_{t=1}^{N} (s_j^m d_t)^2 \leqslant \text{BED}_{\text{dv}}^m$ for at least $n_m - K_m$ voxels; this gives us the flexibility to pick and choose *which* voxels satisfy this upper bound. Since the objective function in (OPTFRAC1) is increasing in each d_t, for $t = 1, 2, \ldots, N$, the largest objective function value is obtained by enforcing this upper bound on $n_m - K_m$ voxels with the smallest sparing factors. That is, we should enforce this upper bound only on the voxels with the smallest, the second smallest, the third smallest, ... and the $(n_m - K_m)$th smallest sparing factors. But this is equivalent to enforcing the upper bound on a single voxel with the $(n_m - K_m)$th smallest sparing factor. Using standard order statistic subscript notation, we thus define the effective sparing factor for normal tissue $m \in \mathcal{M}_3$ as

$$s_{\text{dv}}^m \triangleq s_{(n_m - K_m)}^m. \tag{26}$$

Thus, constraint (21) reduces to the much simpler constraint

$$\sum_{t=1}^{N} s_{\mathrm{dv}}^{m} d_t + \rho_m \sum_{t=1}^{N} (s_{\mathrm{dv}}^{m} d_t)^2 \leqslant \mathrm{BED}_{\mathrm{dv}}^{m}. \tag{27}$$

Consequently, we rewrite problem (OPTFRAC1) equivalently as

$$\text{(OPTFRAC2)} \quad \max_{N,\,\vec{d}} \quad \alpha_0 \sum_{t=1}^{N} d_t + \beta_0 \sum_{t=1}^{N} d_t^2 - \tau(N), \tag{28}$$

$$\text{subject to} \quad \sum_{t=1}^{N} (s_{\max}^{m} d_t) + \rho_m \sum_{t=1}^{N} (s_{\max}^{m} d_t)^2 \leqslant \mathrm{BED}_{\max}^{m}, \quad m \in \mathcal{M}_1, \tag{29}$$

$$\sum_{t=1}^{N} s_{\mathrm{mean}}^{m} d_t + \rho_m \sum_{t=1}^{N} (s_{\mathrm{mean}}^{m} d_t)^2 \leqslant B_{\mathrm{mean}}^{m}, \quad m \in \mathcal{M}_2, \tag{30}$$

$$\sum_{t=1}^{N} s_{\mathrm{dv}}^{m} d_t + \rho_m \sum_{t=1}^{N} (s_{\mathrm{dv}}^{m} d_t)^2 \leqslant \mathrm{BED}_{\mathrm{dv}}^{m}, \quad \text{for } m \in \mathcal{M}_3, \tag{31}$$

$$\vec{d} \geqslant 0, \tag{32}$$

$$N \geqslant 1, \text{ integer}. \tag{33}$$

Observe that (OPTFRAC2) is a non-convex QCQP in the decision vector $\vec{d}$ and also includes the integer-valued decision variable N. As a result, exact solution of this optimization problem appears computationally challenging. Fortunately, we are able to derive two sufficient conditions under which an equal-dosage schedule is optimal to this problem. Furthermore, we expect these sufficient conditions to hold in many commonly studied tumours such as head-and-neck and prostate. This motivates the equal-dosage reformulation of (OPTFRAC2) that we study in Sections 6 and 7 of this paper.

5. Sufficient conditions for optimality of equal-dosage fractionation

Recall that the set of all normal tissues is denoted by $\mathcal{M} = \{1, \ldots, M\}$ and assume for simplicity of exposition that $\mathcal{M}_1 \cap \mathcal{M}_2 = \mathcal{M}_2 \cap \mathcal{M}_3 = \mathcal{M}_1 \cap M_3 = \emptyset$. Clinically, this means that a normal tissue can have either maximum dose constraints, or mean dose constraints, or dose-volume constraints, but cannot have two types of constraints simultaneously. However, we emphasize that by using a more complicated notation than what we will use below, this assumption can be relaxed without any mathematical difficulties. Then we define C_m to equal the right hand side in the quadratic inequality constraint in (OPTFRAC2) for normal tissue $m \in \mathcal{M}$; similarly, we define σ_m as the effective sparing factor for normal tissue $m \in \mathcal{M}$. Then (OPTFRAC2) can be compactly written as

$$\text{(OPTFRAC2)} \quad \max_{N,\vec{d}} \quad \alpha_0 \sum_{t=1}^{N} d_t + \beta_0 \sum_{t=1}^{N} d_t^2 - \tau(N) \tag{34}$$

$$\sigma_m \sum_{t=1}^{N} d_t + \sigma_m^2 \rho_m \sum_{t=1}^{N} d_t^2 \leqslant C_m, \quad m \in \mathcal{M}, \tag{35}$$

$$\vec{d} \geqslant 0, \tag{36}$$

$$N \geqslant 1, \text{ integer}. \tag{37}$$

Moreover, for any fixed integer $N \geqslant 1$, we formulate

$$(\text{OPTFRAC2}(N)) \quad \max_{\vec{d}} \quad \alpha_0 \sum_{t=1}^{N} d_t + \beta_0 \sum_{t=1}^{N} d_t^2 - \tau(N) \tag{38}$$

$$\sigma_m \sum_{t=1}^{N} d_t + \sigma_m^2 \rho_m \sum_{t=1}^{N} d_t^2 \leqslant C_m, \quad m \in \mathcal{M}, \tag{39}$$

$$\vec{d} \geqslant 0. \tag{40}$$

First note that the objective function and the constraint functions in this problem are continuous and the feasible region is bounded. Thus, by the extreme value theorem of Weierstrass (Luenberger, 1969), the problem has an optimal solution.

We next establish two conditions under which equal-dosage fractionation is optimal in (OPTFRAC2(N)).

PROPOSITION 5.1 Suppose $\alpha_0/\beta_0 \geqslant \max_{m\in\mathcal{M}}(\alpha_m/\beta_m/\sigma_m)$. Then it is optimal in (OPTFRAC2(N)) to deliver dose

$$d^* = \min_{m\in\mathcal{M}} \left(\frac{-1 + \sqrt{1 + 4\rho_m C_m/N}}{2\sigma_m \rho_m} \right) \tag{41}$$

in each one of the N fractions.

We provide two different proofs of this result in the Appendix—one is algebraic and the other uses KKT conditions. We expect the condition in this proposition to hold in tumours similar to head-and-neck cancer where the tumour α/β ratio is believed to be much bigger than that of surrounding normal tissues (Williams *et al.*, 1985; Fowler, 2007, 2008). We also emphasize that the key result of Proposition 5.1 is the conclusion that it is optimal to use an identical dose in every session. Once this is established, it is easy to derive the actual value of the dose per session by solving a quadratic equation as has been done in most of the existing literature with a single normal tissue. Finally, in the special case where there is a single normal tissue, our condition correctly reduces to the condition that $(\alpha/\beta)_{\text{tumour}} \geqslant (\alpha/\beta)_{\text{normal tissue}}/s_{\text{effective}}$ as in the existing literature.

REMARK 5.2 We note that if $\alpha_0/\beta_0 > \max_{m\in\mathcal{M}}(\alpha_m/\beta_m/\sigma_m)$ as in Proposition 5.1 then it is not optimal to administer less than $1 + T_{\text{lag}}$ fractions. This is because the tumour-BE $g^*(N) = \min_{m\in\mathcal{M}} g_m(N)$ as defined later in (A.62) is increasing over the range $1 \leqslant N < 1 + T_{\text{lag}}$. This holds because each $g_m(N)$ has this same property as shown in our proof of Lemma B.1 in the case of a single normal tissue. A further discussion of this issue is included in Section 7.5.

PROPOSITION 5.3 Suppose $\alpha_0/\beta_0 \leqslant \min_{m\in\mathcal{M}}(\alpha_m/\beta_m/\sigma_m)$. Then it is optimal in OPTFRAC2(N) to deliver dose

$$d^* = \min_{m\in\mathcal{M}} \left(\frac{-1 + \sqrt{1 + 4\rho_m C_m}}{2\sigma_m \rho_m} \right) \tag{42}$$

in any one of the N fractions and set the doses in the other $N - 1$ fractions to zero. Since this dosing schedule is optimal for every $N \geqslant 1$, it is optimal to use a single fraction in (OPTFRAC2).

We provide two different proofs of this result in the Appendix—one is algebraic and the other uses KKT conditions. We expect this condition to hold in tumours similar to prostate cancer where the tumour α/β ratio is believed to be no larger than that of surrounding normal tissues (Fowler *et al.*, 2001; Wang *et al.*, 2003; Dasu, 2007). Thus, the clinical significance of this proposition is that, based on the LQ model, it suggests using short treatment schedules for tumours that behave similar to prostate cancer even in the presence of multiple normal tissues. The key contribution of Proposition 5.3 is the conclusion that it is optimal to deliver a single dose. Once this is established, it is again easy to derive the value of this dose by solving a quadratic equation. Finally, in the special case where there is a single normal tissue, our condition correctly reduces to the condition that $(\alpha/\beta)_{\text{tumour}} \leqslant (\alpha/\beta)_{\text{normal tissue}}/s_{\text{effective}}$ as in the existing literature.

6. Optimal equal-dosage fractionation

We now focus on an equal-dosage variation of (OPTFRAC2). When the condition in Proposition 5.1 holds, this equal-dosage variation yields an optimal solution to (OPTFRAC2). When the condition in Proposition 5.3 holds, no further calculations are necessary as we know that it is optimal to administer a single fraction with dose given by (42). The equal-dosage variation is obtained by setting $d_1 = d_2 = \ldots = d_N$ and calling this dose d. Thus, (OPTFRAC2) now reduces to the two-variable problem

$$(P) \quad \max_{N,d} \ \alpha_0 N d + \beta_0 N d^2 - \tau(N) \tag{43}$$

$$N(\sigma_m d + \rho_m(\sigma_m^2 d^2)) \leqslant C_m, \quad m \in \mathcal{M}, \tag{44}$$

$$d \geqslant 0, \tag{45}$$

$$N \geqslant 1, \ \text{integer}. \tag{46}$$

Because $\sigma_m d \geqslant 0$, the quadratic constraints (44) can be equivalently written as the linear constraints

$$d \leqslant b_m(N) \triangleq \frac{-1 + \sqrt{1 + 4\rho_m C_m/N}}{2\sigma_m \rho_m}, \quad m \in \mathcal{M}. \tag{47}$$

Moreover, since the objective function in (P) is increasing in d for each fixed $N \geqslant 1$, we can eliminate d from (P) by simply substituting the largest possible value of d that is allowed by these linear constraints. We denote this largest possible value of d by $d^*(N)$ and note that it is given by

$$d^*(N) \triangleq \min_{m \in \mathcal{M}} b_m(N). \tag{48}$$

For each $N \geqslant 1$, we call dose $d^*(N)$ and any corresponding normal tissue (that achieves the minimum in (48)) 'limiting'. Now, with $d^*(N)$ as defined in (48) at our disposal, problem (P) simplifies to

$$(P) \ \max \ g^*(N) \triangleq \alpha_0 N(d^*(N)) + \beta_0 N(d^*(N))^2 - \tau(N) \tag{49}$$

$$N \geqslant 1, \ \text{integer}. \tag{50}$$

In practice, there is an implicit upper bound $N_{\max}$ on the number of fractions that can be administered based on logistical considerations (Fowler & Ritter, 1995). Consequently, the optimal number of fractions can be found by evaluating $g^*(N)$ defined in (49) for every integer $1 \leqslant N \leqslant N_{\max}$ and then picking the best. In Theorem 6.3 in the next section, we go beyond this brute-force yet practical approach and

perform a careful mathematical analysis of the structure of the dose $d^*(N)$ in (48) and the corresponding BE $g^*(N)$ in (49). This provides additional insights into problem (P) and leads to a more streamlined procedure for solving (49 and 50).

6.1 *Characterization of the limiting dose and the optimal tumour-BE as functions of the number of fractions*

LEMMA 6.1 The limiting dose $d^*(N)$ defined in (48) is strictly decreasing in N. When viewed as a function of all real numbers $N \geqslant 1$, $d^*(N)$ is piecewise strictly convex. See Fig. 2.

The rate at which $d^*(N)$ decreases affects the structure of $g^*(N)$. Our next lemma puts an N-dependent upper bound on $d^*(N)$.

LEMMA 6.2 There exists a positive bounding constant $\bar{B}$ such that $d^*(N) \leqslant \bar{B}/N$ for all $N \geqslant 1$.

The next theorem is a key result of this paper as it characterizes the structure of the optimal tumour-BE $g^*(N)$.

THEOREM 6.3 The BE $g^*(N)$ in the objective function (49) of (P) has the following properties:

Property 1. there exists an $\hat{N} \geqslant 2$ such that $g^*(\hat{N}-1) > g^*(\hat{N})$;

Property 2. $g^*(N)$ is either (i) non-increasing in N for all integers $N \geqslant 1$ or (ii) non-decreasing in N up to a positive integer and then non-increasing after this positive integer; and

Property 3. suppose $\hat{N} \geqslant 2$ is the smallest fraction N with the property that $g^*(N-1) > g^*(N)$ (note that such an $\hat{N}$ exists by Property 1), then $\hat{N}-1$ is the optimal number of fractions.

The above theorem motivates the following streamlined procedure for solving problem (P).

A calculation procedure for solving (P)

Step 1. Set $N = 1$.
Step 2. Compute the optimal dose per fraction $d^*(N)$ using (48). Use this $d^*(N)$ to compute the BE $g^*(N)$ as in (49).
Step 3. If $N = 1$, then set $N = 2$ and go to Step 2 above.
Step 4. If $N > 1$ then STOP if $g^*(N-1) > g^*(N)$, report $N-1$ as the optimal number of fractions, $d^*(N-1)$ as the optimal average tumour-dose per fraction, and finally, $u_{\text{nominal}} d^*(N-1)/d_{\text{nominal}}$ as the fluence-map as per the linear dose deposition model; otherwise, set $N = N+1$ and go to Step 2 above.

The following corollary of Theorem 6.3 establishes the correctness of this procedure.

COROLLARY 6.4 The stopping condition in Step 4 of the above procedure is eventually met; i.e. the procedure terminates finitely. Moreover, it correctly identifies the optimal number of fractions in (P).

We acknowledge the possibility that neither the condition in Proposition 5.1 nor that in Proposition 5.3 holds; i.e. $\min_{m\in\mathcal{M}}(\alpha_m/\beta_m/\sigma_m) < \alpha_0/\beta_0 < \max_{m\in\mathcal{M}}(\alpha_m/\beta_m/\sigma_m)$. We provide an example here.

$$\max \quad d_1 + 0.2d_1^2 + d_2 + 0.2d_2^2 \tag{51}$$

$$d_1 + 0.1666d_1^2 + d_2 + 0.1666d_2^2 \leqslant 44.8762 \tag{52}$$

$$d_1 + 0.3571d_1^2 + d_2 + 0.3571d_2^2 \leqslant 79.5918 \tag{53}$$

$$d_1, d_2 \geqslant 0. \tag{54}$$

It turns out that this problem has exactly two optimal solutions: the first where $d_1 \approx 1.04, d_2 \approx 13.46$ and the second where $d_1 \approx 13.46, d_2 \approx 1.04$. Thus, there is no equal-dosage solution that is optimal. Also, it is not optimal to set one dose to zero and the other to a positive value. Note that this scenario cannot arise when there is a single normal tissue as in the existing literature. In these situations, our simple procedure for obtaining an optimal solution to the equal-dosage formulation (P) yields an approximate solution to (OPTFRAC2).

7. Numerical results

We are now ready to present numerical experiments and sensitivity analyses. We first describe in detail the test cases that we used in our experiments. Then, in Section 7.2, we demonstrate that the limiting normal tissue can switch as the number of fractions is varied and that, consequently, tumour-BE without proliferation may not be monotonic. In Section 7.3, we illustrate that the optimal number of fractions cannot be found by solving separate optimization problems for different normal tissues one by one. Then in Section 7.4, we investigate the sensitivity of the optimal number of fractions and the optimal tumour-BE to T_{double}. We find that the optimal number of fractions is smaller for smaller values of T_{double}, which represent faster growing tumours. Also, the optimal tumour-BE itself is smaller for these faster growing tumours. We demonstrate that the optimal number of fractions can be more sensitive to T_{double} when the tumour alpha over beta ratio is larger relative to that of the normal tissues. We also find that for each fixed value of T_{double}, and especially when this value is small, the range for optimal number of fractions over all our head-and-neck cases and for a wide range of other problem parameters is quite narrow. In Section 7.5, we study the sensitivity of the optimal number of fractions to T_{lag}. We find that it is insensitive to T_{lag} other than the fact that $1 + T_{\text{lag}}$ provides a lower bound on the optimal number of fractions as in Remark 5.2 under the sufficient condition in Proposition 5.1. We comment that to avoid overcrowding in our figures and in the interest of brevity, we have included below plots for a single representative case; we believe that this is sufficient because the qualitative trends in all cases were similar.

7.1 *Description of test cases*

In this section, we first describe the ten test cases that were used in our numerical experiments. Five of these were head-and-neck cases and the other five were prostate. All test cases were generated using our in-house software Phantom Creator (PhanC) written in MATLAB (Saberian & Kim, 2014). These test cases were three-dimensional and were carefully developed to be representative of clinical scenarios in terms of geometry and size. Specifically, our test cases were similar to those discussed in Gu *et al.* (2009), Men *et al.* (2009), Romeijn *et al.* (2006), Romeijn & Dempsey (2008) and Salari & Romeijn

TABLE 1 *Description of the geometry used in head-and-neck cancer cases.*

Case #	# of beamlets (k)	# of tumour-voxels (n)	# of normal tissues voxels
1	3910	27576	67386
2	3888	31930	67270
3	4128	36320	76160
4	3003	22372	53176
5	3256	28638	64713

TABLE 2 *Tolerance doses for various normal tissues in our head-and-neck test cases where the dose is administered in $N_{\text{conv}} = 35$ equal-dose fractions. Recall that for dose-volume type constraints no more than a volume fraction ϕ of the normal tissue can receive dose more than D_{dv}.*

Normal tissue	$D_{\max}$ (Gy)	D_{mean} (Gy)	D_{dv} (Gy), ϕ
Spinal cord	45	N/A	N/A
Brainstem	50	N/A	N/A
Left and right parotids	N/A	28	N/A
Unspecified normal tissue	77	N/A	70, 0.05

(2012). All cases used equally spaced coplanar beams and the beamlet resolution was $5 \times 5\,\text{mm}^2$. All voxels were $3 \times 3 \times 3\ \text{mm}^3$. All computer simulations were carried out on a 3.1 GHz iMac desktop with 16 GB RAM using MATLAB.

7.1.1 *Head-and-neck cancer cases.* All cases used seven beams and included spinal cord, brainstem, left and right parotids and unspecified normal tissue between these critical organs. The total number of voxels in the head-and-neck target and in the normal tissues, and the total number of beamlets is shown in Table 1.

The conventional fractionation schedule was assumed to include $N_{\text{conv}} = 35$ fractions. While formulating problem (P), we included maximum dose constrains for spinal cord, brainstem and unspecified normal tissue. A dose-volume constraint for unspecified normal tissue was also added. Mean dose constrains were used for left and right parotids. The tolerance dose values for various normal tissues were similar to Emami *et al.* (1991), Kehwar (2005), Marks *et al.* (2010) and Mavroidis *et al.* (2011) and are listed in Table 2.

We needed a nominal fluence-map u_{nominal} as a starting point to calculate the sparing factors for different normal tissues as explained in Section 4.2. In practice, a spatially optimized fluence-map that conforms to an individual hospital's treatment protocol would be readily available from a commercial treatment planning system and could be used for this purpose. We obtained u_{nominal} by solving a standard fluence-map optimization problem where the objective was to minimize the total squared-deviation of tumour-voxel doses from a prescription-dose of 70 Gy subject to maximum dose constraints for spinal cord and brainstem and mean dose constraints for parotids with dose tolerance levels as in Table 2 (see, e.g. Shepard *et al.*, 1999). To ensure that the intensity profile would be deliverable in practice using a multi-leaf collimator, we put linear smoothness constraints on each radiation field. These smoothness constraints ensured that the absolute percentage difference between intensity values of two adjacent

TABLE 3 *Description of the geometry used in prostate cancer cases.*

Case #	# of beamlets (k)	# of tumour-voxels (n)	Number of normal tissues voxels
1	938	6180	145545
2	847	7225	326703
3	935	4628	367656
4	930	4956	314544
5	870	4840	269450

beamlets was smaller than a threshold. The resulting optimization problem was convex and it was solved using the MATLAB toolbox CVX (Grant & Boyd, 2009).

Head-and-neck tumour was assumed to have $\alpha = 0.35\,\text{Gy}^{-1}$ based on Fowler (1990, 2001, 2007) and Fowler (2008). For unspecified normal tissue, α/β value was fixed at 3 Gy based on Fowler (1984), Joiner and van der Kogel (2009) and Thames *et al.* (1989). Our model had four remaining groups of parameters: α/β ratio for tumour; α/β ratios for spinal cord, brainstem, left and right parotids; and T_{double} and T_{lag} for tumour. We performed sensitivity analyses with several values for these parameters. The α/β ratio for head-and-neck tumour was selected from the set $\{8, 10, 12\}$ based on Fowler (2007), Qi *et al.* (2012) and Williams *et al.* (1985). The α/β ratios for left and right parotids were assumed to be equal and the value of this ratio was selected from the set $\{3, 4, 5, 6\}$ based on Fowler (1990, 2007) and Yang & Xing (2005). The α/β ratios for spinal cord and brainstem were chosen from the set $\{2, 3, 4, 5, 6\}$ based on Fowler (1990, 2007) and Yang & Xing (2005). The doubling time T_{double} for head and neck was selected from the set $\{2, 3, 5, 8, 10, 20, 40, 50\}$ based on Fowler (1990, 2007), Qi *et al.* (2012) and Yang & Xing (2005); and T_{lag} was selected from the set $\{7, 14, 21, 28, 35\}$ based on Fowler (2007).

7.1.2 *Prostate cancer cases.* All cases used five beams and included rectum, bladder, left and right femurs and unspecified normal tissue between these critical organs. The total number of voxels in the prostate target and in the normal tissues, and the total number of beamlets is shown in Table 3.

The conventional fractionation schedule was assumed to include $N_{\text{conv}} = 45$ fractions. In our formulation of problem (P), we included a maximum dose constraint for unspecified normal tissue, and included dose-volume constraints for all normal tissues. There were no mean dose constraints. The dose-volume constraints for all normal tissues were similar to Emami *et al.* (1991), Kehwar (2005), Marks *et al.* (2010) and Mavroidis *et al.* (2011) and are listed in Table 4.

We needed a nominal fluence-map u_{nominal} to calculate sparing factors for different normal tissues. As in the head-and-neck cases, u_{nominal} was obtained by solving a standard fluence-map optimization problem where the goal was to minimize the total squared-deviation of tumour-voxel doses from a prescription-dose of 81 Gy with maximum dose constraints of 85, 89 and 65 Gy for rectum, bladder, and femurs, respectively. Smoothness constraints were also included as explained in the head-and-neck case. Again, the resulting optimization problem was convex and it was solved using the MATLAB toolbox CVX (Grant & Boyd, 2009).

Prostate tumour was assumed to have $\alpha = 0.15\,\text{Gy}^{-1}$. For unspecified normal tissue, α/β value was fixed at 3 Gy again based on Fowler (1984), Joiner and van der Kogel (2009) and Thames *et al.* (1989). After fixing these values, our model had four remaining groups of parameters: α/β ratio for tumour; α/β ratios for rectum, bladder and femurs; and T_{double} and T_{lag} for tumour. We performed sensitivity analyses with several values for these parameters. The α/β ratio for prostate was selected from the set

TABLE 4 *Dose-volume constraints for various normal tissues when dose is administered in $N_{\text{conv}} = 45$ equal-dose fractions—no more than a volume fraction ϕ of the normal tissue can receive dose more than D_{dv}.*

Normal tissue	D_{dv} (Gy), ϕ
Rectum	50, 0.5
	85, 0.05
Bladder	50, 0.5
	70, 0.35
	75, 0.25
	80, 0.15
	89, 0.05
Femurs	65, 0.05
Unspecified normal tissue	81, 0.05

$\{2, 3, 4, 6\}$ based on Brenner & Hall (1999), Fowler *et al.* (2001), Fowler & Ritter (1995) and Wang *et al.* (2003). The α/β ratios for left and right femurs were assumed to be equal and the value of this ratio was selected from the set $\{3, 4, 5, 6\}$ (Fowler & Ritter, 1995; Brenner, 2004; Joiner and van der Kogel, 2009). The α/β ratios for rectum and bladder were chosen to be equal and from the set $\{3, 4, 5, 6\}$ again based on Brenner (2004), Fowler & Ritter (1995), Joiner and van der Kogel (2009) and Marzi *et al.* (2009). The doubling time T_{double} for prostate was selected from the set $\{5, 20, 40, 60, 80\}$ based on Fowler & Ritter (1995), Fowler *et al.* (2001), Gao *et al.* (2010) and Haustermans (1997); and T_{lag} was selected from the set $\{7, 14, 21, 28, 35\}$ based on Gao *et al.* (2010) and Haustermans (1997).

7.2 *Switching of the limiting normal tissues and non-monotonic behavior of the tumour-BE without proliferation*

As stated in Section 1, the dose-limiting normal tissue of course does not change with the number of fractions when the optimization model includes only one normal tissue to begin with. We demonstrate by example here that this is no longer the case when the optimization model includes multiple normal tissues. In particular, Fig. 3(a) plots the limiting doses corresponding to individual normal tissues as functions of N. The overall limiting dose itself is the minimum of such individual limiting doses. The figure shows that the limiting normal tissue switches from the unspecified normal tissue to the right parotid as the number of fractions is varied from 35 to 45 in this case. The qualitative trend in this figure is similar to our schematic in Fig. 2(b) and is consistent with Lemma 6.1. This type of switching between limiting normal tissues is essentially why it is not possible to develop a closed-form formula for the optimal number of fractions with multiple normal tissues. A related point is that, as stated in Section 1, the tumour-BE without proliferation is either increasing or decreasing in N when the problem formulation includes only one normal tissue. We demonstrate in Fig. 3(b) that this is no longer the case when the optimization model includes multiple normal tissues. In fact, the figure illustrates that whether or not the tumour-BE without proliferation is monotonic depends on problem parameters. Moreover, since some believe (Fowler *et al.*, 2001; Wang *et al.*, 2003) that the α/β ratio for prostate tumours could be quite low, and in particular, it could be smaller than that of nearby normal tissues, there are even more possibilities for the qualitative trend in prostate tumour-BE as a function of the number of fractions. These are illustrated in Fig. 4(a and b).

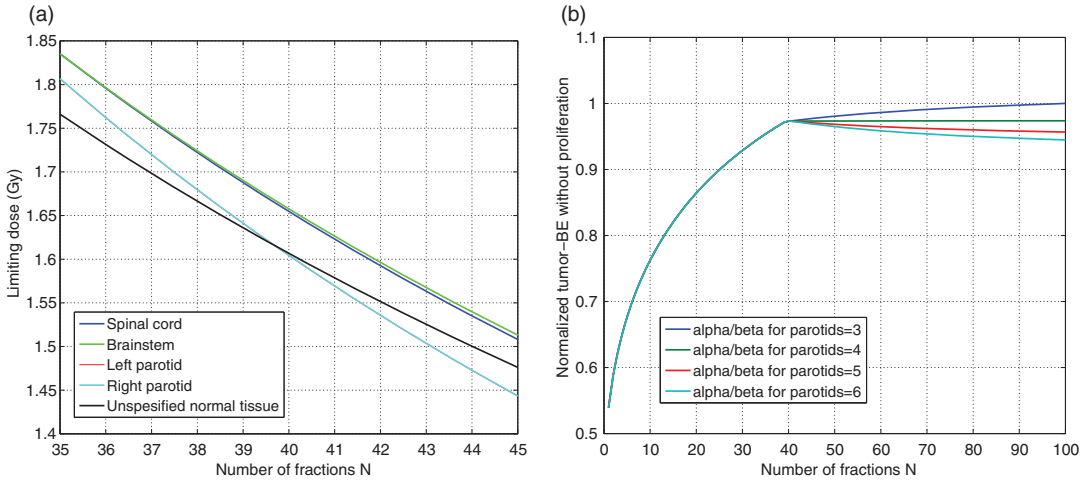

FIG. 3. This figure was generated for our head-and-neck case 1 with tumour $\alpha/\beta = 8$ Gy. The α/β ratios for all normal tissues were fixed at 3 Gy, except for the parotids. (a) Switching of the limiting normal tissue with varying number of fractions; the α/β ratio for the parotids was fixed at 6 Gy; (b) Monontonic and non-monotonic behavior of the tumour-BE without proliferation with varying number of fractions for different values of the α/β ratio for the parotids. Also note that the kink in the tumour-BE without proliferation occurs exactly where there is a switch in the limiting normal tissue.

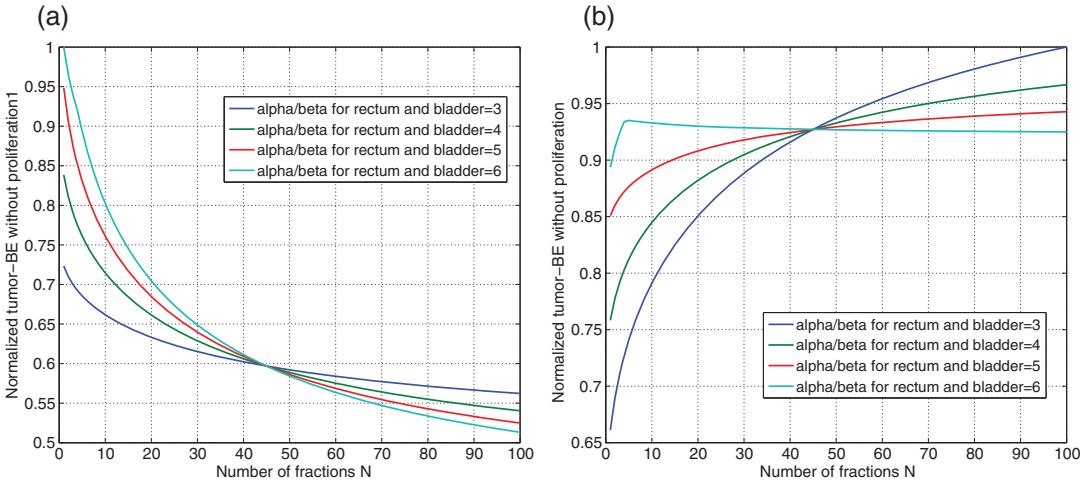

FIG. 4. This figure was generated for our prostate case 2. The α/β ratios for all normal tissues were fixed at 3 Gy, except for rectum and bladder for which the ratios were fixed as shown in the figure. (a) The α/β ratio for the prostate tumour was fixed at 2 Gy; (b) The α/β ratio for the prostate tumour was fixed at 6 Gy. When the α/β for the rectum and the bladder was fixed at 6 Gy, the limiting normal tissue switched at $N = 4$ as suggested by the kink in the corresponding tumour-BE plot; there was no switching for the other values of this α/β ratio.

7.3 *Comparison with results obtained by separately considering each normal tissue one by one*

Since a closed-form formula for the optimal number of fractions is available for the single normal tissue case (see, e.g. Jones *et al.*, 1995; Armpilia *et al.*, 2004; Bortfeld *et al.*, 2013 and other references

TABLE 5 *A comparison of the optimal number of fractions obtained from our model with the optimal number of fractions that would have resulted if we had used a model with a single normal tissue and then had applied our results in Appendix* B. *The table was generated for our head-and-neck case 1. The α/β ratio for the tumour was fixed at* 10 Gy; *the α/β ratios for spinal cord and brainstem were fixed at* 3 Gy. *The α/β ratio for the parotids was varied as shown in the table. T_{lag} was set at* 7 *days.*

Parotid $\alpha/\beta = 3$ Gy		T_{double} (days)							
		2	3	5	8	10	20	40	50
	Our model	12	20	35	39	39	44	71	82
One normal tissue	Spinal cord	9	14	23	34	41	70	112	130
	Brainstem	9	15	25	38	46	79	126	146
	Left parotid	8	8	14	22	26	44	71	82
	Right parotid	8	8	14	22	26	44	71	81
	Unspec. normal tissue (max)	12	20	35	55	67	117	192	223
	Unspec. normal tissue (dv)	12	20	34	52	63	109	178	206
Parotid $\alpha/\beta = 4$ Gy		T_{double} (days)							
		2	3	5	8	10	20	40	50
	Our model	12	20	35	39	39	40	40	45
One normal tissue	Spinal cord	9	14	23	34	41	70	112	130
	Brainstem	9	15	25	38	46	79	126	146
	Left parotid	8	8	8	11	13	23	39	45
	Right parotid	8	8	8	11	13	23	39	45
	Unspec. normal tissue (max)	12	20	35	55	67	117	192	223
	Unspec. normal tissue (dv)	12	20	34	52	63	109	178	206
Parotid $\alpha/\beta = 5$ Gy		T_{double} (days)							
		2	3	5	8	10	20	40	50
	Our model	12	20	35	39	40	40	40	40
One normal tissue	Spinal cord	9	14	23	34	41	70	112	130
	Brainstem	9	15	25	38	46	79	126	146
	Left parotid	1	1	1	1	1	1	8	8
	Right parotid	1	1	1	1	1	1	1	8
	Unspec. normal tissue (max)	12	20	35	55	67	117	192	223
	Unspec. normal tissue (dv)	12	20	34	52	63	109	178	206

therein, as well as Appendix B), the question arises as to whether the optimal number of fractions in the multiple normal tissue case can be obtained somehow by separately applying the closed-form formula to each normal tissue one by one. We show in Table 5 that this is not the case. In particular, the columns corresponding to different values of T_{double} in Table 5 with parotid $\alpha/\beta = 3$ Gy show that there is no single normal tissue that can always correctly predict the optimal number of fractions obtained from our model. For instance, when $T_{\text{double}} = 2, 3, 5$ days, unspecified normal tissue correctly predicts the number of fractions from our model; when $T_{\text{double}} = 8$ days, brainstem best-predicts the number of fractions from

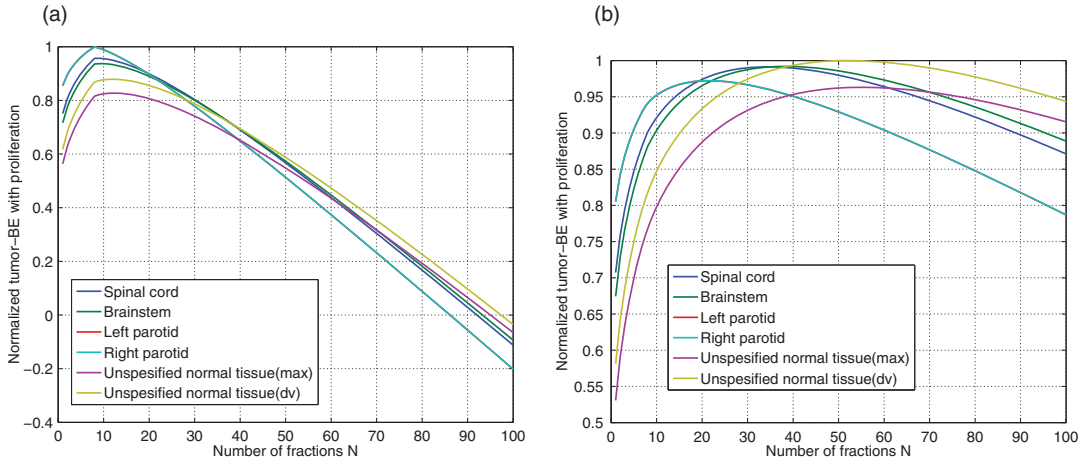

FIG. 5. Tumour-BE with proliferation obtained by considering different normal tissues one by one. The actual tumour-BE with proliferation is given by the lower-envelop of all curves. These figures were generated using our head-and-neck case 1. Tumour α/β ratio was fixed at 10 Gy and normal tissue α/β ratios were fixed at 3 Gy; T_{lag} was fixed at 7 days. (a) T_{double} was 2 days. In this case, the optimal number of fractions obtained from our model happened to equal the number that would have resulted if only unspecified normal tissue with a maximum dose constraint were included in a single normal tissue model. More generally speaking, the maximizer of the lower-envelop occurs at the same location as that of one of the normal tissues. (b) T_{double} was 8 days. Here, the maximizer of the lower-envelop occurs far from the maximizers of any of the individual normal tissue curves.

our model; when $T_{\text{double}} = 10$ days, spinal cord is the best predictor; and finally, when $T_{\text{double}} = 20, 40, 50$ days, parotids are the best predictors. Moreover, in some cases, the optimal number of fractions from our model does not equal any of the numbers obtained from the single normal tissue model; for instance, see the column of $T_{\text{double}} = 20$ with parotid $\alpha/\beta = 4$ Gy, and the columns of $T_{\text{double}} = 20, 40, 50$ with parotid $\alpha/\beta = 5$ Gy.

The intuition behind why models with a single normal tissue cannot predict the optimal number of fractions from our model is illustrated in Fig. 5. In particular, the actual tumour-BE as a function of N in our model is the 'lower-envelop' of the tumour-BEs from the group of models that include a single normal tissue one by one as formalized in (A.62). As a result, the maximizer of this lower-envelop can be far from the maximizers of the individual tumour-BE curves for different normal tissues.

7.4 *Sensitivity to tumour doubling time*

It is known from previous work on optimization models with a single normal tissue that shorter treatment courses would be better when the tumour grows fast as characterized by smaller values of the tumour doubling time T_{double} (see, e.g. Fowler, 1990, Fig. 1). We observed this trend with multiple normal tissues as well. As an example, the tumour-BE with proliferation is plotted against N for various values of T_{double} in Figs 6(a and b) for our head-and-neck case 1 and prostate case 1. The figures show that the optimal number of fractions is smaller for faster growing tumours. Also, the optimal tumour-BE itself is smaller for faster growing tumours.

The figures also show that the tumour-BE can sometimes grow very slowly with N before it starts decreasing. When this is the case, the benefit of continuing treatment beyond a certain point, all the way up to the optimal number of fractions, is negligible especially when compared with other factors such as financial cost of treatment and logistical inconvenience to the patient. For each T_{double} value, we

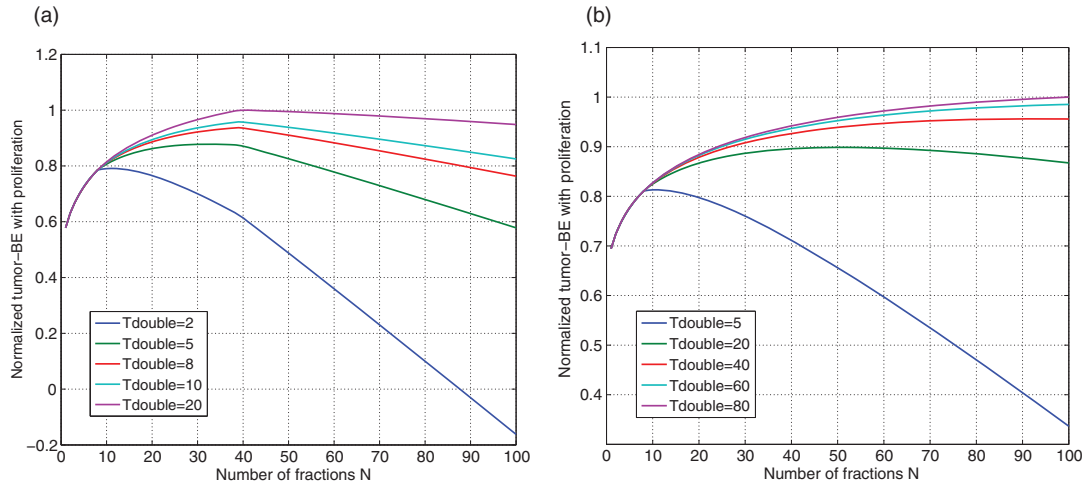

FIG. 6. Sensitivity of the optimal number of fractions and the optimal tumour-BE to tumour doubling time T_{double} (days) when $T_{\text{lag}} = 7$ days. (a) Head-and-neck case 1; tumour α/β was 8 Gy, and the α/β ratio for all normal tissues was 3 Gy. (b) Prostate case 1; tumour α/β was 6 Gy, and the α/β ratio for all normal tissues was 3 Gy. Some believe that the prostate α/β is much smaller than 6, and in particular, even smaller than 3; in most of those cases, as suggested by Proposition 5.1, it is optimal to administer a single fraction. We used this higher value of the α/β ratio here mainly to generate a broader range of curves for completeness.

TABLE 6 *The range of values of N^*_{99} for five head-and-neck cases with $T_{\text{lag}} = 7$ days over all combinations of tumour and normal tissue α/β ratios. Here, in case N^*_{99} was less than $1 + T_{\text{lag}}$ we reset it to $1 + T_{\text{lag}}$ as we know that this is a lower bound on the optimal number of fractions as in Remark 5.2.*

	T_{double} (days)							
Head-and-neck case #	2	3	5	8	10	20	40	50
1	8–10	13–16	23–28	33–35	35–36	37–38	37–50	37–56
2	8–10	14–17	24–29	28–35	29–36	31–42	34–63	34–70
3	8–10	13–16	24–28	37–41	41–44	44–47	45–57	45–65
4	8–10	13–16	24–28	36–39	38–41	41–43	42–59	42–66
5	8–10	14–17	24–29	34–37	36–39	38–40	39–56	39–63

therefore tracked the smallest number of fractions at which the tumour-BE reached 99% of its optimal value. We denote this number of fractions by N^*_{99}. The range of this number over all combinations of tumour α/β ratios and normal tissue α/β ratios for our five head-and-neck cases is reported in Table 6.

The table shows that for each T_{double} value, the range is quite narrow over all cases and over all combinations of α/β ratios. This observation may be helpful in practice given the uncertainty in and difficulty in estimating α/β values.

The table also shows that the effect of increasing T_{double} values on N^*_{99} seems to be saturating beyond $T_{\text{double}} \approx 10$ (there are only a very few combinations for which the right-end of the range is a relatively high number such as 56 or 63).

All optimal fractionation numbers in this table are at least 8, i.e. $1 + T_{\text{lag}}$ is a lower bound on the optimal number of fractions (see Remark 5.2).

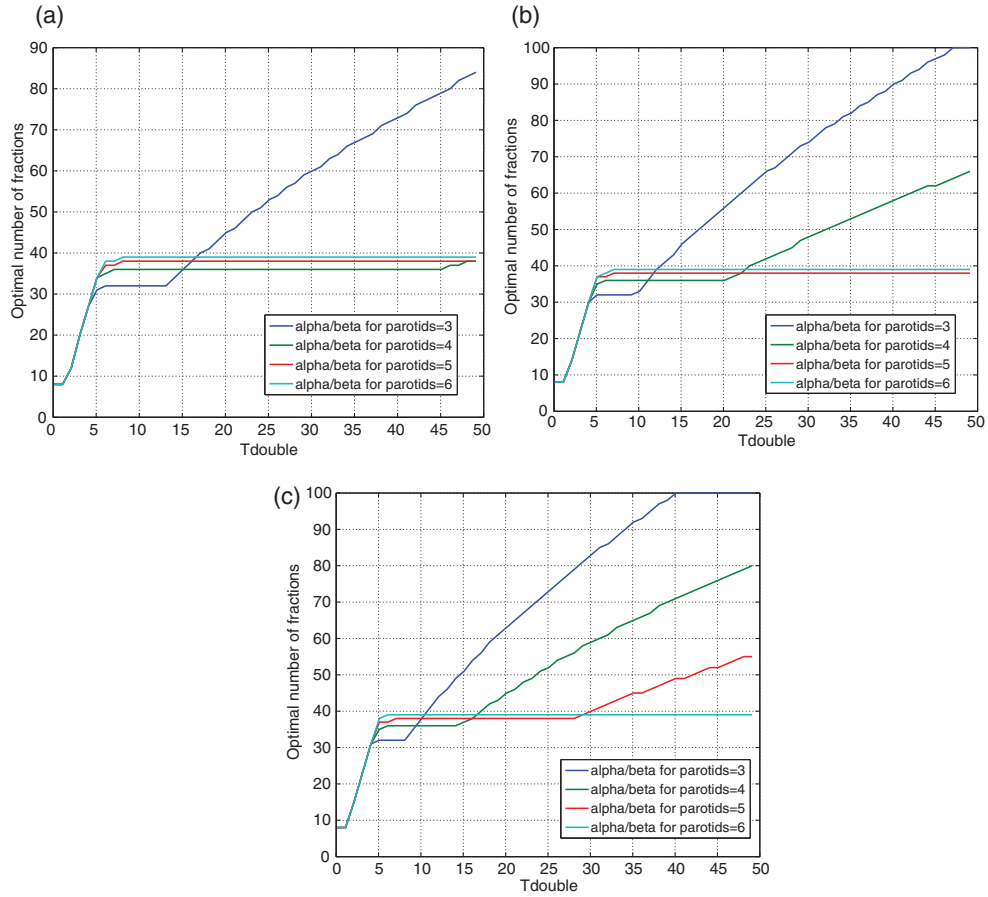

FIG. 7. Sensitivity of the optimal number of fractions to T_{double} as a function of the relative difference between the α/β ratios for the tumour and that for the parotids. This figure was generated for our head-and-neck case 2. The α/β ratios for the spinal cord and the brainstem were fixed at 3 Gy. T_{lag} was 7 days. (a) The α/β ratio for the tumour was 8 Gy; (b) the α/β ratio for the tumour was 10 Gy and (c) the α/β ratio for the tumour was 12 Gy.

We also investigated how the sensitivity of the optimal number of fractions to T_{double} depended on the relative values of the α/β ratios for the head-and-neck tumour and the normal tissues. To achieve this, we fixed the α/β ratios of the spinal cord and the brainstem at 3 Gy, and varied the α/β ratios for the parotids over the set $\{3, 4, 5, 6\}$. We also varied the head-and-neck tumour α/β ratio over the set $\{8, 10, 12\}$. The optimal number of fractions was plotted against T_{double} in Fig. 7. This figure shows that the sensitivity of the optimal number of fractions to T_{double} increases with increasing α/β ratio for the tumour. The qualitative trend in N_{99}^{*} was similar although it was less sensitive to T_{double}.

7.5 *Sensitivity to time-lag before proliferation begins*

We now investigate sensitivity of the optimal number of fractions to T_{lag}. Our Remark 5.2 shows that the optimal number of fractions is not sensitive to T_{lag} other than the fact that $1 + T_{\text{lag}}$ is a lower bound

TABLE 7 *The range of values of N^*_{99} for five head-and-neck cases with different combinations of T_{double} and T_{lag} over all combinations of tumour and normal tissue α/β ratios.*

$T_{\text{lag}} = 14$ (days)	T_{double} (days)							
Head-and-neck case #	2	3	5	8	10	20	40	50
1	15	15–16	23–28	33–35	35–36	37–38	37–50	37–56
2	15	15–17	23–29	28–35	29–36	31–41	34–63	34–70
3	15	15–16	23–28	37–41	41–44	44–47	45–57	45–65
4	15	15–16	23–28	36–39	38–41	41–43	42–59	42–66
5	15	15–17	24–29	34–37	36–39	38–40	39–56	39–63
$T_{\text{lag}} = 21$ (days)	T_{double} (days)							
Head-and-neck case #	2	3	5	8	10	20	40	50
1	22	22	23–28	33–35	35–36	37–38	37–50	37–56
2	22	22	23–29	28–35	29–36	31–41	34–63	34–70
3	22	22	23–28	37–41	41–44	44–47	45–57	45–65
4	22	22	23–28	36–39	38–41	41–43	42–59	42–66
5	22	22	24–29	34–37	36–39	38–40	39–56	39–63
$T_{\text{lag}} = 28$ (days)	T_{double} (days)							
Head-and-neck case #	2	3	5	8	10	20	40	50
1	29	29	29	33–35	35–36	37–38	37–50	37–56
2	29	29	29	29–35	29–36	31–41	34–63	34–70
3	29	29	29	36–41	40–44	44–47	45–57	45–65
4	29	29	29	35–39	38–41	41–43	42–59	42–66
5	29	29	29	34–37	36–38	38–40	39–56	39–63
$T_{\text{lag}} = 35$ (days)	T_{double} (days)							
Head-and-neck case #	2	3	5	8	10	20	40	50
1	36	36	36	36	36	36–38	37–50	37–56
2	36	36	36	36	36	36–41	36–63	36–70
3	36	36	36	36–41	40–44	44–47	45–57	45–65
4	36	36	36	36–39	38–41	41–43	42–59	42–66
5	36	36	36	36–37	36–38	38–40	39–56	39–63

on the optimal number of fractions when the sufficient condition in Proposition 5.1 holds. A closer look at our closed-form formula (B.3) for the optimal number of fractions in the single normal tissue case and the statement of Lemma B.1 shows that this lower bound is relevant mainly when T_{double} is small. Although we do not have a closed-form formula for the optimal number of fractions in the case of multiple normal tissues, we intuitively expect this logic to hold. To obtain further insight into this matter, we performed sensitivity analyses for head-and-neck cases as summarized in Table 7. Our analyses confirmed that for small values of T_{double} relative to T_{lag}, the optimal number of fractions is equal to $1 + T_{\text{lag}}$; for larger value of T_{double}, the optimal number of fractions is not sensitive to T_{lag}. Again, we did not include a similar table for our prostate cases because the optimal number of fractions was one in most instances and thus it was not sensitive to T_{lag}.

8. Discussion

As stated in Section 1, several alternative fractionation schedules have been tested clinically. For example, one popular schedule involves administering once daily treatment only over weekdays; another schedule involves twice daily treatment only over weekdays. Our formulation (P) can be easily modified to accommodate such fractionation schemes. To see how this can be achieved, suppose that the N fractions are well-separated and are administered in $T(N)$ days. Here, the mathematical meaning of the phrase well-separated is that there exists a time-interval of Δ days such that $T(N+1) - T(N) \geqslant \Delta$ for all $N \geqslant 1$. Thus, $T(N)$ is a strictly increasing function of N. Note also that $T(1) = 0$. For example, if treatment is administered once daily as in Sections 4 and 7, then $T(N) = N - 1$ for all $N \geqslant 1$. Similarly, if treatment is administered once daily but only over weekdays, then $T(N) = N - 1$ for $N = 1, 2, \ldots, 5$; $T(N) = N + 1$ for $N = 6, 7, \ldots, 10$; and so on. We emphasize here that this notation is general enough to encompass essentially any plausible fractionation schedule. For instance, consider a hypothetical schedule that administers two fractions on the first day with a 12-h gap in between and then skips the next day and administers a third fraction forty eight hours after the second fraction; in this case, we would have $T(1) = 0$, $T(2) = 1/2$ a day and $T(3) = 2.5$ days. On the other hand, if the gap between the first two fractions is 48 h and the third fraction is administered 12 h after the second fraction, then we would have $T(1) = 0$, $T(2) = 2$ days and $T(3) = 2.5$ days. To accommodate such schedules in our formulation (P), we only need to generalize the definition of $\tau(N)$ in (2) as

$$\tau(N) = \frac{[T(N) - T_{\text{lag}}]^+ \ln 2}{T_{\text{double}}}. \tag{55}$$

Lemmas 6.1 and 6.2, and Property 1 in Theorem 6.3 continue to hold in this general case without change. However, we have not been able to prove that Properties 2 and hence Property 3 in Theorem 6.3 hold. The reason is that our proof of Property 2 relies on convexity of $\tau(N)$ as defined in (2) but the generalization in (55) may not be convex. Thus, our algorithm in Section 6.1 will become a heuristic. However, this is not a limitation in practice. As stated earlier, the optimal number of fractions can be found by a brute-force method that utilizes an upper bound $N_{\max}$ on the number of fractions.

We also comment that the effect of intra-tumour oxygen partial pressure on the tumour's radiobiological parameters α_0 and β_0 is typically incorporated into the LQ model using the so-called Oxygen Enhancement Ratio (Wouters & Brown, 1997; Sovik *et al.*, 2007; South *et al.*, 2008). This approach involves scaling the values of α_0 and β_0 by a factor that depends on oxygen partial pressure. Since α_0 and β_0 serve as input to our optimization formulation, our model can be used to determine fractionation schedules for hypoxic tumours.

In this paper, as in all existing literature on determining the optimal number of fractions using the LQ model with tumour proliferation, we treated the spatial and the temporal components of the problem separately. That is, a fluence-map u_{nominal} was assumed to be available and the optimal number of fractions was determined by scaling this fluence-map. Such separation of the two components is suboptimal. Unkelbach *et al.* (2013) and Kim *et al.* (2012) have recently proposed optimization models where the fluence-map is optimized also using the BED concept; our problem (OPTFRAC2(N)) can be seen as a special case of their general models. However, their general models were not solved to optimality owing to computational difficulties rooted in non-convexity and the large size of their problems. Those two papers also did not explicitly include tumour proliferation and hence did not optimize the number of fractions. An interesting future research direction would be to extend our work here to simultaneously optimize the fluence-map and the number of fractions. However, that formulation will be a large-scale

(with several thousand variables and tens of thousands of constraints), non-convex, mixed-integer optimization problem and hence it will be computationally challenging to solve. As a result, it will not be possible to rigorously guarantee optimality. So the research there will focus on the development of an efficient algorithm for approximate solution.

Funding

This research was funded in part by the National Science Foundation through grant #CMMI 1054026.

References

Ahamad, A., Rosenthal, D. I. & Ang, K. K. (2005) Squamous cell head and neck cancer: recent clinical progress and prospects for the future. *Current Clinical Oncology* (D. J. Adelstein ed.). Totowa: Humana Press.

Arcangeli, G., Saracino, B., Gomellini, S., Petrongari, M. G., Arcangeli, S., Sentinelli, S., Marzi, S., Landoni, V., Fowler, J. & Strigari, L. (2010) A prospective phase III randomized trial of hypofractionation versus conventional fractionation in patients with high-risk prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.*, **78**, 11–18.

Armpilia, C. I., Dale, R. G. & Jones, B. (2004) Determination of the optimum dose per fraction in fractionated radiotherapy when there is delayed onset of tumour repopulation during treatment. *Br. J. Radiol.*, **77**, 765–767.

Bertuzzi, A., Bruni, C., Papa, F. & Sinisgalli, C. (2013) Optimal solution for a cancer radiotherapy problem. *J. Math. Biol.*, **66**, 311–349.

Bortfeld, T., Ramakrishnan, J., Tsitsiklis, J. N. & Unkelbach, J. (2013) Optimization of radiotherapy fractionation schedules in the presence of tumor repopulation. http://pages.discovery.wisc.edu/ jramakrishnan/BRT2013_repop.pdf.

Boyd, S. & Vandenberghe, L. (2004) *Convex Optimization*. Cambridge: Cambridge University Press.

Brenner, D. J. (2004) Fractionation and late rectal toxicity. *Int. J. Radiat. Oncol. Biol. Phys.*, **60**, 1013–1015.

Brenner, D. J. & Hall, E. J. (1999) Fractionation and protraction for radiotherapy of prostate carcinoma. *Int. J. Radiat. Oncol. Biol. Phys.*, **43**, 1095–1101.

Brenner, D. J., Hlatky, L. R., Hahnfeldt, P. J., Hall, E. J. & Sachs, R. K. (1995) A convenient extension of the linear-quadratic model to include redistribution and reoxygenation. *Int. J. Radiat. Oncol. Biol. Phys.*, **32**, 379–390.

Burman, C., Chui, C.-S., Kutcher, G., Leibel, S., Zelefsky, M., LoSasso, T., Spirou, S., Wu, Q., Yang, J., Stein, J., Mohan, R., Fuks, Z. & Ling, C. C. (1997) Planning, delivery, and quality assurance of intensity-modulated radiotherapy using dynamic multileaf collimator: A strategy for large-scale implementation for the treatment of carcinoma of the prostate. *Int. J. Radiat. Oncol. Biol. Phys.*, **39**, 863–873.

Dasu, A. (2007) Is the alpha/beta value for prostate tumours low enough to be safely used in clinical trials? *Clin. Oncol.*, **19**, 289–301.

Ehrgott, M., Guler, C., Hamacher, H. W. & Shao, L. (2008) Mathematical optimization in intensity modulated radiation therapy. *4OR*, **6**, 199–262.

Eisbruch, A. (2002) Intensity-modulated radiotherapy of head-and-neck cancer: encouraging early results. *Int. J. Radiat. Oncol. Biol. Phys.*, **53**, 1–3.

Emami, B., Lyman, J., Brown, A., Coia, L., Goiten, M. & Munzenride, J. E. *et al.* (1991) Tolerance of normal tissue to therapeutic radiation. *Int. J. Radiat. Oncol. Biol. Phys.*, **21**, 109–122.

Fowler, J. F. (1984) Non-standard fractionation in radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.*, **10**, 755–759.

Fowler, J. F. (1990) How worthwhile are short schedules in radiotherapy?: A series of exploratory calculations. *Radiother. Oncol.*, **18**, 165–181.

Fowler, J. F. (2001) Biological factors influencing optimum fractionation in radiation therapy. *Acta Oncol.*, **40**, 712–717.

Fowler, J. F. (2007) Is there an optimal overall time for head and neck radiotherapy? a review with new modeling. *Clin. Oncol.*, **19**, 8–27.

Fowler, J. F. (2008) Optimum overall times ii: Extended modelling for head and neck radiotherapy. *Clin. Oncol.*, **20**, 113–126.

Fowler, J. F., Chappell, R. & Ritter, M. (2001) Is alpha/beta for prostate tumors really low? *Int. J. Radiat. Oncol. Biol. Phys.*, **50**, 1021–1031.

Fowler, J. F. & Ritter, M. A. (1995) A rationale for fractionation for slowly proliferating tumors such as prostatic adenocarcinoma. *Int. J. Radiat. Oncol. Biol. Phys.*, **32**, 521–529.

Fu, K. K., Pajak, T. F., Trotti, A., Jones, C. U., Spencer, S. A., Phillips, T. L., Garden, A. S., Ridge, J. A., Cooper, J. S. & KianAng, K. (2000) A radiation therapy oncology group (RTOG) phase iii randomized study to compare hyperfractionation and two variants of accelerated fractionation to standard fractionation radiotherapy for head and neck squamous cell carcinomas: first report of RTOG 9003. *Int. J. Radiat. Oncol. Biol. Phys.*, **48**, 7–16.

Gao, M., Mayr, N. A., Huang, Z., Zhang, H. & Wang, J. Z. (2010) When tumor repopulation starts? the onset time of prostate cancer during radiation therapy. *Acta Oncol.*, **49**, 1269–1275.

Garden, A. S. (2001) Altered fractionation for head and neck cancer. *Oncology*, **15**, 1326–1332.

Grant, M. & Boyd, S. (2009) CVX: MATLAB software for disciplined convex programming (web page and software).

Gu, X., Choi, D., Men, C., Pan, H., Majumdar, A. & Jiang, S. B. (2009) GPU-based ultrafast dose calculation using a finite size pencil beam model. *Phys. Med. Biol.*, **54**, 6287–6297.

Hall, E. J. & Giaccia, A. J. (2005) *Radiobiology for the Radiologist*. Philadelphia, Pennsylvania, USA: Lippincott Williams & Wilkins.

Haustermans, K. M. G., Hofland, I., Van Poppel, H., Oyen, R., Van de Voorde, W., Begg, A. C. & Fowler, J. F. (1997) Cell kinetic measurements in prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.*, **37**, 1067–1070.

Ho, K. F., Fowler, J. F., Sykes, A. J., Yap, B. K., Lee, L. W. & Slevin, N. J. (2009) Imrt dose fractionation for head and neck cancer: Variation in current approaches will make standardisation difficult. *Acta Oncol.*, **3**, 48.

Horiot, J. C., Bontemps, P., Van Den Bogaert, W., Le, F. R., Van Den Weijngaert, D., Bolla, M., Bernier, J., Lusinchi, A., Stuschke, M., Lopez-Torrecilla, J., Begg, A. C., Pierart, M. & Collette, L. (1997) Accelerated fractionation (AF) compared to conventional fractionation (CF) improves loco-regional control in the radiotherapy of advanced head and neck cancers: results of the eortc 22851 randomized trial. *Radiother. Oncol.*, **44**, 111–121.

Horiot, J. C., Le Fur, R., N'Guyen, T., Chenal, C., Schraub, S., Alfonsi, S., Gardani, G., Van Den Bogaert, W., Danczak, S. & Bolla, M. (1992) Hyperfractionation versus conventional fractionation in oropharyngeal carcinoma: final analysis of a randomized trial of the eortc cooperative group of radiotherapy. *Radiother. Oncol.*, **25**, 231–241.

Jeraj, R. & Keall, P. (1999) Monte carlo-based inverse treatment planning. *Phys. Med. Biol.*, **44**, 1885–1896.

Joiner, M. & van der Kogel, A. (2009) *Basic Clinical Radiobiology*, 4th edn. London: Hodder Arnold.

Jones, B., Tan, L. T. & Dale, R. G. (1995) Derivation of the optimum dose per fraction from the linear quadratic model. *Br. J. Radiol.*, **68**, 894–902.

Kader, H. A., Mydin, A. R., Wilson, M., Alexander, C., Shahi, J., Pathak, I., Wu, J. S. & Truong, P. T. (2011) Treatment outcomes of locally advanced oropharyngeal cancer: A comparison between combined modality radio-chemotherapy and two variants of single modality altered fractionation radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.*, **80**, 1030–1036.

Kehwar, T. S. (2005) Analytical approach to estimate normal tissue complication probability using best fit of normal tissue tolerance doses into the ntcp equation of the linear quadratic model. *J. Cancer Res. Ther.*, **1**, 168–179.

Keller, H., Meier, G., Hope, A. & Davison, M. (2012) Fractionation schedule optimization for lung cancer treatments using radiobiological and dose distribution characteristics. *Med. Phys.*, **39**, 3811–3811.

Kim, M. (2010) A Mathematical Framework for Spatiotemporal Optimality in Radiation Therapy. *Ph.D. Thesis*, University of Washington, Seattle, Washington, USA.

Kim, M., Ghate, A. & Phillips, M. H. (2012) A stochastic control formalism for dynamic biologically conformal radiation therapy. *Eur. J. Oper. Res.*, **219**, 541–556.

Langer, M., Lee, E. K., Deasy, J. O., Rardin, R. L. & Deye, J. A. (2003) Operations research applied to radiotherapy, an NCI-NSF-sponsored workshop February 7-9, 2002. *Int. J. Radiat. Oncol. Biol. Phys.*, **57**, 762–768.

Luenberger, D. G. (1969) *Optimization by Vector Space Methods*. New York: John Wiley and Sons.

Luo, Z.-Q., Ma, W.-K., So, A. M. C. & Ye, Y. (2010) Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Process. Mag.*, **27**, 20–34.

Marks, L. B., Yorke, E. D., Jackson, A., Ten Haken, R. K., Constine, L. S., Eisbruch, A., Bentzen, S. M., Nam, J. & Deasy, J. O. (2010) Use of normal tissue complication probability models in the clinic. *Int. J. Radiat. Oncol. Biol. Phys.*, **76**, S10–S19.

Marzi, S., Saracino, B., Petrongari, M., Arcangeli, S., Gomellini, S., Arcangeli, G., Benassi, M. & Landoni, V. (2009) Modeling of alpha/beta for late rectal toxicity from a randomized phase ii study: conventional versus hypofractionated scheme for localized prostate cancer. *J. Exp. Clin. Cancer Res.*, **28**, 117–124.

Mavroidis, P., Ferreira, B. C. & Lopes, M. do. C. (2011) Response-probability volume histograms and iso-probability of response charts in treatment plan evaluation. *Med. Phys.*, **38**, 2382–2397.

Men, C., Gu, X., Choi, D., Majumdar, A., Zheng, Z., Mueller, K. & Jiang, S. B. (2009) GPU-based ultrafast IMRT plan optimization. *Phys. Med. Biol.*, **54**, 6565–6573.

Mizuta, M., Takao, S., Date, H., Kishimoto, N., Sutherland, K. L., Onimaru, R. & Shirato, H. (2012) A mathematical study to select fractionation regimen based on physical dose distribution and the linear-quadratic model. *Int. J. Radiat. Oncol. Biol. Phys.*, **84**, 829–833.

Qi, X. S., Yang, Q., Lee, S. P., Li, X. A. & Wang, D. (2012) An estimation of radiobiological parameters for head-and-neck cancer cells and the clinical implications. *Cancers*, **4**, 566–580.

Rockwell, S. (1998) Experimental radiotherapy: a brief history. *Radiat. Res.*, **150(Supplement)**, S157–S169.

Romeijn, H. E., Ahuja, R. K., Dempsey, J. F. & Kumar, A. (2006) A new linear programming approach to radiation therapy treatment planning problems. *Oper. Res.*, **54**, 201–216.

Romeijn, H. E. & Dempsey, J. F. (2008) Intensity modulated radiation therapy treatment plan optimization. *TOP*, **16**, 215–243.

Saberian, F. & Kim, M. (2014) *Phantom Creator (PhanC): a MATLAB software for creating phantom test cases for IMRT optimization, working draft of the user's manual*. University of Washington.

Salari, E. & Romeijn, H. E. (2012) Quantifying the trade-off between IMRT treatment plan quality and delivery efficiency using direct aperture optimization. *INFORMS J. Comput.*, **24**, 518–533.

Shepard, D. M., Ferris, M. C., Olivera, G. H. & Mackie, T. R. (1999) Optimizing the delivery of radiation therapy to cancer patients. *SIAM Rev.*, **41**, 721–744.

Siebers, J. V., Tong, S. D., Lauterbach, M., Wu, Q. W. & Mohan, R. (2001) Acceleration of dose calculations for intensity-modulated radiotherapy. *Med. Phys.*, **28**, 903–910.

South, C. P., Partridge, M. & Evans, P. M. (2008) A theoretical framework for prescribing radiotherapy dose distributions using patient-specific biological information. *Med. Phys.*, **35**, 4599–4611.

Sovik, A., Malinen, E., Bruland, O. S., Bentzen, S. M. & Olsen, D. R. (2007) Optimization of tumor control probability in hypoxic tumors by radiation dose redistribution: a modeling study. *Phys. Med. Biol.*, **52**, 499–513.

Spirou, S. V. & Chui, C. S. (1998) A gradient inverse planning algorithm with dose-volume constraints. *Med. Phys.*, **25**, 321–333.

Thames, H. D., Bentzen, S. M., Turesson, I., Overgaard, M. & van den Bogaert, W. (1989) Fractionation parameters for human tissues and tumors. *Int. J. Radiat. Biol.*, **56**, 701–710.

Tian, Z., Zarepisheh, M., Jia, X. & Jiang, S. B. (2013) The fixed-point iteration method for IMRT optimization with truncated dose deposition coefficient matrix. *Technical report*, arXiv:1303.3504.

Trotti, A., Fu, K. K., Pajak, T. F., Jones, C. U., Spencer, S. A., Phillips, T. L., Garden, A. S., Ridge, J. A., Cooper, J. S. & Ang, K. K. (2005) Long term outcomes of RTOG 90–03: A comparison of hyperfractionation and two variants of accelerated fractionation to standard fractionation radiotherapy for head and neck squamous cell carcinoma. *Int. J. Radiat. Oncol. Biol. Phys.*, **63(Supplement 1)**, S70–S71.

Unkelbach, J., Craft, D., Saleri, E., Ramakrishnan, J. & Bortfeld, T. (2013) The dependence of optimal fractionation schemes on the spatial dose distribution. *Phys. Med. Biol.*, **58**, 159–167.

UNKELBACH, J., ZENG, C. & ENGELSMAN, M. (2013) Simultaneous optimization of dose distributions and fractionation schemes in particle radiotherapy. *Med. Phys.*, **40**, 091702.
WANG, J. Z., GUERRERO, M. & LI, X. A. (2003) How low is the alpha/beta ratio for prostate cancer? *Int. J. Radiat. Oncol. Biol. Phys.*, **55**, 194–203.
WATERHOUSE, W. C. (1983) Do symmetric problems have symmetric solutions? *Am. Math. Mon.*, **90**, 378–387.
WEBB, S. (2010) *Contemporary IMRT: Developing Physics and Clinical Implementation*. Bristol: IOP Publishing.
WEBB, S. & OLDHAM, M. (1996) A method to study the characteristics of 3D dose distributions created by superposition of many intensity-modulated beams delivered via a slit aperture with multiple absorbing vanes. *Phys. Med. Biol.*, **41**, 2135–2153.
WILLIAMS, M. V., DENEKAMP, J. & FOWLER, J. F. (1985) A review of alpha/beta ratios for experimental tumors: implications for clinical studies of altered fractionation. *Int. J. Radiat. Oncol. Biol. Phys.*, **11**, 87–96.
WITHERS, H. R. (1985) Biologic basis for altered fractionation schemes. *Cancer*, **55**, 2086–2095.
WOUTERS, B. G. & BROWN, J. M. (1997) Cells at intermediate oxygen levels can be more important than the 'hypoxic fraction' in determining tumor response to fractionated radiotherapy. *Radiat. Res.*, **147**, 541–550.
YANG, Y. & XING, L. (2005) Optimization of radiotherapy dose-time fractionation with consideration of tumor specific biology. *Med. Phys.*, **32**, 3666–3677.

Appendix A. Proofs of technical results

A.1 *An algebraic proof of Proposition* 5.1

Suppose there exists an optimal solution $\vec{d}^* = (d_1^*, d_2^*, \ldots, d_N^*)$ where $d_i^* \neq d_j^*$ for some distinct i and j chosen from the set $\{1, 2, \ldots, N\}$. We will show that it is possible to construct an equal-dosage solution with the same tumour-BE as $\vec{d}^*$. Since N is fixed, $\tau(N)$ has no effect on optimal doses; we thus ignore $\tau(N)$ in this proof.

Since $\vec{d}^*$ is optimal, at least one of the constraints in (39) must be active. Without loss of generality, we assume that it is the constraint with $m = 1$. Thus, we have,

$$\sum_{t=1}^{N} (d_t^*)^2 = \frac{C_1 - \sigma_1 \sum_{t=1}^{N} d_t^*}{\sigma_1^2 \rho_1}. \tag{A.1}$$

Substituting this in the objective function, we see that the optimal objective function value is given by

$$\mathrm{BE}^* \triangleq \frac{C_1 \beta_0}{\sigma_1^2 \rho_1} + \left(\alpha_0 - \frac{\beta_0}{\rho_1 \sigma_1}\right) \sum_{t=1}^{N} d_t^*. \tag{A.2}$$

To obtain a contradiction to the optimality of $\vec{d}^*$, we will now construct an alternative feasible solution with a larger objective value than BE^*. Towards this end, we first define dose

$$\delta = \frac{\sum_{t=1}^{N} d_t^*}{N}, \tag{A.3}$$

and the dose vector $\vec{\delta} = (\delta, \delta, \ldots, \delta)$. We claim that this dose vector is *strictly* feasible to all constraints in OPTFRAC2(N). To see this, we first define the notation $\|\vec{d}\|_1$, for any non-negative dose vector $\vec{d}$, to denote the l_1 norm of $\vec{d}$. Similarly, $\|\vec{d}\|_2$ denotes the l_2 norm of $\vec{d}$. We now evaluate the LHS in (39) for

some any arbitrary $m \in \mathcal{M}$ by substituting the dose vector $\vec{\delta}$. We get,

$$\sigma_m \sum_{t=1}^{N} \delta + \sigma_m^2 \rho_m \sum_{t=1}^{N} \delta^2 = N\sigma_m \delta + \sigma_m^2 \rho_m N \delta^2 = \sigma_m \sum_{t=1}^{N} d_t^* + \sigma_m^2 \rho_m N \frac{(\sum_{t=1}^{N} d_t^*)^2}{N^2} \tag{A.4}$$

$$= \sigma_m \sum_{t=1}^{N} d_t^* + \sigma_m^2 \rho_m \frac{(\|\vec{d}^*\|_1)^2}{N} < \sigma_m \sum_{t=1}^{N} d_t^* + \sigma_m^2 \rho_m (\|\vec{d}^*\|_2)^2 \tag{A.5}$$

$$= \sigma_m \sum_{t=1}^{N} d_t^* + \sigma_m^2 \rho_m \sum_{t=1}^{N} (d_t^*)^2 \leqslant C_m. \tag{A.6}$$

Here, the strict inequality follows from the well-known norm-inequality $\|\vec{d}\|_1 \leqslant \sqrt{N}\|\vec{d}\|_2$ (with equality holding only when all components of $\vec{d}$ are identical; note that at least two components of our vector $\vec{d}^*$ are distinct by assumption). The second inequality holds because $\vec{d}^*$ is feasible to OPTFRAC2(N). Owing to the *strict* feasibility of $\vec{\delta}$ to OPTFRAC2(N), there is some slack in every constraint. In particular, there exists an $\epsilon > 0$ such that delivering dose $\delta + \epsilon$ in each one of the N fractions is feasible. In fact, we will choose ϵ to be large enough such that at least one of the constraints becomes active. Specifically, we let $\epsilon = \min_{m \in \mathcal{M}} \epsilon_m$, where $\epsilon_m > 0$ for $m \in \mathcal{M}$ is uniquely defined such that

$$N\sigma_m(\delta + \epsilon_m) + N\sigma_m^2 \rho_m (\delta + \epsilon_m)^2 = C_m. \tag{A.7}$$

Suppose without loss of generality that $\epsilon = \epsilon_2$. That is, the dose delivered in each fraction now equals $\delta + \epsilon_2$ and

$$N\sigma_2(\delta + \epsilon_2) + N\sigma_2^2 \rho_2 (\delta + \epsilon_2)^2 = C_2, \text{ and hence,} \tag{A.8}$$

$$N(\delta + \epsilon_2)^2 = \frac{C_2 - N\sigma_2(\delta + \epsilon_2)}{\sigma_2^2 \rho_2}. \tag{A.9}$$

Now we evaluate the objective function value BE for this new dosing schedule. We have,

$$\text{BE} = \alpha_0 N(\delta + \epsilon_2) + \beta_0 N(\delta + \epsilon_2)^2 = \frac{C_2 \beta_0}{\sigma_2^2 \rho_2} + \left(\alpha_0 - \frac{\beta_0}{\sigma_2 \rho_2}\right) N(\delta + \epsilon_2) \tag{A.10}$$

$$= \frac{C_2 \beta_0}{\sigma_2^2 \rho_2} + \left(\alpha_0 - \frac{\beta_0}{\sigma_2 \rho_2}\right) \sum_{t=1}^{N} d_t^* + \left(\alpha_0 - \frac{\beta_0}{\sigma_2 \rho_2}\right) N\epsilon_2. \tag{A.11}$$

Now we show that BE = BE*. Since $\vec{d}^*$ is feasible for constraint (39) with $m = 2$, we have,

$$\sigma_2 \sum_{t=1}^{N} d_t^* + \sigma_2^2 \rho_2 \sum_{t=1}^{N} d_t^{*2} \leqslant C_2. \tag{A.12}$$

Now, by substituting for $\sum_{t=1}^{N} d_t^{*2}$ from (A.1), the above inequality simplifies as

$$\sigma_2 \sum_{t=1}^{N} d_t^* + \sigma_2^2 \rho_2 \left(\frac{C_1 - \sigma_1 \sum_{t=1}^{N} d_t^*}{\sigma_1^2 \rho_1}\right) = \frac{\sigma_2^2 \rho_2 C_1}{\sigma_1^2 \rho_1} + \left(\sigma_2 - \frac{\sigma_2^2 \rho_2}{\sigma_1 \rho_1}\right) \sum_{t=1}^{N} d_t^* \leqslant C_2. \tag{A.13}$$

Now, by multiplying both sides of this inequality by the positive term $\frac{\beta_0}{\rho_2\sigma_2^2}$, we get,

$$\frac{\beta_0 C_1}{\sigma_1^2\rho_1} + \frac{\beta_0}{\sigma_2^2\rho_2}\left(\sigma_2 - \frac{\sigma_2^2\rho_2}{\sigma_1\rho_1}\right)\sum_{t=1}^{N} d_t^* \leqslant \frac{\beta_0 C_2}{\sigma_2^2\rho_2}. \tag{A.14}$$

Now by adding $(\alpha_0 - \beta_0/\sigma_2\rho_2)\sum_{t=1}^{N} d_t^*$ to both sides of this inequality, we obtain,

$$\mathrm{BE}^* = \frac{C_1\beta_0}{\sigma_1^2\rho_1} + \left(\alpha_0 - \frac{\beta_0}{\sigma_1\rho_1}\right)\sum_{t=1}^{N} d_t^* \leqslant \frac{C_2\beta_0}{\sigma_2^2\rho_2} + \left(\alpha_0 - \frac{\beta_0}{\sigma_2\rho_2}\right)\sum_{t=1}^{N} d_t^* \leqslant \mathrm{BE}. \tag{A.15}$$

Here, the last inequality follows from the definition of BE in (A.11) because $\epsilon_2 > 0$ and $(\alpha_0 - \beta_0/\sigma_2\rho_2) = (\alpha_0 - \beta_0(\alpha_2/\beta_2)/\sigma_2) \geqslant 0$ by our hypothesis in the statement of the proposition. But since BE^* is the optimal tumour-BE, the above inequality implies that $\mathrm{BE} = \mathrm{BE}^*$. This shows that there is no loss of optimality in only considering equal-dosage fractionation schedules in (OPTFRAC2(N)) under the hypothesis of this proposition. The fact that the dose in an optimal equal-dosage schedule is given by (41) follows by finding the smallest from the M positive roots of M quadratic equations defined by the constraints (39).

Remark A.1 The idea of using the average of the components of a vector optimal solution as in (A.3), has often been used to prove optimality of symmetric solutions to *convex* problems (see, e.g. Waterhouse, 1983 for an interesting historical account of this topic). In fact, we used this idea in the context of a convex problem in radiotherapy in Proposition 4.2.1 in Kim (2010). Since OPTFRAC2(N) is *not* a convex problem, our proof of Proposition 5.1 relies on a sufficient condition under which a modification of this 'average of components' method works. Below we also provide a different proof based on KKT conditions.

A.2 *Proof of Proposition 5.1 using KKT conditions*

Problem OPTFRAC2(N) is not convex. However, it is easy to construct a strictly feasible dose vector. Furthermore, the feasible region is convex. In particular, this problem satisfies Slater's constraint qualification (Boyd & Vandenberghe, 2004). Consequently, KKT conditions are necessary for optimality. To write these, we associate Lagrange multipliers $\lambda_1, \lambda_2, \ldots, \lambda_M$ with the inequality constraints in OPTFRAC2(N) and Lagrange multipliers $\mu_1, \mu_2, \ldots, \mu_N$ with the non-negativity constraints on doses. We let $\vec{e}_t$, for $t = 1, 2, \ldots, N$, denote the N-dimensional unit vector with a one as the tth element and zeros everywhere else. Then, the KKT conditions are given by

$$-\begin{bmatrix} \alpha_0 + 2\beta_0 d_1 \\ \alpha_0 + 2\beta_0 d_2 \\ \vdots \\ \alpha_0 + 2\beta_0 d_N \end{bmatrix} + \sum_{m=1}^{M} \lambda_m \sigma_m \begin{bmatrix} 1 + 2\rho_m\sigma_m d_1 \\ 1 + 2\rho_m\sigma_m d_2 \\ \vdots \\ 1 + 2\rho_m\sigma_m d_N \end{bmatrix} - \sum_{t=1}^{N} \mu_t \vec{e}_t = \vec{0}, \tag{A.16}$$

$$\vec{d} \geqslant 0, \tag{A.17}$$

$$\sigma_m \sum_{t=1}^{N} d_t + \sigma_m^2 \rho_m \sum_{t=1}^{N} (d_t)^2 \leqslant C_m, \quad m \in \mathcal{M}, \tag{A.18}$$

$$\left(\sigma_m \sum_{t=1}^{N} d_t + \sigma_m^2 \rho_m \sum_{t=1}^{N} (d_t)^2 - C_m\right) \lambda_m = 0, \quad m \in \mathcal{M}, \tag{A.19}$$

$$\mu_t d_t = 0, \quad t = 1, 2, \ldots, N, \tag{A.20}$$

$$\lambda_m \geqslant 0, \quad m \in \mathcal{M}, \tag{A.21}$$

$$\mu_t \geqslant 0, \quad t = 1, 2, \ldots, N. \tag{A.22}$$

Now, for each $k = 1, 2, \ldots, N$, let $\vec{d}(k) \triangleq (d_1, d_2, \ldots, d_k, 0, 0, \ldots, 0)$, where $d_1 > 0$, $d_2 > 0$, …, $d_k > 0$. That is, $\vec{d}(k)$ is a dose vector where the doses in the first k fractions are positive but the doses in the other fractions are zero. We use the above KKT conditions to characterize these dose vectors. Note that it suffices to consider dose vectors where the leading elements are positive and the tail elements are zero; i.e. we need not separately consider vectors such as $(0, d_2, 0, 0, \ldots, 0)$. This is because the objective function and the constraints are symmetric with respect to permutations of elements of dose vectors.

For $\vec{d}(k)$, KKT conditions (A.20) imply that $\mu_1 = \mu_2 = \ldots = \mu_k = 0$. Thus, (A.16) reduce to

$$-(\alpha_0 + 2\beta_0 d_t) + \sum_{m=1}^{M} \lambda_m \sigma_m (1 + 2\rho_m \sigma_m d_t) = 0, \quad t = 1, 2, \ldots, k, \tag{A.23}$$

$$\mu_t = \sum_{m=1}^{M} \lambda_m \sigma_m - \alpha_0, \quad t = k+1, \ldots, N. \tag{A.24}$$

We rewrite the first group of equations above as

$$2d_t \left(-\beta_0 + \sum_{m=1}^{M} \lambda_m \rho_m \sigma_m^2\right) = \alpha_0 - \sum_{m=1}^{M} \lambda_m \sigma_m, \quad t = 1, 2, \ldots, k. \tag{A.25}$$

Since the above right hand side does not depend on t, the left hand sides must also be equal for different t. Therefore, we have,

$$d_s \left(-\beta_0 + \sum_{m=1}^{M} \lambda_m \rho_m \sigma_m^2\right) = d_t \left(-\beta_0 + \sum_{m=1}^{M} \lambda_m \rho_m \sigma_m^2\right), \quad s, t = 1, 2, \ldots, k. \tag{A.26}$$

We now consider two cases: (i) $(\alpha_0/\beta_0) > \max_{m \in \mathcal{M}} (\alpha_m/\beta_m)/\sigma_m$ and (ii) $(\alpha_0/\beta_0) = \max_{m \in \mathcal{M}} (\alpha_m/\beta_m)/\sigma_m$.

Case 1: $(\alpha_0/\beta_0) > \max_{m \in \mathcal{M}} (\alpha_m/\beta_m)/\sigma_m$. We claim that $\beta_0 \neq \sum_{m=1}^{M} \lambda_m \rho_m \sigma_m^2$. This would imply that $d_1 = d_2 = \ldots = d_k \triangleq d^*(k)$ for some $d^*(k)$ from the above equation. We proceed by contradiction. So suppose that $\beta_0 = \sum_{m=1}^{M} \lambda_m \rho_m \sigma_m^2$. Substituting this into (A.25), we get, $\alpha_0 = \sum_{m=1}^{M} \lambda_m \sigma_m$. Thus, $(\beta_0/\alpha_0) = (\sum_{m=1}^{M} \lambda_m \rho_m \sigma_m^2)/(\sum_{m=1}^{M} \lambda_m \sigma_m)$. But by (A.21), $\lambda_m \geqslant 0$ for $m \in \mathcal{M}$. Thus, (β_0/α_0) is a convex combination of $\sigma_m \beta_m/\alpha_m$. But this is a contradiction since $(\alpha_0/\beta_0) > (\alpha_m/\beta_m)/\sigma_m$ for all m by assumption. Now, since the right hand side in (A.24) does not depend on t, we must have that $\mu_{k+1} = \mu_{k+2} = \ldots = \mu_N = \mu$ for some μ. Then, (A.24) implies that $\sum_{m=1}^{M} \lambda_m \sigma_m = \alpha_0 + \mu$. Substituting

this in any one of the group (A.23), we get,

$$\mu = 2d^*(k)\left(\beta_0 - \sum_{m=1}^{M} \lambda_m \sigma_m^2 \rho_m\right). \tag{A.27}$$

From this, we claim that if $\mu \geqslant 0$ then $d^*(k) < 0$ and hence the only possible KKT point is where $\mu_t = 0$ for $t = 1, 2, \ldots, N$. Towards this end, suppose $\mu \geqslant 0$. Then, from (A.24), we have that $\sum_{m=1}^{M} \lambda_m \sigma_m \geqslant \alpha_0$. Without loss of generality, suppose that $\min_{m\in\mathcal{M}} \sigma_m \rho_m = \sigma_1 \rho_1$. We have, $\sum_{m=2}^{M} \lambda_m \sigma_m \geqslant \alpha_0 - \lambda_1 \sigma_1$. That is, $-\sum_{m=2}^{M} \lambda_m \sigma_m \leqslant \lambda_1 \sigma_1 - \alpha_0$. Thus,

$$\beta_0 - \sum_{m=1}^{M} \lambda_m \sigma_m^2 \rho_m \leqslant \beta_0 - \sigma_1 \rho_1 \sum_{m=1}^{M} \lambda_m \sigma_m = \beta_0 - \lambda_1 \sigma_1^2 \rho_1 - \sigma_1 \rho_1 \sum_{m=2}^{M} \lambda_m \sigma_m \leqslant \beta_0 - \alpha_0 \sigma_1 \rho_1 < 0. \tag{A.28}$$

Here, the last strict inequality follows because $\alpha_0/\beta_0 > 1/(\rho_m \sigma_m)$ for all m, and in particular, $\alpha_0/\beta_0 > 1/(\rho_1 \sigma_1)$, by assumption. Substituting this in (A.27) implies that $d^*(k) < 0$ as claimed. In particular, $\vec{d}^*(N) \triangleq (d^*(N), d^*(N), \ldots, d^*(N))$ is the only possible KKT point and hence an equal-dosage solution must be optimal. The actual value of $d^*(N)$ can be computed by solving quadratic equations as explained in the algebraic proof above.

Case 2: $(\alpha_0/\beta_0) = \max_{m\in\mathcal{M}}(\alpha_m/\beta_m)/\sigma_m$. We assume without loss of generality that there are no ties in this maximum. In particular, assume without loss of generality that $(\alpha_0/\beta_0) = (\alpha_1/\beta_1)/\sigma_1 = \max_{m\in\mathcal{M}}(\alpha_m/\beta_m)/\sigma_m$. Now suppose as above that $\mu \geqslant 0$ and by similar algebra, we see that $\beta_0 = \sum_{m=1}^{M} \lambda_m \rho_m \sigma_m^2 \leqslant 0$. Then, we consider two further subcases: (a) $\beta_0 = \sum_{m=1}^{M} \lambda_m \rho_m \sigma_m^2$ and (b) $\beta_0 < \sum_{m=1}^{M} \lambda_m \rho_m \sigma_m^2$.

Case 2a: In this case, by using the above convex combination argument, we see that $\lambda_1 > 0$ and $\lambda_m = 0$ for all $m \in \mathcal{M}$, $m \neq 1$. Then (A.25) implies that $\lambda_1 = \alpha_0/\sigma_1 > 0$, and in particular, (A.19) implies that $\sigma_1 \sum_{t=1}^{k} d_t + \sigma_1^2 \rho_1 \sum_{t=1}^{k} (d_t^2) = C_1$. That is, the corresponding inequality constraint is active and we can eliminate $\sum_{t=1}^{k} (d_t^2)$ from the objective function by substituting

$$\sum_{t=1}^{k} (d_t^2) = \frac{C_1 - \sigma_1 \sum_{t=1}^{k} d_t}{\sigma_1^2 \rho_1}. \tag{A.29}$$

But since $(\alpha_0/\beta_0) = 1/(\rho_1 \sigma_1)$, this reduces the objective function to the constant $\beta_0 C_1/(\sigma_1^2 \rho_1)$. Similarly, we can eliminate $\sum_{t=1}^{k} (d_t^2)$ from all other constraints and this reduces OPTFRAC2(N) to a problem with a single non-negative variable $z = \sum_{t=1}^{k} d_t$. Then, an optimal solution to this simplified problem can be found such that $d_1 = d_2 = \ldots = d_k$. Furthermore, the objective function value itself is independent of the value of $k = 1, 2, \ldots, N$. That is, an equal-dosage solution is optimal.

Case 2b: In this case, we can show that $d_i < 0$ for $i = 1, 2, \ldots, k$ by algebra similar to above and hence this case cannot lead to a KKT solution.

A.3 *An algebraic proof of Proposition* 5.3

Suppose that dose vector $\vec{d}^* = (d_1^*, d_2^*, \ldots, d_N^*)$ is optimal, and that there exist distinct i and j chosen from the set $\{1, 2, \ldots, N\}$ such that $d_i^* > 0$ and $d_j^* > 0$. Without loss of generality, we set $i = 1$ and $j = 2$

in the rest of this proof. We will construct a dosing schedule where the dose in the first fraction is positive but the dose in the second fraction is zero; we will show that this schedule has the same tumour-BE as that of $\vec{d}^*$. Then, the claim in this proposition follows by applying this procedure recursively. Since N is fixed, $\tau(N)$ has no effect on optimal doses; we thus ignore $\tau(N)$ in this proof.

Since $\vec{d}^*$ is optimal, at least one of the constraints in (39) must be active. Without loss of generality, we assume that it is the constraint with $m = 1$. Thus, we have,

$$\sigma_1(d_1^* + d_2^* + d_3^* + \cdots + d_N^*) + \sigma_1^2\rho_1((d_1^*)^2 + (d_2^*)^2 + (d_3^*)^2 + \cdots + (d_N^*)^2) = C_1. \tag{A.30}$$

Thus, as in the proof of Proposition 5.3, we again define the optimal objective function value as

$$\mathrm{BE}^* \triangleq \frac{C_1\beta_0}{\sigma_1^2\rho_1} + \left(\alpha_0 - \frac{\beta_0}{\rho_1\sigma_1}\right)\sum_{t=1}^{N} d_t^*. \tag{A.31}$$

Now consider an alternative feasible dose vector $\vec{\delta} \triangleq (d_1^* + \epsilon, 0, d_3^*, \ldots, d_N^*)$, where $\epsilon = \min_{m\in\mathcal{M}} \epsilon_m$, and $\epsilon_m > 0$ for all $m \in \mathcal{M}$ are chosen such that

$$\sigma_m(d_1^* + \epsilon_m + d_3^* + \ldots + d_N^*) + \sigma_m^2\rho_m((d_1^* + \epsilon_m)^2 + (d_3^*)^2 + \ldots + (d_N^*)^2) = C_m. \tag{A.32}$$

We assume without loss of generality that $\epsilon = \epsilon_2$. Therefore,

$$(d_1^* + \epsilon_2)^2 + (d_3^*)^2 + \cdots + (d_N^*)^2 = \frac{C_2 - \sigma_2(d_1^* + \epsilon_2 + d_3^* + \cdots + d_N^*)}{\sigma_2^2\rho_2}. \tag{A.33}$$

Note that because $\vec{\delta}$ is feasible to constraint (39) for $m = 1$ and that this constraint is assumed to be active at $\vec{d}^*$ (recall (A.30)), we have,

$$\sigma_1(d_1^* + \epsilon_2 + d_3^* + \ldots + d_N^*) + \sigma_1^2\rho_1((d_1^* + \epsilon_2)^2 + (d_3^*)^2 + \cdots + (d_N^*)^2) \tag{A.34}$$

$$\leqslant \sigma_1(d_1^* + d_2^* + d_3^* + \cdots + d_N^*) + \sigma_1^2\rho_1((d_1^*)^2 + (d_2^*)^2 + (d_3^*)^2 + \cdots + (d_N^*)^2). \tag{A.35}$$

After simplifying the above inequality, we get

$$\epsilon_2 + \sigma_1\rho_1(\epsilon_2^2 + 2\epsilon_2 d_1^*) \leqslant d_2^* + \sigma_1\rho_1(d_2^*)^2, \tag{A.36}$$

which implies that

$$\epsilon_2 + \sigma_1\rho_1\epsilon_2^2 < d_2^* + \sigma_1\rho_1(d_2^*)^2 \tag{A.37}$$

because $2\epsilon_2 d_1^* > 0$. This shows that

$$\epsilon_2 < d_2^*. \tag{A.38}$$

The objective function value at $\vec{\delta}$ is given by

$$\alpha_0(d_1^* + \epsilon_2 + d_3^* + \ldots + d_N^*) + \beta_0((d_1^* + \epsilon_2)^2 + (d_3^*)^2 + \ldots + (d_N^*)^2) \tag{A.39}$$

$$= \alpha_0(d_1^* + \epsilon_2 + d_3^* + \ldots + d_N^*) + \beta_0\frac{C_2 - \sigma_2(d_1^* + \epsilon_2 + d_3^* + \ldots + d_N^*)}{\sigma_2^2\rho_2} \tag{A.40}$$

$$= \frac{\beta_0 C_2}{\sigma_2^2 \rho_2} + \left(\alpha_0 - \frac{\beta_0}{\sigma_2 \rho_2}\right)(d_1^* + \epsilon_2 + d_3^* + \ldots + d_N^*) \tag{A.41}$$

$$\geqslant \frac{\beta_0 C_2}{\sigma_2^2 \rho_2} + \left(\alpha_0 - \frac{\beta_0}{\sigma_2 \rho_2}\right)(d_1^* + d_2^* + d_3^* + \ldots + d_N^*) \geqslant \mathrm{BE}^*. \tag{A.42}$$

Here, the first inequality follows because $(\alpha_0 - \beta_0/\sigma_2\rho_2) \leqslant 0$ by our hypothesis in the statement of the proposition and $\epsilon_2 < d_2^*$ from (A.38). The proof to show that the second inequality holds is the same as the corresponding part in the proof for Proposition 5.1. But since BE^* is the optimal tumour-BE, the above inequality implies that $\mathrm{BE} = \mathrm{BE}^*$. This shows that there is no loss of optimality in administering a single fraction with positive dose in (OPTFRAC2(N)) under the hypothesis of this proposition. The value of this single dose is easily obtained by finding the smallest among the M positive roots of the M quadratic equations implied by constraints (39) just as in the proof of Proposition 5.1.

A.4 *Proof of Proposition 5.3 using KKT conditions*

We only provide a brief sketch of this proof to avoid repetition. Using a convex combination argument identical to the proof of Proposition 5.1, we can again show that $\beta_0 \neq \sum_{m=1}^{M} \lambda_m \rho_m \sigma_m^2$ if $(\alpha_0/\beta_0) < \min_{m \in \mathcal{M}}(\alpha_m/\beta_m)/\rho_m$. In this case, using algebra similar to the proof of Proposition 5.1, we can show that $\vec{d}^*(k) \triangleq (\underbrace{d^*(k), \ldots, d^*(k)}_{k \text{ times}}, 0, \ldots, 0)$, for some $d^*(k) > 0$, and for $k = 1, 2, \ldots, N$, are the only possible KKT points. In fact, the best value of $d^*(k)$ is defined precisely as in (48) by

$$d^*(k) = \min_{m \in \mathcal{M}} b_m(k), \tag{A.43}$$

where $b_m(k)$ is defined as in (47). That is,

$$b_m(k) \triangleq \frac{-1 + \sqrt{1 + 4\rho_m C_m/k}}{2\sigma_m \rho_m}, \quad m \in \mathcal{M}. \tag{A.44}$$

Among the N different dosing schedules $\vec{d}^*(k)$, the one that maximizes the tumour-BE must be optimal. The fact that $\vec{d}^*(1)$ is this tumour-BE maximizing schedule, can be seen briefly as follows. We define the tumour-BE without proliferation of $\vec{d}^*(k)$ as

$$f^*(k) \triangleq \alpha_0 k(d^*(k)) + \beta_0 k(d^*(k))^2 = \alpha_0 k \left(\min_{m \in \mathcal{M}} b_m(k)\right) + \beta_0 k \left(\min_{m \in \mathcal{M}} b_m(k)\right)^2 \tag{A.45}$$

$$= \min_{m \in \mathcal{M}} f_m(k), \tag{A.46}$$

where $f_m(k) \triangleq \alpha_0 k(b_m(k)) + \beta_0 k(b_m(k))^2$. Now, following our calculus based approach for the single normal tissue case in Appendix B, we can see that $f_m(k)$ is a decreasing function of k for each $m \in \mathcal{M}$. Consequently, $f^*(k)$ is also a decreasing function of k. This shows that $\vec{d}^*(1)$ maximizes tumour-BE. That is, a single-dosage solution is optimal.

In the other case where $(\alpha_0/\beta_0) = \min_{m \in \mathcal{M}}(\alpha_m/\beta_m)/\sigma_m$, we again let this minimum equal $(\alpha_1/\beta_1)/\sigma_1$ without loss of generality. Then there are two further subcases: (a) $\beta_0 = \sum_{m=1}^{M} \lambda_m \rho_m \sigma_m^2$, and (b) $\beta_0 > \sum_{m=1}^{M} \lambda_m \rho_m \sigma_m^2$. The proof of subcase (a) is similar to that of Case 2(a) in the proof of

Proposition 5.1 and the proof of subcase (b) is similar to the proof of the first case above. In both scenarios, we can conclude that a single-dosage solution is optimal.

A.5 *Proof of Lemma 6.1*

Observe that dose $b_m(N)$ as defined in (47) is a strictly decreasing functions of N for each $m \in \mathcal{M}$. Since $d^*(N)$ is the smallest among these doses, it is also strictly decreasing.

Now we view $b_m(N)$, for $m \in \mathcal{M}$, as functions of all real numbers $N \geqslant 1$ and show that they are strictly convex. Thus, it suffices to show that $a_m(N) \triangleq \sqrt{1 + 4\rho_m C_m/N}$ is strictly convex for each $m \in \mathcal{M}$. After some algebraic simplification, we obtain,

$$\frac{\mathrm{d}a_m(N)}{\mathrm{d}N} = \frac{-2\rho_m C_m}{\sqrt{N^4 + 4\rho_m C_m N^3}}. \tag{A.47}$$

As a result, we get,

$$\frac{\mathrm{d}^2 a_m(N)}{\mathrm{d}N^2} = \frac{\rho_m C_m (4N^3 + 12\rho_m C_m N^2)}{(N^4 + 4\rho_m C_m N^3)^{3/2}}. \tag{A.48}$$

The above second derivative is positive, and hence, $a_m(N)$ is strictly convex. If the dose-limiting normal tissue is invariant over all real numbers $N \geqslant 1$, then $d^*(N)$ is strictly convex (and hence trivially piecewise strictly convex). So we suppose that this is not the case. Then, owing to the continuity of doses $b_m(N)$ in N, there exist real numbers $1 < i_1 < i_2 < i_3, \ldots$ such that the dose-limiting normal tissues do not change within the intervals $[1, i_1), [i_1, i_2), [i_2, i_3), \ldots$. Furthermore, in each of these intervals, the limiting dose $d^*(N)$ equals one of the strictly convex functions $b_m(N)$. Therefore, $d^*(N)$ is piecewise strictly convex.

A.6 *Proof of Lemma 6.2*

We show that

$$\bar{B} = \max_{m \in \mathcal{M}} (C_m/\sigma_m). \tag{A.49}$$

works. To see this, first consider any $m \in \mathcal{M}$. We have,

$$1 + 4\rho_m C_m/N \leqslant 1 + 4\sigma_m \rho_m \bar{B}/N \leqslant 1 + 4\sigma_m \rho_m \bar{B}/N + 4\sigma_m^2 \rho_m^2 \bar{B}^2/N^2. \tag{A.50}$$

Here the first inequality follows from the definition of $\bar{B}$ in (A.49) and the second inequality follows because $4\sigma_m^2 \rho_m^2 \bar{B}^2/N^2 > 0$. Then, by taking square roots of both sides, we obtain

$$\sqrt{1 + 4\rho_m C_m/N} \leqslant 1 + 2\sigma_m \rho_m \bar{B}/N, \tag{A.51}$$

or equivalently,

$$\frac{-1 + \sqrt{1 + 4\rho_m C_m/N}}{2\sigma_m \rho_m} = b_m(N) \leqslant \frac{\bar{B}}{N}. \tag{A.52}$$

This shows from (48) that $d^*(N) \leqslant \bar{B}/N$ for all N.

A.7 *Proof of Theorem* 6.3

We prove the three properties one by one.

Proof of Property 1. To establish the existence of an $\hat{N}$ with the said property, it suffices to focus on $N \geqslant 1 + T_{\text{lag}}$.

We consider two cases: the first case is where there exists an $\hat{N} \geqslant 1 + T_{\text{lag}}$ such that $g^*(\hat{N}) + \tau(\hat{N}) \leqslant g^*(\hat{N}-1) + \tau(\hat{N}-1)$; the second case is where such an $\hat{N}$ does not exist.

In the first case, we have,

$$g^*(\hat{N}) + \tau(\hat{N}) \leqslant g^*(\hat{N}-1) + \tau(\hat{N}-1), \tag{A.53}$$

$$\tau(\hat{N}-1) < \tau(\hat{N}), \tag{A.54}$$

where the second inequality follows because $\tau(N)$ is strictly increasing for $N \geqslant 1 + T_{\text{lag}}$. Adding inequalities (A.53) and (A.54) yields that $g^*(\hat{N}) < g^*(\hat{N}-1)$ as required.

The second case is more difficult and we resort to an asymptotic analysis over N to establish our claim. We have that $g^*(N) + \tau(N) > g^*(N-1) + \tau(N-1)$ for all $N \geqslant 1 + T_{\text{lag}}$. We use $z^*(N)$ to denote $g^*(N) + \tau(N)$; thus, $z^*(N)$ is a strictly increasing sequence for all $N \geqslant 1 + T_{\text{lag}}$.

From Lemma 6.2, we have that

$$z^*(N) = g^*(N) + \tau(N) = \alpha_0 N d^*(N) + \beta_0 N (d^*(N))^2 \leqslant \alpha_0 \bar{B} + \beta_0 \frac{\bar{B}^2}{N} \leqslant \alpha_0 \bar{B} + \beta_0 \bar{B}^2, \tag{A.55}$$

for all $N \geqslant 1$. That is, $z^*(N)$ is a bounded sequence. Since $z^*(N)$ is strictly increasing, it must then converge. As a result, there exists an $\bar{N}$ large enough such that

$$z^*(N) - z^*(N-1) < \ln 2 / T_{\text{double}}, \quad \forall N \geqslant \bar{N}. \tag{A.56}$$

For all $N \geqslant 2 + T_{\text{lag}}$, the definition of $\tau(N)$ in (2) implies that

$$\tau(N-1) = \frac{((N-2) - T_{\text{lag}}) \ln 2}{T_{\text{double}}}, \tag{A.57}$$

$$\tau(N) = \frac{((N-1) - T_{\text{lag}}) \ln 2}{T_{\text{double}}} \quad \text{and} \tag{A.58}$$

$$\tau(N) - \tau(N-1) = \frac{\ln 2}{T_{\text{double}}}. \tag{A.59}$$

Let $\bar{\bar{N}} = \max\{\bar{N}, 2 + T_{\text{lag}}\}$. Then, (A.56) and (A.59) imply that $z^*(N) - z^*(N-1) < \tau(N) - \tau(N-1)$ for all $N \geqslant \bar{\bar{N}}$. Then using the fact that $z^*(N) = g^*(N) + \tau(N)$ and $z^*(N-1) = g^*(N-1) + \tau(N-1)$, we get,

$$g^*(N) < g^*(N-1), \quad \forall N \geqslant \bar{\bar{N}}. \tag{A.60}$$

Then setting $\hat{N} = \bar{\bar{N}}$, we see that $g^*(\hat{N}) < g^*(\hat{N}-1)$ as required. □

Proof of Property 2. We define the tumour-BE corresponding to dose $b_m(N)$ per fraction over N fractions as

$$g_m(N) \triangleq \alpha_0 N b_m(N) + \beta_0 N b_m^2(N) - \tau(N), \quad m \in \mathcal{M}. \tag{A.61}$$

In this proof of Property 2, we view $d^*(N)$ in (48), the above tumour-BE, as well as $g^*(N)$ in (49) as functions of all real numbers $N \geqslant 1$. Based on the definition of $d^*(N)$ in (48) and also of $g^*(N)$ in (49), we observe that

$$g^*(N) = \min_{m \in \mathcal{M}} g_m(N). \tag{A.62}$$

We will show that $g_m(N)$ is quasiconcave in N over $N \geqslant 1$ for each fixed $m \in \mathcal{M}$. This would imply that $g^*(N)$ is also quasiconcave because it is the minimum of a collection of quasiconcave functions (see Boyd & Vandenberghe, 2004, Section 3.4.4). According to Boyd & Vandenberghe (2004, Section 3.4.4) there are only three possibilities for the structure of a quasiconcave function such as $g^*(N)$: it is non-decreasing; or it is non-increasing; or it is non-decreasing first and then non-increasing. The first possibility is not feasible in view of Property 1. This would prove Property 2.

Consider any fixed $m \in \mathcal{M}$. To show that $g_m(N)$ is quasiconcave, we consider the function

$$f(N) \triangleq \alpha_0 N b_m(N) + \beta_0 N b_m^2(N). \tag{A.63}$$

For algebraic simplicity, we define $D(N) \triangleq N b_m(N)$ as the total tumour-dose delivered in N fractions when the tumour-dose per fraction equals $b_m(N)$. For all $N > 0$, we note that

$$\frac{\mathrm{d}f(N)}{\mathrm{d}N} = \alpha_0 \frac{\mathrm{d}D(N)}{\mathrm{d}N} + \beta_0 \frac{2ND(N)(dD(N)/dN) - D^2(N)}{N^2} \tag{A.64}$$

$$= \alpha_0 \frac{\mathrm{d}D(N)}{\mathrm{d}N} + 2\beta_0 b_m(N) \frac{\mathrm{d}D(N)}{\mathrm{d}N} - \beta_0 b_m^2(N). \tag{A.65}$$

We have,

$$\frac{\mathrm{d}D(N)}{\mathrm{d}N} = \frac{1}{2\sigma_m \rho_m} \left(-1 + \frac{2N + 4\rho_m C_m}{2\sqrt{N^2 + 4\rho_m C_m N}} \right) \tag{A.66}$$

$$= \frac{1}{2\sigma_m \rho_m} \frac{(\sqrt{1 + 4\rho_m C_m/N} - 1)^2}{2\sqrt{1 + 4\rho_m \rho_m/N}} \tag{A.67}$$

for all $N > 0$. We use the notation $r_m \triangleq 1/(2\sigma_m \rho_m)$ and also $\chi(N) \triangleq \sqrt{1 + 4\rho_m C_m/N} - 1$ for brevity. With this notation, we have $b_m(N) = r_m \chi(N)$ and $\mathrm{d}D(N)/\mathrm{d}N = r_m \chi^2(N)/(2(\chi(N) + 1))$. Then substituting these two into (A.64) we obtain,

$$\frac{\mathrm{d}f(N)}{\mathrm{d}N} = \alpha_0 \frac{r_m \chi^2(N)}{2(\chi(N) + 1)} + 2\beta_0 \frac{r_m^2 \chi^3(N)}{2(\chi(N) + 1)} - \beta_0 r_m^2 \chi^2(N) = \frac{(\alpha_0 - 2\beta_0 r_m) r_m \chi^2(N)}{2(\chi(N) + 1)} \tag{A.68}$$

$$= \frac{(\alpha_0 - \alpha_m \beta_0/\sigma_m \beta_m) r_m \chi^2(N)}{2(\chi(N) + 1)}. \tag{A.69}$$

We now consider two cases. The first case is where $\sigma_m \alpha_0/\beta_0 \leqslant \alpha_m/\beta_m$ and the second case is when $\sigma_m \alpha_0/\beta_0 > \alpha_m/\beta_m$. □

Case 1: $\sigma_m\alpha_0/\beta_0 \leqslant \alpha_m/\beta_m$. In this case, (A.69) implies that $\mathrm{d}f/\mathrm{d}N \leqslant 0$ for all $N > 0$. That is, $f(N)$ is non-increasing for all $N > 0$, and in particular, for all $N \geqslant 1$. Since $\tau(N)$ is non-decreasing, this implies that the objective function $g_m(N) = f(N) - \tau(N)$ is non-increasing over $N \geqslant 1$.

Case 2: $\sigma_m\alpha_0/\beta_0 > \alpha_m/\beta_m$. Now, from (A.69), for all $N > 0$, we have,

$$\frac{\mathrm{d}^2 f(N)}{\mathrm{d}N^2} = \frac{\mathrm{d}}{\mathrm{d}N}\left(\frac{\mathrm{d}f(N)}{\mathrm{d}N}\right) = \frac{\mathrm{d}}{\mathrm{d}\chi}\left(\frac{\mathrm{d}f(N)}{\mathrm{d}N}\right)\frac{\mathrm{d}\chi(N)}{\mathrm{d}N} \tag{A.70}$$

$$= \frac{(\alpha_0 - \alpha_m\beta_0/\sigma_m\beta_m)r_m}{2}\frac{2\chi(N)(\chi(N)+1) - \chi^2(N)}{(\chi(N)+1)^2}\frac{\mathrm{d}\chi(N)}{\mathrm{d}N} \tag{A.71}$$

$$= \frac{(\alpha_0 - \alpha_m\beta_0/\sigma_m\beta_m)r_m}{2}\frac{\chi^2(N) + 2\chi(N)}{(\chi(N)+1)^2}\frac{\mathrm{d}\chi(N)}{\mathrm{d}N}. \tag{A.72}$$

Now recall that $\chi(N) = \sqrt{1 + 4C_m\rho_m/N} - 1$ and hence $\chi(N) > 0$ for all $N > 0$. Moreover, $\mathrm{d}\chi(N)/\mathrm{d}N < 0$ for all $N > 0$. Therefore, $\mathrm{d}^2 f(N)/\mathrm{d}N^2 < 0$ and hence $f(N)$ is strictly concave for all $N > 0$ (a proof of strict concavity based on a second derivative was also developed independently by Bortfeld *et al.*, 2013 in their work on a single normal tissue; we nevertheless decided to present our derivatives using our notation here for the sake of completeness). We now have

$$\tau(N) = \frac{[(N-1) - T_{\mathrm{lag}}]^+ \ln 2}{T_{\mathrm{double}}} = \frac{\max((N-1) - T_{\mathrm{lag}}, 0)\ln 2}{T_{\mathrm{double}}} \tag{A.73}$$

and hence $\tau(N)$ is a convex function. As a result, $g_m(N) = f(N) - \tau(N)$ is a concave function over $N \geqslant 1$.

The above two cases show that $g_m(N)$ is either non-increasing or concave and hence it is quasiconcave as claimed. This completes the proof of Property 2.

Proof of Property 3. This property is an immediate consequence of Property 2. □

A.8 *Proof of Corollary* 6.4

The first claim follows from Property 1 in Theorem 6.3. The second claim follows from Property 3 in Theorem 6.3.

Appendix B. The special case of a single normal tissue

In this appendix, we study a special case of our general model in the main body of this paper; it includes a single normal tissue. In this special case, we are able to derive a closed-form formula for the optimal number of fractions. The formula is valid for a normal tissue with maximum dose, mean dose, or dose-volume type constraints. The formula is consistent with other similar closed-form formulas in the existing literature (reviewed in Section 2 of the main body of this paper), after accounting for minor idiosyncratic differences between some of the existing single normal tissue models and our single normal tissue model.

Consider a special case of problem (P) in the main text of this paper with a single normal tissue. Thus, $\mathcal{M}$ now equals the singleton set $\{1\}$ and hence we denote various normal tissue parameters with

a subscript of '1'. Then, problem (P) reduces to

$$(P_{\text{single}}) \quad \max \alpha_0 N b_1(N) + \beta_0 N b_1^2(N) - \tau(N) \tag{B.1}$$

$$N \geqslant 1, \text{ integer}, \tag{B.2}$$

where $b_1(N)$ is now the limiting dose defined as $(-1+\sqrt{1+4\rho_1 C_1/N})/(2\sigma_1\rho_1)$ from (47). We reiterate that problem (P_{single}) simultaneously, although in some cases only slightly, generalizes the models studied by Fowler and his co-authors, Jones *et al.*, Unkelbach *et al.*, Keller *et al.* and Bortfeld *et al.* We next state and prove one lemma that provides an optimal fractionation schedule for problem (P_{single}). This result is not among the main contributions of this paper. We include it here mainly for the sake of completeness, notational consistency, and to ensure that our slightly general result correctly recovers a previously known formula in the literature. Moreover, our formula here is used in Section 7.3 of the main body of this paper while comparing the optimal number of fractions obtained by separately considering different normal tissues one by one with those from our multiple normal tissue model. Before we state our formula, we emphasize here that if $\alpha_0/\beta_0 \leqslant \alpha_1/\beta_1/\sigma_1$, then Proposition 5.3 implies that it is optimal to administer a single fraction. We therefore focus here on the only other possibility for these parameters of our single normal tissue model, namely, wherein $\alpha_0/\beta_0 > \alpha_1/\beta_1/\sigma_1$.

LEMMA B.1 Suppose that $\alpha_0/\beta_0 > \alpha_1/\beta_1/\sigma_1$. Then it is not optimal to administer less than $1+T_{\text{lag}}$ fractions. Let $r_1 \triangleq 1/(2s_{\text{effective}}\rho_1)$ and $\eta \triangleq \ln 2/T_{\text{double}}$. We also define

$$N^* \triangleq \frac{4\rho_1 C_1}{((\eta+\sqrt{\eta^2+2\eta r_1(\alpha_0-\alpha_1\beta_0/\sigma_1\beta_1)})/r_1(\alpha_0-\alpha_1\beta_0/\sigma_1\beta_1)+1)^2-1}. \tag{B.3}$$

If $N^* < 1+T_{\text{lag}}$, then it is optimal to administer $1+T_{\text{lag}}$ fractions; otherwise it is optimal to administer either $\lfloor N^* \rfloor$ or $\lceil N^* \rceil$ fractions depending on which one of these two integers maximizes the objective function in (P_{single}).

Proof. Since $\tau(N)=0$ for all $1 \leqslant N \leqslant 1+T_{\text{lag}}$, the objective function in problem (P_{single}) equals $f(N) \triangleq \alpha_0 N b_1(N) + \beta_0 N b_1^2(N)$ over this range of N. Then with algebra identical to the proof of Property 2 in Theorem 6.3, we obtain,

$$\frac{\mathrm{d}f(N)}{\mathrm{d}N} = \frac{(\alpha_0-\alpha_1\beta_0/\sigma_1\beta_1)r_1\chi^2(N)}{2(\chi(N)+1)}, \tag{B.4}$$

where $\chi(N) = -1+\sqrt{1+4\rho_1 C_1/N}$. Since $\sigma_1\alpha_0/\beta_0 > \alpha_1/\beta_1$, and $\chi(N)+1 = \sqrt{1+4\rho_1 C_1/N} > 0$, the above derivative is positive. This implies that $f(N)$ is strictly increasing over $1 \leqslant N < 1+T_{\text{lag}}$. Thus, it is not optimal to administer less than $1+T_{\text{lag}}$ fractions. We investigate the objective function in (P_{single}) for $N \geqslant 1+T_{\text{lag}}$. Let $\delta = (1+T_{\text{lag}})\ln 2/T_{\text{double}}$. Then, the objective function is given by

$$g^*(N) \triangleq f(N) - \eta N + \delta, \quad \text{for } N \geqslant 1+T_{\text{lag}}, \tag{B.5}$$

where $f(N) = \alpha_0 N b_1(N) + \beta_0 N b_1^2(N)$. Then, similar to (B.4), we get

$$\frac{\mathrm{d}g^*(N)}{\mathrm{d}N} = \frac{(\alpha_0-\alpha_1\beta_0/\sigma_1\beta_1)r_1\chi^2(N)}{2(\chi(N)+1)} - \eta. \tag{B.6}$$

Also, similar to (A.72), we get,

$$\frac{\mathrm{d}^2 g^*(N)}{\mathrm{d}N^2} = \frac{(\alpha_0 - \alpha_1\beta_0/\sigma_1\beta_1)r_1}{2} \frac{\chi^2(N) + 2\chi(N)}{(\chi(N)+1)^2} \frac{\mathrm{d}\chi(N)}{\mathrm{d}N}. \tag{B.7}$$

Now recall that $\chi(N) = \sqrt{1 + 4\rho_1 C_1/N} - 1$ and hence $\chi(N) > 0$. Moreover, $\mathrm{d}\chi(N)/\mathrm{d}N < 0$. Therefore, $\mathrm{d}^2 g^*(N)/\mathrm{d}N^2 < 0$ and hence $g^*(N)$ is concave. Moreover, by Property 1 in Theorem 6.3, we know that $g^*(N)$ has a stationary point. If $g^*(N)$ has a stationary point that is at least $1 + T_{\mathrm{lag}}$, then this stationary point maximizes the objective function in our problem over all real numbers $N \geqslant 1 + T_{\mathrm{lag}}$. If, on the other hand, $g^*(N)$ does not have such a stationary point (i.e. the stationary point is smaller than $1 + T_{\mathrm{lag}}$), then $N = 1 + T_{\mathrm{lag}}$ maximizes the objective function in our problem over all real numbers $N \geqslant 1 + T_{\mathrm{lag}}$. Thus, we need to first find a stationary point of $g^*(N)$. In particular, (B.6) implies that we need to solve the quadratic equation

$$r_1(\alpha_0 - \alpha_1\beta_0/\sigma_1\beta_1)\chi^2(N) - 2\eta\chi(N) - 2\eta = 0 \tag{B.8}$$

for $\chi(N)$. Since $\chi(N)$ must be positive, we investigate whether the above equation has a positive solution. Solutions of this equation are given by

$$\chi(N) = \frac{\eta \pm \sqrt{\eta^2 + 2\eta r_1(\alpha_0 - \alpha_1\beta_0/\sigma_1\beta_1)}}{r_1(\alpha_0 - \alpha_1\beta_0/\sigma_1\beta_1)}, \tag{B.9}$$

and since $\sigma_1\alpha_0/\beta_0 > \alpha_1/\beta_1$,

$$\chi^*(N) = \frac{\eta + \sqrt{\eta^2 + 2\eta r_1(\alpha_0 - \alpha_1\beta_0/\sigma_1\beta_1)}}{r_1(\alpha_0 - \alpha_1\beta_0/\sigma_1\beta_1)} \tag{B.10}$$

is the unique positive solution. Moreover, since $\chi^*(N) = \sqrt{1 + 4\rho_1 C_1/N^*} - 1$, where N^* is the unique stationary point of $g^*(N)$, (B.10) implies that

$$N^* = \frac{4\rho_1 C_1}{((\eta + \sqrt{\eta^2 + 2\eta r_1(\alpha_0 - \alpha_1\beta_0/\sigma_1\beta_1)})/r_1(\alpha_0 - \alpha_1\beta_0/\sigma_1\beta_1) + 1)^2 - 1}. \tag{B.11}$$

Finally, because $g^*(N)$ is concave, and the optimal number of fractions must be an integer, we substitute $N = \lfloor N^* \rfloor$ and $N = \lceil N^* \rceil$ into $g^*(N)$ and find which one of these two yields the higher BE. That is, the optimal number of fractions. □

Recall from Section 1 that Jones *et al.* did not prove that setting the derivative of BE to zero was sufficient for global optimality in their model. Moreover, they ignored the integer restriction on N. Lemma B.1 overcomes these limitations. Also recall that the optimality of an infinite number of fractions when $\sigma_1\alpha_0/\beta_0 > \alpha_1/\beta_1$ was proven in Unkelbach *et al.* and Keller *et al.* However, those papers did not model tumour proliferation. In a special case of our model where there is no tumour proliferation, we have, $\tau(N) = 0$ for all N; thus, the objective function in problem (P_{single}) is given simply by $f(N)$ and the above proof shows that it is strictly increasing over N for all N when $\sigma_1\alpha_0/\beta_0 > \alpha_1/\beta_1$. That is, an infinite number of fractions is optimal and hence Lemma B.1 recovers the result in Unkelbach *et al.* and Keller *et al.* Bortfeld *et al.* also independently proved strict concavity of tumour-BE using a second derivative similar to ours. One very minor difference between our result here and that of Bortfeld *et al.* is that we allow for a time-lag before proliferation begins; setting $T_{\mathrm{lag}} = 0$ thus correctly recovers their result.