An objective function for radiation treatment optimization based on local biological measures

Phys. Med. Biol. **44** (1999) 479–493. Printed in the UK PII: S0031-9155(99)96381-9

An objective function for radiation treatment optimization based on local biological measures

M Alber and F Nüsslin

Abteilung für Medizinische Physik, Universitätsklinikum Tübingen, D-72076 Tübingen, Germany

Received 27 July 1998, in final form 24 November 1998

Abstract. The implementation of biological optimization of radiation treatment plans is impeded by both computational and modelling problems. We derive an objective function from basic model assumptions which includes the normal tissue constraints as interior penalty functions. For organs that are composed of parallel subunits, a mean response model is proposed which leads to constraints similar to dose-volume constraints. This objective function is convex in the case when no parallel organs lie in the treatment volume. Otherwise, an argument is given to show that a number of local minima may exist which are near degenerate to the global minimum. Thus, together with the measure quality of the objective function, highly efficient gradient algorithms can be used. The number of essential biological model parameters could be reduced to a minimum. However, if the optimization constraints are given as TCP/NTCP values, Lagrange multiplier updates have to be performed by invoking comprehensive biological models.

1. Introduction

In inverse treatment planning, considerable effort has recently been directed towards treatment optimization by biological parameters (Brahme 1995, Bortfeld *et al* 1996, Brahme and Lind 1997, Mohan and Wang 1996, Mohan *et al* 1996, Webb 1997, Peacock *et al* 1998). While the pioneering efforts of Webb (1992), Källman *et al* (1992) and Gustafsson *et al* (1994) demonstrated the feasibility of biological treatment optimization, the subject remains topical mainly because of its dependence on semiempirical models and the clinical database.

The inadequacy of purely dose-to-volume based treatment objectives was indicated by Mohan *et al* (1994). Subsequently, a great number of methods to express the basics of biological and clinical knowledge in the language of physical quantities have been proposed for use in inverse treatment planning (Webb 1997, Raphael 1992, Bortfeld *et al* 1997, Cho *et al* 1998, Hristov and Fallone 1997, 1998, Langer and Kijewski 1991, Rowbottom *et al* 1997, Spirou and Chui 1998, Wang *et al* 1995, Preiser *et al* 1997, DeWagter *et al* 1998). With our development, we aim to derive from a set of fundamental biological model assumptions an objective function of numerical expediency.

Since the emphasis of this paper is placed on practical treatment planning, our development reflects the quest for computational efficiency. For this to be achieved, it is of great importance either to formulate the problem as a convex objective function or to devise efficient strategies to deal with local optimal solutions if they arise. Usually, non-convex optimization problems have to be dealt with by stochastic and rather time-consuming algorithms which have the capability of escaping local minima. However, local minima do not have to result from non-convexity, and if they do, they can be virtually indistinguishable from the global solution, which is often not

0031-9155/99/020479+15$19.50

unique in our setting. In the following, we call this property degeneracy. If a problem gives rise to degenerate local minima only, the use of stochastic algorithms is not warranted and fast gradient methods can be employed without the risk of getting trapped in far from optimal solutions. Thus, the use of gradient algorithms can be placed beyond reasonable doubt if an objective function can be devised *without* loss of biological significance which is convex or degenerate.

In its most elementary form, optimization is the choice in favour of the better of two alternatives; in our case this implies that the optimization algorithm has to induce a faithful order on the set of treatment plans according to some biological considerations. This requirement is a precondition, but is not equivalent to the capability of providing predictive assays of treatment outcome. Usually, the latter creates the difficulty of needing to consider interpatient heterogeneity.

Our concept of biological optimization relies on a classification of complications as ultraparallel, serial and parallel. We derive expressions for the objective function for each of these response types. Of particular importance for optimization is the result that in the former two cases the optimization problem can be shown to be convex, whereas the latter case gives rise to a degenerate non-convex problem.

2. Unconstrained optimization with interior penalty functions

From the point of view of numerics, the handiness of dose as an objective stems from one single feature: dose is a mathematical measure, i.e. it is a density function. At the same time, the local character of a measure is critical if non-local effects like the volume effect have to be taken into account. If it can be proven that a complication mechanism is essentially local, then a measure could be a good descriptor of biological effects. If the contrary is true, further assumptions about the nature of radiation damage have to be made. This non-local volume effect has been conjectured for a number of tissues such as lung, kidney or liver, where the bulk of the organ comprises a reserve for partial failure (Emami *et al* 1991, Burman *et al* 1991, Jackson *et al* 1995). It is commonly held that the inclusion of volume effects into optimization leads to non-convex problems (Deasy 1997, Webb 1997, Niemierko 1996). The effects of local minima have been reported to be of minor importance if dose-volume constraints were employed (Bortfeld *et al* 1997, Cho *et al* 1998, Langer and Kijewski 1991, Spirou and Chui 1998, Langer and Leong 1987, Langer *et al* 1990), suggesting a degenerate problem.

It is our aim to show that there is a biologically significant measure for each of the common TCP/NTCP models which can be the basis of an objective function. Although there is little doubt that biological response is intrinsically nonlinear and non-convex†, the situation may be locally convex if one is chiefly interested in the low-NTCP limit. Ideally, the non-local components of the complication mechanism can be dissociated from the local ones by a separation of length scales.

Additional difficulties arise from taking into account the variability of radiation response in the patient population. As we will demonstrate, the measures are invariant under inclusion of interpatient averages. Hence, for the time being, we regard the integral complication probability P_i as the probability of one given individual suffering some complication i.

† Note that from the power law for partial homogeneous irradiation the effective dose

$$d_{\text{eff}} = \left(\frac{1}{V} \int_V d^{1/n} \, \mathrm{d}V \right)^n \qquad n < 1 \tag{1}$$

can be derived (Kutcher and Burman 1989) which does constitute a non-convex function of dose. However, the $1/n$th power of d_{eff} is a convex function! Non-convexity comes into play also by the sigmoid effective dose to complication probability relation.

Since we are chiefly interested in an order on the space of dose distributions, we introduce the concept of a ranking function which can be tailored for the needs of optimization.

Definition. Let $\mathcal{D}$ be the set of all dose distributions $\boldsymbol{d}$. A function

$$W : \mathcal{D} \to \mathbb{R}^+ \tag{2}$$

is called a ranking function if W is continuous and monotonic, and $\inf_{d\in\mathcal{D}}\{W(\boldsymbol{d})\} = 0$.

In the context of this paper, the dose space $\mathcal{D}$ is chosen to be the span of the dose distributions of incident intensities Φ_j, $j = 1, \ldots, n$ of a finite number n of predefined beam ports to ensure convexity of $\mathcal{D}$

$$\mathcal{D} = \text{span}\{T\Phi_j\} \qquad \Phi_j : \mathbb{R}^2 \to \mathbb{R}^+ \tag{3}$$

T being the TERMA operator. In practical computation $\boldsymbol{d}$ is sampled on some grid to become the vector $(\boldsymbol{d})_j^T = (d_1, \ldots, d_N)$.

In the limit of low NTCP, one way to establish W is by

$$P(\boldsymbol{d}) = \frac{W(\boldsymbol{d})}{1 + W(\boldsymbol{d})}. \tag{4}$$

Ranking functions can be used to define constraints similar to minimum TCP or maximum NTCP constraints. For constrained optimization, the result has to meet with all constraints $P_i(\boldsymbol{d}) \leqslant p_i$, $i = 1, \ldots, n$ being the index of the complication. The constraints can also be taken into account by creating a common objective function, where Lagrangian multipliers, $\lambda_i \in \mathbb{R}^+$ mediate the competing treatment objectives.

$$F(\boldsymbol{d}) = \prod_{i=1}^{n}(1 - P_i(\boldsymbol{d}))^{\lambda_i},\ 0 \leqslant P_i < 1. \tag{5}$$

The optimum dose distribution $\hat{\boldsymbol{d}}$ maximizes F. If the P_i are given as NTCP and (1-TCP) values respectively, the objective function takes the form of a maximum likelihood estimator. The Lagrange multipliers λ_i may then be interpreted as the relative frequencies with which complication i occurs. In the context of this paper, we treat the λ_i as generic multipliers, i.e. any constant may be absorbed in λ_i without further notice.

Practically, the logarithm of $1/F$ is minimized rather than F:

$$\begin{aligned} -\log F &= -\sum_{i=1}^{n}\lambda_i \log(1 - P_i) \\ &= \sum_i \lambda_i \log(1 + W_i) \\ &\approx \sum_i \lambda_i W_i \end{aligned} \tag{6}$$

in the limit of small complication probabilities, by virtue of equation (4).

If the intersection of all sets

$$\mathcal{D}_i = \{\boldsymbol{d} \in \mathcal{D} : P_i(\boldsymbol{d}) \leqslant p_i\} \tag{7}$$

commonly called the feasibility space, is not empty, both approaches yield the same dose distribution $\hat{\boldsymbol{d}}$ (save degeneracy) as the extremal value of F. If the intersection is empty, i.e. the problem is overconstrained, the latter method still delivers a result whereas the former naturally does not. Although this result does not meet with all constraints, it is an approximation to the final result which is obtained when the constraints are relaxed depending on the treatment objectives. By altering the set of Lagrange multipliers, a sequence of $\hat{\boldsymbol{d}}^j$ can

be obtained which converges to the desired final result. Notice that although the objective function is of the form $\lambda_i W_i$, the multiplier updates can be performed using computations of NTCP values if there exists a one-to-one relationship between these two treatment descriptors. Hence, one can choose the most expedient ranking function which faithfully maintains this order.

Commonly, Lagrange multipliers are used as 'penalties' or 'weighting factors' in connection with exterior penalty functions, which are devised to penalize the violation of some constraint. These functions are zero within the feasible region. One popular example is the quadratic overdosage penalty function for some maximum dose constraint. Contrary to this approach we deploy interior penalty functions which increase when the constraint is approached from within the feasible region. This reflects the fact that any small dose is potentially harmful.

3. Biological modelling: terminology and concepts

The available clinical data on NTCP makes it difficult to obtain reliable parameters for use in biologically based objective functions. Our aim is to demonstrate that for a biologically valid ranking function much less is needed than the comprehensive information necessary for TCP/NTCP calculations.

A ranking function has to provide an isoeffect number† for a set of concurring dose distributions for one particular set-up. The ranking of treatment plans need only be a faithful one-to-one mapping onto NTCP, with the NTCP becoming a mere translation into a clinically meaningful figure. Of course, this implies that a ranking function must be insensitive to parameter uncertainties arising from interpatient heterogeneity.

For this to work, the ranking function has to model exactly those microscopic tissue dynamics which are only considered by a bulk average in NTCP statistics. However, this does not replace the need for sophisticated NTCP models for assays of treatment outcome. In fact, our development relies on the future availability of these models. Whereas the algorithm can evaluate a plan with the help of its ranking function, the multiplier updates and the evaluation by the therapist has to be based on reliable NTCP figures.

The response of normal tissue to radiation is organized on a number of length scales. Microscopically, repair mechanisms and apoptosis factors, as well as cell migration and repopulation, among a multitude of further factors, have to be considered. Essential to our development is the assumption that there also exists a distinct mesoscopic length scale, defined by the interactions which take place within this range. This may be seen as the characteristic volume, which can be repopulated without loss of structure, it may be identified with structural functional entities like nephrons, or the volume to which a mesoscopic pathological process can be confined. This interpretation is inspired by what was commonly termed a functional subunit, and it includes the notion that for the same organ, yet different complications, there may be different definitions of these mesoscopic units (Withers *et al* 1988). We put forth the following.

Definition. A functional unit (FU) for the complication mechanism i is the smallest structure with a sigmoidal (damage) dose response relationship $p_i(d)$‡ which is essential to the expression of the complication. The average size of a FU is called l_ϕ and defines the transition from microscopic to mesoscopic interactions.

† Common dose-volume histogram reduction algorithms (Kutcher and Burman 1989, Kutcher *et al* 1991, Lyman 1985, Lyman and Wolbarst 1987, 1989) are a special case of a ranking function.

‡ $p_i(d)$ is a function of the chosen fractionation scheme.

Note that inherent to this definition is the notion that the dose is uniform across the size of the FU, and hence p_i is a function, not a functional. The sigmoid dose response (with a point of inflexion at positive dose) implies a certain degree of functional redundancy in the organization of an FU. In some sense, an FU is the smallest essential volume with some reserve towards partial damage.

It has to be assumed that there exist collective effects of FUs on a mesoscopic length scale. For example, one might think of an FU for the endpoint brain infarction as capillaries where the branching structure induces some non-locality to the effects of FU failure. Pneumonitis is another example, since it certainly would not stay confined to the most heavily damaged FUs and will spread. It is clear that there is a multitude of imaginable effects which couple the reaction of individual FUs and influence the course of the complication. Nevertheless, we can make the following distinction.

Definition. A normal tissue integral probability function $P_i(\boldsymbol{d})$ is called serial, when for all volumes v, diam $v > l_\phi$ and $\boldsymbol{d}(\boldsymbol{x} \in v) = \infty$, the relation

$$\lim_{\text{diam } v \to l_\phi} P_i(\boldsymbol{d}) = 1 \tag{8}$$

is true. Likewise, W_i is serial if

$$\lim_{\text{diam } v \to l_\phi} W_i(\boldsymbol{d}) = \infty. \tag{9}$$

All integral probabilities and ranking functions which do not conform to this definition are called parallel.

Note that this definition extends to complications rather than organs. It is possible that an organ can suffer both from parallel and serial complications (for example, pneumonitis/fibrosis). This definition attempts to capture the clinical observation, that within reason there are organs which could be irradiated to extremely high doses in very small volumes without expression of complications, whereas there are also organs which will always suffer from radiation no matter how small the volume destroyed. Candidates for the latter group are, besides any kind of nerve or intestines, also bones and joints, skin or other membranes, and blood vessels. It is obvious, regardless of the details of the objective function, that only for complications identified as serial must there be a whole volume maximum dose constraint.

4. The ultraparallel model of tumour response

For the purposes of treatment optimization, it is sufficient for a tumour response model to reproduce the correct volume and dose dependences of the tumour response. It can be said that despite its simplicity, the Poisson statistics model offers plentiful opportunities to consider intratumour variability and serves as a solid lower bound on the actual tumour control probability if parameters are chosen on the safe side. It is not likely that the basic volume dependency of this model will change with inclusion of clonogen density, variable radiation sensitivity, time dependences, adjuvant therapies, secondary effects (such as necrosis or hypoxia), dose inhomogeneity and interpatient heterogeneity. Even under consideration of these effects, the Poisson assumption is sufficiently justified.

The FU for the complication endpoint tumour regrowth according to our definition is a set of two clonogens; that is if the linear-quadratic law of cell kill is assumed. The standard Poisson TCP function (Nahum and Tait 1992, Webb 1993, Webb and Nahum 1993, Brenner

1993) is given by:

$$1 - P_t = \mathrm{TCP}(\boldsymbol{d}) = \exp\left(- \int_V \sigma(\boldsymbol{x}) \exp(-\alpha(\boldsymbol{x}) d(\boldsymbol{x}))\, \mathrm{d}x^3 \right) \tag{10}$$

where local inhomogeneities in cell density σ and radiation sensitivity α are taken into account†. For use in the log-objective function equation (6) one needs

$$\lambda_t \log(1 - P_t) = -\lambda_t \sum_{j=1}^{N} \sigma_j \exp(-\alpha_j d_j) \tag{11}$$

in the discretized form on a grid of N boxes. If the cell density and radiation sensitivity are assumed to be constant across the tumour volume, the expression can be simplified to

$$\lambda_t \log(1 - P_t) = -\lambda_t \sum_{j=1}^{N} \exp(-\alpha d_j). \tag{12}$$

This function can be interpreted as the probability measure of clonogen survival. In the appendix we show that interpatient heterogeneity does not change the functional form of this measure.

Next we assume that a standard dose level of d_p was prescribed to the tumour. In the ultraparallel model modest overdosage has no detrimental and only little beneficial effect (Suit *et al* 1992, Goitein and Niemierko 1996) if the TCP level is already high, whereas underdosage greatly diminishes chances for tumour control. Hence we find under neglectance of overdosage effects:

$$\lambda_t \log(1 - P_t) \approx -\lambda_t \left(\sum_{d_j < d_p} \exp[\alpha(d_p - d_j)] - 1 + N \right). \tag{13}$$

This equation reduces to the well established quadratic underdosage penalty function when the exponential is expanded to second order.

5. The critical element model for normal body tissues

One of the first models for normal tissue response was the integral dose model by Schultheiss *et al* (1983), which was later termed the critical element model (Wolbarst 1984, Niemierko and Goitein 1991). In our terminology, the critical element model is serial in that tissue damage becomes manifest if one of a set of independent, identical FUs is destroyed.

The critical element model has the unique feature that the mesoscopic response can be related to the macroscopic, so that in principle the FU dose response can be derived from clinical patient data. After Schultheiss *et al* (1983) and Niemierko and Goitein (1991) we write

$$1 - P_s = \prod_j (1 - p(1, d_j))^{v_j} \tag{14}$$

where v_j is the volume fraction irradiated to dose d_j and $p(1, d)$ is the macroscopic probability of organ damage after whole volume irradiation with dose d. Clearly, if $p(1, d_j)$ equals unity, the complication is induced with certainty.

Since the volume effect of the critical element model is a consequence of the augmentation of independent probabilities rather than some interaction between subvolumes of the organ,

† Notice that the influence of the fractionation scheme can be absorbed via the linear-quadratic model by an effective dose $d' = d + \beta/\alpha d^2$.

we term this a *volume effect of the second kind*. Equation (14) amounts to the assumption of the existence of a log-probability measure.

Whereas for modelling purposes this equation is quite valuable, it should not be taken for granted when clinical data are fed in for the $p(1, d)$ function since this dose response is naturally a population average. Furthermore, it is yet an open question to which degree the prerequisites of identity and independence among the FUs are satisfied. However, for small NTCP levels these effects are of little importance. In the limit of small NTCP, <0.2, the precise shape of the FU dose response is virtually meaningless, because the lowest survival probability of all FUs is an upper bound on the survival probability of the whole organ:

$$1 - P_s = \prod_j q(d_j) \leqslant \min_j\{q(d_j)\} = q\left(\max_j\{d_j\}\right) \tag{15}$$

with $q(d) = 1 - p(d)$. Thus, under inhomogeneous irradiation, the NTCP of a serial type organ is chiefly dictated by the volume receiving maximum dose, and accurate knowledge of the FU dose response at or beyond the point of inflexion is irrelevant.

It is possible to derive an expression for the FU dose response in the limit of small FU complication probabilities. For a single FU, there are a great number of possible pathways the destruction can take, all of them equally improbable in the low dose limit. By Poisson statistics we find $q = \exp(-\mu)$, where the expected value μ is dominated by those fatal mechanisms which occur with the greatest probability. Since this probability must be linked to cell kill, one has: $\mu \propto [1 - \exp(-\alpha d - \beta d^2)]^k$, which in the low-dose regime becomes $\mu \propto d^k$. The parameter k takes the smallest possible value and is related to the number of essential structures *within* the FU involved in the collapse. Hence we have

$$q(d) = \exp\left[-\left(\frac{d}{d_s}\right)^k\right] \tag{16}$$

where d_s is some proportionality factor.

Thus, for the ranking function we find (equation (6)):

$$\lambda_s W_s \approx -\lambda_s \log(1 - P_s) \tag{17}$$

$$= -\lambda_s \sum_j \log q_j^{\sigma}(d_j) \tag{18}$$

$$= -\lambda_s \sum_j \log\left\{\exp\left[-\left(\frac{d_j}{d_s}\right)^k\right]\right\} \tag{19}$$

$$= \lambda_s \sum_j \left(\frac{d_j}{d_s}\right)^k \tag{20}$$

where σ is the number of FUs per dose grid cube. The last expression is equivalent to the dose-volume histogram reduction technique proposed by Kutcher and Burman (1989) (see also equation (1))†. The constant d_s can also be absorbed in the Lagrange multiplier, hence

$$\lambda_s W_s = \lambda_s \sum_j d_j^k. \tag{21}$$

This is the kth order moment of the dose distribution. We find that W_s is convex as a consequence of the existence of a local probability measure for elementary complication. This formula has been used to some success in treatment optimization (Webb 1992, Rowbottom *et al* 1997). By comparison with standard objective functions, where the quadratic deviation from a given dose is usually considered, we find that the prescribed dose should be zero (any

† Note that W_s is still serial despite the small NTCP limit approximations made.

dose is potentially harmful) and fluctuations should be handled more stringently (local failure is total failure) (see figure 1). It is remarkable that a single tissue-specific parameter should suffice. In the appendix, we deal with the robustness of this objective function with respect to interpatient heterogeneity.

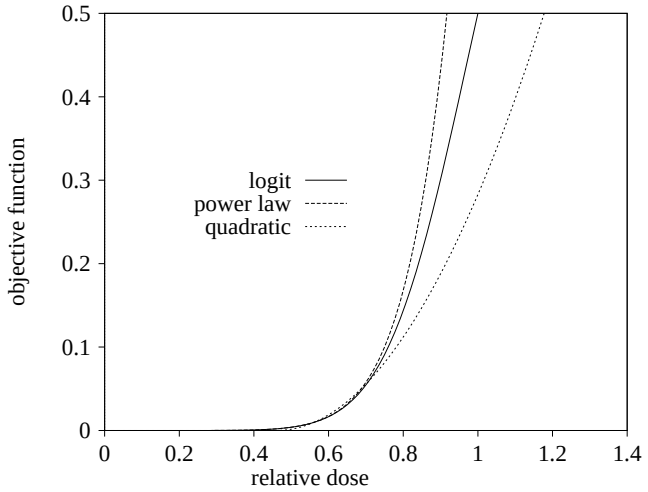

Figure 1. Comparison of three objective functions for serial complications. For shape parameter $k = 8$ both logit and power law functions are steeper than the quadratic overdosage penalty, which in this case was fitted to $p = 0.1\%$ and 5%.

For use in objective functions, we deem it safest to obtain this parameter k in equation (16) from small NTCP volume effect data. With $q(d_v)$ we denote the probability of FU survival after irradiation with a dose d_v. For equal levels of complication probability at partial volume irradiation one obtains from equation (14)

$$q^v(d_v) = q(d_1). \tag{22}$$

Hence with equation (16)

$$d_v = v^{-1/k} d_1 \tag{23}$$

in accordance with the results of Niemierko and Goitein (1991). Thus, the volume effect power law is valid as long as equation (16) holds, which is independent of volume, yet not of the complication level. It has been shown in biological data (Schultheiss *et al* 1983, Niemierko and Goitein 1991) that for smaller volume fractions and high complication levels this power law does not hold. This is chiefly a consequence of the high level of complication occurrence which was considered in the studies and reflects the fact that the low-q condition no longer holds†.

6. A mean response model for normal body tissue

Treatment plan optimization would not be such a difficult problem if there were no tissues with a distinct volume effect. It is apparent that the volume effect of parallel complications is of a

† The functions $p = (d/d_0)^k$ and $p = 1 - [1 + (d/d_0)^k]^{-1}$ are virtually indistinguishable (see figure 1) for low probabilities and yield the same result for the volume effect power law. Since a distinction on the basis of experimental data does not seem to be possible at present, we do not venture to extend the validity of our statements to the high-probability regime.

different nature from that of serial ones. A local probability measure does not seem to exist; instead, the fundamental quantity appears to be a local damage density. In distinction to the former, we call this a *volume effect of the first kind*. Furthermore, the partial volume power law cannot be used to elicit mesoscopic relations since it does not include effects such as the threshold volume and leads to a serial ranking function.

The modelling of parallel complications derives from two assumptions about the complication mechanism which are in accordance with our prior definitions. The first assumption concerns the existence of a local damage density which carries out a local average. As a consequence, the central local quantity is $\rho(d) = \langle p(d) \rangle$ instead of p. It is evident that this necessitates an upper bound on fluctuations of $\boldsymbol{d}$ on the length scale l_ϕ. Technically, this is little more than a mapping of the dose distribution onto the local damage density, and is quite similar to the risk histograms introduced by Jackson *et al* (1993), yet still retains spatial resolution.

The second assumption states that the mesoscopic interactions are well-behaved and can be described by some function of the bulk average. While this is without doubt the case when the dose distribution is homogeneous, any mesoscopic interactions which alter the course of the complication when the damage distribution is inhomogeneous are ignored.

These assumptions lead to a description of the parallel complication in analogy to phase transitions in equilibrium thermodynamics. Let the mean m_1 of $\rho(\boldsymbol{x})$ be given by

$$m_1 = \frac{1}{V} \int_V \rho(\boldsymbol{x}) \, \mathrm{d}^3 x \tag{24}$$

then under negligence of long-range interactions between FUs (i.e. in the limit of homogeneous dose) and low complication probability we expect that

$$1 - P_p \propto \begin{cases} \left(\dfrac{\mu - m_1}{\mu} \right)^\kappa & m_1 < \mu \\ 0 & m_1 \geqslant \mu \end{cases} \tag{25}$$

for some $\kappa > 0$. Notice that if $\kappa \to 0$, equation (25) becomes

$$P_p = \Theta(m_1 - \mu) \tag{26}$$

which was introduced by Jackson *et al* (1993), Yorke *et al* (1993) and Niemierko and Goitein (1992), with $\Theta(x)$ being the Heaviside step function. The parameter μ denotes some mean-field quantity which was termed functional reserve. For homogeneous dose distributions, this parameter is equivalent to the partial volume threshold.

The major shortcoming of models based on these assumptions will be close to the 'critical point', i.e. close to NTCP ≈ 1, where dose inhomogeneity and interactions can no longer be neglected. As a consequence of our development, the slope of the integral response probability will be finite at $P_p = 1$ at contrast to the well known smooth NTCP dose responses. Although this feature could be salvaged by population averaging (Jackson *et al* 1993), we remark that it is essentially an unbiological consequence of the neglect of interactions. Thus, our development should be seen as tailored for the rare incidence of complication.

Given that $\rho \propto d^{1/n}$ and the volume irradiated is larger than μ, we recover the standard power law volume effect for low doses. Commonly, a large volume effect is expressed by the parameter n being in the range of $\approx 0.4 \ldots 0.9$ (Emami *et al* 1991), which means in turn that the FU response is rather shallow with exponents of $\approx 1.2 \ldots 2.5$. Considering that ρ is sigmoidal in shape, the true fractional volume relationship should deviate from the power law already for $v \approx 2\mu$ to higher isodoses for smaller volume fractions. The data on this subject are indecisive. Interpatient heterogeneity certainly assumes a considerable influence on the magnitude of the volume effect. The main reason for this might be that patients while having

comparable FU dose response have vastly differing functional reserves μ due to additional confounding factors, so that μ has to be chosen with care.

For the ranking function one has with equation (25)

$$-\lambda_p \log(1 - P_p) = -\lambda_p \log\left(1 - \frac{m_1}{\mu}\right) \tag{27}$$

$$= -\lambda_p \log\left(1 - \frac{\sum_j \rho(d_j)}{\mu}\right) \tag{28}$$

$$\approx \lambda_p\left(C + \frac{1}{\mu - m_1} \sum_j \rho(d_j)\right). \tag{29}$$

The last equation is quite instructive when compared with equation (21), the corresponding equation for serial complications. Whereas in the serial case the objective function tends to infinity term by term when the constraint is violated locally, in the parallel case the objective function increases collectively by means of a prefactor. Thus, mean response allows for local violation of a whole organ maximum dose constraint in a partial volume of at most $v = \mu$. This trait makes it quite similar to dose-volume constraints, with the distinction that the overdose volume is a function of the dose applied to the remaining volume.

With equation (27) we have transformed the parallel NTCP constraint into *two* interior penalty functions with *one* Lagrange multiplier. In addition to the measure-like term, the first factor penalizes the irradiated volume and limits it to the maximum μ. This interpretation is valid if a fixed number of beam ports is preselected, since the irradiated volume must be assumed to suffer nearly total damage in sensitive organs such as lung. Hence, the mean damage is closely related to the fraction of the total volume which receives a dose higher than the dose at the point of inflexion of the FU dose response. Note that this prefactor could also be absorbed in the Lagrange multiplier and updated every few iterations during optimization.

If d_{PI} is the dose at the point of inflection of the FU dose response ρ and $\rho(d_{\mathrm{PI}}) \approx 1/2$, then the problem is convex if $\boldsymbol{d} < d_{\mathrm{PI}}$. In case the dose exceeds this level, the problem is most likely degenerate as we will show below.

7. Summary of objective functions

In the previous sections our goal was to demonstrate that the problem of biological modelling can be cut down in complexity and number of parameters to yield an expedient objective function. We summarize the required biological model parameters:

(i) *Tumour tissue: ultraparallel.* In its basic form for homogeneous cell density, one parameter suffices: cell radiation sensitivity α. The objective function is given by equation (12). A TCP model is needed for tuning the Lagrange multiplier during iterated optimization.

(ii) *Normal tissue: serial.* Accounting for serial tissue requires one parameter: the partial volume power law exponent n. The objective function is given by equation (21). Multiplier tuning can be performed by the critical element model without knowledge of mesoscopic parameters.

(iii) *Normal tissue: parallel.* An NTCP model is an indispensable requirement. Two constraints are blended to yield the objective equation (27). These constraints correspond to a local maximum dose constraint to limit damage to FUs, and a total damage constraint to limit the gross damage to the functional reserve. At present, parameters for semiempirical models for the FU response should be chosen with care. However, variations in the functional form are unlikely to show up in the results. The functional reserve parameter is much harder to estimate and has to be set according to clinical practice. In most cases, a total of three

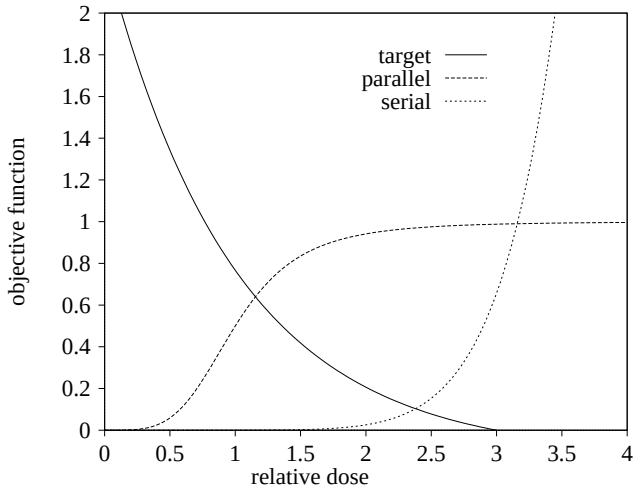

Figure 2. The three components of an objective function with three volumes of interest. The target dose is set to three times the 'tolerance' dose of the parallel and the serial organ. The point of inflexion dose d_{PI} is set to 1. Objectives are not drawn to scale.

parameters suffices: the TD_{50} and slope γ_{50} for the FU response and the organ functional reserve parameter μ.

Examples for the objective function constituents are given in figure 2. Serial organs in the treatment volume result in a convex objective function. If the treatment volume contains parallel organs, the objective function can potentially become non-convex if the maximum dose d_{IP} is exceeded in this volume. Although we see no way to preclude the existence of local minima, there is a fast method to check if the problem is degenerate. Recall that the source of the non-convexity is the sigmoid FU dose-response $\rho(d)$. If we replaced $\rho(d)$ by $\breve{\rho}(d)$ which for doses greater than d_{PI} follows the tangent on ρ in the point of inflexion, only the local minimum $\breve{\boldsymbol{d}}$ with smallest NTCP of the parallel complication prevails. Solving this altered optimization problem and returning to ρ delivers a robust upper bound on the objective function value of the global minimum. Hence, any solution with objective function value greater than this value has to be discarded. If the objective function is minimized again with $\breve{\boldsymbol{d}}$ as a starting point, the gradient algorithm will most likely converge to the local minimum with the lowest NTCP for the parallel complication.

The level of degeneracy of the problem is related to the number of voxels in the parallel organ with a dose around d_{PI}. If the algorithm can redistribute the dose among these voxels without incurring 'costs' due to other objectives, a certain number of degenerate local minima will arise. Thus, the 'separation' of these minima is given by the contribution of these voxels to the objective function by virtue of the measure property. In the shallow, piecewise-constant regions of the FU dose response the objective function is determined by the steeper objectives of target volume and serial complications which effectively convexify the objective function. For few beam ports, a very low level of degeneracy can be expected. With an increasing number of beam ports, degeneracy increases while the minimum objective function value saturates.

Notice, that the non-convex problem which would arise from an attempt to optimize beam angles is of an entirely different quality and can by no means be convexified or computed by gradient algorithms. The reason for this is that any set of finite sets of beam ports is a non-convex subset of the set of infinite sets of beam ports.

8. Conclusion

Our development intends to integrate the basic current biological modelling into an objective function. With the emphasis on computational expediency, we reduced the modelling to the essential aspects and expressed the problem in a convex function.

This was achieved mainly by three steps. Firstly, we moved on to a log-probability objective function which includes the normal tissue constraints as interior penalty functions. Secondly, we restricted the development to the low-NTCP limit, which simplifies the objective function, particularly for serial complications in the critical element model. Thirdly, we demonstrated that interpatient averaging can be absorbed in the objective function. As a consequence, the objective function has the properties of a biological measure which is of great importance in devising efficient algorithms for finding an optimum set of beam ports.

The modelling of parallel complications followed an extension of the critical volume mean-response model and led to an objective function similar to dose-volume constraints for which positive results have been reported by many authors (Bortfeld *et al* 1997, Cho *et al* 1998, Langer and Leong 1987, Langer *et al* 1990, Preiser *et al* 1997, Spirou and Chu 1998, Wang *et al* 1995). We gave a heuristic argument for the experimental finding that no severe problems arise from the non-convexity in practice.

The advantages of this objective function have to be paid for by the need to perform regular updates of Lagrange multipliers which carry significant biological information. Therefore, if treatment objectives are specified as TCP/NTCP targets, comprehensive models with the full number of parameters have to be used. Together with an ever improving clinical data base, fine tuning to therapists' experience will increase the gain of optimization results.

Acknowledgments

The authors gratefully acknowledge the share Thomas Bortfeld and the referees had in improving the manuscript.

Appendix

Taking into account the inescapable interpatient heterogeneity it is necessary to link biological models to predictive assays such as TCP and NTCP. However, for optimizing purposes it is sufficient to show that the result is robust with respect to parameter fluctuations in the population.

For want of more detailed knowledge, a Gaussian distribution is assumed for the relevant biological parameters. If the averaging is invoked on the log-probabilities used in the derivation of the objective function, we find that the functional structure is invariant, yet parameters change.

The Poisson model for TCP

The patient averaged objective for ultra parallel complications reads

$$\langle -\log F\rangle = 1/N \sum_i \int_0^\infty \exp\left(\frac{(\alpha-\alpha_0)^2}{2\sigma_\alpha^2}\right)\exp(-\alpha d_i)\,\mathrm{d}\alpha \tag{30}$$

$$\approx \sum_i \exp[-(\alpha_0 - \sigma_\alpha^2/2\,d_i)d_i] \tag{31}$$

N being a normalization factor. The approximation is valid if $\sigma < \alpha_0$. In this case, it is possible to replace $\alpha_0 - \sigma_\alpha^2/2\,d_i$ with $\alpha_0' = \alpha_0 - \sigma_\alpha^2/2\,d_p$. Note that the second parameter, the variability in cell density is absorbed in the 'population averaged' Lagrange multiplier, i.e. is taken into account at the multiplier updates according to a TCP calculation.

The critical element model

Recall that in the very low NTCP limit the FU dose response can be written as $(d/d_0)^k$ so that the factor d_0^k can be absorbed in a 'population averaged' Lagrange multiplier. For the population average of the objective function we find

$$\langle -\log F\rangle = 1/N \sum_i \int_0^\infty \exp\left(\frac{(k-k_0)^2}{2\sigma_k^2}\right) d_i^k \,\mathrm{d}k \tag{32}$$

$$\approx \sum_{d_i > \exp(-k_0/\sigma_k^2)} d_i^{k_0+\sigma_k^2/2\ln d_i} \tag{33}$$

if the normalized dose $d_i > \exp(-k_0/\sigma_k^2)$. Again, for $k_0 \approx 10\sigma_k^2$, the exponent can be approximated by k_0.

The mean-response model

Although the FU dose response is likely to be the only individual quantity accessible (Boersma *et al* 1994, 1995), it is not available for planning. Hence, a patient average has to be used from the start. At any rate, the uncertainty about the functional reserve is to determine the treatment outcome more dominantly than any other factor. The complication will be expressed by the fraction of the population with functional reserve smaller than the threshold μ, thus rendering a population average of the volume prefactor $1/(\mu - m_1)$ obsolete.

References

Boersma L J, Damen E M F, de Boer R W, Muller S H, Roos C M, Valdes Olmos R A, van Zandwijk N and Lebesque J V 1994 Dose-effect relations for local functional and structural changes of the lung after irradiation for malignant lymphoma *Radiother. Oncol.* **32** 201–9

——1995 Estimation of overall pulmonary function after irradiation using dose-effect relations for local functional injury *Radiother. Oncol.* **36** 15–23

Bortfeld T, Schlegel W, Dykstra C, Levegrün S and Preiser K 1996 Physical vs biological objectives for treatment plan optimization *Radiother. Oncol.* **40** 185

Bortfeld T, Stein J and Preiser K 1997 Clinically relevant intensity modulation optimization using physical criteria *Proc. 12th Int. Conf. on the Use of Computers in Radiation Therapy (Salt Lake City)* ed D D Leavitt and G Starkschall (Madison, WI: Medical Physics Publishing) pp 1–4

Brahme A 1995 Treatment optimization using physical and radiobiological objective functions *Radiation Therapy Physics* ed A R Smith (Berlin: Springer) pp 209–46

Brahme A and Lind B K 1997 The importance of biological modeling in intensity modulated radiotherapy optimization *Proc. 12th Int. Conf. on the Use of Computers in Radiation Therapy (Salt Lake City)* ed D D Leavitt and G Starkschall (Madison, WI: Medical Physics Publishing) pp 5–8

Brenner D J 1993 Dose, volume, and tumor-control predictions in radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **26** 171–9

Burman C, Kutcher G J, Emami B and Goitein M 1991 Fitting of normal tissue tolerance data to an analytic function *Int. J. Radiat. Oncol. Biol. Phys.* **21** 123–35

Cho P S, Lee S, Marks II R J, Oh S, Sutlief S G and Phillips M H 1998 Optimization of intensity modulated beams with volume constraints using two methods: Cost function minimization and projections onto convex sets *Med. Phys.* **25** 435–43

Deasy J O 1997 Multiple local minima in radiotherapy optimization problems with dose-volume constraints *Med. Phys.* **24** 1157–61

De Wagter C, Colle C O, Fortan L G, Van Duyse B B, Van den Berge D L and De Neve W J 1998 3D conformal intensity-modulated radiotherapy planning: interactive optimization by constrained matrix inversion *Radiother. Oncol.* **47** 69–76

Emami B, Lyman J, Brown A, Coia L, Goitein M, Munzenrieder J E, Shank B, Solin L J and Wesson M 1991 Tolerance of normal tissue to therapeutic irradiation *Int. J. Radiat. Oncol. Biol. Phys.* **21** 109–22

Goitein M and Niemierko A 1996 Intensity modulated therapy and inhomogeneous dose to the tumor: a note of caution *Int. J. Radiat. Oncol. Biol. Phys.* **36** 519–22

Gustafsson A, Lind B K and Brahme A 1994 A generalized pencil beam algorithm for optimization of radiation therapy *Med. Phys.* **21** 343–56

Hristov D and Fallone D 1997 An active set algorithm for treatment planning optimization *Med. Phys.* **24** 1455–64

——1998 A continuous penalty function method for inverse treatment planning. *Med. Phys.* **25** 208–23

Jackson A, Kutcher G J and Yorke E D 1993 Probability of radiation-induced complications for normal tissues with parallel architecture subject to non-uniform irradiation *Med. Phys.* **20** 613–25

Jackson A, Ten Halen R K, Robertson J M, Kessler M L, Kutcher G J and Lawrence T S 1995 Analysis of clinical complication data for radiation hepatitis using a parallel architecture model *Int. J. Radiat. Oncol. Biol. Phys.* **31** 883–91

Källman P, Lind B K and Brahme A 1992 An algorithm for maximizing the probability of complication free tumor control in radiation therapy *Phys. Med. Biol.* **37** 871–90

Kutcher G J and Burman C 1989 Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method *Int. J. Radiat. Oncol. Biol. Phys.* **16** 1623–30

Kutcher G J, Burman C, Brewster L, Goitein M and Mohan R 1991 Histogram reduction method for calculating complications probabilities for three-dimensional treatment planning evaluations *Int. J. Radiat. Oncol. Biol. Phys.* **21** 137–46

Langer M, Brown R, Urie M, Leong J, Stracher M and Shapiro J 1990 Large scale optimization of beam weights under dose volume restrictions *Int. J. Radiat. Oncol. Biol. Phys.* **18** 887–93

Langer M and Kijewski P 1991 CCRT for non-small cell lung cancer: sensitivity of clinical gains to organ tolerance restrictions *Int. J. Radiat. Oncol. Biol. Phys.* **22** 325–32

Langer M and Leong J 1987 Optimization of beam weights under dose-volume restrictions *Int. J. Radiat. Oncol. Biol. Phys.* **13** 1255–60

Lyman J T 1985 Complication probabilities as assessed from dose-volume histograms *Radiat. Res.* **104** S13–S19

Lyman J T and Wolbarst A B 1987 Optimisation of radiation therapy iii: a method of assessing complication probabilities from dose-volume histograms *Int. J. Radiat. Oncol. Biol. Phys.* **13** 103–9

——1989 Optimisation of radiation therapy iv: a dose-volume histogram reduction algorithm *Int. J. Radiat. Oncol. Biol. Phys.* **17** 433–6

Mohan R, Ling C C, Stein J and Wang X-H 1996 The number of beams in intensity-modulated treatments: in response to Drs Söderström and Brahme *Int. J. Radiat. Oncol. Biol. Phys.* **34** 758–9

Mohan R and Wang X-H 1996 Response to Bortfeld *et al* re physical vs biological objectives for treatment plan optimization *Radiother. Oncol.* **40** 186–7

Mohan R, Wang X, Jackson A, Bortfeld T, Boyer A L, Kutcher G J, Leibel S A, Fuks Z and Ling C C 1994 The potential and limitations of the inverse radiotherapy technique *Radiother. Oncol.* **32** 232–48

Nahum A E and Tait D M 1992 Maximising tumour control by customized dose prescription for pelvic tumours *Advanced Radiation Therapy: Tumour Response Monitoring and Treatment Planning* ed A Breit (Heidelberg: Springer) pp 425–3

Niemierko A 1996 A selection of objective functions for optimization of intensity modulated beams *Med. Phys.* **23** 1172

Niemierko A and Goitein M 1991 Calculation of normal tissue complication probability and dose-volume histogram reduction schemes for tissues with a critical element architecture *Radiother. Oncol.* **20** 166–76

——1992 Modeling of normal tissue response to radiation: the critical volume model *Int. J. Radiat. Oncol. Biol. Phys.* **25** 135–45

Peacock J H, Nahum A E and Steel G G 1998 Normal-tissue effects in radiotherapy: physics meets biology: report on a workshop held at Hartsfield Manor, Betchworth, Surrey, UK, 14–16 July 1997 *Int. J. Radiat. Biol.* **73** 341–4

Preiser K, Bortfeld T, Hartwig K, Schlegel W and Stein J 1997 A new program for inverse treatment planning *Proc. 12th Int. Conf. on the Use of Computers in Radiation Therapy (Salt Lake City)* ed D D Leavitt and G Starkschall (Madison, WI: Medical Physics Publishing) pp 425–8

Raphael C 1992 Mathematical modelling of objectives in radiation therapy treatment planning *Phys. Med. Biol.* **37** 1297–311

Rowbottom C G, Webb S and Oldham M 1997 Determination of the optimum beam configurations in radiotherapy treatment planning *Proc. 12th Int. Conf. on the Use of Computers in Radiation Therapy (Salt Lake City)* ed D D Leavitt and G Starkschall (Madison, WI: Medical Physics Publishing) pp 306–9

Schultheiss T E, Orton C G and Peck R A 1983 Models in radiotherapy: volume effects *Med. Phys.* **10** 410–15

Spirou S V and Chui C-S 1998 A gradient inverse planning algorithm with dose-volume constraints *Med. Phys* **25** 321–33

Suit H, Skates S, Taghian A, Okunieff P and Efird J T 1992 Clinical implications of heterogeneity of tumor response to radiation therapy *Radiother. Oncol.* **25** 251–60

Wang X-H, Mohan R, Jackson A, Leibel S A, Fuchs Z and Ling C C 1995 Optimization of intensity-modulated 3d conformal treatment plans based on biological indices *Radiother. Oncol.* **37** 140–52

Webb S 1992 Optimization by simulated annealing of three-dimensional conformal treatment planning for radiation fields defined by a multileaf collimator: 2. inclusion of two-dimensional modulation of the x-ray intensity *Phys. Med. Biol.* **37** 1689–704

——1993 *The Physics of Three-dimensional Radiation Therapy: Conformal Radiotherapy, Radiosurgery and Treatment Planning* (Bristol: IOP Publishing)

——1997 *The Physics of Conformal Radiotherapy: Advances in Technology* (Bristol: IOP Publishing)

Webb S and Nahum A E 1993 A model for calculating tumour control probability in radiotherapy including the effects of inhomogeneous distributions of dose and clonogenic cell density *Phys. Med. Biol.* **38** 653–66

Withers H R, Taylor J M G and Maciejewski B 1998 Treatment volume and tissue tolerance *Int. J. Radiat. Oncol. Biol. Phys.* **14** 751–9

Wolbarst A B 1984 Optimization of radiation therapy: ii, the critical-voxel model *Int. J. Radiat. Oncol. Biol. Phys.* **10** 741–5

Yorke E D, Kutcher G J, Jackson A and Ling C C 1993 Probability of radiation-induced complications in normal tissues with parallel architecture under conditions of uniform whole or partial organ irradiation *Radiother. Oncol.* **26** 226–37