Treatment planning optimisation methods for individualised dose prescription in intensity-modulated radiation therapy

Aswin L. Hoffmann

Treatment planning optimisation methods for individualised dose prescription in intensity-modulated radiation therapy

Development and application of mathematical optimisation methods for inter- and intra-patient dose prescription customisation

Treatment planning optimisation methods for individualised dose prescription in intensity-modulated radiation therapy

Development and application of mathematical optimisation methods
for inter- and intra-patient dose prescription customisation

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op
dinsdag 17 december 2013
om 10:30 uur precies

door

Aswin Louis Hoffmann
geboren op 27 februari 1970
te Roosendaal en Nispen

Promotores

Prof. dr. J.H.A.M. Kaanders
Prof. dr. ir. D. den Hertog   Tilburg University

Manuscriptcommissie

Prof. dr. B. van Ginneken
Prof. dr. B.J.M. Heijmen   Erasmus Universiteit Rotterdam
Prof. dr. D. De Ruysscher   Universitair Ziekenhuis Leuven

In memory of my parents

Contents

List of abbreviations

3D-CRT	three-dimensional conformal radiation therapy
AAPM	American Association of Physicists in Medicine
BGRT	biologically-guided radiation therapy
CT	computed tomography
DNA	deoxyribonucleic acid
DV	dose-volume
DVH	dose-volume histogram
FMO	fluence map optimisation
HATD	highest achievable tumour dose
IMAT	intensity-modulated arc therapy
IMRT	intensity-modulated radiation therapy
LQ	linear-quadratic
MCO	multi-criteria optimisation
MLC	multi-leaf collimator
MRI	magnetic resonance imaging
NSCLC	non-small-cell lung cancer
NTCP	normal tissue complication probability
OAR	organ at risk
PEF	Pareto efficient frontier
PET	positron emission tomography
QoL	quality of life
QUANTEC	QUantitative Analysis of Normal Tissue Effects in the Clinic
SABR	stereotactic ablative body radiotherapy
SBRT	stereotactic body radiation therapy
TCP	tumour control probability
TG	task group
TOC	therapeutic operating characteristic
TPS	treatment planning system
WHO	World Health Organization
WIF	Withers iso-effect formula

List of papers

I. H. Huizenga, **A.L. Hoffmann**, J.H.A.M. Kaanders, P.F.M. Stalmeijer, and J.J. van Tol-Geerdink. Inclusion of individual patient preferences into the radiotherapy treatment planning process. In J.P. Bissonnette, editor, *The Radiotherapy Dynamic*, volume 1 of *Proceedings of the XVth International Conference on the Use of Computers in Radiation Therapy (ICCR)*, pages 362–5, Toronto, 2007.

II. **A.L. Hoffmann**, H. Huizenga, and J.H.A.M. Kaanders. Employing the therapeutic operating characteristic (TOC) graph for individualised dose prescription. *Radiat Oncol*, **8**:55, 2013.

III. **A.L. Hoffmann**, E.G.C. Troost, H. Huizenga, J.H.A.M. Kaanders, and J. Bussink. Individualized dose prescription for hypofractionation in advanced non-small-cell lung cancer radiotherapy: an *in silico* trial. *Int J Radiat Oncol Biol Phys*, **83**:1596–602, 2012.

IV. **A.L. Hoffmann** and A.E. Nahum. Fractionation in normal tissues: the $(\alpha/\beta)_{\text{eff}}$ concept can account for dose heterogeneity and volume effects. *Phys Med Biol*, **58**:6897–914, 2013.

V. **A.L. Hoffmann**, A.Y.D. Siem, D. den Hertog, J.H.A.M. Kaanders, and H. Huizenga. Derivative-free generation and interpolation of convex Pareto optimal IMRT plans. *Phys Med Biol*, **51**:6349–69, 2006.

VI. A.Y.D. Siem, D. den Hertog, and **A.L. Hoffmann**. A method for approximating univariate convex functions using only function value evaluations. *INFORMS J Comput*, **23**:591–604, 2011.

VII. R. Monshouwer, **A.L. Hoffmann**, M. Kunze-Busch, J. Bussink, J.H.A.M. Kaanders, and H. Huizenga. A practical approach to assess clinical planning tradeoffs in the design of individualized IMRT treatment plans. *Radiother Oncol*, **97**:561–6, 2010.

VIII. **A.L. Hoffmann**, D. den Hertog, A.Y.D. Siem, J.H.A.M. Kaanders, and H. Huizenga. Convex reformulation of biologically-based multi-criteria intensity-modulated radiation therapy optimization including fractionation effects. *Phys Med Biol*, **53**:6345–62, 2008.

IX. A.Y.D. Siem, D. den Hertog, and **A.L. Hoffmann**. The effect of transformations on the approximation of univariate (convex) functions with applications to Pareto curves. *Eur J Oper Res*, **189**:347–62, 2008.

X. **A.L. Hoffmann**, E.G.C. Troost, J. Bussink, and J.H.A.M. Kaanders. Physico-biological treatment plan optimisation for homogeneous and heterogeneous individualised dose escalation in advanced non-small-cell lung cancer radiotherapy. *Radiother Oncol*, 2013. [Submitted].

General introduction

1

Radiation therapy

1.1 Principles and practice

According to the Global Health Observatory of the World Health Organization (WHO), cancer is a leading cause of death worldwide and accounted for 7.6 million deaths (13% of all deaths) in 2008 [1]. The WHO regions for Europe and the Americas had the highest incidence of all types of cancer combined for both sexes. In these regions, cancer has been the predominant mortality risk since 2008. Lung, breast, stomach, liver and colorectal cancers cause the majority of deaths.

Surgery, systemic agents and ionising radiation are the three main cancer treatment modalities. In many solid tumours, surgery is the primary option for local treatment of the tumour and/or regional lymph node metastases with a curative intent. Systemic treatments (*e.g.* chemotherapy, biological agents) globally affect tumour cell survival mechanisms and are mostly administered in combination with surgery or radiotherapy in a curative setting, or as single modality for palliative intent. Radiation therapy, also called 'radiotherapy', is considered the treatment of choice in approximately 50% of the cases. It is also applied for the treatment of solid tumours, and is specifically suitable for surgically inaccessible tumours or for organ-preserving treatment strategies. Radiotherapy is administered with curative or palliative intent, and may be applied solely or in combination with surgery and/or chemotherapy.

The goal of curative radiotherapy is to deliver a therapeutic dose of ionising radiation to tumourous target tissues in order to sterilise the proliferation of clonogenic cells while minimising the risks of complications to the surrounding normal tissues. In a palliative setting, the dose administered is restricted and the aim is to alleviate pain symptoms in order to achieve the highest possible quality of life. Radiotherapy is a common treatment for many different types of malignancies, such as cancer of the prostate, breast, lung, rectum, brain, skin and head-and-neck region.

In radiotherapy, the patient is either exposed to an internally or externally applied radiation source, indicated as brachytherapy (derived from the ancient Greek words βραχυζ and θεραπεια meaning 'short distance' and 'healing', respectively) and teletherapy (derived from the ancient Greek word τηλε, meaning 'distant'), respectively. In brachytherapy, a sealed source containing a relatively low-energy photon emitting radioactive isotope is placed within or in close proximity to the tumour in order to provide local radiation therapy. This approach typically delivers very high doses to relatively small target volumes that can be reached via natural or artificial orifices. In modern teletherapy, the source is either based on a high-energy ray emitting radionuclide (*e.g.* ^{60}Co tomotherapy [2]) or a particle accelerator electronically generating highly energetic rays of particles (electrons, protons, neutrons, heavy or light ions) that are collimated to produce a beam of ionising radiation, see *e.g.* [3]. When the external beam is pointed onto the patient's body and penetrates through the skin and the tissues underneath it, the energetic particles interact with matter by electrical forces and lose kinetic energy while depositing dose (measured in J/kg) into the tissue at a certain depth.

1.2   External beam radiotherapy: development of delivery techniques

Most contemporary radiotherapy is carried out with linear accelerators (linacs) that produce megavolt photon X-ray beams, which are directed to the target volume (*i.e.* the gross tumour volume plus a margin for spatial uncertainties) from several different orientations around the patient to adequately cover the often deeply situated target volume while preventing from overdosing surrounding healthy normal tissues. The linac has several degrees of freedom to achieve this, typically comprising of a rotating gantry, beam collimator and treatment couch (Figure 1.1). Furthermore, the beam energy and the beam shape can be altered.
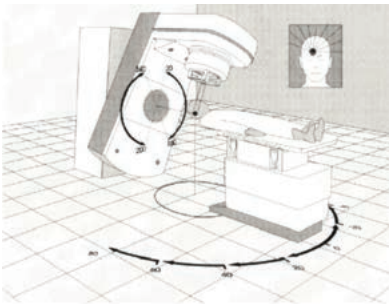
Figure 1.1: Schematic representation of rotational degrees of freedom of a linear accelerator treating a patient's brain tumour from different directions. (Reprinted with permission from [4].)

In conventional conformal radiation therapy a small number of different geometrical beam shapes are adapted to conform to the anatomy of the individual target volume by partially blocking the beam with customised apertures. Each of the apertures could be used with either a uniform or coarsely modified radiation intensity level across the beam, the latter being re-

alised by pre-fabricated wedge filters. Since the intersection of these shaped beams defines the volume that receives the highest radiation dose, there are limited possibilities to fit the shape of the resulting dose distribution to the shape of the target, especially in the case of irregularly shaped concave target volumes.

Intensity-modulated radiation therapy

Improved beam shaping was accomplished through technological advances in hardware and software that took place over the past two decades when linear accelerators were equipped with a computer-controlled multi-leaf collimator (MLC) system mounted in the gantry head (Figure 1.2). In three-dimensional conformal radiation therapy (3D-CRT), the pairwise-opposed tungsten leaves are independently positioned for each beam to match the projection of the target volume onto the fluence plane of the beam. By dynamically blocking different parts of the beam during irradiation with the sliding leaves, the MLC is able to produce segmented fluence intensity-modulated fields that use a much larger number of apertures than in conventional conformal radiotherapy. This form of multi-field intensity-modulated radiation therapy (IMRT) allows to create complex, highly conformal uniform target dose distributions that encouraged efforts of dose escalation and paved the road for non-uniform 'dose sculpting' [5–7]. In recent years, this technique has been further developed and was converted from multi-field
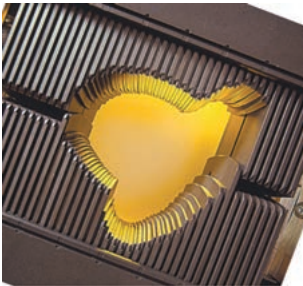
Figure 1.2: Multi-leaf collimator used for IMRT delivery. (Image courtesy of Varian Medical Systems, Inc., Palo Alto, USA.)

IMRT with static beam directions into continuously rotating intensity modulation, originally designated as intensity-modulated arc therapy (IMAT) [8–10], but nowadays known as 'conformal rotational radiotherapy' or 'conformal arc-therapy', making treatment delivery faster than for conventional IMRT.

2

The IMRT planning optimisation problem

Alongside the extensive technological developments in hardware, also significant advances in software were made. Computerised treatment planning systems (TPSs) were not only able to perform accurate dose computations, but also different optimisation approaches were introduced for the following reasons: 1) to deal with the vastly increased number of degrees of freedom that came with IMRT and 2) because new imaging modalities (*e.g.* computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET)) provided 3D morphological and physiological images of the patient's anatomy and the tumour's biology, respectively, allowing for improved normal tissue sparing, whole-target dose escalation and selective intra-tumoural dose (de-)escalation (*i.e.* dose painting) [11]. Following the latter approach, there is a need to spatially adapt the dose distribution not only to the physical conformality (*i.e.* using anatomical data), but also to the biological conformality (*i.e.* using metabolic, functional, genotypic and phenotypic data).

Designing an adequate IMRT plan that handles the inherently conflicting goals of delivering a sufficiently high therapeutic dose to the target volume while avoiding injurious doses to the normal tissue, comprises choices to be made principally relating to medical, biological and physical parameters. While the decisions for a particular treatment regimen (*e.g.* tumour dose prescription, normal-tissue tolerance dose limits, time-dose fractionation, etc.) are predominantly determined by consensus opinions established through empirical knowledge and experience of the radiation oncologist and results of clinical trials, the physical characteristics of the treatment (*i.e.* radiation modality, treatment setup geometry, beam arrangement, beam settings, etc.) are mainly defined by the expertise of the medical physicist. Finding optimal balances in these decisions challenges both the radiation oncologist and the medical physicist in their daily practice.

To introduce the main contributions of the work described in this thesis, the problem statement is illuminated through a multi-disciplinary review, including perspectives from the fields of physics, mathematics, biology, and the clinic.

2.1 Physical perspective: the reverse engineering approach

Computerised treatment plan design

It has been acknowledged that with the advancement of IMRT and the associated level of complexity in designing treatment plans, it was crucial to make an interdisciplinary effort to integrate sophisticated optimisation techniques into the treatment planning process and to explore methods to optimise the clinical outcome of radiotherapy [12]. The major problem, however, is that clinical knowledge may be difficult to acquire and formulate into computer-interpretable format. Physics expertise related to tuneable delivery-related parameters, however, seemed easier to integrate into computerised routines. The problem of designing optimal beam numbers, beam directions, MLC-based beam intensity profiles (also called fluence maps) and apertures prominently lends itself to a mathematical optimisation approach. Especially the fluence map optimisation (FMO) problem and the beam aperture modulation problem have received much attention during the past years from both medical physicists as well as operations researchers. A comprehensive literature overview can be found in *e.g.* [13–15]. Figure 2.1 shows a schematic representation of a projected MLC-aperture onto the total resulting fluence map, contributing to the total dose distribution of a prostate cancer patient.
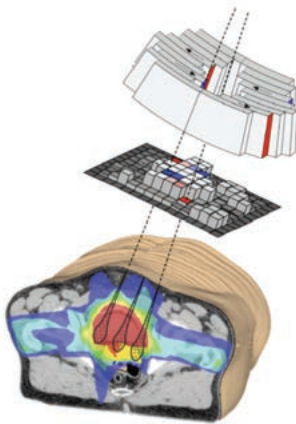
Figure 2.1: Schematic representation of a projected MLC-aperture onto the total resulting fluence map, contributing to the total dose distribution for a prostate cancer patient. (Image courtesy of RaySearch AB, Stockholm, Sweden.)

Forward and inverse treatment planning

In conventional conformal radiation therapy and 3D-CRT the treatment planning problem has been solved by a 'forward' approach, where all design parameters of the plan were manually set in the TPS by the planner. Subsequently, the dose distribution in the patient was calculated by the TPS. If the planner was not satisfied with the resulting dose distribution, the design param-

eters were manually altered and a dose re-computation was done. This approach is therefore essentially a trial-and-error procedure.

With the advent of IMRT, it became impossible to create dose distributions by forward treatment planning due to the increased dimension of the design space and the lack of guarantee that the solution obtained is truly optimal [16]. Therefore, an 'inverse treatment planning' approach has been explored, where tuneable machine-dependent treatment delivery parameters need to be found to meet pre-set requirements of the dose distribution using computerised optimisation models. This means that the solution is obtained from the specification of ideal properties of the dose distribution, rather than from a syntactical description of the design parameters and the application of some compiling recipe [4]. The major difference to the forward approach is that the inverse approach provides a more systematic solution procedure, and hence lends itself for a mathematical optimisation approach.

2.2   Mathematical perspective: formulation as multi-criteria optimisation problem

Unfortunately, the terms 'optimal' and 'optimisation' have different interpretations for the different stakeholders involved in the treatment planning process. Whereas clinicians may interpret optimisation relatively loosely as (iteratively) improving the treatment plan to obtain a certain acceptable trade-off between treatment benefit and injury, physicists and mathematicians use a more rigorous definition: striving for nothing less than the best possible solution within certain pre-set bounds. In the next subsections, the formal terminology of a mathematical optimisation problem is introduced and methods to find solutions are discussed.

Terminology

Following the terminology from [12], a more formal description of the inverse treatment planning optimisation problem can be given. This is essential for the remainder of this thesis. IMRT treatment planning typically involves the selection of certain delivery parameters (*e.g.* beam angles, field segments, monitor units). The basic components of the optimisation problem are the *design* or *decision variables* (*i.e.* delivery parameters), the *objective function* to be maximised (or minimised), and *constraints* on the decision choices. A *solution algorithm* is designed to find design variable values such that the objective function value is maximal (or minimal) and all constraints are satisfied.

In *multi-objective* optimisation, the objective function is a composite of multiple functions and can be either a vector-valued function or a univariate function representing the weighted sum of the constituent functions. Typically, some form of a penalty function is used to reflect the clinical desirability of the target dose distribution (*e.g.* minimum dose and dose heterogeneity), penalising undesirable doses. For normal tissues, either an objective function (*e.g.* maximum dose deposited in a certain volume) or constraint (*e.g.* the tolerance dose level) is included to reflect the avoidance of excessive doses. The set of all possible solutions that do

not violate the constraints is called the *feasible design space*. A solution is *optimal* if it is feasible and the objective function achieves the best possible value among the feasible solutions.

Finding the 'best compromise' solution

A formal mathematical representation of the problem is required for any algorithm to yield candidate treatment plans for evaluation. The definition of objectives and constraints to fulfil the clinical aims is however challenging, because the goals are often contradictory, mutually dependent, case specific and subjective. By way of example, consider target coverage and dose homogeneity *versus* normal-tissue sparing and dose conformality. As a result of this, the initial aims cannot be perfectly achieved and a trade-off is required. Unfortunately and inherently, the 'best compromise' solution is unknown in advance. Finding a solution therefore requires a *decision-making strategy* and some form of *user/algorithm interaction* to guide the search and selection process. To facilitate this, the planner is asked to articulate *preference information*, while the algorithm can use *structure information* about the mathematical properties (*e.g.* curvature, convexity, degeneracy) of the optimisation problem.

Scalarisation method

The classical approach is to use *weighting factors* that (should) capture the judgement about the relative importance or priority of the respective objectives. By combining these into a weighted average of the constituent objective functions, a single composite objective function arises. The composite function is then used as a single objective in the optimisation problem. Disadvantages of this approach are that the weighting factors usually have no clinical meaning, and are defined on arbitrary and incommensurable scales. Another drawback is that the sensitivity of the resulting solution to perturbations in the weights is unknown beforehand. Hence, the optimal weighting factors must be determined by an iterative trial-and-error process, which involves multiple optimisation runs and renders the treatment planning process to be inefficient and not intuitive.

Constrained optimisation methods

By realising that delivering the prescribed dose to the target volume is only limited by the organ-at-risk (OAR) constraints, another approach is to refine the constraints until the treatment plan is overall acceptable. As clinicians and treatment planners often have a hierarchical conception of the planning goals, *constrained optimisation* can be exploited to address the goals in a stepwise manner. In *prioritised goal programming*, highest-order goals are considered first. In subsequent steps, the achievements of the previous steps are turned into constraints, and a single new goal is incorporated as objective function. This approach has been applied successfully to IMRT inverse treatment planning optimisation by several investigators [17–19]. User interaction has been incorporated by quantifying the impact of small perturbations of the constraints on the resulting dose distribution, helping to reduce the human iteration loop [20, 21]. Several variants of this approach have been introduced, using for example, *pre-emptive goal*

programming [17], *lexicographical ordering* [18] and *successive relaxation* [19]. In the first two methods, the objectives are handled one by one and in a predefined order, whereas in the latter method the algorithm tries to find a solution fulfilling all constraints while subsequently relaxing limiting (lower-priority) constraints if a solution cannot be found. Since there is no tuning of weighting factors, it is an intuitive and straightforward approach. Nevertheless, multiple optimisation runs are required and only one final solution is presented. Hence, it hampers the possibility to explore nearby candidate solutions that can give the decision maker a sense of how small concessions can be exchanged for large gains in case of regret about the preceding assignment of priorities.

Pareto optimisation: an *a posteriori* planning approach

In the approaches discussed so far, *a priori* preference information had to be articulated, either by assigning weighting factors or priority levels. However, it is unclear in advance how dependent the objectives and constraints are. Without a thorough understanding of the degree in which the criteria are competing, it is difficult to define an intelligent basis for *a priori* quantification, not ranking, of the inherent trade-off between these criteria. Hence, the weighting factors or priority levels are often determined empirically and are based on advancing experience of the treatment planner. Furthermore, since no sensitivity information of the solution obtained is presented, the user remains unacquainted with the existence of a more appropriate operating point if small changes in the choice of values for weights or priority levels were effectuated.

Clinicians often experience difficulty in defining a complete, unique representation of the treatment planning optimisation problem, and typically differ in how they formulate it. They are, however, perfectly capable of ranking individually prepared treatment plans [22]. Taking this "*I know it when I see it*" concept as a starting-point to decouple the optimisation process from the decision-making process, several investigators explored the potential of multi-criteria optimisation (MCO) techniques using an *a posteriori* preference method to solve the FMO problem in IMRT [23–29].

Among the different methods, so-called *Pareto optimisation* is a versatile approach that is able to quantify the trade-off between the objective functions and allows the decision-maker to select an appropriate treatment plan from a pre-computed set of best-compromise solutions. It has been applied on many optimisation problems in economics and engineering involving multiple objectives (see *e.g.* [30]). Instead of aggregating all objective functions into a single composite objective function, the constituent objectives are considered simultaneously as an objective vector function. Since no single best solution exists for such vector optimisation problems, *optimality* has to be redefined. The Italian-Swiss socio-economist Vilfredo Pareto (1848–1923) introduced the notion that a solution is (strongly) *Pareto efficient* or *Pareto optimal* if no single objective can be further improved without deteriorating at least one other objective [31]. A more formal definition is given in textbooks (*e.g.* [32]). As a result, a set of best-compromise solutions exists, that represents the *Pareto efficient frontier* (PEF) when being evaluated in the space of objective functions.

Different strategies have been applied to generate the PEF for IMRT treatment planning problems. Initially, a brute-force approach was used to systematically generate feasible plans and subsequently test these for Pareto optimality [24]. The large computational burden of this approach has stimulated others to develop and apply more sophisticated PEF-generating methods keeping the computation times within clinically acceptable limits. Since the mathematical complexity depends on the number of objective functions, first the case with 2 objectives is considered. Work presented in this thesis (**Papers V, VI**) employs structural information of the FMO problem to provide a Sandwich-type algorithm for adaptive approximation of the PEF, which allows for efficient and dynamic generation of Pareto optimal IMRT plans.

Another aspect of *a posteriori* optimisation is the need for a technique to aid the decision-maker in the selection of a final solution from the Pareto set representation. This is often called 'navigation'. Different interfaces for navigation through Pareto optimal IMRT plans have been developed [33, 34]. This is beyond the scope of the work presented in this thesis.

2.3 Clinical perspective: balancing treatment benefit and risk trade-offs

Population-based dose prescription

In today's clinical practice, IMRT optimisation is routinely used as a 'meta-optimisation' procedure [35]. Treatment planners do not really expect the TPS to devise truly optimal treatment plans. Instead, they require the system to create plans that closely match the prescription and tolerance dose levels specified for the tumour and normal tissues, respectively. There is no incentive for the planner to do any better, as a link to quantifiable clinical outcome is lacking in the TPSs.

An underlying reason for this is that in contemporary radiotherapy consensus guidelines for treatment of specific patient categories are 'frozen' in local or (inter)national protocols, that specify for a given beam modality (*i.e.* photons or protons) and tumour type the precise radiation dose (within internationally established uniformity limits [36–38]), the fractionation regimen and the normal-tissue dose-volume limiting constraints that shall be used. Hence, a compromise is established balancing between the probability of tumour control that can be achieved and the risks of normal-tissue complications that are considered acceptable.

As discussed previously, IMRT has introduced a vast number of degrees of freedom to shape the dose distribution and has thereby increased the possibilities to make physician specific trade-offs between target coverage and sparing of normal tissues. Physicians may differently rate the various aspects involved in this trade-off and usually include other information in their decision-making when balancing the treatment risks and benefit. These include the patient's condition, age, social circumstances, type of complications expected, options for salvage treatment and patient preferences. The widespread implementation of IMRT will emphasise these differences and calls for decision-support tools to assist in balancing the treatment risk-benefit trade-offs to find and select a 'customised' plan that is considered optimal for the particular patient.

Dose escalation without adaptation for inter-patient diversity

Advances in the IMRT delivery technology, the availability of imaging data and its application in software systems for inverse treatment planning have been translated into improved normal tissue sparing, which consequently allows to escalate the dose to the tumour aiming for improved tumour control. Despite these developments, the way in which the tumour dose is prescribed has hardly changed: the (escalated) dose is still prescribed to the *same* level for *every* patient in the particular risk category. Such a 'one-size-fits all' approach will lead to relative underdosage in individuals with a favourable geometrical relationship between target volume and organs at risk, who are able to tolerate a higher tumour prescription dose. Possibilities to fully exploit dose escalation in these patients are left unemployed. This hampers the potential to 'customise' the dose prescription in relation to the anatomical diversity of patients.

Another aspect that is ignored in the current 'one-size-fits all' dose prescription approach is the diversity in outcome related risk-taking predilections between the physician and the patient, but also between patients. The physician's perception of what constitutes an acceptable risk level for (severe) side effects often prevails, with perhaps less emphasis on the associated level of tumour control. Such a conservative strategy is consistent with the Hippocratic "*dictum primum non nocere*" ("first, do no harm"), which outweighs "cure at all costs". In practice, this means the treatment plan is designed to maximise benefit while constraining the risks to predefined maximum acceptance levels. Other treatment plans with higher values of both benefit and risks always exist (*e.g.* by simultaneously scaling the monitor units of all beams), but they are rarely considered, and the patient does not have any significant input to the decision-making process.

The patient's perception of the treatment outcome shows a great diversity as well. Some patients may prefer a higher dose aiming for increased tumour control, while others may attach more importance to specific quality of life (QoL) related aspects. Results from a prospective trial conducted in patients with localised prostate carcinoma scheduled for 3D-CRT who were offered an *a priori* choice between two (fixed) dose prescription levels have confirmed this hypothesis: the majority (79%) of patients opted to be involved in the choice of their own treatment, and 75% of those patients chose the lower dose arm [39].

As patients differently rate the increase in treatment benefit from the increase in treatment risk, reflecting their willingness to accept higher toxicity in exchange for increased benefit, no assumptions with regard to their risk-taking preferences should be made prior to designing the treatment plan. Instead, they should be offered the possibility to *a posteriori* choose from a set of 'best compromise' solutions based on their specific anatomical (and possibly biological) characteristics.

The ultimate way to accomplish this is to exploit sufficiently reliable estimates of the tumour control probability (TCP) and the normal tissue complication probability (NTCP) based on the patient's individually optimised 3D dose distribution and fractionation regimen. This is challenging, because the treatment response measures also depend on the specific tumour characteristics (*e.g.* spatial distribution and density of clonogenic cells, differential radiosensitivity of cells during the cell cycle, acute and chronic hypoxia, cell proliferation) and the

normal-tissue properties (*e.g.* functionally disabled subregions and fractionation sensitivity), which are often uncertain or unknown for an individual patient. Moreover, the dose distributions that enter into the models to quantify treatment outcome in terms of TCP and NTCP mostly neglect that the planned and the delivered dose distribution differ due to treatment delivery uncertainties, *e.g.* inter- and intra-fractional anatomic variations and patient set-up inaccuracies. Nevertheless, it is considered important and useful to explore the potential of outcome-related metrics in treatment plan evaluation and optimisation in order to assess and rank the therapeutic performance of rival plans.

Patient empowerment: shared decision-making

In an era of individualised cancer therapy, where patient empowerment enters into the clinic, radiotherapy should move on towards *customised treatment planning* where individualisation is not restricted to adapting the spatial dose distribution to the patient's anatomy alone, but also involves *shared decision-making* between the radiation oncologist and the patient in an attempt to balance the trade-off in treatment benefit and injury probabilities with the individual preferences.

This requires that the current treatment planning systems should not only have the ability to compute TCP and NTCP estimates from 3D dose distributions, but additionally should also be equipped with functionality to assess the best achievable trade-off between these estimates for a variety of treatment plans with different tumour dose prescriptions, fractionation regimens and relative dose distributions. Software tools to intelligently generate these best achievable treatment plans in a systematic and methodologically correct way, and present these alternatives in an intuitive manner to the physician or the patient are currently lacking.

Ultimately, the TPSs should be interfaced with a multi-factorial decision-support system, where all available prognostic data (*i.e.* factors related to a measurable characteristic on the likely outcome of the cancer disease in an untreated individual) and predictive data (*i.e.* factors related to the identification of a characteristic on the likely benefit from treatment) from physical, clinical and biological origin is aggregated into prediction models to estimate treatment outcome with the highest accuracy, and allow to *truly* optimise the treatment plan for the individual patient [40]. Probably the most challenging part will be to acquire relevant information from the individual patient's biology (*i.e.* genotypic or phenotypic data).

2.4   Biological perspective: improving treatment response

Quantification of treatment plan quality by dose-volume metrics

Careful evaluation of the dose distributions in terms of isodose contour shape, target dose heterogeneity and normal tissue dose-volume (DV) factors is mandatory to verify against the protocol requirements and for documentation to correlate with clinical outcome. Taking full advantage of IMRT to shape the dose distribution critically relies on radiobiological information, including the specification of dose-volume constraints, to safely guide prescriptions and information about the usefulness of heterogeneous target dose distributions [41]. Although

this data is still sparse, there is a need to pursue this further, as extrapolating historic data obtained from 3D-CRT dose distributions to predict the outcome of IMRT dose distributions is speculative and inherently risky. To quote Paliwal *et al.* [42]: "*the science of optimising therapeutic gain lags significantly behind the capability to deliver IMRT dose distributions*". This is partly due to the fact that the quality of a treatment plan hitherto has been judged by correlating the above physical quantities to biological response rather than directly by TCP and NTCP estimates. However, it is widely recognised that DV parameters are merely surrogate measures of biological response, and preferable should be replaced by biological indices to more closely reflect the clinical goals of radiation therapy [43].

At present, treatment plan evaluation and optimisation mainly rely on DV-based criteria. Inherent limitations associated with DV-based inverse treatment planning have been reported in literature, comprising: 1) the need to specify multiple DV objectives or constraints per OAR to represent the different portions of the dose-volume histogram (DVH) curve that correlate with risk of complications, 2) the difficulty of selecting values and weights for these criteria to obtain optimal outcome estimates in terms of TCP and NTCP; 3) the increased computational complexity and non-convexity of the resulting optimisation problem, which implies that search algorithms may get trapped in local minima, potentially leading to less favourable dose distributions [44, 45]. The replacement of DV criteria by TCP and NTCP criteria may overcome these limitations.

Quantification of treatment response by TCP and NTCP models

Various models and parameter sets for estimating the treatment response in terms of TCP and NTCP have been proposed in the past two decades. These models either have a mechanistic biophysical foundation (*i.e.* the quantification of the effects of both unrepairable cell damage and repairable cell damage susceptible to misrepair after irradiation) or are phenomenological in nature (see *e.g.* [46]). For a summary of these models see *e.g.* [47–49]. Their application has mainly been restricted to the *evaluation* of 3D dose distributions from existing treatment plans. Currently, available dose-volume outcome data is analysed to determine the 'best fit' parameter values for these models, as for example the QUANTEC review data that has been used to update and refine normal-tissue dose-volume tolerance guidelines along with a description of the most commonly used NTCP models and their parameters [50].

Fitting the dose-response models to the clinical outcome data requires statistical learning techniques that produce parameter estimates with their associated confidence intervals. Different methods for the statistical validation of dose-response models have been published, especially with application to NTCP models (*e.g.* [51, 52]). Although the confidence intervals are usually fairly wide, uncertainty issues with current radiobiological models should not prevent them from being exploited for modern radiotherapy. One reason for this is that outcome models are continually improving as more data become available, and thus will not always be considered 'unreliable'. Another reason is that the only way to answer the question whether "*one could have produced another treatment plan which is better according to outcome measures*" is to actually use the radiobiological criteria instead of only relying on the dosimetric

surrogates to steer the inverse treatment planning optimisation process for IMRT.

Nahum [53] pointed out that the use of current TCP and NTCP models is hybrid in nature; model parameters are derived from *population-averaged* patient data, but are subsequently applied to the *individual* patient's dose distribution. Therefore, the models should be used judiciously, *i.e.* within the range of their applicability and with the apprehension that the parameter values have been derived under specific conditions (*e.g.* patient population, treatment technique, dose-time fractionation, etc.). Guidelines on the use and quality assurance of biologically related models for treatment planning have recently been reported by the Biological Effects Subcommittee Task Group #166 (TG-166) of the Therapy Physics Committee of the American Association of Physicists in Medicine (AAPM) [49, 54].

Desired statistical and mathematical properties (*e.g.* predictive power and convexity, respectively) of the radiobiological models may differ when being used for either plan evaluation or plan optimisation. Plan optimisation only requires a model with the ability to steer the optimisation process in the desired direction, while for plan evaluation the accuracy of the predictions is of paramount importance, especially when absolute values are used [55]. At this moment, absolute estimates of TCP and NTCP as main indicators of plan quality are not warranted [54]. Treatment plan evaluation should therefore be performed in conjunction with established DV criteria. Review of the 3D dose distribution should always be part of this process.

Radiobiological optimisation

The meaning of 'radiobiological optimisation' is generally defined as: the attempt to produce the 'best' treatment plan in terms of radiobiological (outcome) criteria. Miller [56] was the first to report on the use of radiobiological criteria to drive multi-field radiotherapy treatment planning. Almost one decade later, it was suggested to use radiobiological models as objective functions to be incorporated into the inverse treatment planning process for IMRT [57–59]. Since then, a number of papers have been published discussing this topic, see the reviews by *e.g.* [41, 48].

DV metrics are known to correlate with TCP or NTCP indices, and may hence be considered *effectively* radiobiological criteria. In spite of this, the direct use of TCP and NTCP models as objective/constraint functions may increase the radiobiological quality of treatment plans as these models more efficiently drive the shape of the DVH curves in the direction of improvements than (a set of) fixed DV constraints do. The TCP and NTCP objective/constraint functions could potentially be combined with DV constraints to prevent unacceptably large hot and cold spots in the target, or small hot spots in normal tissues.

According to Nahum [53], different levels of radiobiological optimisation can be distinguished, either based on the improvement of the therapeutic performance of an *existing* treatment plan through (*a posteriori*) customisation of the dose prescription or on utilising biological indices for the design optimisation of a *new* treatment plan, in approximate order of increasing sophistication:

. **No dose customisation**: the (change in) dose (*i.e.* either the prescribed tumour dose, D_{pres}, or the number of fractions, N_{pres}) is prescribed in the *same* way for *every* patient;

I. **Customisation of prescription dose under iso-toxic conditions**: individualisation of D_{pres} on an iso-toxic (*i.e.* equal NTCP level) basis;

II. **Customisation of prescription dose and fraction number under iso-toxic conditions**: individualidation of not only D_{pres} but also N_{pres} on an iso-toxic basis;

III. **Radiobiologically guided inverse planning**: radiobiological functions are used as objectives/constraints to drive the inverse treatment planning process; D_{pres} and N_{pres} are liberated, resulting in a target dose distribution with statistics that are entirely determined by the spatial interplay of these functions. Dosimetric requirements (*e.g.* minimum target dose, uniformity) can additionally be used to elevate the clinical acceptance;

IV. **Functional image guided inverse planning**: patient-specific information from biological and functional imaging is added to radiobiologically guided inverse planning for selective intra-tumoural dose (de-)escalation to facilitate a 'dose-painting' strategy;

V. **Full-blown biology guided inverse planning**: patient-specific biological information from *e.g.* genomics and proteomics is added to any of the previous levels.

A strategy for full clinical implementation of biologically-guided radiation therapy (BGRT) along the above steps was already presented by Stewart and Li [60] back in 2007. The work in this thesis (**Papers II, III, V, VII, X**) contributes to methodological aspects of the partial implementation of BGRT treatment planning related to Levels I to III.

Biological conformality

It is well known that the spatial distribution of biological parameters (*e.g.* spatial distribution and density of clonogenic cells, differential radiosensitivity of cells during the cell cycle, acute and chronic hypoxia, tumour cell proliferation) in most tumours and normal tissues is heterogeneous. Recent progress in functional and biological imaging has made it possible to map these distributions. The ability of IMRT to deliver deliberately non-uniform target dose distributions has stimulated radiotherapy researchers to transcend the limits of physical conformality and explore the concept of biological conformality to further improve the therapeutic efficacy of cancer radiation treatment [11, 60, 61].

Several approaches have been suggested to quantitatively incorporate the spatial biology data into IMRT inverse treatment planning. Voxel-based and sub-volume based prescriptions have both been pursued as two prototypical strategies for 'dose painting' (see *e.g.* [62, 63]). Subvolume boosting involves the selection and surdosage of an intra-tumoural target, and is related to image-based target volume selection and delineation. Examples of randomised multi-center phase II/III trials currently recruiting participants for integrated boosting of the primary tumour are: the FLAME study (NCT01168479) for advanced adenocarcinoma of the prostate, the PET-Boost study (NCT01024829) for inoperable/irresectable non-small-cell lung

carcinoma (NSCLC), and the ARTFORCE study (NCT01504815) for advanced squamous cell head-and-neck cancer. Dose painting by numbers is a voxel-level dose prescription based on a transformation that maps the image data intensities into prescribed dose. However, the functional form of this transformation is currently unknown.

For either of the two strategies, it is believed that TCP models are required that include either direct information on the aforementioned biological parameters or indirect information about the risk classification of the voxels or the subvolumes [64, 65]. This emphasises the significance to incorporate such models into the treatment planning process.

Dose escalation by increased target dose heterogeneity

Despite technological advances that facilitate the delivery of deliberately non-uniform target dose distributions, the way in which the tumour dose is prescribed in contemporary IMRT still complies with strict homogeneity constraints according to international guidelines that were originally developed for conventional radiotherapy and 3D-CRT. These guidelines state that the target volume should receive between 95% and 107% of the prescribed dose level, aiming for a uniform target dose distribution [36–38].

Escalation of the prescribed tumour dose has been shown to improve local control and progression-free or overall survival in different tumours, *e.g.* adenocarcinoma of the prostate [66] and advanced-stage NSCLC [67]. However, remaining faithful to the dogma of target dose homogeneity may hamper dose escalation, because the tolerated dose or NTCP level of a neighbouring radiation-sensitive OAR may be exceeded. Allowing a higher degree of target dose heterogeneity by loosening the maximum dose constraint seems an effective method to achieve dose escalation with IMRT. This was first confirmed for NSCLC patients in a retrospective treatment planning study by Schwarz *et al.* [68] comparing 3D-CRT and IMRT techniques, either by prescribing a homogeneous dose to the target volume or allowing enlarged dose heterogeneity. In this study, a re-optimisation approach using a fixed fraction size was followed to assess the highest achievable tumour dose (HATD). Radiobiological optimisation combined with dosimetric constraints (*i.e.* so-called 'physico-biological optimisation') can accomplish this in a more efficient way (**Paper X**).

Although exploiting the spatial biological heterogeneity to selectively steer the high-dose to relevant sub-volumes or voxels seems an obvious strategy, it remains to be proven whether the paradigm of biological conformality indeed yields improved clinical outcome. As a matter of fact, it cannot be excluded that increasing the integral dose to the target volume in a spatially random way rather than in a spatially selective way may have the same effect on the TCP. Although existing TCP models do not predict this to be true, it cannot be ruled out that this is due to their limitations. These models typically only incorporate direct effects of cellular response to radiation in their mathematical descriptions, and are merely based on the assumption that the probability of cell survival at a given point is a function solely of the dose deposited to that point [46]. However, there is an increasing amount of evidence that while direct damage to the DNA plays an important role, also a variety of indirect processes (*i.e.* alterations in the tumour microenvironment, intercellular communication, tumour-stroma interactions)

impacts on cellular responses to radiation [69]. These radiation-induced 'bystander' effects are believed to becoming increasingly significant when complex spatially highly modulated dose distributions are delivered through techniques such as IMRT [70], even though the dose gradients do not take place on a cellular scale. Models describing the *in vitro* response of cells to modulated field exposures including intercellular signalling effects have recently appeared in literature [71] and may possibly apply to *in vivo* tissue structures. As long as a refinement of existing radiobiological models to incorporate non-targeted effects and modulated dose distributions has not been effectuated and validated, existing models purely relying on local doses are considered the designated means to use for the comparison and the ranking of competing 3D dose distributions.

Apart from the spatial biological heterogeneity, also temporal effects play a role. It is known that tumour biology will change under the influence radiation therapy (*e.g.* acute hypoxia). Currently, it is not clear to what extent a single pre-treatment snapshot of the spatial distribution of some relevant tumour biological feature will be representative for the dynamic behaviour of the tumour microenvironment during treatment. This is considered to be a very difficult issue, and is beyond the scope of the work presented in this thesis.

From the different perspectives illuminated in this work before, it is worthwhile to investigate how moving away from the homogeneity requirements in 'traditional' IMRT plans will affect the potential therapeutic gain when the tumour dose for an individual patient is maximised to the highest physically achievable level while satisfying pre-set maximally tolerable normal-tissue dose or NTCP constraints. The work in this thesis (**Paper X**) describes the application of a physico-biological optimisation approach to efficiently assess this trade-off for individual patients with advanced-stage NSCLC by generating and comparing Pareto efficient frontiers in the TCP/NTCP space when the target dose uniformity is either restricted to 'traditional' constraints for both the minimum and the maximum dose or is unrestrained for the maximum dose only.

Improved normal tissue sparing: consequences of fractionation effects

Most radiation treatments are delivered in multiple fractions. The underlying biological rationale is that most late responding normal tissues have better repair capacities to recover between fractions than most tumours (see *e.g.* [72]). This is reflected in the difference between the tissue's fractionation sensitivities (expressed by the α/β value). Fractionation is meant to selectively spare normal tissues, allowing for deposition of a therapeutic dose in the tumour while not causing too much irreparable damage to the surrounding normal tissues that unavoidably get irradiated as well.

Modern IMRT planning and delivery techniques have the ability to produce highly conformal homogeneous target dose distributions with much lower doses to adjacent normal tissues than in conventional radiation therapy or 3D-CRT. When the stringent target dose uniformity limits are relaxed, and non-uniform target dose distributions are allowed, normal tissue sparing might even further be improved. Nevertheless, conventional multi-fraction regimens are still widely used, whereas there seems to be less need to fractionate if normal tissues are

receiving low doses. Reducing the number of fractions (while increasing the fraction size to achieve tumour iso-effect) would certainly not only offer an opportunity to further increase the 'therapeutic ratio' of radiotherapy, but also would reduce the social burden for patients, the logistical load for clinics and the economic costs of radiation treatment in general.

Hypofractionation regimens (*i.e.* schedules with few fractions and high dose per fraction) have shown to be safely deliverable in specific situations where: 1) there is evidence that the tumour's α/β-ratio is lower than for the dose-limiting normal tissue (*e.g.* prostate [73] and breast [74] carcinoma), and 2) a small target volume is treated with highly conformal irradiation techniques such as stereotactic body radiation therapy (SBRT) or stereotactic ablative body radiotherapy (SABR) and the dose-limiting normal tissue is exposed to low but nonzero dose, as *e.g.* in inoperable early-stage NSCLC [75–77]. These cases seem to contradict, because normal (healthy uninvolved) lung tissue usually has a lower α/β-ratio than the tumour and according to conventional radiobiological knowledge hypofractionation is not considered beneficial in terms of normal tissue sparing in this case [72]. However, the successful clinical results for early-stage NSCLC treated with SBRT/SABR have proven otherwise. Evidently, the volume effect plays a vital role. Although in the case of stereotactic lung irradiation the non-involved healthy lung tissue is seriously injured, the fact that the injury is limited to a small volume makes that there is very little toxicity.

Different bio-mathematical modelling studies have recently shown an increasing number of indications suggesting that for tumour sites in which the dose-limiting normal tissue receives a substantially lower dose than the tumour, hypofractionation is less problematic for the normal tissue in question than conventionally thought and could become radiobiologically preferable, even if the α/β-ratio of the tumour is larger than the α/β-ratio of the normal tissue [78–82]. Another aspect that these studies have revealed is that there is interdependence between the normal-tissue dose heterogeneity and the optimal fractionation regimen. This aspect is a topic of current research and has not been described in full extent for heterogeneous dose distributions and arbitrary volume effect.

Radiation oncology practitioners wanting to modify fractionation regimens on the basis of this hypothesis will safely do this under conditions that ensure normal-tissue iso-effect. Basically two methods are available to calculate new dose prescriptions with a different number of fractions and/or fraction size: either the Withers iso-effect formula (WIF) [83] or an NTCP model. The latter is considered the correct, but more complex way as it takes into account the whole dose distribution and the volume effect of the normal tissue. Although few commercially available TPSs have the ability to calculate NTCP scores from 3D dose distributions, their capability to use these models to calculate alternative normal-tissue iso-effective schemes with less fractions is limited [54]. Additional software exists that enables 'iso-toxic optimisation' as long as only the prescription dose is *rescaled* and the fraction number and/or fraction size are altered, but also requires the full 3D dose distribution to be available (in the form of a DVH) [84].

Instead, the more simple former method is often applied, which is based on the well-established linear-quadratic model of cell killing [72] and incorporates only the fractionation sensitivity parameter α/β and the dose in the normal tissue. However, it is unclear what nor-

mal tissue dose should be used when exploiting the WIF, especially when heterogeneous dose distributions are considered. A universally adopted practice is to use the prescribed tumour dose as a conservative estimate of the maximum normal tissue dose. Strictly speaking, WIF only applies to normal tissues receiving this maximum dose uniformly or when the normal-tissue response is solely determined by this dose. For all other situations the use of the WIF to derive a new regimen that is intended to be iso-effective will effectively overestimate the NTCP and thus result in highly conservative prescriptions when the number of fractions is reduced. Hence, the WIF seems too primitive and inappropriate for fractionation optimisation of highly conformal dose distributions delivered with modern radiation therapy techniques. The conclusion is that to determine the optimal fractionation regimen, the relationship between the degree of normal-tissue dose heterogeneity, the volume effect associated with the clinical endpoint studied and the fractionation sensitivity should be known. The work in this thesis (**Paper IV**) contributes to the unravelling of this relationship.

Main contributions and summary of appended papers

Compared to conventional radiotherapy and 3D-CRT, modern IMRT planning optimisation and delivery techniques have substantially improved the ability to spare normal tissue structures that are at risk when the target volume is irradiated with a prescribed dose that is meant to achieve a certain therapeutic benefit. However, the way in which the dose is prescribed to patients with a certain tumour belonging to the same risk category has not changed. Uniformity in tumour dose prescription between these patients does not imply uniformity in side effects. Current dose prescription protocols do not take into account the anatomical variety in tumour size and localisation relative to surrounding normal tissues amongst individual patients. This variety offers the opportunity to *customise* the dose prescription. Different strategies can be followed, of which maximally tolerable, iso-toxic and risk-adaptive approaches are part, depending on the decision-making roles of the physician and the patient. Current TPSs do not accommodate decision-support tools to investigate these approaches in a systematic manner. Advanced planning optimisation methods are required to supply this deficiency. The ten papers included in this thesis describe different aspects of new methodologies and their application for clinically relevant cases.

The first paper (**Paper I**) introduces the rationale and importance to breach the dogma of uniform dose prescription for patients belonging to the same treatment group, by including individual risk-taking predilections in the selection of an 'optimal' treatment plan from a range of pre-computed alternatives with acceptable TCP and NTCP estimated obtained from 3D dose distributions. The following three papers (**Papers II, III, IV**) deal with dose prescription customisation approaches to improve the therapeutic performance of an *existing* treatment plan by *re-normalisation*, *i.e.* without the need to redesign the relative dose distribution. The last six papers pertain to dose prescription customisation approaches where dose distributions are newly designed through intelligent *re-optimisation* methods, either by using only *dosimetric* criteria (**Papers V, VI, VII**) or by using *radiobiological* objectives (**Papers VIII, IX, X**). In the last paper, concepts developed in the preceding papers are combined to compare the

re-normalisation and re-optimisation approaches when individualised dose escalation under restricted and unrestricted target dose heterogeneity properties is applied to IMRT of NSCLC patients.

3.1 Inclusion of patient preferences into radiotherapy treatment planning

A first strategy to accomplish customised dose prescription is to take the risk-taking preferences of individual patients into account. In **Paper I**, it is proposed to include patient predilections into the treatment planning process to *individualise* the trade-off between TCP and NTCP based on the planned 3D dose distribution. IMRT facilitates to generate treatment plans with different prescription doses over a range of TCP and NTCP estimates. The acceptable range of treatment plans, from which a single 'best compromise' solution needs to be chosen, can be explored in various degrees of sophistication. Evaluating these solutions by exploiting the *therapeutic operating characteristic* (TOC) graph and the *Pareto efficient frontier* (PEF) seems very attractive for previously and newly designed treatment plans, respectively. From this exploration, the physician may choose treatment options for discussion of treatment preferences with the patient. The differences between these options can be assessed with estimates for TCP and NTCPs. This requires new tools for decision-support to be developed and integrated with existing radiation treatment planning systems.

3.2 Customisation of prescription dose using the TOC graph

In **Paper II**, the concept of the TOC graph is applied as a decision-aid to quantify the trade-off between TCP and NTCP estimates using an *a posteriori* treatment planning optimisation approach. For an existing treatment plan, the dose prescription is re-normalised by 'Level I' optimisation (see page 17), either by changing the fraction size while keeping the fraction number fixed or *vice versa*, to assess the range of TCP and NTCP estimates from the 3D dose distribution. In contrast to the criticised concept of uncomplicated tumour control probability, P_+, no assumptions with regard to *a priori* risk-taking preferences between TCP and NTCP are made to find the 'optimum' prescription dose. The TOC concept is illustrated by a clinical example of prostate cancer where the trade-off between 5-year biochemical no evidence of disease, late gastrointestinal morbidity and genitourinary morbidity is studied for a cohort of patients and for an individual patient.

The area under the TOC graph is proposed as a new index for the therapeutic power of a treatment technique or plan, independent of the prescribed dose level. A comparison between 3D-CRT and IMRT plans designed for the same patient anatomy shows the effect of increased therapeutic power that can be achieved by modern inverse treatment planning optimisation and delivery techniques. The results suggest that increasing the target dose conformity and heterogeneity broadens the therapeutic window and augments the therapeutic power of individualised radiotherapy.

3.3 Customisation of fraction dose and number in NSCLC radiotherapy

An alternative way to customise the dose prescription is via individualised dose escalation strategies. Here, the physician predefines the maximally tolerable dose or NTCP limits for relevant OARs. The prescription dose of an existing treatment plan is escalated for an individual patient until the first normal-tissue limit is met. Recently, iso-toxic clinical protocols have been put in practice for advanced-stage NSCLC radiotherapy with promising results. In some of these protocols, the fraction number is increased while keeping the fraction size constant, whereas in others the opposite is carried through. In most of these protocols only the tolerance doses of the uninvolved lung tissue and the spinal cord have been taken into account.

In **Paper III**, the dose prescription is individualised in a retrospective *in silico* study in 38 advanced-stage NSCLC patients, taking into account dose constraints not only for the uninvolved lung tissue and the spinal cord, but also for the oesophagus, the brachial plexus and the heart. 'Level II' optimisation (see page 17) is applied by adjusting both the fraction size and number for both a maximally tolerable and an iso-toxic dose escalation approach. It is shown that in 79% of the cases significant dose escalation with therapeutic gain is possible. The maximum oesophageal dose is found to be the dominant dose-limiting constraint in most patients.

3.4 Normal-tissue fractionation accounting for heterogeneous dose and volume effect: the effective α/β concept

There is renewed interest in hypofractionation since a growing amount of evidence has shown that it can be safely applied using highly conformal IMRT techniques in specific situations. The standard method to derive tumour dose prescriptions for new (hypofractionation) regimens under iso-effect conditions is to exploit the well-known linear-quadratic based Withers iso-effect formula (WIF), incorporating the intrinsic fractionation sensitivity measures α/β for the tumour and the most relevant dose-limiting normal tissue. The ratio of these measures is used to assess the potential for hypofractionation. Recently obtained insights into radiobiological modelling have shown that tumour localisations where the anatomy allows substantially lower dose to the dose-limiting normal tissue than to the tumour, may have the potential for hypofractionation even if the abovementioned ratio seems unfavourable.

In **Paper IV**, we show that the WIF yields conservative hypofractionation prescriptions with non-exact normal-tissue iso-effect estimates for tissues with a (quasi-)parallel architecture receiving a relatively low dose. We propose a *generalised* WIF for exact normal-tissue iso-effect calculations, which retains the tumour prescription doses, but replaces the conventional fractionation sensitivity measure α/β with an *effective* fractionation sensitivity measure, $(\alpha/\beta)_{\text{eff}}$, which takes into account the non-trivial interdependency between the normal-tissue dose heterogeneity and volume effect associated with the clinical endpoint in question. Situations where hypofractionation can be safely applied can be easily identified using the new $(\alpha/\beta)_{\text{eff}}$ concept. Examples are shown for DVH data of the uninvolved lung tissue in NSCLC patients treated with IMRT and SABR plans.

3.5 Efficient generation of Pareto optimal IMRT plans

Designing the fluence maps for IMRT can be formulated as a multi-objective inverse treatment planning optimisation problem. Among the different methods to solve such problems, Pareto optimisation is the only approach that quantifies the trade-off between inherently conflicting plan objectives and allows the decision maker to select an 'optimal' plan from a pre-computed set of best compromises. This set is represented by the PEF in the objective space and contains an infinitely number of elements. Generally, no closed form of the PEF exists. One way to obtain the PEF is by discrete approximation. This requires multiple optimisation problems to be solved. To confine the computational burden, only methods are considered that are guaranteed to solve the problem to global Pareto optimality and avoid the unnecessary generation of non-Pareto optimal plans.

In **Paper V**, a new algorithm for iterative discrete approximation of the PEF up to some predefined error is developed using an ε-constraint optimisation approach. The convexity of the optimisation problem is exploited to construct piecewise-linear upper and lower bounds to approximate the PEF from a small initial set of Pareto optimal treatment plans. A new 'Sandwich algorithm' is presented (see **Paper VI**) in which these bounds are used with three iterative strategies to determine the location of the next Pareto optimal solution such that the uncertainty in the estimated PEF is maximally reduced. It is shown that an intelligent initial solution for a new Pareto optimal plan can be obtained by fluence map interpolation of neighbouring Pareto optimal plans. The method is applied to a simplified clinical head-and-neck case with two objectives to map the trade-off between target dose heterogeneity and normal-tissue sparing. The results show that with all three strategies representative estimates of the PEF can be produced in an efficient and effective manner.

3.6 Efficient approximation of the convex Pareto efficient frontier

For the case of a nonlinear bi-objective convex optimisation problem, the PEF is a univariate convex function. Discrete approximation is an established method to reconstruct this function. So-called 'Sandwich algorithms' have been proposed for univariate approximation of convex functions. In these algorithms, piecewise-linear upper and lower bounds of the convex function are constructed, making use of derivative information. However, this type of information may be absent. Instead, only function evaluation information may be available. This is the case with ε-constraint optimisation.

In **Paper VI**, the proof is presented that the univariate PEF is a decreasing convex function. Piecewise-linear upper and lower bounds for the univariate convex PEF are derived that are only based on function evaluation information. Complete proofs for these bounds are given. New 'Sandwich algorithms' are proposed that iteratively add new Pareto optimal points in a systematic way until a desired accuracy of the PEF approximation is obtained. The new algorithms show linear convergence. If in addition derivative information is available, quadratic convergence can be shown under certain conditions. Furthermore, it is proven that lower bounds resulting from the combination of function evaluation and derivative information are

tighter than those derived from function evaluation information only. Numerical examples are given that illustrate the usefulness of the algorithms.

3.7 Practical approach to assess IMRT plan trade-offs in NSCLC radiotherapy

In existing TPSs for inverse planning of IMRT, the dose and dose-volume based criteria to steer the optimisation algorithm differ from those used for evaluation of the resulting 3D dose distribution. Clinically established DVH criteria are often used to evaluate the quality of a (given) 3D dose distribution. DVH-based objective functions that are used for optimisation typically have been replaced with quadratic penalties measuring the deviation from some prescribed dose level per voxel. The correlation between the objective values and the plan evaluation criteria is weak in DVH-based optimisation models. This may hamper the use of Pareto optimisation, because the PEF shows the (quantitative) trade-offs in the plan objective space instead of in the plan evaluation space. Furthermore, these DVH-based objective functions are non-convex. In practice, repeated re-optimisation is often required to achieve acceptable solutions that fulfil the clinically relevant plan requirements, which renders this approach inefficient.

To circumvent this problem, we describe in **Paper VII** a software tool that was interfaced with the standard TPS to automatically generate a range of deliverable IMRT plans by systematically varying the parameters of the optimisation functions and to facilitate the analysis of the trade-off between clinically relevant dosimetric parameters for target coverage and normal tissue sparing. The tool is applied to investigate this trade-off for 5 patients with advanced-stage NSCLC who are scheduled to receive IMRT.

Linear relationships between pair-wise clinically relevant dosimetric parameters are established and shown to be dependent on the particular case. This is used as a meta-model to support the decision-making of an individualised treatment plan.

3.8 Convex reformulation of radiobiological optimisation for IMRT planning

When radiobiological treatment plan evaluation criteria based on TCP and NTCP models are to be included as objective and constraint functions in 'Level III' optimisation (see page 17), suitable mathematical transformations have to be applied to make these functions strictly convex/concave, depending on the criterion to be minimised/maximised. Otherwise, the optimisation problem is hard to solve and it cannot be guaranteed that the unique, globally best solution is found.

In **Paper VIII**, we derive transformations for several well-known radiobiological models taking dose fractionation effects into account and establish conditions under which transformed functions result in equivalent convex criteria that do not change the set of Pareto optimal treatment plans. In addition, we show that applying increasing and concave transformations to convex objective functions is beneficial for the piecewise-linear approximation of the Pareto efficient frontier.

3.9   Effect of transformations on approximation of Pareto efficient frontier

As shown in the preceding paper, Pareto optimisation with radiobiological objective and constraint functions, requires mathematical transformations to make existing TCP and NTCP functions strictly convex/concave. We note that transformations of different *quality* exist. Certain transformations yield transformed objective functions that are less convex than other transformations. This is particularly useful for the approximation of the Pareto efficient frontier by piecewise-linear upper and lower bounds.

In **Paper IX**, we mathematically prove that if the objective function is already convex, an increasing and concave transformation that yields a less convex function results in tighter upper and lower bounds of the PEF that is to be approximated than the original functions. For transformed radiobiological criteria analysed before by others, we note that such increasing and concave transformations may exist.

3.10   Physico-biological optimisation of IMRT plans for individualised dose escalation in NSCLC radiotherapy

The 'best' achievable dose distribution for radiation treatment depends on the patient anatomy, the underlying biological heterogeneity, the physical limitations of the radiation technique in question, and the risk-taking predilections of the physician or the patient. Within the context of this multi-dimensional space a 'best compromise' solution needs to be found which is believed to yield the best achievable trade-off between treatment benefit and injury. It is desirable that the TPS can generate a set of 'best compromise' treatment plans, *i.e.* with the highest achievable TCP conditional to a chosen NTCP level, while avoiding the unnecessary effort in search for non-achievable plans.

Paper X describes a retrospective *in silico* dose escalation study where optimisation maximising TCP subject to both dosimetric and NTCP constraints is used to generate these solutions for patients with advanced-stage NSCLC. To assess the full potential of 'Level III' optimisation (see page 17) for IMRT, a comparison is made between homogeneous and heterogeneous target dose escalation. For the homogeneous case, 'Level III' optimisation is compared to 'Level I' optimisation by re-normalisation of the fraction size. The overall results suggest that substantial increases in local tumour control can be achieved over a range of tolerable risk levels for all patients in the study when the between-patient and within-target uniformity in dose prescription is no longer preserved.

General discussion and future perspectives

After a concise recapitulation of the previous sections, a reflection of some aspects is given in the light of future research efforts that are required to bring the concepts that have been developed in the framework of this thesis into clinical practice.

4.1 From population-based dose prescription by physician consensus to individualised dose prescription by patient preferences

In current radiation treatment protocols, patients with similar clinical characteristics, such as tumour localisation and stage, co-morbidity, performance status, age, and social circumstances receive the same dose prescription regardless of their willingness to accept the risk of particular side-effects in exchange for reducing the risk of recurrence or metastases. Clinical practice patterns have shown that treatment judgment is predominantly reserved to physicians. Although physician's decisions are expected to be congruent with best clinical evidence, this does not guarantee to yield the 'best' result for the individual patient. In their treatment strategy, physicians rely on *consensus* guidelines that strive for maximisation of TCP for a given, pre-selected level of NTCP. To achieve this goal, uniform dose prescriptions are employed for patients with tumours belonging to the same risk category. However, the variety in patient anatomy, biological heterogeneity and risk-taking *preferences* of patients is not taken into account in the current dose prescription paradigm. Although alternative dose prescriptions with different combinations of TCP and NTCPs exist, they are generally ignored. This has constituted an 'iatrocratic' regime, where the physician may unintentionally withhold the patient to undergo the 'optimal' treatment. What is considered 'optimal' is however difficult to ascertain, especially when the balancing of treatment benefit and risks involves subjective decision-making and when the decision criteria (*i.e.* TCP and NTCP estimates) are cursed with uncertainties.

The treatment planning studies based on bio-effect models presented in this thesis support the observation that inter-patient and intra-patient uniformity in dose prescription does not yield optimal therapeutic performance in current radiation therapy practices. We hypothesise that if sophisticated imaging, advanced treatment planning and delivery techniques, and the best available bio-mathematical treatment response models are judiciously used, the concept of a *population-based* dose prescription in today's treatment protocols can be replaced by *individualised* dose prescription. The methodologies described in this thesis can provide valuable tools for this purpose.

By empowering the patient to actively participate in the decision-making process, higher levels of adherence to and satisfaction with their treatment may be achieved, eventually resulting in a better short- and long-term quality of life. Individualised dose prescription by inclusion of patient risk-taking preferences is proposed as a new paradigm to improve the treatment outcome of modern radiation therapy. This would involve a new way of medical counselling, as physicians have difficulty predicting the treatment decisional preferences of their patients. Treatment decision discussions are challenging, not only for physicians to transfer unbiased medical information to patients, but also for patients to conceptualise the benefits and risks and to form a treatment decision. Decision-aids that are adequate and intuitive enough for this task are required for this. More research in this field is needed to establish evidence-based methods that focus on assessing the determinants of decision-making according to patients' individual needs (*e.g.* knowledge, values and support), influencing the quality of their decisions. In addition, indicators of decision quality measuring expectations, satisfaction, regret, and concordance between the patients' values and the chosen treatment should be investigated.

4.2   Dose prescription by patient preferences: achievements so far

Fundamental problems in individualised IMRT planning comprise of: 1) the assessment of the trade-off between mutually dependent goals and 2) the balancing of risks associated with exposure of one organ against another.

Multi-objective optimisation techniques using the Pareto optimality concept have the ability to *objectively* assess the trade-off and present a range of achievable 'best compromise' plans in an intuitive way via the PEF. The work presented in this thesis mainly considers bi-objective problems, comprising one objective related to the target volume and one related to a relevant OAR. As multiple organs and different morbidities may be involved in clinically realistic situations, the PEF constitutes a multi-dimensional surface. Recently, an efficient method has been developed to approximate the multi-dimensional PEF by a sequential algorithm [85] that permits parallelised computation [86, 87]. Advanced multi-objective optimisation methods based on these techniques have recently been implemented in next generation commercially available treatment planning systems. Effective and efficient techniques to 'navigate' over the multi-dimensional PEF need to be evaluated. This requires intuitive decision-support tools to be interfaced with the TPS, allowing a full evaluation of the 3D dose distribution to be part of the final decision-making process.

Balancing the treatment benefit and risk involves subjective decision-making, as different patients may have different levels of acceptance for injuries. Modelling the way in which physicians or patients handle the trade-off between different severities of treatment-related side-effects is difficult, if not impossible. Therefore, it is suggested to present a range of treatment plans with different risk-related metrics to the decision-makers. The methods presented in this thesis provide different ways to incorporate intuitive (radiobiologically relevant) risk metrics into the treatment planning process to optimise the treatment plans in terms of the dose prescription, thereby aiming to improve the clinical outcome for the individual patient. This is either accomplished by altering the prescribed dose per fraction and/or the number of fractions, or by releasing the target dose uniformity constraint. The range of 'best compromise' treatment plans can be assessed using the Pareto optimisation concept. The absolute values of the resulting risk-benefit predictions should be interpreted with care as long as model calibration against clinical endpoints is cumbersome. Nevertheless, it entails the best information available so far, and can probably provide reasonably reliable estimates of outcome *differences* between treatment plans being generated with the same dose delivery technique exploiting conventional fractionation regimens.

4.3   Integration of the new methodology in clinical practice

For 'Level I' and 'Level II' optimisation, implementation into clinical practice seems to be straightforward. Tolerance of normal tissue is the major dose-limiting factor for the administration of radiotherapy. Current knowledge of radiation side-effects has reached a level where predictions for conventional fractionation regimens are considered sufficiently reliable for certain relevant organs/tissues and endpoints (*e.g.* rectal bleeding, pneumonitis, xerostomia) such that protocols for tumour dose individualisation have already been developed and applied. One example is our single-institution IDEAL-VMAT Phase II trial (NCT01577212) for individualised hypofractionated radiotherapy in patients with NSCLC, which was initiated based on the findings reported in **Paper III**.

For 'Level III' optimisation, the clinical implementation may be more challenging since the traditional way of dose prescription (in terms of total dose, number of fractions, target dose heterogeneity) is liberated. Although this can be accomplished in a controlled way, for example by successively releasing the conventional target dose heterogeneity constraints, some level of confidence in the radiobiological models steering both the treatment plan design and the decision-making process is required. Some notes of caution are in order before exact predictions of treatment outcome can be trustworthy enough to serve quantitative decision-making.

Firstly, the methods and models employed do not account for uncertainties. Although the parameters of the radiobiological models may be cursed with large confidence intervals, this may not have major effect on the *solution* (*i.e.* the fluence map that defines the treatment plan), while the effect on the quantitative *evaluation* (*i.e.* the predicted TCP or NTCP value) of the treatment plan could be large. The rationale for this is that during maximisation of TCP, for example, the optimiser strives to escalate the target dose as much as possible given the constraints imposed. Preliminary results showed that variation of the model parameters

could barely improve the TCP value, suggesting the solution to be robust with respect to the model parameters. Clearly, more research is needed to generalise this finding.

Secondly, the radiobiological models employed are based on the linear-quadratic (LQ) expression, which links the cell survival fraction to the absorbed dose. At high doses per fraction, approximately above 6–10 Gy, the LQ model may overpredict cell killing, which will result in an *under*prediction of TCP and an *over*prediction of NTCP. More sophisticated models need to be developed and applied to become universally accepted. For large fraction sizes, cell killing may not be the dominant process mediating the irradiation response for both early and late effects. There is an increasing amount of evidence suggesting that vascular and immune reactions also substantially contribute to cell death in the high dose-per-fraction range. Bio-mathematical models should therefore be adapted to properly account for these *in vivo* effects. However, bio-mathematical modelling may never be able to completely describe the complex biological processes involved in the response of high-dose fractionated radiotherapy. Nevertheless, it facilitates the ability to design treatment plans with improved therapeutic power once sufficient clinical outcome data has been obtained.

Thirdly, the geometrical uncertainties during treatment delivery are not taken into account in the optimisation process. Treatment planning of heterogeneous target dose distributions requires that geometrical uncertainties, which are currently included in (population-based) treatment planning margins, are directly incorporated into the treatment planning optimisation to obtain robust treatment plans. Hence, more research into robust optimisation or probabilistic treatment planning is recommended.

Fourthly, the predictive value of the trade-off between TCP and NTCP(s) needs to be validated. Efforts are needed to improve the quality, reliability and robustness of the treatment risk prediction models. Therefore, quality control of the treatment planning and delivery process by an independent *in vivo* dosimetry system is required to enable the evaluation of the delivered dose distribution and the correlation with established clinical goals as quantified through the risk measures. Hence, development and implementation of 3D *in vivo* dosimetry in combination with dose-response modelling incorporating the latest advances in areas such as imaging and genetics is considered compulsory.

Fifthly, many of the contemporary prediction models only include dosimetric data and lack clinical, treatment, imaging and molecular factors. In order to be able to more accurately and precisely predict treatment outcomes, it is required that such data is acquired via standardised procedures. This allows for assessment of the robustness, reproducibility and clinical utility of the data and stimulates the development of validated multi-factorial decision-support systems that are considered important for the implementation of truly individualised radiation oncology [40].

4.4 Consequences for clinical trials

Prescribing doses to individual patients taking into account their risk-taking preferences will affect the way in which randomised clinical trials can be conducted, especially when the study is designed to compare a population-based dose prescription approach to an individualised

dose prescription approach. In the classical trial design, where treatment arms with different population-based dose prescriptions are compared, the primary endpoint (*e.g.* disease free survival, overall survival, QoL) is defined unequivocally. However, this is not the case in trials with a treatment arm involving individualised dose prescription, as certain patients in the study cohort will attach more weight to QoL related aspects than to tumour control, for example. In trials where individualisation has already been incorporated in the design (*e.g.* comparing individualised radiotherapy with or without chemotherapy) this may not be problematic. The challenge remains how to build a reasonable level of evidence when selection bias (by the patient) is explicitly added to the treatment selection process. This is clearly a field where more research is needed.

References

1. J. Ferlay, H.R. Shin, F. Bray, D. Forman, C. Mathers, and D.M. Parkin. GLOBOCAN 2008 v2.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. **10**, 2010.

2. L.J. Schreiner, C.P. Joshi, J. Darko, A. Kerr, G. Salomons, and S. Dhanesar. The role of Cobalt-60 in modern radiation therapy: Dose delivery and image guidance. *J Med Phys*, **34**:133–6, 2009.

3. P. Mayles, A. Nahum, and J.C. Rosenwald, editors. *Handbook of Radiotherapy Physics: Theory and Practice*. Taylor & Francis, London, 2007.

4. K.-H. Küfer, M. Monz, A. Scherrer, P. Süss, F. Alonso, A.S.A. Sultan, Th. Bortfeld, D. Craft, and Chr. Thieke. Multicriteria optimization in intensity modulated radiotherapy planning. Berichte des Fraunhofer ITWM **77**, Fraunhofer Institut für Techno- und Wirtschaftsmathematik, Kaiserslautern, 2005.

5. A. Brahme, J.E. Roos, and I. Lax. Solution of an integral equation encountered in rotation therapy. *Phys Med Biol*, **27**:1221–9, 1982.

6. S. Webb. The physical basis of IMRT and inverse planning. *Br J Radiol*, **76**:678–89, 2003.

7. S. Webb. Conformal and intensity-modulated radiotherapy. In P. Mayles, A. Nahum, and J.C. Rosenwald, editors, *Handbook of Radiotherapy Physics: Theory and Practice*, pages 943–85. Taylor & Francis, London, 2007.

8. C.X. Yu. Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy. *Phys Med Biol*, **40**:1435–49, 1995.

9. K. Otto. Volumetric modulated arc therapy: IMRT in a single gantry arc. *Med Phys*, **35**:310–7, 2008.

10. S. Webb and D. McQuaid. Some considerations concerning volume-modulated arc therapy: a stepping stone towards a general theory. *Phys Med Biol*, **54**:4345–60, 2009.

11. C.C. Ling, J. Humm, S. Larson, H. Amols, Z. Fuks, S. Leibel, and J.A. Koutcher. Towards multidimensional radiotherapy (MD-CRT): biological imaging and biological conformality. *Int J Radiat Oncol Biol Phys*, **47**:551–60, 2000.

12. M. Langer, E.K. Lee, J.O. Deasy, R.L. Rardin, and J.A. Deye. Operations research applied to radiotherapy, an NCI-NSF-sponsored workshop February 7–9, 2002. *Int J Radiat Oncol Biol Phys*, **57**:762–8, 2003.

13. D.M. Shepard, M.C. Ferris, G.H. Olivera, and T.R. Mackie. Optimizing the delivery of radiation therapy to cancer patients. *SIAM Rev*, **41**:721–44, 1999.

14. T. Bortfeld, K.-H. Küfer, M. Monz, A. Scherrer, C. Thieke, and H. Trinkaus. Intensity-Modulated Radiotherapy: a large scale multi-criteria programming problem. Berichte des Fraunhofer ITWM **43**, Fraunhofer Institut für Techno- und Wirtschaftsmathematik, Kaiserslautern, 2003.

15. R. Reemtsen and M. Alber. Continuous optimization of beamlet intensities for intensity modulated photon and proton radiotherapy. In P.M. Pardalos and H.E. Romeijn, editors, *Handbook of Optimization in Medicine*, volume 26 of *Springer Optimization and Its Applications*, pages 83–122. Springer, Berlin, 2009.

16. S. Webb. *The Physics of Conformal Radiotherapy: Advances in Technology*. Series in Medical Physics and Biomedical Engineering. Taylor & Francis, London, 1997.

17. J.J. Wilkens, J.R. Alaly, K. Zakarian, W.L. Thorstad, and J.O. Deasy. IMRT treatment planning based on prioritizing prescription goals. *Phys Med Biol*, **52**:1675–92, 2007.

18. K.-W. Jee, D.L. McShan, and B.A. Fraass. Lexicographic ordering: intuitive multicriteria optimization for IMRT. *Phys Med Biol*, **52**:1845–61, 2007.

19. S. Breedveld, P.R.M. Storchi, M. Keijzer, A.W. Heemink, and B.J.M. Heijmen. A novel approach to multi-criteria inverse planning for IMRT. *Phys Med Biol*, **52**:6339–53, 2007.

20. M. Alber, M. Birkner, and F. Nüsslin. Tools for the analysis of dose optimization: II. Sensitivity analysis. *Phys Med Biol*, **47**:N265–70, 2002.

21. B. Sobotta, M. Söhn, M. Pütz, and M. Alber. Tools for the analysis of dose optimization: III. Pointwise sensitivity and perturbation analysis. *Phys Med Biol*, **53**:6337–43, 2008.

22. H.I. Amols, M. Zaider, M.K. Hayes, and P.B. Schiff. Physician/patient-driven risk assignment in radiation oncology: reality or fancy? *Int J Radiat Oncol Biol Phys*, **38**:455–61, 1997.

23. K.-H. Küfer, H.W. Hamacher, and T. Bortfeld. A multicriteria optimization approach for inverse radiotherapy planning. In W. Schlegel and T. Bortfeld, editors, *The Use of Computers in Radiation Therapy*, pages 26–8. Springer, Heidelberg, 2000.

24. C. Thieke. *Multicriteria Optimization in Inverse Radiotherapy Planning*. PhD thesis, Ruperto-Carola University of Heidelberg, Heidelberg, 2003.

25. D. Craft, T. Halabi, and T. Bortfeld. Exploration of tradeoffs in intensity-modulated radiotherapy. *Phys Med Biol*, **50**:5857–68, 2005.

26. D. Craft, T. Halabi, H.A. Shih, and T. Bortfeld. An approach for practical multiobjective IMRT treatment planning. *Int J Radiat Oncol Biol Phys*, **69**:1600–7, 2007.

27. C. Thieke, K.-H. Küfer, M. Monz, A. Scherrer, F. Alonso, U. Oelfke, P.E. Huber, J. Debus, and T. Bortfeld. A new concept for interactive radiotherapy planning with multicriteria optimization: First clinical evaluation. *Radiother Oncol*, **85**:292–8, 2007.

28. T. Spalke, D. Craft, and T. Bortfeld. Analyzing the main trade-offs in multiobjective radiation therapy treatment planning databases. *Phys Med Biol*, **54**:3741–54, 2009.

29. D.L. Craft, T.S. Hong, H.A. Shih, and T.R. Bortfeld. Improved planning time and plan quality through multicriteria optimization for intensity-modulated radiotherapy. *Int J Radiat Oncol Biol Phys*, **82**:e83–90, 2012.

30. R.T. Marler and J.S. Arora. Survey of multi-objective optimization methods for engineering. *Struct Multidisc Optim*, **26**:369–95, 2004.

31. V. Pareto. *Manuale di Economica Politica*. Societa Editrice Libraria, Milano, 1906.

32. K.A. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, Boston, 1999.

33. M. Monz, K.-H. Küfer, T.R. Bortfeld, and C. Thieke. Pareto navigation—algorithmic foundation of interactive multi-criteria IMRT planning. *Phys Med Biol*, **53**:985–98, 2008.

34. A. Fredriksson and R. Bokrantz. Deliverable navigation for multicriteria intensity-modulated radiation therapy planning by combining shared and individual apertures. Technical Report TRITA-MAT-2013-OS4, KTH Royal Institute of Technology, Stockholm, 2013.

35. C.G. Orton, T.R. Bortfeld, A. Niemierko, and J. Unkelbach. The role of medical physicists and the AAPM in the development of treatment planning and optimization. *Med Phys*, **35**:4911–23, 2008.

36. ICRU. Prescribing, Recording, and Reporting Photon Beam Therapy (Report 50). Technical report, International Commission on Radiation Units & Measurements, 1993.

37. ICRU. Prescribing, Recording and Reporting Photon Beam Therapy (Report 62); Supplement to ICRU Report 50. Technical report, International Commission on Radiation Units & Measurements, 1999.

38. ICRU. Prescribing, Recording, and Reporting Proton-Beam Therapy (ICRU Report 78). Technical report, International Commission on Radiation Units & Measurements, 2007.

39. J.J. van Tol-Geerdink, P.F.M. Stalmeier, E.N.J.T. van Lin, E.C. Schimmel, H. Huizenga, W.A.J. van Daal, and J.W. Leer. Do patients with localized prostate cancer treatment really want more aggressive treatment? *J Clin Oncol*, **24**:4581–6, 2006.

40. Ph. Lambin, R.G.P.M. van Stiphout, M.H.W. Starmans, E. Rios-Velazquez, G. Nalbantov, H.J.W.L. Aerts, E. Roelofs, W. van Elmpt, P.C. Boutros, P. Granone, V. Valentini, A.C. Begg, D. De Ruysscher, and A. Dekker. Predicting outcomes in radiation oncology—multifactorial decision support systems. *Nat Rev Clin Oncol*, **10**:27–40, 2013.

41. J.O. Deasy and J.F. Fowler. Radiobiology of IMRT. In A.J. Mundt and J.C. Roeske, editors, *Intensity Modulated Radiation Therapy: A Clinical Perspective*, pages 53–74. BC Decker, Hamilton, 2005.

42. B.R. Paliwal. IMRT may be used to excess because of its higher reimbursement from medicare: For the proposition. *Med Phys*, **31**:1–2, 2004.

43. C.C. Ling, X.A. Li, and W.R. Hendee. Over the next decade the success of radiation treatment planning will be judged by the immediate biological response of tumor cells rather than by surrogate measures such as dose maximization and uniformity. *Med Phys*, **32**:2189–92, 2005.

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

45. Q. Wu and R. Mohan. Multiple local minima in IMRT optimization based on dose-volume criteria. *Med Phys*, **29**:1514–27, 2002.

46. E.J. Hall and A.J. Giaccia. *Radiobiology for the Radiologist*. Lippincott Williams & Wilkins, Philadelphia, 7th edition, 2012.

47. A. Nahum and G. Kutcher. Biological evaluation of treatment plans. In P. Mayles, A. Nahum, and J.C. Rosenwald, editors, *Handbook of Radiotherapy Physics: Theory and Practice*, pages 731–71. Taylor & Francis, London, 2007.

48. V.A. Semenenko and X.A. Li. Treatment planning using biologically based models. In X.A. Li, editor, *Adaptive Radiation Therapy*, Imaging in Medical Diagnosis and Therapy, pages 85–104. CRC Press, Taylor & Francis Group, Boca Raton, 2011.

49. X.A. Li, M. Alber, J.O. Deasy, A. Jackson, K.-W. Jee, L.B. Marks, M.K. Martel, C. Mayo, V. Moiseenko, A.E. Nahum, A. Niemierko, V.A. Semenenko, and E.D. Yorke. The use and QA of biologically related models for treatment planning. Report of AAPM Task Group 166 of the Therapy Physics Committee, American Association of Physicists in Medicine, 2012.

50. L.B. Marks, R.K. Ten Haken, and M.K. Martel. Guest editor's introduction to QUANTEC: a users guide. *Int J Radiat Oncol Biol Phys*, **76**:S1–2, 2010.

51. P. van Luijk, T.C. Delvigne, C. Schilstra, and J.M. Schippers. Estimation of parameters of dose-volume models and their confidence limits. *Phys Med Biol*, **48**:1863–84, 2003.

52. C.-J. Xu, A. van der Schaaf, A.A. Van't Veld, J.A. Langendijk, and C. Schilstra. Statistical validation of normal tissue complication probability models. *Int J Radiat Oncol Biol Phys*, **84**:e123–9, 2012.

53. A.E. Nahum and J. Uzan. (Radio)biological optimization of external-beam radiotherapy. *Comput Math Methods Med*, **2012**:329214, 2012.

54. X.A. Li, M. Alber, J.O. Deasy, A. Jackson, K.-W. Jee, L.B. Marks, M.K. Martel, C. Mayo, V. Moiseenko, A.E. Nahum, A. Niemierko, V.A. Semenenko, et al. The use and QA of biologically related models for treatment planning: Short report of the TG-166 of the Therapy Physics Committee of the AAPM. *Med Phys*, **39**:1386–1409, 2012.

55. B. Choi and J.O. Deasy. The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning. *Phys Med Biol*, **47**:3579–89, 2002.

56. D.W. Miller. Optimization of attenuator shapes for multiple field radiation treatment. In B.P. Paliwal, D.W. Herbert, and C.G. Orton, editors, *Optimization of Cancer Radiotherapy*, Proceedings of the Second Intl. Conf. on Dose, Time and Fractionation in Radiation Oncology, pages 493–7, Madison, 1984.

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

58. A. Brahme. Treatment optimization using physical and radiobiological objective functions. In A.R. Smith, editor, *Radiation Therapy Physics*, Springer Series in Computational Physics, pages 209–46. Springer, New York, 1995.

59. A. Brahme. Biologically based treatment planning. *Acta Oncol*, **38**:61–8, 1999.

60. R.D. Stewart and X.A. Li. BGRT: biologically guided radiation therapy—the future is fast approaching! *Med Phys*, **34**:3739–51, 2007.

61. Y. Yang and L. Xing. Towards biologically conformal radiation therapy (BCRT): selective IMRT dose escalation under the guidance of spatial biology distribution. *Med Phys*, **32**:1473–84, 2005.

62. S.M. Bentzen. Theragnostic imaging for radiation oncology: dose-painting by numbers. *Lancet Oncol*, **6**:112–7, 2005.

63. G. Meijer, J. Steenhuijsen, M. Bal, K. De Jaeger, D. Schuring, and J. Theuws. Dose painting by contours versus dose painting by numbers for stage II/III lung cancer: practical implications of using a broad or sharp brush. *Radiother Oncol*, **100**:396–401, 2011.

64. D. Thorwarth and M. Alber. Individualised radiotherapy on the basis of functional imaging with FMISO PET. *Z Med Phys*, **18**:43–50, 2008.

65. Y. Kim and W.A. Tomé. Dose-painting IMRT optimization using biological parameters. *Acta Oncol*, **49**:1374–84, 2010.

66. S.T.H. Peeters, W.D. Heemsbergen, P.C.M. Koper, W.L.J. van Putten, A. Slot, M.F.H. Dielwart, J.M.G. Bonfrer, L. Incrocci, and J.V. Lebesque. Dose-response in radiotherapy for localized prostate cancer: results of the Dutch multicenter randomized phase III trial comparing 68 Gy of radiotherapy with 78 Gy. *J Clin Oncol*, **24**:1990–6, 2006.

67. M. Partridge, M. Ramos, A. Sardaro, and M. Brada. Dose escalation for non-small cell lung cancer: analysis and modelling of published literature. *Radiother Oncol*, **99**:6–11, 2011.

68. M. Schwarz, M. Alber, J.V. Lebesque, B.J. Mijnheer, and E.M.F. Damen. Dose heterogeneity in the target volume and intensity-modulated radiotherapy to escalate the dose in the treatment of non–small-cell lung cancer. *Int J Radiat Oncol Biol Phys*, **62**:561–70, 2005.

69. C. Mothersill and C. Seymour. Changing paradigms in radiobiology. *Mutat Res*, **750**:85–95, 2012.

70. K.M. Prise and J.M. O'Sullivan. Radiation-induced bystander signalling in cancer therapy. *Nat Rev Cancer*, **9**: 351–60, 2009.

71. S.J. McMahon, K.T. Butterworth, C.K. McGarry, C. Trainor, J.M. O'Sullivan, A.R. Hounsell, and K.M. Prise. A computational model of cellular response to modulated radiation fields. *Int J Radiat Oncol Biol Phys*, **84**:250–6, 2012.

72. J.F. Fowler. The linear-quadratic formula and progress in fractionated radiotherapy. *Br J Radiol*, **62**:679–94, 1989.

73. R. Miralbell, S.A. Roberts, E. Zubizarreta, and J.H. Hendry. Dose-fractionation sensitivity of prostate cancer deduced from radiotherapy outcomes of 5,969 patients in seven international institutional datasets: $\alpha/\beta = 1.4$ (0.9–2.2) Gy. *Int J Radiat Oncol Biol Phys*, **82**:e17–24, 2012.

74. START Trialists' Group. The UK Standardisation of Breast Radiotherapy (START) Trial A of radiotherapy hypofractionation for treatment of early breast cancer: a randomised trial. *Lancet Oncol*, **9**:331–41, 2008.

75. C. Grau, M. Høyer, J. Lindegaard, and J. Overgaard. The emerging evidence for Stereotactic Body Radiotherapy. *Acta Oncol*, **45**:771–4, 2006.

76. R.D. Timmerman. An overview of hypofractionation and introduction to this issue of *Seminars in Radiation Oncology*. *Semin Radiat Oncol*, **18**:215–22, 2008.

77. G.R. Borst, M. Ishikawa, J. Nijkamp, M. Hauptmann, H. Shirato, R. Onimaru, M.M. van den Heuvel, J. Belderbos, J.V. Lebesque, and J.-J. Sonke. Radiation pneumonitis in patients treated for malignant pulmonary lesions with hypofractionated radiation therapy. *Radiother Oncol*, **91**:307–13, 2009.

78. A.E. Nahum and J.D. Chapman. The interaction between tissue radiobiological properties ("architecture"), dose distribution and optimal fraction size—profound implications for conformal and heavy-particle radiation therapy. *Int J Radiat Oncol Biol Phys*, **55**:446, 2003.

79. I.S. Vogelius, D.C. Westerly, G.M. Cannon, and S.M. Bentzen. Hypofractionation does not increase radiation pneumonitis risk with modern conformal radiation delivery techniques. *Acta Oncol*, **49**:1052–7, 2010.

80. J.-Y. Jin, F.-M. Kong, I.J. Chetty, S. Ajlouni, S. Ryu, R. Ten Haken, and B. Movsas. Impact of fraction size on lung radiation toxicity: hypofractionation may be beneficial in dose escalation of radiotherapy for lung cancers. *Int J Radiat Oncol Biol Phys*, **76**:782–8, 2010.

81. R.J. Myerson. Normal tissue dose conformality measures to guide radiotherapy fractionation decisions. *Med Phys*, **38**:1799–805, 2011.

82. J. Unkelbach, D. Craft, E. Salari, J. Ramakrishnan, and T. Bortfeld. The dependence of optimal fractionation schemes on the spatial dose distribution. *Phys Med Biol*, **58**:159–67, 2013.

83. H.R. Withers, H.D. Thames Jr, and L.J. Peters. A new isoeffect curve for change in dose per fraction. *Radiother Oncol*, **1**:187–91, 1983.

84. J. Uzan and A.E. Nahum. Radiobiologically guided optimisation of the prescription dose and fractionation scheme in radiotherapy using BioSuite. *Br J Radiol*, **85**:1279–86, 2012.

85. G. Rennen, E.R. van Dam, and D. den Hertog. Enhancement of sandwich algorithms for approximating higher-dimensional convex Pareto sets. *INFORMS J Comput*, **23**:493–517, 2011.

86. R. Bokrantz and A. Forsgren. An algorithm for approximating convex Pareto surfaces based on dual techniques. *INFORMS J Comput*, **25**:377–93, 2013.

87. R. Bokrantz. Distributed approximation of Pareto surfaces in multicriteria radiation therapy treatment planning. *Phys Med Biol*, **58**:3501–16, 2013.

Papers

Treatment plan individualisation by inclusion of patient preferences

Inclusion of individual patient preferences into the radiotherapy treatment planning process

H Huizenga*, AL Hoffmann, JHAM Kaanders, PFM Stalmeier, JJ van Tol-Geerdink
Radboud University Nijmegen Medical Centre, Nijmegen, NL

Treatment prescription and planning in radiotherapy include a trade-off between TCP and NTCPs. The question is, whether it is possible to *individualize* the trade-off between TCP and NTCP in the treatment planning process, based on the 3D dose distribution. The importance follows from a study in which patients with a primary localized prostate carcinoma were offered a choice between a 70 Gy and 74 Gy 3D-CRT treatment. Once these patients were informed about the existence of the trade-off via an independent decision aid, 80% wanted to choose and 75% of those patients chose the lower dose option. However, the actual TCP and NTCPs are expected to vary significantly dependent on the type of IMRT and IGRT methods used nowadays. Uniformity in dose prescription and margins does not imply uniformity in NTCPs. Irrespective of the equipment and treatment technique in a particular institution, a range of treatment solutions can be found with a range of estimates for TCP and NTCPs. The acceptable range of treatment plans, from which a choice needs to be made, can be explored in various degrees of sophistication, of which evaluating the Pareto frontier is the most attractive. From this exploration the physician may choose treatment options for discussion of treatment preferences with the patient. The *differences between these two plans* can be characterized with estimates for TCP and NTCPs. The estimated *differences* in TCP and NTCPs estimates between optional plans is expected to be relatively accurate, since the 3D dose distributions are relatively similar, in the sense that the same treatment equipment and beam geometries are involved. The proposal circumvents the uncertainty in treatment outcome for the large variety of treatment methods applied nowadays. New tools in treatment planning systems are required, e.g. for the generation of Pareto frontiers.

Introduction

The treatment prescription and treatment plan in radiotherapy is well known to be a trade-off between the probability to achieve loco-regional tumor control (TCP) and the probability to cause complications or severe side effects to normal tissues (NTCPs). Here, TCP and NTCP is short for clinical outcome measures as life expectancy and probabilities of specifically stated side effects, and does not refer to the use of a specific model. The chosen compromise is mostly hidden in local, national or international treatment protocols and is based on consensus about what is considered to be the best compromise for a specific patient group. An important question however is whether the trade-off between TCP and NTCP can be *individualized*, i.e. whether patient preferences can be included in the treatment planning process. E.g. Amols *et al* (1997) proposed to take doctors' or patients' predilections for TCP or NTCP into account by ranking of treatment plans based on a TCP and NTCP based figure of merit. In the present paper we summarize results obtained from a recent clinical decision-making trial that we conducted, and we suggest a method to implement patient preferences in treatment planning on an *individualized* basis.

Recent results from a clinical trial

From 2000-2005 patients with a primary localized prostate carcinoma to be treated with three-dimensional conformal radiotherapy (3D-CRT), were included in a study in which they were offered a choice between a 70 Gy and a 74 Gy treatment (van Tol *et al* 2006b, 2006c). Probabilities presented for cure (local control, overall and disease-free survival bNED, life expectancy) and side-effects (erectile dysfunction and severe late gastro-intestinal and genitourinary effects) were based on a systematic literature review of the clinical effects of 3D conformal radiotherapy (van Tol *et al* 2006a).

Once these patients were informed about the existence of the trade-off between TCP and NTCPs for their treatment, via an independent decision aid, 80% wanted to choose between the two treatment arms (van Tol *et al* 2006b), and 75% of those patients chose the lower dose option (Van Tol *et al* 2006c). It was the first study in which, to the knowledge of the authors, the patients' choice between two radiation treatments was actually implemented. From the study of Van Tol *et al* (2006bc) it is clear that most prostate cancer patients want to be involved in the decision-making and actually want to make a choice, once a decision aid is provided. Some prefer a higher dose with enhanced life-expectancy, while others attach more weight to specific quality of life aspects (e.g. gastro-intestinal toxicity).

Discussion

The choice between the 70 Gy and 74 Gy treatment in this study (van Tol *et al* 2006abc) can neither be generalized to other institutions nor be included in national or international protocols. The actual probability of tumor control and complications is generally expected to be dependent on the sophistication of the treatment, e.g. on
- the amount of imaging data used to define the target volume,
- possible means to separate target volume from organs at risk, e.g. by a rectal balloon in prostate cancer,
- the chosen margins,
- the treatment technique and equipment applied, e.g. various types of beam configuration and IMRT, or tomotherapy,
- the amount of image-guidance during treatment delivery, e.g. in prostate cancer the use of lasers and/or EPIDs and gold markers or cone-beam CT based adaptive strategies on the accelerator to verify the daily target volume position and shape.

The quest for (inter)national consensus on treatment protocols is also hampered by the fact that treatment techniques and processes start to vary much more than in the past. Uniformity in dose prescription and margins does not imply uniformity in probabilities of side-effects.

The type of IMRT and IGRT determines the width of the therapeutic window, and thus the possibility to allow a higher dose to the target volume and/or to achieve a lower dose to the organs at risk. As such, the actual choice offered by Van Tol *et al* cannot be implemented in its present form in institutions that have embarked on methods as IMRT, IGRT, dose escalation, hypofractionation and/or tomotherapy. The only way to take into account this variety, is to predict TCP and NTCP on the basis of patients' individual 3D dose distribution and fractionation scheme. This is quite a challenge, since the improvement in clinical outcome for this variety of treatments offered by various institutions is not yet known from clinical trials, and thus the uncertainty in the probabilities is significant.

In the design of the study of Van Tol *et al* it was clearly decided that only treatment options would be offered to the patients that fall into the range of adequate treatment options, according to local, national or international treatment protocols. The only difference between both arms was the possibility for a subjective trade-off between TCP and NTCP, but within the range of adequate treatment options. TCP and NTCPs presented to the patients were based on literature; the planned patients' 3D dose distribution was not used to calculate estimates for TCP or NTCPs. The trade-off was implemented as a choice between two dose level prescriptions, which was considered possible because the treatment technique in both arms was the same. As stated above, not the dose level itself, but all details of the treatment determine the clinical outcome. As such, it seems more appropriate to characterize choice options by estimates for TCP and NTCPs based on the planned 3D dose distribution of a patient.

Specific TCP and NTCP models with specific parameter sets (e.g. Webb and Nahum, 1993, Burman *et al* 1991, Ten Haken 2001) have often been criticized. Indeed, the simple models cannot describe the complex biological behavior of cells and tissues exposed to radiation. Moreover, the dose distributions that enter in these models mostly neglect that the delivered dose distribution to volumes might differ from the planned dose distribution, due to inter- and intrafractional movement and patient set-up uncertainty. Nevertheless, it is generally believed that these models are adequate to rank treatment plans, *if and only if* these models are used to interpolate between dose levels of clinical known end points, and refrain from extrapolation.

Proposal

Irrespective of the radiotherapy equipment and treatment technique available in a particular institution, almost always a range of treatment solutions can be found, some with a slightly higher and some with a slightly lower probability of cure, with an unavoidable related slightly higher or lower probability of adverse side effects. Most of the modern treatment planning systems can provide TCP- and NTCP-estimates for these treatment solutions, based on models and the 3D dose distribution, with the model calibrated against clinical endpoints.

The acceptable range of treatment plans, from which a choice needs to be made, can be explored in various degrees of sophistication:
   a) by scaling the dose distribution to a slightly lower or higher prescription (Lind *et al* 1999, Sanchez-Nieto *et al* 2001)
   b) by adaptation of the weights of objectives in dose-based inverse IMRT treatment planning (e.g. Craft *et al* 2005)
   c) by adaptation of weights in biological-based or physico-biological inverse IMRT treatment planning (Hoffmann *et al* 2003, 2005)
   d) by exploring the Pareto-efficient frontier in a few dimensions (Yu *et al* 1997). Craft *et al* (2006) and Hoffmann *et al* (2006) have shown that these frontiers can be generated dynamically and efficiently.

From this exploration the physician may choose two (or more) treatment options for discussion of treatment preferences with the patient. Options should fall into the range of adequate treatments, but differ in the "weights" attached to TCP and NTCPs. The *differences between optional plans* can be characterized with TCP and NTCP measures in terms of life expectancy and probability of severe side effects. The estimated *difference* in calculated TCP and NTCP between these plans is expected to be relatively accurate, since the 3D dose distributions are relatively similar, in the sense that the same treatment equipment and beam geometries are involved, and optional plans are within the range of adequate treatments according to medical protocols. Unfortunately, the absolute values for TCP and NTCPs will be less accurate. Anyhow, it is the best information available and also serves as the basis for evaluation by the physician.

The first (a) and fourth method (d) do not require a trade-off prior to the treatment planning process. Once the treatment planning process is finished, the physician might choose scaling (a) or pick a plan from the Pareto frontier (d). Possibilities (b) and (c) require an iterative treatment planning process, and are thus inefficient. Ultimately, we prefer method (d), because it directly provides insight into the effect of changes.

Conclusion

The inclusion of individual patient preferences in radiotherapy treatment planning is a way to act upon our finding that prostate cancer patients are - and other groups of radiotherapy patients might be - more willing to be involved in decision-making than has been thought in the past (Van Tol *et al* 2006b). The proposal to focus on *differences* between optional plans with more or less the same technique circumvents the uncertainty about treatment outcome for the large variety of actual treatment methods applied nowadays. For (d) new tools in radiotherapy treatment planning systems are required.

References

Amols HI, Zaider M, Hayes MK and Schiff PB 1997 Physician/patient-drive risk assignment in radiation oncology: reality or fancy? *Int. J. Radiat. Oncol.Biol. Phys.* **38** 455-61

Burman C, Kutcher GJ, 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

Craft D, Halabi T and Bortfeld T 2005 Exploration of trade-offs in IMRT *Phys. Med. Biol.* **50** 5857-68

Craft DL, Halabi TF, Shih HA and Bortfeld TR 2006 Approximating convex pareto surfaces in multiobjective radiotherapy planning *Med. Phys.* **33** 3399-407

Hoffmann A, Larsson M, Van Lin E, Löf J and Huizenga H 2003 Possible routes towards radiobiological-based treatment plan optimisation for clinical IMRT *Radiother. Oncol.* **68** (Suppl. 1) S102

Hoffmann A, Kaanders H and Huizenga H 2005 From a priori to a posteriori decision-making in the design and selection of an optimal IMRT treatment plan *Radioth. Oncol.* **76** (Suppl.2) S107

Hoffmann AL, Siem AYD, Den Hertog D, Kaanders JHAM and Huizenga H 2006 Derivative-free generation and interpolation of convex Pareto optimal IMRT plans *Phys. Med. Biol.* **51** 6349 -69

Lind BK, Mavroidis P, Hyodynmaa S and Kappa C 1999 Optimization of the dose level for a given treatment plan to maximize the complication-free tumor cure *Acta Oncol.* **38** 787-98

Sanchez-Nieto B, Nahum AE and Dearnaley DP 2001 Individualization of dose prescription based on normal-tissue dose-volume and radiosensitivity data *Int. J. Radiat. Oncol. Biol. Phys.* **49** 487-99

Ten Haken RK (ed.) 2001 Partial organ irradiation *Sem. Radiat. Oncol.* **11** 181-268

Van Tol-Geerdink JJ, Stalmeier PFM, Pasker-de Jong PCM, Huizenga H, Van Lin ENJT, Schimmel EC, Leer JW and Van Daal WAJ 2006a Systematic review of the effect of radiation dose on tumor control and morbidity in the treatment of prostate cancer by 3D-CRT *Int. J. Radiat. Oncol. Biol. Phys.* **64** 534-43

Van Tol-Geerdink JJ, Stalmeier PFM, Van Lin ENJT, Schimmel EC, Huizenga H, Van Daal WAJ and Leer JW 2006b Do prostate cancer patients want to choose their own radiation treatment? *Int. J. Radiat. Oncol. Biol. Phys.* **66** 1105-11

Van Tol-Geerdink JJ, Stalmeier PFM, van Lin ENJT, Schimmel EC, Huizenga H, Van Daal WAJ and Leer JW 2006c Do patients with prostate cancer really want more aggressive radiation treatment? *J. Clin. Oncol.* **24** 4581-86

Webb S and Nahum AE, 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

Yu Y., 1997, Multi-objective decision theory for computational optimization in radiation therapy. *Med. Phys.* **24** 1455-54

II

Dose prescription customisation using the TOC graph

RADIATION ONCOLOGY

RESEARCH   **Open Access**

Employing the therapeutic operating characteristic (TOC) graph for individualised dose prescription

Aswin L Hoffmann[1,2*], Henk Huizenga[1] and Johannes HAM Kaanders[1]

Abstract

Background: In current practice, patients scheduled for radiotherapy are treated according to 'rigid' protocols with predefined dose prescriptions that do not consider risk-taking preferences of individuals. The therapeutic operating characteristic (TOC) graph is applied as a decision-aid to assess the trade-off between treatment benefit and morbidity to facilitate dose prescription customisation.

Methods: Historical dose-response data from prostate cancer patient cohorts treated with 3D-conformal radiotherapy is used to construct TOC graphs. Next, intensity-modulated (IMRT) plans are generated by optimisation based on dosimetric criteria and dose-response relationships. TOC graphs are constructed for dose-scaling of the optimised IMRT plan and individualised dose prescription. The area under the TOC curve (AUC) is estimated to measure the therapeutic power of these plans.

Results: On a continuous scale, the TOC graph directly visualises treatment benefit and morbidity risk of physicians' or patients' choices for dose (de-)escalation. The trade-off between these probabilities facilitates the selection of an individualised dose prescription. TOC graphs show broader therapeutic window and higher AUCs with increasing target dose heterogeneity.

Conclusions: The TOC graph gives patients and physicians access to a decision-aid and read-out of the trade-off between treatment benefit and morbidity risks for individualised dose prescription customisation over a continuous range of dose levels.

Keywords: Radiotherapy, Treatment planning, Individualisation, Dose-response relations, Decision-making

Background

The main task in radiation dose prescription and treatment planning is to maximise the tumour control probability (TCP) while maintaining an acceptable normal tissue complication probability (NTCP). Currently, this compromise is 'frozen' in treatment protocols, which are based on consensus opinions about what is considered to be the best trade-off for a specific patient population. However, in an era where patient empowerment enters into clinical practice, subjective criteria reflecting the physician's or individual patient's risk-taking preferences

should inherently be taken into account to establish a 'customised' treatment. Therefore, individualised treatment prescription and planning requires decision-making based on TCP and NTCP scores rather than on dosimetric criteria alone.

Amols *et al.* proposed an *a posteriori* decision-making approach to rank existing treatment plans having different combinations of TCP and NTCP based on a single figure of merit quantifying the physician's preferences [1]. Another approach is to extract patient preferences prior to treatment. In a recent prospective trial involving patients with localised prostate carcinoma scheduled for three-dimensional conformal radiotherapy (3D-CRT), patients were offered an *a priori* choice between treatments with two alternative dose levels resulting in different probabilities for tumour control and side-effects [2,3]. From this study, it

* Correspondence: aswin.hoffmann@maastro.nl
[1]Department of Radiation Oncology, Radboud University Nijmegen Medical Center, P.O. Box 9101, Nijmegen 6500 HB, The Netherlands
[2]Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, P.O. Box 1588, Maastricht 6201 BN, The Netherlands

became clear that the majority (79%) of patients opted to be involved in the choice of their treatment once a decision-aid was provided, and most patients attached a higher weight to quality-of-life related aspects than to tumour control [4].

The availability of modern treatment techniques provides the ability to design a variety of treatment plans with divergent trade-offs between benefit and morbidity. The ultimate way to take this variety into account is to predict TCP and NTCP scores for the individual patient based on the 3D dose distribution and fractionation scheme applied. This is challenging, since the clinical outcome measures for these treatments are not well known. Furthermore, specific tumour characteristics determining radiation response (*e.g.*, intrinsic radiosensitivity and hypoxia) are often unknown for the individual patient. Nevertheless, various models and parameter sets to estimate TCP and NTCP from a 3D dose distribution have been developed and applied (*e.g.*, [5-8]). It is generally believed that these models are adequate to rank rival treatment plans provided they are used over the dose range from which they have been derived. Some of these models have been incorporated into modern treatment planning systems, enabling a 'radiobiological' evaluation of treatment plans. However, the TCP/NTCP trade-off is difficult to assess in current treatment planning systems, even for a simple 'dose scalarisation' approach where only the prescription dose of a given treatment plan is changed for either a fixed number of fractions or for a fixed fraction dose. The concept of maximizing the probability of uncomplicated tumour control, P_+, as a function of dose has been proposed to find the single optimum dose level for this approach [9,10]. The criticism against this measure is that an *a priori* 'rigid' trade-off between TCP and NTCP is assumed without knowing their interrelationship over the range of potential dose prescriptions. As such, it is not suitable as a single measure for dose prescription customisation.

The aim of the present paper is to apply the concept of the therapeutic operating characteristic (TOC) graph to assess the trade-off in the TCP/NTCP domain over a continuous range of prescribed dose levels for given treatment plans. The TOC graph is presented as an interactive tool for dose prescription customisation of a treatment plan for an individual patient. We compare the P_+ and TOC graph and discuss their value for individualised dose prescription optimisation. The concept is illustrated by a clinical example of prostate cancer where the trade-off between 5-year biochemical no evidence of disease (bNED$_5$), late gastrointestinal (GI) and genitourinary (GU) morbidity is studied.

Methods

Therapeutic operating characteristic (TOC) graph

The TOC is a parametric plot of TCP *vs.* NTCP with the prescribed dose as a continuous independent parameter

[11-15]. As TCP and NTCP increase with total dose, their interrelationship presents an ascending curve in the benefit-injury decision space. The TOC graph can be used to estimate the optimal level of therapeutic effect and the associated dose level, based on individual risk-taking preferences.

So far, few quantitative measures of therapeutic window or therapeutic power have been published. Both refer to a quality index for radiotherapy to achieve loco-regional tumour control and to prevent severe late side effects. We propose to use the area under the curve (AUC) of the TOC graph as an index of the therapeutic power of a treatment technique or plan, independent of consensus on the prescribed dose level. This is by analogy with the AUC of the ROC graph used in diagnostic radiology [16].

TOC graph for a patient population: modelling results from clinical studies

The TOC graph was first applied to clinical outcome data from a systematic literature review on the effects of radiation dose on tumour control and morbidity in the treatment of prostate cancer [3]. The trade-off between TCP (*i.e.*, bNED$_5$), NTCP$_{GI}$ and NTCP$_{GU}$ (late GI and GU morbidity Grade ≥ 2 RTOG) was assessed for three-dimensional conformal radiotherapy (3D-CRT) techniques of combined data from dose escalation studies. The models and parameters used to describe the dose-effect relationship for the patient population studied are summarised in Additional file 1.

TOC graph for an individual patient: technique assessment and dose prescription customisation

TCP and NTCP models derived from the literature were applied to a 3D dose distribution of a given initial treatment plan to construct TOC graphs by *a posteriori* variation of the total prescription dose. This was either accomplished by variation of the number of fractions (at constant dose-per-fraction assuming no tumour cell repopulation correction for overall treatment time) or by variation of the fraction dose (while keeping the number of fractions constant). TOC graphs were generated after treatment planning to allow for TCP/NTCP balancing and selection of the preferred prescription for therapy delivery. Taking the individual's risk-taking preferences into account, a treatment plan with a customised dose prescription can be selected as a point on the TOC graph. We illustrate the concept of dose-level scaling for a typical prostate cancer patient using different treatment delivery and plan optimisation techniques in a step-wise approach. Firstly, TCP/NTCP evaluation was done for a forward planned 3D-CRT dose distribution that had initially been designed for standard fractionation. Secondly, a TOC graph was generated for an intensity-modulated radiotherapy (IMRT) plan obtained by inverse

planning using dose and dose-volume (*i.e.*, physical) objectives with equivalent initial prescribed dose (IMRT$_{phys}$), as it was expected that the NTCP of the IMRT$_{phys}$ plan was lower under tumour iso-effective conditions than for the 3D-CRT plan. Then a third plan (IMRT$_{biol}$) was obtained from the IMRT$_{phys}$ plan by physico-biological optimisation, improving the TCP under isotoxic conditions. The hypothetical benefit of scaling the IMRT$_{biol}$ was also assessed. TOC graphs and AUCs of the three plans were compared.

Organ segmentation and standard treatment plans

Both the 3D-CRT and IMRT$_{phys}$ plans had the same dose prescription recipe: 78 Gy in 39 fractions satisfying the 95% and 107% under- and overdosage criteria according to ICRU 50 criteria. The planning target volume (PTV) encompassed the prostate gland and base of the seminal vesicles plus a 5 mm isotropic margin. The rectum and bladder were delineated as organs at risk (OARs). The 3D-CRT plan comprised a wedged 10 MV co-planar 4-beam arrangement. The IMRT$_{phys}$ plan encompassed a 10 MV co-planar 5-beam geometry and was generated for a maximum of 60 step-and-shoot segments by inverse treatment planning (Pinnacle[3] version 7.6c; Philips Radiation Oncology Systems, Fitchburg, USA) with direct machine parameter optimisation (DMPO; RaySearch Laboratories AB, Stockholm, Sweden).

Physico-biological treatment plan optimisation

For isotoxic optimisation, values for NTCP$_{GI}$ and NTCP$_{GU}$ were calculated from the IMRT$_{phys}$ plan, and constituted upper limits for the IMRT$_{biol}$ plan at fixed fraction number ($N = 39$). The physical objectives from the IMRT$_{phys}$ plan were converted into constraints, resulting in the following physico-biological optimisation problem:

$$\text{maximise} \quad \text{TCP}\left(\underline{D}\right)$$
$$\text{subject to} \quad \text{NTCP}_{GI}\left(\underline{D}\right) \leq ntcp_{GI}$$
$$\text{NTCP}_{GU}\left(\underline{D}\right) \leq ntcp_{GU}$$
$$\text{DVH}_i\left(D, V_j\right) \leq dvh_{i,j}$$

where $\underline{D}$ is the dose distribution to be optimised, and $ntcp_{GI}$, $ntcp_{GU}$, $dvh_{i,j}$ are constraint values for organ i and dose-volume constraint j as obtained from the IMRT$_{phys}$ plan after it had been generated by inverse planning. Non-clinical research software (ORBIT Workstation, version 1.5; RaySearch Laboratories AB, Stockholm, Sweden) was used to solve this problem with DMPO [17]. The dose-response models used are summarised in Additional file 2.

For TOC analysis, treatment plans were retrieved from Pinnacle[3] and ORBIT Workstation into an in-house developed software tool (MATLAB version 7.6.0; The MathWorks Inc., Natick, USA). For the tumour a conservative $(\alpha/\beta)_T = 2$ Gy adopted from [18] was applied, while for both OAR endpoints generally accepted values of $(\alpha/\beta)_{OAR} = 3$ Gy and 6 Gy were adopted for Grade ≥2 late GI [19] and late GU toxicity [20], respectively. These (α/β) ratios were used to account for voxel-based fractionation correction prior to calculating TCP and NTCP scores.

Results
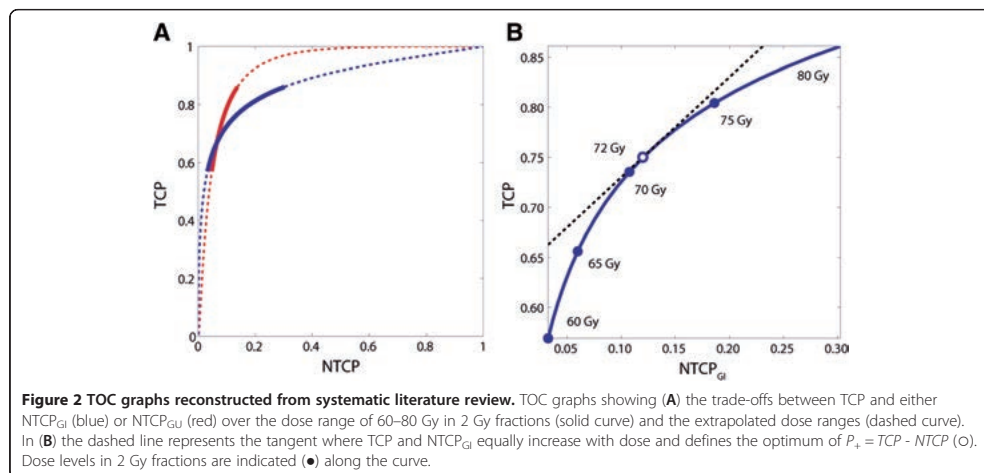
TOC graph for patient population

Figure 1 illustrates the population-averaged dose-response graphs for TCP and NTCP as a function of the prescribed dose level in the 2 Gy equivalent dose (EQD$_2$) range of 60–80 Gy, obtained from a systematic literature review [3].

In Figure 2A, two TOC graphs depict the trade-offs between the TCP, NTCP$_{GI}$, and NTCP$_{GU}$ of Figure 1 when plotted against each other. Figure 2B illustrates the TOC graph zoomed in on the EQD$_2$ range of 60–80 Gy for GI morbidity. In this graph, the dose level for which the increase in TCP and NTCP with dose is equal, and hence $P_+ = TCP - NTCP$ achieves its maximum value, is 72 Gy with associated TCP = 75% and NTCP = 12%. Below this level, the gain in TCP per unit dose is larger than the increase in NTCP, whereas the converse is true beyond this level.

In Figure 3, the interdependence between TCP, NTCP$_{GI}$ and NTCP$_{GU}$ as a function of the prescribed dose is shown in a 3D TOC graph together with its 2D projections on the
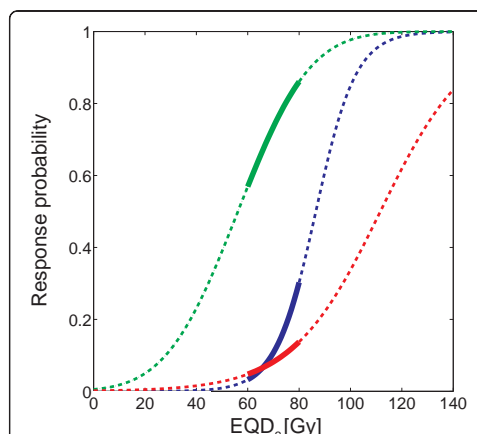
Figure 1 Dose response graphs obtained from systematic literature review. TCP (green), NTCP$_{GI}$ (blue) and NTCP$_{GU}$ (red) as a function of the prescribed total dose in 2 Gy fractions, obtained from a systematic literature review over the dose range of 60–80 Gy for 3D-CRT [3]. Dashed curves represent the relationship over extrapolated dose ranges.
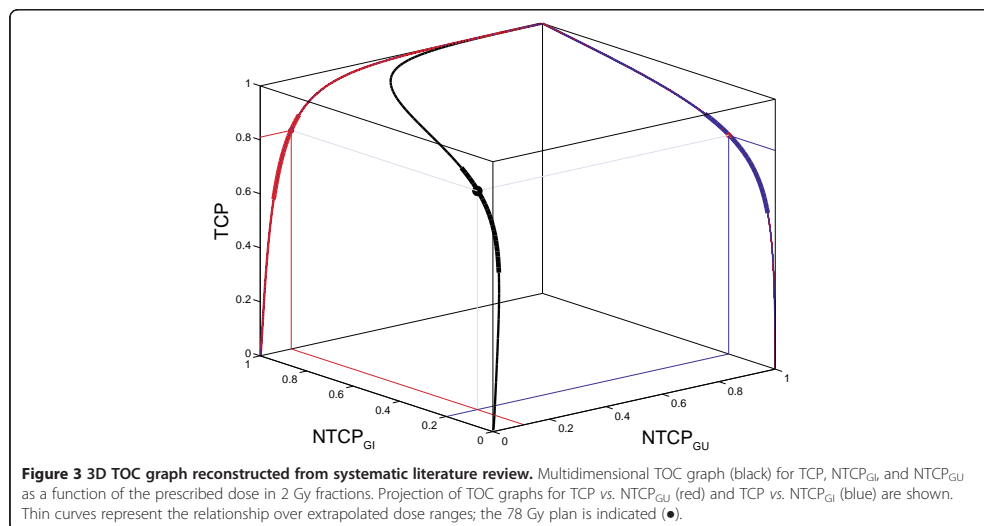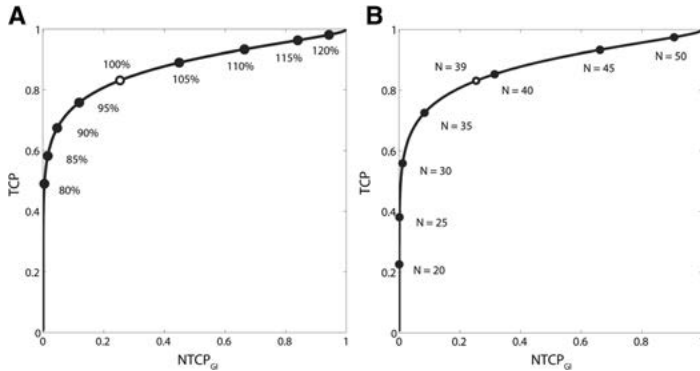
Figure 2 TOC graphs reconstructed from systematic literature review. TOC graphs showing (**A**) the trade-offs between TCP and either NTCP$_{GI}$ (blue) or NTCP$_{GU}$ (red) over the dose range of 60–80 Gy in 2 Gy fractions (solid curve) and the extrapolated dose ranges (dashed curve). In (**B**) the dashed line represents the tangent where TCP and NTCP$_{GI}$ equally increase with dose and defines the optimum of $P_+ = TCP - NTCP$ (O). Dose levels in 2 Gy fractions are indicated (●) along the curve.

TCP/NTCP$_{GI}$ and TCP/NTCP$_{GU}$ space that is shown in Figure 2A.

TOC graphs for individual patient
In Figures 4A and 4B, TOC graphs are shown for the initial (*i.e.*, 39 × 2 Gy) 3D-CRT treatment plan (TCP = 83%, NTCP = 25%) as a function of the scaled dose per fraction and the number of fractions, respectively. Dose statistics for renormalised plans are given in Table 1. It is evident that both curves coincide at the nominal dose prescription

of 2 Gy per fraction or $N = 39$ fractions. Their relative position in the TCP/NTCP space is insensitive to the model parameters (TD$_{50}$, γ_{37}, m, a), and only depends on the ratio of $(\alpha/\beta)_T$ to $(\alpha/\beta)_{OAR}$. Since $(\alpha/\beta)_T < (\alpha/\beta)_{OAR}$, escalating the total dose beyond the original 78 Gy level by adding fractions will increase the TCP/NTCP ratio less than by increasing the fraction dose. This becomes apparent from Table 1 when comparing the TCP and NTCP scores for the plans with a 5% nominally higher prescription dose of 82 Gy, obtained either from scaling the fraction size to 2.1

Figure 3 3D TOC graph reconstructed from systematic literature review. Multidimensional TOC graph (black) for TCP, NTCP$_{GI}$, and NTCP$_{GU}$ as a function of the prescribed dose in 2 Gy fractions. Projection of TOC graphs for TCP *vs.* NTCP$_{GU}$ (red) and TCP *vs.* NTCP$_{GI}$ (blue) are shown. Thin curves represent the relationship over extrapolated dose ranges; the 78 Gy plan is indicated (●).

Figure 4 TOC graphs for fraction size and fraction number variation. TOC graphs for prostate 3D-CRT plan showing trade-off between TCP and $NTCP_{GI}$ as a function of (**A**) the dose-per-fraction scale factor (at constant fraction number) and (**B**) the number of fractions N (at constant fraction dose). The initial plan (39×2 Gy) is marked (O) at the 100% dose level and at $N = 39$, respectively.

Gy (with 39 fractions) or from increasing the number of fractions to 41 (with 2 Gy per fraction). The opposite holds in case of dose de-escalation below 78 Gy and for $(\alpha/\beta)_T > (\alpha/\beta)_{OAR}$. When plotted in the same graph, both TOC curves would coincide only if $(\alpha/\beta)_T = (\alpha/\beta)_{OAR}$.

In Figure 5, TOC graphs for the 3D-CRT, $IMRT_{phys}$, and $IMRT_{biol}$ plans for the same patient anatomy are shown. Dose statistics for the initial plans are given in Table 2. By comparing the TOC graphs for the 3D-CRT and the IMRT plans, it is obvious that the $IMRT_{phys}$ plan has a more conformal dose distribution than the 3D-CRT plan, reducing NTCP (from 25% to 13%) at constant TCP = 83%. Comparison of the TOC graphs for the $IMRT_{phys}$

and the $IMRT_{biol}$ plan suggests that an increase in TCP (from 83% to 87%) can be obtained at constant NTCP =13%. It can be seen that the three TOC graphs do not cross and have different AUCs (3D-CRT: 0.87, $IMRT_{phys}$: 0.91, $IMRT_{biol}$: 0.93). The $IMRT_{biol}$ plan outperforms the $IMRT_{phys}$ plan over the whole range of prescribed dose levels and fractionation schemes, whereas the latter outperforms the 3D-CRT plan. By comparing the TOC graphs, its use for prescription dose customisation becomes apparent; when rescaling a given treatment plan does not fulfill the TCP/NTCP trade-off requirements, only re-optimization (with more direct steering of the TCP/NTCP criteria) will improve the quality of the plan.

Table 1 Dose and response statistics for renormalised 3D-CRT plans

		Dose-per-fraction ($N = 39$)			Number of fractions (2 Gy/fraction)		
		95%	100%	105%	$N = 37$	$N = 39$	$N = 41$
prostate							
	D_{min}	73.2	77.1	80.9	73.1	77.1	81.1
	D_{mean}	74.3	78.2	82.1	74.2	78.2	82.2
	D_{max}	75.1	79.1	83.1	75.0	79.1	83.2
	TCP	76%	83%	89%	78%	83%	87%
rectum							
	D_{mean}	45.3	47.7	50.1	45.3	47.7	50.1
	D_{max}	74.9	78.9	82.8	74.9	78.9	82.9
	$NTCP_{GI}$	13%	25%	44%	16%	25%	38%
bladder							
	D_{mean}	31.6	33.3	34.9	31.6	33.3	35.0
	D_{max}	74.0	77.9	81.8	73.9	77.9	81.9
	$NTCP_{GU}$	13%	15%	16%	13%	15%	19%

Abbreviations: N, number of fractions; D_{min}, minimum dose; D_{mean}, mean dose; D_{max}, maximum dose, TCP, tumour control probability, $NTCP_{GI}$, probability of late gastrointestinal morbidity; $NTCP_{GU}$, probability of late genitourinary morbidity.

Figure 5 TOC graphs for 3D-CRT, IMRT$_{phys}$ and IMRT$_{biol}$ plans.
TOC graphs for TCP *vs.* NTCP$_{GI}$ obtained by dose level scaling of a
prostate 3D-CRT plan (solid curve), the IMRT$_{phys}$ plan (dashed curve),
and the IMRT$_{biol}$ plan (dotted curve) at constant fraction number,
$N = 39$. Symbols ($\bullet$, $\circ$, $\square$) represent the TCP and NTCP scores of
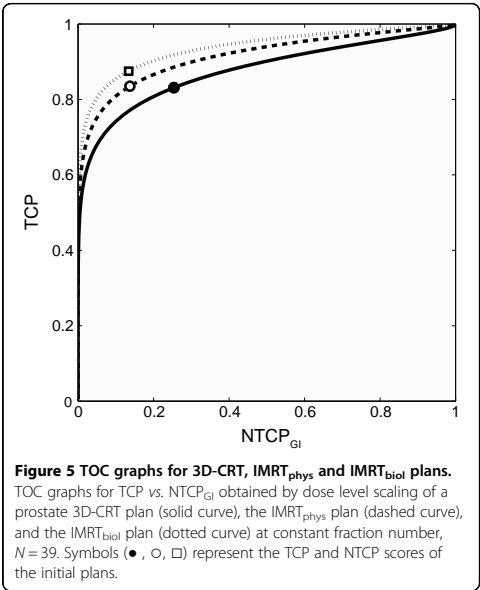the initial plans.

Figure 5 clearly shows that both the TOC's shape and
its AUC help to estimate the therapeutic power of a
treatment plan over a continuous range of TCP and
NTCP scores. Finally, the TOC graph of the best
performing (IMRT$_{biol}$) plan can be presented to the

**Table 2 Dose and response statistics for initial treatment
plans (39 fractions)**

		3D-CRT	IMRT$_{phys}$	IMRT$_{biol}$
prostate				
	D$_{min}$	77.1	75.4	74.1
	D$_{mean}$	78.2	78.8	81.8
	D$_{max}$	79.1	81.8	97.0
	TCP	83%	83%	87%
rectum				
	D$_{mean}$	47.7	36.8	36.3
	D$_{max}$	78.9	80.1	84.7
	NTCP$_{GI}$	25%	13%	13%
bladder				
	D$_{mean}$	33.3	25.1	17.7
	D$_{max}$	77.9	82.4	93.8
	NTCP$_{GU}$	15%	14%	14%

Abbreviations: 3D-CRT, three-dimensional conformal radiation therapy; *IMRT$_{phys}$*,
physically optimised IMRT plan; *IMRT$_{biol}$*, physico-biologically optimised IMRT
plan; *D$_{min}$*, minimum dose; *D$_{mean}$*, mean dose; *D$_{max}$*, maximum dose, *TCP*,
tumour control probability, *NTCP$_{GI}$*, probability of late gastrointestinal
morbidity; *NTCP$_{GU}$*, probability of late genitourinary morbidity.

physician and/or patient to assess the trade-off between
treatment benefit and morbidity and to choose from the
TOC graph the dose-level that best fits their preferences.

Discussion

In an era of individualised cancer therapy, radiotherapy
should move towards customised dose prescription in
order to maximise individual patient's outcome and
quality of life. The currently applied 'rigid' treatment
protocols neither take into account the anatomical diver-
sity of patients within the risk group nor their individual
risk-taking preferences. The standard strategy will lead
to relative underdosage in individuals who are willing to
tolerate a higher radiation dose aiming for tumour
control, whilst relatively overdosing those not willing to
accept the possible adverse effects resulting from the
predefined radiation dose. Hence, there is a need to
move toward customised treatment planning where
individualisation is not restricted to adapting the spatial
dose distribution to the patient's anatomy, but also
involves balancing of treatment benefit and morbidity in
terms of TCP and NTCP.

As an overture towards individualised TCP and NTCP
risk balancing, we conducted a prospective decision-
making trial in patients with localised prostate carcinoma
scheduled for 3D-CRT, who were offered an *a priori*
restricted choice between treatments with two alternative
pre-selected dose levels [2]. To take this one step further,
a tool to depict a continuum of clinically relevant dose
levels and their corresponding TCP and NTCP indices
would be helpful to assist selecting an optimum dose level
after an initial treatment plan has been generated for the
individual patient.

In previous work by Lind *et al.*, the concept of maximiz-
ing the probability of uncomplicated tumour control, P_+, as
a function of dose has been proposed to find the optimum
dose level [10]. It should be noted that the TOC graph pro-
vides unbiased information in comparison to the P_+ graph
when plotted as a function of dose. This becomes clear
when the general expression for P_+ is considered:

$$P_+ = (TCP - NTCP + \delta\{NTCP(1 - TCP)\}), \qquad (1)$$

where δ is the estimated fraction of patients for which
tumour and normal tissue response are statistically
independent ($0 \leq \delta \leq 1$) [9]. Expression (1) assumes that
an implicit *a priori* trade-off between TCP and NTCP is
made. For example, with $\delta = 0$ equation (1) reduces to
$P_+ = TCP - NTCP$, assuming that the risk of recurrence is
equally important as the risk of suffering from (severe)
side effects. Therefore, P_+ has often been criticised as a
measure that does not reflect the clinical reality that
reductions in TCP are rated differently from the risk of
complications by clinicians and patients. By plotting TCP

vs. NTCP with the prescribed dose as an independent parameter, no assumption with regard to *a priori* risk-taking preferences between TCP and NTCP is made. Instead, their interrelationship is assessed over the whole range of potential dose levels to facilitate the selection of an optimum that satisfies the risk-taking preferences of the individual clinician or patient. Using the full TOC graph instead of selecting the single dose level where P_+ is maximal pre-empts the criticism against *a priori* balancing the TCP/NTCP trade-off.

In the 1970's Moore and Mendelsohn proposed the TOC curve as a method to optimise treatment levels in cancer therapy [11] and later on, it was used in studies on radiation therapy for head and neck tumours [12-14]. However, this concept has only been employed to determine an optimum dose level for a patient population in a 'one size fits all' approach. We emphasize that with the current possibilities and knowledge of inverse treatment planning techniques (*e.g.*, by explicitly incorporating dose-response relationships into the optimization as objective and/or constraint functions) the application options for individualized dose prescription strategies may be more important and clinically relevant than before.

The TOC graph used in the current work provides a means to visualise and explore the trade-off between TCP and NTCP in an intuitive manner to be used as a tool for *a posteriori* dose prescription customisation of an initial treatment plan for an individual patient. Plans with different TCP/NTCP trade-offs can be generated from the same underlying relative dose distribution by scaling of the prescribed dose or fractionation scheme. Since no re-planning is required, the TOC graph can be generated off-line. Furthermore, this approach facilitates the interactive balance between TCP and NTCP of a given plan *after* treatment planning and plan optimisation and does not require risk-taking predilections to be articulated *a priori*, as is the case in today's inverse treatment planning approaches. Additionally, by exploiting the AUC, a quantitative definition of the therapeutic power is provided independent of consensus on the dose level.

The clinical application of changing the prescribed dose and fractionation after an initial treatment plan has been generated has recently gained renewed interest as part of individualised dose prescription strategies that escalate the tumour dose until maximally tolerable NTCP limits are reached in, for example, non-small cell lung cancer radiotherapy [6,7,21,22]. Current treatment planning systems lack the means to assess the effects of dose or fractionation variation of a given treatment plan in terms of TCP and NTCP indices and do not provide insight in their interrelationship. Consequently, it is common practice to completely re-design and re-calculate a treatment plan once the dose description or fractionation schedule has changed. Our re-scaling

approach together with the TOC concept brings individualised dose prescription into clinical practice by providing an intuitive and easy-to-apply tool to find the preferred prescription dose for either a fixed number of fractions (by changing the dose per fraction) of for fixed fraction dose (by changing the number of fractions) which yields pre-selected NTCP limit(s) for the OAR(s).

Practical use of TOC graph
To exploit our concept in the clinical workflow, first, a TOC graph is generated from historical dose-response data that were derived for a *group* of patients treated with the same irradiation technique but with different dose prescriptions. The *individual* patient and the radiation oncologist interactively choose TCP/NTCP coordinates from the TOC graph in a first decision-making step, which results in an *intended* dose prescription. Subsequently, an initial treatment plan is designed for the intended dose prescription based on the individual patient's anatomy. Hence, the population-based TOC can be used to establish an evidence-based 'customised' dose prescription for a given treatment technique prior to designing a treatment plan for the *intended* dose prescription level. Finally, a second decision-making step is required to further customise the dose prescription by re-normalisation and to achieve at least the desired, but likely superior, TCP/NTCP scores.

Physico-biological treatment plan optimisation
We showed that constrained optimisation in inverse planning based on TCP and NTCP models resulted in an iso-toxic treatment plan with improved TCP. As the TCP model allows for some degree of dose heterogeneity in the PTV, a dose reduction in parts of the PTV is exploited to reduce the NTCP, whereas in other parts the dose is escalated. A steep dose fall-off at the border of the PTV keeps the TCP constant, and allows for better sparing of the OAR, which facilitates dose escalation resulting in the same NTCP with higher TCP.

Relation of TOC graph to Pareto efficient frontier
The TOC graph should not be confused with the Pareto efficient frontier (PEF) that has been discussed in radiotherapy lately (*e.g.*, [23]). The principles of TOC and PEF are fundamentally different. Whereas the TOC graph has been obtained from scaling a single treatment plan, a PEF would have required the physico-biological optimisation problem to be repeatedly solved to optimality for different *a priori* set NTCP constraint values. A comparison between the TOC and PEF is however beyond the scope of this paper and will be addressed in another paper by the authors.

Due to uncertainties in the radiobiological models, the choice for a final dose level should not solely be based on the TOC graph, but should always include the underlying

3D dose distribution. Preferably, the TOC graph should be presented together with the model parameters, TCP/NTCP values, and confidence levels [13].

However, confidence intervals for all model parameters are not available or may be unreliable for IMRT dose distributions, as they have mainly been derived from 3D conformal radiotherapy dose distributions. As more clinical data comes available the TOC-based estimate of the optimal dose level will get more trustworthy. We therefore recommend to carefully introduce the TOC into clinical practice, by gradually releasing the interval of dose-level re-scaling towards values where the uncertainties are largest. The true value of the TOC graph and underlying TCP/NTCP models should be assessed in clinical trials comparing the predicted and actual TCP and NTCP values.

Conclusions

For an individual patient, the TOC graph can be exploited as an *a posteriori* decision aid in risk-adapted dose prescription customisation of a given treatment plan as a function of the prescribed dose level or the number of fractions. It provides physicians and patients with a decision aid for individual risk-taking preferences in terms of TCP/NTCP trade-off. The AUC is a dose level independent measure of the therapeutic power of a treatment plan.

Additional files

Additional file 1: TCP and NTCP models used in systematic literature review.

Additional file 2: TCP and NTCP models used by ORBIT Workstation.

Abbreviations

AUC: Area under the curve; bNED$_5$: 5-year genitourinary biochemical no evidence of disease; DVH: Dose-volume histogram; EQD$_2$: Equivalent dose in 2 Gy fractions; GI: Gastrointestinal; GU: Genitourinary; ICRU: International commission on radiation units and measurements; IMRT: Intensity-modulated radiation therapy; IMRT$_{phys}$: Physically optimised IMRT plan; IMRT$_{biol}$: Physico-biologically optimised IMRT plan; MV: Megavolt; NTCP: Normal tissue complication probability; OAR: Organ at risk; P_+: Uncomplicated tumour control probability; PEF: Pareto efficient frontier; PTV: Planning target volume; ROC: Receiver operating characteristic; RTOG: Radiation therapy oncology group; TCP: Tumour control probability; TOC: Therapeutic operating characteristic; 3D-CRT: Three-dimensional conformal radiation therapy.

Competing interests

All authors declare to have no competing interests.

Authors' contributions

ALH, conceived and designed the study, carried out the data collection, performed the data analysis and interpretation, and drafted the manuscript. HH, participated in the design of the study and the interpretation of the data, and helped to draft the manuscript. JHK, contributed substantially to the critical revision of the paper. All authors read and approved the final manuscript.

References

1. Amols HI, Zaider M, Heyes MK, Schiff PB: **Physician/patient-driven risk assignment in radiation oncology: reality or fancy?** *Int J Radiat Oncol Biol Phys* 1997, **38**:455–461.
2. Van Tol-Geerdink JJ, Stalmeier PF, Van Lin EN, Schimmel EC, Huizenga H, Van Daal WA, Leer JW: **Do prostate cancer patients want to choose their own radiation treatment.** *Int J Radiat Oncol Biol Phys* 2006, **66**:1105–1111.
3. Van Tol-Geerdink JJ, Stalmeier PFM, Pasker-de Jong PCM, Huizenga H, Van Lin EN, Schimmel EC, Leer JW, Van Daal WA: **Systematic review of the effect of radiation dose on tumour control and morbidity in the treatment of prostate cancer by 3D-CRT.** *Int J Radiat Oncol Biol Phys* 2006, **64**:534–543.
4. Van Tol-Geerdink JJ, Stalmeier PF, Van Lin EN, Schimmel EC, Huizenga H, Van Daal WA, Leer JW: **Do patients with localised prostate cancer treatment really want more aggressive treatment?** *J Clin Oncol* 2006, **24**:4581–4586.
5. Warkentin B, Stavrev P, Stvreva N, Field C, Fallone BG: **A TCP-NTCP estimation module using DVHs and known radiobiological models and parameter sets.** *J Appl Clin Med Phys* 2004, **5**:50–63.
6. Van Baardwijk A, Bosmans G, Bentzen S, Boersma L, Dekker A, Wanders R, Wouters BG, Lambin P, De Ruysscher D: **Radiation dose prescription for non-small-cell lung cancer according to normal tissue dose constraints: an *in silico* clinical trial.** *Int J Radiat Oncol Biol Phys* 2008, **71**:1103–1110.
7. Hoffmann AL, Troost EG, Huizenga H, Kaanders JH, Bussink J: **Individualized dose prescription for hypofractionation in advanced non-smal-cell lung cancer radiotherapy: an *in silico* trial.** *Int J Radiat Oncol Biol Phys* 2012, **83**:1596–1602.
8. Van Baardwijk A, Wanders S, Boersma L, Borger J, Öllers M, Dingemans AM, Bootsma G, Geraedts W, Pitz C, Lunde R, Lambin P, De Ruysscher D: **Mature results of an individualized radiation dose prescription study based on normal tissue constraints in stage I to III non-small-cell lung cancer.** *J Clin Oncol* 2010, **28**:1380–1386.
9. Brahme A: **Treatment optimization using physical and radiobiological objective functions.** In *Radiation Therapy Physics*. Edited by Smith AR. Berlin: Springer; 1995:209–246.
10. Lind BK, Mavroidis P, Hyödynmaa S, Kappas C: **Optimisation of the dose level for a given treatment plan to maximise the complication-free tumour cure.** *Acta Oncol* 1999, **38**:787–798.
11. Moore DH, Mendelsohn ML: **Optimal treatment levels in cancer therapy.** *Cancer* 1972, **30**:97–106.
12. Tokars RP, Griem ML: **Carcinoma of the nasopharynx and optimisation of radiotherapeutic management for tumour control and spinal cord injury.** *Int J Radiat Oncol Biol Phys* 1979, **5**:1741–1748.
13. Metz CE, Tokars RP, Kronman HB, Griem ML: **Maximum likelihood estimation of dose-response parameters for therapeutic operating characteristic (TOC) analysis of carcinoma of the nasopharynx.** *Int J Radiat Oncol Biol Phys* 1982, **8**:1185–1192.
14. Ågren A, Brahme A, Turesson I: **Optimisation of uncomplicated control for head and neck tumours.** *Int J Radiat Oncol Biol Phys* 1990, **19**:1077–1085.
15. Andrews JR: **Benefit, risk and optimisation by ROC analysis in cancer therapy.** *Int J Radiat Oncol Biol Phys* 1985, **11**:1557–1562.
16. Hanley JA, McNeil BJ: **The meaning and use of the area under a receiver operating characteristic (ROC) curve.** *Radiology* 1982, **143**:29–36.
17. Eriksson K, Rehbinder H: **Advanced tools for radiobiological evaluation and optimisation of treatment plans.** In *Proceedings of XVth International Conference on the Use of Computers in Radiation Therapy (ICCR). Volume II: 4–7 June 2007*. Toronto: University of Toronto; 2007:60–64.
18. Miralbell R, Roberts SA, Zubizarreta E, Hendry JH: **Dose-fractionation sensitivity of prostate cancer deduced from radiotherapy outcomes of 5,969 patients in seven international institutional datasets: α/β = 1.4 (0.9 2.2) Gy.** *Int J Radiat Oncol Biol Phys* 2012, **82**:e17–e24.

19. Michalski JM, Gay H, Jackson A, Tucker SL, Deasy JO: **Radiation dose-volume effects in radiation-induced rectal injury.** *Int J Radiat Oncol Biol Phys* 2010, **76**(Suppl 3):S123–S129.

20. Dale E, Hellebust TP, Skjønsberg A, Høgberg T, Olsen DR: **Modeling normal tissue complication probability from repetitive computed tomography scans during fractionated high-dose-rate brachytherapy and external beam radiotherapy of the uterine cervix.** *Int J Radiat Oncol Biol Phys* 2000, **47**:963–971.

21. Uzan J, Nahum AE: **Radiobiologically guided optimisation of the prescription dose and fractionation scheme in radiotherapy using BioSuite.** *Br J Radiol* 2012, **85**:1279–1286.

22. Vinogradskiy Y, Tucker SL, Bluett JB, Wages CA, Liao Z, Martel MK: **Prescribing radiation dose to lung cancer patients based on personalized toxicity estimates.** *J Thorac Oncol* 2012, **7**:1676–1682.

23. Ottosson RO, Engstrom PE, Sjöstrom D, Behrens CF, Karlsson A, Knöös T, Ceberg C: **The feasibility of using Pareto fronts for comparison of treatment planning systems and delivery techniques.** *Acta Oncol* 2009, **48**:233–237.

TCP and NTCP models used in systematic literature review

In [3], a Probit model has been used to describe the tumour control probability (TCP) as a function of the prescribed dose level:

$$TCP(D) = \Phi\left(\frac{D - TD_{50}}{m \cdot TD_{50}}\right),$$

where

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-t^2/2)\mathrm{d}t.$$

is the standard normal cumulative distribution function, D is the prescribed equivalent total dose in 2 Gy fractions, TD_{50} is the effective dose at the 50% response level and m determines the slope of the sigmoidal function Φ. The model parameters used are TD_{50} = 56.18 Gy and m = 0.39 for the overall bNED. These values were derived from a fit over the dose domain of 60–80 Gy in 2 Gy fractions.

The normal tissue complication probability (NTCP) model used in [3] is a logistic regression model:

$$NTCP(D) = \frac{1}{1 + \exp(-a - bD)},$$

where D is the prescribed equivalent total dose in 2 Gy fractions, and a and b are fitting parameters. Parameter values a = −11.08 and b = 0.06 gave the best fit over the relevant dose domain for gastrointestinal toxicity, while for genitourinary toxicity parameter values a = −6.47 and b = 0.03 resulted from the fitting procedure.

TCP and NTCP models used by ORBIT Workstation

The linear-quadratic Poisson model was used to calculate the TCP for a heterogeneous dose distribution $\mathbf{D} = \{D_i\}$ [7]:

$$TCP(\mathbf{D}) = \prod_i \left(\exp\left(-\exp\left(e \cdot \gamma_{37} - (e \cdot \gamma_{37} - \ln(\ln 2))\frac{EQD_2(D_i)}{TD_{50}}\right)\right)^{v_i}\right),$$

where TD_{50} is the dose at the 50% response level, γ_{37} is the normalized dose-response gradient at the 37% response level, and D_i and v_i are the dose bins and relative volumes of the dose-volume histogram (DVH), respectively. The model parameters TD_{50} = 56.18 Gy and γ_{37} = 1.3 and α/β = 2 Gy perfectly match the dose-response of the above Probit TCP model over the dose domain of 60–80 Gy in 2 Gy fractions.

For NTCP calculation, the Lyman-Kutcher-Burman model was used with 2 Gy fractionation correction based on the linear-quadratic cell survival model [24, 25]:

$$NTCP(\mathbf{D}) = \Phi\left(\frac{gEUD_{2,a}(\mathbf{D}) - TD_{50}}{m \cdot TD_{50}}\right),$$

where $\Phi(z)$ is the standard normal cumulative distribution function, and

$$gEUD_{2,a}(\mathbf{D}) = \left[\sum_i v_i EQD_2(D_i)^a\right]^{1/a}$$

is the generalized equivalent uniform dose for a 2 Gy fractionation-corrected DVH with dose bins D_i and relative volume v_i. Model parameters for Grade $\geq$ 2 late rectal toxicity (TD_{50} = 76.9 Gy, m = 0.13, a = 12.5) and α/β = 3 Gy were adopted from [19]. Model parameters for Grade $\geq$ 2 late bladder toxicity (TD_{50} = 62 Gy, m = 0.11, a = 7.7) and α/β = 6 Gy were adopted from [20].

III

Customisation of fraction dose and number in NSCLC radiotherapy

Clinical Investigation: Thoracic Cancer

Individualized Dose Prescription for Hypofractionation in Advanced Non-Small-Cell Lung Cancer Radiotherapy: An *in silico* Trial

Aswin L. Hoffmann, M.Sc., Esther G.C. Troost, M.D., Ph.D., Henk Huizenga, Ph.D., Johannes H.A.M. Kaanders, M.D., Ph.D., and Johan Bussink, M.D., Ph.D.

Radboud University Nijmegen Medical Centre, Department of Radiation Oncology, Nijmegen, The Netherlands

Summary

An *in silico* trial of 38 patients with advanced non-small-cell lung cancer treated by intensity-modulated radiation therapy investigated the therapeutic gain of individualized dose prescription with dose escalation based on normal tissue dose constraints for various hypofractionation schemes. This trial demonstrated that individualized dose escalation based on scaling of existing treatment plans with standard fractionation enables a therapeutic gain in 79% of the cases.

Purpose: Local tumor control and outcome remain poor in patients with advanced non-small-cell lung cancer (NSCLC) treated by external beam radiotherapy. We investigated the therapeutic gain of individualized dose prescription with dose escalation based on normal tissue dose constraints for various hypofractionation schemes delivered with intensity-modulated radiation therapy.

Methods and Materials: For 38 Stage III NSCLC patients, the dose level of an existing curative treatment plan with standard fractionation (66 Gy) was rescaled based on dose constraints for the lung, spinal cord, esophagus, brachial plexus, and heart. The effect on tumor total dose (TTD) and biologic tumor effective dose in 2-Gy fractions (TED) corrected for overall treatment time (OTT) was compared for isotoxic and maximally tolerable schemes given in 15, 20, and 33 fractions. Rescaling was accomplished by altering the dose per fraction and/or the number of fractions while keeping the relative dose distribution of the original treatment plan.

Results: For 30 of the 38 patients, dose escalation by individualized hypofractionation yielded therapeutic gain. For the maximally tolerable dose scheme in 33 fractions (MTD_{33}), individualized dose escalation resulted in a 2.5−21% gain in TTD. In the isotoxic schemes, the number of fractions could be reduced with a marginal increase in TED. For the maximally tolerable dose schemes, the TED could be escalated up to 36.6%, and for all patients beyond the level of the isotoxic and the MTD_{33} schemes (range, 3.3−36.6%). Reduction of the OTT contributed to the therapeutic gain of the shortened schemes. For the maximally tolerable schemes, the maximum esophageal dose was the dominant dose-limiting constraint in most patients.

Conclusions: This modeling study showed that individualized dose prescription for hypofractionation in NSCLC radiotherapy, based on scaling of existing treatment plans up to normal tissue dose constraints, enables dose escalation with therapeutic gain in 79% of the cases. © 2012 Elsevier Inc.

Keywords: Radiotherapy, Non-small-cell lung cancer, Hypofractionation, Dose escalation, Radiobiological modeling

---

Reprint requests to: Johan Bussink, M.D., Ph.D., Radboud University Nijmegen Medical Centre, Department of Radiation Oncology, PO Box 9101, 6500 HB, Nijmegen, The Netherlands. Tel: +31-24-3613665; Fax: +31-24-3610792; E-mail: j.bussink@rther.umcn.nl

Introduction

In patients with advanced non-small-cell lung cancer (NSCLC) treated by external beam radiation therapy (EBRT) alone, local tumor control and outcome remain poor. It is known that failure at the primary tumor site adversely influences progression-free, metastases-free, and overall survival. Better local control rates can be achieved through treatment modification by the addition of concurrent chemotherapy, intensification of the radiotherapy schedule or dose escalation. Concurrent chemotherapy has been shown to improve treatment outcome but also to increase acute (esophageal) side effects (1−3). Several Phase I/II trials explored altered EBRT fractionation schedules that increased the biologic effective dose to the primary tumor and reduced the local relapse rate (4−6). Intensification of the irradiation schedule by continuous, hyperfractionated, accelerated radiotherapy showed an absolute improvement in 2-year survival of 9% (5), but it is not available in many centers because of its heavy logistic load. In patients treated with concurrent chemoradiation, radiotherapy dose escalation facilitated by advancing technology resulted in an improved 3-year overall survival from 9% to 31% with no significant difference in side effects (4). Recently, van Baardwijk *et al.* published an individualized dose prescription study in a group of 166 Stage III NSCLC patients (6). Using a three-dimensional conformal radiotherapy (3D-CRT) technique with an accelerated scheme of 1.8-Gy fractions twice daily, patients were treated to the maximally tolerable dose (MTD) by increasing the number of fractions until normal tissue constraints for the healthy lung and spinal cord were met. Delivering a median tumor total dose (TTD) of 64.8 ± 11.4 Gy in a median overall treatment time (OTT) of 25 ± 5.8 days, the 1- and 2-year overall survival was favorable with acceptable toxicity.

It is evident that dose escalation in NSCLC patients is hampered by radiation-sensitive normal tissues, such as the surrounding healthy lung, spinal cord, esophagus, heart, and brachial plexus (1, 3, 7−10). Both local control and complication rates have improved over the past decade because of technological progress in conformal dose delivery. Further benefit from organ-sparing technologies (*e.g.*, intensity-modulated radiation therapy [IMRT]) can be expected when combined with innovative (*e.g.*, individualized), short fractionation schemes.

It has been estimated that the best local control rate could be achieved by delivering the total dose in an OTT of less than 5 weeks (11, 12). Hypofractionation strategies delivering a high TTD within a short OTT meet these demands. However, clinical trials and outcome data of hypofractionation schedules for advanced NSCLC are to be awaited. Most trials use a population-based dose prescription protocol, in which a rigid predefined dose prescription will lead to underdosage in some individuals and to adverse effects from overdosage in others.

Therefore, the aim of the present study was to pursue the potential implications of changing a widely used population-based treatment schedule (33×2 Gy) into a more intensified, individualized hypofractionation schedule with acceptable toxicity. The effect on tumor total dose was studied for both isotoxic and maximally tolerable hypofractionation schemes.

Methods and Materials

Patient characteristics

Thirty-eight consecutive patients with Stage IIIA/B NSCLC treated with curative intent at our institution (by sequential chemotherapy

and radiotherapy, concurrent schedules, or radiation therapy alone) were retrospectively included in this study (Table 1).

Organ segmentation and treatment planning technique

For all 38 patients, original treatment plans with standard fractionation were retrieved from the treatment planning system (TPS) archive (Pinnacle[3], version 8.0h; Philips Radiation Oncology Systems, Fitchburg, WI). The gross tumor volume (GTV_T) had been defined on a contrast-enhanced (^{18}F-fluorodeoxyglucose positron emission tomography) computed tomography scan, and the planning target volume was created following the institute's guidelines. Contouring of the lungs and heart had been done automatically by the TPS and was manually corrected if necessary. The esophagus had been delineated from the lower border of the cricoid cartilage to the gastroesophageal junction. The spinal cord had been considered to be at the inner margin of the entire bony thoracic spinal canal. All plans were generated using coplanar 10-MV photon beams and multisegment fields for step-and-shoot IMRT delivery. All plans had been normalized to a mean dose of

Table 1 Patient characteristics ($n = 38$)

Age (years; median [range])	66.1 [47.5−78.6]
Tumor location (no. of patients)	
Left upper lobe	8
Left lower lobe	7
Left hilus	2
Left lung	1
Right upper lobe	10
Right middle lobe	2
Right lower lobe	5
Right hilus	2
Unknown*	1
Stage (no. of patients)	
IIIA	23
IIIB	15
T stage (no. of patients)	
T0*	1
T1	4
T2	14
T3	7
T4	12
N stage (no. of patients)	
N0	2
N1	2
N2	29
N3	5
Organ volume (cm³; median [range])	
GTV_T	125 [5−802]
Lungs	4,193 [2,663−6,587]
Lungs−GTV_T	4,165 [2,655−6,571]
Treatment (no. of patients)	
Concurrent RT and CHT	22
Sequential RT and CHT	13
RT alone	3

Abbreviations: GTV_T = gross primary tumor volume; RT = radiation therapy; CHT = chemotherapy.

* Primary tumor not detected by computed tomography, positron emission tomography, or pathology.

66 Gy in 33 fractions satisfying the International Commission on Radiation Units and Measurements 50/62 guidelines.

Customized dose prescription based on normal tissue dose constraints

For each original treatment plan, the mean normalized total dose (NTD_{mean}) was calculated for the total volume of the paired lungs excluding the GTV_T (indicated as lungs−GTV_T) using an α/β ratio of 4 Gy (8). The NTD_{mean} is the mean of the doses delivered to lungs−GTV_T, adjusted for the dose per fraction (13). The NTD_{mean} was calculated using

$$NTD_{mean} = \overline{D} \cdot \frac{\alpha/\beta + \overline{D}/N \cdot \left(1 + \left(\frac{\sigma_D}{\overline{D}}\right)^2\right)}{\alpha/\beta + 2}, \qquad (1)$$

where $\overline{D}$ and σ_D are the mean and the standard deviation of the physical total dose distribution in lungs−GTV_T, respectively (see Appendix 1, available online at www.redjournal.org). Because the NTD_{mean} is strongly correlated with the incidence of Grade ≥ 2 radiation-induced pneumonitis (7), an $NTD_{mean} \leq 19$ Gy_4 was used as a constraint. All patients had a pretreatment forced expiratory volume at 1 s of more than 1.5 L and a diffusion capacity of lungs for carbon monoxide higher than 40%.

For organs at risk with a predominant "serial" structure, the maximum normalized total dose, NTD_{max}, was calculated as the maximum dose in 2-Gy fractions using the EQD_2 formula:

$$NTD_{max} = D_{max} \cdot \frac{\alpha/\beta + D_{max}/N}{\alpha/\beta + 2}, \qquad (2)$$

where D_{max} is the maximum total dose in the organ at risk, N is the number of fractions, and α/β is the fractionation sensitivity of the organ.

The monitor units of all beams of the original treatment plans were simultaneously rescaled using an individualized prescribed MTD based on maximum normal tissue dose constraints for the lung ($NTD_{mean} \leq 19$ Gy_4 for low-medium grade pneumonitis) (7); esophagus ($NTD_{max} \leq 80$ Gy_3 and $NTD_{mean} \leq 34$ Gy_3 for late esophagitis/esophageal stricture; $\alpha/\beta = 3$ Gy) (3); spinal cord ($NTD_{max} \leq 50$ Gy_2 for myelopathy; $\alpha/\beta = 2$ Gy) (10); brachial plexus ($NTD_{max} \leq 66$ Gy_2 for plexopathy; $\alpha/\beta = 2$ Gy) (10); heart ($NTD_{mean} \leq 26$ Gy_3 for pericarditis; $\alpha/\beta = 3$ Gy) (9). As a conservative estimate of the MTD, we used an $\alpha/\beta = 3$ Gy for late esophageal toxicity instead of $\alpha/\beta = 8-10$ Gy for acute esophageal toxicity. Acute esophagitis is burdensome during and shortly after completion of (chemo)radiation, but it does not influence daily life in the long term.

Rescaling was accomplished by altering the fraction size and/or the number of fractions while keeping the same beam arrangement and the relative dose distribution of the original treatment plan. To investigate the value of rescaling on TTD and the effect of the OTT, five schedules were assessed for each patient based on the original treatment plan (TP_{orig}) using these two approaches.

1. The isotoxic dose approach aimed at obtaining isotoxic schedules (ITD_{15} and ITD_{20}) by rescaling the fraction size of TP_{orig} at a fixed fraction number ($N = 15$ and $N = 20$) while applying the normal tissue NTD levels calculated from TP_{orig} as constraints.
2. The maximally tolerable dose approach aimed at dose-escalation of ITD_{15}, ITD_{20}, and TP_{orig} to obtain MTD_{15}, MTD_{20}, and MTD_{33}, respectively, by further scaling the

fraction size depending on the previously mentioned *maximal* normal tissue dose constraints.

In both approaches, it was the first-met dose limit that stopped the escalation.

The tumor effective dose (TED) was calculated as the normalized tumor total dose, corrected for OTT:

$$TED = N \cdot d \cdot \frac{\alpha/\beta + d}{\alpha/\beta + 2} - D_{prolif} \cdot \max[0, (t - t_k)], \qquad (3)$$

where N is the number of fractions, d the dose per fraction, α/β the fractionation sensitivity, D_{prolif} the dose recovered daily in 2-Gy fractions, t the OTT, and t_k is the onset point for tumor cell repopulation in days since start of the treatment. D_{prolif} was assumed to be 0.45 Gy (14), whereas t_k was chosen at 21 days (11). The hypothesized effect of additional chemotherapy on accelerated tumor cell repopulation was not taken into account as conclusive data are lacking. No temporal corrections for OTT or incomplete repair were made for the organs at risk. In this study, the maximum TTD (TTD_{max}) was restricted to 82.5 Gy in 33 fractions (*i.e.*, a maximum increase of 25% in total dose relative to the standard scheme), which is in line with the dose levels applied in other (radiation only) dose escalation studies using 3D-CRT (4, 15).

Sensitivity analysis

Because the benefit of reducing OTT is mainly driven by D_{prolif} and t_k, we investigated how robust the results were depending on these parameters. We varied D_{prolif} over the interval 0.45–0.60 Gy/day at constant $t_k = 21$ days, and varied t_k over the interval 14–28 days at constant $D_{prolif} = 0.45$ Gy/day. We investigated worst-case scenarios for D_{prolif} and t_k to give no improvement in TED with respect to standard fractionation. The α/β ratios of the dominant dose limiting toxicities were perturbed to assess the effect on TED.

Statistical analysis

Results were analyzed for each patient in the cohort individually. The percentage of patients that would benefit from the individualized dose escalation schemes was calculated. Median dose values for TTD, TED, NTD_{mean}, and NTD_{max} were calculated for the cohort. The Wilcoxon signed-rank test was used to compare the cohort results.

Results

Benefit for individual patients

For 12 of 38 patients, either the maximally tolerable dose to the heart ($n = 4$) or the brachial plexus ($n = 8$) had already been reached in the original treatment plan. In the former 4 patients, a 5% therapeutic gain in TED could only be obtained under isotoxic conditions at 15 or 20 fractions, whereas in the latter 8 patients, no gain could be obtained at all. For the remaining 26 patients, dose escalation always yielded therapeutic gain (Table 2). For the isotoxic schemes ITD_{15} and ITD_{20}, the number of fractions could be safely reduced for these patients while a marginal increase of 0.2–8.3% and of 2.9–8.3% in TED was obtained, respectively. For the maximally tolerable schemes, dose escalation was feasible for all 26 patients, resulting in a gain of 2.5–21% in TTD and in tumor fraction dose for the MTD_{33}

Table 2 Median and (range) of dose to tumor in patient cohort for different schedules

	TP_{orig} ($n = 30$)	ITD_{20} ($n = 30$)	ITD_{15} ($n = 30$)	MTD_{33} ($n = 26$)	MTD_{20} ($n = 26$)	MTD_{15} ($n = 26$)
TTD (Gy)	66.0 —	56.7 (55.4−57.8)	51.5 (49.8−53.1)	74.2* (67.5−79.9)	64.2 (56.9−68.5)	58.0 (51.0−63.6)
TFD (Gy)	2.00 —	2.84 (2.77−2.89)	3.43 (3.32−3.54)	2.25* (2.05−2.42)	3.21 (2.85−3.43)	3.86 (3.40−4.24)
TED (Gy_{10})	55.2 —	58.4* (56.8−59.8)	57.7 (55.3−59.8)	64.9* (57.0−71.9)	68.3* (58.6−74.4)	67.0* (57.0−75.4)

Abbreviations: TP_{orig} = original treatment plan; ITD_{20} = isotoxic dose plan in 20 fractions; ITD_{15} = isotoxic dose plan in 15 fractions; MTD_{33} = maximally tolerable dose plan in 33 fractions; MTD_{20} = maximally tolerable dose plan in 20 fractions; MTD_{15} = maximally tolerable dose plan in 15 fractions; TTD = tumor total dose; TFD = tumor fraction dose; TED = tumor effective dose in 2-Gy fractions corrected for overall treatment time.
   * Significant difference from TP_{orig}; $p < 0.001$.

schedule, and of 3.3−36.6% and 6.2−34.8% in TED for MTD_{15} and MTD_{20}, respectively. Hence, 30/38 cases (79%) had benefited from either isotoxic or maximally tolerable dose escalation (68.5%) or from isotoxic dose escalation alone (10.5%). In 8/38 cases (21%), no dose escalation was feasible at all.

Comparison of schemes for patient cohort

Table 2 shows the median TTD, tumor fraction dose, and TED for all six schemes. Both isotoxic and maximally tolerable schemes showed a higher TED at 15 and 20 fractions than the standard scheme at 33 fractions. There was a significant difference between the TED of the ITD_{15} and ITD_{20} schemes ($p < 0.001$), and of the MTD_{15} and MTD_{20} schemes ($p = 0.0001$). The TED of the MTD_{33} schedule was significantly lower than for the MTD_{15} or MTD_{20} schemes ($p < 0.001$). The results suggest that the TED can be safely escalated by 5.8% and 23.7% relative to the standard scheme for the isotoxic and maximally tolerable schemes given in 20 fractions, respectively.

Dose to tumor: Effect of repopulation

In Figure 1, TEDs for both the individually escalated isotoxic and maximally tolerable schemes are graphed against the number of fractions for a typical patient. The ascending phase shows the

TED to increase with time, whereas at 16 fractions (and 21 days) the repopulation effect begins to subtract from the effective dose. The increase in TED then becomes slower with repopulation and reaches a maximum at 20 fractions, after which the TED falls off because of prolongation of the OTT.

In 24 of the 30 patients (80%), the maximum TED was reached at 20 fractions for both the isotoxic and maximally tolerable schemes, whereas in 6 patients (20%) the maximum was at 15 fractions.

Sensitivity analysis

Increasing D_{prolif} from 0.45 Gy/day to 0.60 Gy/day caused an increase in TED of 6.5% for both ITD_{15} and ITD_{20} and of 7.2% for MTD_{15} and MTD_{20}. Increasing D_{prolif} did not change the proportion of patients with the largest benefit from MTD_{15} or MTD_{20}, but more patients benefitted from ITD_{15} as opposed to ITD_{20}. Perturbation of t_k showed that the loss in TED was more sensitive to delayed onset than the gain in TED is for advanced onset of tumor cell repopulation. The uncertainty in TED from t_k was estimated to be less than -6% for the 15 and 20 fraction schedules. For most patients, we found TED to not improve for ITD_{15} and ITD_{20} relative to the standard schedule in case t_k <15 days and $D_{prolif} \approx 0.25$ Gy/day. Because these values seem low in relation to literature (11, 14), we believe our estimates of therapeutic gain are conservative.
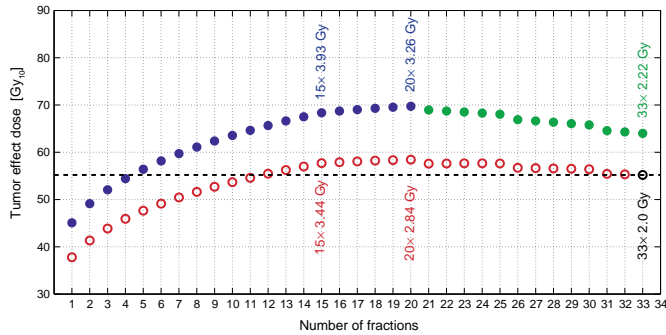

Fig. 1.   Typical example of tumor effective dose as a function of the number of fractions for isotoxic (open dots) and maximally tolerable (filled dots) schemes. Colors represent the dominant dose-limiting constraint ($\bigcirc$ (red)= NTD_{max}(cord); $\bullet$ (blue)= NTD_{max}(plexus); $\bullet$ (green)= NTD_{mean}(lungs−GTV_T)). The thick dashed horizontal line represents the 55.2 Gy_{10} tumor effective dose of the standard scheme.

Table 3 Median and (range) of normalized total dose for organs at risk in patient cohorts for different schedules

	TP_{orig} ($n = 30$)	ITD_{20} ($n = 30$)	ITD_{15} ($n = 30$)	MTD_{33} ($n = 26$)	MTD_{20} ($n = 26$)	MTD_{15} ($n = 26$)
NTD_{max} (cord)	34.9 (19.1–40.6)	34.9 (18.9–40.5)	34.9 (18.8–40.7)	$39.5^{\dagger}$ (22.7–49.4)	$39.8^{\dagger}$ (21.8–49.7)	$39.7^{\dagger}$ (21.3–49.8)
NTD_{mean} (lungs–GTV_T)	14.6 (10.3–18.5)	13.8* (9.4–17.3)	$13.2^{\dagger}$ (8.9–16.6)	$16.9^{\dagger}$ (10.7–19.0)	15.8* (9.8–19.0)	15.1 (9.3–19.0)
NTD_{max} (plexus)	1.6 (0.2–63.4)	1.4 (0.2–63.4)	1.3 (0.2–63.4)	2.1 (0.2–66.0)	1.8 (0.2–66.0)	1.7 (0.1–66.0)
NTD_{mean} (esophagus)	20.6 (10.0–37.3)	20.4 (9.5–36.8)	20.3 (9.2–36.6)	$24.2^{\dagger}$ (11.7–34.0)	23.6* (11.5–34.0)	23.2* (11.4–34.0)
NTD_{max} (esophagus)	66.3 (43.3–70.8)	67.8 (43.5–70.9)	68.3 (43.6–71.0)	$80.0^{\dagger}$ (52.1–80.0)	$80.0^{\dagger}$ (54.1–80.0)	$80.0^{\dagger}$ (55.7–80.0)
NTD_{mean} (heart)	13.4 (3.4–30.9)	12.9 (3.1–29.9)	12.7 (2.9–29.4)	14.6* (3.9–26.0)	13.9* (3.5–26.0)	13.5* (3.2–26.0)

Abbreviations: TP_{orig} = original treatment plan; ITD_{20} = isotoxic dose plan in 20 fractions; ITD_{15} = isotoxic dose plan in 15 fractions; MTD_{33} = maximally tolerable dose plan in 33 fractions; MTD_{20} = maximally tolerable dose plan in 20 fractions; MTD_{15} = maximally tolerable dose plan in 15 fractions; NTD_{mean} = mean normalized total dose (in Gy); NTD_{max} = maximum normalized total dose (in Gy).
Significant difference from TP_{orig}: *$p < 0.05$, $^{\dagger}p < 0.01$, and $^{\ddagger}p < 0.001$.

Varying the α/β for myelopathy over the range of $1-4$ Gy, we found the median TED to change $\pm4\%$ for ITD_{15} and ITD_{20}. For the MTD schemes, the median TED changed $\pm7\%$ when varying the α/β for late esophageal toxicity over the range of $1-6$ Gy. We estimate that the added uncertainties in quadrature do not exceed the 10% level.

Dose to normal tissues

Table 3 shows the median NTD values of the relevant organs at risk for all schemes. No significant differences were observed for the median NTD values of the ITD_{15} and ITD_{20} schemes in comparison with the standard scheme, except for the residual healthy lungs, where a lower NTD_{mean} is found. The median NTD values of the three MTD schemes were significantly greater than those of the standard scheme, except for the NTD_{mean} of the healthy lungs for MTD_{15}. As expected, the median NTD_{max} of the brachial plexus was low due to the distant primary tumor localization relative to the brachial plexus in the majority of patients. For all organs at risk, the median NTD values were below the respective tolerance dose levels. For the esophagus, the median NTD_{max} values for the three MTD schemes all reached the tolerance dose level for late esophageal toxicity.

Dominant dose-limiting organ

In Table 4, the frequency distribution of the dominant dose-limiting constraints is shown. The maximum cord dose was the dominant dose-limiting factor in 25 (83%) of the isotoxic plans, but in none of the three MTD schemes. The maximum dose to the brachial plexus was dose-limiting in 13% of the ITD plans. In roughly 20% of the MTD plans, the mean lung dose was the dominant constraint. The maximum esophageal dose was dose-limiting in all three MTD schemes: in roughly 60% of cases for MTD_{33} and in 70% for MTD_{15} and MTD_{20}. The mean doses to the esophagus and heart were dose-limiting in 1 patient for MTD_{15} and MTD_{20}, and in 2 patients for MTD_{33}. The TTD_{max} was not reached in any of the cases.

Discussion

Individualized dose prescription

In current clinical practice, all patients within a certain risk group scheduled for radiation therapy are treated according to "rigid" protocols and receive a predefined identical dose scheme. This strategy does not take into account the diversity in patients and tumors within the group leading to relative underdosage in some individuals in order to prevent the occurrence of adverse effects caused by overdosage in the entire group. Several steps have already been taken for individualization of beam shapes, beam weights, nonuniform beam intensities according to the tumor shape, and 3D patient anatomy. In an era of individualized cancer therapy, there is a need to move toward customized radiation treatment planning including individualized dose prescription. Acknowledging this, Van Baardwijk *et al.* studied the approach of individual dose prescription based on an accelerated scheme of 1.8-Gy fractions twice daily, and recently reported favorable overall survival rates with limited toxicity (6, 16).

Hypofractionation has regained interest, partially because outcome of clinical trials has shown a smaller than conventionally

Table 4 Frequency distribution of dominant normal tissue dose constraints

	ITD_{20} ($n = 30$)	ITD_{15} ($n = 30$)	MTD_{33} ($n = 26$)	MTD_{20} ($n = 26$)	MTD_{15} ($n = 26$)
NTD_{max} (cord)	25	25	0	0	0
NTD_{mean} (lungs$-GTV_T$)	0	0	6	5	5
NTD_{max} (plexus)	4	4	1	1	1
NTD_{mean} (esophagus)	1	1	2	1	1
NTD_{max} (esophagus)	0	0	15	18	18
NTD_{mean} (heart)	0	0	2	1	1

Abbreviations: ITD_{20} = isotoxic dose plan in 20 fractions; ITD_{15} = isotoxic dose plan in 15 fractions; MTD_{33} = maximally tolerable dose plan in 33 fractions; MTD_{20} = maximally tolerable dose plan in 20 fractions; MTD_{15} = maximally tolerable dose plan in 15 fractions; NTD_{mean} = mean normalized total dose; NTD_{max} = maximum normalized total dose.

expected difference between the fractionation sensitivities of tumors and normal tissues, but also because of the ability of modern treatment delivery techniques to produce highly conformal dose distributions. For organs with a marked volume effect such as the healthy lungs, moderate hypofractionation using modern delivery techniques may not substantially increase radiation pneumonitis risk even though the tumor fractionation sensitivity is smaller than that of the critical normal tissue (17, 18). Recently, two groups independently reported that hypofractionation tends to decrease the damage to the volume of the healthy lung for highly conformal dose distributions (18, 19). Based on a conventional fractionation schedule (*i.e.*, 60 Gy in 30 fractions), Vogelius *et al.* produced tumor isoeffective hypofractionation plans using 3D-CRT and tomotherapy, and found the latter to decrease the mean pneumonitis risk by 2.3% relative to the former when using 20 fractions. Consequently, historical trials of hypofractionation applying 3D-CRT techniques may overestimate the relative toxicity when compared with modern highly conformal treatment delivery techniques. Hence, with the increasing cost of cancer therapy, hypofractionation is attractive from an economic, logistic and patient convenience perspective, and may even allow for dose escalation with little detrimental effect to the therapeutic ratio.

Radiobiological modeling has shown that increased biologically effective dose to tumors can be achieved by shortening the radiation delivery schedule and increasing the dose per fraction. Based on these findings, we investigated the possibility of either isotoxic or maximally tolerable hypofractionation schedules escalating the radiation dose up to normal tissue dose constraints. For 26/38 patients, dose escalation was feasible in both isotoxic and maximally tolerable hypofractionation schemes with a shortened or conventional overall treatment time. As expected, the extent of dose escalation varied between individual patients mainly based on tumor localization relative to the organs at risk. In the modeling presented in this study, the organs-at-risk dose recovered per day was neglected. The D_{prolif} of 0.45 Gy/day is lower than 0.6 Gy/day assumed by others (14), resulting in conservative estimates of therapeutic gain. Furthermore, the onset of accelerated tumor cell repopulation t_k was assumed to be 21 days, although there is weak evidence for this, as the onset might start earlier. Sensitivity analysis has shown that the actual values of D_{prolif} and t_k are considered noncritical for the viability of our trial.

Clinical trials on treatment-related toxicity

As opposed to broad knowledge on pulmonary and neurological toxicity, no firm recommendations on the maximally tolerable esophageal dose exist (Radiation Therapy Oncology Group) and no

hypofractionation trials for primary esophageal tumors have been conducted. Therefore, two 3D-CRT based Phase I/II trials have been approved in the UK investigating individualized dose escalation based on normal tissue dose constraints in patients with Stage II or III NSCLC. In the multicenter Phase I/II trial, Isotoxic Dose Escalation and Acceleration in Lung Cancer ChemoRadiotherapy (IDEAL-CRT; ISRCTN12155469), concurrent chemotherapy is combined with a total radiation dose of 63−73 Gy given in 30 fractions over 6 weeks. Patients are allocated to one of seven total dose levels such that each patient has the same risk of Grade ≥ 2 radiation pneumonitis (based on an estimated relative normalized total dose [$rNTD_{mean}$] to the lung) and keeping within tolerance limits for the spinal cord or esophagus. The second trial, IsoToxic Accelerated RadioTherapy (I-START; CRUK/10/005), has recently been approved for patients with locally advanced NSCLC to be treated with radiation only. According to this protocol, the dose to an individual patient ranges from 58−65 Gy in 20 fractions over 4 weeks, taking into account established dose constraints to the organs at risk. The primary end point of both trials is to establish the maximally tolerable dose to the esophagus during and shortly after treatment. In these 3D-CRT based trials, the maximally achievable lung dose was estimated using the $rNTD_{mean}$ as a surrogate measure of healthy lung volume irradiated (13, 20), whereas in our study an exact calculation based on NTD_{mean} of Equation 1 was exploited in combination with highly conformal IMRT. Besides, we were able to calculate both isotoxic and maximally tolerable schedules with different numbers of fractions also incorporating corrections for overall treatment time. Noteworthy, our methodology allows the calculation of the individual prescription dose directly from the preset NTD_{mean} and NTD_{max} dose levels.

In the United States, the University of Wisconsin is conducting a helical tomotherapy-based hypofractionation study (NCT00214123) with pulmonary toxicity (pneumonitis Grade 3 lasting for more than 2 weeks) as primary end point. The purpose of this trial is to pilot reducing the duration of radiation treatment for lung cancer patients from 6 to 5 weeks using tomotherapy. Specific patient doses are based on tumor volume being treated and are prescribed using $rNTD_{mean}$ according to five dose bins, treating with an increasing dose per fraction at a fixed number of fractions. Results of an interim analysis have been reported (17, 18), and further data of these trials are to be awaited.

Conclusions

Our modeling study showed that individualized dose prescription for hypofractionation in Stage III NSCLC radiotherapy, based on

rescaling of existing treatment plans with standard fractionation up to normal tissue dose constraints, enables dose escalation with a therapeutic gain in 79% of the cases. For dose escalation beyond the presented levels, esophagus-sparing dose delivery techniques are mandatory.

References

1. Fenwick JD, Nahum AE, Malik ZI, *et al*. Escalation and intensification of radiotherapy for stage III non-small cell lung cancer: Opportunities for treatment improvement. *Clin Oncol (R Coll Radiol)* 2009; 21:343−360.

2. Rowell NP, O'Rourke NP. Concurrent chemoradiotherapy in non-small cell lung cancer. *Cochrane Database Syst Rev* 2004;CD002140.

3. Werner-Wasik M, Yorke E, Deasy J, *et al*. Radiation dose-volume effects in the esophagus. *Int J Radiat Oncol Biol Phys* 2010;76: S86−S93.

4. Nakayama H, Satoh H, Kurishima K, *et al*. High-dose conformal radiotherapy for patients with stage III non-small-cell lung carcinoma. *Int J Radiat Oncol Biol Phys* 2010;78:645−650.

5. Saunders M, Dische S, Barrett A, *et al*. Continuous, hyperfractionated, accelerated radiotherapy (CHART) versus conventional radiotherapy in non-small cell lung cancer: Mature data from the randomised multicentre trial. CHART Steering committee. *Radiother Oncol* 1999;52:137−148.

6. van Baardwijk A, Wanders S, Boersma L, *et al*. Mature results of an individualized radiation dose prescription study based on normal tissue constraints in stages I to III non-small-cell lung cancer. *J Clin Oncol* 2010;28:1380−1386.

7. Kwa SL, Lebesque JV, Theuws JC, *et al*. Radiation pneumonitis as a function of mean lung dose: An analysis of pooled data of 540 patients. *Int J Radiat Oncol Biol Phys* 1998;42:1−9.

8. Marks LB, Bentzen SM, Deasy JO, *et al*. Radiation dose-volume effects in the lung. *Int J Radiat Oncol Biol Phys* 2010;76:S70−S76.

9. Schultz-Hector S, Sund M, Thames HD. Fractionation response and repair kinetics of radiation-induced heart failure in the rat. *Radiother Oncol* 1992;23:33−40.

10. van der Kogel AJ. The nervous system: Radiobiology and experimental pathology. In: Scherer E SC, Trott K-R, editor. *Radiopathology of Organs and Tissues*. Berlin: Springer, 1991.

11. Withers HR, Taylor JM, Maciejewski B. The hazard of accelerated tumor clonogen repopulation during radiotherapy. *Acta Oncol* 1988; 27:131−146.

12. Fowler JF, Chappell R. Non-small cell lung tumors repopulate rapidly during radiation therapy. *Int J Radiat Oncol Biol Phys* 2000;46: 516−517.

13. Lebesque JV, Keus RB. The simultaneous boost technique: The concept of relative normalized total dose. *Radiother Oncol* 1991;22:45−55.

14. Bentzen SM, Saunders MI, Dische S. From CHART to CHARTWEL in non-small cell lung cancer: Clinical radiobiological modelling of the expected change in outcome. *Clin Oncol (R Coll Radiol)* 2002;14: 372−381.

15. Bradley J, Graham MV, Winter K, *et al*. Toxicity and outcome results of RTOG 9311: A phase I-II dose-escalation study using three-dimensional conformal radiotherapy in patients with inoperable non-small-cell lung carcinoma. *Int J Radiat Oncol Biol Phys* 2005;61: 318−328.

16. van Baardwijk A, Bosmans G, Bentzen SM, *et al*. Radiation dose prescription for non-small-cell lung cancer according to normal tissue dose constraints: An in silico clinical trial. *Int J Radiat Oncol Biol Phys* 2008;71:1103−1110.

17. Adkinson JB, Khuntia D, Bentzen SM, *et al*. Dose escalated, hypo-fractionated radiotherapy using helical tomotherapy for inoperable non-small cell lung cancer: Preliminary results of a risk-stratified phase I dose escalation study. *Technol Cancer Res* 2008;7:441−447.

18. Vogelius IS, Westerly DC, Cannon GM, *et al*. Hypofractionation does not increase radiation pneumonitis risk with modern conformal radiation delivery techniques. *Acta Oncol* 2010;49:1052−1057.

19. Jin JY, Kong FM, Chetty IJ, *et al*. Impact of fraction size on lung radiation toxicity: Hypofractionation may be beneficial in dose escalation of radiotherapy for lung cancers. *Int J Radiat Oncol Biol Phys* 2010;76:782−788.

20. Panettieri V, Malik ZI, Eswar CV, *et al*. Influence of dose calculation algorithms on isotoxic dose-escalation of non-small cell lung cancer radiotherapy. *Radiother Oncol* 2011;97:418−424.

Appendix 1: Analytical formula for mean normalized total dose

Conventionally, the mean normalized total dose, NTD_{mean}, is calculated from conversion of the differential dose-volume histogram (DVH) using:

$$NTD_{\text{mean}} = \sum_i v_i NTD_i, \qquad (\text{III.1})$$

where v_i is the fractional volume of the organ irradiated to a normalized total dose NTD_i (7), defined as the biologically equivalent dose in 2 Gy fractions of the corresponding physical total dose D_i given in n fractions (13):

$$NTD_i = D_i \frac{\alpha/\beta + D_i/n}{\alpha/\beta + 2}. \qquad (\text{III.2})$$

Substitution of (III.2) into (III.1) and rearrangement of sum terms yields:

$$NTD_{\text{mean}} = \frac{\alpha/\beta}{\alpha/\beta + 2} \sum_i v_i D_i + \frac{1/n}{\alpha/\beta + 2} \sum_i v_i D_i{}^2. \qquad (\text{III.3})$$

By noting that $\sum_i v_i D_i$ and $\sum_i v_i D_i{}^2$ are the weighted first and second sample moments of the physical total dose distribution $\mathbf{D} = \{D_i\}$, respectively, and the weighted sample variance of $\mathbf{D}$ is defined as: $\sigma_{\mathbf{D}}^2 = \sum_i v_i D_i{}^2 - (\sum_i v_i D_i)^2$, equation (III.3) can be rewritten in the form:

$$NTD_{\text{mean}} = \bar{D} \frac{\alpha/\beta + \bar{D}/n \left(1 + \left(\frac{\sigma_{\mathbf{D}}}{\bar{D}}\right)^2\right)}{\alpha/\beta + 2}, \qquad (\text{III.4})$$

where $\bar{D} = \sum_i v_i D_i$ is the weighted mean dose of the dose distribution $\mathbf{D}$.

From (III.4) it can be concluded that NTD_{mean} can be calculated from the mean total dose, $\bar{D}$, and the standard deviation $\sigma_{\mathbf{D}}^2$, provided that the fractionation sensitivity measure α/β and the number of fractions n are given. Since most treatment planning systems provide these two statistical metrics, the method to convert the physical dose in the DVH bins into corresponding NTD doses to calculate NTD_{mean} has become obsolete. The newly established analytical relation between NTD_{mean} and $\bar{D}$ makes experimentally determined first-order corrections of the latter out-dated (7, 16).

IV

Normal-tissue effective fractionation sensitivity

Fractionation in normal tissues: the $(\alpha/\beta)_{\text{eff}}$ concept can account for dose heterogeneity and volume effects

Aswin L Hoffmann[1] **and Alan E Nahum**[2]

[1] Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, 6201 BN Maastricht, The Netherlands
[2] Physics Department, Clatterbridge Cancer Centre, Bebington CH63 4 JY, UK

E-mail: aswin.hoffmann@maastro.nl and alan.nahum@clatterbridgecc.nhs.uk

Abstract

The simple Linear–Quadratic (LQ)-based Withers iso-effect formula (WIF) is widely used in external-beam radiotherapy to derive a new *tumour* dose prescription such that there is *normal-tissue* (NT) iso-effect when changing the fraction size and/or number. However, as conventionally applied, the WIF is invalid unless the normal-tissue response is solely determined by the tumour dose. We propose a *generalized* WIF (gWIF) which retains the tumour prescription dose, but replaces the *intrinsic* fractionation sensitivity measure (α/β) by a new concept, the normal-tissue *effective* fractionation sensitivity, $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$, which takes into account both the dose heterogeneity in, and the volume effect of, the late-responding normal-tissue in question. Closed-form analytical expressions for $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ ensuring exact normal-tissue iso-effect are derived for: (i) uniform dose, and (ii) arbitrary dose distributions with volume-effect parameter $n = 1$ from the normal-tissue dose–volume histogram. For arbitrary dose distributions and arbitrary n, a numerical solution for $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ exhibits a weak dependence on the number of fractions. As n is increased, $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ increases from its intrinsic value at $n = 0$ (100% serial normal-tissue) to values close to or even exceeding the tumour (α/β) at $n = 1$ (100% parallel normal-tissue), with the highest values of $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ corresponding to the most conformal dose distributions. Applications of this new concept to inverse planning and to highly conformal modalities are discussed, as is the effect of possible deviations from LQ behaviour at large fraction sizes.

1. Introduction

Fractionation was introduced early in the history of radiation therapy (Thames and Hendry 1987, Steel 2007). Coutard (1929) established what we would now recognize as standard fractionation for treating head and neck cancer: large numbers of relatively small doses of radiation (typically around 2 Gy), delivered each weekday, yielded the best 'therapeutic ratio', i.e. the highest chance of local tumour control (even cure) for acceptably low complication rates.

Five decades later, these essentially empirical findings were put on a theoretical basis via the Linear–Quadratic (LQ) model of cell killing (Fowler 1989, Chapman 2003, Steel 2007), that incorporates a fractionation sensitivity parameter α/β as ratio of the intrinsic radiosensitivities α and β. A mixture of animal experiments and empirical observations indicated that iso-effect curves (connecting total dose and dose per fraction) for 'late' normal-tissue (NT) complications were characterized by an α/β ratio of around 3 Gy (though with quite wide variations), whereas the corresponding curves for tumour iso-effect were best described by $\alpha/\beta \approx 10$ Gy (Thames *et al* 1982). It followed that for a given level of tumour control (assuming negligible clonogen proliferation) the effects on normal tissues would continually decrease as the number of fractions increased and the fraction size decreased.

1.1. Withers iso-effect formula

Making direct use of the LQ model, Withers *et al* (1983) derived a simple formula for converting from a reference fractionation regimen (total dose D_{ref}, fraction size d_{ref}) to a new one (total dose D_{new}, fraction size d_{new}), in order to ensure iso-effect for the tissue of interest, characterized by its α/β ratio:

$$\frac{D_{new}}{D_{ref}} = \frac{\alpha/\beta + d_{ref}}{\alpha/\beta + d_{new}}. \tag{1}$$

This equation, known as the Withers iso-effect formula (WIF), became the standard method for relating one fractionation regimen to another. It has been taught to generations of radiation oncologists, and can be found in numerous textbooks (Steel 2007, Joiner and van der Kogel 2009, Hall and Giaccia 2011). Generally one assumes $\alpha/\beta = 3$ Gy for late-responding normal tissues but the formula functions equally well for tumours when α/β is usually set equal to 10 Gy.

If (1) is used to convert a conventional multi-fraction regimen (e.g. 33 fractions of 2.0 Gy) to a small number of much larger fractions for normal-tissue iso-effect, then the new regimen will result in a drastically reduced tumour control probability (TCP). Equally well, if one instead derives a new total dose and larger fraction size for tumour iso-effect, a significant increase in complication probability would be expected. Therefore in order to maximize the 'therapeutic ratio' radiotherapy must be delivered in a large number of small fractions, which is one of the tenets of classical radiobiology.

There is increasing evidence that curative radiotherapy using large fraction sizes can be safely delivered using highly conformal irradiation techniques in specific situations e.g. early stage lung tumours, often referred to as stereotactic body radiotherapy or stereotactic ablative radiotherapy (SABR) (Blomgren *et al* 1995, Fowler *et al* 2004, Grau *et al* 2006, Timmerman 2008, Borst *et al* 2009). The application of (1) to such regimen (typically 3 × 15–20 Gy) strongly suggests that they are very far from normal-tissue iso-effective compared to standard well-tolerated regimens e.g. 20 × 2.75 Gy. However, estimates of normal-tissue complication probability (NTCP) for these very large fraction sizes are consistent with the acceptably low complication rates observed clinically. All the above suggests that there is a major problem with the Withers expression in certain clinically relevant situations involving modern, highly conformal dose-delivery techniques (Nahum and Chapman 2003, Calandrino *et al* 2005, Vogelius *et al* 2010, Jin *et al* 2010, Myerson 2011).

1.2. Aim of this paper

In this paper, starting from a critical examination of the WIF, we develop an alternative analysis of normal-tissue fractionation sensitivity. Our analysis explicitly takes into account both the dose heterogeneity present in most normal tissues and the organ-specific nature of the so-called

'volume effect', as addressed by Nahum and Kutcher (2007) and many other investigators. A modified form of the WIF is derived, and we show that this hetero- or generalized Withers expression explains why certain hypo-fractionated regimen, such as the ones referred to above, are within normal-tissue tolerance. Furthermore, the new expression provides a methodology to readily identify other situations where hypo-fractionation can or cannot be safely applied.

2. Theory

2.1. Re-derivation of Withers iso-effect formula

For N equal fractions of dose d, assuming that sublethal lesion repair is completed in the inter-fraction interval, the LQ-based surviving fraction (SF) is given by

$$
\begin{aligned}
\mathrm{SF}(N, d; \alpha, \beta) &= \exp(-\alpha d - \beta d^2)^N \\
&= \exp(-\alpha N d - \beta N d^2).
\end{aligned}
\tag{2}
$$

Identifying Nd as the total dose D, this becomes

$$
\begin{aligned}
\mathrm{SF}(D, d; \alpha, \beta) &= \exp(-\alpha D - \beta\, d D) \\
&= \exp\left[-\alpha D \left(1 + \frac{d}{\alpha/\beta}\right)\right].
\end{aligned}
\tag{3}
$$

Two fractionation regimens are said to be *iso-effective* if their surviving fractions are equal. Thus to convert from a reference regimen ($N_{\mathrm{ref}}, d_{\mathrm{ref}}$) to a new regimen ($N_{\mathrm{new}}, d_{\mathrm{new}}$) we equate the exponents:

$$
\alpha D_{\mathrm{new}} \left(1 + \frac{d_{\mathrm{new}}}{\alpha/\beta}\right) = \alpha D_{\mathrm{ref}} \left(1 + \frac{d_{\mathrm{ref}}}{\alpha/\beta}\right),
$$

and recover the WIF of (1), written alternatively as:

$$
\frac{D_{\mathrm{new}}}{D_{\mathrm{ref}}} = \left(1 + \frac{d_{\mathrm{ref}}}{\alpha/\beta}\right)\left(1 + \frac{d_{\mathrm{new}}}{\alpha/\beta}\right)^{-1}.
\tag{4}
$$

2.2. WIF in clinical practice: implicit assumptions

The application of the WIF to tumours is generally unproblematic. In most clinical situations the tumour receives an almost uniform dose and thus the tumour total dose D and the tumour fraction dose d are unambiguously defined. However, the WIF is more commonly used to convert from the reference regimen to one with a different number of fractions and/or fraction size which is *iso-effective* for normal tissue rather than for the tumour. Thus, $\alpha/\beta = 3$ Gy is typically used but it is not obvious what the normal-tissue doses D^{NT} and d^{NT} should be. The universally adopted practice is to use the prescribed total and fractional dose values for the *tumour*, D^{T} and d^{T}, respectively. Now the most critical normal tissue will usually be positioned adjacent to the tumour and therefore the *maximum* values of D^{NT} and d^{NT} will certainly be close to D^{T} and d^{T}.

   This use of the WIF is therefore strictly correct either for normal tissues receiving this maximum dose uniformly (almost never the case) or where normal-tissue response is solely determined by the maximum dose it receives, i.e. for 100% serial organs such as the spinal cord. For all other situations, i.e. the overwhelming majority encountered in clinical practice, the inescapable conclusion is that the WIF as conventionally applied is invalid.

   A clinical example of this invalidity is normal lung tissue surrounding a lung tumour. There is a large body of evidence that radiation pneumonitis is correlated with (radiobiologically adjusted) mean lung dose (MLD) in the paired non-involved lung (Marks *et al* 2010), and the MLD of a typical lung treatment plan is a certain percentage of the prescribed total dose D^{T};

for modern dose-delivery techniques this is often around 20% (Hoffmann *et al* 2012, Roelofs *et al* 2012). The WIF as conventionally applied implicitly assumes that a small high-dose region totally controls the complication response of the rest of the lung tissue. Therefore, the conventional use of the WIF to derive a new regimen that is intended to be iso-effective for radiation pneumonitis will effectively overestimate the normal-tissue complication probability (NTCP) and thus result in conservative prescriptions when the number of fractions is decreased. An example is given in section 3.1.

2.3. The concept of effective fractionation sensitivity, $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$

From the previous section, it is clear that the conventional WIF is invalid for normal-tissue iso-effect in the way it is generally used. Nevertheless, we wish to retain the WIF expression for its simplicity. We will now develop an alternative form of the WIF that explicitly takes into account both the dose heterogeneity present in most normal tissues and the organ-specific nature of the volume effect. The only change to the conventional WIF is the replacement of the *intrinsic* α/β ratio by a new concept: the *effective* α/β ratio.

For reasons of comprehensibility, we start with the hypothetical case where the normal tissue receives a uniform dose. Subsequently, we consider the general and more realistic case of a heterogeneous dose distribution, where we exploit a dose–volume histogram (DVH) reduction technique to represent the heterogeneous dose distribution by a uniform equivalent normal-tissue dose.

2.3.1. Effective α/β ratio for uniform normal-tissue dose. Suppose that under the reference regimen the normal tissue receives a uniform dose with fraction size, $d^{\mathrm{NT}}_{\mathrm{ref}}$, and that this dose is different to the tumour fraction size, $d^{\mathrm{T}}_{\mathrm{ref}}$. We now change the fractionation regimen iso-effectively to produce a new one with $d^{\mathrm{T}}_{\mathrm{new}}$ and $d^{\mathrm{NT}}_{\mathrm{new}}$ for N_{new} fractions. Further it is assumed that the normal tissue and tumour responses are characterized by $(\alpha/\beta)^{\mathrm{NT}}$ and $(\alpha/\beta)^{\mathrm{T}}$, respectively.

Applying the WIF for normal-tissue iso-effect to this situation, the ratio of the total doses for the new and reference regimens can be written in the same format as the WIF of (4):

$$\frac{D^{\mathrm{NT}}_{\mathrm{new}}}{D^{\mathrm{NT}}_{\mathrm{ref}}} = \left(1 + \frac{d^{\mathrm{NT}}_{\mathrm{ref}}}{(\alpha/\beta)^{\mathrm{NT}}}\right)\left(1 + \frac{d^{\mathrm{NT}}_{\mathrm{new}}}{(\alpha/\beta)^{\mathrm{NT}}}\right)^{-1}. \tag{5}$$

Note that the ratio of total doses must be the same in the tumour as in the normal tissues, as they both have to scale in the same way for a given treatment plan, i.e. the *relative* dose distribution is not being modified; it is only re-scaled.

It is important to realize that the above expression differs in one significant respect from the conventional use of WIF. In (5) we are using the respective doses in the normal tissue, and not those in the tumour. Suppose now that we wish to retain the way in which WIF is conventionally employed, i.e. we want the expression to involve the respective doses to the tumour; note that $D^{\mathrm{NT}}_{\mathrm{new}}/D^{\mathrm{NT}}_{\mathrm{ref}}$ must be equal to $D^{\mathrm{T}}_{\mathrm{new}}/D^{\mathrm{T}}_{\mathrm{ref}}$. In other words, we wish to find an 'effective' α/β ratio, $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$, such that

$$\frac{D^{\mathrm{NT}}_{\mathrm{new}}}{D^{\mathrm{NT}}_{\mathrm{ref}}} = \left(1 + \frac{d^{\mathrm{T}}_{\mathrm{ref}}}{(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}}\right)\left(1 + \frac{d^{\mathrm{T}}_{\mathrm{new}}}{(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}}\right)^{-1}, \tag{6}$$

where the left-hand side is given by (5). This will be the case provided that

$$(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}} = \frac{d^{\mathrm{T}}_{\mathrm{ref}}}{d^{\mathrm{NT}}_{\mathrm{ref}}}(\alpha/\beta)^{\mathrm{NT}} \tag{7}$$

and

$$(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}} = \frac{d^{\mathrm{T}}_{\mathrm{new}}}{d^{\mathrm{NT}}_{\mathrm{new}}}(\alpha/\beta)^{\mathrm{NT}} \tag{8}$$

for the reference and new regimen, respectively.

Identities (7) and (8) are equivalent because $d_{\text{ref}}^{\text{T}}/d_{\text{ref}}^{\text{NT}} = d_{\text{new}}^{\text{T}}/d_{\text{new}}^{\text{NT}}$ as we are merely re-scaling the reference dose distribution to produce the new one for a different number of fractions.

Summarizing, for this special case of a uniform dose delivered to the normal tissue for which we require iso-effect, the conventional WIF of (1) will still apply, using the reference and the new *tumour* fractional and total doses, but employing an effective α/β ratio given by

$$(\alpha/\beta)_{\text{eff}}^{\text{NT}} = \frac{d^{\text{T}}}{d^{\text{NT}}}(\alpha/\beta)^{\text{NT}}, \qquad (9)$$

where we have removed the subscripts 'ref' and 'new', since it does not matter which regimen is used as long as d^{T} and d^{NT} are taken from the same regimen. We introduce here the adjective 'intrinsic' for $(\alpha/\beta)^{\text{NT}}$ to distinguish it from the new concept $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$.

2.3.2. Effective α/β ratio for heterogeneous normal-tissue dose.   Now we will deal with the more complicated situation of a heterogeneous dose distribution. We consider that the correct way to describe normal-tissue iso-effectiveness is to use NTCP models that take into account the whole dose distribution and the volume effect. Our analysis is based on the assumption that the power-law model for local dose–volume effect is correct.

Suppose that under the reference regimen with N_{ref} fractions the normal tissue receives a heterogeneous fractional dose distribution, $(N_{\text{ref}}, \mathbf{d}_{\text{ref}}^{\text{NT}})$, where $\mathbf{d}_{\text{ref}}^{\text{NT}}$ is a vector of dose per voxel values. Re-scaling the dose distribution produces a new fractionation regimen, $(N_{\text{new}}, \mathbf{d}_{\text{new}}^{\text{NT}})$, for a different number of fractions, N_{new}. For normal-tissue iso-effect of the dose distributions $(N_{\text{ref}}, \mathbf{d}_{\text{ref}}^{\text{NT}})$ and $(N_{\text{new}}, \mathbf{d}_{\text{new}}^{\text{NT}})$, their NTCPs must to be equal. The NTCP can be expressed as a monotonically increasing function g of a generalized tissue damage function F, i.e.

$$\text{NTCP}(N, \mathbf{d}^{\text{NT}}) = g(F(N, \mathbf{d}^{\text{NT}})), \qquad (10)$$

where F is separable into a local damage function f weighting the biological effect of the dose per voxel for a given end-point:

$$F(N, \mathbf{d}^{\text{NT}}) = f^{-1}\left[\sum_{i=1}^{V} v_i f\left(E(N, d_i^{\text{NT}})\right)\right],$$

and f accounts for the volume effect of the effective dose $E(N, d_i)$ (see below) in voxel i having a fractional volume v_i, and $\sum_{i=1}^{V} v_i = 1$ is the normalized total volume over all V voxels of the whole organ (Romeijn *et al* 2004, Myerson 2011).

In the Lyman–Kutcher–Burman (LKB) NTCP model (Lyman 1985, Kutcher *et al* 1991) $g(z) = 1/\sqrt{2\pi} \int_{-\infty}^{z} e^{-t^2/2} \, \mathrm{d}t$ is the standard normal cumulative distribution function and $F(N, \mathbf{d}) = \text{gEUD}_a(N, \mathbf{d})$ is the generalized equivalent uniform dose (Niemierko 1999), which is based on the power-law relationship $f(D) = D^a$ relating dose and volume to local biological effect. The parameter a characterizes the volume-dependence of the tissue, and is related to the original Lyman (1985) formulation via $a = 1/n$. For normal tissues with a serial architecture the value of n is close to 0, whereas for organs with a parallel architecture it is closer to 1. To account for fractionation effects in a tissue with a given *intrinsic* fractionation sensitivity α/β, the function E is the biologically effective dose $\text{BED}(N, d_i; \alpha/\beta) = Nd_i(1 + d_i/(\alpha/\beta))$ (Fowler 1989).

We derive an expression for the effective α/β ratio in the context of the gEUD_a model with the BED-based fractionation correction. This generates a radiobiologically corrected gEUD_a, which is also known as the 'modified equivalent uniform dose' (Park *et al* 2005) or 'generalized equivalent uniform BED' (gEUBED_a) (Hoffmann *et al* 2008). Note that the same results would have been obtained for any NTCP model that satisfies identity (10).

We first consider the $n = 1$ case and then extend our methodology to $0 \leqslant n \leqslant 1$.

Normal tissue with 'parallel' architecture. In case of a 'parallel' organ, $n = 1$, $gEUBED_a$ reduces to the radiobiologically adjusted mean dose (BED_{mean}). As an example, we refer to non-small cell lung cancer (NSCLC) radiotherapy, where the BED_{mean} has been shown to correlate with radiation pneumonitis (Marks *et al* 2010).

Assuming normal-tissue iso-effect conditions, the ratio of the mean total doses for the new and reference regimens can be found by equating their analytical expressions for BED_{mean} (see Appendix A), analogous to the derivation of the ratio of total doses in (5):

$$\frac{\bar{D}_{new}^{NT}}{\bar{D}_{ref}^{NT}} = \left(1 + \frac{h \cdot \bar{d}_{ref}^{NT}}{(\alpha/\beta)^{NT}}\right)\left(1 + \frac{h \cdot \bar{d}_{new}^{NT}}{(\alpha/\beta)^{NT}}\right)^{-1}, \tag{11}$$

where $h = 1 + (\sigma_{\mathbf{d}}^{NT}/\bar{d}^{NT})^2$ is a normal-tissue dose heterogeneity factor, $\bar{d}^{NT}$ is the fractional mean dose, and $\sigma_{\mathbf{d}}^{NT}$ is the standard deviation of the fractional dose distribution $\mathbf{d}^{NT}$. Since the relative dose distribution is invariant under scaling, the dose heterogeneity factor h is equal for the new and reference regimens. Expressing the ratio of normal-tissue doses $\bar{D}_{new}^{NT}/\bar{D}_{ref}^{NT}$ in terms of (6), we obtain

$$\frac{\bar{D}_{new}^{NT}}{\bar{D}_{ref}^{NT}} = \left(1 + \frac{d_{ref}^{T}}{(\alpha/\beta)_{eff}^{NT}}\right)\left(1 + \frac{d_{new}^{T}}{(\alpha/\beta)_{eff}^{NT}}\right)^{-1}. \tag{12}$$

Equating (11) and (12), and solving for $(\alpha/\beta)_{eff}^{NT}$ yields

$$(\alpha/\beta)_{eff}^{NT} = \frac{1}{1 + \left(\sigma_{\mathbf{d}_{ref}}^{NT}/\bar{d}_{ref}^{NT}\right)^2}\frac{d_{ref}^{T}}{\bar{d}_{ref}^{NT}}(\alpha/\beta)^{NT} \tag{13}$$

and

$$(\alpha/\beta)_{eff}^{NT} = \frac{1}{1 + \left(\sigma_{\mathbf{d}_{new}}^{NT}/\bar{d}_{new}^{NT}\right)^2}\frac{d_{new}^{T}}{\bar{d}_{new}^{NT}}(\alpha/\beta)^{NT} \tag{14}$$

for the reference and new regimen, respectively. Identities (13) and (14) are equivalent because $d_{ref}^{T}/\bar{d}_{ref}^{NT} = d_{new}^{T}/\bar{d}_{new}^{NT}$ as we are merely re-scaling the reference dose distribution to produce the new one for a different number of fractions. For a homogeneous dose distribution, $\sigma_{\mathbf{d}_{ref}}^{NT} = \sigma_{\mathbf{d}_{new}}^{NT} = 0$, and these two identities reduce to (7) and (8), respectively.

Following the same reasoning as in subsection 2.3.1 and removing the subscripts 'ref' and 'new', we arrive at

$$(\alpha/\beta)_{eff}^{NT} = \frac{1}{1 + \left(\sigma_{\mathbf{d}}^{NT}/\bar{d}^{NT}\right)^2}\frac{d^{T}}{\bar{d}^{NT}}(\alpha/\beta)^{NT}. \tag{15}$$

Summarizing, for this special case of a heterogeneous dose distribution in a purely 'parallel' normal tissue for which we require iso-effect, the conventional WIF of (1) will still apply, using the tumour doses, providing that the intrinsic $(\alpha/\beta)^{NT}$ is replaced by the effective α/β ratio from (15).

Normal tissue with arbitrary architecture. We now address the general case of $0 \leqslant n \leqslant 1$. Suppose that a fictitious normal-tissue structure exists, which receives the tumour prescription dose (N_{ref}, d_{ref}^{T}). We now require the new prescription (N_{new}, d_{new}^{T}) such that the two fractionation regimens are iso-effective for an *effective* normal-tissue fractionation sensitivity $(\alpha/\beta)_{eff}^{NT}$ and for equal NTCPs of the corresponding normal-tissue dose distributions ($N_{ref}, \mathbf{d}_{ref}^{NT}$) and ($N_{new}, \mathbf{d}_{new}^{NT}$). Hence, the BEDs of both regimens and the corresponding $gEUBED_a$s of the dose distributions should be equal, i.e.

$$BED\left(N_{new}, d_{new}^{T}; (\alpha/\beta)_{eff}^{NT}\right) \equiv BED\left(N_{ref}, d_{ref}^{T}; (\alpha/\beta)_{eff}^{NT}\right), \tag{16}$$

where we use $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$, and

$$\text{gEUBED}_a\left(N_{\text{new}}, \mathbf{d}_{\text{new}}^{\text{NT}}; (\alpha/\beta)^{\text{NT}}\right)) \equiv \text{gEUBED}_a\left(N_{\text{ref}}, \mathbf{d}_{\text{ref}}^{\text{NT}}; (\alpha/\beta)^{\text{NT}}\right)), \quad (17)$$

where we use the intrinsic $(\alpha/\beta)^{\text{NT}}$ for the full dose distribution in the real normal tissue.

In order to obtain $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ from equations (16) and (17), we introduce the scaling factor f as:

$$f\left(N_{\text{new}}; N_{\text{ref}}, d_{\text{ref}}^{\text{T}}, (\alpha/\beta)_{\text{eff}}^{\text{NT}}\right) = d_{\text{new}}^{\text{T}}\left(N_{\text{new}}; N_{\text{ref}}, d_{\text{ref}}^{\text{T}}, (\alpha/\beta)_{\text{eff}}^{\text{NT}}\right)/d_{\text{ref}}^{\text{T}}, \quad (18)$$

where

$$d_{\text{new}}^{\text{T}}\left(N_{\text{new}}; N_{\text{ref}}, d_{\text{ref}}^{\text{T}}, (\alpha/\beta)_{\text{eff}}^{\text{NT}}\right) = \frac{1}{2}(\alpha/\beta)_{\text{eff}}^{\text{NT}}\left[-1 + \sqrt{1 + 4\frac{N_{\text{ref}}}{N_{\text{new}}}\frac{d_{\text{ref}}^{\text{T}}}{(\alpha/\beta)_{\text{eff}}^{\text{NT}}}\left(1 + \frac{d_{\text{ref}}^{\text{T}}}{(\alpha/\beta)_{\text{eff}}^{\text{NT}}}\right)}\right]$$
$$(19)$$

is obtained from solving (16). Expression (19) is not new; it is a modification of the one that applies to the conventional WIF as given in textbooks, see e.g. Jones and Morgan (2007), with $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ replacing $(\alpha/\beta)^{\text{NT}}$.

Hence, (17) can be written as:

$$\text{gEUBED}_a\left(N_{\text{new}}, f \cdot \mathbf{d}_{\text{ref}}^{\text{NT}}; (\alpha/\beta)^{\text{NT}}\right)) \equiv \text{gEUBED}_a\left(N_{\text{ref}}, \mathbf{d}_{\text{ref}}^{\text{NT}}; (\alpha/\beta)^{\text{NT}}\right)). \quad (20)$$

Equation (20) links $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ to the intrinsic $(\alpha/\beta)^{\text{NT}}$ for given values of N_{ref} and N_{new}. Although we have not been able to find an analytical solution, a numerical solution was obtained by using Brent's root-finding algorithm (Brent 1973). Examples will be given in section 3.2.2.

In the limiting case of N_{new} approaching N_{ref}, it is possible to derive a closed-form expression for $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ (see appendix B):

$$(\alpha/\beta)_{\text{eff}}^{\text{NT}} = \frac{\sum_{i=1}^{V} v_i \left(1 + \frac{d_i^{\text{NT}}}{(\alpha/\beta)^{\text{NT}}}\right)^{a-1} \left(d_i^{\text{NT}}\right)^a}{\sum_{i=1}^{V} v_i \left(1 + \frac{d_i^{\text{NT}}}{(\alpha/\beta)^{\text{NT}}}\right)^{a-1} \left(d_i^{\text{NT}}\right)^{a+1}}(\alpha/\beta)^{\text{NT}}d^{\text{T}}, \quad (21)$$

where d^{T} and d_i^{NT} apply to the reference regimen.

By substituting $a = 1/n$, it can be shown that (21) reduces to (9) in the case of $n = 0$, and to (15) in the case of $n = 1$.

### 2.4. The generalized Withers iso-effect formula (gWIF)

Summarizing the above derivations, we now present the generalized WIF (gWIF) for normal-tissue iso-effect as the conventional WIF of (1) with the intrinsic α/β ratio replaced by the $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ of (15) in the case of $n = 1$, or in the case of arbitrary n, by (21) for N_{new} close to N_{ref}, and by the numerical solution of (20) if N_{new} is not close to N_{ref}:

$$\frac{D_{\text{new}}^{\text{T}}}{D_{\text{ref}}^{\text{T}}} = \frac{(\alpha/\beta)_{\text{eff}}^{\text{NT}} + d_{\text{ref}}^{\text{T}}}{(\alpha/\beta)_{\text{eff}}^{\text{NT}} + d_{\text{new}}^{\text{T}}}. \quad (22)$$

By noting that the left-hand side of (22) equals $N_{\text{new}}/N_{\text{ref}} \cdot d_{\text{new}}^{\text{T}}/d_{\text{ref}}^{\text{T}}$, $d_{\text{new}}^{\text{T}}$ can be obtained as function of N_{new} for a given reference regimen $(N_{\text{ref}}, d_{\text{ref}}^{\text{T}})$ and a given $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ by making use of (19).

3. Numerical evaluation

In this section we provide some examples of the use of the $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ and gWIF concepts for specific normal-tissue dose distributions.
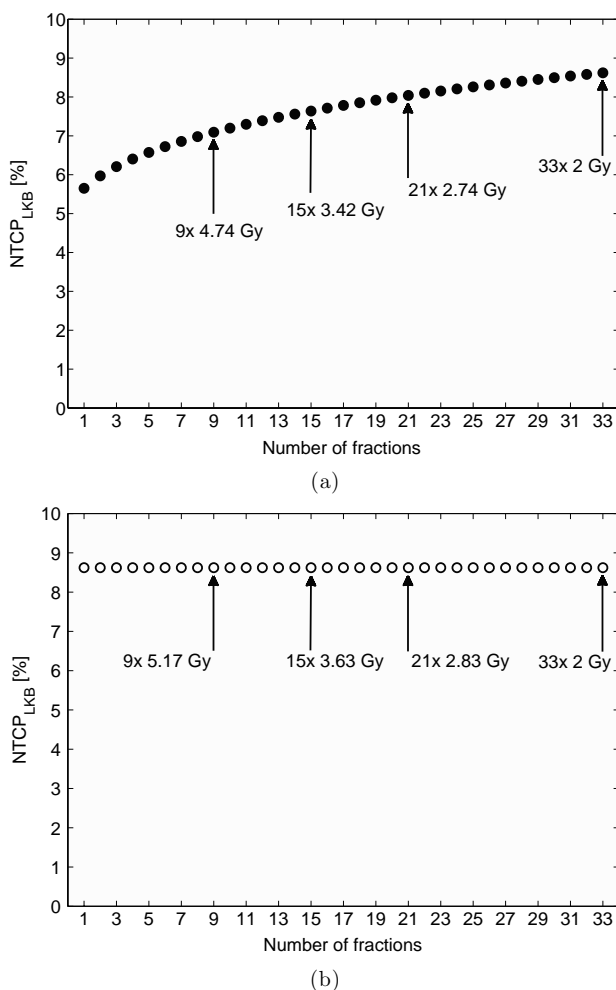
Figure 1. For a parallel normal tissue ($n = 1$) and plan IMRT$_2$ (see table 2) the conventional WIF results in a decreasing NTCP with decreasing fraction number (a), whereas the gWIF results in exact iso-NTCP at any fraction number (b). Arrows indicate tumour dose prescriptions at 33, 21, 15 and 9 fractions.

3.1. Example of gWIF versus WIF

To illustrate the invalidity of WIF and the utility of gWIF, we have used the DVH of a 100% parallel organ (i.e. the paired non-involved lungs) from a given dose distribution at standard fractionation (see IMRT$_2$ plan in subsection 3.2) to compute the NTCP according to the LKB model with parameters ($TD_{50} = 31.4$ Gy, $m = 0.45$, $n = 1$) taken from Marks *et al* (2010). The NTCP of the reference regimen at 33×2.0 Gy is 8.6%.

Figure 1(a) shows the NTCP values that result from the new prescription doses given by the conventional WIF when the number of fractions is reduced relative to the reference
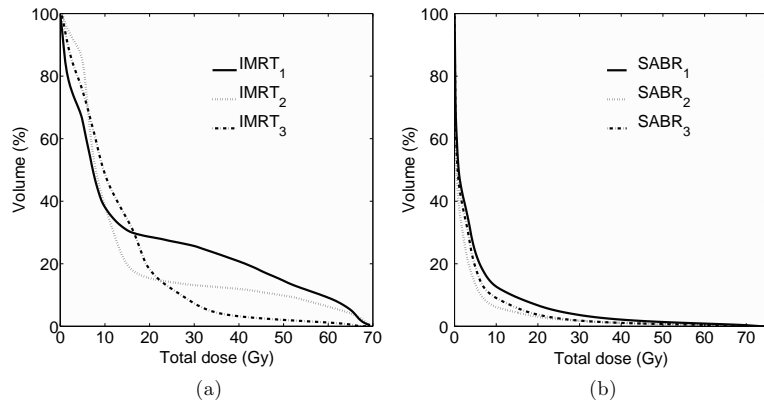
Figure 2. Cumulative dose-volume histograms for uninvolved lung tissue of (a) IMRT and (b) SABR plans having a prescribed dose of 33×2 Gy and 8×7.5 Gy, respectively, for the reference regimen.

Table 1. TNM stage and gross tumour volume characteristics of NSCLC patients.

Plan ID	Stage	TNM	Volume (cm^3)	Location
IMRT$_1$	IIIA	cT2N2M0	57	Left hilus
IMRT$_2$	IIIA	cT4N0M0	330	Left lung
IMRT$_3$	IIIA	cT0N2M0	27	Mediastinum
SABR$_1$	IB	cT2N0M0	6	Left upper lobe
SABR$_2$	IA	cT1N0M0	3	Left lower lobe
SABR$_3$	IA	cT1N0M0	1	Right upper lobe

regimen. Decreasing the number of fractions from 33 to, for example, 21, 15 and 9 fractions, the prescribed tumour fractional doses according to WIF become 2.74, 3.42 and 4.74 Gy, respectively, and the resulting NTCP values are 8.0%, 7.6% and 7.1%, respectively. If the WIF were valid, a constant NTCP of 8.6% would have resulted independently of the number of fractions. This is exactly what is achieved by applying the gWIF (figure 1(b)). Note that the new tumour doses in figure 1(b) are higher than the WIF-derived values in figure 1(a). We have verified the new prescription doses given by gWIF using the BioSuite software for isotoxic dose and fractionation optimization (Uzan and Nahum 2012).

3.2. Patient and treatment plan data

In this subsection we evaluate $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$ for different treatment plans and normal-tissue architecture. Representative DVH data (figure 2) of the uninvolved lung tissue were extracted from six treatment plans for NSCLC patients with different tumour characteristics (table 1): three from IMRT plans with six co-planar 10 MV photon beams using multi-segment fields for step-and-shoot delivery, and three from SABR plans with 10 MV photon beams delivered using volumetric modulated arc therapy with two arcs.

　　　Relevant dosimetric characteristics of the IMRT and SABR plans are summarized in table 2.

Table 2. Dose statistics (in Gy) of IMRT and SABR plans for their respective fractionation regimens; the values of $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ are given for each plan, calculated assuming that the intrinsic $(\alpha/\beta)^{\text{NT}} = 3$ Gy and $n = 1$.

Plan ID	D^{T}	$\bar{D}^{\text{NT}}$	$\sigma_{\mathbf{D}}^{\text{NT}}$	$(\alpha/\beta)_{\text{eff}}^{\text{NT}}$
IMRT$_1$	33×2.0 Gy	18.2	21.6	4.5
IMRT$_2$	33×2.0 Gy	14.9	17.4	5.6
IMRT$_3$	33×2.0 Gy	13.1	11.5	8.5
SABR$_1$	8×7.5 Gy	5.0	10.2	7.0
SABR$_2$	8×7.5 Gy	3.0	7.6	8.1
SABR$_3$	8×7.5 Gy	3.7	7.8	8.9

Abbreviations: D^{T} = prescribed total tumour dose; $\bar{D}^{\text{NT}}$ = mean of normal-tissue total dose; $\sigma_{\mathbf{D}}^{\text{NT}}$ = standard deviation of normal-tissue total dose; $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ = effective α/β ratio of normal tissue.

3.2.1. $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ for normal tissues with 'parallel' architecture. In table 2 the connection between normal-tissue fractionation sensitivity, expressed in terms of our new effective α/β concept, and the degree of normal-tissue sparing is illustrated for the case of $n = 1$. The $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ varies from 4.5 Gy for IMRT$_1$, for which $D^{\text{T}}/\bar{D}^{\text{NT}} = 3.63$ but the normal-tissue dose distribution is very broad ($\sigma_{\mathbf{D}}^{\text{NT}} = 21.6$ Gy), to as high as 8.9 Gy for SABR$_3$ for which $D^{\text{T}}/\bar{D}^{\text{NT}} = 16.22$ with a relatively narrow dose distribution ($\sigma_{\mathbf{D}}^{\text{NT}} = 7.8$ Gy). In the latter case, as the value of 8.9 is very close to 10, the value usually assumed for the tumour, there is almost nothing to be gained from using a large number of fractions.

3.2.2. $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ for normal tissue with arbitrary architecture. In figure 3 we explore the effect on $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ of not only the dose heterogeneity (as in table 2) but also the variation in the value of the volume-effect parameter n and the number of fractions N_{new}. For the three IMRT and three SABR plans, $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ has been calculated for the full spectrum of n values ranging from zero to unity. In this case, the reference and new number of fractions, N_{ref} and N_{new}, respectively, must also be specified. The pair of curves for each plan correspond to the $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ at $N_{\text{new}} = 1$ and to the limiting case of N_{new} approaching N_{ref}.

It can be seen how $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ is close to the intrinsic $(\alpha/\beta)^{\text{NT}}$ value (i.e. 3 Gy) for serial normal-tissue behaviour ($n = 0$), which corresponds to the conditions of validity of the conventional WIF. As n increases $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ increases, with the rate of increase being greater for the more conformal plans. Note that for $n = 0$, the $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ of the IMRT plans in figure 3(a) is 5% below $(\alpha/\beta)^{\text{NT}} = 3$ Gy. This is due to the fact that in these particular treatment plans the normal-tissue maximum dose is 105% of the prescribed tumour dose. The $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ of the SABR plans in figure 3(b) at $n = 0$ is 20% below $(\alpha/\beta)^{\text{NT}} = 3$ Gy, since the maximum normal-tissue dose in these cases is 125% of the (nominal) prescribed tumour dose.

Figure 4 shows how the magnitude of the difference between the pairs of curves in figure 3 depends on the value of N_{new} compared to N_{ref} for $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ evaluated numerically from (20). These differences are clearly largest for $N_{\text{new}} = 1$. As the lower of each pair of curves in figure 3 corresponds to $N_{\text{new}} = 1$, figure 4 demonstrates that these lower curves are lower bounds on the value of $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$.

3.3. Practical use of $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$ for iso-effect calculations

As an example of the practical use of $(\alpha/\beta)_{\text{eff}}^{\text{NT}}$, we give here the prescription for normal-tissue iso-effect when the number of fractions for the IMRT cases is changed from the reference value of 33 to the new value of 20. By applying the conventional WIF, one would have used (19) with the intrinsic $(\alpha/\beta)^{\text{NT}}$ of 3 Gy to obtain a tumour fractional dose for the new regimen
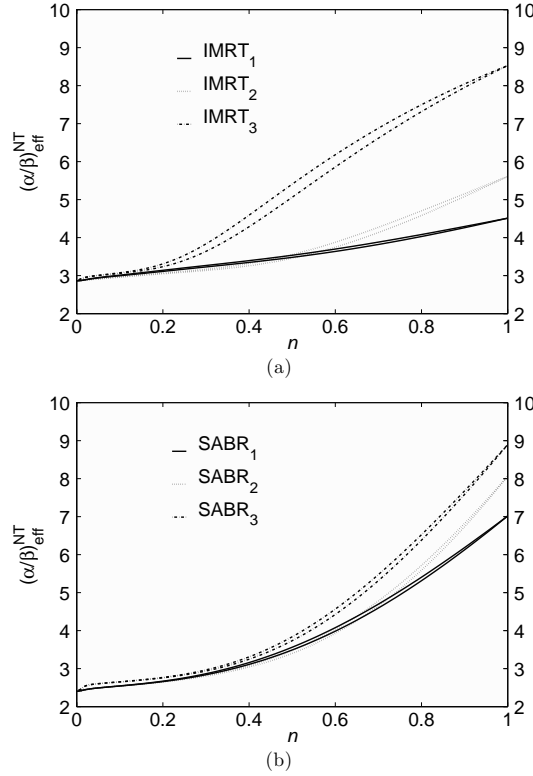
Figure 3. The normal-tissue effective fractionation sensitivity $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$ as function of the volume parameter n for (a) IMRT and (b) SABR plans, assuming an intrinsic $(\alpha/\beta)^{\mathrm{NT}} = 3$ Gy. Lower curves correspond to $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$ at $N_{\mathrm{new}} = 1$ from (20); higher curves correspond to the limiting case of N_{new} approaching N_{ref} from (21).

of 2.83 Gy. If we instead apply gWIF and use (19) with $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}} = 4.5$, 5.6 or 8.5 Gy (see table 2) to account for the normal-tissue dose heterogeneity and volume effect, the tumour fractional doses would now be 2.90, 2.94 and 3.01 Gy, respectively, instead of the constant value of 2.83 Gy which is obtained if the differences between the three treatment plans are ignored. Using $(\alpha/\beta)^{\mathrm{T}} = 10$ Gy for the tumour, this implies increases of 3.0%, 4.8% and 7.9%, respectively, in the 2 Gy equi-effective dose (EQD2$_{10}$) [3] for the tumour relative to the 2.83 Gy per fraction regimen. If additionally the (normalized) gradient of the tumour dose–response relationship were known, the (relative) increase in tumour control probability could be calculated.

3.4. Implications of $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$ for the therapeutic gain

Figure 5 shows how the EQD2$_{10}$ would vary under normal-tissue iso-effect conditions for four different values of $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$; it has been assumed that there is no tumour clonogen proliferation.

---

[3] EQD$X_{\alpha/\beta}$ is the new notation for the equi-effective dose, defined as the biologically equivalent total dose delivered by a reference treatment plan using a fraction size of X Gy (Bentzen *et al* 2012).
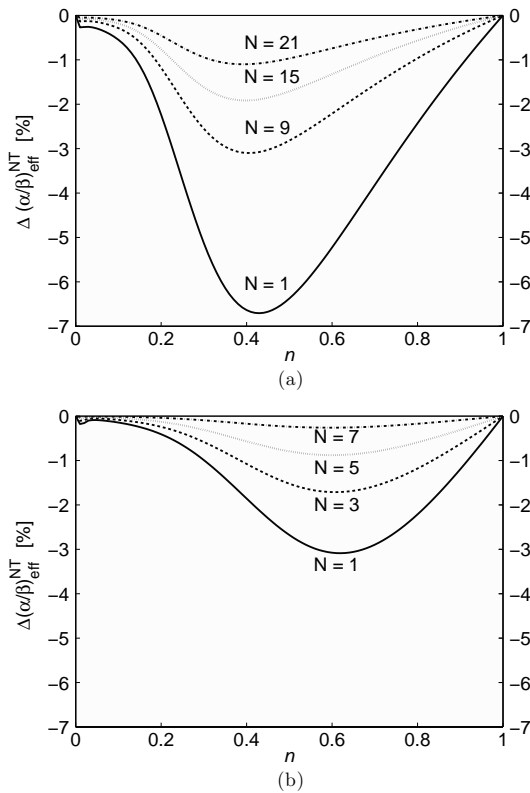
Figure 4. Differences in $(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}$ calculated from (20) relative to (21) for the IMRT$_3$ plan with $N_{\mathrm{ref}} = 33$ and $N_{\mathrm{new}} = N = 1, 9$, 15 or 21 fractions (a), and for the SABR$_3$ plan with $N_{\mathrm{ref}} = 8$ and $N_{\mathrm{new}} = N = 1, 3, 5$ or 7 fractions (b), as a function of volume-effect parameter n; the intrinsic $(\alpha/\beta)^{\mathrm{NT}} = 3$ Gy.

For $(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}} > (\alpha/\beta)^{\mathrm{T}}$, which is possible for an ultra-conformal treatment delivery technique and a parallel organ, the therapeutic gain *increases* with a decreasing number of fractions.

4. Discussion

Radiation oncology practitioners wanting to modify treatment fractionation to ensure normal-tissue iso-effect have hitherto had basically only two tools: either the simple LQ-based WIF or full-blown NTCP modelling. When going from smaller to larger fractions, WIF-derived (tumour) doses are highly conservative, i.e. in many situations they result in decreased complication rates compared to the (small-fraction) reference regimen; this is probably why the use of WIF has continued. A corollary to this is that the use of the WIF to convert from a large-fraction-size regimen which is just within tolerance to a small-fraction-size one could be dangerous in the case of a quasi-parallel normal tissue. An NTCP model-based approach may yield a 'correct' value of the NTCP for a given number of fractions and a given (total) dose distribution, but offers no immediate insight into what would happen if the fractionation
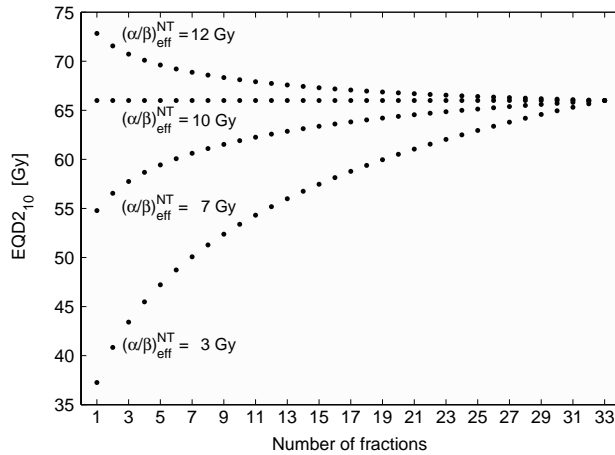
Figure 5. The 2 Gy equi-effective dose for a tumour with $(\alpha/\beta)^{\mathrm{T}} = 10$ Gy as function of the number of fractions under normal-tissue iso-effect conditions ($n = 1$) for a reference regimen of 33×2 Gy. The curves show the effect of different values of $(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}$.

regimen were changed. By contrast, the intrinsic $(\alpha/\beta)^{\mathrm{NT}}$ in the WIF gives users an apparent 'feel' for what is going on. We consider that the new $(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}$ concept will not only allow the user to retain this 'feel' but also extend its validity to any combination of normal-tissue dose distribution and architecture.

The connection between the heterogeneity of a normal-tissue dose distribution and the fractionation sensitivity has not been generally appreciated. The classic experimental iso-effect plot of Thames *et al* (1982), featured in radiobiology textbooks, e.g. Steel (2007), demonstrated that $(\alpha/\beta)^{\mathrm{NT}} \approx 3$ Gy for late-responding normal tissues, but is significantly higher for early reactions. A large body of experimental work supported these findings (Thames and Hendry 1987). The most plausible reason for this is that these early experiments, mostly done on small animals using relatively large fields, corresponded to one of the 'extreme' conditions for which the conventional Withers expression is valid: uniform normal-tissue irradiation. In such situations $(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}$ is equal to the *intrinsic* $(\alpha/\beta)^{\mathrm{NT}}$.

We envisage the new $(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}$ concept being used in several different ways:

 (i) to derive, for exact normal-tissue iso-effect, fractionation regimens differing from the reference one (whether converting from a large to a small number of fractions or vice versa).
 (ii) as a metric for gauging the hypo-fractionation potential of any given treatment plan. This will be particularly attractive for tumours with rapidly proliferating clonogens.
(iii) to quantify the degree of invalidity of the (conventional) WIF in a given situation; this will be given by the difference between the intrinsic and the effective $(\alpha/\beta)^{\mathrm{NT}}$.
(iv) in radiobiological inverse treatment planning with a pre-set number of fractions, maximization of $(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}$ could be added to other objectives. If a high value of $(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}$ is obtained, the optimization might be repeated with a smaller number of fractions to obtain a better overall treatment plan.

Naturally there are uncertainties associated with the numerical values of our new $(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}$ concept. Firstly, this should be viewed in the context of what it replaces, i.e. the WIF with the intrinsic $(\alpha/\beta)^{\mathrm{NT}}$. We have clearly demonstrated that except in the hypothetical case of a

uniform normal-tissue dose equal to the tumour dose, it is only for perfectly 'serial' normal tissues that the use of the WIF is valid. In all other situations, the WIF leads to a reduction of the therapeutic ratio, compared to the reference regimen it replaces, i.e. the use of gWIF will drastically reduce the error inherent in the use of the WIF. Secondly, if a gWIF-derived hypo-fractionation regimen is applied, but n for the normal tissue has been overestimated, then $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$ will be overestimated. This would result in a total tumour dose and fraction size, corresponding to an NTCP slightly greater than that for the reference regimen; the order of magnitude of this over-estimation of $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$ can be judged from figure 3. An additional consideration is the probable over-prediction of cell killing by the LQ expression at large fraction sizes—see below—which will work in the opposite direction, making it highly unlikely that an overestimate of n will result in a significant increase in NTCP in any practical situation.

The WIF is completely based on the LQ model of cell killing and this also applies to the analysis presented here. However, the validity of the LQ model at large doses has been questioned (Kirkpatrick *et al* 2009). At fractional doses above ≈ 10 Gy certain experiments indicate that the LQ expression *over*-predicts cell killing/*under*-predicts cell survival, e.g. Wang *et al* (2010) and the so-called *generalized* LQ model (Carlone *et al* 2005, Wang *et al* 2010) may be more appropriate. If cell killing at large fraction sizes is, at least for some cell types, less than that predicted by the LQ model, then LQ-based radiobiological models will *over*-predict NTCP (Wennberg *et al* 2011). It follows therefore that iso-effective hypo-fractionation regimens based on our expressions for $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$ could *over*-predict complication rates (i.e. be *sub*-iso-effective) but is highly unlikely to *under*-predict them. In other words, hypo-fractionated regimens resulting from the gWIF may in practice be *safer* than the small-fraction reference regimens they were derived from.

More normal-tissue sparing, whether achieved by beam-shaping alone, intensity modulation, stereotaxy or any other modern radiotherapy delivery technique, will result in a higher $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$ and consequently hypo-fractionation may be an option. For spot-scanned proton therapy, which can achieve a high degree of normal-tissue sparing, e.g. Nyström (2010) and Schippers and Lomax (2011), it follows that high $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$ values can be expected for normal tissues with a large volume effect. Proton treatments with a smaller number of fractions would make this modality less expensive per treatment course, with implications for increasing the numbers of patients who could benefit from it.

5. Conclusions

The widely used Withers iso-effect formula, as conventionally applied to late-responding normal tissues is only correct for the 'special' cases of a normal tissue receiving a uniform dose equal to the tumour dose or one with a 100% serial structure where the maximum dose is equal to the tumour dose. This important limitation is removed by using our new effective α/β concept for normal tissues, which makes explicit the relationship between the degree of normal-tissue dose heterogeneity, the volume effect associated with the clinical end-point and the fractionation sensitivity. Using $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$ instead of the intrinsic $(\alpha/\beta)^{\mathrm{NT}}$ in the otherwise unmodified Withers iso-effect formula for calculating a new tumour dose prescription ensures exact normal-tissue iso-effect when changing the fractionation regimen of a given dose distribution. The more conformal the target dose distribution and/or the more 'parallel' the normal-tissue architecture, the higher is $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$ and hence the greater is the potential for hypo-fractionation. The $(\alpha/\beta)^{\mathrm{NT}}_{\mathrm{eff}}$ concept in the well-established Withers iso-effect formula constitutes a user-friendly tool for exploiting the hypo-fractionation potential of modern, highly conformal dose-delivery methods.

Appendix A. Analytical formula for mean biologically effective total dose

The mean biologically effective dose, $\mathrm{BED}_{\mathrm{mean}}$, is calculated from the fractional dose distribution $(N, \mathbf{d})$ and the fractionation sensitivity α/β as a weighted sum over the BED-converted fractional doses d_i delivered over N fractions:

$$\mathrm{BED}_{\mathrm{mean}}(N, \mathbf{d}; \alpha/\beta) = \sum_{i=1}^{V} v_i \mathrm{BED}(N, d_i; \alpha/\beta), \tag{A.1}$$

where v_i is the fractional volume of the organ irradiated to a BED

$$\mathrm{BED}(N, d_i; \alpha/\beta) = N d_i \left(1 + \frac{d_i}{\alpha/\beta}\right), \tag{A.2}$$

and $\sum_{i=1}^{V} v_i = 1$ is the normalized total volume over all V voxels of the whole organ. Substitution of the total dose $D_i = N d_i$ and $\mathrm{BED}(N, d_i; \alpha/\beta)$ from (A.2) into (A.1) and rearrangement of terms yields

$$\mathrm{BED}_{\mathrm{mean}}(N, \mathbf{D}; \alpha/\beta) = \sum_i v_i D_i + \frac{1/N}{\alpha/\beta} \sum_i v_i D_i^2. \tag{A.3}$$

By noting that $\sum_i v_i D_i$ and $\sum_i v_i D_i^2$ are the weighted first and second sample moments of the total dose distribution $\mathbf{D}$, respectively, and its weighted sample variance is defined as $\sigma_{\mathbf{D}}^2 = \sum_i v_i D_i^2 - (\sum_i v_i D_i)^2$, (A.3) can be re-written in the form

$$\mathrm{BED}_{\mathrm{mean}}(N, \mathbf{D}; \alpha/\beta) = \bar{D}\left(1 + \frac{h \cdot \bar{d}}{\alpha/\beta}\right), \tag{A.4}$$

where $\bar{D} = \sum_i v_i D_i$ is the mean total dose, $\bar{d} = \bar{D}/N$ is the mean fractional dose, and $h = 1 + (\sigma_{\mathbf{D}}/\bar{D})^2$ is a dose heterogeneity factor.

Appendix B. Condition for simultaneous tumour and normal-tissue iso-effectiveness at the reference regimen

For iso-effectiveness of the fictitious normal-tissue structure receiving the tumour prescription dose $(N_{\mathrm{ref}}, d_{\mathrm{ref}}^{\mathrm{T}})$ we require a family of fractionation regimens (N, d^{T}) that are iso-effective for an *effective* normal-tissue fractionation sensitivity $(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}$. Hence the surviving fraction $\mathrm{SF}(N, d^{\mathrm{T}})$ should be constant as a function of N. Note that in this case d^{T} is uniquely determined by N, so we can re-write (2) as

$$\mathrm{SF}\left(N; \alpha_{\mathrm{eff}}^{\mathrm{NT}}, \beta_{\mathrm{eff}}^{\mathrm{NT}}\right) = \exp\left[-\alpha_{\mathrm{eff}}^{\mathrm{NT}} N d^{\mathrm{T}}(N)\left(1 + \frac{d^{\mathrm{T}}(N)}{(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}}\right)\right]$$

and solve for $\partial \mathrm{SF}/\partial N = 0$. It is straightforward to show that this gives the differential equation:

$$\frac{\partial d^{\mathrm{T}}}{\partial N} = -\frac{d^{\mathrm{T}}}{N}\left(1 + \frac{d^{\mathrm{T}}}{(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}}\right)\left(1 + \frac{2d^{\mathrm{T}}}{(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}}\right)^{-1}. \tag{B.1}$$

For iso-effectiveness of the normal-tissue dose distribution, we investigate when the generalized tissue damage function gEUBED is a constant function of N, with

$$\mathrm{gEUBED}_a(N, \mathbf{d}^{\mathrm{NT}}; (\alpha/\beta)^{\mathrm{NT}})) = \left\{ \sum_{i=1}^{V} v_i \big[\mathrm{BED}(N, d_i^{\mathrm{NT}}; (\alpha/\beta)^{\mathrm{NT}})\big]^a \right\}^{1/a}, \tag{B.2}$$

and

$$\mathrm{BED}(N, d_i^{\mathrm{NT}}; (\alpha/\beta)^{\mathrm{NT}}) = N d_i^{\mathrm{NT}} \left(1 + \frac{d_i^{\mathrm{NT}}}{(\alpha/\beta)^{\mathrm{NT}}}\right). \tag{B.3}$$

This means setting:

$$\frac{\partial \mathrm{gEUBED}_a}{\partial N} = 0.$$

Substituting the right-hand side of (B.2) in the above, this can be re-written as

$$\frac{1}{a} \big[\mathrm{gEUBED}_a(N, \mathbf{d}^{\mathrm{NT}}; (\alpha/\beta)^{\mathrm{NT}}))\big]^{1-a} \cdot \frac{\partial}{\partial N} \left\{ \sum_{i=1}^{V} v_i \big[\mathrm{BED}(N, d_i^{\mathrm{NT}}; (\alpha/\beta)^{\mathrm{NT}}))\big]^a \right\} = 0,$$

which reduces to

$$\frac{\partial}{\partial N} \left\{ \sum_{i=1}^{V} v_i \big[\mathrm{BED}(N, d_i^{\mathrm{NT}}; (\alpha/\beta)^{\mathrm{NT}}))\big]^a \right\} = 0.$$

This can be re-written as

$$\sum_{i=1}^{V} v_i \big[\mathrm{BED}(N, d_i^{\mathrm{NT}}; (\alpha/\beta)^{\mathrm{NT}}))\big]^{a-1} \cdot \frac{\partial}{\partial N} \left[N d_i^{\mathrm{NT}} \left(1 + \frac{d_i^{\mathrm{NT}}}{(\alpha/\beta)^{\mathrm{NT}}}\right) \right] = 0.$$

By noting that d_i^{NT} is a voxel-dependent fraction λ_i of d^{T}, i.e. $d_i^{\mathrm{NT}}(N) = \lambda_i d^{\mathrm{T}}(N)$, and using (B.1) for iso-effectiveness of the effective normal tissue, this yields

$$\sum_{i=1}^{V} v_i \left(\frac{d^{\mathrm{T}}}{(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}} - \frac{d_i^{\mathrm{NT}}}{(\alpha/\beta)^{\mathrm{NT}}}\right) \left(1 + \frac{d_i^{\mathrm{NT}}}{(\alpha/\beta)^{\mathrm{NT}}}\right)^{a-1} (d_i^{\mathrm{NT}})^a = 0, \tag{B.4}$$

where N as the argument of d^{T} and d_i^{NT} has been omitted for reasons of readability. Solving (B.4) for $(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}}$ and setting $N = N_{\mathrm{ref}}$ it follows that

$$(\alpha/\beta)_{\mathrm{eff}}^{\mathrm{NT}} = \frac{\sum_{i=1}^{V} v_i \left(1 + \frac{d_i^{\mathrm{NT}}}{(\alpha/\beta)^{\mathrm{NT}}}\right)^{a-1} (d_i^{\mathrm{NT}})^a}{\sum_{i=1}^{V} v_i \left(1 + \frac{d_i^{\mathrm{NT}}}{(\alpha/\beta)^{\mathrm{NT}}}\right)^{a-1} (d_i^{\mathrm{NT}})^{a+1}} (\alpha/\beta)^{\mathrm{NT}} d^{\mathrm{T}},$$

where d^{T} and d_i^{NT} apply to the reference regimen.

References

Bentzen S M, Dörr W, Gahbauer R, Howell R W, Joiner M C, Jones B, Jones D T L, van der Kogel A J, Wambersie A and Whitmore G 2012 Bioeffect modeling and equieffective dose concepts in radiation oncology—terminology, quantities and units *Radiother. Oncol.* **105** 266–8

Blomgren H, Lax I, Näslund I and Svanström R 1995 Stereotactic high dose fraction radiation therapy of extracranial tumors using an accelerator. Clinical experience of the first thirty-one patients *Acta Oncol.* **34** 861–70

Borst G R, Ishikawa M, Nijkamp J, Hauptmann M, Shirato H, Onimaru R, van den Heuvel M M, Belderbos J, Lebesque J V and Sonke J J 2009 Radiation pneumonitis in patients treated for malignant pulmonary lesions with hypofractionated radiation therapy *Radiother. Oncol.* **91** 307–13

Brent R P 1973 *Algorithms for Minimization without Derivatives* (Englewood Cliffs, NJ: Prentice-Hall)

Calandrino R, Broggi S, Cattaneo G M, Fiorino C, Nahum A E and Fazio F 2005 Some radiobiological considerations on the potential of IMRT in non-standard fractionation *Radiother. Oncol.* **76** (suppl. 2) S59

Carlone M, Wilkins D and Raaphorst P 2005 The modified linear-quadratic model of Guerrero and Li can be derived from a mechanistic basis and exhibits linear-quadratic-linear behaviour *Phys. Med. Biol.* **50** L9–13; author reply L13–15

Chapman J D 2003 Single-hit mechanism of tumour cell killing by radiation *Int. J. Radiat. Biol.* **79** 71–81

Coutard H 1929 Die Röntgenbehandlung der epithelialen Krebse der Tonsillengegend *Strahlentherapie* **33** 249–52

Fowler J F 1989 The linear-quadratic formula and progress in fractionated radiotherapy *Br. J. Radiol.* **62** 679–94

Fowler J F, Tomé W A, Fenwick J D and Mehta M P 2004 A challenge to traditional radiation oncology *Int. J. Radiat. Oncol. Biol. Phys.* **60** 1241–56

Grau C, Høyer M, Lindegaard J and Overgaard J 2006 The emerging evidence for stereotactic body radiotherapy *Acta Oncol.* **45** 771–4

Hall E J and Giaccia A J 2011 *Radiobiology for the Radiologist* 7th edn (Philadelphia, PA: Williams & Wilkins)

Hoffmann A L, den Hertog D, Siem A Y D, Kaanders J H A M and Huizenga H 2008 Convex reformulation of biologically-based multi-criteria intensity-modulated radiation therapy optimization including fractionation effects *Phys. Med. Biol.* **53** 6345–62

Hoffmann A L, Troost E G C, Huizenga H, Kaanders J H A M and Bussink J 2012 Individualized dose prescription for hypofraction in advanced non-small-cell lung cancer radiotherapy: an in silico trial *Int. J. Radiat. Oncol. Biol. Phys.* **83** 1596–602

Jin J Y, Kong F M, Chetty I J, Ajlouni M, Ryu S, Ten Haken R and Movsas B 2010 Impact of fraction size on lung radiation toxicity: hypofractionation may be beneficial in dose escalation of radiotherapy for lung cancers *Int. J. Radiat. Oncol. Biol. Phys.* **76** 782–8

Joiner M and van der Kogel A (ed) 2009 *Basic Clinical Radiobiology* 4th edn (London: Edward Arnold)

Jones B and Morgan D A L 2007 *Radiobiological Modelling in Radiation Oncology* ed R Dale and B Jones (London: British Institute of Radiology) pp 51–78

Kirkpatrick J P, Brenner D J and Orton C G 2009 Point/Counterpoint. The linear-quadratic model is inappropriate to model high dose per fraction effects in radiosurgery *Med. Phys.* **36** 3381–4

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

Lyman J T 1985 Complication probability as assessed from dose–volume histograms *Radiat. Res. Suppl.* **8** S13–19

Marks L B *et al* 2010 Radiation dose-volume effects in the lung *Int. J. Radiat. Oncol. Biol. Phys.* **76** (suppl. 3) S70–6

Myerson R J 2011 Normal tissue dose conformality measures to guide radiotherapy fractionation decisions *Med. Phys.* **38** 1799–805

Nahum A and Kutcher G 2007 Biological evaluation of treatment plans *Handbook of Radiotherapy Physics—Theory and Practice* ed P Mayles *et al* (London: Taylor and Francis) pp 731–71

Nahum A E and Chapman J D 2003 The interaction between tissue radiobiological properties ('architecture'), dose distribution and optimal fraction size—profound implications for conformal and heavy-particle radiation therapy *Int. J. Radiat. Oncol. Biol. Phys.* **55** 446 (abstract, ICTR 2003)

Niemierko A 1999 A generalized concept of equivalent uniform dose (EUD) *Med. Phys.* **26** 1100 (abstract)

Nyström H 2010 The role of protons in modern and biologically-guided radiotherapy *Acta Oncol.* **49** 1124–31

Park C S, Kim Y, Lee N, Bucci K M, Quivey J M, Verhey L J and Xia P 2005 Method to account for dose fractionation in analysis of IMRT plans: modified equivalent uniform dose *Int. J. Radiat. Oncol. Biol. Phys.* **62** 925–32

Roelofs E *et al* 2012 Results of a multicentric in silico clinical trial (ROCOCO): comparing radiotherapy with photons and protons for non-small cell lung cancer *J. Thorac. Oncol.* **7** 165–76

Romeijn H E, Dempsey J F and Li J G 2004 A unifying framework for multi-criteria fluence map optimization models *Phys. Med. Biol.* **49** 1991–2013

Schippers J M and Lomax A J 2011 Emerging technologies in proton therapy *Acta Oncol.* **50** 838–50

Steel G G 2007 Dose fractionation in radiotherapy *Handbook of Radiotherapy Physics—Theory and Practice* ed P Mayles *et al* (London: Taylor and Francis) pp 163–77

Thames H D and Hendry J H 1987 *Fractionation in Radiotherapy* (London: Taylor and Francis)

Thames H D Jr, Withers H R, Peters L J and Fletcher G H 1982 Changes in early and late radiation responses with altered dose fractionation: implications for dose-survival relationships *Int. J. Radiat. Oncol. Biol. Phys.* **8** 219–26

Timmerman R D 2008 An overview of hypofractionation and introduction to this issue of seminars in radiation oncology *Semin. Radiat. Oncol.* **18** 215–22

Uzan J and Nahum A E 2012 Radiobiologically guided optimisation of the prescription dose and fractionation scheme in radiotherapy using BioSuite *Br. J. Radiol.* **85** 1279–86

Vogelius I S, Westerly D C, Cannon G M and Bentzen S M 2010 Hypofractionation does not increase radiation pneumonitis risk with modern conformal radiation delivery techniques *Acta Oncol.* **49** 1052–7

Wang J Z, Huang Z, Lo S S, Yuh W T C and Mayr N A 2010 A generalized linear-quadratic model for radiosurgery, stereotactic body radiation therapy, and high-dose rate brachytherapy *Sci. Transl. Med.* **2** 39–48

Wennberg B M *et al* 2011 NTCP modelling of lung toxicity after SBRT comparing the universal survival curve and the linear quadratic model for fractionation correction *Acta Oncol.* **50** 518–27

Withers H R, Thames H D Jr and Peters L J 1983 A new isoeffect curve for change in dose per fraction *Radiother. Oncol.* **1** 187–91

V

Efficient generation of Pareto optimal IMRT plans

Derivative-free generation and interpolation of convex Pareto optimal IMRT plans

Aswin L Hoffmann[1]**, Alex Y D Siem**[2]**, Dick den Hertog**[2]**,
Johannes H A M Kaanders**[1] **and Henk Huizenga**[1]

[1] Department of Radiation Oncology, Radboud University Nijmegen Medical Centre,
PO Box 9101, 6500 HB Nijmegen, The Netherlands
[2] Department of Econometrics and Operations Research/Center for Economic Research
(CentER), Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

E-mail: a.hoffmann@rther.umcn.nl

Abstract

In inverse treatment planning for intensity-modulated radiation therapy
(IMRT), beamlet intensity levels in fluence maps of high-energy photon beams
are optimized. Treatment plan evaluation criteria are used as objective functions
to steer the optimization process. Fluence map optimization can be considered
a multi-objective optimization problem, for which a set of Pareto optimal
solutions exists: the Pareto efficient frontier (PEF). In this paper, a constrained
optimization method is pursued to iteratively estimate the PEF up to some
predefined error. We use the property that the PEF is convex for a convex
optimization problem to construct piecewise-linear upper and lower bounds
to approximate the PEF from a small initial set of Pareto optimal plans. A
derivative-free Sandwich algorithm is presented in which these bounds are used
with three strategies to determine the location of the next Pareto optimal solution
such that the uncertainty in the estimated PEF is maximally reduced. We show
that an intelligent initial solution for a new Pareto optimal plan can be obtained
by interpolation of fluence maps from neighbouring Pareto optimal plans. The
method has been applied to a simplified clinical test case using two convex
objective functions to map the trade-off between tumour dose heterogeneity
and critical organ sparing. All three strategies produce representative estimates
of the PEF. The new algorithm is particularly suitable for dynamic generation
of Pareto optimal plans in interactive treatment planning.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

In inverse treatment planning for intensity-modulated radiation therapy (IMRT), fluence maps of a finite number of high-energy photon beams are optimized for delivering a sufficiently high dose to the tumour volume and an adequate sparing of surrounding healthy tissue structures (Webb 2001). To guide the search process, a fluence map optimization (FMO) model based on a set of conflicting treatment plan evaluation criteria is generally used (Reemtsen and Alber 2004). Traditionally, the model has been formulated as a minimization of a composite objective function being the weighted sum of constituent criteria (Bortfeld 1995, Brahme 1995). However, assigning weights or penalty factors to the constituent criteria prior to optimization imposes an *a priori* trade-off between the conflicting criteria. Unfortunately, there exists no intelligent basis for an *a priori* quantification of the trade-off without a proper understanding of which criteria are competing and non-competing. Therefore, the weighting factors must be determined by a trial-and-error process that involves multiple optimization problems to be solved repeatedly. This is often done in a human iteration loop, where the weights are altered in case the solution does not satisfy the clinical goals (Hunt *et al* 2002). This deteriorates the planning efficiency and does not allow for interactive treatment planning. Nowadays, the weighting factors are often determined empirically and are based on advancing clinical experience. However, the weighting factors have no direct relation to the clinical characteristics of the treatment plan. Furthermore, since the sensitivity of the optimization result to changes in the weighting factors is unknown beforehand, preferably these weighting factors should be avoided. To circumvent these problems, it was proposed to decouple the optimization and decision-making process and formulate the FMO model as a multi-objective optimization problem (Hamacher and Küfer 2002, Küfer *et al* 2003, Thieke 2003). In this approach, the weights have been eliminated by considering the optimization task as a simultaneous minimization over the set of conflicting objectives. Multiple solutions to such a problem exist, each of which represents an ultimate compromise between the objectives.

In this paper, we restrict ourselves to the so-called Pareto optimal (PO) or efficient solutions (Pareto 1906). These solutions represent the best attainable compromises and have the property that improving the value of a single objective cannot be accomplished without worsening at least one other objective value. Treatment plans that possess this property are referred to as PO plans. Solving the multi-objective optimization problem in inverse treatment planning for IMRT thus entails characterizing its set of PO treatment plans, which in the objective space is represented by the Pareto efficient frontier (PEF). The generation of the PO plans does not require user interaction and a database of plans can be precomputed off-line. Once the PEF has been generated, the decision maker (i.e., usually the physician) can select a single PO plan according to the clinical characteristics of the treatment plan and the patient's risk-taking preferences. In this way, an *a posteriori* trade-off is made between competing treatment plan evaluation criteria.

Clinical rationale. In radiation oncology, consensus guidelines for treatment of specific patient groups are laid down in local or national treatment protocols, specifying preferred dose levels for the tumour volume and dose limiting constraints for the organs at risk (e.g., the RTOG protocols). IMRT has introduced a large number of degrees of freedom to shape the dose distribution and thereby has increased the possibilities to make physician/patient specific trade-offs between target coverage and the probability of severe side effects. Physicians may rate differently the various aspects involved in the trade-off and usually include other information in balancing the treatment risks and benefit, such as the patient's condition, age, social circumstances, type of complications expected, possible salvage treatment of complications and patient preference (Amols *et al* 1997, van Tol-Geerdink *et al* 2006). The

introduction of IMRT will emphasize these differences and calls for tools to assist in balancing patient- and physician-specific trade-offs.

Among the different methods to solve multi-objective optimization problems, Pareto optimization is the only approach that actually shows the trade-offs and allows the decision maker to select an 'optimal' plan from the precomputed set of best compromises. Typically, inverse planning of IMRT involves multiple dosimetric criteria, either formulated as objectives or constraints for the tumour volume and organs at risk. These manifold objectives define a high-dimensional PEF. It is our experience that in practice the clinically relevant trade-offs are often limited to two or three objectives at most. Typical examples of bi-objective problems encountered in IMRT are: (1) sparing of the parotid gland(s) versus coverage of the elective treatment volume in head-and-neck radiotherapy and (2) sparing of the rectal wall versus dorsal treatment margin in prostate radiotherapy. The other optimization criteria are in fact considered as constraints (although sometimes implemented as objectives) to be able to focus on the most relevant trade-offs. This is also necessary since in a (clinical) decision-making process it is not conceivable to successfully weigh more than three or ultimately four variables (Graeme *et al* 2005). Treating all optimization criteria as objectives would make the treatment plan selection process intractable. Therefore, we start off with a bi-objective problem and elaborate on how to possibly extend to higher dimensions.

Computational motivation. It can be shown that the PEF is convex when only convex treatment plan evaluation criteria are considered (Romeijn *et al* 2004). The same authors also proved convexity for a number of commonly used treatment plan evaluation criteria and showed that under certain conditions non-convex criteria can be transformed into equivalent convex criteria. In this case, identification of PO treatment plans relies on solving a family of convex FMO models to global optimality. However, this is a computationally intensive task since multiple optimization runs have to be performed repeatedly, each starting from scratch. Hence, it is desirable to have a method that is able to generate a representative set of PO plans in as few as possible optimization runs. To achieve this goal, we incorporate *a priori* knowledge about the convexity of the PEF into a new method that approximates the PEF by piecewise-linear bounds and that enables to dynamically generate an initial solution for a new PO plan. As we show, the initial fluence maps for this new PO plan can be obtained by interpolation of the fluence maps of neighbouring PO plans.

Structure of the paper. It is the aim of the present paper to provide an efficient method to dynamically generate a representative set of PO treatment plans using convexity bounds, starting from a small number (e.g., three or four) of PO plans that have already been generated. This paper is organized as follows. In section 2, first a formal introduction to the fluence map optimization problem is given, and a commonly applied method for single-objective optimization is described. Secondly, the concept of multi-objective optimization is presented and the so-called ε-constraint method to generate Pareto optimal solutions is discussed. In section 3, the new method to approximate the PEF for (nonlinear) convex bi-objective optimization problems is proposed. Upper and lower bounds for the PEF are constructed based on the property that the PEF is convex for convex objective functions (Siem *et al* 2006a). We show that by iteratively minimizing a cost function, representing uncertainty between the estimated and the true PEF, an improvingly better piecewise-linear approximation of the PEF can be obtained dynamically. For each new PO plan to be generated, we demonstrate that a good initial solution can be obtained by fluence map interpolation, which significantly speeds up the optimization process. In section 4, we show proof of principle on a simplified clinical case using a patient with head-and-neck cancer. The discussion and conclusions follow in section 5.

2. Fluence map optimization problem

2.1. Definition and notation

It is supposed that the patient's body is discretized into a number of volume elements, denoted as *voxels*. Let m be the number of voxels, and let d_i, $i = 1, \ldots, m$, denote the dose deposited in voxel i. The dose d_i for a certain particle fluence is given as the weighted sum of the so-called *pencil beams* (Gustafsson *et al* 1994). The beam element related to a specific unit pencil beam is denoted as a *bixel*. A *fluence map* of a beam is the union of the weights of the bixels for the specific beam. A total number of n bixel weights $w_j \geqslant 0$, $j = 1, \ldots, n$, are the design variables (also called the optimization variables) in the FMO problem under consideration. The beam and tissue interaction is described by an *influence matrix*, $\mathbf{P}$, that relates the bixel weights to the dose in the voxels. Each bixel has its own pencil beam that is dependent on the source-to-surface distance, angle of incidence and the patient's anatomy-dependent electron density characteristics. The dose distribution $\mathbf{d}(\mathbf{w})$ is related to the influence matrix $\mathbf{P}$ and the bixel weights $\mathbf{w}$ through a linear relation

$$\mathbf{d}(\mathbf{w}) = \mathbf{P}\mathbf{w}.$$

In tensor notation, the dose in voxel i can be denoted as a weighted sum over all bixel weights:

$$d_i(\mathbf{w}) = \sum_{j=1}^{n} P_{ij} w_j.$$

2.2. Single-objective optimization

To steer the optimization algorithm, a single *composite objective function* $F : \mathbb{R}_+^m \mapsto \mathbb{R}_+$ is used to quantify the plan evaluation score as a function of the dose distribution $\mathbf{d}(\mathbf{w})$. Examples of commonly used objective functions for inverse treatment planning are discussed in section 3.1. It is the task of the optimization algorithm to find a set of bixel weights that minimize $F(\mathbf{d}(\mathbf{w}))$:

$$\begin{aligned} \min_{\mathbf{w}} \quad & F(\mathbf{d}(\mathbf{w})) \\ \text{s.t.} \quad & \mathbf{w} \geqslant 0. \end{aligned} \tag{1}$$

The composite objective function usually is a convex combination of constituent objective functions $F_k : \mathbb{R}_+^m \mapsto \mathbb{R}_+, k = 1, \ldots, l$, where l is the number of objective functions (typically, at least one objective per organ) and the kth objective is assigned an objective importance weight $\lambda_k \geqslant 0$ such that $\sum_{k=1}^{l} \lambda_k = 1$:

$$F(\mathbf{d}(\mathbf{w})) := \sum_{k=1}^{l} \lambda_k F_k(\mathbf{d}(\mathbf{w})). \tag{2}$$

Since in general the objectives F_k are conflicting, and a solution that minimizes all the objective functions simultaneously does not exist, the weights λ_k, $k = 1, \ldots, l$, are introduced to make an *a priori* trade-off between the separate objectives. As already mentioned above, there does not exist an intelligent basis for making the trade-off between the objectives prior to optimization. For example, in clinical inverse treatment planning practice it is often the case to balance homogeneity of the dose distribution in the target volume against the mean or maximum dose to an organ at risk (OAR). Unfortunately, the interrelationship between these contradictory goals is unknown beforehand. A proper balancing of the objectives prior to optimization is therefore impracticable.

2.3. Multi-objective optimization and Pareto optimality

Multi-objective optimization decouples the optimization and decision-making process by first analysing all feasible candidate solutions and subsequently presents the trade-offs between them to a human decision maker. This allows the decision maker to articulate individual preference relations between alternative solutions and to select an optimal solution. Such an approach has been widely applied to solve engineering design problems where cost–quality trade-offs need to be made between multiple conflicting and possibly incommensurable (e.g., having different units) criteria. For a general discussion of this approach see the textbook by Miettinen (1999) and the survey paper by Marler and Arora (2004). From this perspective, the problem to design and select an optimal IMRT treatment plan that is tuned to the physician's predilections has been considered suitable to be solved with the multi-objective approach. Initial work from Küfer and co-workers has demonstrated the applicability of this concept for optimization of inverse treatment planning in IMRT (Hamacher and Küfer 2002, Küfer *et al* 2003, Thieke 2003).

In solving the FMO problem with the multi-objective optimization approach, a vector $\mathbf{F}$ of all l objective functions $F_k : \mathbb{R}_+^m \mapsto \mathbb{R}_+, k = 1, \ldots, l$, is to be minimized simultaneously (Miettinen 1999, p 5):

$$\min_{\mathbf{w}} \quad \mathbf{F}(\mathbf{d}(\mathbf{w})) = \begin{pmatrix} F_1(\mathbf{d}(\mathbf{w})) \\ F_2(\mathbf{d}(\mathbf{w})) \\ \vdots \\ F_l(\mathbf{d}(\mathbf{w})) \end{pmatrix} \tag{3}$$

$$\text{s.t.} \quad \mathbf{w} \geqslant 0.$$

Because of the contradiction and possible incommensurability of the objective functions, a single solution that would be optimal (i.e., minimal) for all the objectives simultaneously does not exist in general. Instead, multiple solutions exist, and therefore a criterion to define optimality in the multi-objective context is required. Here, we restrict to those solutions that have the property that no single-objective value can be improved without deteriorating at least one other objective value. Solutions that comply with this definition are called *Pareto optimal* (also called Edgeworth–Pareto optimal, efficient, nondominated or noninferior; see Pareto (1906)). A more formal definition of Pareto optimality is the following:

Definition 2.1. *An optimization (variable) vector* $\mathbf{x}^* \in \mathcal{X}$ *(in our case,* $\mathbf{x}^* = \mathbf{d}(\mathbf{w}^*)$*) is Pareto optimal (PO) for problem (3) if there does not exist another optimization vector* $\mathbf{x} \in \mathcal{X}$ *such that* $F_i(\mathbf{x}) \leqslant F_i(\mathbf{x}^*)$ *for all* $i = 1, \ldots, l$ *and* $F_j(\mathbf{x}) < F_j(\mathbf{x}^*)$ *for at least one index* j.

Here, $\mathcal{X}$ denotes the feasible design space (often called the feasible decision space), which is a subset of the optimization variable space $\mathbb{X}$, defined as

$$\mathcal{X} := \{\mathbf{x} \in \mathbb{X} \mid G_i(\mathbf{x}) \leqslant 0 \quad (i = 1, \ldots, q)\},$$

where $G_i(\mathbf{x})$ denotes a constraint function. The treatment plan resulting from the PO bixel vector $\mathbf{w}^*$ is called the *Pareto optimal treatment plan*. Usually, an infinite number of PO solutions exist. The set of PO solutions is commonly referred to as the *Pareto optimal set* and is denoted by $\mathcal{X}^*$. The image of the feasible design space, $\mathbf{F}(\mathcal{X})$, defines the feasible objective space $\mathcal{Z}$, which is a subset of the objective space $\mathbb{Z}$ (also called the feasible criterion space). In the objective space, the image of the Pareto optimal set, $\mathbf{F}(\mathcal{X}^*)$, is mapped to the so-called *Pareto efficient frontier* (PEF), denoted by $\mathcal{Z}^*$. It is evident that all Pareto optimal solutions lie on the lower boundary of the feasible objective space $\mathcal{Z}$ (figure 1). A Pareto optimal treatment
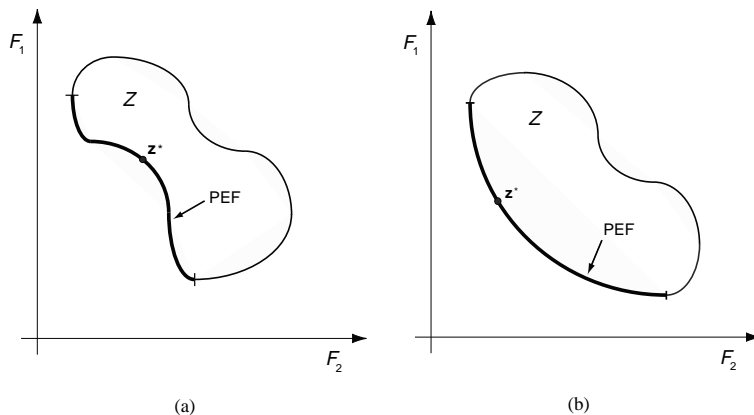
Figure 1. The sets $\mathcal{Z}$ and $\mathcal{Z}^*$ represent a feasible objective region with a non-convex (a) and a convex (b) Pareto efficient frontier (PEF), respectively, for a bi-objective problem. A Pareto optimal objective vector $\mathbf{z}^*$ is on the boundary $\mathcal{Z}^*$.

plan that is characterized by its objective vector, $\mathbf{z}^*$, is on this boundary. Analogous to the formal definition of Pareto optimality in the optimization variable space, Pareto optimality has also been defined formally in the objective space.

Definition 2.2. *An objective vector* $\mathbf{z}^* \in \mathcal{Z}$ *is Pareto optimal (PO) if there does not exist another objective vector* $\mathbf{z} \in \mathcal{Z}$ *such that* $z_i \leqslant z_i^*$ *for all* $i = 1, \ldots, l$ *and* $z_j < z_j^*$ *for at least one index* j.

Equivalently, if $\mathbf{z}^*$ is Pareto optimal, then the associated solution, $\mathbf{x}^*$, for which $\mathbf{z}^* = \mathbf{F}(\mathbf{x}^*)$ is Pareto optimal.

The purpose of multi-objective optimization is to determine those solutions corresponding to the PEF. Because the objective space is usually of a lower dimension than the optimization variable space, the PEF can be used by the decision maker to navigate efficiently through the PO solutions and select an optimal compromise. Moreover, the physician will only use the values of the treatment plan evaluation criteria in the decision-making process to select a final solution and will not be interested in the (design) values of the underlying bixel weights.

2.4. Strategies to find Pareto optimal solutions

In a first attempt to apply a multi-objective approach for the generation of Pareto optimal IMRT plans, Thieke (2003) used a simple 'brute force' strategy to generate feasible candidate plans and subsequently tested the plans for Pareto optimality. As a clinical example, a case with one single target and two organs at risk was used. For both OARs, the two constraint values were varied stepwise over fixed intervals. For all combinations of constraint values the feasibility was determined, resulting in a grid of feasible plans. From this set, the PO plans at the PEF were identified and separated from the dominated plans in the interior of the feasible objective region. The large computational burden to generate plans without the guarantee of being Pareto optimal and the fact that the majority of the plans generated are rejected make this strategy highly inefficient for use in a clinical environment. Furthermore, it is evident that such an ad hoc method does not suffice for cases with multiple targets and more

than two OARs. Therefore, it was recommended by Thieke (2003) to implement and apply more sophisticated PEF-generating methods keeping the computation time within clinically acceptable limits.

From the field of operations research a variety of methods to generate Pareto optimal solutions for multi-objective optimization and decision-making problems are available (Marler and Arora 2004). In contrast to the generate-and-test method exploited by Thieke (2003), nonlinear programming tools capable of generating a Pareto optimal solution for every single optimization run should be applied to solve the multi-objective FMO problem. These tools solve a multi-objective optimization problem by *scalarization*, where it is converted into a single or a series of single-objective optimization problems. The optimal solutions of multi-objective optimization problems can be characterized as solutions of certain single-objective optimization problems. The two most frequently used deterministic generation methods where only *a priori* articulation of preference information is used are the so-called weighted-sum (WS) method and the ε-constraint (EC) method (Miettinen 1999).

In the weighted-sum method, each objective function, F_k, is associated with a weighting factor, λ_k, and the weighted sum of the objectives is minimized. Under the assumption that the constraint set $\mathcal{X}$ is convex, the objective functions $F_k(\mathbf{x})$ are convex and the weighting factors $\lambda_k \geqslant 0$, a Pareto optimal solution is obtained by solving the *weighting problem* of (1), (2):

$$\min_{\mathbf{x}} \quad \sum_{k=1}^{l} \lambda_k F_k(\mathbf{x}) \tag{4}$$
$$\text{s.t.} \quad \mathbf{x} \in \mathcal{X}.$$

A WS-method-based Sandwich algorithm that uses derivative information to approximate the convex Pareto efficient frontier in three- and four-dimensional IMRT objective space has recently been proposed by Craft *et al* (2006).

In this paper, we apply the EC method in combination with a derivative-free Sandwich algorithm to compute well-distributed points on the convex Pareto efficient frontier. In the ε-constraint method, one of the objective functions is selected to be minimized while all the others are converted into constraints by setting an upper bound on each of them (Haimes *et al* 1971). The single-objective optimization problem to be solved is now of the form

$$\min_{\mathbf{x}} \quad F_k(\mathbf{x}) \tag{5}$$
$$\text{s.t.} \quad F_j(\mathbf{x}) \leqslant \varepsilon_j \qquad \text{for all} \quad j = 1, \ldots, l, \quad j \neq k, \quad \mathbf{x} \in \mathcal{X},$$

where $k \in \{1, \ldots, l\}$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_{k-1}, \varepsilon_{k+1}, \ldots, \varepsilon_l)^T$ is a vector of fixed constraint values.

As can be proven, the ε-constraint method finds every Pareto optimal solution of any multi-objective optimization problem, regardless of the convexity of the problem (Miettinen 1999, theorem 3.2.2., p 85). To ensure that a solution is Pareto optimal, either l different optimization problems have to be solved or a unique solution has to be obtained. In general, the uniqueness is not easy to verify. However, if the problem is convex and the objective function to be minimized is strictly convex, a unique solution is guaranteed (Chankong and Haimes 1983a, p 131).

To generate a Pareto optimal solution, the ε-constraint method is used as an *a priori* method, where the decision maker specifies the single-objective function to be minimized together with the upper bounds on the other objectives. Systematic ways to perturb the upper bounds in order to obtain a representative set of Pareto optimal solutions have been suggested in Chankong and Haimes (1983b). These authors propose an ad hoc strategy to solve multiple ε-constraint problems on a preselected grid of constraint values ε and attempt

curve fitting to find a PO solution as a function of ε. Unfortunately, their method does not provide information on the accuracy of the approximation.

In the next section, we demonstrate how the ε-constraint method can be applied as part of our algorithm to dynamically generate a representative set of Pareto optimal solutions.

3. Approximation of the Pareto efficient frontier by convexity bounds

3.1. Convex objective functions

Evaluation criteria that are commonly used to measure treatment plan quality can be classified into physical (i.e., dose-based) and radiobiological (i.e., response-based) score functions (Brahme 1995). The first class typically measures deviations from a prescribed dose level in an organ, while the second class attempts to model the radiobiological effect of irradiating the organ. In general, for a treatment plan determined by a bixel weight vector $\mathbf{w}$, the kth evaluation function F_k applied to an organ structure V maps the corresponding dose distribution $\mathbf{d}(\mathbf{w})$ to a positive real value $F_k(\mathbf{d}(\mathbf{w}))$. Typically, a structure-based treatment plan evaluation criterion is averaged over the doses to the individual voxels in V and can be expressed as

$$F_k(\mathbf{d}(\mathbf{w})) = \sum_{i \in V} \mu_i f_k(d_i(\mathbf{w})),$$

where f_k is a voxel-based evaluation function that is convex in d_i and μ_i is a voxel-dependent scaling factor. Often, the scaling factor is chosen to be the voxel volume relative to the total volume of the organ.

Several authors have analysed the convexity properties of commonly used treatment plan evaluation functions with respect to the dose to a voxel, d_i (Bortfeld 1995, Deasy 1997, Llacer *et al* 2003, Romeijn *et al* 2004). It was reported that most of the criteria proposed are convex functions, or can be transformed into equivalent convex functions, except for dose–volume-based criteria. The latter type have proven to be non-convex (Deasy 1997) and will therefore be left out of consideration in the present paper.

For the sake of completeness, it should be mentioned that since the dose to a voxel d_i is a linear combination of the bixel weights w_j, the convexity properties of the objective functions also hold for the bixel weights that are to be optimized. This is due to the property that a convex function remains convex after an affine mapping.

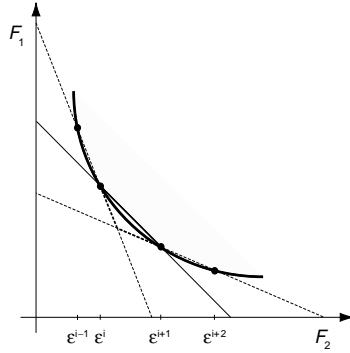### 3.2. Upper and lower bounds for the Pareto efficient frontier

In this paragraph we show that in case the PEF is convex, and a number of PO solutions is available, piecewise-linear upper and lower bounds for the PEF can be derived. These bounds provide a local approximation of the true PEF and provide a measure of uncertainty.

As indicated in section 2.4, a point on the PEF can be found by solving the ε-constraint problem (5). Solving this problem yields a unique PO solution, provided that the objective function F_k to be minimized is strictly convex, the constraint functions F_j are convex for all $j \neq k$, and the feasible design region $\mathcal{X}$ is convex (Chankong and Haimes 1983a, p 131).

In the present paper, we only consider a convex bi-objective optimization problem and formulate it as an ε-constraint problem:

$$p(\varepsilon) := \min_{\mathbf{x}} \quad F_1(\mathbf{x})$$
$$\text{s.t.} \quad F_2(\mathbf{x}) \leqslant \varepsilon \quad \mathbf{x} \in \mathcal{X}, \tag{6}$$

Figure 2. Upper (solid line) and lower (dashed lines) bounds for the Pareto efficient frontier on the interval $[\varepsilon^i, \varepsilon^{i+1}]$ when only function evaluation information in the Pareto optimal points is available.

where $\mathcal{X}$ is a convex set, and F_1 and F_2 are strictly convex and convex functions, respectively. Then, the PEF associated with problem (6) is a univariate convex function $p : \mathcal{E} \mapsto \mathbb{R}_+$, for which $p(\varepsilon) = F_1(\mathbf{x}^*(\varepsilon))$ and $\mathbf{x}^*(\varepsilon)$ is the solution of (6) with $\varepsilon \in \mathcal{E}$ and

$$\mathcal{E} := \{\varepsilon \in \mathbb{R}_+ : \exists \mathbf{x} \in \mathcal{X} : F_2(\mathbf{x}) \leqslant \varepsilon\},$$

the set of feasible constraint values. The complete proof that the PEF is a decreasing convex function is presented in a related paper by Siem *et al* (2006a).

Now suppose that n PO solutions $p(\varepsilon^1), \ldots, p(\varepsilon^n)$ are given for the points $\varepsilon^1 < \cdots < \varepsilon^n$ that partition the interval $[\varepsilon^1, \varepsilon^n] \in \mathcal{E}$. Furthermore, suppose that a PO solution is required for an intermediate constraint value ε on the subinterval $[\varepsilon^i, \varepsilon^{i+1}]$, with $2 \leqslant i \leqslant n - 2$. Then, it can be shown that the line segment connecting the points $(\varepsilon^i, p(\varepsilon^i))$ and $(\varepsilon^{i+1}, p(\varepsilon^{i+1}))$, for $1 \leqslant i \leqslant n-1$, is an upper bound for $p(\varepsilon)$ on this subinterval. The proof is presented in (Siem *et al* 2006a). Construction of the upper bound is schematically illustrated in figure 2.

For the derivation of the lower bounds, a classification of methods according to the availability of derivative information is adopted. In the next two paragraphs, first a derivative-free approximation method is presented for the case no derivative information in the PO points is available. Secondly, a derivative-based method is described. We prove that the derivative-based approximation method provides tighter lower bounds than the derivative-free method.

3.2.1. Derivative-free approximation methods.   These methods only use function evaluation information to estimate the lower bounds.   For IMRT optimization, derivative-free approximation methods may be preferred over derivative-based approximation methods, as the solvers commonly used in clinical inverse treatment planning systems usually do not provide derivative information or may produce derivative information that is numerically unstable.

Similar to the derivation of the upper bound, it can be shown that both the line segment connecting the points $(\varepsilon^{i-1}, p(\varepsilon^{i-1}))$ and $(\varepsilon^i, p(\varepsilon^i))$, for $2 \leqslant i \leqslant n-1$, and the line segment connecting $(\varepsilon^{i+1}, p(\varepsilon^{i+1}))$ and $(\varepsilon^{i+2}, p(\varepsilon^{i+2}))$, for $1 \leqslant i \leqslant n-2$, are lower bounds for $p(\varepsilon)$ on the interval $[\varepsilon^i, \varepsilon^{i+1}]$. The complete proof for the lower bounds is also presented in (Siem *et al* 2006a). In figure 2, the bounds are depicted for a typical univariate convex PEF.
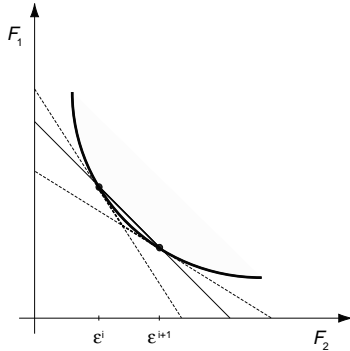
Figure 3. Upper (solid line) and lower (dashed lines) bounds for the Pareto efficient frontier on the interval $[\varepsilon^i, \varepsilon^{i+1}]$ when derivative information in the Pareto optimal points is available.

3.2.2. Derivative-based approximation methods. In addition to function evaluation information, these methods use derivative information to estimate the lower bounds. The WS method is preferably used, as the weighting factors define a supporting hyperplane, which by definition provides the derivative information (Boyd and Vandenberghe 2004, Craft *et al* 2006). In case the EC method is used, Lagrange multipliers may provide the derivative information. For the latter case, we refer to Alber *et al* (2002), who applied sensitivity information in the form of Lagrange multipliers to balance the gradients of the constraints and the objective function. Their figure 2 even shows the subdifferentials in the points on the PEF which is shown in their figure 1.

Now suppose that not only n PO solutions $p(\varepsilon^1), \ldots, p(\varepsilon^n)$ are given for the points $\varepsilon^1 < \cdots < \varepsilon^n$, but that also the subdifferentials $s(\varepsilon^1), \ldots, s(\varepsilon^n)$ in these points are given. We use subdifferentials instead of derivatives because $p(\varepsilon)$ may be non-differentiable. Let $\varepsilon \in [\varepsilon^i, \varepsilon^{i+1}]$. Then, the straight line tangent to the point $(\varepsilon^i, p(\varepsilon^i))$ is a lower bound for $p(\varepsilon)$, that is,

$$p(\varepsilon) \geqslant p(\varepsilon^i) + s(\varepsilon^i)(\varepsilon - \varepsilon^i). \tag{7}$$

Within the subinterval $[\varepsilon^i, \varepsilon^{i+1}]$ the line segments of the tangent lines through the points $(\varepsilon^i, p(\varepsilon^i))$ and $(\varepsilon^{i+1}, p(\varepsilon^{i+1}))$ define a lower bound for $p(\varepsilon)$ on this subinterval. This is schematically depicted in figure 3. In (Siem *et al* 2006a), it is shown that these lower bounds are tighter than the lower bounds derived from function evaluation information only.

Because Lagrange multipliers could not be accessed from our clinical treatment planning system, we did not use derivative information and concentrated on the derivative-free approximation of the Pareto efficient frontier instead.

3.3. Iterative strategies to generate new Pareto solutions

In the literature, the so-called Sandwich algorithms have been proposed for iteratively approximating univariate convex functions (Fruhwirth *et al* 1989, Burkard *et al* 1991, Rote 1992, Yang and Goh 1997). In these methods, upper and lower bounds are constructed making use of the convexity of the function to be approximated. However, the methods of Fruhwirth *et al* (1989) and Rote (1992) rely on the availability of derivative information, which may not be available when using the ε-constraint method to solve a multi-objective optimization

problem. Either the information is not available or not accurate due to numerical errors. In constraint-based routine inverse treatment planning optimization such an approach would therefore be impracticable. In Yang and Goh (1997), a derivative-free optimization problem has to be solved in case no derivative information is available. This usually costs many function evaluations, which may be time consuming.

In the present paper, we present a new Sandwich-type algorithm to approximate a univariate convex Pareto curve up to some predescribed error δ using function evaluation information only. The construction of Pareto curves is usually time consuming since multiple optimization problems have to be solved that each may be large and difficult to solve. Our method provides an efficient strategy to approximate the Pareto curve and reduce the time required to solve the optimization problems by use of a new initial solution.

3.3.1. Sandwich algorithms. Sandwich algorithms generate upper and lower bounds iteratively, starting from a small number of data points $(x^1, y(x^1)), \ldots, (x^n, y(x^n))$ with $x^1 < \cdots < x^n$ that define a set of intervals, $I = \{[x^1, x^2], [x^2, x^3], \ldots, [x^{n-1}, x^n]\}$. Now, let $J \subseteq I$ denote the set of intervals for which the error $\delta_j > \delta$, with $j \in J$. Different kinds of error measures can be used, and some examples are presented below. An interval $[a, b]$ from J is selected and partitioned according to some partitioning rule, resulting in two subintervals $[a, c]$ and $[c, b]$. The function evaluation $y(c)$ and, if possible, the derivative $y'(c)$ are calculated. Whenever the error of any of the two subintervals is greater than δ, this subinterval is added to J. The procedure is repeated until all intervals have an error smaller than δ (i.e., until $J = \emptyset$).

Different error measures and partition rules have been proposed in the literature. The most commonly used error measures are (Rote 1992) as follows:

  (i) *maximum error* on interval: $\delta_{[a,b]} := \max\limits_{x \in [a,b]} \{u(x) - l(x)\}$,

 (ii) *area* enclosed by upper and lower bounds on interval: $\delta_{[a,b]} := \int_{[a,b]} [u(x) - l(x)] \, dx$,

(iii) *Hausdorff distance* on interval: $\delta_{[a,b]} := \max \big\{ \sup\limits_{v \in L} \inf\limits_{w \in U} \|w - v\|, \sup\limits_{w \in U} \inf\limits_{v \in L} \|w - v\| \big\}$,

where $[a, b]$ is the interval-of-interest, $u(x)$ and $l(x)$ are the upper and lower bounds, respectively, $L := \{(x, l(x)) | x \in [a, b]\}$ and $U := \{(x, u(x)) | x \in [a, b]\}$. An advantage of the last two error measures over the first one is that they do not discriminate between the two coordinate directions.

Among the derivative-free partition rules, the most frequently used are the *interval bisection* rule (i.e., partition the interval into two equal parts) and the *maximum error* rule (i.e., partition the interval at the point where the maximum error is attained) (Rote 1992).

The most important difference between our Sandwich algorithm applied in combination with the bounds described in section 3.2 and the traditional Sandwich algorithms is that adding a new point with our algorithm reduces the error not only in the interval where the point is added, but also in the adjacent neighbouring intervals. This is not the case when lower bounds based on derivative information are used.

3.3.2. New methods for dynamic Pareto point generation. Recently, four new iterative strategies to select a new data point to be evaluated have been suggested by Siem *et al* (2006a). In the current paper, we apply three of these strategies to iteratively add a new Pareto optimal solution to the PEF calculated so far:

  (i) maximal *maximum error* measure with the *interval bisection* rule, where the new data point is selected exactly in the middle of the interval-of-interest,

(ii) maximal *Hausdorff distance* measure with the *interval bisection* rule, where the new data point is selected exactly in the middle of the interval-of-interest,

(iii) maximal *total uncertainty area* measure with the *maximum error* rule, where the new data point is selected to minimize the maximal total area of uncertainty.

In the next section, we apply these methods to a convex bi-objective FMO problem using a simplified clinical test case and show proof of principle.

3.4. Approximation of a high-dimensional Pareto efficient frontier

Convexity bounds. The upper and lower bounds as presented in section 3.2 can be extended to the higher dimensional case with more than two convex objective functions. Some ingredients have already been developed by Siem *et al* (2006b).

Suppose that n PO solutions $p(\varepsilon^1), \ldots, p(\varepsilon^n)$ of the ε-constraint problem (5) are given for the points $\varepsilon^1, \ldots, \varepsilon^n \in \mathcal{E}^{l-1} \subseteq \mathbb{R}^{l-1}$, where

$$\mathcal{E} := \left\{ \varepsilon \in \mathbb{R}^{l-1} : \exists x \in \mathcal{X} : F_j(x) \leqslant \varepsilon_j^i, \forall j = 2, \ldots, l \right\}$$

is the set of feasible constraint values. For any value of ε in the convex hull of the points $\varepsilon^1, \ldots, \varepsilon^n$, we can calculate upper and lower bounds for $p(\varepsilon)$. For $\varepsilon \in \mathrm{conv}\{\varepsilon^1, \ldots, \varepsilon^n\}$ the upper bound can be found by solving the linear program (LP):

$$u(\varepsilon) := \min_{\alpha_1, \ldots, \alpha_n} \quad \sum_{i=1}^{n} \alpha_i \, p(\varepsilon^i) \tag{8}$$
$$\text{s.t.} \quad \varepsilon = \sum_{i=1}^{n} \alpha_i \varepsilon^i \quad 0 \leqslant \alpha_i \leqslant 1 \quad \sum_{i=1}^{n} \alpha_i = 1.$$

For the lower bound, we can distinguish between the case that no derivative information is available and the case that this is available. If no derivative information is available, the lower bound can be found by solving

$$\ell^0(\varepsilon) := \max_{k=1,\ldots,n} \left\{ \begin{array}{rl} \max_{z^k, z_i^k, t^k} & t^k y(\varepsilon^k) - \sum_{i \neq k} z_i^k y(\varepsilon^i) \\ \text{s.t.} & \varepsilon^k t^k = \sum_{i \neq k} z_i^k \varepsilon^i + z^k \varepsilon \\ & \sum_{i \neq k} z_i^k + z^k = t^k \\ & z_i^k \geqslant 0 \\ & z^k = 1. \end{array} \right\}. \tag{9}$$

In (9), we take the maximum of n solutions of LPs. Note that not necessarily all of these LPs are feasible. If an LP is infeasible, we take $-\infty$ as optimal value. The complete derivation of (8) and (9) can be found in Siem *et al* (2006b). So to obtain the upper and lower bounds of $p(\varepsilon)$, for $\varepsilon \in \mathcal{E}$, we must solve $n + 1$ LPs, which can be solved quickly compared to the ε-constraint problem (5). It can be shown that $u : \mathrm{conv}\{\varepsilon^1, \ldots, \varepsilon^n\} \mapsto \mathbb{R}$ and $\ell^0 : \mathrm{conv}\{\varepsilon^1, \ldots, \varepsilon^n\} \mapsto \mathbb{R}$ are piecewise linear on their domain.

Furthermore, if derivative (or subdifferential) information is available by either using the WS method or by calculating the Lagrange multipliers associated with the ε-constraint method, we can generalize (7). Then, we obtain

$$p(\varepsilon) \geqslant p(\varepsilon^i) + s(\varepsilon^i)^T (\varepsilon - \varepsilon^i),$$

where $s(\varepsilon^i)$ is a subdifferential of p in the point ε^i. So,

$$\ell^1(\varepsilon) = p(\varepsilon^i) + s(\varepsilon^i)^T (\varepsilon - \varepsilon^i)$$

is a lower bound of $p(\varepsilon)$.

Since we also know that the Pareto surface $p(\varepsilon)$ is decreasing, i.e., for all $\tilde{\varepsilon}, \bar{\varepsilon} \in \mathcal{E}$ with $\tilde{\varepsilon} \leqslant \bar{\varepsilon}$, $p(\tilde{\varepsilon}) \geqslant p(\bar{\varepsilon})$, we obtain another lower bound:

$$p(\varepsilon) \geqslant p(\varepsilon^i), \qquad \forall (\varepsilon \in \mathcal{E} | \varepsilon \leqslant \varepsilon^i), \quad \forall i = 1, \ldots, n.$$

Iterative strategy. Extension of the Sandwich algorithms as presented in section 3.3 to the multivariate case is not straightforward, since there are no intervals to be considered anymore.

However, a suitable value for $\varepsilon \in \mathcal{E}$ to evaluate can be found by constructing a grid $G \subseteq \mathbb{R}^{\ell-1}$, calculating the values of $u(\varepsilon)$ and $\ell^0(\varepsilon)$ on $G \cap \text{conv}\{\varepsilon^1, \ldots, \varepsilon^n\}$, and taking the value of ε for which the uncertainty is maximal, i.e., we determine

$$\max_{\varepsilon \in G \cap \text{conv}\{\varepsilon^1, \ldots, \varepsilon^n\}} (u(\varepsilon) - \ell^0(\varepsilon)). \tag{10}$$

Let us denote the value of ε, which maximizes (10) by ε^0. Next, we calculate $p(\varepsilon^0)$ by solving the ε-constraint problem (5), and again calculate the values of the new upper and lower bounds. We repeat the procedure until we find a sufficiently good approximation of the Pareto surface. In case we have derivative information we take $\ell^1(\varepsilon)$, instead of $\ell^0(\varepsilon)$.

Note that the new values of ε^0, to evaluate, always lie inside the convex hull of the already evaluated values $\varepsilon^1, \ldots, \varepsilon^n$. Therefore, the initial set of values $\varepsilon^1, \ldots, \varepsilon^n$ should be chosen such that its convex hull $\text{conv}\{\varepsilon^1, \ldots, \varepsilon^n\}$ is sufficiently large.

3.5. Fluence map interpolation

The process to dynamically generate a new PO treatment plan can be accelerated by exploiting the convexity property of the PEF. Initial fluence maps for the new PO plan can be obtained by linear interpolation of the fluence maps of neighbouring PO plans. This was first postulated by Thieke (2003), and is now further developed in the present paper. The interpolated solution lies in the feasible region and is a first-order approximation of the true intermediate PO solution. The piecewise-linear upper and lower bounds that were derived in section 3.2 provide limits for the true objective function values.

In case a new PO treatment plan $\mathbf{w}^*$ has to be generated for a constraint $F_2(\mathbf{d}(\mathbf{w}^*)) \leqslant \tilde{\varepsilon}$, a good initial solution can be obtained by interpolation of two neighbouring PO plans, $\mathbf{w}_i^*$ and $\mathbf{w}_{i+1}^*$, for which $F_2(\mathbf{d}(\mathbf{w}_i^*)) = \varepsilon_i$ and $F_2(\mathbf{d}(\mathbf{w}_{i+1}^*)) = \varepsilon_{i+1}$ and $\varepsilon_i \leqslant \tilde{\varepsilon} \leqslant \varepsilon_{i+1}$. Clearly, there exists a $\lambda \in [0, 1]$ such that $\tilde{\varepsilon} = \lambda \varepsilon_i + (1 - \lambda)\varepsilon_{i+1}$. It is straightforward to show that the interpolated plan, $\tilde{\mathbf{w}} := \lambda \mathbf{w}_i^* + (1 - \lambda)\mathbf{w}_{i+1}^*$, is mapped into the shaded region of the objective space that is schematically depicted in figure 4.

The interpolated fluence maps can advantageously be used as an initial solution that is close to the set of PO plans. Starting an optimization run from the interpolated solution will drastically reduce the computation time in comparison to a new optimization run that starts from scratch.

4. Clinical case study

4.1. Description of the case study

A simplified clinical head-and-neck case is used to test the approximation algorithms on a bi-objective FMO problem. The case is represented by an oropharynx tumour (i.e., clinically T3N2bM0 squamous cell carcinoma originating from the left tonsil) with a planning target volume (PTV) comprising the gross tumour volume plus unilaterally affected regional lymph nodes with margins and the elective treatment volume of bilateral cervical lymph nodes (regions II−V and retropharyngeal nodes). Both parotid salivary glands, the spinal cord and
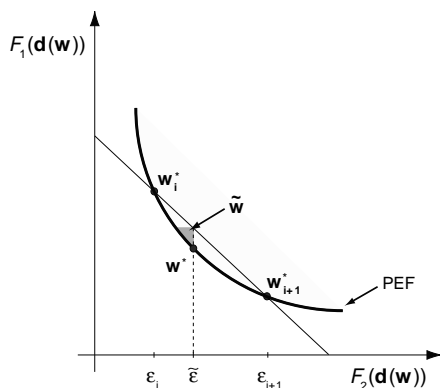
Figure 4. The plan $\tilde{\mathbf{w}}$ interpolated from neighbouring Pareto optimal plans $\mathbf{w_i^*}$ and $\mathbf{w_{i+1}^*}$ is mapped to the shaded region in the objective space. Hence, $\tilde{\mathbf{w}}$ is a good initial estimate for the true intermediate Pareto optimal plan $\mathbf{w}^*$.

(a)   (b)
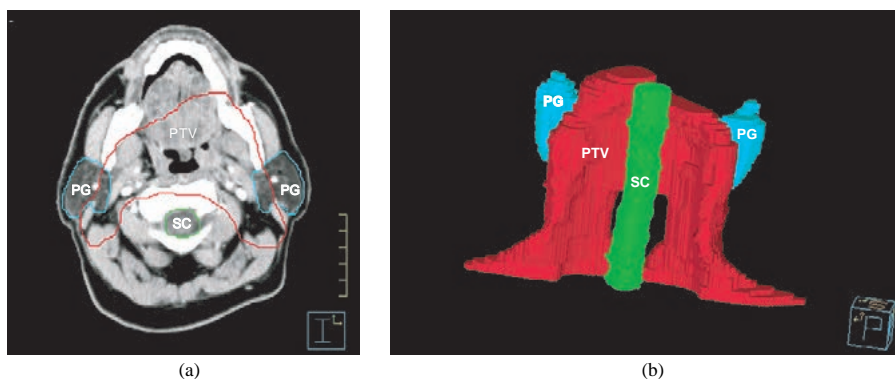
Figure 5. Delineation of planning target volume (PTV) and organs at risk (PG = parotid glands, SC = spinal cord + brainstem) on a transversal CT slice (a) and the corresponding three-dimensional model (b) of the delineated structures for the tested clinical head-and-neck case with oropharyngeal cancer in posterior oblique view.

the brainstem are organs at risk (see figure 5). For the sake of simplicity, the spinal cord and brainstem are considered to be one OAR, as is also the case for both the left and right parotids.

The optimization geometry consists of seven equiangular co-planar 6 MV photon beams containing 5800 bixels of 5×5 mm^2 each. The dose matrix comprises $60 \times 60 \times 40$ voxels of $4 \times 4 \times 4$ mm^3 each. As optimization objective, the mean dose to the parotids is minimized subject to different constraint values for the target dose heterogeneity (measured in terms of the relative standard deviation of the dose distribution in the PTV). Minimizing the parotid gland mean dose is a radiobiologically relevant planning goal if substantial sparing of the gland function is required (Eisbruch *et al* 1999), while striving for a homogeneous dose distribution in the PTV is a clinically relevant but conflicting objective, because of the partial overlap between the two volumes (see figure 5(a)). In order to generate plans that are clinically

recognizable, additional constraints were imposed to prevent hot spots in the surrounding of the PTV and achieve conformality. Plans were normalized to a mean PTV dose of 46 Gy. Boosting of the gross tumour volume up to 70 Gy was omitted for simplicity of the proof of principle pursued in this paper.

The objective function for mean parotid dose (MPD) was calculated by the generalized equivalent uniform dose (gEUD_a) function of Niemierko (1999) with $a = 1$:

$$F_1(\mathbf{d}(\mathbf{w})) := \text{gEUD}_1(\mathbf{d}(\mathbf{w})) = \sum_{i \in \text{PG}} d_i(\mathbf{w}) \Delta v_i, \tag{11}$$

where Δv_i is the voxel volume relative to the total volume of the organ. The constraint function for target dose heterogeneity (TDH) is defined as the squared ratio of the PTV dose variance against the PTV mean dose, $\bar{d}_{\text{PTV}}$, expressed by

$$F_2(\mathbf{d}(\mathbf{w})) := \sum_{i \in \text{PTV}} \left(\frac{d_i(\mathbf{w}) - \bar{d}_{\text{PTV}}}{\bar{d}_{\text{PTV}}} \right)^2 \Delta v_i. \tag{12}$$

With regard to the convexity of both functions used in the optimization, Choi and Deasy (2002) have proven that the gEUD_a function is convex for $a \geqslant 1$ and concave for $a \leqslant 1$. Hence, the MPD is a convex (and linear) function of dose and bixel weight. The TDH function is convex as well, because its second derivative is constant, provided that $\bar{d}_{\text{PTV}}$ is constant.

We used a beta version of the Pinnacle[3] Radiation Therapy Planning software (Version 7.7a, not for clinical use) from Philips Medical Systems (Madison, USA) with the Research Interface facility (Version 1.0) from RaySearch Laboratories AB (Stockholm, Sweden) to the inverse treatment planning module P[3] IMRT (Release 2.0). This module is coupled to the quasi-Newton sequential quadratic programming solver NPSOL from Stanford University (Stanford, USA).

The three strategies mentioned in section 3.3.2 are applied to iteratively add nine new Pareto optimal solutions to the PEF calculated so far. For comparison, we also consider the case in which 12 data points have been chosen equidistantly. An initial set of three Pareto optimal plans was determined by generating the two anchor points that bound the extent of the PEF in the two objective dimensions and a single interior point. The two anchor points were found by individually optimizing the two objective functions F_1 and F_2 of (11) and (12), respectively.

4.2. Results of the case study

The two anchor points found are (6.2%, 33.4 Gy) and (58.6%, 1.9 Gy), indicating a wide interval-of-interest in the feasible objective space. Both objective intervals are considered not to be clinically relevant over the whole range of feasible objective values. It reflects the extreme sparing of the parotids at the interface with the target volume. Nevertheless, the case is used to show proof of principle. Solving the constrained optimization problem (6) for TDH $\leqslant$ 11.3% yielded a third point (11.3%, 12.7 Gy) in the initial set of Pareto optimal solutions. The computational times required to solve the constrained optimization problem on a SunFire 250 workstation ranged from 45 to 135 min with an average of 100 ± 10 (1 SD) min, when a stopping tolerance of 10^{-5} was applied.

Two-dimensional dose distributions of the three corresponding Pareto optimal plans are shown in figure 6. From this figure, it can be easily appreciated that an improved sparing of both parotid glands can be accomplished by reducing dose in the target volume near the interface with the glands. This comes at the cost of an increased heterogeneity of the dose distribution in the remaining target volume, causing an increased overall TDH. This effect is
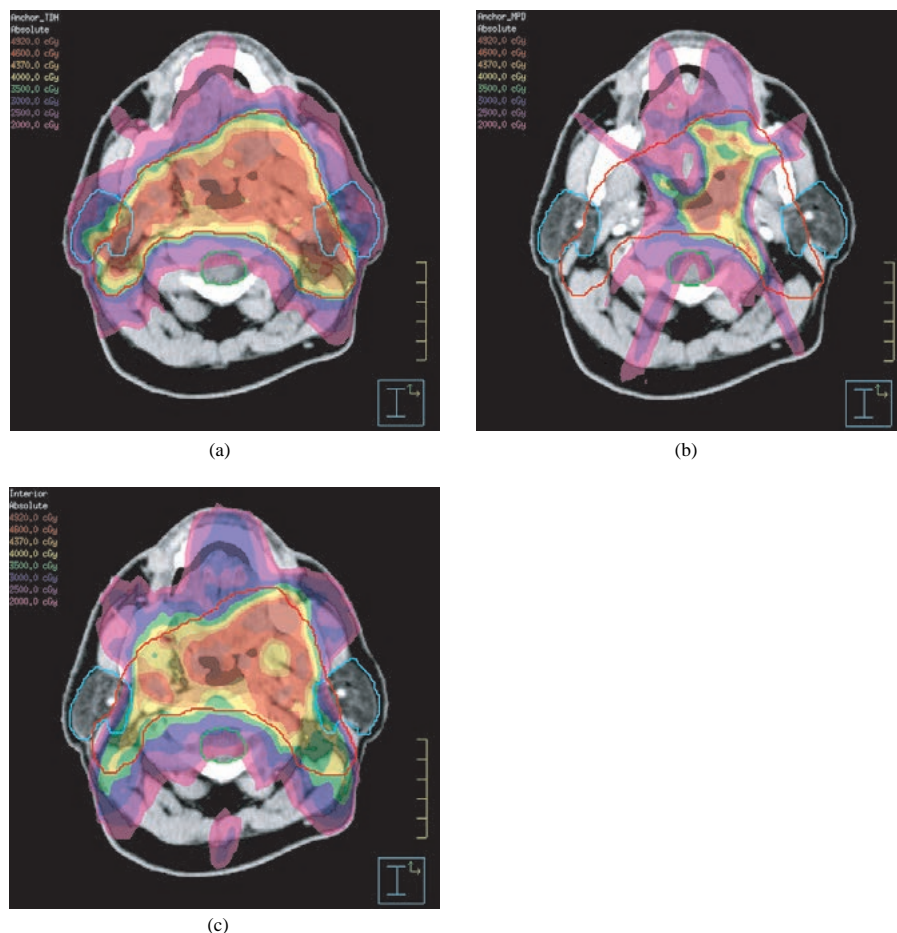
Figure 6. Two-dimensional dose distribution of Pareto optimal plans corresponding to (a) the leftmost anchor point, (b) the rightmost anchor point and (c) the interior point having a target dose heterogeneity of 11.3%. All plans have been normalized to a mean target dose of 46 Gy.

most dominantly present in those regions where the PTV and parotid glands overlap. More caudally, the target dose heterogeneity is significantly smaller.

Starting from the initial set of three Pareto optimal solutions, we constructed upper and lower bounds to determine a new coordinate for solving another constrained optimization problem with a new TDH constraint value ε. The results for the three PEF-generating strategies are shown in table 1. For each of the three strategies, the error measures of section 3.3.2 are tabulated for each iteration. The equidistant approach gave 15.3, 43.7 and 3.5 as error measures for the *maximum error*, *total uncertainty area* and *Hausdorff distance* PEF-generating strategies, respectively. As expected, all three iterative strategies give better results than the equidistant approach. Strategies based on the *Hausdorff distance* and the *total uncertainty area* have an almost equal performance.
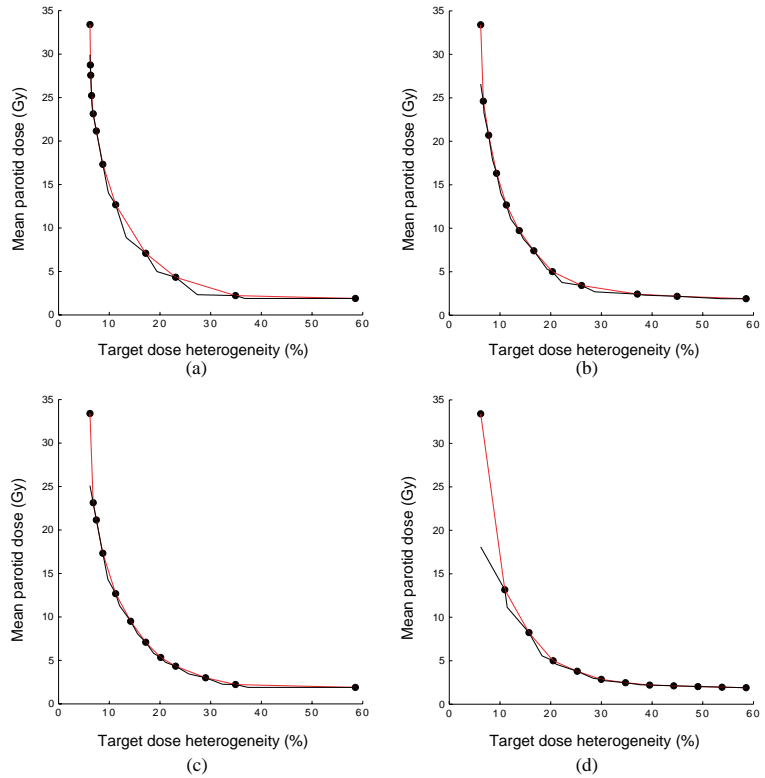
Figure 7. Estimated PEF with upper (red) and lower (black) bounds for (a) the *maximum error*, (b) the *total uncertainty area* and (c) the *Hausdorff distance* criterion after nine iterations, and (d) for the 12 equidistant points.

Table 1. Error measures after each iteration for the *maximum error* (ME), *total uncertainty area* (MA) and *Hausdorff distance* (MH) PEF-generating strategies.

	ME			MA			MH		
Iteration	ME	MA	MH	ME	MA	MH	ME	MA	MH
0	19.6	290.5	9.9	19.6	290.5	9.9	19.6	290.5	9.9
1	11.4	241.7	9.2	17.6	121.8	6.2	18.5	157.8	8.3
2	9.5	231.7	9.2	15.8	78.8	3.8	17.1	87.8	4.4
3	8.4	103.6	7.1	9.1	51.3	1.9	15.9	61.7	3.8
4	8.3	100.1	7.1	9.1	36.1	1.8	11.4	36.4	1.8
5	7.7	98.6	7.1	9.1	27.8	1.5	8.4	27.3	1.3
6	6.1	40.6	3.5	8.3	22.1	1.0	8.4	22.8	1.2
7	4.3	39.8	3.5	6.8	17.4	0.9	8.4	17.2	0.9
8	1.8	21.6	1.3	6.8	14.0	0.9	8.3	13.8	0.8
9	1.7	21.4	1.3	6.8	11.2	0.7	8.3	11.7	0.6

In figure 7, the piecewise-linear upper and lower bounds after nine iterations are shown for the PEFs generated with the three iterative strategies. From this figure, it can be seen that a PEF-generating strategy based on either the *Hausdorff distance* or the *total uncertainty area*

is to be preferred over the *maximum error* criterion. After all, the latter criterion tends to add new data points to the steepest part of the PEF, where the difference between the upper and lower bounds is likely to be largest. The former two criteria are not biased towards any of the two coordinate directions and will thus populate the PEF with points that are better spread according to the curvature of the underlying graph.

5. Discussion and conclusion

Formulating the fluence map optimization model in inverse treatment planning for IMRT as a multi-objective optimization problem eliminates the *a priori* selection of weighting factors (Thieke 2003) and facilitates the understanding of trade-offs between conflicting treatment plan evaluation criteria (Craft *et al* 2005). The quantification of these trade-offs comes down to finding the Pareto efficient frontier in the objective space. However, no closed form of the PEF exists in general, the set of Pareto optimal solutions is infinitely large and each solution requires a separate single-objective optimization problem to be solved. Therefore, it is most practical to generate a discrete subset of representative solutions that characterizes the full PEF.

In this paper, we present an algorithm to dynamically compute a small set of globally Pareto optimal treatment plans that approximates the PEF up to some predefined error by solutions that are well distributed over the PEF and take its curvature into account. This method relies on the convexity of treatment plan evaluation criteria that are used as objective functions to steer the optimization process. Previously, it has been shown that many treatment plan evaluation criteria commonly used in existing inverse treatment planning systems are convex or can be transformed into convex criteria (Romeijn *et al* 2004).

The current paper is mainly restricted to a bi-objective FMO problem and describes a method to extend to higher dimensions. Nevertheless, it is our experience that in practice the clinically relevant trade-offs are often limited to two or three objectives at most. Physicians lower the dimensionality of the problem by handling objectives as constraints.

Solutions were found by using a constraint-based method that minimizes one of the objectives while the other objective is handled as constraint by setting an upper bound value. We exploit the property that the resulting univariate PEF is convex to approximate it by constructing improvingly tighter piecewise-linear upper and lower bounds as new points on the PEF are generated according to an iterative strategy. We have applied the method for different error measures and partitioning rules on a simplified clinical case. By applying different iterative strategies, three estimates for the Pareto curve have been obtained and compared. Strategies based on error measures that are unbiased with respect to the coordinate directions were found to have the best performance. Normalizing the objective dimensions into the interval [0, 1] such that error measures are not dominated by the large-magnitude objective function(s) might be applied to overcome the bias.

A non-Sandwich algorithm based approach was explored in a recent paper by Craft *et al* (2005) considering two-dimensional trade-offs between tumour dose homogeneity versus OAR sparing. To populate the PEF with individual points, these authors applied the standard method of altering weighting factors in a weighted-sum (WS) objective scalarization approach as well as the normalized normal constraint (NC) method (Messac *et al* 2003). For the WS method the challenge is to systematically choose the weights such that a representative set of Pareto optimal solutions is obtained. The NC method is designed to resolve this problem by using constraints and thus an even spread of points over the PEF can be guaranteed. However, both the WS and NC method do not exploit the convexity of the PEF to compute the next Pareto optimal point, and hence ignore the accumulated curvature information that is contained in

the set of points computed so far. Our algorithm overcomes this drawback by adding a new point to the PEF where the uncertainty between the lower and upper bound approximation is maximally reduced. This guarantees an efficient and effective placement of points on the PEF.

Others have recently developed a multi-dimensional PEF-generating recursive Sandwich algorithm (called PGEN) that can be applied in combination with any WS-method-based inverse treatment planning system (Craft *et al* 2006). The rationale behind this algorithm is in essence the same as for our method, except that it exploits the mean value theorem in addition to convexity properties. This facilitates a systematic choice of weighting factors and the use of derivative information to construct upper and lower bounds of the PEF.

In the present paper, we proved that lower bounds resulting from the combination of function evaluation and derivative information are tighter than those derived from function evaluation information only. In that respect, the WS-method-based Sandwich algorithm developed by Craft *et al* (2006) has a clear advantage over our EC-method-based Sandwich algorithm. Another advantage of the WS method is that it is suitable for use with existing single-objective solvers. Comparison of the weighted-sum approach versus the constrained-based approach remains to be established in a future study. In the first step of the PGEN algorithm, anchor points are found by individually optimizing the objective functions. It is highly plausible that these anchor points fall beyond the limits of the clinical interval-of-interest. Because weighting factors that correspond to the extreme points of the clinical interval-of-interest are unknown beforehand, constraints should be added to obtain clinically feasible solutions. In this regard, our EC method is cleaner.

Our method has at least three unique advantages over the other PEF-generating methods, which speed up the process to generate a representative set of Pareto optimal plans. In the first place, our derivative-free Sandwich algorithm also reduces the error in neighbouring intervals when adding a new point to the Pareto curve. Secondly, fluence map interpolation can be used to provide an initial solution for a new Pareto point to be added to the PEF. Thirdly, our method allows us to restrict to a clinically relevant subregion of the PEF by optimizing on the upper and lower limit of the interval-of-interest of a criterion as set by the constraint.

Our findings have shown that a PEF can be generated efficiently and effectively with a derivative-free Sandwich algorithm that is based on the ε-constraint method. Trade-offs between tumour coverage and critical organ sparing have been quantified and mapped for a clinical case with two objectives. The automatic generation of a set of representative Pareto plans avoids the human iteration loop in current inverse treatment planning and facilitates *a posteriori* decision making in treatment planning.

Acknowledgments

References

Alber M, Birkner M and Nusslin F 2002 Tools for the analysis of dose optimization: II. Sensitivity analysis *Phys. Med. Biol.* **47** N265–70

Amols H I, Zaider M, Hayes M K and Schiff P B 1997 Physician/patient-driven risk assignment in radiation oncology: reality or fancy? *Int. J. Radiat. Oncol. Biol. Phys.* **38** 455–61

Bortfeld T 1995 Dosiskonformation in der Tumortherapie mit externer ionisierender Strahlung: Physikalische Möglichkeiten und Grenzen (Dose conformation in radiotherapy with external ionizing radiation: physical possibilities and limitations) *Doctoral Dissertation,* Ruperto-Carola University of Heidelberg

Boyd S and Vandenberghe L 2004 *Convex Optimization* (Cambridge: Cambridge University Press)

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

Burkard R, Hamacher H W and Rote G 1991 Sandwich approximation of univariate convex functions with an application to separable convex programming *Nav. Res. Logist.* **38** 911–24

Chankong V and Haimes Y Y 1983a *Multiobjective Decision Making Theory and Methodology* (New York: Elsevier)

Chankong V and Haimes Y Y 1983b Optimization-based methods for multiobjective decision-making: an overview *Large Scale Syst.* **5** 1–33

Choi B and Deasy J O 2002 The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning *Phys. Med. Biol.* **47** 3579–89

Craft D, Halabi T and Bortfeld T 2005 Exploration of tradeoffs in intensity-modulated radiotherapy *Phys. Med. Biol.* **50** 5857–68

Craft D L, Halabi T F, Shih H A and Bortfeld T R 2006 Approximating convex Pareto surfaces in multiobjective radiotherapy planning *Med. Phys.* **33** 3399–407

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

Eisbruch A, Ten Haken R K, Kim H M, Marsh L H and Ship J A 1999 Dose, volume, and function relationships in parotid salivary glands following conformal and intensity-modulated irradiation of head and neck cancer *Int. J. Radiat. Oncol. Biol. Phys.* **45** 577–87

Fruhwirth B, Burkard R and Rote G 1989 Approximation of convex curves with application to the bi-criteria minimum cost flow problem *Eur. J. Oper. Res.* **42** 326–38

Graeme S, Baker R, McCredden J E and Bain J D 2005 How many variables can humans process *Psychol. Sci.* **16** 70–6

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

Haimes Y Y, Lasdon L S and Wismer D A 1971 On a bicriterion formulation of the problems of integrated system identification and system optimization *IEEE Trans. Syst. Man Cybern.* **1** 296–7

Hamacher H W and Küfer K-H 2002 Inverse radiation therapy planning—a multiple objective optimization approach *Discrete Appl. Math.* **118** 145–61

Hunt M A, Hsiung C Y, Spirou S V, Chui C S, Amols H I and Ling C C 2002 Evaluation of concave dose distributions created using an inverse planning system *Int. J. Radiat. Oncol. Biol. Phys.* **54** 953–62

Küfer K-H, Scherrer A, Monz M, Alonso F, Trinkaus H, Bortfeld T and Thieke C 2003 Intensity-modulated radiotherapy—a large scale multi-criteria programming problem *OR Spectr.* **25** 223–49

Llacer J, Deasy J O, Bortfeld T R, Solberg T D and Promberger C 2003 Absence of multiple local minima effects in intensity modulated optimization with dose–volume constraints *Phys. Med. Biol.* **48** 183–210

Marler R T and Arora J S 2004 Survey of multi-objective optimization methods for engineering *Struct. Multidisc Optim.* **26** 369–95

Messac A, Ismail-Yahaya A and Mattson C A 2003 The normalized normal constraint method for generating the Pareto frontier *Struct. Multidisc Optim.* **25** 86–98

Miettinen K 1999 *Nonlinear Multiobjective Optimization* (Boston: Kluwer)

Niemierko A 1999 A generalized concept of equivalent uniform dose *Med. Phys.* **26** 1100 (Abstract)

Pareto V 1906 *Manuale di Economica Politica, Societa Editrice Libraria, Milan. Translated into English as Manual of Political Economy by Schwier A S* ed A S Schwier and A N Kelley (New York, 1971)

Reemtsen R and Alber M 2004 Continuous optimization of beamlet intensities for photon and proton radiotherapy (http://www.math.tu-cottbus.de/INSTITUT/lsing1/publications/reemtsen_e.html)

Romeijn H E, Dempsey J F and Li J G 2004 A unifying framework for multi-criteria fluence map optimization models *Phys. Med. Biol.* **49** 1991–2013

Rote G 1992 The convergence rate of the Sandwich algorithm for approximating convex functions *Computing* **48** 337–61

Siem A Y D, den Hertog D and Hoffmann A L 2006a A method for approximating univariate convex functions using only function evaluations with applications to Pareto curves *Eur. J. Oper. Res.* submitted

Siem A Y D, den Hertog D and Hoffmann A L 2006b *Multivariate Convex Approximation and Least-Norm Convex Data-Smoothing* ed M Gavrilova, O Gervasi, V Kumar, C J K Tan, D Taniar, A Laganà, Y Mun and H Choo (Berlin: Springer) pp 812–21 ICCSA 2006, LNCS 3982

Thieke C 2003 Multicriteria optimization in inverse radiotherapy planning *Doctoral Dissertation,* Ruperto-Carola University of Heidelberg

van Tol-Geerdink J J, Stalmeier P F M, van Lin E N J T, Schimmel E C, Huizenga H, van Daal W A J and Leer J W 2006 Do prostate cancer patients want to choose their own radiation treatment? *Int. J. Radiat. Oncol. Biol. Phys.* **66** 1105–11

Webb S 2001 *Intensity-Modulated Radiation Therapy (Series in Medical Physics and Biomedical Engineering)* (Bristol: Institute of Physics Publishing)

Yang X and Goh C 1997 A method for convex curve approximation *Eur. J. Oper. Res.* **97** 205–12

VI

Efficient approximation of the convex Pareto efficient frontier

A Method for Approximating Univariate Convex Functions Using Only Function Value Evaluations

A. Y. D. Siem, D. den Hertog
Department of Econometrics and Operations Research/Center for Economic Research (CentER), Tilburg University,
5000 LE Tilburg, The Netherlands {a.y.d.siem@uvt.nl, d.denhertog@uvt.nl}

A. L. Hoffmann
Department of Radiation Oncology, Radboud University Nijmegen Medical Centre,
6525 GA Nijmegen, The Netherlands, a.hoffmann@rther.umcn.nl

In this paper, piecewise-linear upper and lower bounds for univariate convex functions are derived that are only based on function value information. These upper and lower bounds can be used to approximate univariate convex functions. Furthermore, new sandwich algorithms are proposed that iteratively add new input data points in a systematic way until a desired accuracy of the approximation is obtained. We show that our new algorithms that use only function value evaluations converge quadratically under certain conditions on the derivatives. Under other conditions, linear convergence can be shown. Some numerical examples that illustrate the usefulness of the algorithm, including a strategic investment model, are given.

Key words: approximation; convexity; metamodel; analysis of algorithms; sandwich algorithm
History: Accepted by Robert Fourer, Area Editor for Modeling: Methods and Analysis; received March 2009; revised January 2010, July 2010; accepted August 2010. Published online in *Articles in Advance* December 29, 2010.

1. Introduction

In the field of discrete approximation, we are interested in approximating a function given a certain discrete data set. We sometimes know beforehand that the function that is to be approximated has some characteristics. It could be known, for example, that it is a nonnegative, increasing, or convex function. However, the approximation does not necessarily inherit these characteristics. In Siem et al. (2008a), e.g., nonnegativity preserving (trigonometric) polynomials and rational functions are studied. In that paper, monotonicity preserving polynomials are also studied. In Burkard et al. (1991), Fruhwirth et al. (1989), Rote (1992), and Yang and Goh (1997) the so-called sandwich algorithms are proposed for univariate approximation of convex functions. In these algorithms, upper and lower bounds of the convex function are constructed. The methods in Burkard et al. (1991), Fruhwirth et al. (1989), and Rote (1992) make use of derivative information, which is not always available, especially in case of black-box functions. In Yang and Goh (1997), a derivative-free optimization problem has to be solved in case there is no derivative information. This requires many function value evaluations, which may be time consuming. In §3 we treat these sandwich algorithms in more detail.

In this paper we present a methodology to find approximations of univariate convex functions via upper and lower bounds. Besides convexity, we

assume that the function value can be calculated accurately for each point in the interval of interest. The methods proposed in this paper are especially valuable for those applications for which one function evaluation is time consuming (say, minutes or even hours per evaluation). An important difference with the methods studied in Burkard et al. (1991), Fruhwirth et al. (1989), and Rote (1992) is that our methodology uses only function value evaluations. Based on convexity, we construct upper and lower bounds of a convex univariate function $y \colon \mathbb{R} \mapsto \mathbb{R}$, which is only known on a finite set of points $x^1, \ldots, x^n \in \mathcal{U} \subseteq \mathbb{R}$ with values $y(x^1), \ldots, y(x^n) \in \mathbb{R}$, and for which no derivative information is known. In den Boef and den Hertog (2007), this kind of bound is used for efficient line searching of convex functions. We show that if derivative information is available, tighter lower bounds can be obtained than if this information is not available. In our previous paper (Siem et al. 2008b), it is shown that under certain conditions, these upper and lower bounds can be improved by using suitable transformations. Furthermore, we present iterative strategies that determine in each iteration which new input data point is best to be evaluated next, until a desired accuracy is met. Different criteria can be used to select this new input data point. The iterative strategies that we use also belong to the class of sandwich algorithms. However, these sandwich algorithms are based on derivative

592

information. Therefore, in §3 we introduce a version of the sandwich algorithm that can be used when only function value information is available. Moreover, we introduce two other iterative strategies, based on function value information only. For these two strategies, we do not give convergence proofs. In §4 we give convergence proofs for our new sandwich algorithms. Under certain conditions on the derivatives of $y(x)$, we can show quadratic convergence for different variants of our sandwich algorithms. Under other conditions, linear convergence can be shown for our sandwich algorithms. With some numerical examples, we compare different variants of our new iterative strategies and show that our methods give better results than choosing the input data equidistantly. Also, we apply these methods to approximate the convex optimal value function of a strategic investment model.

The methods described in this paper can be used in several ways.

• First, our method can be used for optimizing a black-box function that is time consuming to evaluate and for which no derivative information is available. This function, for example, could be represented by a deterministic computer simulation. As described by Zabinsky et al. (2003), in engineering and science it is often necessary to estimate univariate black-box functions based on a small number of function value evaluations (without getting derivative or subgradient information). In that paper, a derivative-free approximation method is developed for Lipschitz functions. The approximation of the black-box function can, for example, be used for optimization. See, e.g., Baran (2004) and Hansen et al. (1991) for univariate applications. Our approximation method can also be used for integration. In Baran et al. (2008), an approximation method is developed for integration of univariate black-box functions for which no derivative information is available. As in Zabinsky et al. (2003), they assume that the function is Lipschitz continuous and that the Lipschtitz constant is known.

• Second, our methods can be used for obtaining parametric curves in an efficient way. Mathematical modeling packages (e.g., AIMMS) often offer the possibility of approximating parametric curves and are often based on optimization runs for parameter values on an equidistant grid for the parameter range. Especially when an optimization run is time consuming, our method may reduce the time needed significantly. Note that for our methods it is necessary to know that the parametric curve is convex. For many classes of problems, convexity of the parametric curves can indeed be proven. In Slack and Lewis (2002), it is shown that many parametric curves in the area of operations strategy are convex; see also Bitran and Morabito (1999) for (convex)

parametric curve analysis of manufacturing system designs. In Demeulemeester et al. (1998), the approximation of convex time/cost trade-off curves in project networks is discussed. Many engineering problems also lead to convex parametric curves. The parametric curve example in den Boef and den Hertog (2007) stemming from resource management in an in-home network is a specific example in this area. In this application the so-called buffer device costs should be approximated as a function of the data transportation capacity. Approximations for convex buffer-bandwith parametric curves are also developed in Kumaran and Mandjes (2001). It is well known that parametric curves are often nondifferentiable.

• The third possible application concerns the approximation of a convex Pareto frontier, resulting from, for example, a biobjective optimization problem. In particular, if calculating one Pareto point is already time consuming, our methods may significantly reduce the time needed. For example, in intensity-modulated radiation therapy (IMRT), the efficient approximation of convex Pareto frontiers is a very important topic. The aim is to deliver the right radiation dose to the tumour while sparing the healthy organs. This leads to (convex) multiobjective optimization problems; see, e.g., Kufer et al. (2003), Craft et al. (2006), and Hoffmann et al. (2006). Again, we note that to apply our methods, we should know that the Pareto frontier is convex. In general, this cannot be guaranteed. However, for several important classes of problems, convexity is guaranteed. Moreover, in certain cases, transformations of the objective functions are possible such that the functions become convex, and consequently, the resulting Pareto frontier is convex. For example, see Romeijn et al. (2004) and Hoffmann et al. (2008), in which it is shown that most of the biological objectives used in IMRT can be made convex using appropriate transformations. Another nontrivial example of convex Pareto frontiers are the exchange curves in inventory management; see, e.g., Silver et al. (1998). In Diakonikolas and Yannakakis (2009), many other biobjective Pareto examples are given. It is well known that parametric curves are often nondifferentiable.

• The fourth possible application concerns the efficient approximation of convex (or concave) functions in optimization problems. A well-known method is to linearize such convex functions, such that the problem gets a simpler format, e.g., network programming or mixed-integer linear programming. The aim is to find an accurate linearization with minimal number of linear pieces. This is also the aim in Guérin et al. (2006). Moreover, as will be explained later in this paper, our method also yields a piecewise-linear convex underestimator that may be used in many optimization algorithms, especially in global

optimization. Our new approximation method may, for example, be used in the DIRECT global optimization algorithm (see Chiter 2006) as an alternative for the selecting and dividing strategy of hyperrectangles.

The remainder of this paper is organized as follows. In §2, we show how we can obtain upper and lower bounds to approximate univariate convex functions and show that if derivative information is available, we can obtain even tighter bounds. In §3, we discuss iterative strategies for determining new data points to be evaluated. In §4, we consider convergence results. In §5, we study numerical examples. Finally, in §6, we give our conclusions and discuss possible directions for further research.

2. Approximating Convex Functions

2.1. Bounds Based on Function Value Evaluations

Let $y: \mathbb{R} \to \mathbb{R}$ be a convex function. Suppose that n input data points $x^1, \ldots, x^n \in [x^1, x^n] \subseteq \mathbb{R}$, are given, together with the n corresponding output data points $y(x^1), \ldots, y(x^n) \in \mathbb{R}$. It is well known that the straight line through the points $(x^i, y(x^i))$ and $(x^{i+1}, y(x^{i+1}))$, for $1 \leq i \leq n-1$, is an upper bound of the curve $y(x)$, for $x \in [x^i, x^{i+1}]$; see Figure 1. Furthermore, it is known that the straight lines through the points $(x^{i-1}, y(x^{i-1}))$ and $(x^i, y(x^i))$, for $2 \leq i \leq n-1$, and $(x^{i+1}, y(x^{i+1}))$ and $(x^{i+2}, y(x^{i+2}))$, for $1 \leq i \leq n-2$, are lower bounds of the curve $y(x)$, for $x \in [x^i, x^{i+1}]$; again see Figure 1. We summarize this in the following theorem. In the following, we define

$$y_i(x) := \frac{x^{i+1} - x}{x^{i+1} - x^i} y(x^i) + \frac{x - x^i}{x^{i+1} - x^i} y(x^{i+1})$$
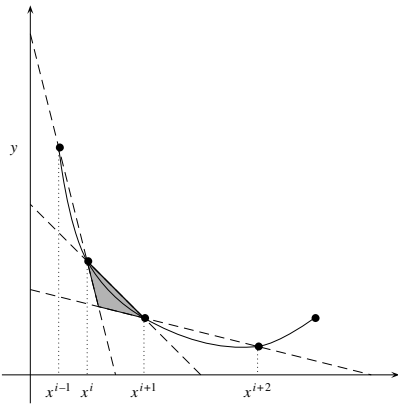$$\text{for } 1 \leq i \leq n-1. \quad (1)$$

Theorem 1. *Let n input/output data points $(x^1, y(x^1))$, $\ldots, (x^n, y(x^n))$, with $x^1 < x^2 < \cdots < x^n$ be given, and let $y(x)$ be convex. Suppose furthermore that $x^i \leq x \leq x^{i+1}$, for some $1 \leq i \leq n-1$. Then*

$$y(x) \leq y_i(x), \quad (2)$$

and

$$y(x) \geq y_j(x) \quad \text{for } 1 \leq j < i \text{ and } i < j \leq n-1. \quad (3)$$

2.2. Bounds Based on Derivatives

In addition to the bounds described in §2.1, we can also use derivative information (if present) to obtain lower bounds. Suppose that $y(x)$ is convex and differentiable, and that not only are the n data points $(x^1, y(x^1)), \ldots, (x^n, y(x^n))$ given, but also the derivative information $(x^1, y'(x^1)), \ldots, (x^n, y'(x^n))$. Then we have

$$y(x) \geq y(x^i) + y'(x^i)(x - x^i) \quad \forall x \in [x^1, x^n],$$
$$\forall i = 1, \ldots, n. \quad (4)$$

This lower bound is schematically shown in Figure 2. The following theorem states that these lower bounds are tighter than the lower bounds derived in the previous subsection, which do not use derivative information. We define the functions

$$t_i(x) := y(x^i) + y'(x^i)(x - x^i) \quad \text{for } 1 \leq i \leq n-1. \quad (5)$$

Theorem 2. *Let n input/output data points $(x^1, y(x^1))$, $\ldots, (x^n, y(x^n))$, with $x^1 < x^2 < \cdots < x^n$ be given, and let $y(x)$ be differentiable and convex. Suppose furthermore that $x^i \leq x \leq x^{i+1}$, for some $1 \leq i \leq n-1$. Then*

$$t_i(x) \geq y_{i-1}(x), \quad \text{if } i \geq 2, \quad (6)$$

and

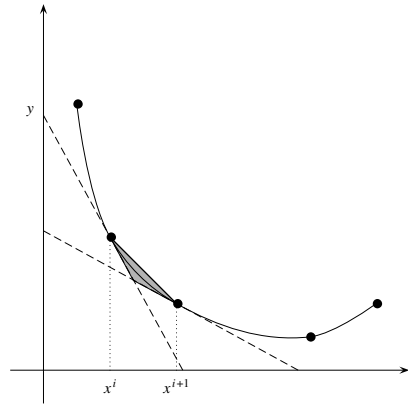$$t_{i+1}(x) \geq y_{i+1}(x), \quad \text{if } i \leq n-2. \quad (7)$$

Figure 1   Upper and Lower Bounds for a Convex Function on the Interval $[x^i, x^{i+1}]$ Using Only Function Value Evaluations

Figure 2   Upper and Lower Bounds for a Convex Function on the Interval $[x^i, x^{i+1}]$, Using Derivative Information

594

3. Iterative Strategies

In this section, we deal with iterative strategies to approximate univariate convex functions. These methods select a new input data point to evaluate in every iteration, until a desired accuracy is met. In §3.1, we consider the so-called sandwich algorithms that are already known from literature. These sandwich algorithms can be used in combination with the lower bound based on derivative information (4). In §3.2, we introduce a version of the sandwich algorithm that can be used in combination with the lower bounds based on function value evaluations only (3). Furthermore, we propose two other iterative strategies to add new input data points.

3.1. Sandwich Algorithms Using Derivative Information

In this section, we consider sandwich algorithms based on derivative information to construct approximations that satisfy a prescribed accuracy δ. There is a vast literature on these sandwich algorithms; see Burkard et al. (1991), Fruhwirth et al. (1989), Rote (1992), and Yang and Goh (1997). In these sandwich algorithms, upper and lower bounds are generated in an iterative way. We start with evaluating the function that is to be approximated at a "small" number of input data points, $x^1, \ldots, x^n \in [x^1, x^n] \subseteq \mathbb{R}$; i.e., we calculate $y(x^1), \ldots, y(x^n) \in \mathbb{R}$, and the derivative values $y'(x^1), \ldots, y'(x^n) \in \mathbb{R}$. Then we calculate the associated upper and lower bounds (2) and (4). The input data points $x^1, \ldots, x^n$, with $x^1 < \cdots < x^n$ define a set of intervals $I = \{[x^1, x^2], [x^2, x^3], \ldots, [x^{n-1}, x^n]\}$. Let δ_j denote the error for interval j, and let $J \subseteq I$ denote the set of intervals for which the error $\delta_j > \delta$. We can use different kinds of error measures, which we mention below. Next, we partition an arbitrary interval in the set J according to some of the partition rules, which we mention below, and calculate the output value y and its derivative y' at the input value x^0, where the interval is partitioned; i.e., we calculate $y(x^0)$ and $y'(x^0)$. Then we determine the new upper and lower bounds. Whenever the error of any of the two subintervals is greater than δ, we add this interval to the set J. We repeat this procedure until all intervals in J have an error smaller than δ, i.e., until $J = \varnothing$. This procedure is summarized in Algorithm 1.

Algorithm 1 (Sandwich algorithm with derivative information)

```
INPUT:
  An initial set of intervals J, for which
    δⱼ > δ, for all j ∈ J.
WHILE J ≠ ∅ DO
  SELECT interval [a, b] ∈ J.
  J := J\{[a, b]}.
  PARTITION [a, b] into two subintervals
```

```
  [a, c] and [c, b].
  Calculate y(c) and y'(c).
    Calculate the new upper and lower bounds.
  IF δ_[a, c] > δ
    J := J ∪ {[a, c]}.
  ENDIF
  IF δ_[c, b] > δ
    J := J ∪ {[c, b]}.
  ENDIF
ENDWHILE
```

Different error measures and different partition rules have been proposed in literature. The error measures as mentioned in Rote (1992) are as follows:

1. *Maximum error* on interval (∞-norm): $\delta_{[a, b]}^{\infty} = \max_{x \in [a, b]} \{u(x) - l(x)\}$;

2. *Uncertainty area* enclosed by bounds on interval (1-norm): $\delta_{[a, b]}^1 = \int_{[a, b]} (u(x) - l(x)) \, dx$;

3. *Hausdorff distance* on interval:

$$\delta_{[a, b]}^H = \max \left\{ \sup_{v \in L} \inf_{w \in U} |w - v|, \sup_{w \in U} \inf_{v \in L} |w - v| \right\},$$

where $[a, b]$ is the interval of interest, $u(x)$ is the upper bound, $l(x)$ is the lower bound, $L = \{(x, l(x)) \mid x \in [a, b]\}$, and $U = \{(x, u(x)) \mid x \in [a, b]\}$. An advantage of the last two error measures is that they do not discriminate between the two coordinate directions.

The partition rules as mentioned in Rote (1992) are as follows:

1. *Interval bisection*: Interval is partitioned into two equal parts.

2. *Maximum error*: Interval is partitioned at the point where the maximum error is attained.

3. *Slope bisection*: Find the supporting line whose slope is the mean value of the slopes of the tangent lines at the endpoints. Partition the interval at the point where this line is tangent to the graph of the function.

4. *Chord rule*: Find the supporting tangent line whose slope is equal to the slope of the line connecting the two endpoints. Partition the interval at the point where this line is tangent to the graph of the function.

3.2. Iterative Strategies Using Only Function Value Information

We cannot use the sandwich algorithms as described in §3.1 in combination with the lower bounds based on function value evaluations only as described in (3) because we do not have derivative information. If we use the lower bounds from (3), adding a new point reduces the error not only in the interval where the point is added, but most possibly also in the neighbouring intervals. This is not the case if we use lower bounds based on derivative information. Therefore, in this section we adjust Algorithm 1 such that it can be applied in combination with the lower bounds

(3) based on function value evaluations only. The adjusted procedure is summarized in Algorithm 2. An important difference is that in Algorithm 2, we have to update the set J in a different way. We have to check whether the neighbouring intervals still belong to J. Furthermore, another difference is that we select the new input data point in the interval in which the error measure is largest instead of selecting an arbitrary interval. This may cause the error to decrease faster. Note that for the sandwich algorithm in §3.1, by selecting the interval where the error is maximal, the accuracy δ is not reached earlier than an arbitrary interval J is selected.

**Algorithm 2** (Sandwich algorithm with only function value information)

```
INPUT:
   An initial set of intervals J, for which
      δ_j > δ, for all j ∈ J.
WHILE J ≠ ∅ DO
   SELECT interval [a, b] ∈ J for which δ_[a, b] is
      maximal.
   J := J \ {[a, b]}
   PARTITION [a, b] into two subintervals [a, c]
      and [c, b].
   Calculate y(c).
   Calculate the new upper and lower bounds.
   IF δ_[a, c] > δ
      J := J ∪ {[a, c]}
   ENDIF
   IF δ_[c, b] > δ
      J := J ∪ {[c, b]}
   ENDIF
   Check if the errors of neighbouring
      intervals are still larger than δ, and
      if not, remove them from the set J.
ENDWHILE
```

Note that we use all three error measures, as mentioned in §3.1. With respect to the four partition rules mentioned in §3.1, we will only use the interval bisection rule because we focus on cases for which there is function value information only (no derivative or subgradient information).

Finally, we introduce two other iterative strategies. These add a new input data point such that the uncertainty area after adding that input data point is minimized until the uncertainty area is below a certain level δ. However, we do not know the uncertainty area after adding a new data point, because we do not know the output value y of that input data point. We solve this problem as follows. Suppose we have the input/output data points $(x^1, y(x^1)), \ldots, (x^n, y(x^n))$, with the corresponding upper and lower bounds; see Figure 3.

Then, if we locate the $(n + 1)$st point at (x^0, y^0), the uncertainty area after adding this point to our data reduces. A new data point must be located on
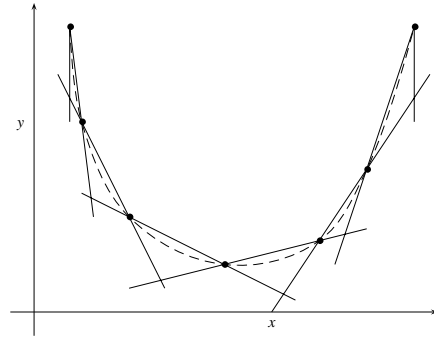
Figure 3   Upper and Lower Bounds for a Convex Function, Based on Function Value Evaluations

the line segment between $(x^0, l(x^0))$ and $(x^0, u(x^0))$, but it needs not to be identified now. Locating a test point on the segment, namely, at the point (x^0, y^0) where $l(x^0) \leq y^0 \leq u(x^0)$, we can estimate the uncertainty area as a function of x^0 and y^0. Therefore, a first approach is that by taking the average with respect to y^0 in the range $[l(x^0), u(x^0)]$, we calculate the average uncertainty area as a function of x^0. A second approach is that we calculate the worst case with respect to y^0 in the range $[l(x^0), u(x^0)]$, which is again a function of x^0.

Thus, we can evaluate the next data point x^0, according to the following rules:

• *Average area rule*: We take the value of x^0, where the average uncertainty area after the addition is minimal.

• *Worst-case area rule*: We take the value of x^0, where the maximal uncertainty area after the addition is minimal.

Let us now describe this more mathematically. Let us denote the upper bound after adding the point (x^0, y^0) as $u(x; (x^0, y^0))$ and the lower bound as $l(x; (x^0, y^0))$. Then the area between the upper bound and the lower bound is given by

$$A(x^0, y^0) = \int_X [u(x; (x^0, y^0)) - l(x; (x^0, y^0))] \, dx, \quad (8)$$

where $X = [x^1, x^n]$ is the total interval. As mentioned above, we define the average uncertainty area by

$$\bar{A}(x^0) = \frac{1}{u(x^0) - l(x^0)}$$
$$\cdot \int_{Y(x^0)} \int_X [u(x; (x^0, y^0)) - l(x; (x^0, y^0))] \, dx \, dy^0,$$
$$(9)$$

where $Y(x^0) = \{y \in \mathbb{R} \mid l(x^0) \leq y \leq u(x^0)\}$, and $u(x^0)$ and $l(x^0)$ are the bounds, based on the original data, before

adding a new point. The maximal uncertainty area is defined by

$$A_{\max}(x^0)$$
$$= \max_{y^0 \in Y(x^0)} \int_X \left[u(x; (x^0, y^0)) - l(x; (x^0, y^0)) \right] dx. \quad (10)$$

We are now interested in finding the value of $x^0 \in X$ for which $\bar{A}(x^0)$ or $A_{\max}(x^0)$ is minimal. This means that we solve the following problem

$$\min_{x^0 \in X} \bar{A}(x^0), \quad (11)$$

or

$$\min_{x^0 \in X} A_{\max}(x^0). \quad (12)$$

We repeat this until the error is below a desired accuracy level δ. In Figure 4, the optimal choices for the average area and the worst-case area rules are illustrated. The "x" denotes the next point that is evaluated.

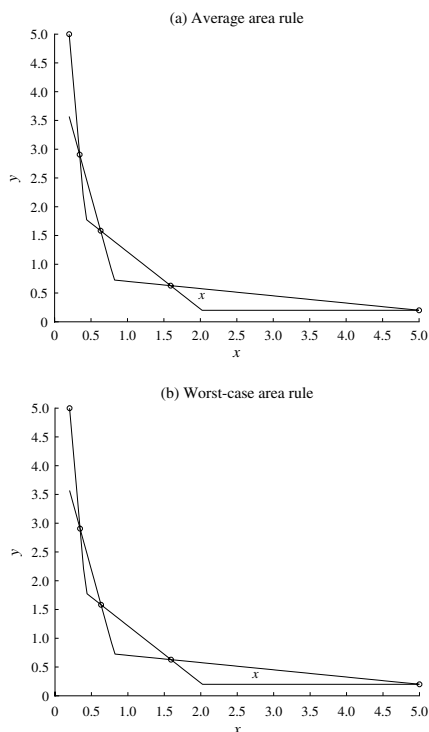Note that the use of barycentric coordinates simplifies the computation of the integrals in (9) and (10)

(a) Average area rule

(b) Worst-case area rule

Figure 4 **Illustration of Average Area Rule and Worst-Case Area Rule**

a lot. For a description of how to use barycentric coordinates, we refer to the proof of Theorem 6. The minimization problem in (11) and (12) is solved using a brute-force method, i.e., an exhaustive search on a fine grid.

In §4, we present convergence results for the (sandwich) Algorithm 2, and in §5, we show some numerical examples to illustrate and compare the different iterative strategies.

4. Convergence

In this section, we consider the convergence of Algorithm 1 and present new convergence results of Algorithm 2.

4.1. Sandwich Algorithms

Concerning convergence proofs for sandwich algorithms, Fruhwirth et al. (1989) proved that (sandwich) Algorithm 1 in the case of Hausdorff distance is of order $O(1/n^2)$, where n denotes the number of evaluation points. Burkard et al. (1991) obtained the same order for the maximum error (∞-norm). All these convergence results require that the right derivative in the left endpoint and the left derivative in the right endpoint of the interval are finite. Guérin et al. (2006) derived an optimal adaptive sandwich algorithm for which they proved $O(1/n^2)$ convergence without assuming bounded right and left derivatives at the left and right endpoints, respectively. Note that these sandwich algorithms use derivative information in each evaluation point. Yang and Goh (1997) propose a sandwich algorithm that uses only function evaluations. In each major iteration they have to minimize $y(x) - k_i x$ for each interval i, in which k_i is a certain constant. They correctly claim that there are derivative-free methods for such problems, but these problems require many function evaluations. Because Yang and Goh (1997) assume that $y(x)$ is easy to compute, these many extra function evaluations are no problem. However, as indicated in §1, we focus on situations in which $y(x)$ is time consuming to compute (say, minutes or even hours per evaluation).

In this section, we prove that our upper and lower bounds that do not use derivative information for equidistant input data points are of order $O(1/n^2)$ for the maximum error (∞-norm), for the uncertainty area (1-norm), and for the Hausdorff distance. These results also require bounded right and left derivatives at the left and right endpoints, respectively. Notice that in the case of approximating a convex Pareto frontier in particular, this assumption may be violated; see, e.g., Example 5.2. When this assumption does not hold, we prove an $O(1/n)$ convergence for our upper and lower bounds for equidistant input data points in the case of Hausdorff distance and the

uncertainty area (1-norm). Note that such a convergence result certainly does not hold for the maximum error (∞-norm). From these results it will follow in this section that (sandwich) Algorithm 2, using the interval bisection partitioning rule, converges at least at the same rate as in the equidistant case for all error measures.

4.2. Approximation Theory

In approximation theory, error bounds are usually given for the ∞-norm and involve some global property of the function. The following results are taken from the book by Szabados and Vértesi (1990), which contains a detailed survey on error bounds for univariate interpolation. The best error bound (in the ∞-norm) known is $O(1/\sqrt{n})$ for Lipschitz f for n function evaluations, obtained by Bernstein approximation. This improves to $O(1/n)$ if f' is Lipschitz. If the approximation is allowed to be nonconvex, then an $O((\log n)/n)$ error bound is obtained for Lipschitz f by Lagrange interpolation at the Chebyschev nodes and $O(1/n^2)$ if f' is continuous. Note that our convergence results improve Bernstein's convergence results for convex approximation.

4.3. Convergence Rates

In Theorems 3, 4, and 5, we give convergence results for equidistant input data points for three different error measures. For simplicity, we write $y^i = y(x^i)$.

THEOREM 3. *Suppose that* $y: [x^1, x^n] \mapsto \mathbb{R}$ *is convex and is known on the equidistant input data points* $x^1, \dots, x^n$. *Furthermore, suppose that the right derivative* y'_+ *in* x^1 *and the left derivative* y'_- *in* x^n *exist. Then, we have for the maximum error* $\delta^{\infty}_{[x^1, x^n]}$ *between the upper and lower bounds* $u(x)$ *and* $l(x)$ *of Theorem 1 that*

$$\delta^{\infty}_{[x^1, x^n]} \leq \frac{x^n - x^1}{n - 1}(y'_-(x^n) - y'_+(x^1)).$$

Furthermore, suppose that y *is* $C^2[x^1, x^n]$. *Then, we have for the maximum error* $\delta^{\infty}_{[x^1, x^n]}$ *between the upper and lower bounds* $u(x)$ *and* $l(x)$ *of Theorem 1 that*

$$\delta^{\infty}_{[x^1, x^n]} \leq \frac{(x^n - x^1)^2}{(n - 1)^2} \|y''\|_{\infty}.$$

PROOF. Let $\lambda^i(x) = (x^{i+1} - x)/h$, where h is the length of the interval $[x^i, x^{i+1}]$. For the intervals $[x^i, x^{i+1}]$, with $i = 1, \dots, n-2$, we subtract the "right" lower bound (3) from the upper bound (2):

$$\Delta y_i(x) = \lambda^i(x)(y^i - 2y^{i+1} + y^{i+2}). \quad (13)$$

Assuming that the right derivative y'_+ in x^1 exists, the left derivative y'_- in x^n exists, and using the convexity of $y(x)$, we obtain $y^{i+2} - y^{i+1} \leq y^n - y^{n-1} \leq hy'_-(x^n)$ and

$y^i - y^{i+1} \leq y^1 - y^2 \leq -hy'_+(x^1)$. Substituting these into (13) gives

$$\Delta y_i(x) \leq \delta^{\infty}_{[x^1, x^n]} \leq h(y'_-(x^n) - y'_+(x^1))$$
$$\leq \frac{x^n - x^1}{n - 1}(y'_-(x^n) - y'_+(x^1)). \quad (14)$$

For the interval $[x^{n-1}, x^n]$, we can also obtain (14) by subtracting the "left" lower bound (3) from the upper bound (2).

If we assume that y is $C^2[x^1, x^n]$, using Taylor's remainder formula, we have that $y^{i+2} = y^{i+1} + hy'(x^{i+1}) + \frac{1}{2}h^2 y''(\xi_1)$, where $\xi_1 \in [x^{i+1}, x^{i+2}]$, and $y^i = y^{i+1} - hy'(x^{i+1}) + \frac{1}{2}h^2 y''(\xi_2)$, where $\xi_2 \in [x^i, x^{i+1}]$. Substituting these into (13) gives

$$\Delta y_i(x) \leq \delta^{\infty}_{[x^1, x^n]} \leq \frac{1}{2}h^2(y''(\xi_1) + y''(\xi_2))$$
$$\leq \frac{(x^n - x^1)^2}{(n - 1)^2} \|y''\|_{\infty}. \quad (15)$$

For the interval $[x^{n-1}, x^n]$, we can also obtain (15) by subtracting the left lower bound (3) from the upper bound (2). $\square$

THEOREM 4. *Suppose that* $y: [x^1, x^n] \mapsto \mathbb{R}$ *is convex and is known on the equidistant input data points* $x^1, \dots, x^n$. *Then, we have for the total uncertainty area* $\delta^1_{[x^1, x^n]}$ *between the upper and lower bounds* $u(x)$ *and* $l(x)$ *of Theorem 1 that*

$$\delta^1_{[x^1, x^n]} \leq \frac{x^n - x^1}{n - 1}(y^{\max} - y^{\min}), \quad (16)$$

in which $y^{\max} = \max_{x \in [x^1, x^n]} y(x)$ *and* $y^{\min} = \min_{x \in [x^1, x^n]} y(x)$. *Furthermore, suppose that the right derivative* y'_+ *in* x^1 *exists and that the left derivative* y'_- *in* x^n *exists. Then, we have for the total area* $\delta^1_{[x^1, x^n]}$ *between the upper and lower bounds* $u(x)$ *and* $l(x)$ *of Theorem 1 that*

$$\delta^1_{[x^1, x^n]} \leq \frac{(x^n - x^1)^2}{2(n - 1)^2}(y'_-(x^n) - y'_+(x^1)). \quad (17)$$

PROOF. As in the proof of Theorem 3, let $\lambda^i(x) = (x^{i+1} - x)/h$, where h is the length of the interval $[x^i, x^{i+1}]$. For the intervals $[x^i, x^{i+1}]$, with $i = 1, \dots, n-2$, by (13) we have

$$A_i \leq \int_{x^i}^{x^{i+1}} \lambda^i(x)(y^i - 2y^{i+1} + y^{i+2}) \, dx$$
$$= \frac{1}{2}h(y^i - 2y^{i+1} + y^{i+2}), \quad (18)$$

where A_i denotes the uncertainty area on $[x^i, x^{i+1}]$. In the derivation of inequality (18), we only used the right lower bound. For the interval $[x^{n-1}, x^n]$, we do the same, but then with the left lower bound. We then obtain

$$A_{n-1} = \int_{x^{n-1}}^{x^n} \lambda^{n-1}(x)(y^{n-2} - 2y^{n-1} + y^n) \, dx$$
$$= \frac{1}{2}h(y^{n-2} - 2y^{n-1} + y^n).$$

598

Then the total uncertainty area is given by

$$\delta^1_{[x^1, x^n]} = \sum_{i=1}^{n-1} A_i \leq \tfrac{1}{2} h(y^1 - y^2 - y^{n-1} + y^n)$$

$$\leq h(y^{\max} - y^{\min}), \qquad (19)$$

which shows (16).

Now we assume that the right derivative y'_+ in x^1 exists and that the left derivative y'_- in x^n exists. Using the convexity of $y(x)$, we have $y^n - y^{n-1} \leq hy'_-(x^n)$, and $y^1 - y^2 \leq -hy'_+(x^1)$.

Then, instead of (19), we obtain

$$\delta^1_{[x^1, x^n]} \leq \frac{1}{2} h^2(y'_-(x^n) - y'_+(x^1))$$

$$\leq \frac{(x^n - x^1)^2}{2(n-1)^2}(y'_-(x^n) - y'_+(x^1)),$$

which shows (17). □

THEOREM 5. *Suppose that* $y: [x^1, x^n] \mapsto \mathbb{R}$ *is convex and is known on the equidistant input data points* $x^1, \dots, x^n$. *Furthermore, suppose that the right derivative* y'_+ *in* x^1 *exists and that the left derivative* y'_- *in* x^n *exists. Then, we have for the Hausdorff distance* $\delta^H_{[x^i, x^{i+1}]}$ *between the upper and lower bounds* $u(x)$ *and* $l(x)$ *of Theorem 1 on the interval* $[x^i, x^{i+1}]$ *that*

$$\delta^H_{[x^i, x^{i+1}]} \leq \frac{x^n - x^1}{n-1}(y'_-(x^n) - y'_+(x^1)). \qquad (20)$$

Furthermore, suppose that y *is* $C^2[x^1, x^n]$. *Then, we have for the Hausdorff distance* $\delta^H_{[x^i, x^{i+1}]}$ *between the upper and lower bounds* $u(x)$ *and* $l(x)$ *of Theorem 1 on the interval* $[x^i, x^{i+1}]$ *that*

$$\delta^H_{[x^i, x^{i+1}]} \leq \frac{(x^n - x^1)^2}{(n-1)^2}\|y''\|_\infty. \qquad (21)$$

PROOF. It is well known (see Fruhwirth et al. 1989) that the Hausdorff distance is always less than or equal to the maximum error. Therefore, (20) and (21) follow immediately from Theorem 3. □

In the following corollary, we show that the results of Theorems 3, 4, and 5 imply that (sandwich) Algorithm 2 with the interval bisection rule converges at least at the same rate.

COROLLARY 1. *If we apply Algorithm 2 in combination with the interval bisection rule instead of equidistant input data points, it will converge at least at the same rate as in the results in Theorems 3, 4, or 5, if, respectively,* δ^∞, δ^1, *or* δ^H *is used.*

PROOF. Let us first concentrate on the δ^∞ measure. Suppose that we want to obtain an approximation with level of uncertainty δ; i.e., the sandwich algorithm stops when $\delta^\infty \leq \delta$. Then, Theorem 3 gives us

the number (say, N) of equidistant points that are required to achieve that uncertainty.

Let us now define the equidistant grid of $N_k = 2^k + 1$ points, with $k \in \mathbb{N}$; i.e., these N_k equidistant points define a grid obtained by consecutively halving each interval k times. Now define $\bar{k}$ as the smallest value for k, such that $N_k \geq N$. The claim is now that the points that are generated by (sandwich) Algorithm 2 are a subset of the equidistant grid of $N_{\bar{k}}$ points. To see this, first observe that the points generated by our sandwich algorithm are a subset of the equidistant grid defined by N_l points, for sufficiently large l, because interval bisection is used. Let $\bar{l}$ be the smallest value for l such that the points generated by the sandwich algorithm are a subset of the N_l points. We have to show that $\bar{l} \leq \bar{k}$. Suppose on the contrary that $\bar{l} > \bar{k}$. This means that the set of evaluated points contains at least three points (say, z_1, z_2, and z_3) that are neighbours in the $N_{\bar{k}}$ equidistant grid, and a point w in-between z_1 and z_2 or in-between z_2 and z_3. In the first case (w is in-between z_1 and z_2), we use the right lower and upper bounds as in the proof of Theorem 3, especially for the result in (14), to show that $\delta^\infty_{[z_1, z_2]} \leq \delta$, which contradicts the fact that the sandwich algorithm continued to evaluate w. For the second case (w is in-between z_2 and z_3), we can use the left lower and upper bounds to obtain the same contradiction. This proves that $\bar{l} \leq \bar{k}$. This means that the sandwich algorithm needs at most $N_{\bar{l}} \leq N_{\bar{k}} \leq 2N$ points to obtain a δ accurate solution. Hence, the sandwich algorithm converges at the same rate as in the equidistant case of Theorem 3. The proofs for δ_1 and δ^H are similar. □

4.4. Area Reduction per Iteration

Next, we consider Algorithm 2 using the uncertainty area as error measure and the interval bisection partitioning rule. We give a more precise result on the area reduction per iteration. By adding a point in Algorithm 2, the triangle in which the data point is added is divided into two triangles. In the following lemma, we show that the total area of the two "new" triangles is at most half the area of the "original" triangle. We denote the area of the original triangle by A_t, and we denote the area of the new triangles by A_1 and A_2.

THEOREM 6. *Let* $y(x)$ *be convex. Suppose we use Algorithm 2 to approximate* $y(x)$ *and that we use the interval bisection partitioning rule and the uncertainty area as error measure. Then, we have that*

$$\frac{A_1 + A_2}{A_t} \leq \frac{1}{2}.$$

PROOF. First, we construct a parameterization for a general triangle, which captures all possible triangles that can occur in the algorithm. The chosen parameterization is shown in Figure 5. In this figure, the triangle ΔOAB represents the original triangle. Suppose
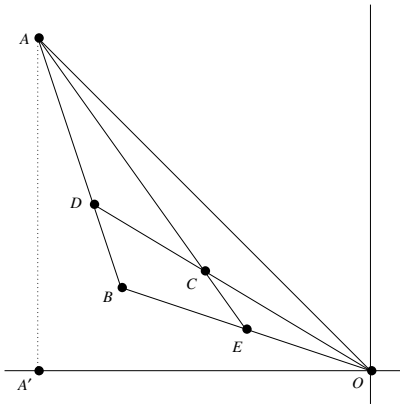
Figure 5   **Parameterization for a General Triangle Occurring in Algorithm 2**

that Algorithm 2 is applied for the approximation of a univariate convex and decreasing function $y(x)$. Then, the line OA is an upper bound of the function $y(x)$ on the interval $[x_A, 0]$. Suppose that there is a data point P on the left-hand side of data point A, and that there is a data point Q on the right-hand side of the data point O. Then, both PA and OQ are lower bounds for the function $y(x)$ on the interval $[x_A, 0]$. We denote the point where both lines intersect by B.

Using barycentric coordinates (α, β, γ) with respect to the triangle BOA (or OBA depending on the slope of the line OB), it can be verified that

$$\eta := \frac{A_1 + A_2}{A_t} = \alpha \left(\frac{\beta}{1-\beta} + \frac{\gamma}{1-\gamma} \right). \qquad (22)$$

Furthermore, C is on the line

$$(\alpha, \beta, \gamma) = (0, 1/2, 1/2) + u(1-v, v-1/2, -1/2),$$

with $(1-v, v, 0)$ being the barycentric coordinates of the intersection point of the vertical line through C and the edge BO. Then $0 \le u \le 1$ and $0 \le v \le 1/2$. Substituting this into the formula for η yields a rational function in u and v. The global extrema of this function can easily be calculated. From those it can be seen that the maximal value for η is $1/2$ for $u = 1$ and $v = 1/2$. $\square$

Using Theorem 6, we can also show that Algorithm 2 converges at least linearly using the uncertainty area as error measure and the interval bisection partitioning rule. Suppose that we add the data points such that we halve the areas of all triangles instead of choosing the interval with the largest area of uncertainty. In this way, the rate of convergence

can only become smaller. Suppose that we need k halvings to make the total area smaller than the prescribed δ; then,

$$A_t^k \le \left(\tfrac{1}{2} \right)^k A_t^0 < \delta, \qquad (23)$$

where A_t^k is the area of uncertainty after k halvings, and A_t^0 is the initial area of uncertainty. Note that k halvings require $N = \sum_{i=1}^{k} 2^{i-1} = 2^k - 1$ function value evaluations. Substituting this into (23), we obtain

$$\frac{A_t^0}{\delta} - 1 < N.$$

Therefore, at most

$$N = \frac{A_t^0}{\delta} - 1$$

iterations are needed to obtain a total uncertainty area smaller than δ.

5. Numerical Examples

In this section, we treat some numerical examples to illustrate the methodology proposed in this paper.

Example 5.1 (Artificial Data). In this example, we apply four different iterative methods that we discussed in §3.2, and we compare them with the case that we choose the input variables equidistantly. In the first method, we use the interval bisection rule in combination with the maximum error measure. In the second method, we use the interval bisection rule in combination with the Hausdorff distance error measure. In the third method, we select the new point such that the average uncertainty area after addition is minimized, i.e., the value of x that solves optimization problem (11); and in the fourth method we select the new point such that the worst-case uncertainty area is minimized, i.e., the value of x that solves optimization problem (12).

We consider the approximation of the function $y(x) = 1/x$ on the interval $[0.2, 5]$. As the initial data set we take two data points: $(0.2, 5)$ and $(5, 0.2)$. In Figure 6, the upper and lower bounds after several iterations for the worst-case area method are given. We measure the maximum error (ME), the uncertainty area (UA), and the Hausdorff distance (H) after each iteration. The results are shown in Table 1. As expected, all four new methods give better results than when we use the equidistant approach. Furthermore, as expected, if we use the maximum error or the Hausdorff distance as a measure to select a new point, the maximum error or the Hausdorff distance, respectively, in general decreases quicker than if we use the other criteria. Also, if we use the average area rule or the worst-case area rule, the total area decreases
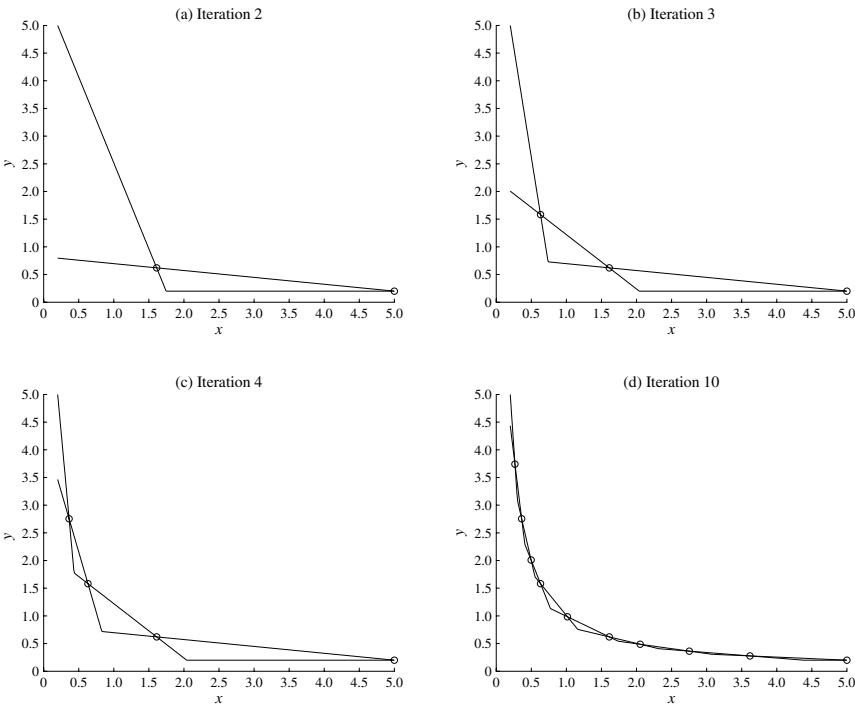
600

Figure 6 Upper and Lower Bounds of the Function $y = 1/x$ on the Interval $[0.2, 5]$ After Several Iterations of the Iterative Strategy Selecting a New Input Point According to the Worst-Case Area Rule

quicker than if we use the maximum error measure. In Figure 7, the upper and lower bounds after 10 iterations for all strategies are compared.

Next, we again approximate this function on the same interval, but now using derivative information, which is in fact Rote's algorithm. This was done

to compare the performance of our algorithm with this conventional algorithm. The results are shown in Table 2. The results for the algorithm that uses derivatives are better than when no derivative information is available. This is easy to understand, because Theorem 2 tells us that using derivative information (when

Table 1 Maximum Error, Total Uncertainty Area, and Hausdorff Distance After Each Iteration in Example 5.1 on Interval $[0.2, 5]$ Using the Maximum Error with Interval Bisection (ME/IB), Hausdorff Distance with Interval Bisection (H/IB), Minimal Average Area (MAA), Minimal Maximal Area (MMA), and Equidistant Iterative Strategies for Upper and Lower Bounds Based on Function Value Information

|	ME/IB			H/IB			MAA			MMA			Equidistant		
It.	ME	UA	H	ME	UA	H	ME	UA	H	ME	UA	H	ME	UA	H
0	4.80	11.52	3.39	4.80	11.52	3.39	4.80	11.52	3.39	4.80	11.52	3.39	4.80	11.52	3.39
1	4.43	5.53	2.04	4.43	5.53	2.04	4.23	3.77	1.35	4.20	3.62	1.27	4.43	5.53	2.04
2	3.96	2.67	1.07	3.96	2.67	1.07	3.11	1.65	0.51	2.99	1.64	0.53	4.18	3.52	1.42
3	3.21	1.33	0.51	3.21	1.33	0.51	2.62	1.35	0.35	1.43	1.17	0.48	3.96	2.54	1.07
4	2.25	0.74	0.22	2.25	0.74	0.22	2.62	1.03	0.35	1.43	0.72	0.43	3.75	1.95	0.85
5	1.29	0.52	0.20	1.29	0.52	0.20	1.48	0.61	0.25	1.43	0.50	0.15	3.56	1.57	0.70
6	0.58	0.45	0.20	1.29	0.44	0.15	1.48	0.46	0.25	1.43	0.40	0.15	3.38	1.30	0.59
7	0.32	0.43	0.20	1.29	0.35	0.11	1.03	0.37	0.25	1.14	0.31	0.13	3.21	1.09	0.51
8	0.25	0.38	0.17	1.29	0.26	0.11	1.03	0.28	0.11	0.41	0.25	0.13	3.06	0.94	0.44
9	0.23	0.37	0.17	1.29	0.23	0.09	1.03	0.23	0.11	0.41	0.20	0.13	2.92	0.82	0.39

601
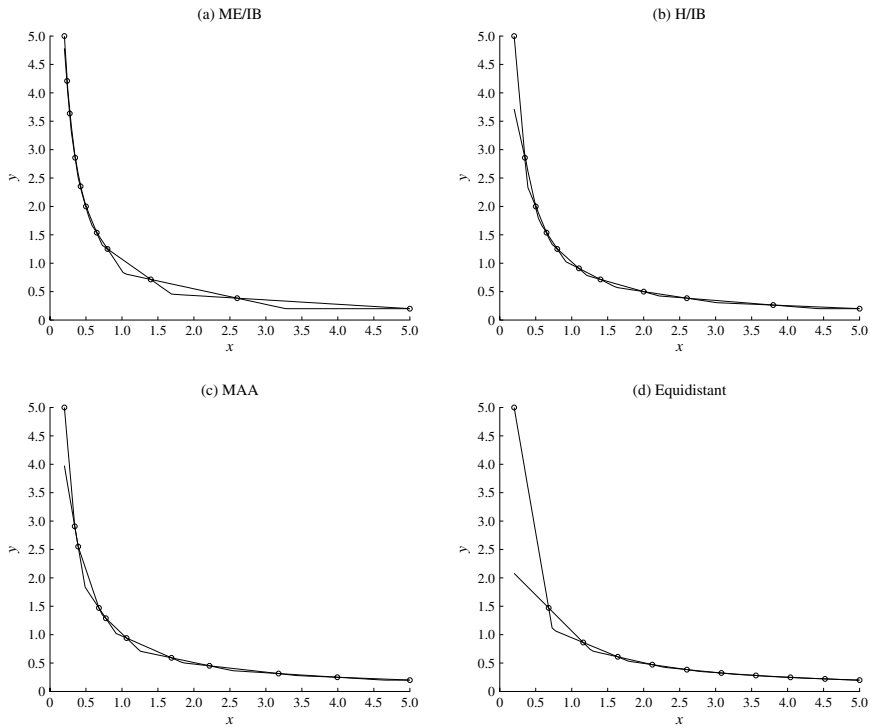
Figure 7   **Upper and Lower Bounds of the Function** $y = 1/x$ **on the Interval** $[0.2, 5]$ **After 10 Iterations of Different Sandwich Algorithms and Iterative Strategies**

available) leads to better lower and upper bounds. It is also interesting to observe that if one evaluation in Rote's algorithm counts for two evaluations (function value and derivative), then the results of our algorithm are (much) better. This can be seen by comparing the results of Tables 1 and 2 with the same

number of "evaluations," which is $2 + \mathrm{it}$ and $2(2 + \mathrm{it})$, respectively.

We also analyzed whether the proved convergence rates for the sandwich algorithm can be observed from the numerical results. First of all, we indeed observed that the upper bounds for the errors given

Table 2   **Maximum Error, Total Uncertainty Area, and Hausdorff Distance After Each Iteration in Example 5.1 on Interval** $[0.2, 5]$
Using the Maximum Error with Interval Bisection (ME/IB), Hausdorff Distance with Interval Bisection (H/IB), and Equidistant Iterative Strategies for Upper and Lower Bounds Based on Function Value and Derivative Information

|	ME/IB			H/IB			Equidistant		
It.	ME	UA	H	ME	UA	H	ME	UA	H
0	4.43	10.63	3.13	4.43	10.63	3.13	4.43	10.63	3.13
1	3.96	4.82	1.83	3.96	4.82	1.83	3.96	4.82	1.83
2	3.21	2.06	0.87	3.21	2.06	0.87	3.56	2.92	1.20
3	2.25	0.85	0.36	2.25	0.85	0.36	3.21	2.01	0.87
4	1.29	0.39	0.13	1.29	0.39	0.13	2.92	1.48	0.66
5	0.58	0.25	0.11	0.58	0.25	0.11	2.67	1.14	0.52
6	0.22	0.22	0.11	0.58	0.22	0.10	2.44	0.91	0.43
7	0.17	0.22	0.11	0.58	0.18	0.06	2.25	0.74	0.36
8	0.15	0.20	0.11	0.58	0.16	0.06	2.08	0.62	0.30
9	0.15	0.19	0.11	0.58	0.11	0.04	1.93	0.52	0.26

602
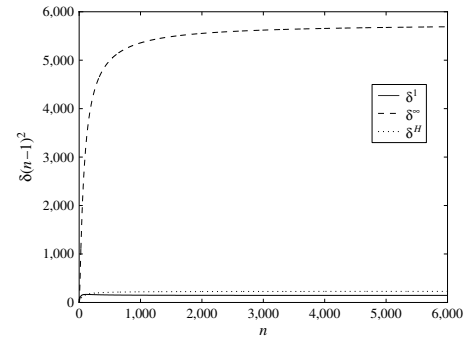
Figure 8 $(n-1)^2\delta$ **for Different δ-Measures**

in Theorems 3, 4, and 5 were really valid in this case. Moreover, we have run the methods up to 6,000 iterations and indeed observed that $\delta^\infty(n-1)^2$, $\delta^1(n-1)^2$, and $\delta^H(n-1)^2$ are more or less constant for large values of n. This illustrates the validity of the proved convergence rates; see also Figure 8.

Next, we again approximate the function $y(x) = 1/x$, but now only on the interval $[1, 2]$ with the points $(1, 1)$ and $(2, 0.5)$ as initial data set. The results are given in Table 3. We can see from this table that in this case, if we look to the area, choosing the inputs equidistantly does not perform significantly worse than the four more sophisticated methods. This could be explained by the shape of the two different functions that are to be approximated. On the interval $[0.2, 5]$, the function has much more curvature than on the interval $[1, 2]$. However, if we look at the maximum error and the Hausdorff distance, our four new methods perform better than the equidistant approach.

EXAMPLE 5.2 (STRATEGIC INVESTMENT MODEL). In this example we consider a strategic investment model. There exist many sorts of investment categories, such as deposits, saving accounts, bonds, stocks, real estate, commodities, foreign currencies, and derivatives. Each category has its own expected return, and its own risk characteristic. The strategic investment model models how top management could spread an overall budget over several investment categories. The objective is to minimize the portfolio risk (measured by the variance of the return), such that a certain minimal desired expected return is achieved. The model was introduced by Markowitz (1952) and is given by

$$y(M) := \min_x \ x^T V x$$
$$\text{s.t.} \quad r^T x \geq M,$$
$$e_p^T x = 1, \tag{24}$$
$$x \in \mathbb{R}_+^p,$$

where V is a positive semidefinite covariance matrix consisting of elements V_{ij} of covariances between investment categories i and j, r is the vector consisting of elements r_i of expected return of investment category i, M is the desired expected portfolio return, e_p is the p-dimensional all-one vector, x is the vector with elements x_i of fractions of the budget invested in each category, and p is the number of investment categories. Note that, in general, optimal value functions are nondifferentiable. In this strategic investment case, we are interested in the optimal value as a function of M, which can indeed be shown to be nondifferentiable for several values of M.

In Table 4, some data are given, which we took from Bisschop (2000). It contains three investment categories: stocks, bonds, and real estate.

The optimum in (24), can be seen as a function $y(M)$. It can be shown that y is convex and increasing. We carried out the same experiment as in Example 5.1. We applied the same four different

Table 3 Maximum Error, Total Uncertainty Area, and Hausdorff Distance After Each Iteration in Example 5.1 on Interval $[1, 2]$ **Using the Maximum Error with Interval Bisection (ME/IB), Hausdorff Distance with Interval Bisection (H/IB), Minimal Average Area (MAA), Minimal Maximal Area (MMA), and Equidistant Iterative Strategies for Upper and Lower Bounds Based on Function Value Information**

	ME/IB			H/IB			MAA			MMA			Equidistant		
It.	ME	UA	H	ME	UA	H	ME	UA	H	ME	UA	H	ME	UA	H
0	0.5000	0.2500	0.4472	0.5000	0.2500	0.4472	0.5000	0.2500	0.4472	0.5000	0.2500	0.4472	0.5000	0.2500	0.4472
1	0.1667	0.0625	0.1387	0.1667	0.0625	0.1387	0.1317	0.0638	0.1230	0.1376	0.0651	0.1283	0.1667	0.0625	0.1387
2	0.0667	0.0275	0.0593	0.0667	0.0275	0.0593	0.0932	0.0269	0.0741	0.0852	0.0263	0.0673	0.1000	0.0278	0.0800
3	0.0625	0.0206	0.0593	0.0667	0.0154	0.0521	0.0712	0.0182	0.0566	0.0430	0.0161	0.0411	0.0667	0.0154	0.0521
4	0.0222	0.0087	0.0181	0.0222	0.0087	0.0181	0.0302	0.0103	0.0289	0.0263	0.0091	0.0237	0.0476	0.0097	0.0366
5	0.0205	0.0075	0.0181	0.0222	0.0066	0.0172	0.0215	0.0062	0.0159	0.0241	0.0059	0.0180	0.0357	0.0066	0.0271
6	0.0179	0.0054	0.0172	0.0222	0.0046	0.0166	0.0145	0.0044	0.0108	0.0124	0.0046	0.0110	0.0278	0.0048	0.0209
7	0.0110	0.0035	0.0103	0.0110	0.0035	0.0103	0.0145	0.0033	0.0107	0.0124	0.0036	0.0102	0.0222	0.0036	0.0166
8	0.0080	0.0025	0.0065	0.0080	0.0025	0.0065	0.0145	0.0026	0.0107	0.0087	0.0027	0.0080	0.0182	0.0028	0.0135
9	0.0065	0.0020	0.0051	0.0065	0.0020	0.0051	0.0112	0.0022	0.0083	0.0070	0.0021	0.0054	0.0152	0.0023	0.0112

Table 4   Expected Returns and Covariances

| | | | V_{ij} | | |
| | | | j | | |
Category	i	r_i	1	2	3
Stocks	1	10.8	2.250	−0.120	0.450
Bonds	2	7.600	−0.120	0.640	0.336
Real estate	3	9.500	0.450	0.336	1.440

iterative strategies and calculated the maximum error, the uncertainty area, and the Hausdorff distance after each iteration. We compared the results with the case in which we choose the input data points equidistantly. The results are given in Table 5. As we expected, in Table 5 all of the four iterative strategies perform better than when we choose the input data points equidistantly. In Figure 9, the upper and lower bounds are shown after iteration 9 of the sandwich
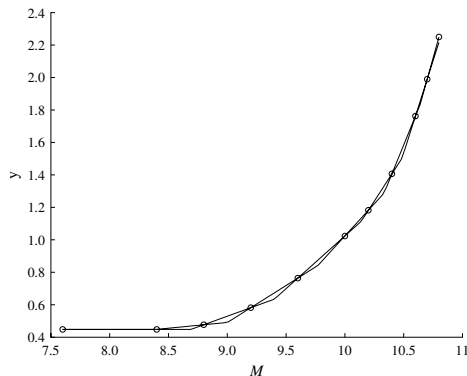
Figure 9   Upper and Lower Bounds of the Function $y(M)$ **on the Interval** [7.6, 10.8] **After Iteration 9 of the Sandwich Algorithm of Using the Hausdorff Distance for Example 5.2**

algorithm using the Hausdorff distance as the error measure.

6.   Conclusions and Further Research

In this paper we proposed piecewise-linear upper and lower bounds for approximation of univariate convex functions. For the approximation of univariate convex functions, we can construct piecewise-linear upper and lower bounds, based on function value evaluations only. These bounds can be given explicitly. The difference between the upper and lower bounds can be seen as a measure of accuracy. We may use the so-called sandwich algorithms to select new input values to be evaluated and obtain good approximations. We introduced a new variant of the sandwich algorithm, and we also introduced two new iterative strategies that minimize the area of uncertainty of the approximation. It can be shown that our new sandwich algorithms that do not use derivative information are of order $O(1/n^2)$ for the 1-norm, the ∞-norm, and the Hausdorff distance. These results require assumptions on the derivatives of $y(x)$. If these assumptions do not hold, it can be shown that under other conditions, we have $O(1/n)$ convergence for these sandwich algorithms. We applied these new algorithms to an artificial example and a practical example. It turned out that our algorithms perform better than when we choose the input data points equidistantly. This is especially the case if the function to be approximated has large curvature.

For further research, we are interested in generalizing this methodology to more dimensions, i.e., approximating functions of two or more variables. This is partly done in Siem et al. (2006). In this paper we also assume that the function value evaluations are accurate. An idea for further research may be to use the convexifying methods described in Siem et al. (2008a) and then apply the approximation methods described in this paper. In §1, we also mentioned the

Table 5   Maximum Error, Uncertainty Area, and Hausdorff Distance After Each Iteration in Example 5.2 Using the Maximum Error with Interval Bisection (ME/IB), Hausdorff Distance with Interval Bisection (H/IB), Minimal Average Area (MAA), Minimal Maximal Area (MMA), and Equidistant Iterative Strategies for Upper and Lower Bounds Based on Function Value Information

| | ME/IB | | | H/IB | | | MAA | | | MMA | | | Equidistant | | |
It.	ME	UA	H	ME	UA	H	ME	UA	H	ME	UA	H	ME	UA	H
0	1.8018	2.8828	1.5700	1.8018	2.8828	1.5700	1.8018	2.8828	1.5700	1.8018	2.8828	1.5700	1.8018	2.8828	1.5700
1	1.5347	1.3261	1.0624	1.5348	1.3261	1.0624	1.1625	0.9926	0.6849	1.1203	0.9870	0.6450	1.5347	1.3261	1.0624
2	0.7847	0.5060	0.4287	0.7847	0.5060	0.4287	0.7899	0.5915	0.4654	0.4994	0.5771	0.3590	1.0251	0.6979	0.6190
3	0.4606	0.2763	0.1974	0.4606	0.2763	0.1974	0.4108	0.3279	0.1944	0.4994	0.2672	0.2162	0.7847	0.4397	0.4287
4	0.1738	0.2024	0.1522	0.1738	0.2024	0.1522	0.4108	0.2011	0.1674	0.1792	0.1824	0.1618	0.6734	0.3064	0.3384
5	0.1321	0.1454	0.1087	0.1321	0.1454	0.1087	0.2538	0.1350	0.1034	0.1137	0.1192	0.0852	0.6065	0.2298	0.2838
6	0.1090	0.1337	0.1087	0.1321	0.0911	0.0835	0.2538	0.1003	0.1034	0.1137	0.0835	0.0852	0.5334	0.1779	0.2371
7	0.0897	0.0794	0.0835	0.1321	0.0675	0.0648	0.0865	0.0689	0.0674	0.1100	0.0609	0.0492	0.4606	0.1407	0.1974
8	0.0846	0.0644	0.0835	0.1321	0.0524	0.0501	0.0646	0.0531	0.0361	0.0676	0.0494	0.0492	0.3935	0.1127	0.1639
9	0.0565	0.0409	0.0372	0.0565	0.0409	0.0372	0.0646	0.0392	0.0361	0.0676	0.0384	0.0377	0.3332	0.0911	0.1358

use of the obtained piecewise-linear convex underestimator in (global) optimization. Note that this underestimator can be obtained by connecting the tips of the underestimating triangles. With respect to δ^∞, the approximation is not worse than the one presented in this paper because the worst approximation is always at the tips of the triangles. It is a subject of further research to investigate the effect of using this convex underestimator in global optimization methods.

Acknowledgments

References

Baran, I. 2004. Adaptive algorithms for problems involving black-box Lipschitz functions. Master's thesis, Massachusetts Institute of Technology, Cambridge.

Baran, I., E. D. Demaine, D. A. Katz. 2008. Optimally adaptive integration of univariate Lipschitz functions. *Algorithmica* **50**(2) 255–278.

Bisschop, J. 2000. *AIMMS Optimization Modeling*. Paragon Decision Technology, Haarlem, The Netherlands.

Bitran, G. R., R. Morabito. 1999. An overview of tradeoff curves in manufacturing systems design. *Production Oper. Management* **8**(1) 56–75.

Burkard, R. E., H. W. Hamacher, G. Rote. 1991. Sandwich approximation of univariate convex functions with an application to separable convex programming. *Naval Res. Logist.* **38** 911–924.

Chiter, L. 2006. DIRECT algorithm: A new definition of potentially optimal hyperrectangles. *Appl. Math. Comput.* **179**(2) 742–749.

Craft, D. L., T. F. Halabi, H. A. Shih, T. R. Bortfeld. 2006. Approximating convex Pareto surfaces in multiobjective radiotherapy planning. *Medical Phys.* **33**(9) 3399–3407.

Demeulemeester, E., B. De Reyck, B. Foubert, W. Herroelen, M. Vanhoucke. 1998. New computational results on the discrete time/cost trade-off problem in project networks. *J. Oper. Res. Soc.* **49**(11) 1153–1163.

den Boef, E., D. den Hertog. 2007. Efficient line search methods for convex functions. *SIAM J. Optim.* **18**(1) 338–363.

Diakonikolas, I., M. Yannakakis. 2009. Small approximate Pareto sets for biobjective shortest paths and other problems. *SIAM J. Comput.* **39**(4) 1340–1371.

Fruhwirth, B., R. E. Burkard, G. Rote. 1989. Approximation of convex curves with application to the bi-criteria minimum cost flow problem. *Eur. J. Oper. Res.* **42**(3) 326–338.

Guérin, J., P. Marcotte, G. Savard. 2006. An optimal adaptive algorithm for the approximation of concave functions. *Math. Programming* **107**(3) 357–366.

Hansen, P., B. Jaumard, S.-H. Lu. 1991. On the number of iterations of Piyavskii's global optimization algorithm. *Math. Oper. Res.* **16**(2) 334–350.

Hoffmann, A. L., D. den Hertog, A. Y. D. Siem, J. H. A. M. Kaanders, H. Huizenga. 2008. Convex reformulation of biologically-based multi-criteria IMRT optimization including fractionation effects. *Phys. Med. Biol.* **53**(22) 6345–6362.

Hoffmann, A. L., A. Y. D. Siem, D. den Hertog, J. H. A. M. Kaanders, H. Huizenga. 2006. Derivative-free generation and interpolation of convex Pareto optimal IMRT plans. *Phys. Med. Biol.* **51**(24) 6349–6369.

Kufer, K. H., A. Scherrer, M. Monz, F. Alonso, H. Trinkaus, T. R. Bortfeld. 2003. Intensity-modulated radiotherapy: A large scale multi-criteria programming problem. *OR Spectrum* **25**(2) 223–249.

Kumaran, K., M. Mandjes. 2001. The buffer-bandwidth trade-off curve is convex. *Queueing Systems* **38**(4) 471–483.

Markowitz, H. M. 1952. Portfolio selection. *J. Finance* **7**(1) 77–91.

Romeijn, H. E., J. F. Dempsey, J. G. Li. 2004. A unifying framework for multi-criteria fluence map optimization models. *Phys. Med. Biol.* **49**(10) 1991–2013.

Rote, G. 1992. The convergence rate of the sandwich algorithm for approximating convex functions. *Computing* **48**(3–4) 337–361.

Siem, A. Y. D., D. den Hertog, A. L. Hoffmann. 2006. Multivariate convex approximation and least-norm convex data-smoothing. M. Gavrilova, O. Gervasi, C. J. K. Tan, D. Taniar, A. Lagunà, Y. Mun, H. Choo, eds. *ICCSA 2006*. Lecture Notes in Computer Science, Vol. 3982. Springer, Berlin, 812–821.

Siem, A. Y. D., E. de Klerk, D. den Hertog. 2008a. Discrete least-norm approximation by nonnegative (trigonometric) polynomials and rational functions. *Structural Multidisciplinary Optim.* **35**(4) 327–339.

Siem, A. Y. D., D. den Hertog, A. L. Hoffmann. 2008b. The effect of transformations on the approximation of univariate (convex) functions with applications to Pareto curves. *Eur. J. Oper. Res.* **189**(2) 347–362.

Silver, E. A., D. F. Pyke, R. Peterson. 1998. *Inventory Management and Production Planning and Scheduling*. John Wiley & Sons, New York.

Slack, N., M. Lewis. 2002. *Operations Strategy*. Pearson Education Limited, Essex, UK.

Szabados, J., P. Vértesi. 1990. *Interpolation of Functions*. World Scientific, Singapore.

Yang, X. Q., C. J. Goh. 1997. A method for convex curve approximation. *Eur. J. Oper. Res.* **97**(1) 205–212.

Zabinsky, Z. B., R. L. Smith, B. P. Kristinsdottir. 2003. Optimal estimation of univariate black-box Lipschitz functions with upper and lower error bounds. *Comput. Oper. Res.* **30**(10) 1539–1553.

VII

Practical approach to assess IMRT plan trade-offs in NSCLC radiotherapy

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Quality assurance

A practical approach to assess clinical planning tradeoffs in the design of individualized IMRT treatment plans

René Monshouwer *, Aswin L. Hoffmann, Martina Kunze-Busch, Johan Bussink, Johannes H.A.M. Kaanders, Henk Huizenga

Department of Radiation Oncology, Radboud University Nijmegen Medical Centre, The Netherlands

ARTICLE INFO

ABSTRACT

Background and purpose: To investigate the tradeoffs between organ at risk sparing and tumour coverage for IMRT treatment of lung tumours, and to develop a tool for clinical use to graphically represent these tradeoffs.
Material and methods: For 5 patients with inoperable non-small cell lung cancer (NSCLC) different IMRT plans were generated using a standard TPS. The plans were automatically generated for a range of IMRT settings (weights and dose levels of the objective functions) and were systematically evaluated, focusing on the tradeoffs between organ at risk (OAR) dose and target coverage. A method to analyze and visualize planning tradeoffs was developed and evaluated.
Results: Lung and oesophagus were identified as the critical organs at risk for NSCLC, the sparing of which strongly influences PTV coverage. Systematically analyzing the tradeoffs between these organs revealed that the sparing of these organs was approximately linearly related to PTV coverage parameters. Using this property, a tool was developed to graphically present the tradeoffs between the sparing of these organs at risk and the PTV coverage. The tool is an effective method to visualize the tradeoffs.
Conclusions: A tool was developed to assist IMRT plan design and selection. The clear presentation of the tradeoffs between OAR dose and coverage facilitates the optimization process and offers additional information to the clinician for a patient specific choice of the optimal IMRT plan.

© 2010 Elsevier Ireland Ltd. All rights reserved. Radiotherapy and Oncology 97 (2010) 561–566

In many radiotherapy institutions, an increasing number of treatments are performed using intensity modulated radiation therapy (IMRT). Even though IMRT functionality has been available in treatment planning systems (TPS) for many years, the method for designing and optimizing an IMRT plan has changed only marginally. Generally, the implementation of IMRT for new tumour sites starts with the design of one or a few 'class solutions'. This 'class solution' consists of beam parameters (the number of beams, beam angles, beam energy, etc.), and mathematically formulated criteria for the dose distribution. These criteria consist of objective or constraint functions with associated parameters such as the weights and dose levels (further referred to as 'IMRT parameters'). The inverse planning module of the TPS uses the mathematical functions to design IMRT beams that result in a dose distribution that is a compromise between the different objectives. In most cases the result of such a class solution is not considered to be an optimum for individual patients, and the IMRT parameters need to be 'tuned' to produce a clinically acceptable solution. The rela-

tionship between the IMRT parameters and the resulting dose distribution is complicated and *a priori* it is not clear what tradeoffs are between e.g., coverage of the planning target volume (PTV) and sparing of organs at risk. Several approaches have been suggested to make the planning procedure more time-efficient and intuitive. One solution is to improve the 'input' side of the optimization algorithm, i.e., to add additional (clinical) *a priori* information to the optimizer to arrive at a clinically acceptable solution [1–3]. The disadvantage of these methods is that the exact nature of the tradeoffs between the objectives is not known beforehand, which may influence the prioritization.

An alternative approach is to explore many feasible (usually Pareto optimal) plans by building a library of so called 'best compromise' or Pareto optimal plans [4–10]. The advantage of this method is that the user gains insight in the quantitative tradeoffs that are inherent to the optimization problem, and can make decisions based on clinical grounds. For the characterization of the full range of possible IMRT solutions a vast number of optimizations are needed, although algorithms have been developed to limit the number of necessary optimizations by systematically choosing the IMRT parameters for each simulation [4,11,12]. Also it was recently found that surprisingly few plans are needed to obtain an accurate description of the Pareto efficient frontier [9].

* Corresponding author. Address: Radboud University Nijmegen Medical Centre, Department of Radiation Oncology, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands.

E-mail address: R.Monshouwer@rther.umcn.nl (R. Monshouwer).

The aim of this study is to bring these concepts closer to clinical practice. Firstly we recognise that only the properties of the planned dose distributions are clinically interesting for plan evaluation. Therefore the focus in this paper is not on the Pareto surface in classical sense as others [3–7,15], but on the surface created in the space of plan evaluation parameters. Assessing the plan evaluation criteria has strong advantages since some evaluation parameters are not available as objective functions. Secondly the plans are generated including segmentation of the fluence into deliverable IMRT fields. Therefore, the plans compared are representative for the treatment, and any selected plan can directly be used to treat the patient. Thirdly the complexity of the analysis is limited by focussing on the two most important regions of interest only. For the example site studied in this paper (NSCLC), the mathematical relationship between the evaluation criteria turns out to be simple and enables a clear and intuitive presentation of the clinical tradeoffs on a patient specific basis.

Material and methods

Patients and clinical objectives

Five patients diagnosed with stage II to III non-small cell lung cancer (NSCLC) were selected for this study. These patients were treated with concomitant radiotherapy (66 Gy in fractions of 2 Gy) and chemotherapy (Cisplatin plus Etoposide). The gross tumour volume (GTV) was delineated using contrast enhanced CT scans. An isotropic margin of 1.5 cm in the lung and 1 cm in the mediastinum was used to expand the GTV to the PTV. The PTV was located in the caudal right lung for patients 1 and 2, cranial right for patient 3, caudally left for patient 4 and cranially left for patient 5. The oesophagus was delineated from the hypopharynx to the diaphragm. Absolute planning constraints to the organs at risk were: a maximum dose of 50 Gy to the spinal cord, a V_{20} of the lungs below 35% and for the heart a V_{40} of less than 60%. The lung evaluation criteria were calculated from the healthy lung volume (both lungs minus the GTV).

IMRT optimization and database generation

IMRT planning was done with the Pinnacle[3] TPS (Version 8.0h Philips Medical Systems, Fitchburg, USA) using direct machine parameter optimization (DMPO) which produces deliverable plans with multiple beam segments. Plans were limited to less than 80 segments with a minimum segment size of 4 cm^2. An IMRT class solution was designed to get adequate coverage of the PTV, a spinal cord dose constrained to less than 50 Gy and objective functions that limit the dose to the heart, both lungs and the oesophagus. A software system using Pinnacle[3] scripting and Matlab was implemented to automatically generate, save and analyze the dose–volume histogram (DVH) data of multiple IMRT plans. Using this application it was possible to view all individual DVHs and to analyze and compare different plan evaluation criteria for all plans in the library.

The starting point for building the library of plans was the plan of the initial IMRT class solution. Subsequently, the IMRT parameters (weights and dose levels) of all objective functions, including the PTV, were kept constant and only parameters of the objective functions of the lungs and the oesophagus were varied. The range in which the parameters were varied was chosen such that a broad, but clinically relevant range of IMRT plans was generated.

The objective function for each lung was a DVH objective (10% volume, 20 Gy), of which the weights were varied simultaneously from 0 to 100. For the oesophagus an objective function was used aiming to keep the equivalent uniform dose ($a = 17$) below a prescribed dose level. The dose level was varied between 12 and

42 Gy. For the cord a maximum dose constraint of 45 Gy was taken. For the heart (minus PTV + 18 mm margin) a maximum mean dose of 25 Gy was used.

Plan evaluation

The plans were evaluated based on criteria derived from the DVH. In the analysis software, the user can choose from several plan evaluation criteria for both quantification of the PTV coverage and the organ at risk dose. Only the criteria used in this paper are discussed below. Target coverage is measured using the D_{99}, defined as the minimum dose received by 99% of the volume of the PTV (99% of the volume was chosen instead of 100% to make the calculation more robust). For adequate coverage of the PTV we require that D_{99} is larger than 95% of the prescribed dose level.

For the oesophagus, a recent meta-analysis by Rose et al. [13] shows that there is little consistency in the literature on parameters correlated with acute esophagitis. In the current study we consider the mean oesophagus dose (MD_E) as a surrogate parameter for toxicity of the oesophagus. However the principle of the method used in this study is equally valid when using for instance V_{35}. There are a large number of studies on the complication probability of the lung [14–16], with usually radiation pneumonitis as endpoint. Generally either dose or dose–volume based parameters have been used. In agreement with our clinical practice, we estimate the toxicity using the V_{20} as a surrogate parameter.

Results

The class solution for IMRT of NSCLC consisted of 6 equidistant coplanar photon beams (10 MV), arranged on the ipsilateral side. The objective functions and parameters of the IMRT plan were optimized iteratively to generate a clinically acceptable dose distribution. In Fig. 1 (solid line) the DVHs are shown for one of the patients. The D_{99} of the PTV is equal to 64.5 Gy, which is well above 95% of the prescribed dose of 66 Gy, and shows the excellent coverage that can be achieved using IMRT.
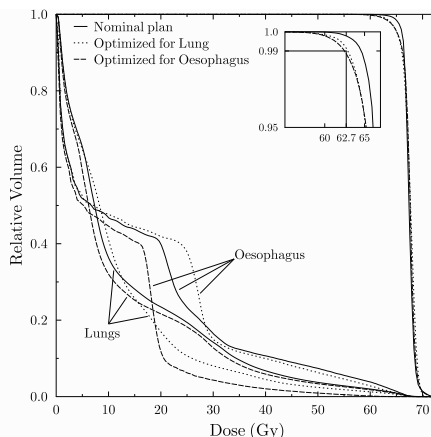
Fig. 1. DVH curves for three different IMRT plans of the same patient (see legend). For the plan optimized for lung sparing, the weight of the lung objective was increased from 1 to 50. For the plan optimized for oesophagus sparing, the dose level of the oesophagus objective function was lowered from 42 to 18 Gy (see text). The inset shows the DVH enlarged around 62.7 Gy (95% of the prescribed dose).

For all five patients in this study the dosimetric constraints to the heart and the spinal cord were easily met, indicating that the objective functions of these OARs are sufficient to achieve the planning criteria. Furthermore, changing the spinal cord and heart IMRT parameters by moderate amounts did not significantly change the results in terms of PTV coverage and dose to the oesophagus or lung. This indicates that for these cases the spinal cord and heart objectives are not strongly competing with the coverage of the PTV and with other OARs. We, therefore, focussed on the trade-off between the coverage of the PTV, the dose to the lungs, and the dose to the oesophagus.

To illustrate the tradeoffs and to show that the initial plan is only one of many possible solutions, the IMRT parameters were changed (see Fig. 1). The two plans have almost identical coverage of the PTV (D_{99} = 62.5 Gy and 63.1 Gy) but have either reduced dose to the lung (V_{20} = 0.19) or reduced dose to the oesophagus (MD_E = 10.9 Gy). Using the software described above, a series of 36 optimizations were done for this case and DVH parameters were extracted. Total computation time for this series was approximately 3 h.

In Fig. 2 the D_{99} of the PTV is plotted versus the weight of the lung objective that was used to generate the plan. We see that D_{99} is reduced with increasing the weight of the lung objective due to the competition of the coverage of the PTV with the sparing of the OAR. Reducing the dose level of the oesophagus objective function has a negative effect on the D_{99} of the PTV for the same reason. It is also clear that the dependence is not continuous and appears 'noisy' (see Discussion).

From a clinical point of view the relationship between the D_{99} and the weights of the OAR objective functions is not relevant. More interesting are the relationships between plan evaluation criteria, as shown in Fig. 3A, where the relationship between D_{99} of the PTV and V_{20} of the lung is shown for the same dataset as in Fig. 2.

In contrast to Fig. 2, the relationship between the coverage parameter D_{99} and the V_{20} appears to be approximately linear. A similar graph for the mean dose to the oesophagus is shown in Fig. 3B, also showing an approximately linear relationship between MD_E and D_{99}.
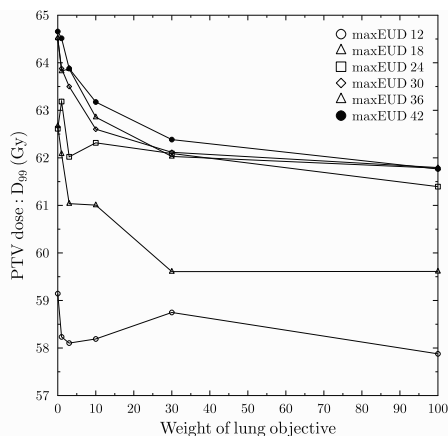
Fig. 2. DVH data for 36 different IMRT plans with different parameters for the lung and oesophagus objectives (see text). Plotted is the D_{99} of the PTV as a function of the weight of the lung objective. Calculations were done for six different MaxEUD values of the oesophagus objective function, ranging from 12 Gy to 42 Gy in intervals of 6 Gy.
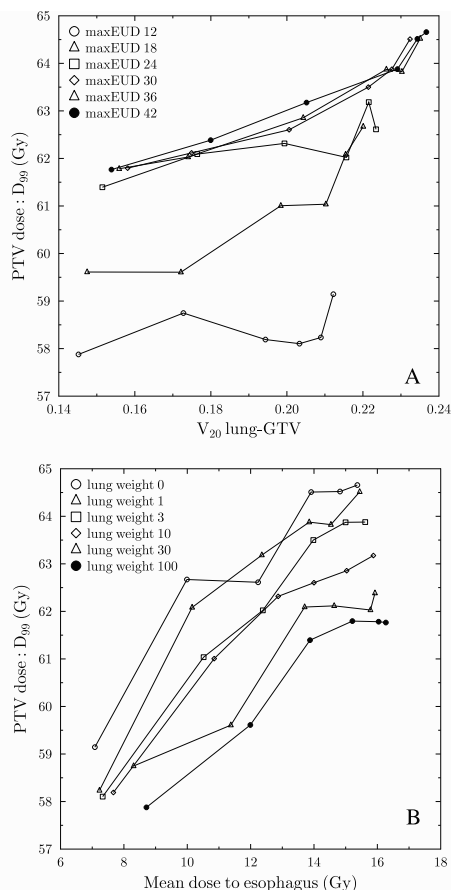
Fig. 3. Based on the same data as Fig. 2. Each data point represents an IMRT plan with different settings of the lung and oesophagus objectives (see text). (A) The D_{99} of the PTV versus the V_{20} of the total lung-GTV. (B) The D_{99} of the PTV versus the mean dose to the oesophagus.

Given the observed relationship between the PTV and OAR evaluation criteria, D_{99} was approximated using the following (first order) function:

$$D_{(99)} = a \cdot MD_E + b \cdot V_{20} + c \qquad (1)$$

Here D_{99} is the minimum dose received by 99% of the PTV volume, MD_E is the mean dose to the oesophagus and V_{20} is the volume fraction of the lung receiving 20 Gy. The coefficients a, b and c are parameters that define a linear relationship between the plan evaluation criteria for a particular case. Eq. (1) was used to fit the 'surface' of the D_{99} versus V_{20} and MD_E using a standard non weighted linear least square fit. Residuals show a deviation smaller than 0.8 Gy (1σ) for D_{99} for all patients. Given the approximation in Eq. (1) it is possible to generate a figure that presents the relationship between the plan evaluation criteria in a clear and intuitive way.

In Fig. 4, each dot represents a plan from the library. The coordinates of a dot indicate the OAR evaluation criteria of the plan. The diagonal lines are 'iso-coverage' lines, indicating the approximate value of D_{99} of the PTV based on the fit using Eq. (1). From this graph we can directly identify the region of OAR plan evaluation criteria that can be realized while maintaining a certain minimum coverage criterion. For instance, all plans where 99% of the volume receives a minimum of 95% of the prescribed dose are located in the triangle in the top right corner above the diagonal dashed line that describes the 62.7 Gy iso-coverage line. A number of plans have a $D_{99} > 62.7$ Gy, and within that group of acceptable plans, the trade-off between the oesophagus mean dose and the lung V_{20} can be varied.

Fig. 5 shows that the observed graphs are strongly patient dependent. For some patients the region of plans with $D_{99} > 62.7$ Gy can be very small (patient 3) whereas for other patients the choice is much larger (patient 4). The slope of the lines indicates the 'balance' in the trade-off between both OAR evaluation criteria. For patient 5 for instance, the PTV coverage competes stronger with the lung V_{20} than with the oesophagus mean dose. Associated with each dot in the graph is a set of IMRT parameters (weight and dose level of the oesophagus objective functions) that can be directly used in the TPS to yield the associated plan. This makes clinical implementation of the selected treatment plan straightforward. As a reference the plan evaluation parameters of the original plans are shown in the graph. In general the clinical plans are better than the 'lower limit' for coverage. Based on the graphs, the clinician could decide to trade some coverage for sparing of the OAR.

Discussion

In the design of a radiotherapy treatment plan, there is an inevitable conflict between the proper treatment of the designated target volume and sparing of the surrounding normal tissue and critical organs. For inverse planning of IMRT, the clinical aims
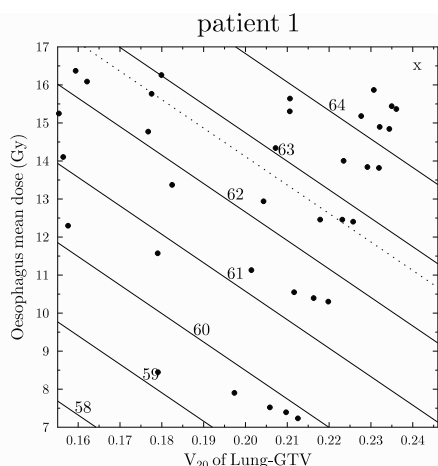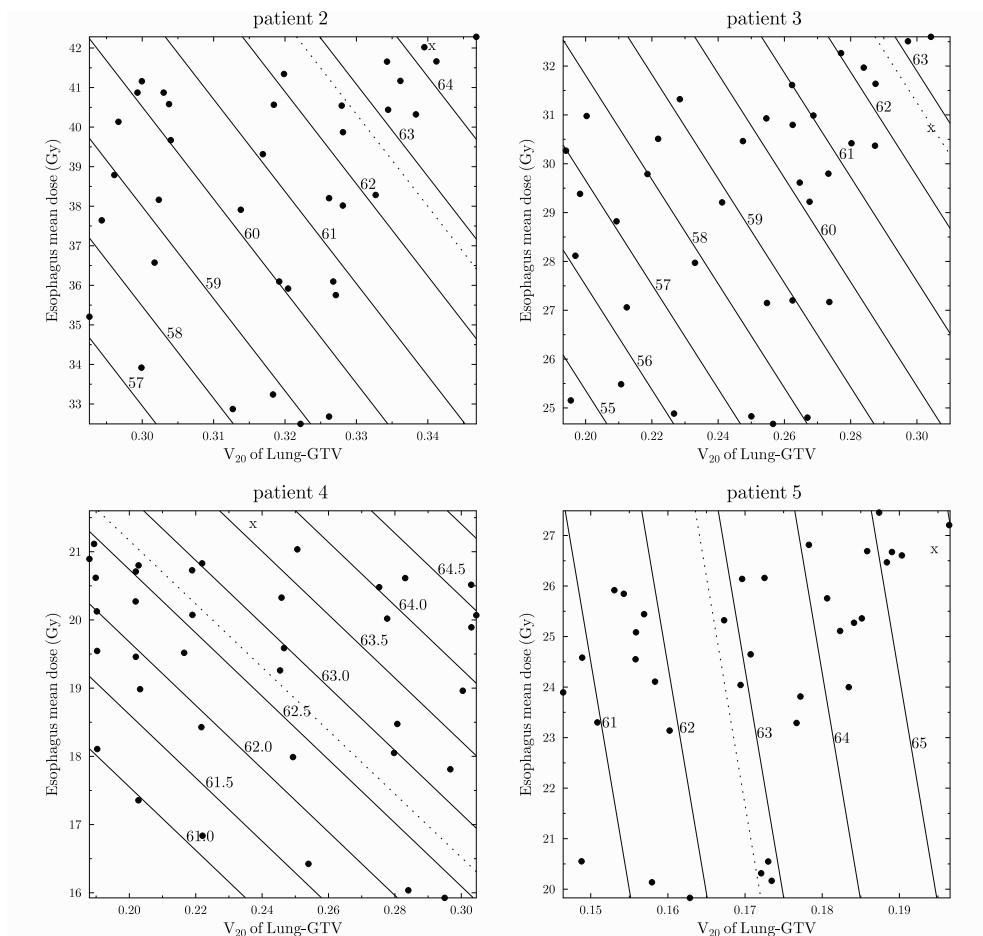
and restrictions for the plan are formulated in the form of mathematical objective functions and their parameters (weights and dose levels). These functions reflect how well the treatment goals and restrictions are satisfied. The optimization module of the treatment planning system subsequently finds a single optimal compromise between the different objectives. Due to the complex nature of the optimization problem no clear relationship exists between these mathematically formulated objective functions and the actual plan it results in. This makes it difficult to gain insight in the (quantitative) tradeoffs that are inherent to the IMRT problem. Therefore the subsequent manual optimization of the IMRT plan to achieve a more clinically acceptable solution can be time consuming. This problem has been recognized by others [1–3,5–9,17] and one of the most attractive solutions is to investigate a large range of possible IMRT plans by building an offline 'library' of possible IMRT plans.

In this paper a practical tool has been developed to investigate and represent the tradeoffs. To achieve this, the IMRT planning problem was simplified by limiting it to one tumour site only and by focusing on the two most relevant organs at risk. Note that this does not imply that the other organs at risk are not a part of the set of objective functions, however, the 'weight' with which they contribute to the multi-objective optimization is kept constant. For the patients in this study this resulted in satisfactory dose levels for all other organs at risk. Furthermore, the library generation is done using a standard, clinically used TPS including segmentation of the fluence into deliverable beam segments. Therefore, the IMRT plan chosen from the library can directly be used to treat the patient.

The multi-objective treatment planning concept for IMRT has been studied from a more mathematical point of view by several other investigators in the past few years [4,7,12]. The emphasis of these studies is on the Pareto surface in objective space. Generally these objectives do not coincide with the parameters that are used to evaluate the plan. For example, minimum, maximum and uniformity objective functions are used as objective functions during IMRT optimization to ensure proper coverage of the PTV, but subsequently the plan evaluation is conducted in terms of the D_{99}, which is considered a clinically relevant criterion for the PTV. Therefore in this study we focussed on the Pareto surface in the plan evaluation space. A drawback is that the plan is not guaranteed to be Pareto optimal with respect to the plan evaluation parameters. The latter could be solved by having objective functions that are identical to the evaluation parameters. However, in practice, this is not always possible, either because multiple objectives are used for one 'goal' or because the objective functions are not present in the optimizer software. For the simplified two-dimensional analysis of the IMRT plans in this paper, the relationship between the clinically relevant plan evaluation criteria was found to be close to linear (see Eq. (1)). Based on this linear relationship, a very simple and intuitive way of presenting the tradeoffs between the plan evaluation criteria in the form of a simple graph was developed (Figs. 4 and 5). In this way, the tradeoffs can be clearly presented to the user. The graphs in Fig. 5 show that the tradeoffs between the OAR dose and the PTV coverage are strongly patient dependent. Using this type of graph, an individualized choice for the clinically optimal treatment plan can be made by the clinician. It is evident that for the final selection of a treatment plan, the associated DVH and 3D dose distribution are taken into account.

The linear fit used to model the output parameters (i.e., the plan evaluation criteria) constitutes a first-order approximation and can be seen as a meta-model for visualization and optimization of an individualized treatment plan. The linear relation found is related to the fact that only a small region of the solution space is described, and that generally the Pareto surface is relatively smooth

patient 1

Fig. 4. Plot showing the OAR plan properties of all plans generated during the simulation (dots). The diagonal lines are iso-coverage lines connecting the points with equal D_{99} based on the linear model given in Eq. (1). The dashed line represents the 62.7 Gy iso-coverage that satisfies the norm that $D_{99} > 95\%$ of the prescribed dose. The x in the top right corner denotes the original (clinical) plan parameters.

Fig. 5. Graphs showing the trade-off between the OAR plan evaluation criteria and the coverage of the PTV for patients 2–5 (see Fig. 4 and the text for details). Clearly the graphs are strongly dependent on the patient and tumour geometry.

[9,10]. To test whether the found linear relation is specific for lung tumors, simulations were done for prostate cases as well. The D_{99} of the PTV was fitted as a function of the mean dose to the rectal-wall and the mean dose to the bladder. These simulations show that for all patients the relationship is again roughly linear (standard deviation of about 0.5 Gy of the D_{99}, indicating that the linear relationship is not specific for lung tumors, although for each tumor site and patient the quality of the fit should be verified.

Different from work by others [4,6–9,11], we used a treatment planning optimization algorithm that includes segmentation of the beam into deliverable segments as well as use of non-convex objective functions. The use of non-convex objective functions is commonly avoided in multi-objective optimization, since the optimization algorithm is not guaranteed to reach the global optimum. However, non-convex functions were used in this paper, since we

find that for our TPS this results in better IMRT plans. Segmentation of the beam into deliverable segments is used since in that case the calculated dose distribution is representative of the clinical situation. The consequence of this choice is shown in Figs. 2 and 3 illustrating the noise-like behaviour of the plan evaluation criteria. The noise like behaviour is caused by the fact that segmentation of the fluence and the use of non-convex objective functions can cause the optimizer to get stuck in local minima [18]. Therefore, the prediction by the linear model is only accurate to within ~1 Gy, and much more than the minimum of 3 generated plans is necessary to properly estimate the coefficients of the linear fit presented in Eq. (1). However, we believe that the gain in plan quality and ease of integration in the clinic justifies this choice. Furthermore, these effects are not specifically related to the method described in this paper and are inherent to the use of this TPS

also when using the normal planning procedure. To limit the computation time, the library generation was limited to only two dimensions. Since the actual parameter space has more than 20 dimensions, care should be taken to start the analysis with a plan which is already close to optimal. The library can be seen as an exploration of the (two dimensional) region 'around' this plan, where actually Eq. (1) constitutes a two-dimensional sensitivity analysis of the region surrounding this point.

The work in this paper is related to work by Ottoson et al. [10] and Hunt et al. [19]. In their work a (commercial) TPS is used to systematically review the relationship between the objective space and the final plan criteria to evaluate class solutions, but also for instance the influence of segmentation on the resulting plans. The method presented in this paper, however, is specifically aimed at providing a method to aid the planner and clinician on a patient specific basis with choosing the optimal IMRT plan.

The analysis presented was limited to one type of tumour (tumour site) and to only two OARs, enabling the straightforward graphical representation of the tradeoffs as shown in Figs. 4 and 5. Since only a limited region of the solution space is scanned, the initial class solution should be chosen correctly. For the lung tumors we find that the same class solution can be applied in most cases. For more complex cases (such as head and neck cancer) this might be different, and in this case the two dimensional 'scan' is insufficient. This will be subject of further study.

Conclusion

A method was developed where a standard TPS is used to generate a library of deliverable IMRT plans that are used to present the inherent tradeoffs to the user in terms of clinically relevant plan evaluation criteria. Furthermore, for the two dimensional case studied in this paper, the relationship between the plan evaluation criteria is close to linear, making it possible to present the trade-off in a clear and intuitive manner.

Acknowledgement

We thank P. van Kollenburg for developing the initial 'class solution' used in this study.

References

[1] Xing L, Li JG, Donaldson S, Le QT, Boyer AL. Optimization of importance factors in inverse planning. Phys Med Biol 1999;44:2525–36.
[2] Jee KW, McShan DL, Fraass BA. Lexicographic ordering: intuitive multicriteria optimization for IMRT. Phys Med Biol 2007;52:1845–61.
[3] Breedveld S, Storchi PR, Keijzer M, Heemink AW, Heijmen BJ. A novel approach to multi-criteria inverse planning for IMRT. Phys Med Biol 2007;52:6339–53.
[4] Craft DL, Halabi TF, Shih HA, Bortfeld TR. Approximating convex pareto surfaces in multiobjective radiotherapy planning. Med Phys 2006;33:3399–407.
[5] Rosen I, Liu HH, Childress N, Liao Z. Interactively exploring optimized treatment plans. Int J Radiat Oncol Biol Phys 2005;61:570–82.
[6] Craft D, Halabi T, Shih HA, Bortfeld T. An approach for practical multiobjective IMRT treatment planning. Int J Radiat Oncol Biol Phys 2007;69:1600–7.
[7] Craft D, Halabi T, Bortfeld T. Exploration of tradeoffs in intensity-modulated radiotherapy. Phys Med Biol 2005;50:5857–68.
[8] Thieke C, Kufer KH, Monz M, et al. A new concept for interactive radiotherapy planning with multicriteria optimization: first clinical evaluation. Radiother Oncol 2007;85:292–8.
[9] Craft D, Bortfeld T. How many plans are needed in an IMRT multi-objective plan database? Phys Med Biol 2008;53:2785–96.
[10] Ottosson RO, Engstrom PE, Sjostrom D, et al. The feasibility of using Pareto fronts for comparison of treatment planning systems and delivery techniques. Acta Oncol 2009;48:233–7.
[11] Hoffmann AL, Siem AY, den Hertog D, Kaanders JH, Huizenga H. Derivative-free generation and interpolation of convex Pareto optimal IMRT plans. Phys Med Biol 2006;51:6349–69.
[12] Siem AYD, den Hertog D, Hoffmann AL. The effect of transformations on the approximation of univariate (convex) functions with applications to Pareto curves. Eur J Operational Res 2008;189:347–62.
[13] Rose J, Rodrigues G, Yaremko B, Lock M, D'Souza D. Systematic review of dose–volume parameters in the prediction of esophagitis in thoracic radiotherapy. Radiother Oncol 2009;91:282–7.
[14] Semenenko VA, Li XA. Lyman-Kutcher-Burman NTCP model parameters for radiation pneumonitis and xerostomia based on combined analysis of published clinical data. Phys Med Biol 2008;53:737–55.
[15] Kwa SL, Theuws JC, Wagenaar A, et al. Evaluation of two dose–volume histogram reduction models for the prediction of radiation pneumonitis. Radiother Oncol 1998;48:61–9.
[16] Theuws JC, Kwa SL, Wagenaar AC, et al. Dose-effect relations for early local pulmonary injury after irradiation for malignant lymphoma and breast cancer. Radiother Oncol 1998;48:33–43.
[17] Hoffmann AL, den Hertog D, Siem AY, Kaanders JH, Huizenga H. Convex reformulation of biologically-based multi-criteria intensity-modulated radiation therapy optimization including fractionation effects. Phys Med Biol 2008;53:6345–62.
[18] Carlsson F, Forsgren A. Iterative regularization in intensity-modulated radiation therapy optimization. Med Phys 2006;33:225–34.
[19] Hunt MA, Hsiung CY, Spirou SV, et al. Evaluation of concave dose distributions created using an inverse planning system. Int J Radiat Oncol Biol Phys 2002;54:953–62.

VIII

Convex reformulation of radiobiological optimisation for IMRT planning

Convex reformulation of biologically-based multi-criteria intensity-modulated radiation therapy optimization including fractionation effects

Aswin L Hoffmann[1]**, Dick den Hertog**[2]**, Alex Y D Siem**[2]**, Johannes H A M Kaanders**[1] **and Henk Huizenga**[1]

[1] Department of Radiation Oncology, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands
[2] Department of Econometrics and Operations Research/Center for Economic Research (CentER), Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

E-mail: a.hoffmann@rther.umcn.nl

Abstract
Finding fluence maps for intensity-modulated radiation therapy (IMRT) can be formulated as a multi-criteria optimization problem for which Pareto optimal treatment plans exist. To account for the dose-per-fraction effect of fractionated IMRT, it is desirable to exploit radiobiological treatment plan evaluation criteria based on the linear-quadratic (LQ) cell survival model as a means to balance the radiation benefits and risks in terms of biologic response. Unfortunately, the LQ-model-based radiobiological criteria are nonconvex functions, which make the optimization problem hard to solve. We apply the framework proposed by Romeijn *et al* (2004 *Phys. Med. Biol.* **49** 1991–2013) to find transformations of LQ-model-based radiobiological functions and establish conditions under which transformed functions result in equivalent convex criteria that do not change the set of Pareto optimal treatment plans. The functions analysed are: the LQ-Poisson-based model for tumour control probability (TCP) with and without inter-patient heterogeneity in radiation sensitivity, the LQ-Poisson-based relative seriality *s*-model for normal tissue complication probability (NTCP), the equivalent uniform dose (EUD) under the LQ-Poisson model and the fractionation-corrected Probit-based model for NTCP according to Lyman, Kutcher and Burman. These functions differ from those analysed before in that they cannot be decomposed into elementary EUD or generalized-EUD functions. In addition, we show that applying increasing and concave transformations to the convexified functions is beneficial for the piecewise approximation of the Pareto efficient frontier.

139

1. Introduction

It is generally acknowledged that designing fluence maps for high-energy photon beams in intensity-modulated radiation therapy (IMRT) can be posed as a constrained optimization problem (Börgers 1997, Küfer *et al* 2003, Reemtsen and Alber 2006). The aim of the optimization problem is to find beam fluence maps that deposit a sufficiently high dose to the target volume and simultaneously spare organs at risk (OARs) and other surrounding normal tissue as much as possible. A mathematical fluence map optimization (FMO) model is required to guide the inverse treatment planning process through the search space of possible solutions. Such models are typically based on a set of conflicting treatment plan evaluation criteria that reflect how well the treatment goals and restrictions are satisfied.

These criteria can either be formulated as physical criteria, i.e. on measurable physical quantities like doses and volumes, or as biological criteria that reflect the responses of the different tissues to dose distributions (Brahme 1995). Physical criteria are often implemented as a dose-dependent function calculating the mean-squared deviation from a prescribed target dose level. The radiobiological rationale of these criteria is questionable since positive and negative deviations from the target dose have different biological consequences, but are weighted equally using a quadratic penalty function. In this paper, we focus on biological criteria only, because they have been shown to predict the response of tumours and healthy tissues more adequately than physical criteria by taking the underlying radiation biology into account. The relevance of biological treatment goals, leading to nonlinear criteria as tumour control probability (TCP), normal tissue complication probability (NTCP), (generalized) equivalent uniform dose ((g)EUD) or dose–volume based criteria, has generally been acknowledged during the past years (Niemierko 2005). Formulation of clinically relevant radiobiological FMO problems often requires the inequality constraints to be formulated in terms of clinically prescribed tolerance bounds, e.g., NTCP $\leqslant \epsilon$, where ϵ is an acceptable probability of injury.

Direct application of such nonlinear and nonconvex biological criteria as objective and constraint functions can make the FMO problem very hard to solve, especially when the number of optimization variables (i.e., the beamlet weight intensities) is large. According to Hindi (2004) various sources exist for the difficulties encountered in large-scale nonconvex optimization. First, the search space may be riddled with multiple local optima. Second, the feasible set could be empty. Third, stopping criteria used in general optimization algorithms are often arbitrary. Fourth, the problem may be degenerate, in which case the same optimal value is attained for multiple solutions (see, e.g., Alber *et al* (2002)). Fifth, optimization algorithms might have poor convergence rates. Sixth, numerical problems could cause the optimization algorithm to stop or roam.

Fortunately, in the case of minimization, it is known that if all objective functions and upperbound-inequality constraint functions are *convex*, the first three difficulties disappear: any local minimum is necessarily a global minimum; the solution set is convex and therefore comprises either a single solution or is infinite; and very precise stopping criteria are available using *duality* (see, e.g., Bertsekas (1999)). However, convergence rate and numerical stability remain a potential problem. If, in addition to convexity, the objective and constraint functions allow for *self-concordant* barriers, the issues of convergence and numerical sensitivity could be avoided using interior point methods (Nesterov and Nemirovski 1994). Hence, it is possible to solve a large class of convex optimization problems with great efficiency using local solvers like gradient-based algorithms. Convexity of the criteria is also sufficient to guarantee that the Pareto efficient frontier is convex in case the FMO model is multi-criteria (Romeijn *et al* 2004). This is of particular interest in case piecewise linear approximation techniques like

Sandwich-type algorithms are used to estimate the Pareto efficient frontier (Craft *et al* 2006, Hoffmann *et al* 2006).

Radiobiological criteria like TCP and NTCP are often sigmoidal functions of dose and hence are inherently nonlinear and nonconvex. Consequently, their direct implementation necessarily leads to nonconvex optimization problems. Recently, Romeijn *et al* (2004) presented a unifying framework for multi-criteria fluence map optimization problems that establishes conditions under which well-known nonconvex radiobiological treatment plan evaluation criteria can be transformed into convex criteria while preserving the set of Pareto efficient solutions. In particular, they showed that transformations of criteria such as TCP, NTCP and sigmoidal functions of (g)EUD exist that are equivalent to criteria formulated in terms of (g)EUD only, concluding that only two distinct Pareto efficient frontiers exist. Others have explored the convexity properties of transformed radiobiological treatment plan evaluation criteria as well (Deasy 1997, Choi and Deasy 2002). However, to the best of our knowledge the majority of convexity analyses reported so far are limited to simple, single-hit *linear* cell survival models that do not take into account fractionation effects, whereas in clinical practice almost all radiation treatments are delivered over multiple fractions.

As an extension of the work by Romeijn *et al* (2004) we explicitly include the dose-per-fraction effect in the convexity analysis of commonly occurring radiobiological criterion functions by using the *linear-quadratic* (LQ) cell survival model (see, e.g., Fowler (1989)). We analyse TCP, NTCP and EUD criteria that are not related to the elementary (linear-Poisson) EUD or (power-law) gEUD model, and thus yield different Pareto efficient frontiers from the criteria analysed by Romeijn *et al* (2004). More specifically, we establish transformations of the following criterion functions and investigate the conditions under which these criteria are *strictly* convex/concave depending on the criterion to be minimized/maximized, respectively: the LQ-Poisson model based TCP function, LQ-Poisson model based relative seriality NTCP function, LQ-Poisson model based EUD function and the fractionation-corrected Probit-model-based NTCP function according to Lyman, Kutcher and Burman. Strict convexity/concavity of the objective function is an interesting property, as it is a sufficient condition for the existence of a unique solution (i.e., minimizer/maximizer) that rules out the possibility of multiple local optima (i.e., minima/maxima) (see, e.g., Bertsekas (1999)). Furthermore, we present a detailed convexity analysis of the (transformed) population-averaged TCP function that incorporates inter-patient radiosensitivity heterogeneity, for which Choi and Deasy (2002) could not rule out the possibility of multiple local minima.

2. Mathematical definitions

2.1. Multi-criteria optimization for IMRT

To introduce the problem, we confine to the essential mathematics and refer to Hoffmann *et al* (2006) for a more comprehensive enunciation. In IMRT optimization, the dose distribution $\mathbf{D} = (D_1, \ldots, D_M)$ is a linear mapping of the bixel weights $\mathbf{w} = (w_1, \ldots, w_N)$ by the dose deposition operator $\mathbf{P}$,

$$\mathbf{D}(\mathbf{w}) = \mathbf{P} \cdot \mathbf{w}.$$

Hence, the dose in voxel i can be denoted as a weighted sum over all N bixel weights

$$D_i(\mathbf{w}) = \sum_{j=1}^{N} P_{ij} w_j,$$

where P_{ij} is the dose deposited in voxel i from bixel j at unit intensity. This represents the discretized form of the Fredholm integral equation of the first kind that is commonly

encountered in solving the inverse problem of radiation treatment planning (Lind 1990). The elements of **P** describe the physics of the beam–tissue interaction, and can be pre-calculated using various dose calculation algorithms.

Given the matrix **P**, it is the aim of the optimizer to find a suitable vector **w** that satisfies the optimization goals. Typically, the goals are formulated by a set of *objective functions* $F_k : \mathbb{R}_+^M \mapsto \mathbb{R}_+, k = 1, \ldots, K$, where K is the number of objective functions. The *constraint functions* $B_l : \mathbb{R}_+^M \mapsto \mathbb{R}_+, l = 1, \ldots, L$, where L is the number of constraint functions, define certain equalities and inequalities that the design variable **w** and the dose distribution **D(w)** must satisfy. Each criterion function quantifies the plan-evaluation score as a function of the dose distribution **D(w)**. Without loss of generality, it is assumed that lower values are preferred to higher values for each of the criterion functions.

Typically, the constrained multi-criteria optimization problem in IMRT is formulated as

$$
\min_{\mathbf{w}} \quad \mathbf{F}(\mathbf{D}(\mathbf{w})) = \begin{pmatrix} F_1(\mathbf{D}(\mathbf{w})) \\ F_2(\mathbf{D}(\mathbf{w})) \\ \vdots \\ F_K(\mathbf{D}(\mathbf{w})) \end{pmatrix}
$$

$$
\begin{aligned}
\text{s.t.} \quad & B_l(\mathbf{D}(\mathbf{w})) \leqslant 0, \qquad l = 1, \ldots, \mathrm{L} \\
& \mathbf{D}(\mathbf{w}) = \mathbf{P} \cdot \mathbf{w} \\
& \mathbf{w} \geqslant 0.
\end{aligned}
\tag{1}
$$

Since problem (1) is associated with multiple solutions, the concept of *Pareto efficiency* (see, e.g., Miettinen (1999)) is applied to restrict to solutions that have the property that no single objective value can be improved without deteriorating at least one other objective value. Solutions that comply with this definition are called *Pareto optimal* (also called Edgeworth–Pareto optimal, efficient, nondominated or noninferior). In the objective space, the set of Pareto optimal plans is mapped to the *Pareto efficient frontier*. Since the dimension of the objective space is much less than that of the solution space, this frontier is used by the decision maker to navigate efficiently through the set of solutions and select a compromise solution that best meets with his/her approval. Therefore, it is important to have an algorithm that generates a discrete subset of Pareto optimal solutions that forms a representative estimate of the true Pareto efficient frontier (Craft *et al* 2006, Hoffmann *et al* 2006).

2.2. Commonly used radiobiological treatment plan evaluation criteria

Several mathematical models have been developed to describe the dose–response relationship for tumours and normal tissues. We analyse a commonly used mechanistic TCP model and two well-known NTCP models; a mechanistic and a phenomenological one.

2.2.1. Mechanistic dose–response relationship based on Poisson model. To calculate the dose response to a heterogeneous dose distribution, the internal infrastructural organization (i.e., volume dependence) of the irradiated tissue is taken into account.

For eradication of all clonogenic cells of a tumour, every individual tumour element has to be eradicated. Hence, the tumour control probability (TCP) is the product of all individual responses. Based on the well-known linear-quadratic (LQ) Poisson model without cell repopulation in time (Fowler 1989), the TCP function is denoted as

$$
\mathrm{TCP}_{\mathrm{LQ}}(\mathbf{D}) = \exp\left[-N_0 \sum_{k=1}^{N} v_k \mathrm{SF}_{\mathrm{LQ}}(D_k) \right],
\tag{2}
$$

where $\mathrm{SF}_{\mathrm{LQ}}(D_k) = \mathrm{e}^{-\alpha D_k - \beta D_k{}^2/n}$ is the *surviving fraction* of cells exposed to dose D_k in voxel k, v_k is the relative volume of voxel k, N_0 is the total initial number of clonogenic cells, α is the *intrinsic radiosensitivity* representing the nonrepairable radiation damage, β represents a repairable type of injury that is responsible for the *dose-per-fraction effect* and n is the number of treatment fractions. The α/β ratio of the LQ-model then determines the tissue's sensitivity to alterations in radiation fraction size (Barendsen 1982).

The relative seriality s-model is the most well-known mechanistic normal tissue dose–response model that is based on the LQ-Poisson survival function (Källman *et al* 1992). For an OAR irradiated with a heterogeneous dose distribution the model is expressed as

$$\mathrm{NTCP}_{\mathrm{RS}}(\mathbf{D}) = \left[1 - \prod_{k=1}^{N}[1 - \mathrm{P}_{\mathrm{LQ}}(D_k)^s]^{v_k} \right]^{\frac{1}{s}}, \tag{3}$$

where $\mathrm{P}_{\mathrm{LQ}}(D_k) = \exp(-N_0 \cdot \mathrm{SF}_{\mathrm{LQ}}(D_k))$ is the Poisson approximation of the binomial probability that no cells survive the dose D_k, and s is the relative seriality parameter that characterizes the internal organization of a tissue. A value of $s \approx 0$ represents a largely parallel organ (e.g., lung, parotid or liver), whereas $s \approx 1$ corresponds to a serial organ (e.g., spinal cord or oesophagus).

Response for patient population with inter-tumoural variation of radiation sensitivity. The LQ-Poisson-based TCP function of (2) describes the tumour response for an individual patient. However, the clinically measured dose response is a population-averaged dose response, which differs from the individual response due to inter-individual variability in radiobiological characteristics. A population-averaged TCP function, $\mathrm{TCP}_{\mathrm{pop}}$, can be constructed by averaging an individual TCP function, $\mathrm{TCP}_{\mathrm{ind}}$, over the range of parameters found in a population (see, e.g., Roberts and Hendry (1998)). Assuming that only the radiosensitivity α is subject to variations, the population TCP function can be expressed as

$$\mathrm{TCP}_{\mathrm{pop}}(\mathbf{D}) = \int_0^{\infty} \phi(\alpha)\mathrm{TCP}_{\mathrm{ind}}(\mathbf{D}, \alpha) \, \mathrm{d}\alpha, \tag{4}$$

where $\phi(\alpha)$ is the probability density function of the parameter α.

For $\phi(\alpha)$ often a normal probability density function has been used, assuming the variation of β to be correlated by a constant value of α/β (e.g., Sanchez-Nieto and Nahum (2000)). Others have applied the log–normal distribution to limit the parameter range of α to biologically meaningful values (Keall and Webb 2007). We analyse both the normal and log–normal averaged $\mathrm{TCP}_{\mathrm{pop}}$ function for log-concavity.

LQ-model-based equivalent uniform dose. The concept of *equivalent uniform dose* (EUD) was first introduced by Niemierko (1997) by equating the TCP of (2) to the TCP of an equivalent homogeneous distribution, i.e., $\mathrm{TCP}_{\mathrm{LQ}}(\mathbf{D}) \equiv \mathrm{TCP}(\mathrm{EUD}_{\mathrm{LQ}}(\mathbf{D}))$, and solving for $\mathrm{EUD}_{\mathrm{LQ}}(\mathbf{D})$. It is easy to show (see, e.g., McGary *et al* (2000)) that EUD under the LQ-Poisson model can be expressed as

$$\mathrm{EUD}_{\mathrm{LQ}}(\mathbf{D}) = -\frac{1}{2}\frac{\alpha}{\beta}n \left[1 - \sqrt{1 - \frac{4\beta}{\alpha^2 n} \ln \overline{\mathrm{SF}}_{\mathrm{LQ}}(\mathbf{D})} \right], \tag{5}$$

with

$$\overline{\mathrm{SF}}_{\mathrm{LQ}}(\mathbf{D}) = \sum_{k=1}^{N} v_k \, \mathrm{e}^{-\alpha D_k - \beta D_k{}^2/n}.$$

Others have shown that the EUD under the (single-hit) linear-Poisson model is concave (Choi and Deasy 2002, Romeijn *et al* 2004). We analyse the concavity of (5) using a direct approach

and compare the result to (Bortfeld *et al* 2008), which has been established under an indirect approach using Hardy *et al* (1952, theorem 106, p 88).

2.2.2. Phenomenological dose–response relationship based on the Probit model. Assuming a normal distribution of intrinsic radiosensitivities, Lyman (1985) and Kutcher and Burman (1989) applied the Probit model to calculate the response of an OAR to a heterogeneous dose distribution according to

$$\text{NTCP}_{\text{LKB}}(\mathbf{D}) = \Phi\left(\frac{\text{gEUD}_a(\mathbf{D}) - D_{50}}{m\,D_{50}}\right), \tag{6}$$

where $\Phi(z) = 1/\sqrt{2\pi}\int_{-\infty}^{z} e^{-t^2/2}\,dt$ is the standard normal cumulative distribution function, D_{50} denotes the uniform dose where a 50% complication probability occurs, m determines the slope of the sigmoidal function Φ and $\text{gEUD}_a(\mathbf{D})$ is the *generalized equivalent uniform dose* (Niemierko 1999) of the dose distribution $\mathbf{D}$,

$$\text{gEUD}_a(\mathbf{D}) = \left[\sum_{k=1}^{N} v_k D_k{}^a\right]^{\frac{1}{a}}, \tag{7}$$

and $a \geqslant 1$ is a tissue-dependent parameter that describes the volume dependence.

To account for the dose-per-fraction effect of a multi-fraction treatment, the dose to each voxel is converted into an iso-effective *biologically effective dose* (BED) (Fowler 1989) prior to calculation of the gEUD,

$$\text{BED}(D_i) = D_i\left(1 + \frac{D_i/n}{\alpha/\beta}\right).$$

Subsequent calculation of the gEUD from the BED-converted dose distribution yields the *generalized equivalent uniform biologically effective dose* (gEUBED) function,

$$\text{gEUBED}_a(\mathbf{D}) = \text{gEUD}_a(\mathbf{BED}(\mathbf{D}))$$
$$= \left[\sum_{k=1}^{N} v_k \text{BED}(D_k)^a\right]^{\frac{1}{a}}. \tag{8}$$

The fractionation-corrected LKB model then becomes

$$\text{NTCP}_{\text{LKB}}(\mathbf{D}) = \Phi\left(\frac{\text{gEUBED}_a(\mathbf{D}) - \text{BED}_{50}}{m\text{BED}_{50}}\right), \tag{9}$$

where $\text{BED}_{50} = \text{BED}(D_{50})$.

2.2.3. Composite evaluation functions. For multiple targets and vital organs, the overall probability of benefit and injury is respectively given by

$$\text{TCP}(\mathbf{D}) = \prod_{j=1}^{T} \text{TCP}_j(\mathbf{D})^{\eta_j}, \tag{10}$$

and

$$\text{NTCP}(\mathbf{D}) = 1 - \prod_{j=1}^{S}[1 - \text{NTCP}_j(\mathbf{D})]^{\xi_j}, \tag{11}$$

where T and S are the number of targets and vital organs, respectively, $\text{TCP}_j(\mathbf{D})$ is the control probability for target j, η_j is the fractional volume of the jth target, $\text{NTCP}_j(\mathbf{D})$ is the

complication probability in organ j and ξ_j is the relative weight of the jth complication, such that $\sum_j^T \eta_j = \sum_j^S \xi_j = 1$. Here, $\text{TCP}_j(\mathbf{D})$ and $\text{NTCP}_j(\mathbf{D})$ can be any of the fractionation-corrected dose–response models from the preceding subsections.

Another composite objective function is the probability of uncomplicated tumour control, P_+, which is a measure to *a priori* balance between treatment benefit and injuries (see, e.g., Brahme (1995)). In generalized form it is defined as

$$P_+^\delta(\mathbf{D}) = (1 - \delta)P_+^0(\mathbf{D}) + \delta P_+^1(\mathbf{D}),$$

where $P_+^0(\mathbf{D}) \equiv \text{TCP}(\mathbf{D}) - \text{NTCP}(\mathbf{D})$, $P_+^1(\mathbf{D}) \equiv \text{TCP}(\mathbf{D})[1 - \text{NTCP}(\mathbf{D})]$ and δ is the fraction of patients for which tumour and normal tissue responses are statistically independent ($0 \leqslant \delta \leqslant 1$). $\delta = 0$ corresponds to the case that the events of tumour control and severe injury are totally correlated (i.e., no complications without achieving tumour control). For $\delta = 1$ it is assumed that tumour control is independent of complications. Note that $P_+^\delta(\mathbf{D})$ is a convex combination of $P_+^0(\mathbf{D})$ and $P_+^1(\mathbf{D})$.

3. Methods

3.1. Convex multi-criteria optimization

A general convex multi-criteria optimization problem is formulated in standard form as

$$
\min_{\mathbf{x}} \quad \mathbf{G}(\mathbf{x}) = \begin{pmatrix} G_1(\mathbf{x}) \\ G_2(\mathbf{x}) \\ \vdots \\ G_K(\mathbf{x}) \end{pmatrix}
\tag{12}
$$

$$
\text{s.t.} \quad C_l(\mathbf{x}) \leqslant 0, \qquad l = 1, \ldots, L
$$
$$
E_l(\mathbf{x}) = 0, \qquad l \in \mathcal{E}.
$$

Here $\mathbf{x} \in \mathbb{R}^M$ is the optimization variable, $G_k : \mathbb{R}^M \mapsto \mathbb{R}$ are convex *objective functions*, $C_l : \mathbb{R}^M \mapsto \mathbb{R}$ are convex *inequality constraint functions*, $E_i : \mathbb{R}^M \mapsto \mathbb{R}$ are affine *equality constraint functions* and $\mathcal{E}$ is the set of indices for equality constraints.

In the case of the FMO problem in IMRT (1), the optimization variable $\mathbf{x}$ is obtained from the dose deposition operator by a linear mapping of the design variable $\mathbf{w}$, which we control to generate fluence maps. If the only constraint on the treatment plan is the non-negativity of bixel weights, then the set $\mathcal{A}$ of allowed bixel weights is convex. Therefore, the set $\mathcal{R}$ of realizable dose distributions (being the image of $\mathcal{A}$ under the linear dose deposition operator $\mathbf{P}$) is convex. For the multi-criteria FMO problem to be cast as a convex problem (12), the objective and constraint functions of (1) should all be convex or transformed into equivalent convex functions.

Equivalence of optimization problem for transformed criterion functions. Direct application of the (nonconvex) criterion functions as objective and constraint functions would result in a nonconvex problem. Romeijn *et al* (2004) have shown that transforming any or all criterion functions of the general (nonconvex) minimization problem (1) via strictly increasing functions leads to an equivalent Pareto efficient frontier. This property can be used to convexify problem (1) such that it can be formulated as a convex problem (12). A sufficient condition is that strictly increasing transformations $h_k : \mathbb{R} \mapsto \mathbb{R}$ ($k = 1, \ldots, K$) and $\tilde{h}_l : \mathbb{R} \mapsto \mathbb{R}$ ($l = 1, \ldots, L$) exist, such that compositions

$$G_k(\mathbf{x}) = (h_k \circ F_k)(\mathbf{x}) \equiv h_k(F_k(\mathbf{x}))$$

and

$$C_l(\mathbf{x}) = (\tilde{h}_l \circ B_l)(\mathbf{x}) \equiv \tilde{h}_l(B_l(\mathbf{x}))$$

are convex functions.

In the following section, we find transformations for the radiobiological criterion functions described in section 2.2 and analyse the conditions under which the transformed criteria are (strictly) convex.

Although the criterion functions have been cast as a function of the dose distribution $\mathbf{D}$, rather than as a function of the bixel weights $\mathbf{w}$, the analysis is conducted in terms of $\mathbf{D}$. Formally, the objective functions $G_k(\mathbf{D})$ and $G_k(\mathbf{w}) = G_k(\mathbf{D}(\mathbf{w})) = G_k(\mathbf{P} \cdot \mathbf{w})$ are different, but since $\mathbf{D}$ is a linear mapping of $\mathbf{w}$, the convexity properties of the functions remain invariant. This is due to the fact that the composition of a convex function with an affine mapping is convex (see, e.g., Boyd and Vandenberghe (2004, p 79)). However, for G_k to be strictly convex with respect to $\mathbf{w}$, the mapping $\mathbf{P}$ should be full rank (and thus have an empty nullspace). This can be derived from the Hessian of G_k with respect to $\mathbf{w}$,

$$\nabla_{ww}^2 G_k(\mathbf{D}(\mathbf{w})) = \mathbf{P}^T \cdot \nabla_{DD}^2 G_k(\mathbf{D}) \cdot \mathbf{P},$$

and using proper statements of positive semidefiniteness from linear algebra. Using the properties of positive (semi)definite symmetric products (see, e.g., Horn and Johnson (1987, p 399)) it can be shown that if $\nabla_{DD}^2 G_k(\mathbf{D})$ is positive definite (and thus G_k is strictly convex in $\mathbf{D}$), and $\mathbf{P}$ is any particular dimensionally compatible matrix that has an empty nullspace, then $\nabla_{ww}^2 G_k(\mathbf{D}(\mathbf{w}))$ is positive definite (and thus G_k is strictly convex in $\mathbf{w}$). It also follows that, if $\mathbf{P}$ is of compatible dimension but rank deficient, then $\nabla_{ww}^2 G_k(\mathbf{D}(\mathbf{w}))$ is positive semidefinite (and thus G_k is convex in $\mathbf{w}$). Clearly, this also holds for the constraint functions C_l. In practice, the dose grid is finer than the bixel grid, i.e., $M \gg N$ and thus $\mathbf{P}$ has full column rank (Carlsson 2008).

3.2. Convexity analysis of transformed radiobiological criterion functions

In general, convexity of a criterion function $G : \mathbb{R}^M \mapsto \mathbb{R}$ can be analysed via different approaches. A direct approach is to check the conditions on its Hessian, $\nabla^2 G$. A necessary and sufficient condition for convexity of G is that its Hessian is positive semidefinite. Without proof, we state the conditions for convexity and concavity more formally:

Definition 1. *A matrix* $\mathbf{A} \in \mathbb{R}^{M \times M}$ *is positive (negative) semidefinite on a set* $\mathcal{S}$ *if* $\mathbf{x}^T \mathbf{A} \mathbf{x} \geqslant (\leqslant) 0$ *for all* $\mathbf{x} \in \mathcal{S}$, *and positive (negative) definite on* $\mathcal{S}$ *if* $\mathbf{x}^T \mathbf{A} \mathbf{x} > (<) 0$ *for* $\mathbf{x} \neq 0$ *and* $\mathbf{x} \in \mathcal{S}$.

Lemma 1. *A real-valued, twice differentiable function G is convex (concave) on a set $\mathcal{S}$ if and only if its Hessian, $\nabla^2 G$, is positive (negative) semidefinite on $\mathcal{S}$. The function G is strictly convex (concave) on the set $\mathcal{S}$ if and only if $\nabla^2 G$ is positive (negative) definite on $\mathcal{S}$.*

An indirect approach is to exploit the rules of *composition* to deduce convexity/concavity of differentiable functions by means of the chain rule. The following composition rules for convexity and concavity can be derived (see, e.g., Boyd and Vandenberghe (2004, p 83–7)).

Lemma 2. *Consider the functions $\mathbf{F} : \mathbb{R}^p \mapsto \mathbb{R}^q$ and $h : \mathbb{R}^q \mapsto \mathbb{R}$ with their composition $G = h \circ \mathbf{F} : \mathbb{R}^p \mapsto \mathbb{R}$, defined by*

$$G(\mathbf{x}) = h(\mathbf{F}(\mathbf{x})) = h(F_1(\mathbf{x}), \dots, F_q(\mathbf{x})),$$

where $F_i : \mathbb{R}^p \mapsto \mathbb{R}$, $p \geqslant 1$ and $q \geqslant 1$. Then,

> *G is convex if h is convex and nondecreasing (nonincreasing) in each argument, and F_i are convex (concave),*

> *G is concave if h is concave and nondecreasing (nonincreasing) in each argument, and F_i are concave (convex).*

Depending on the mathematical form of the function, a transformation of the radiobiological criteria functions presented in section 2.2 is suggested, and one of the two approaches is applied to analyse the convexity of the transformed criterion functions.

4. Results

4.1. TCP function using the LQ-Poisson cell survival model

The concavity of the logarithmically transformed linear-Poisson model based TCP function has been investigated by others (Choi and Deasy 2002, Romeijn *et al* 2004). These authors have applied the strictly increasing transformation $h(z) = \ln z$ to show that the linear-Poisson-based TCP function is log-concave.

Here, the same transformation has been applied to the LQ-Poisson-based TCP function of (2). In appendix A, we show that $-\ln \mathrm{TCP}_{\mathrm{LQ}}(\mathbf{D})$ is strictly convex under the condition that

$$D_k > \sqrt{\frac{1}{2}\frac{(\alpha/\beta)n}{\alpha}} - \frac{1}{2}(\alpha/\beta)n. \tag{13}$$

Since dose is a non-negative quantity, constraint (13) is satisfied if $\alpha^2 n > 2\beta$. In the opposite case, where $\alpha^2 n < 2\beta$, $\mathrm{SF}_{\mathrm{LQ}}(D)$ is sigmoidal for $D \geqslant 0$, and its curvature is only strictly positive if (13) is satisfied. For typical clinical values of α, α/β and n (e.g., for prostate cancer, $\alpha = 0.15 \pm 0.04 \, \mathrm{Gy}^{-1}$, $\alpha/\beta = 3.1 \pm 0.5 \, \mathrm{Gy}$, $n = 30$ (Wang *et al* 2003)), the right-hand side of (13) is negative, and therefore $-\ln TCP_{\mathrm{LQ}}(\mathbf{D})$ is a strictly convex function. Consequently, $\ln \mathrm{TCP}_{\mathrm{LQ}}(\mathbf{D})$ is a strictly concave function.

4.2. TCP function using the linear-Poisson cell survival model

In the case of either a large α/β ratio (i.e., $\beta \approx 0$) or small dose inhomogeneities and a dose-per-fraction close to the reference dose for which the underlying dose–response model was derived, the LQ-Poisson-based TCP function of (2) reduces to the linear-Poisson-based TCP model,

$$\mathrm{TCP}_{\mathrm{L}}(\mathbf{D}) = \exp\left[-N_0 \sum_{k=1}^{N} v_k \, \mathrm{e}^{-\alpha D_k}\right]. \tag{14}$$

Others have shown that $\mathrm{TCP}_{\mathrm{L}}(\mathbf{D})$ is a log-concave function (Choi and Deasy 2002, Romeijn *et al* 2004).

In appendix B, we apply a concave and strictly increasing transformation of the (already) convex function $-\ln \mathrm{TCP}_{\mathrm{L}}(\mathbf{D})$ and show that it is still convex. In particular, we conclude that the function $\ln(-\ln \mathrm{TCP}_{\mathrm{L}}(\mathbf{D}))$ is 'less convex' than $-\ln \mathrm{TCP}_{\mathrm{L}}(\mathbf{D})$. Hence, the former double-logarithmic transformation will produce tighter bounds in the piecewise linear approximation of the Pareto efficient frontier than the single-logarithmic transformation of $\mathrm{TCP}_{\mathrm{L}}(\mathbf{D})$ (Siem *et al* 2008).

4.3. Population-averaged TCP function using the LQ-Poisson cell survival model

Assuming that the inter-patient intrinsic radiosensitivity of a certain tumour type is normally distributed over the population, we show in appendix C that $-\ln \mathrm{TCP}_{\mathrm{pop}}(\mathbf{D})$ is strictly convex

for the LQ-Poisson-based TCP function, provided that the dose in each voxel satisfies the condition

$$D_k > \sqrt{\left[\frac{1}{2}(\alpha/\beta)n\right]^2 + \frac{1}{2}\frac{(\alpha/\beta)n}{\alpha}} - \frac{1}{2}(\alpha/\beta)n. \tag{15}$$

For typical clinical values of α, α/β and n, the right-hand side of (15) is a small positive value, and thus the inequality is easily satisfied for tumours.

In case the inter-patient radiosensitivity has a log-normal probability density function, $\phi(\alpha)$, the population-averaged LQ-Poisson-based TCP function is not necessarily logarithmic concave since $\phi(\alpha)$ is not logarithmic concave (Bagnoli and Bergstrom 2005).

4.4. NTCP function using the relative seriality s-model

For both the (nonfractionation-corrected) gEUD-based reformulation of Lyman's phenomenological NTCP model (6) and the gEUD-based phenomenological NTCP function due to Alber and Nüsslin (2001), Romeijn *et al* (2004) have proven convexity under the strictly increasing transformation $h(z) = -\ln(1 - z)$.

Here, we apply the same transformation to the relative seriality *s*-model of (3), and analyse its convexity properties. Since the relative seriality *s*-model can neither be formulated as a function of EUD nor as a function of gEUD, the application of the model in multi-criteria optimization of IMRT will yield a Pareto frontier that is different from the two aforementioned gEUD-based NTCP models.

The result comprises two steps (see appendix D). In theorem 4, it is shown that applying the strictly increasing transformation $h(z) = -\ln(1 - z^s)$ to (3) yields a strictly convex function for $s \geqslant 0$. In theorem 5, this result is used to prove that $-\ln(1 - \mathrm{NTCP_{RS}}(\mathbf{D}))$ is a strictly convex function, provided that $0 < s < 1$. For most normal tissues this condition is satisfied.

Note that $-\ln(1 - \mathrm{NTCP_{RS}}(\mathbf{D}))$ is a 'more convex' transformation of $\mathrm{NTCP_{RS}}(\mathbf{D})$ than $-\ln(1 - \mathrm{NTCP_{RS}}(\mathbf{D})^s)$. Therefore, the latter transformation is more efficient for the approximation of the Pareto efficient frontier than the former (Siem *et al* 2008). Nevertheless, the convexity of $-\ln(1 - \mathrm{NTCP_{RS}}(\mathbf{D}))$ is used in the convexity analysis of the composite NTCP function (11).

4.5. EUD function using the LQ-Poisson cell survival model

Note that $\mathrm{EUD_{LQ}}(\mathbf{D})$ is concave if and only if its Hessian is negative semidefinite. Unfortunately, the formulae for the second partial derivatives become so complex that we were unable to analytically derive conditions under which the Hessian is negative semidefinite. Nevertheless, numerical evaluation of the Hessian for the two-dimensional case $\mathbf{D} = (50\,\mathrm{Gy}, 40\,\mathrm{Gy})$ with equally-sized voxel volumes and realistic clinical values $\alpha = 0.30\,\mathrm{Gy}^{-1}$, $\alpha/\beta = 10\,\mathrm{Gy}$, $n = 30$ showed that the determinant of the 2×2 Hessian is negative, which implies that the Hessian is indefinite. Hence, $\mathrm{EUD_{LQ}}(\mathbf{D})$ is neither convex nor concave in this point. The same result is obtained for other values of $\mathbf{D}$. This contradicts the results of Bortfeld *et al* (2008), although their concavity criterion $\alpha^2 n > 2\beta$ is fulfilled. Our counter example requires the criterion to be re-evaluated.

4.6. gEUBED function for normal tissues using the LQ-model

By noting that $\mathrm{gEUBED}_a(\mathbf{D})$ can be obtained from (7) via a simple vector transformation, we prove that $\mathrm{gEUBED}_a(\mathbf{D})$ is convex for $a \geqslant 1$ in theorem 6 of appendix E.

4.7. NTCP function using the fractionation-corrected LKB model

In a similar way, Romeijn *et al* (2004) applied the transformation $h(z) = -\ln(1 - z)$ to the (nonfractionation-corrected) gEUD-based LKB model of (6), we apply it to the (fractionation-corrected) gEUBED-based LKB model of (9). Using the result of section 4.6, we show in theorem 7 of appendix E that $-\ln(1 - \text{NTCP}_{\text{LKB}}(\mathbf{D}))$ is a convex function.

4.8. Composite TCP and NTCP functions

In appendix F, it is shown that $-\ln \text{TCP}(\mathbf{D})$ and $-\ln(1 - \text{NTCP}(\mathbf{D}))$ are convex functions for the relevant fractionation-corrected dose–response models presented in section 2.2. This extends the results obtained by Romeijn *et al* (2004) to include TCP and NTCP models that incorporate fractionation effects.

4.9. Uncomplicated tumour control probability

Without proof, we echo the notion of Romeijn *et al* (2004) that from theorems 8 and 9 (see appendix F) it immediately follows that $-\ln P_+^1(\mathbf{D})$ is a strictly convex function, provided that the convexity conditions of the constituent TCP and NTCP functions are satisfied. The analysis gets complicated when the transformation $h(z) = -\ln z$ is applied to $P_+^0(\mathbf{D})$. No guarantee for the convexity of $-\ln P_+^0(\mathbf{D})$ can be derived.

Whether this result also rules out convexity of $-\ln P_+^\delta(\mathbf{D})$ for $0 < \delta < 1$ remains to be shown. Little impact of the δ value on $P_+^\delta(\mathbf{D})$ has been reported (Kim and Tomé 2006). Another argument would be that $P_+^1(\mathbf{D}) - P_+^0(\mathbf{D}) = \text{NTCP}(\mathbf{D})[1 - \text{TCP}(\mathbf{D})]$ is small in the optimum. Therefore, it seems reasonable to use $P_+^1(\mathbf{D})$ instead of $P_+^\delta(\mathbf{D})$, since the former criterion is guaranteed to be log-concave.

Nevertheless, for biologically-based multi-criteria optimization the more elementary composite TCP and NTCP functions are preferred, since they do not impose an *a priori* balance between them.

5. Discussion and conclusions

We have identified transformations of commonly used nonconvex radiobiological treatment plan evaluation criteria that account for the dose-per-fraction effect and analysed conditions under which the transformed criteria are convex functions that do not change the set of Pareto efficient treatment plans of a multi-criteria FMO problem. The dose-per-fraction effect is accounted for by exploiting the well-known log-linear-quadratic cell survival model that has often been applied in fractionated radiotherapy. By using the transformations, we were able to reformulate clinically relevant multi-criteria optimization problems being cast in terms of TCP and NTCP endpoints as convex problems.

In contrast to the work conducted by Romeijn *et al* (2004), our analysis comprised radiobiological criteria that cannot be cast as a transformation of the elementary EUD or gEUD functions. This implies that the treatment plan evaluation criteria analysed in the current work are truly different from those that were analysed before. Therefore, the criteria analysed will yield different Pareto efficient frontiers.

We have shown that the convexity conditions of transformed criterion functions that are based on the mechanistic LQ-Poisson-based dose–response model are directly related to the form of the log-linear-quadratic cell survival function underlying it. In this respect, it should be mentioned that the log-linear cell survival model underlying the criterion functions analysed by Romeijn *et al* (2004) is a limit of the log-linear-quadratic model for either a large α/β

ratio or for small dose inhomogeneities and a dose-per-fraction close to the reference dose for which the dose–response model was derived. Although Bortfeld *et al* (2008) have claimed the LQ-Poisson-based EUD function to be concave under certain conditions, we have provided a numerical counter example that shows that this function is neither convex nor concave.

Another aspect that was introduced as an extension to the work of Romeijn *et al* (2004) is the type of transformation applied. Apart from being strictly increasing, we have noted that transformations of different quality exist. Certain transformations exist that yield transformed criterion functions that are 'less convex' than other transformations. This is particularly useful for the approximation of the Pareto efficient frontier by piecewise linear upper and lower bounds (Siem *et al* 2008). These authors have shown that if the criterion function is already convex, an increasing and concave transformation yields a less convex function, for which tighter upper and lower bounds can be obtained than for the original bounds of the Pareto efficient frontier that is to be approximated. We note that for the transformed criteria analysed by Romeijn *et al* (2004) increasing and concave transformations may exist that are more efficient.

Future research will address the application of convex biologically-based multi-criteria optimization to assess the effect of different radiobiological models on the Pareto efficient frontier and associated dose distributions for a given treatment technique (i.e., beam modality, beam energy, beam configuration and machine characteristics).

Acknowledgments

Appendix A

The following theorem shows that the log-transformed LQ-Poisson-based TCP function is concave under certain conditions.

Theorem 1. *Consider the function* $\text{TCP}_{\text{LQ}}(\mathbf{D})$ *of (2). Let* $h(z) = \ln z$, *which is a strictly increasing function. Then the composition*

$$-h(\text{TCP}_{\text{LQ}}(\mathbf{D})) = -\ln \text{TCP}_{\text{LQ}}(\mathbf{D}) = N_0 \sum_{k=1}^{N} v_k \text{SF}_{\text{LQ}}(D_k) \qquad (\text{A.1})$$

is a strictly convex function on the set of dose distributions

$$\left\{ \mathbf{D} \mid D_k > \sqrt{\frac{1}{2}\frac{(\alpha/\beta)n}{\alpha}} - \frac{1}{2}(\alpha/\beta)n, k = 1, \ldots, M \right\}.$$

Proof. Since the non-negative sum of strictly convex functions is strictly convex, it suffices to prove that $\text{SF}_{\text{LQ}}(D)$ is strictly convex for $-\ln \text{TCP}_{\text{LQ}}(\mathbf{D})$ to be strictly convex. The curvature of $\text{SF}_{\text{LQ}}(D)$ is the second derivative with respect to D,

$$\text{SF}''_{\text{LQ}}(D) = [(\alpha + 2\beta D/n)^2 - 2\beta/n]\text{SF}_{\text{LQ}}(D). \qquad (\text{A.2})$$

From (A.2), it is easy to see that the curvature is strictly positive if

$$D > \sqrt{\frac{1}{2}\frac{(\alpha/\beta)n}{\alpha}} - \frac{1}{2}(\alpha/\beta)n.$$

$\square$

Appendix B

The following theorem shows that $\ln(-\ln \mathrm{TCP_L}(\mathbf{D}))$ is a convex function of $\mathbf{D}$.

Theorem 2. *Consider the function* $-\ln \mathrm{TCP_L}(\mathbf{D})$ *with* $\mathrm{TCP_L}(\mathbf{D})$ *according to (14). Let* $h(z) = \ln z$, *which is a concave and strictly increasing function. Then the composition*

$$h(-\ln \mathrm{TCP_L}(\mathbf{D})) = \ln(-\ln \mathrm{TCP_L}(\mathbf{D})) = \ln N_0 + \ln\left(\sum_{k=1}^{N} v_k\, \mathrm{e}^{-\alpha D_k}\right) \qquad \text{(B.1)}$$

is a convex function on the set of dose distributions $\{\mathbf{D} \mid D_k \geqslant 0, k = 1, \ldots, M\}$.

Proof. It is easy to verify that $\mathrm{e}^{-\alpha D}$ is a log-convex function of D. Since the sum of log-convex functions is log-convex, $\ln\left(\sum_{k=1}^{N} v_k\, \mathrm{e}^{-\alpha D_k}\right)$ is convex (Boyd and Vandenberghe 2004, p 105). Hence, $\ln(-\ln \mathrm{TCP_L}(\mathbf{D}))$ is a convex function of $\mathbf{D}$. $\square$

Appendix C

The following theorem shows that in case the inter-patient distribution of intrinsic radiosensitivities is Gaussian the log-transformed population-averaged LQ-Poisson-based TCP function is concave under certain conditions.

Theorem 3. *Consider the function* $\mathrm{TCP_{pop}}(\mathbf{D})$ *of (4). Let* $\mathrm{TCP_{ind}}(\mathbf{D}, \alpha) = \mathrm{TCP_{LQ}}(\mathbf{D}, \alpha)$ *and let* $\phi(\alpha)$ *be the standard normal probability density function. Furthermore, let* β *be correlated to* α *by a constant value of* α/β. *Let* $h(z) = \ln z$, *which is a strictly increasing function. Then the composition*

$$h(\mathrm{TCP_{pop}}(\mathbf{D})) = \ln \mathrm{TCP_{pop}}(\mathbf{D}) = \ln\left[\int_0^\infty \phi(\alpha)\mathrm{TCP_{ind}}(\mathbf{D}, \alpha)\,\mathrm{d}\alpha\right]$$

is a strictly concave function on the set of dose distributions

$$\left\{\mathbf{D} \mid D_k > \sqrt{\left[\frac{1}{2}(\alpha/\beta)n\right]^2 + \frac{1}{2}\frac{(\alpha/\beta)n}{\alpha}} - \frac{1}{2}(\alpha/\beta)n, k = 1, \ldots, M\right\}.$$

Proof. We use theorem 6 from Prékopa (1973), which states that the function

$$F(\mathbf{D}) = \int_{I_\alpha} f(\mathbf{D}, \alpha)\,\mathrm{d}\alpha,$$

with $I_\alpha \subseteq \mathbb{R}$ a convex set, is logarithmic concave in $\mathbf{D}$ if $f(\mathbf{D}, \alpha)$ is a logarithmic concave function of $\mathbf{D}$ and α. In our case, $f(\mathbf{D}, \alpha) = \phi(\alpha)\mathrm{TCP_{ind}}(\mathbf{D}, \alpha)$ and $I_\alpha = \mathbb{R}_+$. Note that

$$\ln[\phi(\alpha)\mathrm{TCP_{ind}}(\mathbf{D}, \alpha)] = \ln \phi(\alpha) + \ln \mathrm{TCP_{ind}}(\mathbf{D}, \alpha).$$

It is easy to see that $\phi(\alpha)$ is strictly logarithmic concave (see, e.g., Bagnoli and Bergstrom (2005)). Since the non-negative sum of strictly concave functions is strictly concave, it remains to prove that $\ln \mathrm{TCP_{ind}}(\mathbf{D}, \alpha)$ is a strictly concave function of $\mathbf{D}$ and α. Note that since

$$-\ln \mathrm{TCP_{ind}}(\mathbf{D}, \alpha) = N_0 \sum_{k=1}^{N} v_k \mathrm{SF_{LQ}}(D_k, \alpha)$$

is separable in the doses to individual voxels, it suffices to show that $SF_{LQ}(D, \alpha)$ is a strictly convex function of D and α. Therefore, we analyse the Hessian of $SF_{LQ}(D, \alpha)$,

$$\nabla^2 SF_{LQ}(D, \alpha) = SF_{LQ}(D, \alpha) \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \tag{C.1}$$

with $a = (\alpha + 2\beta D/n)^2 - 2\beta/n$, $b = 2\beta D^2/n + \alpha D - 1$ and $c = D^2$. The Hessian (C.1) is positive definite if and only if $a > 0$ and its determinant, $ac - b^2$, is positive. It is easy to show that these two conditions are satisfied if

$$D > \sqrt{\left[\frac{1}{2}(\alpha/\beta)n\right]^2 + \frac{1}{2}\frac{(\alpha/\beta)n}{\alpha}} - \frac{1}{2}(\alpha/\beta)n. \tag{$\square$}$$

Appendix D

The following theorem shows that $-\ln(1 - NTCP_{RS}(\mathbf{D})^s)$ is a strictly convex function on the set of physically meaningful dose distributions.

Theorem 4. *Consider the function* $NTCP_{RS}(\mathbf{D})$ *of (3) with* $s \geqslant 0$. *Let* $h(z) = -\ln(1 - z^s)$, *which is a strictly increasing function. Then the composition*

$$h(NTCP_{RS}(\mathbf{D})) = -\ln(1 - NTCP_{RS}(\mathbf{D})^s)$$
$$= -\sum_{k=1}^{N} v_k \ln(1 - P_{LQ}(D_k)^s) \tag{D.1}$$

is a strictly convex function on the set of dose distributions $\{\mathbf{D} \mid D_k \geqslant 0, k = 1, \ldots, M\}$.

Proof. Note that $h(NTCP_{RS}(\mathbf{D}))$ is separable in doses to individual voxels. Hence, it suffices to show that $g(D) = -\ln(1 - P_{LQ}(D)^s)$ is a strictly convex function of D. By calculating the second derivative of g with respect to D, it can be shown that $g''(D)$ is strictly positive if

$$(\alpha + 2\beta D/n)^2 \left(\frac{sN_0 SF_{LQ}(D)}{1 - P_{LQ}(D)^s} - 1\right) + 2\beta/n > 0. \tag{D.2}$$

Note that the left-hand side of (D.2) can only be nonpositive if

$$sN_0 SF_{LQ}(D) < 1 - P_{LQ}(D)^s. \tag{D.3}$$

By substituting $x = sN_0 SF_{LQ}(D)$ and noting that $P_{LQ}(D)^s = e^{-x}$, it follows that (D.3) is equivalent with

$$x < 1 - e^{-x},$$

which is not true for $x \in \mathbb{R}$. Hence, it follows that (D.2) holds for any $D \geqslant 0$, which implies that $-\ln(1 - NTCP_{RS}(\mathbf{D})^s)$ is a convex function. $\square$

The following theorem shows that $-\ln(1 - NTCP_{RS}(\mathbf{D}))$ is a strictly convex function of the dose distribution, provided that $0 < s < 1$.

Theorem 5. *Let* $F(\mathbf{D}) = -\ln(1 - NTCP_{RS}(\mathbf{D})^s)$ *be the convex function from (D.1) and let* $h(z) = -\ln(1 - (1 - e^{-z})^{1/s})$. *Then the composition*

$$h(F(\mathbf{D})) = -\ln(1 - NTCP_{RS}(\mathbf{D}))$$

is a strictly convex function on the set of dose distributions $\{\mathbf{D} \mid D_k \geqslant 0, k = 1, \ldots, M\}$ *if* $0 < s < 1$.

Proof. From lemma 2 it is known that $G(\mathbf{D}) = h(F(\mathbf{D}))$ is convex if F is convex and h is convex and nondecreasing. Therefore, we only have to prove that h is convex and nondecreasing.

Let $t = 1/s$, with $0 < s < 1$. Then $t > 1$. Differentiating h with respect to z yields

$$h'(z) = \frac{t(1 - e^{-z})^{t-1} e^{-z}}{1 - (1 - e^{-z})^t}.$$

Since $0 < \mathrm{NTCP}_{\mathrm{RS}}(\mathbf{D}) < 1$, it follows that $z > 0$, from which it is easy to see that $h'(z) > 0$ and thus h is nondecreasing. For the second derivative of h to be strictly positive, it can be derived that the inequality

$$t\,e^{-z} - 1 + (1 - e^{-z})^t > 0$$

should hold. By substituting $x = e^{-z}$, and noting that $0 < x < 1$, we obtain

$$p(x) = tx - 1 + (1 - x)^t,$$

for which we have to show that $p(x) > 0$ for $0 < x < 1$. Note that $p(0) = 0$ and that the first derivative of p with respect to x is given by $p'(x) = t(1 - (1 - x)^{t-1})$. Since $p'(x) > 0$ for $0 < x < 1$ and $t > 1$, p is an increasing function of x. From $p(0) = 0$ and $p'(x) > 0$ for $0 < x < 1$, it follows that $p(x) > 0$ for $t > 1$ on the interval $0 < x < 1$, which implies that h is strictly convex. □

Using similar arguments, it can be shown that $p(x)$ is negative for $0 < x < 1$ in the case $s > 1$, which implies that h is concave. However, this does not imply that $-\ln(1 - \mathrm{NTCP}_{\mathrm{RS}}(\mathbf{D}))$ is not convex in $\mathbf{D}$. Unfortunately, the composition rules cannot be used to prove that $-\ln(1 - \mathrm{NTCP}_{\mathrm{RS}}(\mathbf{D}))$ is indeed a convex function, provided that $s > 1$.

Appendix E

The following theorem shows that $\mathrm{gEUBED}_a(\mathbf{D})$ is a convex function of the set of physically meaningful dose distributions.

Theorem 6. *The function* $\mathrm{gEUBED}_a(\mathbf{D})$ *of (8) is convex on the set of dose distributions* $\{\mathbf{D} \mid D_k \geqslant 0, k = 1, \ldots, M\}$ *for* $a \geqslant 1$.

Proof. From its definition, it is clear that (8) is obtained from (7) by a vector transformation of its argument, i.e., $\mathrm{gEUBED}_a(\mathbf{D}) = \mathrm{gEUD}_a(\mathbf{h}(\mathbf{D}))$ with $\mathbf{h}(\mathbf{z}) = (h_1(\mathbf{z}), \ldots, h_N(\mathbf{z}))$ and $h_k(\mathbf{z}) = \mathrm{BED}(z_k) = z_k\left(1 + \frac{z_k/n}{\alpha/\beta}\right)$, where $k = 1, \ldots, N$. The convexity of $\mathrm{gEUBED}_a(\mathbf{D})$ follows from the fact that for $a \geqslant 1$ the function $\mathrm{gEUD}_a(\mathbf{D})$ is convex and nondecreasing in each argument (Choi and Deasy 2002), and h_k are strictly convex. □

The following theorem shows that $-\ln(1 - \mathrm{NTCP}_{\mathrm{LKB}}(\mathbf{D}))$ is a convex function of $\mathbf{D}$.

Theorem 7. *Consider the function* $\mathrm{NTCP}_{\mathrm{LKB}}(\mathbf{D})$ *of (9) with* $a \geqslant 1$. *Let* $h(z) = -\ln(1 - z)$, *which is a strictly increasing function. Then the composition*

$$h(\mathrm{NTCP}_{\mathrm{LKB}}(\mathbf{D})) = -\ln(1 - \mathrm{NTCP}_{\mathrm{LKB}}(\mathbf{D}))$$

is a convex function on the set of dose distributions $\{\mathbf{D} \mid D_k \geqslant 0, k = 1, \ldots, M\}$.

Proof. Note that $-\ln(1 - \mathrm{NTCP}_{\mathrm{LKB}}(\mathbf{D}))$ is a transformation of $\mathrm{gEUBED}_a(\mathbf{D})$,

$$-\ln(1 - \mathrm{NTCP}_{\mathrm{LKB}}(\mathbf{D})) = \Psi\left(\frac{\mathrm{gEUBED}_a(\mathbf{D}) - \mathrm{BED}_{50}}{m\mathrm{BED}_{50}}\right),$$

where $\Psi(t) = -\ln(1 - \Phi(t))$, and $\Phi(t)$ is the standard normal cumulative distribution function. For lemma 2 to be applicable, we use theorem 6 and only have to show that $\Psi(t)$ is an increasing and convex function of t. Since h and Φ are strictly increasing functions, $\Psi(t)$ is strictly increasing in t. To investigate whether $\Psi(t)$ is convex, the first derivative with respect to t is calculated,

$$\Psi'(t) = \frac{\Phi'(t)}{1 - \Phi(t)}.$$

From Bagnoli and Bergstrom (2005) it is known that if $\Phi'(t)$ is logarithmic concave, $\Psi'(t)$ is increasing. Note that it is easy to show that $\Phi'(t)$ is log-concave, since

$$\ln(\Phi'(t)) = \ln\left(\frac{1}{\sqrt{2\pi\sigma}}\right) - \frac{1}{2}\left(\frac{t - \mu}{\sigma}\right)^2$$

is a parabola with strictly negative curvature. Since $\Psi'(t)$ is increasing, the second derivative of Ψ with respect to t is positive, which implies that $\Psi(t)$ is convex.   □

Appendix F

The following theorems show that the log-transformed composite TCP and NTCP functions are concave and convex, respectively, on the set of physically meaningful dose distributions for the relevant dose–response models defined in section 2 of this paper.

Theorem 8. *Consider the function* $\text{TCP}(\mathbf{D})$ *of (10), with* $\text{TCP}_j(\mathbf{D})$ *according to* $\text{TCP}_{\text{LQ}}(\mathbf{D})$ *of (2). Let* $h(z) = \ln z$, *which is a strictly increasing function. Then the composition*

$$-h(\text{TCP}(\mathbf{D})) = -\ln\text{TCP}(\mathbf{D}) = -\sum_{j=1}^{T} \xi_j \ln\text{TCP}_j(\mathbf{D})$$

is a strictly convex function on the set of dose distributions

$$\left\{ \mathbf{D} \mid D_k > \sqrt{\frac{1}{2}\frac{(\alpha/\beta)n}{\alpha}} - \frac{1}{2}(\alpha/\beta)n, k = 1, \dots, M \right\}.$$

Proof. As shown in theorem 1, the function $-\ln\text{TCP}_{\text{LQ}}(\mathbf{D})$ is strictly convex if (13) is satisfied. By noting that the non-negative weighted sum of strictly convex functions is strictly convex, it is evident that $-\ln\text{TCP}(\mathbf{D})$ is strictly convex in $\mathbf{D}$.   □

Theorem 9. *Consider the function* $\text{NTCP}(\mathbf{D})$ *of (11) with* $\text{NTCP}_j(\mathbf{D})$ *according to* $\text{NTCP}_{\text{RS}}(\mathbf{D})$ *of (3) or* $\text{NTCP}_{\text{LKB}}(\mathbf{D})$ *of (9). Let* $h(z) = -\ln(1 - z)$, *which is a strictly increasing function. Then the composition*

$$h(\text{NTCP}(\mathbf{D})) = -\ln(1 - \text{NTCP}(\mathbf{D})) = -\sum_{j=1}^{T} \xi_j \ln(1 - \text{NTCP}_j(\mathbf{D}))$$

is a strictly convex function on the set of dose distributions $\{\mathbf{D} \mid D_k \geqslant 0, k = 1, \dots, M\}$.

Proof. According to theorems 5 and 7, the functions $-\ln(1 - \text{NTCP}_{\text{RS}}(\mathbf{D}))$ and $-\ln(1 - \text{NTCP}_{\text{LKB}}(\mathbf{D}))$ are strictly convex, respectively. By noting that the non-negative weighted sum of strictly convex functions is strictly convex, it is evident that $-\ln(1 - \text{NTCP}(\mathbf{D}))$ is strictly convex in $\mathbf{D}$.   □

References

Alber M, Meedt G, Nüsslin F and Reemtsen R 2002 On the degeneracy of the IMRT optimization problem *Med. Phys.* **29** 2584–9

Alber M and Nüsslin F 2001 A representation of an NTCP function for local complication mechanisms *Phys. Med. Biol.* **46** 439–47

Bagnoli M and Bergstrom T 2005 Log-concave probability and its application *Econ. Theor.* **26** 445–69

Barendsen G W 1982 Dose fractionation, dose rate and iso-effect relationships for normal tissue responses *Int. J. Radiat. Oncol. Biol. Phys.* **8** 1981–97

Bertsekas D P 1999 *Nonlinear Programming* 2nd edn (Nashua, NH: Athena Scientific)

Börgers C 1997 The radiation therapy planning problem *Computational Radiology and Imaging: Therapy and Diagnostics (The IMA Volumes in Mathematics and its Applications) Proc. of IMA Workshop, March 1997* ed C Börgers and F Natterer (Berlin: Springer)

Bortfeld T, Craft D, Dempsey J F, Halabi T and Romeijn H E 2008 Evaluating target cold spots by the use of tail EUDs *Int. J. Radiat. Oncol. Biol. Phys.* **71** 880–9

Boyd S and Vandenberghe L 2004 *Convex Optimization* (Cambridge: Cambridge University Press)

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

Carlsson F 2008 Utilizing problem structure in optimization of radiation therapy *Doctoral Thesis* Royal Institute of Technology, Stockholm, Sweden (ISBN 978-91-7178-920-4)

Choi B and Deasy J O 2002 The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning *Phys. Med. Biol.* **47** 3579–89

Craft D L, Halabi T F, Shih H A and Bortfeld T R 2006 Approximating convex Pareto surfaces in multiobjective radiotherapy planning *Med. Phys.* **33** 3399–407

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

Fowler J F 1989 The linear-quadratic formula and progress in fractionated radiotherapy *Br. J. Radiol.* **62** 679–94

Hardy G H, Littlewood J E and Pólya G 1952 *Inequalities* (Cambridge: Cambridge University Press)

Hindi H 2004 A tutorial on convex optimization *Proc. of the American Control Conf.* vol 4 (Boston: IEEE) pp 3252–65

Hoffmann A L, Siem A Y, den Hertog D, Kaanders J H and Huizenga H 2006 Derivative-free generation and interpolation of convex Pareto optimal IMRT plans *Phys. Med. Biol.* **51** 6349–69

Horn R A and Johnson C R 1987 *Matrix Analysis* (Cambridge: Cambridge University Press)

Källman P, Ågren A and Brahme A 1992 Tumour and normal tissue responses to fractionated non-uniform dose delivery *Int. J. Radiat. Biol.* **62** 249–62

Keall P J and Webb S 2007 Optimum parameters in a model for tumour control probability, including interpatient heterogeneity: evaluation of the log–normal distribution *Phys. Med. Biol.* **52** 291–302

Kim Y and Tomé W A 2006 Risk-adaptive optimization: selective boosting of high-risk tumor subvolumes *Int. J. Radiat. Oncol. Biol. Phys.* **66** 1528–42

Küfer K-H, Scherrer A, Monz M, Alonso F, Trinkaus H, Bortfeld T and Thieke C 2003 Intensity-modulated radiotherapy—a large scale multi-criteria programming problem *OR Spectr.* **25** 223–49

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

Lind B K 1990 Properties of an algorithm for solving the inverse problem in radiation therapy *Inverse Problems* **6** 415–26

Lyman J T 1985 Complication probability as assessed from dose–volume histograms *Radiat. Res. Suppl.* **8** S13–9

McGary J E, Grant W and Woo S Y 2000 Applying the equivalent uniform dose formulation based on the linear-quadratic model to inhomogeneous tumor dose distributions: caution for analyzing and reporting *J. Appl. Clin. Med. Phys.* **1** 126–37

Miettinen K 1999 *Nonlinear Multiobjective Optimization* (Dordrecht: Kluwer)

Nesterov Y and Nemirovski A 1994 Interior-point polynomial methods in convex programming *SIAM Series in Applied Mathematics* vol 13 (Philadelphia, PA : SIAM)

Niemierko A 1997 Reporting and analyzing dose distributions: a concept of equivalent uniform dose *Med. Phys.* **24** 103–10

Niemierko A 1999 A generalized concept of equivalent uniform dose *Med. Phys.* **26** 1100 (abstract)

Niemierko A 2005 Biological optimization *Image-guided IMRT* ed T Bortfeld, R Schmidt-Ullrich, W De Neve and D E Wazer (Berlin: Springer) pp 199–216

Prékopa A 1973 On logarithmic concave measures and functions *Acta Sci. Math.* **34** 335–43

Reemtsen R and Alber M 2006 Continuous optimization of beamlet intensities for photon and proton radiotherapy Available at http://www.math.tu-cottbus.de/INSTITUT/lsing1/publications/SurveyRT0306.ps

Roberts S A and Hendry J H 1998 A realistic closed-form radiobiological model of clinical tumor-control data incorporating intertumor heterogeneity *Int. J. Radiat. Oncol. Biol. Phys.* **41** 689–99

Romeijn H E, Dempsey J F and Li J G 2004 A unifying framework for multi-criteria fluence map optimization models *Phys. Med. Biol.* **49** 1991–2013

Sanchez-Nieto B and Nahum A E 2000 BIOPLAN: software for the biological evaluation of radiotherapy treatment plans *Med. Dosim.* **25** 71–6

Siem A Y, den Hertog D and Hoffmann A L 2008 The effect of transformations on the approximation of univariate (convex) functions with applications to Pareto curves *Eur. J. Oper. Res.* **189** 347–62

Wang J Z, Guerrero M and Li X A 2003 How low is the α/β ratio for prostate cancer? *Int. J. Radiat. Oncol. Biol. Phys.* **55** 194–203

Effect of transformations on approximation of the Pareto efficient frontier

Continuous Optimization

The effect of transformations on the approximation of univariate (convex) functions with applications to Pareto curves

A.Y.D. Siem [a,*], D. den Hertog [a], A.L. Hoffmann [b]

[a] *Department of Econometrics and Operations Research/Center for Economic Research (CentER),
Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands*
[b] *Department of Radiation Oncology, Radboud University Nijmegen Medical Centre, Geert Grooteplein 32, 6525 GA Nijmegen,
The Netherlands*

Abstract

In the literature, methods for the construction of piecewise linear upper and lower bounds for the approximation of univariate convex functions have been proposed. We study the effect of the use of transformations on the approximation of univariate (convex) functions. In this paper, we show that these transformations can be used to construct upper and lower bounds for nonconvex functions. Moreover, we show that by using such transformations of the input variable or the output variable, we obtain tighter upper and lower bounds for the approximation of convex functions than without these approximations. We show that these transformations can be applied to the approximation of a (convex) Pareto curve that is associated with a (convex) bi-objective optimization problem.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Approximation; Convex programming; Convex/concave transformation; Multiple objective programming; Pareto curve

1. Introduction

We consider the approximation of a univariate convex function $y : \mathbb{R} \mapsto \mathbb{R}$, which is only known in a finite set of points $x^1, \ldots, x^n \in \mathbb{R}$ with values $y(x^1), \ldots, y(x^n) \in \mathbb{R}$. In Burkard et al. (1991), Fruhwirth et al. (1989), Rote (1992), Yang and Goh (1997) and Siem et al. (2007), this is done by iteratively constructing piecewise linear upper and lower bounds. For the construction of the bounds discussed in Siem et al. (2007) and Yang and Goh (1997), only function value information, and no derivative information is needed. However, for the construction of the bounds in Burkard et al. (1991), Fruhwirth et al. (1989) and Rote (1992), also derivative information is necessary.

---

[*] Corresponding author. Tel.: +31 13 4663254; fax: +31 13 4663280.
E-mail addresses: a.y.d.siem@uvt.nl (A.Y.D. Siem), d.denhertog@uvt.nl (D. den Hertog), a.hoffmann@rther.umcn.nl (A.L. Hoffmann).

For the approximation of a nonconvex function, these piecewise linear upper and lower bounds cannot be used. However, in this paper, we show that if we can find a transformation of the input variable or an increasing transformation of the output variable such that the nonconvex function becomes convex, we can also obtain upper and lower bounds for this nonconvex function.

Moreover, if the function that is to be approximated is already convex, we show in this paper that by using increasing and concave transformations of the output variable y, we can obtain tighter upper and lower bounds. Furthermore, we show that by using certain transformations of the input variable x, we can also obtain tighter upper and lower bounds. These transformations can be applied in combination with the lower bounds based on only function value information as well as in combination with the lower bounds based on derivative information.

Furthermore, we show the relevance of our methodology for the approximation of a univariate (convex) Pareto curve that is associated with (convex) bi-objective optimization problems. The construction of a Pareto curve may be time-consuming, since the underlying optimization problems may be very large in size; see e.g. Küfer et al. (2003) and Ehrgott and Johnston (2003). The methodology in this paper accelerates the construction of an accurate Pareto curve.

The remainder of this paper is organized as follows. In Section 2, we repeat the expressions for the upper and lower bounds as presented in Siem et al. (2007). In Section 3 we study the effect of transformations of the output variables. In Section 4, we discuss the effect of transformations of the input variables. In Section 5, we show the relevance of the transformations for the approximation of a (convex) Pareto curve for (convex) bi-objective optimization problems, and consider some examples. Finally, in Section 6 we give our conclusions.

2. Approximating convex functions

In this section we summarize some results on piecewise linear upper and lower bounds for approximating convex functions from Siem et al. (2007). We suppose that n input data points $x^1 < \cdots < x^n \in \mathbb{R}$ are given together with the associated output data points $y(x^1), \ldots, y(x^n) \in \mathbb{R}$. Then, it can be shown (see Siem et al., 2007) that the straight line through the points $(x^i, y(x^i))$ and $(x^{i+1}, y(x^{i+1}))$, for $1 \leqslant i \leqslant n-1$, is an upper bound of the convex function $y(x)$, for $x \in [x^i, x^{i+1}]$. Furthermore, it can be shown that the straight lines through $(x^{i-1}, y(x^{i-1}))$ and $(x^i, y(x^i))$, for $2 \leqslant i \leqslant n-1$ and through $(x^{i+1}, y(x^{i+1}))$ and $(x^{i+2}, y(x^{i+2}))$, for $1 \leqslant i \leqslant n-2$, are lower bounds of the convex function $y(x)$, for $x \in [x^i, x^{i+1}]$. In the rest of the paper we define

$$\lambda^i(x) = \frac{x^{i+1} - x}{x^{i+1} - x^i}.$$

Theorem 1. *Let n input/output data points $(x^1, y(x^1)), \ldots, (x^n, y(x^n))$, with $x^1 < x^2 < \cdots < x^n$ be given, and let $y(x)$ be convex. Suppose furthermore that $x^i \leqslant x \leqslant x^{i+1}$, then*

$$y(x) \leqslant \lambda^i(x) y(x^i) + (1 - \lambda^i(x)) y(x^{i+1}) \quad \forall x \in [x^i, x^{i+1}] \; \forall 1 \leqslant i \leqslant n-1, \tag{1}$$

$$y(x) \geqslant (1 - \lambda^{i-1}(x)) y(x^i) + \lambda^{i-1}(x) y(x^{i-1}) \quad \forall x \in [x^i, x^{i+1}], \; 2 \leqslant i \leqslant n-1, \tag{2}$$

and

$$y(x) \geqslant (1 - \lambda^{i+1}(x)) y(x^{i+2}) + \lambda^{i+1}(x) y(x^{i+1}) \quad \forall x \in [x^i, x^{i+1}], \; 1 \leqslant i \leqslant n-2. \tag{3}$$

Furthermore, if we also have derivative information, i.e., we also have the values $y'(x^1), \ldots, y'(x^n)$, then the tangent lines through the data points are also lower bounds. More mathematically we have:

$$y(x) \geqslant y(x^i) + y'(x^i)(x - x^i) \quad \forall x \in [x^1, x^n] \; \forall i = 1, \ldots, n. \tag{4}$$

If y is not differentiable, y' can also be a subgradient. It is shown in Siem et al. (2007), that these lower bounds are tighter than the lower bounds that are only based on function value information as given in Theorem 1. The bounds mentioned in this section are illustrated in Fig. 1.
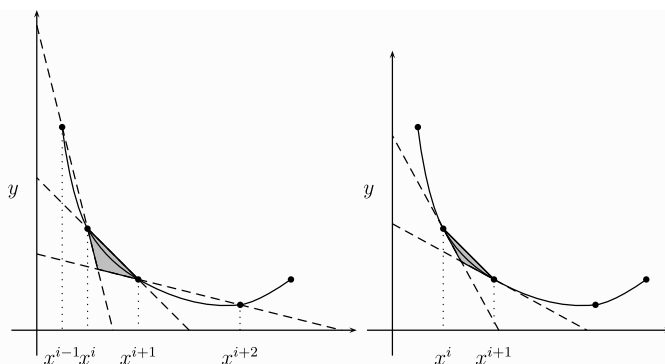
Fig. 1. Upper and lower bounds for a convex function on the interval $[x^i, x^{i+1}]$, using only function value information and using also derivative information.

3. The effect of transformations of the output variable

In this section we study the effect of transformations of the output variable on the upper and lower bounds based on only function evaluations, but also on the lower bounds based on derivative information.

Suppose that we want to construct upper and lower bounds for a function $y(x)$, that is not necessarily convex, and that we know an increasing function $h : \mathbb{R} \mapsto \mathbb{R}$ such that the function $h(y(x))$ is convex. Then, instead of constructing upper and lower bounds for the function $y(x)$, we can construct upper and lower bounds for $h(y(x))$ as mentioned in Section 2. In this section, we show that by applying the inverse transformation h^{-1} to these upper and lower bounds of $h(y(x))$, we obtain bounds for $y(x)$. In this way, we are able to construct upper and lower bounds for nonconvex functions.

Moreover, suppose that $y(x)$ is convex, and that we know an increasing concave function $h : \mathbb{R} \mapsto \mathbb{R}$ such that the function $h(y(x))$ is still convex. In this section, we also show that the bounds that we obtain, after applying the inverse transformation h^{-1} to the upper and lower bounds of $h(y(x))$, are even tighter than the bounds in (1)–(4). Without proof we first give the following well-known result.

Lemma 1. *If* $h : \mathbb{R} \mapsto \mathbb{R}$ *is strictly increasing and concave, then* $h^{-1} : \mathbb{R} \mapsto \mathbb{R}$ *exists, and is strictly increasing and convex.*

If $h : \mathbb{R} \mapsto \mathbb{R}$ *is strictly decreasing and convex, then* $h^{-1} : \mathbb{R} \mapsto \mathbb{R}$ *exists, and is also strictly decreasing and convex.*
If $h : \mathbb{R} \mapsto \mathbb{R}$ *is strictly increasing and convex, then* $h^{-1} : \mathbb{R} \mapsto \mathbb{R}$ *exists, and is strictly increasing and concave.*
Finally, if $h : \mathbb{R} \mapsto \mathbb{R}$ *is strictly decreasing and concave, then* $h^{-1} : \mathbb{R} \mapsto \mathbb{R}$ *exists, and is also strictly decreasing and concave.*

Now we can show our main results. First, we consider the upper bounds, second, we consider the lower bounds based on only function value information, and third, we consider the lower bounds based on derivative information.

Theorem 2. *Let* $h : \mathbb{R} \mapsto \mathbb{R}$ *be strictly increasing and let* $y : \mathbb{R} \mapsto \mathbb{R}$ *be such that* $h(y(x))$ *is convex. Then*

$$y(x) \leqslant h^{-1}[\lambda^i(x)h(y(x^i)) + (1 - \lambda^i(x))h(y(x^{i+1}))] \quad \forall x \in [x^i, x^{i+1}] \ \forall i = 1, \ldots, n-1, \tag{5}$$

i.e., the transformed upper bound is also an upper bound for the (not necessarily convex) function $y(x)$.
In addition, let h *be concave and* y *be convex. Then*

$$y(x) \leqslant h^{-1}[\lambda^i(x)h(y(x^i)) + (1 - \lambda^i(x))h(y(x^{i+1}))] \leqslant \lambda^i(x)y(x^i) + (1 - \lambda^i(x))y(x^{i+1})$$
$$\forall x \in [x^i, x^{i+1}] \ \forall i = 1, \ldots, n-1, \tag{6}$$

i.e., the transformed upper bound is tighter than the original upper bound (1).

Proof. From Theorem 1 and the convexity of $h(y(x))$ it follows that

$$h(y(x)) \leqslant \lambda^i(x)h(y(x^i)) + (1 - \lambda^i(x))h(y(x^{i+1})) \quad \forall x \in [x^i, x^{i+1}] \; \forall i = 1, \ldots, n-1. \tag{7}$$

Note that from Lemma 1, we know that h^{-1} is increasing. Applying h^{-1} on both sides of (7) gives (5). Next, we show (6):

$$y(x) = h^{-1}(h(y(x))) \leqslant h^{-1}[\lambda^i(x)h(y(x^i)) + (1 - \lambda^i(x))h(y(x^{i+1}))] \leqslant \lambda^i(x)y(x^i) + (1 - \lambda^i(x))y(x^{i+1}),$$

where in the first inequality we used (7) and the fact that h^{-1} is increasing, and in the second inequality that h^{-1} is convex. $\square$

Theorem 3. *Let $h : \mathbb{R} \mapsto \mathbb{R}$ be strictly increasing and let $y : \mathbb{R} \mapsto \mathbb{R}$ be such that $h(y(x))$ is convex, then*

$$y(x) \geqslant h^{-1}[\lambda^{i-1}(x)h(y(x^{i-1})) + (1 - \lambda^{i-1}(x))h(y(x^i))] \quad \forall x \in [x^i, x^{i+1}] \; \forall i = 2, \ldots, n-1, \tag{8}$$

$$y(x) \geqslant h^{-1}[\lambda^{i+1}(x)h(y(x^{i+1})) + (1 - \lambda^{i+1}(x))h(y(x^{i+2}))] \quad \forall x \in [x^i, x^{i+1}] \; \forall i = 1, \ldots, n-2, \tag{9}$$

i.e., the transformed lower bound is also a lower bound for the (not necessarily convex) function $y(x)$.
　　In addition, let h be differentiable and concave, and let y be convex. Then

$$y(x) \geqslant h^{-1}\big[\lambda^{i-1}(x)h(y(x^{i-1})) + (1 - \lambda^{i-1}(x))h(y(x^i))\big] \geqslant \lambda^{i-1}(x)y(x^{i-1}) + (1 - \lambda^{i-1}(x))y(x^i)$$
$$\forall x \in [x^i, x^{i+1}] \quad \forall i = 2, \ldots, n-1, \tag{10}$$

$$y(x) \geqslant h^{-1}\big[\lambda^{i+1}(x)h(y(x^{i+1})) + (1 - \lambda^{i+1}(x))h(y(x^{i+2}))\big] \geqslant \lambda^{i+1}(x)y(x^{i+1}) + (1 - \lambda^{i+1}(x))y(x^{i+2})$$
$$\forall x \in [x^i, x^{i+1}] \quad \forall i = 1, \ldots, n-2, \tag{11}$$

i.e., the transformed lower bounds are tighter than the original lower bounds (2) and (3).

Proof. From Theorem 1 and the convexity of $h(y(x))$ it follows that

$$h(y(x)) \geqslant \lambda^{i-1}(x)h(y(x^{i-1})) + (1 - \lambda^{i-1}(x))h(y(x^i)) \quad \forall x \geqslant x^i \; \forall i = 2, \ldots, n-1.$$

Since h^{-1} is increasing (see Lemma 1), we have that

$$y(x) = h^{-1}(h(y(x))) \geqslant h^{-1}[\lambda^{i-1}(x)h(y(x^{i-1})) + (1 - \lambda^{i-1}(x))h(y(x^i))] \quad \forall x \geqslant x^i \; \forall i = 2, \ldots, n-1,$$

which shows (8) and the first inequality of (10).
　　To show the second inequality of (10) we define

$$g_1^i(x) = h^{-1}[\lambda^{i-1}(x)h(y(x^{i-1})) + (1 - \lambda^{i-1}(x))h(y(x^i))] \quad \forall i = 2, \ldots, n-1,$$

and

$$g_2^i(x) = \lambda^{i-1}(x)y(x^{i-1}) + (1 - \lambda^{i-1}(x))y(x^i) \quad \forall i = 2, \ldots, n-1.$$

Note that g_1^i is convex since h^{-1} is a convex function with a linear function as argument. Now define $g^i(x) := g_1^i(x) - g_2^i(x)$. Then $g^i(x)$ is a convex function with zeros for $x = x^{i-1}$ and $x = x^i$. From Theorem 2 we may conclude that

$$h^{-1}[\lambda^{i-1}(x)h(y(x^{i-1})) + (1 - \lambda^{i-1}(x))h(y(x^i))] \leqslant \lambda^{i-1}(x)y(x^{i-1}) + (1 - \lambda^{i-1}(x))y(x^i)$$

for $x \in [x^{i-1}, x^i]$, which means that $g^i(x) \leqslant 0$ for $x \in [x^{i-1}, x^i]$. From the mean value theorem it follows that there exists a $\xi \in [x^{i-1}, x^i]$, for which $(g^i)'(\xi) = 0$. Since g is convex, we may conclude that $(g^i)'(x) \geqslant 0$, for all $x \geqslant x^i$, so also $g^i(x) \geqslant 0$ for all $x \geqslant x^i$, which shows the second inequality. The inequalities in (9) and (11) follow in a similar way. $\square$

　　Next, we show a similar result for the lower bounds based on derivative information.

Theorem 4. *Let $h : \mathbb{R} \mapsto \mathbb{R}$ be continuously differentiable, and strictly increasing. Furthermore, suppose that $y : \mathbb{R} \mapsto \mathbb{R}$, such that $h(y(x))$ is convex. Then*

$$y(x) \geqslant h^{-1}[h(y(x^i)) + h'(y(x^i))y'(x^i)(x - x^i)] \quad \forall x \in [x^1, x^n] \ \forall i = 1, \ldots, n, \tag{12}$$

i.e., the transformed lower bound is also a lower bound for the (not necessarily convex) function y(x).
  In addition, let h be concave and let y be convex. Then

$$y(x) \geqslant h^{-1}[h(y(x^i)) + h'(y(x^i))y'(x^i)(x - x^i)] \geqslant y(x^i) + y'(x^i)(x - x^i) \quad \forall x \in [x^1, x^n] \ \forall i = 1, \ldots, n, \tag{13}$$

i.e., the transformed lower bound is tighter than the original lower bound (4).

Proof. From (4) and the convexity of $h(y(x))$ it follows that

$$h(y(x)) \geqslant h(y(x^i)) + h'(y(x^i))y'(x^i)(x - x^i) \quad \forall x \in [x^1, x^n] \ \forall i = 1, \ldots, n.$$

Since we know from Lemma 1 that h^{-1} is increasing, we have that

$$y(x) = h^{-1}(h(y(x))) \geqslant h^{-1}[h(y(x^i)) + h'(y(x^i))y'(x^i)(x - x^i)] \quad \forall x \in [x^1, x^n] \ \forall i = 1, \ldots, n,$$

which shows (12) and the first inequality of (13). To show the second inequality of (13) we define

$$g_1^i(x) = h^{-1}[h(y(x^i)) + h'(y(x^i))y'(x^i)(x - x^i)] \quad \forall i = 1, \ldots, n,$$

and

$$g_2^i(x) = y(x^i) + y'(x^i)(x - x^i) \quad \forall i = 1, \ldots, n.$$

Note that g_1^i is convex since h^{-1} is a convex function (see Lemma 1) with a linear function as argument. Now define $g^i(x) := g_1^i(x) - g_2^i(x)$. Then $g^i(x)$ is a convex function, which is zero for $x = x^i$. Differentiating $g_1^i(x)$ gives:

$$(g_1^i)'(x) = (h^{-1})'[h(y(x^i)) + h'(y(x^i))y'(x^i)(x - x^i)]h'(y(x^i))y'(x^i)$$

$$= \frac{1}{h'[h^{-1}[h(y(x^i)) + h'(y(x^i))y'(x^i)(x - x^i)]]} h'(y(x^i))y'(x^i),$$

where we used the inverse function theorem. This means that

$$(g_1^i)'(x^i) = (g_1^i)'(x^i) - (g_2^i)'(x^i) = y'(x^i) - y'(x^i) = 0.$$

Since $g^i(x)$ is convex, we have that $(g^i)'(x) \geqslant 0$, for all $x \geqslant x^i$, and $(g^i)'(x) \leqslant 0$, for all $x \leqslant x^i$. This implies that $g^i(x) \geqslant 0$ for all $x \in [x^1, x^n]$, which shows the second inequality of (13). $\quad \square$

In a similar way it can be shown that if $h : \mathbb{R} \mapsto \mathbb{R}$ is strictly increasing and convex, and $h(y(x))$ is convex, the upper and lower bounds that we obtain by applying the inverse transformation h^{-1} to the upper and lower bounds of $h(y(x))$ are looser than the original upper and lower bounds of $y(x)$.

4. The effect of transformations of the input variable

In this section we study the effect of transformations of the input variable on the upper and lower bounds based on only function evaluations, but also on the lower bounds based on derivative information.

Suppose we want to construct upper and lower bounds for a function $y(x)$ that is not necessarily convex. If we know a function $h : \mathbb{R} \mapsto \mathbb{R}$ such that the function $y(h(x))$ is convex, we can construct upper and lower bounds for $y(h(x))$ as mentioned in Section 2. In this section, we show that by applying the inverse transformation h^{-1} to these upper and lower bounds of $y(h(x))$, we obtain bounds for $y(x)$. In this way, we are able to construct upper and lower bounds for nonconvex functions.

If $y(x)$ is convex, and we know a function $h : \mathbb{R} \mapsto \mathbb{R}$ such that the function $y(h(x))$ is still convex, we can also show that under certain conditions, the bounds that we obtain after applying the inverse transformation h^{-1} to the upper and lower bounds of $y(h(x))$, are tighter than the bounds in (1)–(4).

We have to distinguish between the case that $y(x)$ is decreasing and the case that $y(x)$ is increasing. If $y(x)$ is decreasing, then h has to be either strictly increasing and convex, or strictly decreasing and concave to obtain

tighter bounds. If $y(x)$ is increasing, then h has to be either strictly increasing and concave, or strictly decreasing and convex to obtain tighter bounds.

4.1. Decreasing output

In the rest of this paper we define

$$\mu^i(x) = \frac{h^{-1}(x^{i+1}) - h^{-1}(x)}{h^{-1}(x^{i+1}) - h^{-1}(x^i)}.$$

Theorem 5. *Let* $h : \mathbb{R} \mapsto \mathbb{R}$ *and* $y : \mathbb{R} \mapsto \mathbb{R}$ *be such that* $y(h(x))$ *is convex. Then*

$$y(x) \leqslant \mu^i(x)y(x^i) + (1 - \mu^i(x))y(x^{i+1}) \quad \forall x \in [x^i, x^{i+1}] \; \forall i = 1, \ldots, n-1, \tag{14}$$

i.e., the transformed upper bound is also an upper bound for the (not necessarily convex) function $y(x)$.

In addition, let h *be strictly increasing and convex, or strictly decreasing and concave. Let* y *be convex and let* $y(x^i) \geqslant y(x^{i+1}) \; \forall i = 1, \ldots, n-1$. *Then*

$$y(x) \leqslant \mu^i(x)y(x^i) + (1 - \mu^i(x))y(x^{i+1}) \leqslant \lambda^i(x)y(x^i) + (1 - \lambda^i(x))y(x^{i+1}) \quad \forall x \in [x^i, x^{i+1}] \; \forall i = 1, \ldots, n-1, \tag{15}$$

i.e., the transformed upper bounds are tighter than the original upper bounds.

Proof. Since the original dataset is given by $(x^i, y(x^i))$, for all $i = 1, \ldots, n$, the transformed dataset is given by $(h^{-1}(x^i), y(h(h^{-1}(x^i))))$. Note that it is not given by $(x^i, y(h(x^i)))$, since the value of $y(x)$ is not known in $x = h(x^i)$, but in $x = x^i = h(h^{-1}(x^i))$. From Theorem 1 and the convexity of $y(h(x))$ it follows that

$$y(h(x)) \leqslant \frac{h^{-1}(x^{i+1}) - x}{h^{-1}(x^{i+1}) - h^{-1}(x^i)} y(h(h^{-1}(x^i))) + \frac{x - h^{-1}(x^i)}{h^{-1}(x^{i+1}) - h^{-1}(x^i)} y(h(h^{-1}(x^{i+1}))) \quad \forall x \in [h^{-1}(x^i), h^{-1}(x^{i+1})].$$

Applying the transformation h^{-1} to the variable x yields

$$y(x) \leqslant \mu^i(x)y(x^i) + (1 - \mu^i(x))y(x^{i+1}) \quad \forall x \in [x^i, x^{i+1}],$$

which shows (14) and the first inequality of (15). The second inequality in (15) is equivalent with

$$\mu^i(x)(y(x^i) - y(x^{i+1})) \leqslant \lambda^i(x)(y(x^i) - y(x^{i+1})).$$

Since we assumed that $y(x^i) \geqslant y(x^{i+1})$ we only have to show that $\mu^i(x) \leqslant \lambda^i(x)$, i.e.,

$$\frac{h^{-1}(x^{i+1}) - h^{-1}(x)}{h^{-1}(x^{i+1}) - h^{-1}(x^i)} \leqslant \frac{x^{i+1} - x}{x^{i+1} - x^i}. \tag{16}$$

From Lemma 1, it follows that h^{-1} is concave. Let $\ell_i(x)$ be the straight line through the points $(x^i, h^{-1}(x^i))$ and $(x^{i+1}, h^{-1}(x^{i+1}))$, i.e.,

$$\ell_i(x) = h^{-1}(x^i) + \frac{h^{-1}(x^{i+1}) - h^{-1}(x^i)}{x^{i+1} - x^i}(x - x^i).$$

We can now write for $x \in [x^i, x^{i+1}]$

$$\frac{x^{i+1} - x}{x^{i+1} - x^i} = \frac{\ell_i(x^{i+1}) - \ell_i(x)}{\ell_i(x^{i+1}) - \ell_i(x^i)} \geqslant \frac{h^{-1}(x^{i+1}) - h^{-1}(x)}{h^{-1}(x^{i+1}) - h^{-1}(x^i)},$$

where in the inequality we used the concavity of h^{-1}, the fact that $\ell_i(x)$ is linear and that $\ell_i(x^i) = h^{-1}(x^i)$ and $\ell_i(x^{i+1}) = h^{-1}(x^{i+1})$, which implies $\ell_i(x) \leqslant h^{-1}(x) \; \forall x \in [x^i, x^{i+1}]$. $\square$

353

Theorem 6. *Let $h : \mathbb{R} \mapsto \mathbb{R}$ and $y : \mathbb{R} \mapsto \mathbb{R}$ be such that $y(h(x))$ is convex. Then*

$$y(x) \geqslant \mu^{i-1}(x)y(x^{i-1}) + (1 - \mu^{i-1}(x))y(x^i), \tag{17}$$

$$y(x) \geqslant \mu^{i+1}(x)y(x^{i+1}) + (1 - \mu^{i+1}(x))y(x^{i+2}), \tag{18}$$

i.e., the transformed lower bound is also a lower bound for the (not necessarily convex) function $y(x)$.

 In addition, let h be differentiable, and either strictly increasing and convex, or strictly decreasing and concave. Let y be convex and let $y(x^i) \geqslant y(x^{i+1})$ $\forall i = 1, \ldots, n - 1$. Then

$$
\begin{aligned}
y(x) &\geqslant \mu^{i-1}(x)y(x^{i-1}) + (1 - \mu^{i-1}(x))y(x^i) \\
&\geqslant \lambda^{i-1}(x)y(x^{i-1}) + (1 - \lambda^{i-1}(x))y(x^i) \quad \forall x \in [x^i, x^{i+1}] \ \forall i = 2, \ldots, n - 1,
\end{aligned} \tag{19}
$$

$$
\begin{aligned}
y(x) &\geqslant \mu^{i+1}(x)y(x^{i+1}) + (1 - \mu^{i+1}(x))y(x^{i+2}) \\
&\geqslant \lambda^{i+1}(x)y(x^{i+1}) + (1 - \lambda^{i+1}(x))y(x^{i+2}) \quad \forall x \in [x^i, x^{i+1}] \ \forall i = 1, \ldots, n - 2,
\end{aligned} \tag{20}
$$

i.e., the transformed lower bounds are tighter than the original lower bounds.

Proof. From Theorem 1 and the convexity of $y(h(x))$ it follows that

$$y(h(x)) \geqslant \frac{h^{-1}(x^i) - x}{h^{-1}(x^i) - h^{-1}(x^{i-1})} y(h(h^{-1}(x^{i-1}))) + \frac{x - h^{-1}(x^{i-1})}{h^{-1}(x^i) - h^{-1}(x^{i-1})} y(h(h^{-1}(x^i))) \quad \forall x \geqslant h^{-1}(x^i).$$

Applying the transformation $h^{-1}(x)$ yields

$$y(x) \geqslant \mu^{i-1}(x)y(x^{i-1}) + (1 - \mu^{i-1}(x))y(x^i) \quad \forall x \geqslant h^{-1}(x^i),$$

which shows (17) and the first inequality in (19). To show the second inequality in (19), we first define

$$g_1^i(x) = \mu^{i-1}(x)y(x^{i-1}) + (1 - \mu^{i-1}(x))y(x^i) \quad \forall i = 2, \ldots, n - 1,$$

and

$$g_2^i(x) = \lambda^{i-1}(x)y(x^{i-1}) + (1 - \lambda^{i-1}(x))y(x^i) \quad \forall i = 2, \ldots, n - 1.$$

Note that $g_1^i(x)$ is a convex function, since h^{-1} is concave (see Lemma 1) and $y(x^{i-1}) \geqslant y(x^i)$. Now, define $g^i(x) := g_1^i(x) - g_2^i(x)$. Then $g^i(x)$ is a convex function with zeros in $x = x^{i-1}$ and $x = x^i$. From Theorem 5 we may conclude that

$$\mu^{i-1}(x)y(x^{i-1}) + (1 - \mu^{i-1}(x))y(x^i) \leqslant \lambda^{i-1}(x)y(x^{i-1}) + (1 - \lambda^{i-1}(x))y(x^i)$$

for $x \in [x^{i-1}, x^i]$, with $y(x^{i-1}) \geqslant y(x^i)$, which means that $g^i(x) \leqslant 0 \ \forall x \in [x^{i-1}, x^i]$. From the mean value theorem it follows that there exists a $\xi \in [x^{i-1}, x^i]$, for which $(g^i)'(\xi) = 0$. Since g is convex, we may conclude that $(g^i)'(x) \geqslant 0$, for all $x \geqslant x^i$, so also $g^i(x) \geqslant 0$ for all $x \geqslant x^i$, which shows the second inequality. The inequalities in (18) and (20) follow in a similar way. $\square$

 Next, we show a similar result for the lower bound based on derivative information.

Theorem 7. *Let $h : \mathbb{R} \mapsto \mathbb{R}$ and $y : \mathbb{R} \mapsto \mathbb{R}$ be such that $y(h(x))$ is convex. Then*

$$y(x) \geqslant y(x^i) + y'(x^i)h'(h^{-1}(x^i))(h^{-1}(x) - h^{-1}(x^i)) \quad \forall x \in [x^1, x^n] \ \forall i = 1, \ldots, n, \tag{21}$$

i.e., the transformed lower bound is also a lower bound for the (not necessarily convex) function $y(x)$.

 Let h be continuously differentiable, and either strictly increasing and convex, or strictly decreasing and concave. Let $y : \mathbb{R} \mapsto \mathbb{R}$ be convex, and let $y'(x) \leqslant 0 \ \forall x \in [x^1, x^n]$. Then

$$y(x) \geqslant y(x^i) + y'(x^i)h'(h^{-1}(x^i))(h^{-1}(x) - h^{-1}(x^i)) \geqslant y(x^i) + y'(x^i)(x - x^i) \quad \forall x \in [x^1, x^n] \ \forall i = 1, \ldots, n, \tag{22}$$

i.e., the transformed lower bound is tighter than the original lower bound.

Proof. We first consider (21) and the first inequality of (22). From (4) and the convexity of $y(h(x))$ it follows that

$$y(h(x)) \geqslant y(h(h^{-1}(x^i))) + y'(h(h^{-1}(x^i)))h'(h^{-1}(x^i))(x - h^{-1}(x^i)) \quad \forall x \in [h^{-1}(x^1), h^{-1}(x^n)] \ \forall i = 1, \ldots, n.$$

By applying the transformation h^{-1}, we obtain

$$y(x) = y(h(h^{-1}(x))) \geqslant y(x^i) + y'(x^i)h'(h^{-1}(x^i))(h^{-1}(x) - h^{-1}(x^i)) \quad \forall x \in [x^1, x^n] \ \forall i = 1, \ldots, n.$$

We now prove the second inequality in (22). Since $y'(x) \leqslant 0$ we only have to show that

$$h'(h^{-1}(x^i))(h^{-1}(x) - h^{-1}(x^i)) \leqslant x - x^i.$$

Now define

$$g^i(x) = h'(h^{-1}(x^i))(h^{-1}(x) - h^{-1}(x^i)) - (x - x^i).$$

Note that $g^i(x)$ is concave, since it follows from Lemma 1 that h^{-1} is concave. Also note that $g^i(x^i) = 0$. With the inverse function theorem it follows that

$$(g^i)'(x^i) = h'(h^{-1}(x^i)) \frac{1}{h'(h^{-1}(x^i))} - 1 = 0.$$

Since $g^i(x)$ is concave, it follows that $g^i(x) \leqslant 0$. This shows the second inequality in (22). $\quad\square$

In a similar way it can be shown for the case that $y(x^i) \geqslant y(x^{i+1})$, for $i = 1, \ldots, n-1$, that if $h : \mathbb{R} \mapsto \mathbb{R}$ is either strictly increasing and concave, or strictly decreasing and convex, the upper and lower bounds that we obtain by applying the inverse transformation h^{-1} to the upper and lower bounds of $y(h(x))$ are looser than the original upper and lower bounds of $y(x)$.

4.2. Increasing output

We have similar theorems for the case that $y(x^i) \leqslant y(x^{i+1})$. However, to obtain tighter bounds, we now need $h(x)$ to be strictly increasing and concave. We give the theorems without proofs, since they follow in a similar way as Theorems 5–7.

Theorem 8. *Let $h : \mathbb{R} \mapsto \mathbb{R}$ and $y : \mathbb{R} \mapsto \mathbb{R}$ be such that $y(h(x))$ is convex. Then*

$$y(x) \leqslant \mu^i(x)y(x^i) + (1 - \mu^i(x))y(x^{i+1}) \quad \forall x \in [x^i, x^{i+1}] \ \forall i = 1, \ldots, n-1, \tag{23}$$

i.e., the transformed upper bound is also an upper bound for the (not necessarily convex) function $y(x)$.
   In addition, let h be strictly increasing and concave, or strictly decreasing and convex. Let y be convex and let $y(x^i) \leqslant y(x^{i+1}) \ \forall i = 1, \ldots, n-1$. Then

$$y(x) \leqslant \mu^i(x)y(x^i) + (1 - \mu^i(x))y(x^{i+1}) \leqslant \lambda^i(x)y(x^i) + (1 - \lambda^i(x))y(x^{i+1}) \quad \forall x \in [x^i, x^{i+1}] \ \forall i = 1, \ldots, n-1, \tag{24}$$

i.e., the transformed upper bounds are tighter than the original upper bounds.

Theorem 9. *Let $h : \mathbb{R} \mapsto \mathbb{R}$ and $y : \mathbb{R} \mapsto \mathbb{R}$ be such that $y(h(x))$ is convex. Then*

$$y(x) \geqslant \mu^{i-1}(x)y(x^{i-1}) + (1 - \mu^{i-1}(x))y(x^i),$$
$$y(x) \geqslant \mu^{i+1}(x)y(x^{i+1}) + (1 - \mu^{i+1}(x))y(x^{i+2}),$$

i.e., the transformed lower bound is also a lower bound for the (not necessarily convex) function $y(x)$.
   In addition, let h be differentiable, and either strictly increasing and concave, or strictly decreasing and convex. Let y be convex and let $y(x^i) \leqslant y(x^{i+1}) \ \forall i = 1, \ldots, n-1$. Then

$$y(x) \geqslant \mu^{i-1}(x)y(x^{i-1}) + (1 - \mu^{i-1}(x))y(x^i)$$
$$\geqslant \lambda^{i-1}(x)y(x^{i-1}) + (1 - \lambda^{i-1}(x))y(x^i) \quad \forall x \in [x^i, x^{i+1}] \ \forall i = 2, \ldots, n-1, \tag{25}$$

355

Table 1
The effect of transformations h on the upper and lower bounds for different scenarios of the input variable x and output variable y together with the corresponding numbers of the theorems

h	h Strictly increasing			h Strictly decreasing	
	y	x		x	
		y Increasing	y Decreasing	y Increasing	y Decreasing
Convex	Looser	Looser	Tighter 5–7	Tighter 8–10	Looser
Concave	Tighter 2–4	Tighter 8–10	Looser	Looser	Tighter 5–7

$$y(x) \geqslant \mu^{i+1}(x)y(x^{i+1}) + (1 - \mu^{i+1}(x))y(x^{i+2})$$
$$\geqslant \lambda^{i+1}(x)y(x^{i+1}) + (1 - \lambda^{i+1}(x))y(x^{i+2}) \quad \forall x \in [x^i, x^{i+1}] \ \forall i = 1, \ldots, n-2, \tag{26}$$

i.e., the transformed lower bounds are tighter than the original lower bounds.

Theorem 10. Let $h : \mathbb{R} \mapsto \mathbb{R}$ and $y : \mathbb{R} \mapsto \mathbb{R}$ be such that $y(h(x))$ is convex. Then
$$y(x) \geqslant y(x^i) + y'(x^i)h'(h^{-1}(x^i))(h^{-1}(x) - h^{-1}(x^i)) \quad \forall x \in [x^1, x^n] \ \forall i = 1, \ldots, n, \tag{27}$$
i.e., the transformed lower bound is also a lower bound for the (not necessarily convex) function $y(x)$.

In addition, let h be continuously differentiable, and either strictly increasing and concave, or strictly decreasing and convex. Let y be convex, and let $y'(x) \geqslant 0 \ \forall x \in [x^1, x^n]$. Then
$$y(x) \geqslant y(x^i) + y'(x^i)h'(h^{-1}(x^i))(h^{-1}(x) - h^{-1}(x^i)) \geqslant y(x^i) + y'(x^i)(x - x^i) \quad \forall x \in [x^1, x^n] \ \forall i = 1, \ldots, n, \tag{28}$$
i.e., the transformed lower bound is tighter than the original lower bound.

It can be shown in a similar way for the case that $y(x^i) \leqslant y(x^{i+1})$, for $i = 1, \ldots, n-1$, that if $h : \mathbb{R} \mapsto \mathbb{R}$ is either strictly increasing and convex, or strictly decreasing and concave, the upper and lower bounds that we obtain by applying the inverse transformation h^{-1} to the upper and lower bounds of $y(h(x))$, are looser than the original upper and lower bounds of $y(x)$.

In the following example of the approximation of $y(x) = x^2$, we show that it is also possible to find transformations for general convex functions by dividing the function into a decreasing part and an increasing part, provided that it is known for which value of x the function changes from decreasing to increasing.

Example 1. We consider the approximation of the function $y(x) = x^2$ on the interval $[-1, 1]$. Given are the data points $(-1, 1), (0, 0)$, and $(1, 1)$. Note that we can use the transformation $h_1(u) = \sqrt{u}$ on x, for the interval $[0, 1]$, and use the transformation $h_2(u) = \sqrt{-u}$ on x, for the interval $[-1, 0]$, to obtain tighter bounds. For this function we exactly know the value for x, in which the function changes from decreasing to increasing. In practice, such a point is often unknown.

We have summarized a part of the results of Sections 3 and 4 in Table 1.

5. Application to the approximation of the Pareto efficient frontier

An application of the methodology presented in this paper is the approximation of a convex Pareto curve (or Pareto efficient frontier) associated with a bi-objective optimization problem. First, in Section 5.1 we repeat some theory on bi-objective optimization. In Section 5.2, we show how we can apply the theory discussed in Sections 3 and 4 to obtain tighter upper and lower bounds of convex Pareto curves, and also to obtain upper and lower bounds of nonconvex Pareto curves.

5.1. Bi-objective optimization

Bi-objective optimization problems can in general be written in the form
$$\begin{aligned} \min_{v} \quad & f(v) = [f_1(v) \quad f_2(v)]^T \\ \text{s.t.} \quad & v \in S, \end{aligned} \tag{29}$$

where f_1 and f_2 are objective functions, and $S \subseteq \mathbb{R}^p$ is the feasible decision space. We want to minimize both functions f_1 and f_2 simultaneously. However, if there is a conflict between the objectives, this is not possible. In general, this optimization problem does not have a unique solution, since usually there is no solution that minimizes both objectives simultaneously. Actually, we are interested in those objective vectors, of which none of the components can be improved without worsening the other component, i.e., we are interested in the so-called Pareto optimal points.

Definition 1 (*Pareto optimality of a decision vector*). A decision vector $v^* \in S$ is Pareto optimal if there does not exist another decision vector $v \in S$ such that $f_i(v) \leqslant f_i(v^*)$, for all $i = 1, 2$ and $f_j(v) < f_j(v^*)$ for at least one index j.

The set of Pareto optimal points is called the Pareto optimal set and will be denoted by S^*. Let $Z := f(S)$ be the feasible objective space. Now we can define Pareto optimality in the objective space.

Definition 2 (*Pareto optimality of an objective vector*). An objective vector $z^* \in Z$ is Pareto optimal if there does not exist another objective vector $z \in Z$ such that $z_i \leqslant z_i^*$ for all $i = 1, 2$ and $z_j < z_j^*$ for at least one index j.

This means that the vector z^* is Pareto optimal if the corresponding decision vector v^* for which $z^* = f(v^*)$ is Pareto optimal. The image of the Pareto optimal set $f(S^*)$ is called the Pareto curve (or Pareto efficient frontier).

Two common methods to find Pareto optimal points are the weighting method and the ε-constraint method; see e.g. Miettinen (1999). In the latter method, we need to solve the following convex optimization problem for $\ell = 1, 2$:

$$\min_v \quad f_\ell(v)$$
$$\text{s.t.} \quad f_j(v) \leqslant \varepsilon_j \quad \forall j = 1, 2, \; j \neq \ell, \tag{30}$$
$$v \in S.$$

We now give the following theorem, which can be found in Miettinen (1999).

Theorem 11. *A decision vector $v^* \in S$ is Pareto optimal if and only if it is a solution of the ε-constraint problem (30) for every $\ell = 1, 2$, where $\varepsilon_j = f_j(v^*)$ for $j = 1, 2, \; j \neq \ell$.*

Proof. See Miettinen (1999), p. 85. $\square$

Let $\mathscr{E} = \{\varepsilon_2 \in \mathbb{R} : \exists v \in S : f_2(v) \leqslant \varepsilon_2\}$. Now define the function $p : \mathscr{E} \mapsto \mathbb{R}$ as $p(\varepsilon_2) = f_1(v^*(\varepsilon_2))$, where $v^*(\varepsilon_2)$ is the solution of (30) for $\ell = 1, 2$. The following theorem states that the Pareto curve $p(\varepsilon_2)$ is convex, provided that both f_1 and f_2 are convex, and S is a convex set.

Theorem 12. *Suppose that f_1 and f_2 are convex functions and S is a convex set, then the Pareto curve $p : \mathscr{E} \mapsto \mathbb{R}$ corresponding with bi-objective optimization problem (29) is convex.*

Proof. Let $0 \leqslant \lambda \leqslant 1$, $\varepsilon^1, \varepsilon^2 \in \mathscr{E}$. We then have $f_2(v^*(\varepsilon^1)) \leqslant \varepsilon^1$, and $f_2(v^*(\varepsilon^2)) \leqslant \varepsilon^2$. Let $v^0 = \lambda \, v^*(\varepsilon^1) + (1 - \lambda)v^*(\varepsilon^2)$, and $\varepsilon^0 = \lambda\varepsilon^1 + (1 - \lambda)\varepsilon^2$. Then, $f_2(v^0) = f_2(\lambda \; v^*(\varepsilon^1) + (1 - \lambda)v^*(\varepsilon^2)) \leqslant \lambda \; f_2(v^*(\varepsilon^1)) + (1 - \lambda) f_2(v^*(\varepsilon^2)) \leqslant \lambda\varepsilon^1 + (1 - \lambda)\varepsilon^2 = \varepsilon^0$. Also, $v^0 \in S$. Therefore, v^0 is a feasible point for optimization problem (30) with $\varepsilon = \varepsilon^0$ and $\ell = 1$. Furthermore

$$p(\varepsilon^0) = p(\lambda\varepsilon^1 + (1 - \lambda)\varepsilon^2) = f_1(v^*(\lambda\varepsilon^1 + (1 - \lambda)\varepsilon^2)) \leqslant f_1(v^0) = f_1(\lambda v^*(\varepsilon^1) + (1 - \lambda)v^*(\varepsilon^2))$$
$$\leqslant \lambda f_1(v^*(\varepsilon^1)) + (1 - \lambda)f_1(v^*(\varepsilon^2)) = \lambda p(\varepsilon^1) + (1 - \lambda)p(\varepsilon^2),$$

where we used in the first inequality that v^0 is feasible, and in the second inequality that f_1 is convex. Therefore p is convex. $\square$

If the objective functions are not convex, the Pareto curve $p(\varepsilon)$ is not necessarily convex. In Fig. 2 the feasible objective region is shown with both convex and nonconvex Pareto curves. There is a vast literature on methods to find Pareto optimal points; see e.g. Miettinen (1999).

Given a set of Pareto optimal points, we can now use the upper and lower bounds, given in Section 2, to approximate a convex Pareto curve $p(\varepsilon)$. However, Pareto curves are decreasing by definition. As a consequence of this, we can add the additional lower bound $p(\varepsilon^n) \leqslant p(\varepsilon)$, for $\varepsilon^{n-1} \leqslant \varepsilon \leqslant \varepsilon^n$. This is illustrated in Fig. 3.

5.2. Approximation of the Pareto efficient frontier

We can use the results of Sections 3 and 4 to obtain tighter bounds of the Pareto curve, by transforming one or both of the objectives. Suppose we want to minimize f_1 and f_2 simultaneously, and that f_1 and f_2 are convex. If we know an increasing and concave function h, such that $h(f_1(v))$ is convex, then also

$$
\begin{aligned}
h(p(\varepsilon)) = \min_v \; & h(f_1(v)) \\
\text{s.t. } & f_2(v) \leqslant \varepsilon, \\
& v \in S
\end{aligned}
\tag{31}
$$

is convex, and by applying Theorems 2–4, we can obtain tighter bounds for $p(\varepsilon)$.

Furthermore, if we can find a function $\tilde{h} : \mathbb{R} \mapsto \mathbb{R}$, which is either strictly increasing and concave, or strictly decreasing and concave, such that $\tilde{h}(f_2(v))$ is convex, then the function
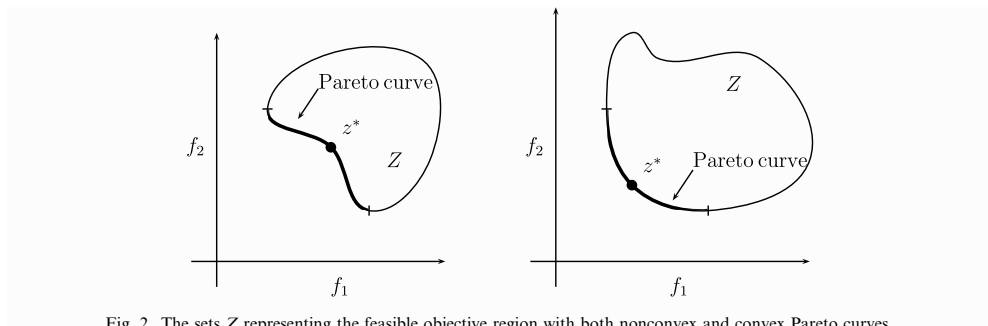
Fig. 2. The sets Z representing the feasible objective region with both nonconvex and convex Pareto curves.
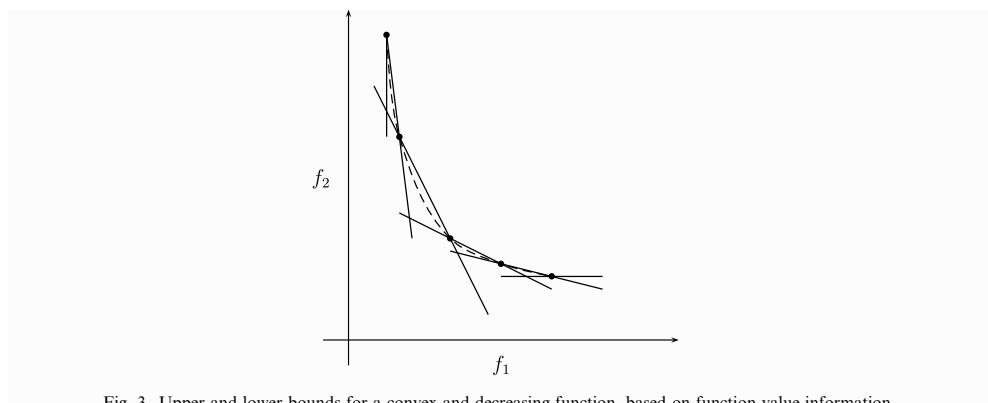
Fig. 3. Upper and lower bounds for a convex and decreasing function, based on function value information.

$$\tilde{p}(\varepsilon) = \min_{v} f_1(v)$$
$$\text{s.t. } \tilde{h}(f_2(v)) \leqslant \varepsilon, \tag{32}$$
$$v \in S$$

is convex. We can rewrite this into

$$p(\tilde{h}^{-1}(\varepsilon)) = \min_{v} f_1(v)$$
$$\text{s.t. } f_2(v) \leqslant \tilde{h}^{-1}(\varepsilon),$$
$$v \in S.$$

Since $\tilde{h}^{-1}$ is either strictly increasing and convex, or strictly decreasing and concave, $p(\tilde{h}^{-1}(\varepsilon))$ is still convex, and $p(\varepsilon^i) \geqslant p(\varepsilon^{i+1})$, for $i = 1, \ldots, n-1$, by applying Theorems 5–7, we can obtain tighter upper and lower bounds for $p(\varepsilon)$.

Furthermore, if f_1 in (31) is not convex, h is increasing and such that $h(f_1(v))$ is convex, then by applying Theorems 2–4, we can obtain upper and lower bounds for the nonconvex Pareto curve. If in (32) f_2 is not convex and h is such that $h(f_2(x))$ is convex, then by applying Theorems 5–7, we can obtain upper and lower bounds for the nonconvex Pareto curve.

Example 2 (*p-norm*). For example, let $f_1(v) = v^T A v$ and $f_2(v) = v^T B v$, with A and B positive semi-definite, both be convex quadratic functions, we can choose $h(u) = \tilde{h}(u) = \sqrt{u}$. Note that both $h(f_1(v))$ and $\tilde{h}(f_2(v))$ are convex, since $h(f_1(v)) = \sqrt{v^T A v}$ and $h(f_2(v)) = \sqrt{v^T B v}$ are norms. Then, the Pareto curve, associated with optimization problem

$$\min_{v} \quad \sqrt{v^T A v}$$
$$\text{s.t.} \quad \sqrt{v^T B v} \leqslant \varepsilon,$$
$$v \in S$$

is convex. After applying the inverse transformation to the constructed bounds, we obtain tighter bounds, than without the transformations $h(x)$ and $\tilde{h}(x)$. More generally we can apply this to convex functions of the form $f(v) = \sum_i (a_i^T v)^p$, where a_i^T is the ith row of a squared matrix A. We can apply the transformation $h(u) = \sqrt[p]{u}$. Since $h(f(v)) = \sqrt[p]{\sum_i (a_i^T v)^p}$ is a norm (known as the p-norm), $h(f(v))$ is convex. The family of functions $f(v) = \sum_i (a_i^T v)^p$ plays an important role in l_p-programming; see Terlaky (1985).

Example 3 (*Strategic investment model*). In this example we consider a strategic investment model. There exist many sorts of investment categories, such as deposits, saving accounts, bonds, stocks, real estate, commodities, foreign currencies, and derivatives. Each category has its own expected return, and its own risk characteristic. The strategic investment model can be used to model how top management could spread an overall budget over several investment categories. The objective is to minimize the portfolio risk (measured by the variance of the return), such that a certain minimal desired expected return is achieved. The model was introduced by Markowitz (1952), and is given by

$$\min_{v} \quad v^T \Sigma v$$
$$\text{s.t.} \quad r^T v \geqslant M,$$
$$e_p^T v = 1, \tag{33}$$
$$v \in \mathbb{R}_+^p,$$

where Σ is a positive semi-definite covariance matrix consisting of elements Σ_{ij} of covariances between investment categories i and j, r is the vector consisting of elements r_i of expected return of investment category i, M is the desired expected portfolio return, e_p is the p-dimensional all-one vector, v is the vector with elements v_i of fractions of the budget invested in each category, and p is the number of investment categories.

In Table 2, a simple problem instance is given, which we took from Bisschop (2000), Chapter 18. It contains three investment categories: stocks, bonds, and real estate. The stochastic variable R_i denotes the expected return of investment category i. Based on four equidistant data points, we calculate the upper and lower bounds (1)–(4). Then, we apply the concave and increasing transformation $h(u) = \sqrt{u}$ to the objective in (33). Note that since the function $f(v) = \sqrt{v^T \Sigma v}$ is convex (it is a norm), the conditions of Theorems 2–4 are satisfied. Then, we calculate the transformed upper and lower bounds from (6), (10), (11), and (13). The transformed and nontransformed bounds are shown in Fig. 4. Indeed, as we can see in Fig. 4, the transformed bounds are tighter than the nontransformed bounds, as we showed in Theorems 2–4.

Next, we exchange the objective and the first constraint in (33):

$$\begin{aligned}
\min_{v} \quad & -r^T v \\
\text{s.t.} \quad & v^T \Sigma v \leqslant M, \\
& e_p^T v = 1, \\
& v \in \mathbb{R}_+^p.
\end{aligned} \tag{34}$$

We calculate the upper and lower bounds (1)–(4). Again, we apply the transformation $h(u)$ to the portfolio risk, i.e., the first constraint in (34). By calculating the transformed bounds as given in (15), (19), (20), and (22), we obtain tighter bounds of the Pareto efficient frontier. The transformed and nontransformed upper and lower bounds are shown in Fig. 5.

Example 4 (*Nonconvex Pareto efficient frontier*). In this example we consider the estimation of a nonconvex Pareto efficient frontier. The objective is to minimize

$$f_1(v) = \left[1 - \prod_{i=1}^{n} [1 - (e^{-\exp(-\alpha v_i - \beta v_i^2/n)})^s]^{\Delta c_i} \right]^{1/s},$$

Table 2
Expected returns and covariances

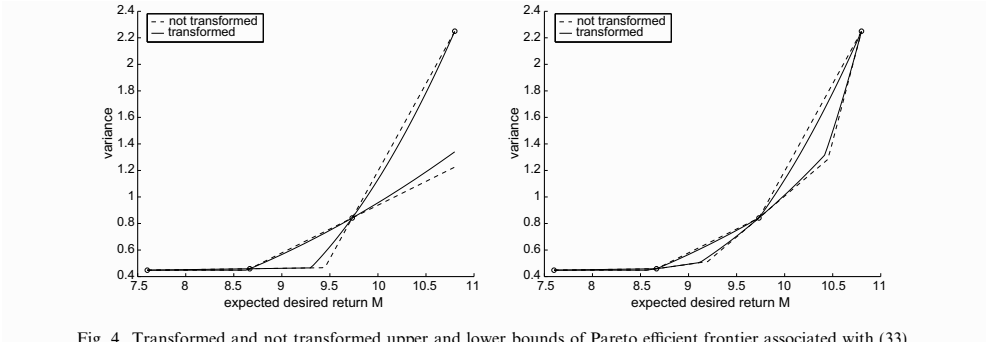| Category	i	ER_i	$\mathrm{Cov}[R_i, R_j]$		
			j		
			1	2	3
Stocks	1	10.8	2.250	−0.120	0.450
Bonds	2	7.600	−0.120	0.640	0.336
Real estate	3	9.500	0.450	0.336	1.440

Fig. 4. Transformed and not transformed upper and lower bounds of Pareto efficient frontier associated with (33).
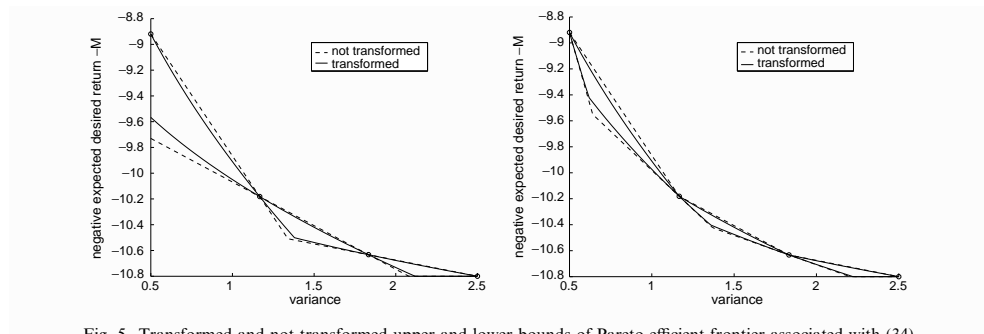
Fig. 5. Transformed and not transformed upper and lower bounds of Pareto efficient frontier associated with (34).

where α, β, s, $\Delta c_i = c_i / \sum_{i=1}^{n} c_i$, and c_i are constants, and at the same time maximize

$$f_2(v) = \sum_{i=1}^{n} v_i.$$

The use of $f_1(v)$ in Intensity Modulated Radiation Therapy can be found in Brahme and Agren (1987). The associated ε-constraint optimization problem is given as

$$
\begin{aligned}
\min_{v} \quad & \left[1 - \prod_{i=1}^{n} [1 - (e^{-\exp(-\alpha v_i - \beta v_i^2/n)})^s]^{\Delta c_i} \right]^{1/s} \\
\text{s.t.} \quad & \sum_{i=1}^{n} v_i \geqslant \varepsilon, \\
& v \in \mathbb{R}_+^n.
\end{aligned}
\tag{35}
$$

It can be shown that $f_1(v)$ is not convex. This implies that the Pareto efficient frontier that is associated with (35) is not necessarily convex. However, by applying the convex and increasing transformation $h(u) = -\log(1 - u^s)$ to f_1 we obtain a convex function $h(f_1(v))$; see Hoffmann et al. (2007). We take $n = 5$, $\alpha = 1$, $\beta = 5$, $s = 2$, $c_1 = 5$, $c_2 = 6$, $c_3 = 4$, $c_4 = 3$, and $c_5 = 8$. Now, we can construct the transformed upper and lower bounds both using only function value information and using also derivative information as given in (5), (8), (9), and (12). The bounds are shown in Fig. 6.
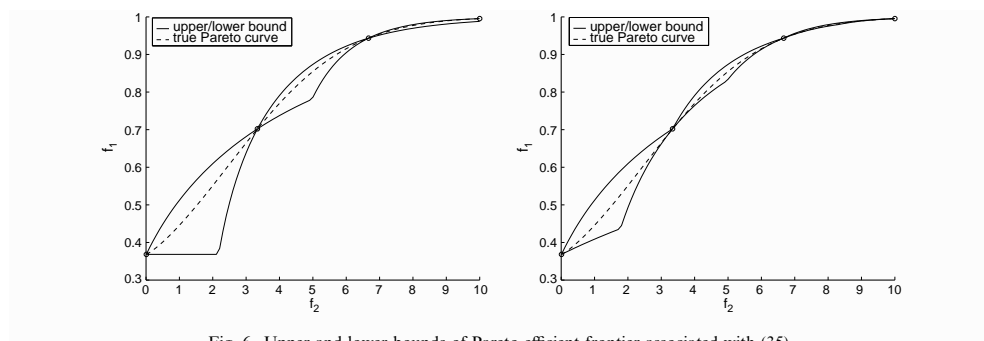

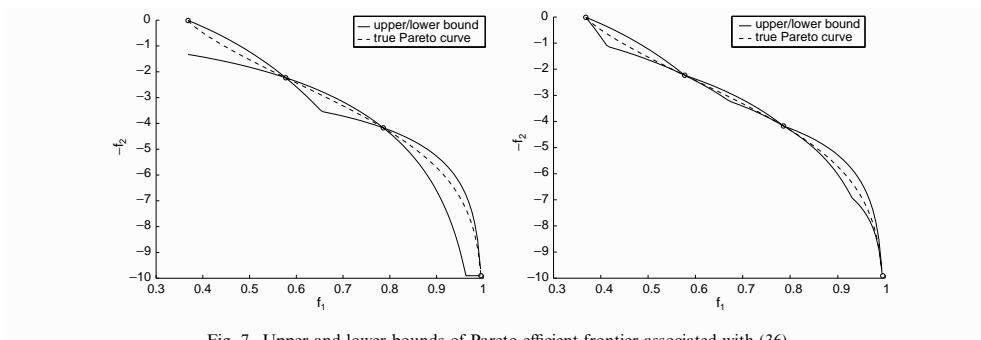Fig. 6. Upper and lower bounds of Pareto efficient frontier associated with (35).

Fig. 7. Upper and lower bounds of Pareto efficient frontier associated with (36).

Next, we consider the ε-constraint problem with f_1 as a constraint and f_2 as objective:

$$
\begin{aligned}
\min_{v} \quad & -\sum_{i=1}^{n} v_i \\
\text{s.t.} \quad & \left[1 - \prod_{i=1}^{n} [1 - (e^{-\exp(-\alpha v_i - \beta v_i^2/n)})^s]^{\Delta c_i} \right]^{1/s} \leqslant \varepsilon, \\
& v \in \mathbb{R}_+^n.
\end{aligned}
\tag{36}
$$

Again, we apply the transformation $h(u)$ to f_1, and construct the transformed upper and lower bounds using only function value information and using also derivative information as given in (14), (17), (18) and (21). These bounds are shown in Fig. 7.

6. Conclusions

In this paper, we studied the effect of transformations on the approximation of univariate (convex) functions. By using transformations on the input or the output variables, we can transform nonconvex functions into convex functions, for which upper and lower bounds are given. We showed that applying the inverse transformation to these upper and lower bounds gives us bounds for the original nonconvex function.

Moreover, we showed that if the function that is to be approximated is convex, we can obtain tighter upper and lower bounds than the original piecewise linear upper and lower bounds. We can achieve this by using increasing and concave transformations of the output variable y and concave or convex transformations of the input variable x.

Furthermore, we applied the developed theory to the approximation of a convex Pareto curve and a nonconvex Pareto curve, associated with bi-objective optimization problems. Finally, we gave some examples of these applications.

Acknowledgement

References

Bisschop, J., 2000. AIMMS optimization modeling. Technical Report, Haarlem.
Brahme, A., Agren, A.K., 1987. Optimal dose distribution for eradication of heterogeneous tumours. Acta Oncologica 26, 377–386.
Burkard, R.E., Hamacher, H.W., Rote, G., 1991. Sandwich approximation of univariate convex functions with an application to separable convex programming. Naval Research Logistics 38, 911–924.

Ehrgott, M., Johnston, R., 2003. Optimisation of beam directions in intensity modulated radiation therapy planning. OR Spectrum 25, 251–264.

Fruhwirth, B., Burkard, R.E., Rote, G., 1989. Approximation of convex curves with application to the bi-criteria minimum cost flow problem. European Journal of Operational Research 42, 326–338.

Hoffmann, A.L., den Hertog, D., Siem, A.Y.D., 2007. Convexity analysis of LQ-model based radiobiological objective functions for radiation therapy planning optimization. Working paper.

Küfer, K.H., Scherrer, A., Monz, M., Alonso, F., Trinkaus, H., Bortfeld, T., Thieke, C., 2003. Intensity-modulated radiotherapy—a large scale multi-criteria programming problem. OR Spectrum 25, 223–249.

Markowitz, H.M., 1952. Portfolio selection. Journal of Finance 7, 77–91.

Miettinen, K., 1999. Nonlinear Multiobjective Optimization. Kluwer Academic Publishers, Boston.

Rote, G., 1992. The convergence rate of the sandwich algorithm for approximating convex functions. Computing 48, 337–361.

Siem, A.Y.D., den Hertog, D., Hoffmann, A.L., 2007. A method for approximating univariate convex functions using function only evaluations. Working paper.

Terlaky, T., 1985. On l_p programming. European Journal of Operational Research 22, 70–100.

Yang, X.Q., Goh, C.J., 1997. A method for convex curve approximation. European Journal of Operational Research 97, 205–212.

Physico-biological optimisation of IMRT plans for NSCLC radiotherapy

A.L. Hoffmann, E.G.C. Troost, J. Bussink and J.H.A.M. Kaanders

Physico-biological treatment plan optimisation for homogeneous and heterogeneous individualised dose escalation in advanced non-small-cell lung cancer

A.L. Hoffmann[1], E.G.C. Troost[1], J. Bussink[2], J.H.A.M. Kaanders[2]

[1]Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, The Netherlands

[2]Department of Radiation Oncology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

Purpose

Radiation dose escalation reduces local relapse and improves overall survival in advanced stage non-small-cell lung cancer (NSCLC) patients. Dose-enhancement of an existing treatment plan can be achieved by re-normalisation until a pre-defined normal-tissue dose constraint is reached, by re-optimisation using a trial-and-error approach or by physico-biological optimisation (PBO) where the tumour control probability (TCP) and normal tissue complication probability, *e.g.* Grade ≥ 2 radiation induced pneumonitis ($NTCP_{RIP}$), drive the inverse treatment planning process.

Methods and Materials

This *in silico* study included ten randomly selected patients with stage III NSCLC. The initial treatment plans at fixed, unchanged number of fractions were re-normalised or re-optimised using PBO for an iso-toxic (IT) or maximum tolerable (MT) approach to determine the highest achievable TCP as function of the $NTCP_{RIP}$, with and without target dose homogeneity constraint, while satisfying dose constraints for the oesophagus, spinal cord and heart.

Results

For the initial treatment plans, the mean TCP in the study cohort was 34% and $NTCP_{RIP}$ varied between 6.4–16.4%. For the homogeneous approaches, IT re-optimisation yielded the least benefit, while MT re-optimisation resulted in a median [range] TCP of 53.5% [37.8–73.8%]. MT heterogeneous re-optimisation produced a median TCP of 72.8% [58.8–91.4%]. Mean fractional doses ± relative standard deviation for the MT homogeneous and heterogeneous re-optimised study arms were 2.2 [2.0–2.3] Gy ± 3.7% [2.4–5.2%] and 2.3 [2.1–2.7] Gy ± 8.8% [7.4–13.1%], respectively.

Conclusions

PBO provides an individual trade-off of TCP versus $NTCP_{RIP}$ as a means to generate optimised treatment plans that are independent of the local institute's dose-prescription protocol, thus enabling 'customised' dose prescription. Substantial increases in local tumour control may be achieved.

Introduction

Patients with advanced stage III non-small-cell lung cancer (NSCLC) treated with external beam radiation therapy (EBRT) have unfavourable outcomes [1]. Failure at the primary tumour site adversely influences progression-free and overall survival in NSCLC. To improve outcome, concurrent chemo-radiotherapy is employed but with minor therapeutic gain at the cost of increased acute toxicity [2]. Intensification of the radiotherapy regimen by reducing

the overall treatment has been shown to be beneficial for local control and survival [1, 3, 4]. Escalating the prescribed dose has been proven promising in the phase I/II RTOG 0117 study [5, 6]. However, the subsequent phase III RTOG 0617 clinical trial, comparing 60 Gy versus 74 Gy with concurrent chemotherapy, was prematurely stopped based on excess cardiac-related deaths in the escalated dose arm [7]. Generally, dose escalation in NSCLC patients is hampered by organs at risk (OARs), with symptomatic Grade ≥ 2 radiation-induced pneumonitis (RIP) being the most commonly reported dose-limiting side-effect [8]. Shrinking-field techniques [10], highly conformal dose-delivery techniques, such as intensity-modulated radiation therapy (IMRT) [11, 12], or inhomogeneous radiation beams [12, 13] enable the radiation treatment plan to be tailored, with a possible increase of the tumour dose, depending on the location and size of the target volume in relation to the organs at risk.

Escalation of the dose to the tumour in an existing treatment plan can be achieved by different approaches [14]. Firstly, by re-normalisation (RN; *i.e.* scaling the monitor units while maintaining the beam configuration, weights and intensity modulation pattern) of the plan by altering the number of fractions at constant fraction dose [15] or by changing the dose per fraction at constant number of fractions [16]. Hence, the original treatment plan can be altered until it reaches pre-defined iso-toxic (IT; *i.e.* sustaining the OAR dose of the original treatment plan) or maximum tolerable (MT; *i.e.* increasing the dose to the OAR while keeping within their tolerance dose level) normal tissue dose constraints. A recent *in silico* study in 38 stage IIIA/B NSCLC patients proved the latter approach to be beneficial for IT or MT re-normalisation with an increase in tumour effective dose up to 8.3% and 36.6%, respectively [16]. A second strategy is re-optimisation (RO) of the initial treatment plan using a trial-and-error approach to find the highest achievable tumour dose (HATD) in keeping with the normal-tissue dose constraints [12, 13, 17]. This approach is neither time-efficient, nor does it guarantee the HATD to be reached. Dose-response relationships have become available from published clinical data to serve as a basis for further dose trials [1, 8] and can potentially be exploited not only to evaluate (existing) dose distributions, but also to create new ones. A third option is *physico-biological optimisation* (PBO) where the tumour control probability (TCP) and normal tissue complication probability (NTCP) are used as objective and constraint functions to drive the inverse treatment planning process [14, 18]; a constraint regarding the homogeneity of the tumour dose distribution is optional.

Available studies on dose escalation in NSCLC patients have explored the potential of a shrinking field technique [10], beam inhomogeneity using a trial-and-error approach [12, 13], or the benefit of advanced treatment planning and delivery techniques [11, 12]. These studies use physical dose for plan optimisation and some of them subsequently calculated TCP values. Having ourselves studied the effects of IT and MT dose re-normalisation in a previous paper [16], this *in silico* study aims to explore the potential of PBO to further escalate the target dose on an individual patient level, both exploiting homogeneous and heterogeneous target doses. The HATD obtained by PBO and estimated TCP are studied in relation to the original treatment plan as a function of the mean lung dose and associated NTCP, which are primary constraints, with and without the target dose homogeneity constraint.

Methods and materials

Study design

An *in silico* study was conducted with the aim of increasing local tumour control by individualised dose escalation, based on either an IT or an MT approach with Grade ≥ 2 RIP as the toxicity endpoint of interest. For all patients the original treatment plan consisted of 66 Gy in 33 daily 2 Gy fractions delivered using IMRT. For this study, they were re-planned for both the IT and MT study arms to determine the HATD for either constant or maximum allowed risk of RIP, respectively. For both arms, the HATD after PBO was determined by homogeneous and fully heterogeneous target dose escalation. See Figure X.1 for a schematic representation of the study design.
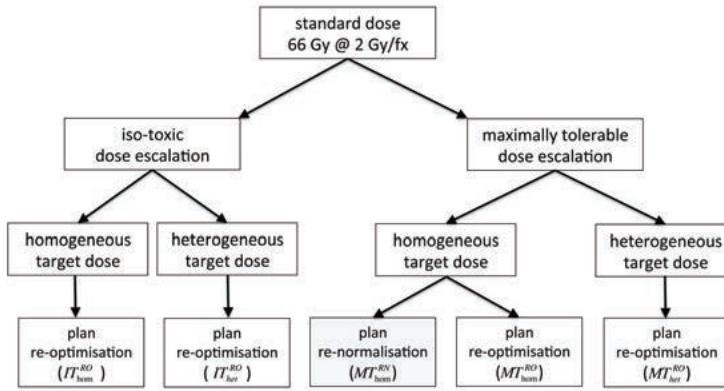
Figure X.1: Schematic representation of the study design. (*Abbreviations*: IT_{hom}^{RO} = iso-toxic treatment plan derived by re-optimisation with uniformity constraint; IT_{het}^{RO} = iso-toxic treatment plan derived by re-optimisation without uniformity constraint; MT_{hom}^{RN} = maximum tolerable treatment plan derived by re-normalisation with uniformity constraint; MT_{hom}^{RO} = maximum tolerable treatment plan derived by re-optimisation with uniformity constraint; MT_{het}^{RO} = maximum tolerable treatment plan derived by re-optimisation without uniformity constraint.)

Patient characteristics

Ten randomly selected patients with Stage IIIA/B NSCLC treated at the Radboud University Nijmegen Medical Centre with curative intent by sequential or concurrent chemoradiotherapy were included in this retrospective study (Table X.1).

Target and normal tissue volume definition

Initial treatment plans were retrieved from the Pinnacle³ treatment planning system (TPS) archive (version 8.oh; Philips Radiation Oncology Systems, Fitchburg, USA). The gross tumour volume of the primary tumour (GTV_T) and of the metastatic hilar/mediastinal lymph nodes (GTV_N) was defined on simultaneously acquired contrast-enhanced and slow-CT scans

Table X.1: Patient characteristics (N = 10). (*Abbreviations*: GTV = gross tumour volume; PTV = planning target volume; RT = radiation therapy; CHT = chemotherapy.)

Age (years; median [range])	66.4 [47.5–76.1]
Tumour location (no. of patients)	
left upper lobe	3
left lower lobe	3
left lung	1
right upper lobe	1
right lower lobe	1
unknown*	1
Stage (no. of patients)	
IIIA	6
IIIB	4
T-stage (no. of patients)	
T0*	1
T1	3
T2	2
T3	0
T4	4
N-stage (no. of patients)	
N0	1
N1	1
N2	6
N3	2
Tumour volumes (cm³; median [range])	
GTV	104.0 [28–331]
PTV	460.5 [165–1020]
Treatment (no. of patients)	
concurrent RT and CHT	6
sequential RT and CHT	4

* primary tumour not detected by CT, PET or pathology.

taking into account the metabolic tumour information of the ^{18}F-fluorodeoxyglucose PET-scan [19–21]. The total GTV was the union of GTV_T and GTV_N. The clinical target volumes (CTV_T and CTV_N) were created by uniform expansion of the respective GTVs by 5 mm, correcting for anatomical boundaries. The planning target volume (PTV) was the union of the CTV_T expanded by 10 mm and the CTV_N expanded by 5 mm.

The contouring of the lungs and heart was done automatically by the TPS and manually corrected if necessary. The oesophagus was delineated from the lower border of the cricoid cartilage to the gastro-oesophageal junction. For the spinal cord the inner border of the entire bony thoracic spinal canal was delineated.

Treatment planning technique
An IMRT treatment plan was generated using a 3D convolution/superposition method for dose calculation, and a standard radiation beam geometry not encompassing the contralateral lung. Multi-segment fields were generated for IMRT delivery on a step-and-shoot linear accelerator (Elekta SLi; Elekta AB, Stockholm, Sweden), using six co-planar 10 MV photon beams and applying the direct machine parameter optimisation algorithm (DMPO, P³IMRT version 8.0h, Philips Radiation Oncology Systems). Plans were limited to 60 segments with a mini-

mum segment area of 6 cm² and at least 10 monitor units. All plans had been normalised to a GTV mean total dose of 66 Gy satisfying the 95% and 107% criteria according to ICRU 50/62. The PTV mean total dose was at least 98% of the prescribed tumour dose. A collapsed-cone algorithm with tissue heterogeneity corrections was applied for accurate dose calculation.

Normal tissue dose constraints
The 2 Gy normalised total dose (NTD) constraints for the relevant organs at risk were: non-involved lung tissue (lungs minus GTV_T): $NTD_{mean} \leq 19$ Gy for Grade ≥ 2 RIP ($\alpha/\beta = 4$ Gy) [8]; heart: $NTD_{mean} \leq 26$ Gy for pericarditis ($\alpha/\beta = 3$ Gy) [8]; oesophagus: $NTD_{max} \leq 80$ Gy and $NTD_{mean} \leq 34$ Gy for late esophagitis/oesophageal stricture ($\alpha/\beta = 3$ Gy) [9]; spinal cord: $NTD_{max} \leq 50$ Gy for myelopathy ($\alpha/\beta = 2$ Gy) [22].

Plan re-normalisation
The easiest method to homogeneously escalate the dose was to re-normalise the total monitor units per fraction of the initial plan until the first normal tissue dose constraint was met [16]. This only applies to the MT approach, indicated as MT_{hom}^{RN} in Figure X.1. All other approaches require the plan to be re-optimised.

Plan re-optimisation
All initial plans from Pinnacle³ were imported into the non-clinical research software OR-BIT Workstation (version 1.5, RaySearch Laboratories AB, Stockholm, Sweden) [23]. As opposed to Pinnacle³, ORBIT Workstation is capable of physico-biological treatment plan optimisation using TCP/NTCP values. In order to validate the quality of the initial plan in OR-BIT Workstation, the imported plans were first re-optimised in ORBIT Workstation maintaining the initial settings (MLC, monitor units per segment) and the same objectives and constraints used in Pinnacle³ (with identical MLC settings and minimum monitor units per segment). Then, they were re-optimised with DMPO using constrained PBO, where the TCP of the GTV was maximised subject to the above normal tissue NTD constraints and a user-adjustable $NTCP_{RIP}$ constraint value for the non-involved lung tissue. This value was varied in steps of 1% between the one obtained for the initial plan and a pre-set maximum of 19% that corresponds to $NTD_{mean} = 19$ Gy for Grade ≥ 2 RIP [8], to implement the IT and MT approach, respectively, keeping within the predefined doses to the other organs at risk. For each $NTCP_{RIP}$ constraint value, a new optimisation problem was solved, and the resulting TCP and $NTCP_{RIP}$ were evaluated. Hence, a trade-off relationship between TCP and $NTCP_{RIP}$ could be obtained for every individual patient at each value of $NTCP_{RIP}$ up to the maximum of 19%. For the homogeneous cases, a target dose uniformity constraint in the form of a maximum allowable relative standard deviation for the PTV dose was added, equalling that of the initial plan. For the heterogeneous cases, the target dose uniformity was unrestricted. A minimum PTV dose constraint of 66 Gy was used to prevent 'cold spots'.

TCP and NTCP models

The following dose-response models and parameters were taken from literature for plan evaluation and optimisation. For both TCP and NTCP, a voxel based NTD was calculated in order to account for fractionation effects. For TCP calculation a linear-quadratic Poisson model for heterogeneous dose distributions [24], as available in ORBIT Workstation, was applied on the GTV's dose distribution $\mathbf{D}^{\mathrm{GTV}}$:

$$TCP(\mathbf{D}^{\mathrm{GTV}}) = \prod_i \left(\exp\left(-\exp\left(e \cdot \gamma_{37} - (e \cdot \gamma_{37} - \ln(\ln 2))\frac{NTD_{10}(D_i^{\mathrm{GTV}})}{TD_{50}}\right)\right) \right)^{\nu_i},$$

where NTD_{10} is the NTD function for α/β = 10 Gy, D_i^{GTV} and ν_i are the physical dose bins and relative volumes of the GTVs dose-volume histogram (DVH), respectively. Model parameters TD_{50} = 72.5 Gy and γ_{37} = 1.6 fitted the frequently applied logistic model of Martel *et al.* [25] for local progression-free survival at 30 months [1]. No correction for overall treatment time was applied, since the number of fractions was kept at 33 for all plans.

For the NTCP calculation, the Lyman-Kutcher-Burman model [26, 27] with fractionation correction was applied to the dose distribution $\mathbf{D}^{\mathrm{NLT}}$ of the non-involved lung tissue (NLT):

$$NTCP(\mathbf{D}^{\mathrm{NLT}}) = \Phi\left(\frac{NTD_{\mathrm{mean}}(\mathbf{D}^{\mathrm{NLT}}) - TD_{50}}{m \cdot TD_{50}} \right),$$

where $\Phi(z)$ is the standard normal cumulative distribution function, and

$$NTD_{\mathrm{mean}}(\mathbf{D}^{\mathrm{NLT}}) = \sum_i \nu_i NTD_{\alpha/\beta}(D_i^{\mathrm{NLT}})$$

is the mean-NTD function converting the physical DVH into an NTD-equivalent DVH. Model parameters for Grade $\geq$ 2 RIP (TD_{50} = 31.4 Gy, m = 0.45, α/β = 4 Gy) were adopted from Marks *et al.* [8].

Statistical analysis

Results were analysed for every patient in the cohort individually and for the entire cohort. Descriptive statistics of the HATD in terms of the mean target dose plus its relative standard deviation were reported. Median values for these parameters were calculated and compared between the study arms using the Wilcoxon signed rank test (GraphPad Prism, version 4.0, GraphPad Software, La Jolla, USA) with $p < 0.05$ being considered a statistically significant difference. The dominant dose-limiting constraint was also assessed.

Results

Individual risk-benefit trade-off

By way of example, the trade-off relationship between TCP and $NTCP_{\mathrm{RIP}}$ is shown in Figure X.2 for patient 3 (cT4N1M0, left lower lobe, GTV = 144 cm³). This figure shows the increase in TCP by re-optimised dose escalation over the $NTCP_{\mathrm{RIP}}$ interval that is considered clinically acceptable. The TCP and $NTCP_{\mathrm{RIP}}$ of the initial plan is 34.5% and 12.9%, respectively. The lung

constraint is neither dose-limiting for the MT_{hom}^{RN} nor for the MT_{hom}^{RO} plan, but dose-limiting for the MT_{het}^{RO} plan. Iso-toxic re-optimisation shows an increase in TCP of 11.6% and 34.5% for the IT_{hom}^{RO} and the IT_{het}^{RO} plan relative to the initial plan, respectively. The TCP of the MT_{hom}^{RN} plan can be improved by 9.7% and 26.6% for the MT_{hom}^{RO} and the MT_{het}^{RO} plans, respectively, at the cost of an increase in $NTCP_{RIP}$ of 2.0% and 2.4%. Corresponding DVHs and relevant dose levels of these plans are reported in Figure X.4 and Table X.5 (see Supplementary Data).
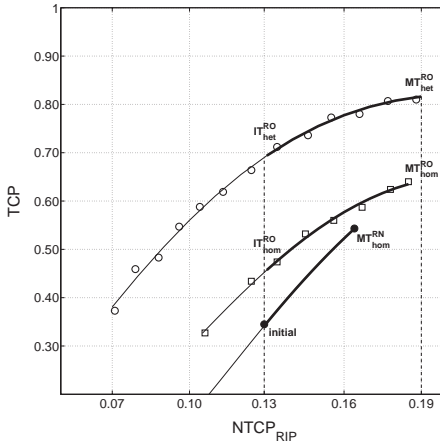

Figure X.2: Trade-off relationships between TCP and $NTCP_{RIP}$ for patient 3, obtained by re-normalisation (●) or re-optimisation (○, □) of the initial treatment plan. Data points IT_{hom}^{RO}, IT_{het}^{RO}, MT_{hom}^{RN}, MT_{hom}^{RO}, MT_{het}^{RO} correspond to outcome scores of the associated treatment plans. (*Abbreviations*: see Figure X.1.)

Benefit for individual patients

Figure X.3 shows the TCP and $NTCP_{RIP}$ for all individual patients and every study arm. For the initial treatment plan, TCP was 34% and $NTCP_{RIP}$ varied between 6.4% and 16.4%. For the homogeneous approaches, the IT_{hom}^{RO} arm showed the least benefit, with TCP ranging from 34−49.5%, with 4/10 patients having an increase less than 5%. For MT_{hom}^{RN} and MT_{hom}^{RO}, only one patient had an increase less than 5%, while TCP ranged from 37.7−65.0% and 37.8−73.8%, respectively. In both IT arms, $NTCP_{RIP}$ values were similar to the initial plan, independent of the target dose uniformity constraint. The MT arms showed two clusters of patients: in 7/10 patients dose escalation concurred with a significant increase in $NTCP_{RIP}$ approaching the maximum tolerable level, and in 3/10 patients the increase was less than 3%. For the latter group, evidently another constraint was dose limiting. Comparing the MT_{hom}^{RN} and MT_{hom}^{RO} arms, it appears that re-optimisation yielded a higher TCP in 4/10 patients than for re-normalisation, whereas for the remaining patients TCP was at least 5% lower with re-optimisation than with re-normalisation. For $NTCP_{RIP}$ there was a statistically significant, although clinically irrelevant, difference between these two arms in favour of MT_{hom}^{RN} ($p = 0.037$). Therefore, the benefit from re-optimisation for the MT approach under homogeneous conditions is questionable for

the individual patient. This does not hold for the heterogeneous arms. For all individual patients, omitting the target dose uniformity constraint resulted in a significant increase in TCP.
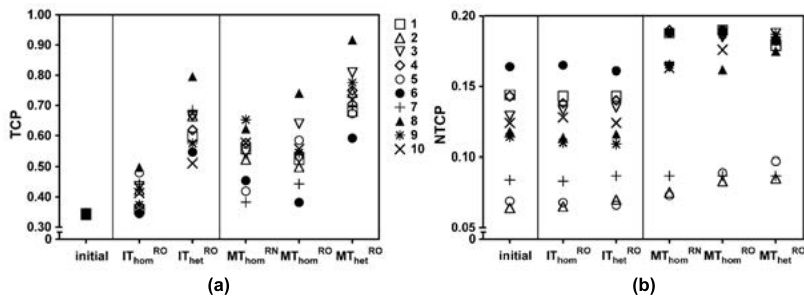
Figure X.3: (a) Tumour control probability and (b) normal tissue complication probability related to Grade $\geq$ 2 radiation-induced pneumonitis for all individual patients (1 to 10) over the 5 study arms, including the initial treatment plan. (*Abbreviations*: see Figure X.1.)

Comparison of approaches for patient cohort

Tables X.2 and X.3 show the median increase in TCP and $NTCP_{RIP}$ for all study arms relative to the initial treatment plan. There was a significant increase in TCP for all study arms relative to the initial plan, and also between the arms, except for MT_{hom}^{RN} and MT_{hom}^{RO} (p = 0.385). By definition, there was no increase of $NTCP_{RIP}$ with the two IT arms, nor between the $NTCP_{RIP}$ of these arms and that of the initial treatment plan. The $NTCP_{RIP}$ increase was not statistically significant amongst the three MT arms (p = 0.371). However, it was significant for all these arms relative to the initial treatment plan.

Table X.2: Increase in tumour control probability (ΔTCP) and normal tissue complication probability (ΔNTCP) between the initial and the re-optimised (RO) treatment plans of the iso-toxic approach (median [range]; in %). (*Abbreviations*: see Figure X.1.)

	IT_{hom}^{RO}	IT_{het}^{RO}
ΔTCP	7.5 [0−15.5]	29.7* [16.7−45.3]
ΔNTCP	0.1 [-0.5−0.4]	0.1[n.s.] [-0.5−0.6]

*p < 0.01; n.s. = non-significant.

Table X.4 (see Supplementary Data) shows relevant dosimetric parameters for the target volume over all study arms. Evidently, there is no significant difference in PTV uniformity between the initial plan and the ones in the homogeneous arms. The median [range] of the PTV mean fractional dose was 2.0 [2.0−2.1] Gy for IT_{hom}^{RO}, 2.2 [2.0−2.4] Gy for MT_{hom}^{RN}, and 2.2 [2.0−2.3] Gy for MT_{hom}^{RO}. Without homogeneity constraint, the PTV uniformity changed significantly between the homo- and heterogeneous study arms, and so did the median [range] of the PTV mean fractional doses: 2.2 [2.0−2.4] Gy for IT_{het}^{RO} and 2.3 [2.1−2.7] Gy for MT_{het}^{RO}.

Table X.3: Increase in tumour control probability (ΔTCP) and normal tissue complication probability (ΔNTCP) between the initial, the re-optimised (RO) and the re-normalised (RN) treatment plans of the maximally-tolerable approach (median [range]; in %). (*Abbreviations*: see Figure X.1.)

	MT^{RN}_{hom}	MT^{RO}_{hom}	MT^{RO}_{het}
ΔTCP	20.7 [3.7–31.0]	19.5$^{n.s.}$ [3.8–39.8]	38.8* [24.8–57.4]
ΔNTCP	3.7 [0.3–5.1]	4.4** [0.3–7.3]	3.8$^{n.s.}$ [0.3–7.3]

* $p < 0.01$; ** $p < 0.05$; n.s. = non-significant.

Dominant dose-limiting organ

For 4/10 patients, the dominant dose-limiting organ did not change over the study arms; this was the lung tissue for three patients and the heart for one. For all ten patients, the dose-limiting organ remained the same when the homogeneity constraint was omitted. For both IT arms, in 8/10 patients the non-involved lung tissue was dose-limiting, whereas for the other 2/10 patients, these were the heart and the oesophagus. For both MT arms that were re-optimised, five, three, and two patients experienced dose limitation by the non-involved lung tissue, the heart and the NTD_{max} of the oesophagus, respectively. For the re-normalised MT arm, the dominant constraint was the NTD_{max} of the oesophagus (5/10 patients), followed by 3/10 patients for who the non-involved lung tissue was dose-limiting. In the other two patients, either the NTD_{mean} for the heart or for the oesophagus impeded dose escalation.

Discussion

Recently, a dose-response relationship for NSCLC based on clinical data has been established from a systematic literature review [1], indicating that (even modest) increases in dose are worthwhile. The challenge, however, is along which path dose escalation can be best achieved, taking into account the relevant dose-limiting organs. An easy to implement approach is the re-normalisation of an existing treatment plan at standard fractionation. Another, more time-consuming option is the re-optimisation of the plan for a different dose prescription. However, for an individualised approach, it is *a priori* not clear what the tumour dose prescription should be and whether this is feasible given the specific EBRT technique and normal-tissue dose constraints.

The traditional way is to escalate the prescription dose to the same level for every patient in the specific patient population, also referred to as 'Level nil' radiobiological optimisation [14]. However, despite the improved normal-tissue sparing and tumour-targeting ability of modern highly-conformal EBRT techniques such a "one-size-fits-all" strategy would hamper the potential to 'customise' the dose prescription in relation to the individual anatomy, *i.e.* the location and size of the target volume relative to the OARs.

So far, population-based dose escalation in clinical trials has been accomplished either by increasing the fraction size [3, 16, 28] or by increasing the fraction number [15] in small increments uniformly over all patients within the study cohort, until a pre-set maximum tolerable

dose limit for one of the OARs was reached. Such an approach has resulted in modest gains in local control rates for an acceptable increase in side effects. The question that remains unanswered is what HATD can be safely realised for an individual patient using a specific inverse treatment planning and dose delivery technique. To answer this question, we have applied a physico-biological inverse treatment planning method that allows a TCP function to be maximised subject to pre-set dosimetric criteria for the PTV and OARs. Since the HATD evidently depends on the degree of target dose heterogeneity, we investigated this effect by varying an adequate PTV uniformity measure between the two extremes: its value for the initial plan that satisfies the ICRU-criteria and infinity. The HATD is also restricted by the NTD_{mean} constraint for the non-involved healthy lung tissue, which is associated with Grade ≥ 2 RIP. Therefore, we varied the NTD_{mean} constraint value over a clinically relevant interval under invariant maximum tolerable dose constraints for the other OARs. In order to compare the HATD for the study arms, TCP scores were calculated and the NTD_{mean} was converted into an NTCP score. Consequently, we were able to construct the Pareto efficient frontier (see *e.g.* [29]) in the TCP/NTCP domain for individual patients, which shows the boundary between physically realisable and non-realisable dose distributions.

Iso-toxic re-optimisation of the initial plan was sub-optimal in terms of HATD at restricted target dose uniformity. PBO has the ability to provide a significantly higher GTV and PTV mean total dose, and guarantees the HATD to be achieved while respecting the relevant constraints imposed. This allows for a new dose prescription paradigm, where the optimal dose level for every individual patient is determined by a protocol- and treatment-planner-independent optimisation methodology. The HATD will then be prescribed once the trade-off in terms of target dose homogeneity/heterogeneity and risk of potential side-effects has been carefully balanced by the physician's and patient's preferences.

Our results suggest that the largest increase in TCP can be achieved by abandoning the homogeneous target dose dogma but it remains to be proven that this yields improved local control and survival. Maybe the maximum benefit will ultimately be obtained if the high dose is selectively steered to regions or voxels that coincide with spatio-temporal information on tumour characteristics (*e.g.* distribution of hypoxia and clonogenic tumour cells) based on functional imaging modalities (*e.g.* [17, 30]). Still, this has to be confirmed.

Our study has some limitations. Firstly, the radiobiological models have uncertain parameter values. Hence, the TCP and NTCP predictions should not be taken as absolute values. Instead, they should be used to compare the relative performance of competing dose distributions. Clinical decision-making should therefore not solely be based on these values, but should always include an evaluation of the 3D dose distribution. Nevertheless, a sensitivity analysis (results not shown) indicated that the HATD is robust under uncertainties in the TCP model parameters. Secondly, we did not investigate the effect of varying the number of fractions on the TCP/NTCP trade-off boundaries. This could be added as an independent variable in the physico-biological treatment planning procedure, which implies that one would need to perform repeated series of optimisation runs, varying the number of fractions each time. Thirdly, we did not take tumour motion due to breathing or systematic/random errors into account which may all influence the dose delivered to the primary tumour [31]. Employing

anisotropic margins compensating for these effects, or implementing techniques controlling the breathing cycle [32] may be ways to reduce the effects on TCP as well as NTCP.

Conclusions

Physico-biological optimisation, maximising a homogeneous or heterogeneous target dose prescription provides a trade-off of TCP versus NTCP as a means to generate optimised treatment plans that are independent of the local institute's dose-prescription protocol, and allows for individualised balancing of treatment benefit and injury. This *in silico* trial suggests that physico-biological optimisation has the ability to produce deliverable treatment plans with an increased highest achievable 'customised' tumour dose that is within a clinically acceptable range and yields a significant increase in estimated TCP.

References

1. M. Partridge, M. Ramos, A. Sardaro, and M. Brada. Dose escalation for non-small cell lung cancer: analysis and modelling of published literature. *Radiother Oncol*, 99(1):6–11, 2011.

2. A. Aupérin, C. Le Péchoux, E. Rolland, W.J. Curran, K. Furuse, P. Fournel, et al. Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small-cell lung cancer. *J Clin Oncol*, 28(13):2181–90, 2010.

3. J.D. Fenwick, A.E. Nahum, Z.I. Malik, C.V. Eswar, M.Q. Hatton, V.M. Laurence, et al. Escalation and intensification of radiotherapy for stage III non-small cell lung cancer: opportunities for treatment improvement. *Clin Oncol (R Coll Radiol)*, 21(4):343–60, 2009.

4. A. van Baardwijk, S. Wanders, L. Boersma, J. Borger, M. Öllers, A.-M. Dingemans, et al. Mature results of an individualized radiation dose prescription study based on normal tissue constraints in stages I to III non-small-cell lung cancer. *J Clin Oncol*, 28(8):1380–6, 2010.

5. J.D. Bradley, J. Moughan, M.V. Graham, R. Byhardt, R. Govindan, J. Fowler, et al. A phase I/II radiation dose escalation study with concurrent chemotherapy for patients with inoperable stages I to III non-small-cell lung cancer: phase I results of RTOG 0117. *Int J Radiat Oncol Biol Phys*, 77(2):367–72, 2010.

6. J.D. Bradley, K. Bae, M.V. Graham, R. Byhardt, R. Govindan, J. Fowler, et al. Primary analysis of the phase II component of a phase I/II dose intensification study using three-dimensional conformal radiation therapy and concurrent chemotherapy for patients with inoperable non-small-cell lung cancer: RTOG 0117. *J Clin Oncol*, 28(14):2475–80, 2010.

7. J.D. Cox. Are the results of RTOG 0617 mysterious? *Int J Radiat Oncol Biol Phys*, 82(3):1042–44, 2012.

8. L.B. Marks, S.M. Bentzen, J.O. Deasy, F.-M. Kong, J.D. Bradley, I.S. Vogelius, et al. Radiation dose-volume effects in the lung. *Int J Radiat Oncol Biol Phys*, 76(3 Suppl):S70–6, 2010.

9. M. Werner-Wasik, E. Yorke, J. Deasy, J. Nam, and L.B. Marks. Radiation dose-volume effects in the esophagus. *Int J Radiat Oncol Biol Phys*, 76(3 Suppl):S86–93, 2010.

10. M. Engelsman, P. Remeijer, M. van Herk, J.V. Lebesque, B.J. Mijnheer, and E.M.F. Damen. Field size reduction enables iso-NTCP escalation of tumor control probability for irradiation of lung tumors. *Int J Radiat Oncol Biol Phys*, 51(5):1290–8, 2001.

11. S.L.A. Govaert, E.G.C. Troost, O.C.J. Schuurbiers, L.-F. de Geus-Oei, A. Termeer, P.N. Span, and J. Bussink. Treatment outcome and toxicity of intensity-modulated (chemo)radiotherapy in stage III non-small cell lung cancer patients. *Radiat Oncol*, 7:150, 2012.

12. M. Guckenberger, A. Kavanagh, and M. Partridge. Combining advanced radiotherapy technologies to maximize safety and tumor control probability in stage III non-small cell lung cancer. *Strahlenther Onkol*, 188(10):894–900, 2012.

13. M. Schwarz, M. Alber, J.V. Lebesque, B.J. Mijnheer, and E.M.F. Damen. Dose heterogeneity in the target volume and intensity-modulated radiotherapy to escalate the dose in the treatment of non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys*, 62(2):561–70, 2005.

14. A.E. Nahum and J. Uzan. (Radio)biological optimization of external-beam radiotherapy. *Comput Math Methods Med*, 2012:329214, 2012.

15. A. van Baardwijk, G. Bosmans, S.M. Bentzen, L. Boersma, A. Dekker, R. Wanders, et al. Radiation dose prescription for non-small-cell lung cancer according to normal tissue dose constraints: an *in silico* clinical trial. *Int J Radiat Oncol Biol Phys*, 71(4):1103–10, 2008.

16. A.L. Hoffmann, E.G.C. Troost, H. Huizenga, J.H.A.M. Kaanders, and J. Bussink. Individualized dose prescription for hypofractionation in advanced non-small-cell lung cancer radiotherapy: an *in silico* trial. *Int J Radiat Oncol Biol Phys*, 83(5):1596–602, 2012.

17. W. van Elmpt, D. De Ruysscher, A. van der Salm, A. Lakeman, J. van der Stoep, D. Emans, E. Damen, M. Öllers, J.-J. Sonke, and J. Belderbos. The PET-boost randomised phase II dose-escalation trial in non-small cell lung cancer. *Radiother Oncol*, 104(1):67–71, 2012.

18. A. Brahme. Biologically based treatment planning. *Acta Oncol*, 38(13):61–8, 1999.

19. D. De Ruysscher, S. Wanders, E. van Haren, M. Hochstenbag, W. Geeraedts, I. Utama, et al. Selective mediastinal node irradiation based on FDG-PET scan data in patients with non-small-cell lung cancer: a prospective clinical study. *Int J Radiat Oncol Biol Phys*, 62(4):988–94, 2005.

20. D. De Ruysscher, R.-H. Bremer, F. Koppe, S. Wanders, E. van Haren, M. Hochstenbag, et al. Omission of elective node irradiation on basis of CT-scans in patients with limited disease small cell lung cancer: a phase II trial. *Radiother Oncol*, 80(3):307–12, 2006.

21. J.S.A. Belderbos, W.D. Heemsbergen, K. De Jaeger, P. Baas, and J.V. Lebesque. Final results of a Phase I/II dose escalation trial in non-small-cell lung cancer using three-dimensional conformal radiotherapy. *Int J Radiat Oncol Biol Phys*, 66(1):126–34, 2006.

22. J.P. Kirkpatrick, A.J. van der Kogel, and T.E. Schultheiss. Radiation dose–volume effects in the spinal cord. *Int J Radiat Oncol Biol Phys*, 76(3 Suppl):S42–49, 2010.

23. K. Eriksson and H. Rehbinder. Advanced tools for radiobiological evaluation and optimisation of treatment plans. In *Proceedings of the XVth International Conference on the Use of Computers in Radiation Therapy (ICCR)*, volume II, pages 60–4, Toronto, 2007.

24. B.K. Lind, P. Mavroidis, S. Hyödynmaa, and C. Kappas. Optimization of the dose level for a given treatment plan to maximize the complication-free tumor cure. *Acta Oncol*, 38(6):787–98, 1999.

25. M.K. Martel, R.K. Ten Haken, M.B. Hazuka, M.L. Kessler, M. Strawderman, A.T. Turrisi, et al. Estimation of tumor control probability model parameters from 3-D dose distributions of non-small cell lung cancer patients. *Lung Cancer*, 24(1):31–7, 1999.

26. G.J. Kutcher, C. Burman, L. Brewster, M. Goitein, and R. Mohan. Histogram reduction method for calculating complication probabilities for three-dimensional treatment planning evaluations. *Int J Radiat Oncol Biol Phys*, 21(1):137–46, 1991.

27. J.T. Lyman and A.B. Wolbarst. Optimization of radiation therapy III: A method of assessing complication probabilities from dose-volume histograms. *Int J Radiat Oncol Biol Phys*, 13(1):103–9, 1987.

28. J.F. Lester, L. Nixon, P. Mayles, H. Mayles, Y. Tsang, A. Ionescu, et al. The I-START trial: ISoToxic Accelerated RadioTherapy in locally advanced non-small cell lung cancer. *Lung Cancer*, 75, Supplement 1:S51, 2012.

29. D. Craft, T. Halabi, and T. Bortfeld. Exploration of tradeoffs in intensity-modulated radiotherapy. *Phys Med Biol*, 50(24):5857–68, 2005.

30. G. Meijer, J. Steenhuijsen, M. Bal, K. De Jaeger, D. Schuring, and J. Theuws. Dose painting by contours versus dose painting by numbers for stage II/III lung cancer: practical implications of using a broad or sharp brush. *Radiother Oncol*, 100(3):396–401, 2011.

31. J. Selvaraj, J. Uzan, C. Baker, and A. Nahum. Loss of local control due to tumor displacement as a function of margin size, dose-response slope, and number of fractions. *Med Phys*, 40(4):041715, 2013.

32. S.R. Bowen, M.J. Nyflot, M. Gensheimer, K.R.G. Hendrickson, P.E. Kinahan, G.A. Sandison, and S.A. Patel. Challenges and opportunities in patient-specific, motion-managed and PET/CT-guided radiation therapy of lung cancer: review and perspective. *Clin Transl Med*, 1(1):18, 2012.
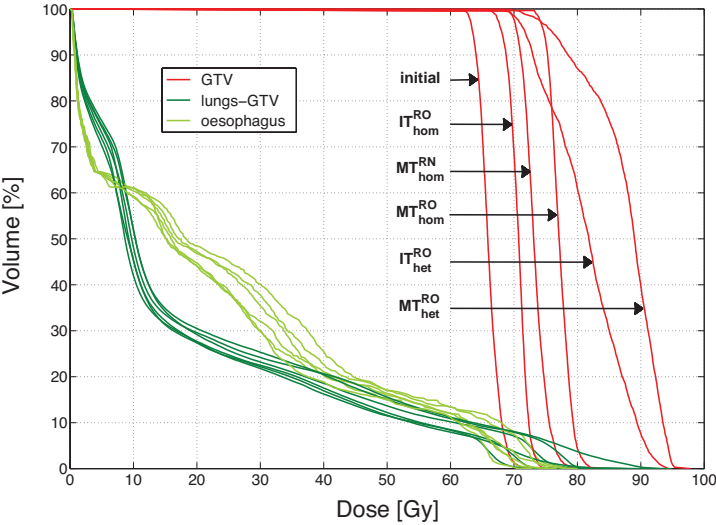
Supplementary data

Figure X.4: Dose-volume histograms of relevant treatment plans of patient 3.

Table X.4: Median and [range] of tumour total dose in patient cohort for different study arms. (*Abbreviations:* D_{mean} = mean total dose; $D_{1\%}$ = minimum total dose to hottest 1% of volume; σ/D_{mean} = relative standard deviation.)

	dose statistic		initial	IT_{hom}^{RO}	IT_{het}^{RO}	MT_{hom}^{RN}	MT_{hom}^{RO}	MT_{het}^{RO}
GTV	D_{mean}	[Gy]	66.0 [66.0–66.0]	68.8* [66.0–71.6]	78.8* [72.6–93.3]	73.9* [67.3–78.4]	72.9* [67.5–82.2]	82.8* [76.0–100.9]
	$D_{1\%}$	[Gy]	70.5 [69.2–72.0]	72.5* [70.9–76.1]	89.3* [78.3–101.2]	79.3* [70.0–83.9]	77.7* [72.2–81.4]	94.1* [86.2–115.7]
PTV	D_{mean}	[Gy]	65.8 [64.7–66.8]	67.1* [64.9–70.6]	71.3* [67.2–78.0]	73.7* [67.6–78.2]	71.8* [66.4–77.3]	76.2* [70.5–89.7]
	σ/D_{mean}	[%]	3.6 [2.4–5.4]	3.6$^{n.s.}$ [2.3–5.9]	8.5* [6.7–12.1]	3.6 [2.4–5.4]	3.7$^{n.s.}$ [2.4–5.2]	8.8* [7.4–13.1]

* $p < 0.01$; n.s. = non-significant.

Table X.5: Tumour and normal tissue structure dose statistics for all six treatment plans of patient 3. (*Abbreviations:* D_{mean} = mean total dose; $D_{1\%}$ = minimum total dose to hottest 1% of volume; σ/D_{mean} = relative standard deviation; NTD_{mean} = mean normalised total dose; NTD_{max} = maximum normalised total dose.)

	dose statistic		initial	IT_{hom}^{RO}	IT_{het}^{RO}	MT_{hom}^{RN}	MT_{hom}^{RO}	MT_{het}^{RO}
GTV	D_{mean}	[Gy]	66.0	70.6	81.5	73.5	77.2	87.7
	$D_{1\%}$	[Gy]	70.2	73.9	92.9	78.2	81.4	95.0
PTV	D_{mean}	[Gy]	66.8	70.4	75.7	74.4	77.3	81.6
	σ/D_{mean}	[%]	3.3	3.3	8.9	3.3	3.3	8.8
lungs–GTV	NTD_{mean}	[Gy]	15.4	15.7*	15.8*	17.6	18.7	18.9*
esophagus	NTD_{max}	[Gy]	68.5	72.0	73.7	79.9*	79.7*	77.6
	NTD_{mean}	[Gy]	19.7	19.6	19.3	22.7	21.5	19.0
cord	NTD_{max}	[Gy]	24.3	31.5	31.8	28.0	25.7	23.2
heart	NTD_{mean}	[Gy]	10.9	10.4	10.1	12.5	12.1	11.4

* Dose-limiting constraint.

Appendices

Samenvatting (Summary in Dutch)

Moderne bestralingstechnieken zoals intensiteitsgemoduleerde radiotherapie (IMRT) hebben het risico op beschadiging van gezonde, normale weefsels in vergelijking met conventionele en drie-dimensionale conformale radiotherapie (3D-CRT) substantieel verkleind, door gebruik te maken van computerintensieve ontwerpmethoden voor optimalisatie van de dosisverdeling. Desondanks is het dosisvoorschrift voor patiënten met een zelfde bestralingsindicatie onveranderd uniform gebleven. Deze uniformiteit van het dosisvoorschrift impliceert echter geen uniformiteit van de bijwerkingen. De huidige behandelprotocollen waarin het dosisvoorschrift is 'bevroren' houden namelijk geen rekening met patiëntspecifieke kenmerken, zoals de anatomische variatie in tumorgrootte en -lokalisatie (ten opzichte van omliggende normale weefsels). Deze variatie biedt thans onbenutte mogelijkheden om het dosisvoorschrift te *individualiseren*. Verschillende strategieën kunnen daarvoor worden aangewend, zoals: maximaal tolerabele, isotoxische en risico-geadapteerde benaderingen. Deze zijn afhankelijk van de rol die de behandelend radiotherapeut-oncoloog en de individuele patiënt spelen bij de besluitvorming omtrent de afweging van voor- en nadelen die met de bestralingsbehandeling gepaard gaan. In de huidige computersystemen waarmee bestralingsplannen worden ontworpen ontbreken beslissingsondersteunende hulpmiddelen, die aangewend kunnen worden voor een systematische verkenning van bovengenoemde strategieën. Geavanceerde methoden voor computerondersteunde optimalisatie van het individuele behandelplan dienen in deze behoefte te voorzien. Het werk in deze dissertatie beschrijft verschillende aspecten van nieuwe rekenkundige methoden en hun toepassing voor klinisch relevante casuïstieken.

Het eerste artikel (**Paper I**) introduceert de rationale en het belang om het dogma van uniforme dosisvoorschriften voor patiënten met een zelfde bestralingsindicatie te doorbreken door voorkeuren met betrekking tot individuele risico-afweging te betrekken bij de selectie van een 'optimaal' behandelplan. Hiervoor worden op basis van verschillende dosisverdelingen schattingen gemaakt van kansen op tumorcontrole en complicaties van relevant geachte normale weefsels. De navolgende drie artikelen (**Papers II, III, IV**) beschrijven verschillende strategieën voor individualisering van het dosisvoorschrift teneinde de therapeutische prestaties van een bestaand behandelplan te verbeteren door middel van *renormalisatie*; door schaling van de relatieve dosisverdeling zonder deze opnieuw te ontwerpen. De laatste zes artikelen hebben betrekking op verschillende strategieën voor individualisering van het dosisvoorschrift, waarbij dosisverdelingen opnieuw worden ontworpen met behulp van intelligente methoden voor *heroptimalisatie*.

Deze methoden omvatten zowel *dosimetrische* (**Papers V, VI, VII**) als *radiobiologische* criteria (**Papers VIII, IX, X**), die respectievelijk de kwaliteit van de dosisverdeling en de te verwachten behandeluitkomst trachten te kwantificeren. In het laatste artikel worden de concepten uit de eerdere artikelen samengevoegd teneinde de renormalisatie en heroptimalisatie benaderingen met elkaar te vergelijken. Bij patiënten met een gevorderd stadium niet-kleincellig longcarcinoom (NSCLC) wordt individuele dosisescalatie middels IMRT toegepast onder condities waarbij de heterogeniteit van de dosisverdeling in het te bestralen doelvolume wordt beperkt, dan wel wordt losgelaten.

Inclusie van patiëntenvoorkeuren bij de bestralingsplanning van geïndividualiseerde radiotherapie

Een eerste strategie om een persoonlijk dosisvoorschrift te bewerkstelligen is risicovoorkeuren van de individuele patiënt mee te wegen. In **Paper I** wordt voorgesteld de voorkeuren van de patiënt in het proces van de bestralingsplanning te includeren met als doel de balans tussen de kans op tumorcontrole (TCP) en de kans op bijwerkingen in normale weefsels (NTCP), zoals geschat op basis van de vooraf berekende dosisverdeling, te individualiseren. IMRT maakt het mogelijk bestralingsplannen met verschillende dosisvoorschriften over een bereik van TCP- en NTCP-waarden te genereren. Het klinisch acceptabele bereik van bestralingsplannen van waaruit één 'beste compromis' gekozen moet worden, kan in verschillende mate van complexiteit worden onderzocht. De evaluatie hiervan door middel van de *therapeutic operating characteristic* (TOC) grafiek en de *Pareto efficient frontier* (PEF) lijkt geschikt voor respectievelijk eerder en nieuw ontworpen bestralingsplannen. Vanuit deze weergave kan de radiotherapeut-oncoloog behandelmogelijkheden kiezen en de risicovoorkeuren met de patiënt bespreken. De verschillen tussen deze behandelmogelijkheden kunnen in schattingen van de TCP en NTCP worden beoordeeld. Dit maakt het noodzakelijk nieuwe methoden voor beslissingsondersteuning te ontwikkelen en in bestaande bestralingsplanningssystemen te integreren.

Aanpassing van het dosisvoorschrift met behulp van de TOC-grafiek

In **Paper II** wordt het concept van de TOC-grafiek als keuzehulp toegepast om de afweging tussen TCP en NTCP te kwantificeren door middel van een *a posteriori* optimalisatie van een eerder ontworpen bestralingsplan. Voor dit bestralingsplan wordt het dosisvoorschrift nadien gerenormaliseerd middels 'Niveau I' optimalisatie (zie pagina 17), hetzij door variatie van de fractiegrootte (bij een constant aantal bestralingsfracties), hetzij door variatie van het aantal bestralingsfracties (bij constante fractiegrootte), teneinde het gehele bereik van TCP- en NTCP-waarden van de desbetreffende dosisverdeling in kaart te brengen. In tegenstelling tot het sterk bekritiseerde P_+ concept voor de 'complicatievrije tumorcontrolekans' worden voor het vinden van het 'optimale' dosisvoorschrift geen veronderstellingen gedaan betreffende de *a priori* voorkeuren voor TCP en NTCP. Het TOC-concept wordt voorgesteld aan de hand van een klinisch voorbeeld van prostaatkanker, waarbij de afwegingen tussen biochemische afwezigheid van tumorrecidief na 5 jaar, late gastro-intestinale en urogenitale morbiditeit voor een patiëntencohort alsmede voor een individuele patiënt zijn onderzocht.

De oppervlakte onder de TOC-grafiek wordt als nieuwe, dosisonafhankelijke index voor het therapeutische vermogen van een bestralingstechniek of -plan voorgesteld. Een vergelijk tussen 3D-CRT en IMRT plannen demonstreert het effect van een verhoogde therapeutische winst dat met moderne optimalisatie- en bestralingstechnieken kan worden bereikt. Voorts suggereren de resultaten dat door toenemende dosisconformaliteit en -heterogeniteit binnen het doelvolume het therapeutische venster zich opent en daarmee de therapeutische winst van geïndividualiseerde radiotherapie wordt vergroot.

Aanpassing van fractiedosis en -aantal bij bestraling van het niet-kleincellig longcarcinoom

Een alternatief om het dosisvoorschrift aan te passen is middels geïndividualiseerde dosis-escalatie, waarbij verschillende strategieën kunnen worden gevolgd. Bij de maximaal tolerabele benadering definieert de radiotherapeut-oncoloog de limieten voor maximaal toelaatbare dosis en/of complicatiekansen voor de relevante risico-organen, en wordt de voorgeschreven dosis van een reeds bestaand bestralingsplan voor een individuele patiënt *a posteriori* verhoogd totdat de meest kritische limiet wordt bereikt. Bij de isotoxische benadering wordt de voorgeschreven dosis individueel verhoogd tot aan het niveau waarop de kans op een bepaalde complicatie gelijk geacht wordt te zijn aan die van de rest van de patiëntengroep. Recentelijk zijn dergelijke behandelstrategieën voor de bestraling van het gevorderd NSCLC met succes in de kliniek geïntroduceerd. In deze protocollen wordt ofwel het aantal bestralingsfracties verhoogd met een gelijkblijvende fractiedosis, dan wel het tegenovergestelde, maar er vindt geen simultane aanpassing van de fractiedosis en het aantal fracties plaats. Tot dusverre werd bij dergelijke behandelstrategieën enkel rekening gehouden met de tolerantiedoses van het niet-geïnvolveerde longweefsel en het myelum, doch niet met die van andere potentiële risico-organen.

Paper III beschrijft de geïndividualiseerde dosisescalatie in een retrospectieve *in silico* studie bij 38 patiënten met gevorderd stadium NSCLC, waarbij niet alleen dosislimieten voor het niet-geïnvolveerde longweefsel en het myelum, maar ook die voor de oesophagus, de plexus brachialis en het hart zijn opgenomen. Zogenaamde 'Niveau II' optimalisatie (zie pagina 17) wordt toegepast door de fractiegrootte en het -aantal simultaan te variëren voor zowel maximaal tolerabele als ook isotoxische dosisescalatie. Aangetoond wordt dat in 79% van de patiënten een significante dosisescalatie met therapeutische winst mogelijk zou zijn geweest. De maximale dosis op de oesophagus blijkt bij de meeste patiënten het dominante dosisli-miterende risico-orgaan te zijn.

Het effectieve α/β concept verdisconteert dosisheterogeniteit en volume-effect bij de fractioneringsgevoeligheid van normale weefsels

Door toenemend bewijs dat radiotherapie met een hoge fractiedosis middels hoog-conformale bestralingstechnieken veilig kan worden toegepast, is hernieuwde interesse voor hypofractionering ontstaan. De gebruikelijke methode om uitgaande van een bestaand dosisvoorschrift een nieuw, iso-effectief bestralingsregime te berekenen maakt gebruik van de zogenaamde *Withers iso-effect formule* (WIF), welke gebaseerd is op het bekende radiobiologische model voor de lineair-kwadratische celoverlevingscurve. Dit model omvat maten voor de intrinsieke fractioneringsgevoeligheid, α/β, van zowel de tumor als van het meest relevant geachte dosislimiterende normale weefsel. De ratio van deze maten dient thans als instrument om de potentie van hypofractionering te beoordelen. Recente inzichten in de modellering suggereren echter dat voor tumorlokalisaties waarbij de specifieke anatomie een substantieel lagere dosis op het dosis-limiterende normale weefsel toelaat dan op de tumor, in potentie geschikt zijn voor hypofractionering ondanks dat voornoemde ratio onvoordelig lijkt te zijn.

In **Paper IV** wordt betoogd dat de WIF tot conservatieve hypofractionering leidt met niet-exacte schattingen voor de kans op iso-effectiviteit van normale weefsels die een (quasi-)parallelle architectuur hebben en een relatief lage bestralingsdosis ontvangen. Wij presenteren een gegeneraliseerd WIF (gWIF) concept voor de *exacte* berekening van deze iso-effectiviteit, waarbij de voorgeschreven (tumor)dosis in de formule behouden blijft en enkel de conventionele α/β als maat voor de fractioneringsgevoeligheid door een *effectieve* waarde, $(\alpha/\beta)_{\text{eff}}$, wordt vervangen. Deze $(\alpha/\beta)_{\text{eff}}$ verdisconteert de niet-triviale afhankelijkheid van de dosisheterogeniteit en het volume-effect van het normale weefsel voor het desbetreffende klinische eindpunt. Met behulp van het nieuwe $(\alpha/\beta)_{\text{eff}}$ concept kan voor complexe dosisverdelingen op eenvoudige wijze worden beoordeeld of hypofractionering in potentie radiobiologisch voordelig is. Het $(\alpha/\beta)_{\text{eff}}$ concept wordt geïllustreerd aan de hand van voorbeelden waarbij gebruik gemaakt wordt van dosis-volume histogram (DVH) gegevens van het niet-geïnvolveerde longweefsel van NSCLC-patiënten die hetzij met IMRT dan wel met stereotactische ablatieve radiotherapie (SABR) werden bestraald.

Efficiënt genereren van Pareto-optimale IMRT plannen

Het ontwerp van bundelfluentieverdelingen ten behoeve van IMRT kan als *multi-objective* optimalisatieprobleem worden geformuleerd. Pareto optimalisatie is een methode om dit probleem met meerdere, veelal conflicterende doelfuncties (*objectives*) op te lossen. Dit betreft een *a posteriori* optimalisatiemethode waarbij eerst de samenhang tussen de conflicterende doelfuncties (van het bestralingsplan) wordt gekwantificeerd als een verzameling van beste compromissen, en vervolgens aan de besluitvormer de mogelijkheid wordt geboden hieruit een enkele 'optimale' oplossing te selecteren. Deze verzameling wordt door de PEF in de *objective space* afgebeeld en omvat een oneindig aantal elementen. In het algemeen bestaat er geen gesloten vorm van de PEF. Een manier om de PEF te verkrijgen is middels discrete approximatie. Hierbij worden successievelijk multipele optimalisatieproblemen opgelost. Elke oplossing is een element van de genoemde verzameling en representeert een punt op de PEF.

Om de rekenintensiviteit te reduceren is het wenselijk methoden te gebruiken die garanderen dat enkel globaal optimale Pareto-oplossingen worden gegenereerd.

In **Paper V** wordt een nieuw algoritme ontwikkeld voor iteratieve, discrete approximatie van de PEF, gebruikmakend van de zogenaamde ε-constraint optimalisatiemethode. De convexiteit van het optimalisatieprobleem wordt benut om stuksgewijs lineaire boven- en ondergrenzen te construeren teneinde de PEF, uitgaande van een kleine verzameling van initiële Pareto-optimale oplossingen, tot op een vooraf vastgestelde nauwkeurigheid te benaderen. Een nieuw '*sandwich* algoritme' wordt gepresenteerd (zie **Paper VI**) waarin deze grenzen middels drie iteratieve strategieën worden gebruikt om de lokatie van de volgende Pareto-optimale oplossing te bepalen zodanig dat de onzekerheid van de geschatte PEF maximaal wordt gereduceerd. Aangetoond wordt dat een intelligente initiële oplossing voor een nieuw Pareto-optimaal bestralingsplan kan worden verkregen door interpolatie van fluentieverdelingen van twee bestaande naburige Pareto-optimale plannen. Deze methode wordt toegepast op een gesimplificeerde klinische hoofd-hals-casus met twee doelfuncties, om de afweging te illustreren tussen dosisheterogeniteit in het doelgebied en het vermogen om normale weefsels te sparen. De resultaten tonen aan dat middels alle drie strategieën een efficiënte en representatieve approximatie van de PEF is te realiseren.

Efficiënte approximatie van de convexe Pareto efficient frontier

Voor het geval van een niet-lineair convex optimalisatieprobleem met twee doelfuncties is de PEF een univariate convexe functie. Discrete approximatie is een bekende methode om deze functie te reconstrueren. Eerder zijn *sandwich* algoritmen voorgesteld voor de univariate approximatie van convexe functies. Door deze algoritmen worden stuksgewijs lineaire boven- en ondergrenzen van de te approximeren convexe functie geconstrueerd op basis van afgeleide-informatie. Dergelijke informatie kan echter ontbreken en uitsluitend informatie met betrekking tot functie-evaluaties kan voorhanden zijn. Dit is het geval bij ε-constraint optimalisatie.

In **Paper VI** wordt het wiskundig bewijs geleverd dat een univariate PEF een dalende convexe functie is. Stuksgewijs lineaire boven- en ondergrenzen voor de univariate convexe PEF worden afgeleid, enkel berustend op informatie met betrekking tot functie-evaluaties. Volledige bewijzen voor deze grenzen worden gegeven. Tevens worden nieuwe *sandwich* algoritmen voorgesteld, welke op een systematische wijze iteratief nieuwe Pareto-optimale punten toevoegen totdat een gewenste nauwkeurigheid van de PEF-approximatie wordt bereikt. De nieuwe algoritmen zijn lineaire convergent. Indien aanvullende afgeleide-informatie beschikbaar is, dan kan onder specifieke condities kwadratische convergentie worden aangetoond. Bewezen wordt dat de ondergrenzen die resulteren uit de combinatie van functie-evaluatie en afgeleide-informatie stringenter zijn dan die welke alleen met behulp van functie-evaluatie informatie worden gegenereerd. De bruikbaarheid van deze algoritmen wordt middels numerieke voorbeelden geïllustreerd.

Praktische benadering voor afwegingen van IMRT planning bij de bestraling van het niet-kleincellig longcarcinoom

In beschikbare computersystemen voor inverse planning van IMRT verschillen de doelfuncties die het optimalisatiealgoritme sturen van de evaluatiefuncties die voor de beoordeling van de resulterende dosisverdeling worden gebruikt. Klinisch afgeleide DVH-criteria worden vaak gebruikt om de kwaliteit van een dosisverdeling te beoordelen. DVH-gebaseerde doelfuncties zijn doorgaans geïmplementeerd als kwadratische functies die de afwijking van een voorgeschreven dosis-volumeniveau kwantificeren. De correlatie tussen de waarden van de doel- en evaluatiefuncties in DVH-gebaseerde optimalisatiemodellen is zwak. Dit zou het gebruik van Pareto optimalisatie nadelig kunnen beïnvloeden, omdat de PEF de kwantitatieve relatie tussen de doelfuncties afbeeldt in plaats van de relatie tussen de evaluatiefuncties. Verder zijn de DVH-gebaseerde doelfuncties niet convex. In de praktijk is herhaaldelijke optimalisatie doorgaans noodzakelijk om acceptabele oplossingen te verkrijgen die aan de klinisch relevant geachte eisen van het bestralingsplan voldoen, hetgeen inefficiënt is.

Om dit probleem te vermijden is in **Paper VII** een software hulpmiddel beschreven dat gekoppeld werd aan het bestralingsplanningssysteem om automatisch een reeks van klinisch toepasbare IMRT plannen te genereren. Hiertoe werden de parameterwaarden van de doelfuncties op een systematische wijze gevarieerd. Tevens faciliteert dit hulpmiddel de analyse van de afwegingen tussen klinisch relevant geachte parameters, zoals de dekkingsgraad van het doelvolume en de sparing van normale weefsels. Om dergelijke afwegingen te onderzoeken wordt dit hulpmiddel retrospectief toegepast bij vijf patiënten met gevorderd stadium NSCLC die middels IMRT zijn bestraald.

Er worden lineaire relaties gevonden tussen klinisch relevante dosimetrische parameters, welke afhankelijk zijn van de specifieke casus. Dit wordt als meta-model gebruikt bij de ondersteuning voor de keuze van een individueel bestralingsplan.

Convexe herformulering van radiobiologische optimalisatie bij IMRT planning

Indien het wenselijk is radiobiologische evaluatiefuncties voor TCP en NTCP als doelfuncties en/of restricties in 'Niveau III' optimalisatie (zie pagina 17) te gebruiken, is het noodzakelijk geschikte wiskundige transformaties toe te passen om deze functies strikt convex/concaaf te maken, afhankelijk of de doelfunctie geminimaliseerd/gemaximaliseerd dient te worden. Anders is het optimalisatieprobleem moeizaam op te lossen en kan niet worden gegarandeerd dat de unieke, globaal beste oplossing wordt gevonden.

In **Paper VIII** worden wiskundige transformaties afgeleid voor een aantal bekende radiobiologische modellen die fractioneringseffecten verdisconteren. Tevens worden condities bepaald waaronder getransformeerde functies in equivalente convexe criteria resulteren zonder de verzameling van Pareto-optimale oplossingen te wijzigen. Voorts wordt aangetoond dat de toepassing van stijgende en concave transformaties op convexe doelfuncties gunstig is voor de stuksgewijs lineaire approximatie van de convexe PEF.

Het effect van transformaties op de approximatie van het Pareto efficient frontier

Zoals in de voorgaande sectie is beschreven vereist Pareto optimalisatie met radiobiologische doelfuncties en restricties dat wiskundige transformaties worden toegepast om bestaande TCP- en NTCP-evaluatiefuncties strikt convex/concaaf te maken. Vastgesteld is echter dat transformaties van verschillende kwaliteit bestaan. Bepaalde transformaties resulteren in getransformeerde doelfuncties die minder convex zijn dan andere transformaties. Dit is vooral voordelig voor de discrete approximatie van de PEF middels stuksgewijs lineaire boven- en ondergrenzen.

In **Paper IX** wordt wiskundig bewezen dat indien de doelfunctie reeds convex is, een stijgende en concave transformatie die een minder convexe functie oplevert, resulteert in boven- en ondergrenzen van de te approximeren PEF die stringenter zijn dan voor de originele functies. Voor getransformeerde radiobiologische evaluatiefuncties die reeds door anderen zijn geanalyseerd, wordt aangetoond dat zulke stijgende en concave transformaties kunnen bestaan.

Fysico-biologische optimalisatie van IMRT plannen voor geïndividualiseerde dosisescalatie bij bestraling van het niet-kleincellig longcarcinoom

De 'best' haalbare dosisverdeling hangt af van de individuele anatomie van patiënt, de onderliggende biologische heterogeniteit, de fysische beperkingen van de bestralingstechniek en de risico-afwegingen die de radiotherapeut-oncoloog en de patiënt samen maken. In de context van deze multi-factoriële ruimte moet een 'beste compromis' oplossing worden gevonden, waarvan wordt verwacht dat deze de best haalbare afweging tussen baat en bijwerkingen van de bestralingsbehandeling representeert. Derhalve is het wenselijk dat het bestralingsplanningssysteem een verzameling van Pareto-optimale bestralingsplannen kan genereren met de hoogst mogelijke kans op tumorcontrole bij een vooraf acceptabel geachte kans op complicaties.

Paper X beschrijft een retrospectieve *in silico* dosisescalatiestudie waarbij maximalisatie van de kans op tumorcontrole onder vooraf gestelde dosimetrische en/of NTCP-restricties wordt toegepast bij patiënten met een gevorderd stadium NSCLC. Om het volledige potentieel van 'Niveau III' optimalisatie voor IMRT te onderzoeken wordt een vergelijking uitgevoerd tussen homogene en heterogene dosisescalatie in het doelvolume. Voor het homogene geval wordt 'Niveau III' optimalisatie vergeleken met 'Niveau I' optimalisatie door toepassing van renormalisatie van de fractiegrootte. De resultaten suggereren dat een substantiële toename in lokale tumorcontrole bereikt kan worden over een interval van klinisch toelaatbare risico's voor alle patiënten in de studie, indien bij het dosisvoorschrift niet langer wordt vastgehouden aan stringente homogeniteitseisen voor de dosisverdeling in het doelvolume en het dosisvoorschrift niet voor alle patiënten binnen de desbetreffende groep uniform is.

Acknowledgements

Curriculum vitae

The author of this thesis was born on February 27th, 1970 in Roosendaal en Nispen, the Netherlands. He received the MSc degree in Electrical Engineering from the Eindhoven University of Technology in 1996. In the following two years, he worked as a trainee patent attorney with EP&C BV in Rijswijk and as a junior scientist in biomedical image and signal processing at the Signal Processing & Pattern Recognition Group of the Delft University of Technology. From 1998 to 2002, he was a research scientist at the Biomedical Engineering Research Centre in the Department of Urology of the Radboud University Nijmegen Medical Centre. From 2002 to 2007, he worked as project physicist at the Department of Radiation Oncology of the same institute, where he was instrumental for the clinical introduction of IMRT. Meanwhile, he did his training as medical physicist, and after completion became certified by the Dutch Society of Medical Physics (NVKF). From 2007 to 2011, he continued to work in Nijmegen as a staff member with special interests in HDR-brachytherapy, mathematical optimisation of IMRT, functional imaging for high-precision radiotherapy of prostate and head-and-neck carcinoma, and modelling of anorectal complications in prostate radiotherapy.

In October 2011, he moved to MAASTRO CLINIC, where he was appointed principal physicist brachytherapy and also became responsible for the introduction of advanced MRI techniques in radiation treatment planning. His research interests are focussed on treatment planning optimisation with (bio-)mathematical models, the development of MRI-guided radiotherapy for non-small-cell lung cancer and the exploration of ultra-high-field MRI for radiotherapy of glioblastoma. He is co-author of more than 30 peer-reviewed scientific publications, is a member of several professional societies (ESTRO, AAPM, ISMRM, NVKF) and has been a reviewer for various international scientific journals since 2009. Currently, he has advisory and supervisory responsibilities for the doctoral theses of six PhD students. Recently, he has been appointed chairman-elect of the board of the medical staff at MAASTRO CLINIC.

List of publications

Original articles

1. **A.L. Hoffmann** and A.E. Nahum. Fractionation in normal tissues: the $(\alpha/\beta)_{\text{eff}}$ concept can account for dose heterogeneity and volume effects. *Phys Med Biol*, **58**:6897–914, 2013.

2. K.M.J. van Gestel, D.J.M. Buurman, R. Pijls, P.A.W.H. Kessler, P.L.A. van den Ende, **A.L. Hoffmann**, and E.G.C. Troost. Locally advanced verrucous carcinoma of the oral cavity: treatment using customized mold HDR brachytherapy instead of hemi-maxillectomy. *Strahlenther Onkol*, 2013. [In press].

3. **A.L. Hoffmann**, H. Huizenga, and J.H.A.M. Kaanders. Employing the therapeutic operating characteristic (TOC) graph for individualised dose prescription. *Radiat Oncol*, **8**:55, 2013.

4. B.L. Gorissen, D. den Hertog, and **A.L. Hoffmann**. Mixed integer programming improves comprehensibility and plan quality in inverse optimization of prostate HDR brachytherapy. *Phys Med Biol*, **58**:1041–57, 2013.

5. A.J.S. Idema, **A.L. Hoffmann**, H.D. Boogaarts, E.G.C. Troost, P. Wesseling, A. Heerschap, W.T.A. van der Graaf, J.A. Grotenhuis, and W.J.G. Oyen. 3'-Deoxy-3'-^{18}F-fluorothymidine PET-derived proliferative volume predicts overall survival in high-grade glioma patients. *J Nucl Med*, **53**:1904–10, 2012.

6. **A.L. Hoffmann**, E.G.C. Troost, H. Huizenga, J.H.A.M. Kaanders, and J. Bussink. Individualized dose prescription for hypofractionation in advanced non-small-cell lung cancer radiotherapy: an *in silico* trial. *Int J Radiat Oncol Biol Phys*, **83**:1596–602, 2012.

7. R.J. Smeenk, W.P.M. Hopman, **A.L. Hoffmann**, E.N.J.Th. van Lin, and J.H.A.M. Kaanders. Differences in radiation dosimetry and anorectal function testing imply that anorectal symptoms may arise from different anatomic substrates. *Int J Radiat Oncol Biol Phys*, **82**:145–52, 2012.

8. R.J. Smeenk, **A.L. Hoffmann**, W.P.M. Hopman, E.N.J.Th. van Lin, and J.H.A.M. Kaanders. Dose-effect relationships for individual pelvic floor muscles and anorectal complaints after prostate radiotherapy. *Int J Radiat Oncol Biol Phys*, **83**:636–44, 2012.

9. M. Chrispijn, F. Nevens, T.J.G. Gevers, R. Vanslembrouck, M.G.H. van Oijen, W. Coudyzer, **A.L. Hoffmann**, H.M. Dekker, R.A. de Man, L. van Keimpema, and J.P.H. Drenth. The long-term outcome of patients with polycystic liver disease treated with lanreotide. *Aliment Pharmacol Ther*, **35**:266–74, 2012.

10. J.F. Langenhuijsen, E.N. van Lin, **A.L. Hoffmann**, I. Spitters-Post, J.A. Witjes, J.H. Kaanders, and P.F. Mulders. Neoadjuvant androgen deprivation for prostate volume reduction: the optimal duration in prostate cancer radiotherapy. *Urol Oncol*, **29**:52–7, 2011.

11. A.Y.D. Siem, D. den Hertog, and **A.L. Hoffmann**. A method for approximating univariate convex functions using only function value evaluations. *INFORMS J Comput*, **23**:591–604, 2011.

12. R. Monshouwer, **A.L. Hoffmann**, M. Kunze-Busch, J. Bussink, J.H.A.M. Kaanders, and H. Huizenga. A practical approach to assess clinical planning tradeoffs in the design of individualized IMRT treatment plans. *Radiother Oncol*, **97**:561–6, 2010.

13. E.G.C. Troost, J. Bussink, **A.L. Hoffmann**, O.C. Boerman, W.J.G. Oyen, and J.H.A.M. Kaanders. ^{18}F-FLT PET/CT for early response monitoring and dose escalation in oropharyngeal tumors. *J Nucl Med*, **51**:866–74, 2010.

14. L. van Keimpema, F. Nevens, R. Vanslembrouck, M.G.H. van Oijen, **A.L. Hoffmann**, H.M. Dekker, R.A. de Man, and J.P.H. Drenth. Lanreotide reduces the volume of polycystic liver: a randomized, double-blind, placebo-controlled trial. *Gastroenterology*, **137**:1661–8.e2, 2009.

15. D.A.X. Schinagl, **A.L. Hoffmann**, W.V. Vogel, J.A. van Dalen, S.M.M. Verstappen, W.J.G. Oyen, and J.H.A.M. Kaanders. Can FDG-PET assist in radiotherapy target volume definition of metastatic lymph nodes in head-and-neck cancer? *Radiother Oncol*, **91**:95–100, 2009.

16. **A.L. Hoffmann** and D. den Hertog. Optimisation of radiation treatment against cancer: Operations Research applied to Radiotherapy. *STAtOR*, **4**:4–8, 2008. [In Dutch].

17. **A.L. Hoffmann**, D. den Hertog, A.Y.D. Siem, J.H.A.M. Kaanders, and H. Huizenga. Convex reformulation of biologically-based multi-criteria intensity-modulated radiation therapy optimization including fractionation effects. *Phys Med Biol*, **53**:6345–62, 2008.

18. A.Y.D. Siem, D. den Hertog, and **A.L. Hoffmann**. The effect of transformations on the approximation of univariate (convex) functions with applications to Pareto curves. *Eur J Oper Res*, **189**:347–62, 2008.

19. D.A.X. Schinagl, W.V. Vogel, **A.L. Hoffmann**, J.A. van Dalen, W.J. Oyen, and J.H.A.M. Kaanders. Comparison of five segmentation tools for ^{18}F-fluoro-deoxy-glucose-positron emission tomography-based target volume definition in head and neck cancer. *Int J Radiat Oncol Biol Phys*, **69**:1282–9, 2007.

20. J.A. van Dalen, **A.L. Hoffmann**, V. Dicken, W.V. Vogel, B. Wiering, T.J. Ruers, N. Karssemeijer, and W.J.G. Oyen. A novel iterative method for lesion delineation and volumetric quantification with FDG PET. *Nucl Med Commun*, **28**:485–93, 2007.

21. **A.L. Hoffmann**, A.Y.D. Siem, D. den Hertog, J.H.A.M. Kaanders, and H. Huizenga. Derivative-free generation and interpolation of convex Pareto optimal IMRT plans. *Phys Med Biol*, **51**:6349–69, 2006.

22. A.Y.D. Siem, D. den Hertog, and **A.L. Hoffmann**. Multivariate convex approximation and least-norm convex data-smoothing. In M. *et al.* Gavrilova, editor, *Computational Science and Its Applications–ICCSA 2006*, volume 3982 of *Lecture Notes in Computer Science*, pages 812–21. Springer, Berlin, 2006.

23. E.N.J.T. van Lin, J.J. Fütterer, S.W.T.P.J. Heijmink, L.P. van der Vight, **A.L. Hoffmann**, P. van Kollenburg, H.J. Huisman, T.W.J. Scheenen, J.A. Witjes, J.W. Leer, et al. IMRT boost dose planning on dominant intraprostatic lesions: gold marker-based three-dimensional fusion of CT with dynamic contrast-enhanced and ^{1}H-spectroscopic MRI. *Int J Radiat Oncol Biol Phys*, **65**:291–303, 2006.

24. J.R. Spermon, **A.L. Hoffmann**, S. Horenblas, A.L.M. Verbeek, J.A. Witjes, and L.A. Kiemeney. The efficacy of different follow-up strategies in clinical stage I non-seminomatous germ cell cancer: a Markov simulation study. *Eur Urol*, **48**:258–68, 2005.

25. E.N.J.Th. van Lin, **A.L. Hoffmann**, P. van Kollenburg, J.W. Leer, and A.G. Visser. Rectal wall sparing effect of three different endorectal balloons in 3D conformal and IMRT prostate radiotherapy. *Int J Radiat Oncol Biol Phys*, **63**:565–76, 2005.

26. **A.L. Hoffmann**, M.P. Laguna, J.J.M.C.H. de la Rosette, and H. Wijkstra. Quantification of prostate shrinkage after microwave thermotherapy: a comparison of calculated cell-kill versus 3D transrectal ultrasound planimetry. *Eur Urol*, **43**:181–7, 2003.

27. **A.L. Hoffmann**, J.J. de la Rosette, and H. Wijkstra. Intraprostatic temperature monitoring during transurethral microwave thermotherapy: status and future developments. *J Endourol*, 14:637–42, 2000.
28. J.A. Kors and **A.L. Hoffmann**. Induction of decision rules that fulfil user-specified performance requirements. *Pat Recogn Let*, 18:1187–95, 1997.

Proceedings of meetings (peer reviewed)

1. H. Huizenga, **A.L. Hoffmann**, J.H.A.M. Kaanders, P.F.M. Stalmeijer, and J.J. van Tol-Geerdink. Inclusion of individual patient preferences into the radiotherapy treatment planning process. In J.P. Bissonnette, editor, *The Radiotherapy Dynamic*, volume 1 of *Proceedings of the XVth International Conference on the Use of Computers in Radiation Therapy (ICCR)*, pages 362–5, Toronto, 2007.
2. **A.L. Hoffmann**, J.J.M.C.H. de la Rosette, and H. Wijkstra. A proposed system for ultrasonic temperature imaging of the human prostate *in vivo* during transurethral microwave thermotherapy: data acquisition and initial experience. In M.F. Insana, editor, *Medical Imaging 2001: Ultrasonic Imaging and Signal Processing*, volume 4325 of *Proceedings of SPIE*, pages 75–86, 2001.

Letters to the editor

1. B.L. Gorissen and **A.L. Hoffmann**. Comparison of optimization algorithms for inverse treatment planning requires objective function evaluation. *Phys Med Biol*, 2013. [In press].
2. E.G.C. Troost, **A.L. Hoffmann**, and J. Bussink. Balancing radiation pneumonitis versus locoregional tumor control in non-small-cell lung cancer. *J Thorac Oncol*, 8:e47, 2013.
3. **A.L. Hoffmann**, J.A. van Dalen, J. Lee, V. Grégoire, W.J.G. Oyen, and J.H.A.M. Kaanders. Regarding Davis *et al.*: Assessment of ^{18}F PET signals for automatic target volume definition in radiotherapy treatment planning. *Radiother Oncol*, 83:102–3, 2007.