CHAPTER 14

Treatment Plan Optimization: An Elusive Goal

Ellen D. Yorke and Chen-Shou Chui

14.2 Elements of the “Optimal” Plan 212
14.2.1 Dose Distribution 212
14.2.2 General Treatment Methods 212
14.2.3 Practicalities 213
14.3 Information for Plan Optimization 216
14.3.1 Target 216
14.3.2 Normal Tissues 219
14.4 Plan Optimization Methods 220
14.4.1 Objective Functions 221
14.4.2 Inverse Planning Algorithms 222
14.5 Optimization Is Not Magic! 223
14.5.1 Steps in Treatment Planning: A Summary 224
14.5.2 Case Study 224
14.6 Dose Delivery Methods 226
14.6.1 Beams at Discrete Angles on Conventional Medical Linacs 226
14.6.2 Accuray CyberKnife® 227
14.6.3 Rotational Delivery Methods 228
14.7 Major Disease Sites Treated with “Optimized” Plans 229
14.8 Plan Optimization for Prostate Cancer 229
14.9 Future Directions in Treatment Plan Optimization 232
14.10 Conclusion 233
14.11 References 233
14.1 Introduction: What Is Treatment Planning?

All cancer radiation therapy—external beam, brachytherapy, unsealed radionuclide therapy—has the common goal of killing clonogenic tumor cells (those capable of replicating uncontrollably) while causing an acceptably low level of damage to normal tissues. Treatment planning is performed to create a deliverable dose distribution that provides an acceptable tumor control probability (TCP) while, simultaneously, adequately limiting the normal tissue complication probability (NTCP) for the various healthy tissues irradiated. Achieving an optimal balance of this sort is sometimes referred to as maximizing the therapeutic ratio, and at other times simply as good treatment planning.

Detailed descriptions of external beam radiotherapy and brachytherapy are found in a chapter in Advances in Medical Physics: 2006 (Court and Chin 2006) and in numerous texts and review papers (Khan 2003; Leibel and Phillips 2004; Perez et al. 2004b; Podgorsak 2005; Wolbarst et al. 2006). Treatment planning is most advanced for external photon beam radiotherapy. Planning for therapy with unsealed radionuclides, though relatively primitive (Perez et al. 2004a), is evolving (Kolbert et al. 1997).

Here we focus on megavoltage (MV) external photon beam radiotherapy. But the problems of arriving at a superior and deliverable treatment plan in a reasonable amount of time by manually or electronically solving a patient-specific, multi-criterion optimization problem to obtain both tumor eradication and normal tissue protection are common to all radiation therapy modalities.

14.2 Elements of the “Optimal” Plan

14.2.1 Dose Distribution

The ideal dose distribution would give lethal dose to each clonogenic tumor cell and zero dose everywhere else. Clinically, however, there is inevitable exposure to normal tissue from entrance, exit, and scatter radiation, and that prevents unlimited escalation of dose to the tumor. Current imaging techniques can localize the clonogenic cells only quite crudely, and there are huge error bars regarding the total dose and dose-per-fraction combination needed for high TCP. The primary objective of treatment planning is to find a good compromise between a tumor dose high enough to give a reasonable chance of local control or long-term tumor growth delay, and at the same time normal tissue doses low enough so that severe complications are unlikely. In practice, the target dose and normal tissue dose limits are prescribed by a physician, whose decisions are based on published evidence, personal and anecdotal experience, and institutional policy, and there are uncertainties inherent in all of this. On complicated cases, the oncologist often consults with medical physicists and treatment planners. So it should come as no surprise that for many radiation therapy treatments, “standard of care” encompasses a wide range of treatment doses, dose distributions, and delivery methods.

14.2.2 General Treatment Methods

Sometimes a technically simple, two-dimensional (2-D) treatment plan is adequate. For example, irradiating the whole brain with right and left lateral beams, while shielding nontarget tissues with attenuators designed from conventional simulator radiographs, is a common technique even in some curative treatments. But technically complex, more precise treatment methods are increasingly available, thanks to two major developments of the past 20 years: conventional three-dimensional conformal radiation therapy (3DCRT) and intensity-modulated radiation therapy (IMRT).

In 3DCRT, the dose distribution is shaped by standard wedges and/or customized apertures formed using attenuating blocks or a multileaf collimator (MLC) (Figures 14–1a,b,c) set at a fixed shape for each beam (Court and Chin 2006). The radiation passes unimpeded through the open areas of the field but is differentially attenuated across the slope of the wedge and strongly attenuated by the block or MLC that shapes the aperture. Only about 4% of the incident intensity is transmitted through a block, and <2% through the middle of an MLC leaf. The cross section of the irradiated volume is geometrically similar to the aperture shape. The choice of beam directions is important in 3DCRT, as adequate target dose and normal tissue sparing are accomplished, in part, by a crossfire method: All beams pass through the target, but relatively few intersect any given normal tissue. Beam directions and weights (the relative dose from each beam) are manually chosen so that, in the total dose distribution arising from all the beams overlapping, high-dose regions conform to the target and normal tissues are spared. This process is called forward planning.

In IMRT the incident radiation intensity is varied (modulated) in a customized fashion to meet planning goals. Usually the intensity modulation is designed by a computerized inverse planning algorithm and delivered by a computer-controlled MLC, but the planner must specify the treatment goals to the algorithm. For some IMRT methods, the planner must also choose the beam directions. Detailed information about IMRT is found in volumes 26, issues 1 and 2 of the journal Medical Dosimetry (2001) and in Palta and Mackie (2003) and Bortfeld et al. (2006). At present, unfortunately, inverse planning does not guarantee a satisfactory compromise between target coverage and normal tissue protection. Without art and skill, the result may be merely acceptable or, if improperly implemented, dangerously inferior.
14.2.3 Practicalities

All complex treatment planning is done with a computer; in the United States, this is usually a U.S. Food and Drug Administration (FDA)-approved commercial treatment planning system.

Before planning can start, the target and critical normal tissues must be delineated, and the treatment goals (target dose and normal tissue limits) and physician preferences (e.g., beam energy) must be prescribed. The planner’s skill and time, the capabilities of the planning system, the medical issues, the physician preferences and departmental guidelines, and the delivery hardware together determine the efficiency of the planning process and, to some extent, the quality of the treatment plan. Plan optimization for a difficult case involves trial and error, so the faster and more user friendly the software and the smoother the interactions between the physician and the treatment planning staff, the more likely it is that a satisfactory plan will arise within a day or less.

Deliverability

A treatment plan may be optimal on paper or in silico (when performed on computer), but if the dose calculations are inaccurate or the treatment is incorrectly delivered by the available hardware, the results might be catastrophic. Animal and clinical data indicate that dose differences of 5% to 10% may result in reduced tumor control or increased normal tissue complication (Brahme 1984; Dische et al. 1993; Kutcher et al. 1994).

Figure 14–2 illustrates a case in which a treatment planning system was pushed beyond its limitations. The field was small (~2*2 cm²) and highly modulated, and delivery was planned for an MLC with 0.5 cm wide leaves, but the beam model in the calculation system had been commissioned only for larger fields. The other beams in this plan showed similar discrepancies, and the plan was deemed unsuitable for clinical delivery.

What delivery accuracy is needed?

Clinical data, although noisy, indicate a sigmoidal increase of both TCP and NTCP, from zero at low dose to 100% at high dose, Figure 14–3. $D_{50}$, the dose at which the TCP or NTCP equals 50% for uniform irradiation, depends on the tumor or complication type, dose per fraction, and other factors. Small human tumors typically have $D_{50} \sim 60$ Gy when delivered in 30 fractions. If dose is normalized to $D_{50}$ and the ordinate is in percent, the slope of the curve at its midpoint divided by 100 is called $\gamma_{50}$. For human tumors, $\gamma_{50} \sim 2$, while normal tissue complications have steeper slopes, with $\gamma_{50}$ between ~2 and 6 (Brahme 1984; Bentzen 1997). This implies that, if an error causes the delivered tumor dose to be systematically 10% below calculations (e.g., planned dose is 60 Gy, delivered dose is 54 Gy), TCP is reduced from the expected 50% to approximately 30%. Similarly, if delivered doses are systematically 10% higher than planned doses, a normal tissue complication with $\gamma_{50}$ of 2.5 would occur in 75% of patients rather than the expected 50%.

Hardware requirements

The above exercise illustrates why the radiation therapy community requires tight agreement between predicted and delivered dose to phantoms under controlled conditions. Anticipating that patient doses are subject to many other uncertainties (such as tumor location, tumor shrinkage or growth, setup error, organ motion), American Association of Physicists in Medicine (AAPM) quality assurance guidelines (Kutcher et al. 1994) recommend a maximum overall dosimetric uncertainty of ±5% and geometric/mechanical uncertainty of ±5 millimeters (mm) for 2D- and 3DCRT. IMRT requires even tighter geometric/mechanical limits (Patla and Mackie 2003).

IMRT makes more stringent demands on dose calculation and delivery systems. For 3DCRT, the MLC merely defines the field edges, and a positional error of 1 to 2 mm is acceptable, especially for fields larger than 5x5 cm². But MLC-based
Figure 14–2. Discrepancy between a calculated treatment planning system (TPS) and measured (film) dose distribution for a small intensity-modulated radiation therapy (IMRT) field; doses are in cGy. Calculations and measurements are at 5 cm depth in water-equivalent phantom. The upper panel shows measured (dotted) and calculated (solid) isodose contours. The lower panel shows the difference between measurement and calculation. There are large cold spots where the measured doses are more than 20 cGy lower than calculations.

IMRT demands submillimeter leaf position accuracy (LoSasso 2003; LoSasso et al. 1998, 2001). For dynamic delivery, in which the MLC leaves move continuously while the beam is on, measurements show that less than 1 mm positional error in the gap between an opposing leaf pair can cause a 4% to 5% dose error for a typical prostate IMRT plan, with the error spread over the entire field-width traversed by the leaves.

For step-and-shoot IMRT, a 1-mm positional error can result in a dose error of ±13–17% that is spatially concentrated in the region of the leaf position (LoSasso 2003). For serial tomotherapy, submillimeter accuracy of couch motion is required (Low et al. 1998; Curran 2003). In short, accurate delivery of an IMRT plan requires robust hardware and a strict physics quality assurance program.
14. TREATMENT PLAN OPTIMIZATION: AN ELUSIVE GOAL

Dose calculation accuracy
For both 3DCRT and IMRT plans delivered on a well-maintained linear accelerator (linac), a major source of discrepancy between the doses to a well-positioned patient and calculations by a modern planning system is tissue inhomogeneity. The delivered dose is predicted most poorly near interfaces between soft tissue (density ~ 1 g/cc) and lung (density ~ 0.3 g/cc), within the first few millimeters from the surface of an air cavity, and in the buildup region below the skin surface (Papanikolaou et al. 2004). In these cases, phantom measurements are demanding, and many practitioners simply distrust computer calculations and rely on past experience and careful clinical observation of the patient. For tumors abutting air cavities or soft-tissue tumors within lung, it is common to use lower-energy beams (4 MV to 8 MV) to minimize the volume of underdosed target.

How to deal with tissue inhomogeneity corrections in treatment plans for lung cancer is controversial. Some practitioners prefer lung plans that handle the entire patient as water equivalent, arguing that past clinical experience comes from uncorrected 2-D plans. Others want the planned dose distribution to be as accurate as possible and apply inhomogeneity correction algorithms. To further complicate the matter, there are several different correction algorithms that differ considerably in accuracy, and the more precise calculations take significantly longer. Papanikolaou et al. (2004, pp. 5–8) provides a summary, with references, of the ongoing debate. The following example demonstrates that tissue inhomogeneity correction can strongly affect the quantities that are used for plan evaluation. For the 3DCRT plan shown in Figure 14–4, the chosen beams have considerable path length through low-density lung. Suppose a treatment plan is designed without consideration of tissue inhomogeneity to give 95% of the target volume at least the full prescription dose ($D_{95} = 100\%$). If the beam-on times (usually called the monitor units or MU) calculated from this plan are used to treat a patient who, of course, has low-density lungs and other tissue inhomogeneities, the differences between doses expected on the basis of calculations which do not account for inhomogeneity versus those with a simple correction method (pencil beam, corrected for radiological path length) are summarized in Table 14–1. The radiological path length–corrected calculations predict higher doses for all the tabulated indices.

Monte Carlo calculations would account most accurately for tissue inhomogeneities. Such calculations are presently unavailable to the authors, but would likely predict the lowest minimum planning target volume (PTV) dose and values for the other indices between the two tabulated values.

For IMRT and for some oddly shaped 3DCRT fields, the accuracy with which the planning system accounts for subtle treatment machine features may also be important. Different MLC and linac models have different radiological characteristics, such as MLC leaf transmission and scatter, linac head-scatter, or effective size of the extended radiation source. If these effects are inaccurately modeled, the predicted and delivered doses can differ by 5% or more over all or parts of a treatment field, as seen in Figure 14–2.

Dosimetry measurements in a phantom are often carried out for a patient’s treatment; but it is sometimes difficult to assess the significance of modest discrepancies between these, which are generally made field by field, and the results
of treatment planning calculations, in which doses from all the fields are added together.

For up-to-date and detailed summaries about issues in dose calculation and delivery accuracy for IMRT, see Palta and Mackie (2003).

14.3 Information for Plan Optimization

For 3DCRT and IMRT, the main source of patient-specific anatomical information is the planning computed tomography (CT) scan. Images are acquired at a simulation session days or weeks before treatment on a CT simulator: a CT scanner that is customized for simulation by the addition of software, a rigid, flat couch top, and alignment lasers. To facilitate reproducible positioning, customized immobilization devices are fabricated in which the patient is scanned and treated; Bentel (1996) has many photographs of such devices. The planning scan provides a geometrically accurate 3-D representation of the patient and, secondarily, the electron density distribution used for inhomogeneity corrections in the dose calculations.

14.3.1 Target

Target definition

The target is usually defined by the physician, assisted by software tools on a treatment planning system or CT-simulator computer, and this can be quite challenging (Giraud et al. 2005; Rasch et al. 2005). The gross target volume (GTV) is the site of overt disease (Figure 14–5). The physician contours it on the planning CT images, typically aided by information from diagnostic imaging studies (CT, MRI, FDG-PET), surgical reports, and the known natural history of the disease. The probable extent beyond the GTV of subclinical or microscopic disease is inferred from population-based studies and patient-specific pathology; these regions are included in the clinical target volume (CTV). When the GTV is well visualized, the CTV is often assumed to form an approximately 5-mm thick shell around it; but sometimes only the CTV is visualized. With prostate cancer, for example, the prostate is the CTV. The GTV is composed of multifocal spots within the gland, but their locations are seldom known, though there are investigational efforts to visualize them with magnetic resonance spectroscopy, PET, and ultrasound techniques.

Thoracic and abdominal tumors can move during treatments (intrafraction) and between treatments (interfraction) because of normal physiological processes such as respiration, movement of bowel gas, and rectal and bladder filling (for a review, see volume 14, issue 1 of Seminars in Radiation Oncology [2004]). And there are always interfractional setup variations caused by hardware differences, patient stresses, and therapist’s actions between simulation and treatment conditions (van Herk 2004). To prevent positional variations from causing severe tumor underdose, the CTV or GTV is expanded to form an internal target volume (ITV) and a planning target volume (PTV). The ITV encompasses the likely tumor locations in the presence of organ motion, and the PTV allows for changes in ITV due to setup error. The margins from CTV to ITV and PTV are usually based on clinical convention or patient population studies, although increasing numbers of studies indicate that individualized margins (van Herk 2004) and adaptive radiotherapy, where the treatment plan is altered based on

Table 14–1. Doses (expressed as percent of prescription) for the 3DCRT plan of Figure 14–4. The “uncorrected” doses and MU are calculated in a completely water-equivalent patient (no correction for lung inhomogeneity). The doses “corrected” for lung inhomogeneity were calculated using a pencil beam algorithm with radiological path length correction (Mohan et al. 1988).

<table>
<thead>
<tr>
<th>Target D95</th>
<th>Target max dose</th>
<th>Target min dose</th>
<th>Spinal cord max dose</th>
<th>Lung max dose</th>
<th>Lung mean dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncorrected</td>
<td>100.0</td>
<td>104.7</td>
<td>94.5</td>
<td>58.1</td>
<td>111.0</td>
</tr>
<tr>
<td>Corrected</td>
<td>127.6</td>
<td>132.4</td>
<td>114.0</td>
<td>72.5</td>
<td>142.1</td>
</tr>
</tbody>
</table>

Figure 14–5. The target volume hierarchy, including the gross target volume (GTV), clinical target volume (CTV), internal target volume (ITV), and planning target volume (PTV).
information gained during the course of treatment, would improve the therapeutic ratio. In the target hierarchy diagram of Figure 14–5, the GTV and CTV dose distributions are of interest, but the treatment planner usually focuses attention on the PTV.

It is not uncommon to have several targets with different dose goals. Achieving them is sometimes accomplished by a multiphase treatment: In the first phase, large fields treat all regions at risk, while subsequent phases cone-down on and boost the volumes of most aggressive disease. For example, in the Radiation Therapy Oncology Group (RTOG), IMRT “dose painting” protocol 0225 (RTOG 2005) for nasopharyngeal carcinoma, there are three subtargets, each of which may be made up of two or three spatially disjoint volumes: the GTV receives 70 Gy, the high-risk nodal CTV receives 59.4 Gy, and the low-risk nodal disease, 50.4 Gy. An ITV is unnecessary (no organ motion in the nasopharynx) but a 5-mm PTV surrounds each CTV.

Target dose considerations

A treatment plan is designed to deliver the prescribed dose to the bulk of the target. There is evidence that TCP for many tumors follows a dose-response qualitatively like that in Figure 14–3. There are approximate mechanistic mathematical models for TCP (Niemierko and Goitein 1993a; Webb 1994; Zaider and Minerbo 2000; Yorke 2003) but the model input parameters—the clonogen density and spatial distribution and their radiobiological parameters—are poorly understood. Estimates of clonogen density within a GTV can differ by orders of magnitude! For prostate cancer, the most common North American male cancer, there is hot debate, which may never be resolved, regarding the radiobiological parameter(s) that determine response to the dose per fraction (Fowler et al. 2001; Nahum et al. 2003). In view of these uncertainties, it is common practice to limit the doses to all the target structures so that the planned normal tissue doses are safe according to available clinical evidence. Phase I dose escalation trials (aimed at determining toxicity) increase dose gradually and terminate if complications are excessive.

Both target and normal tissue dose distributions are often evaluated with the complementary tools of graphical isodose displays and dose-volume histograms (DVHs) (Lyman 1985; Lyman and Wolbarsht 1989; Khan 2003; Purdy 2004). To generate a DVH, the range of doses spanned by the dose distribution in the region of interest is divided into dose bins and the structure itself is divided into voxels small enough so that the dose in each voxel is approximately uniform. The differential DVH is a bar graph with a bar centered at the midpoint of each dose bin. The bar height is proportional to the total volume of the voxels receiving dose covered by that dose bin. The integral or cumulative DVH is also a histogram of volume (y-axis) versus dose (x-axis), where the bar height at each dose bin is proportional to the total volume receiving at least the dose represented by the dose bin. Figures 14–6a and 14–6b are the differential and integral DVHs for a target (prostate), while Figures 14–6c and 14–6d are differential and integral DVHs for a normal tissue (the spinal cord from a lung treatment plan). Common variants of the DVH have a y-axis representing the percent or fractional volume and/or an x-axis representing percent dose, rather than absolute dose. Integral DVHs are much more commonly displayed; see, for example, Figures 14–7b, 14–7c and 14–10c in this chapter.

It would be desirable to increase dose in parts of the target where there is evidence of aggressive disease but, in the absence of patient-specific information about clonogen distribution, a uniform target dose is often assumed to be best. In this case, the target integral DVH is a step function; the dotted curve in Figure 14–7b is the DVH for an approximately uniform target dose generated by an actual IMRT plan. Simple models suggest that if the clonogen density is uniform, small high-dose regions in the target are unlikely to improve TCP but small low-dose regions (cold spots) degrade TCP (Tomé and Fowler 2000, 2002). If there are no normal tissue concerns, it is preferable to avoid cold spots. But if there is risk of unacceptable complications and it is impossible to both spare a critical normal tissue and achieve a uniform target dose, a low-dose region is imposed on the target by the overriding concern, “first do no harm.”

An extreme example is that of Figure 14–7. This patient was planned for a single fraction of stereotactic body radiation therapy (SBRT) for a paraspinal metastasis wrapping around the spinal cord (Figure 14–7a). The cord dose was required to be at most 14 Gy, a value taken in the institution as its tolerance level for SBRT. To interpret the percent-dose, absolute volume DVHs of Figures 14–7a and 14–7b, note that the prescription dose is delivered to the 100% isodose level. Although it was possible to construct an IMRT plan that gave a very uniform PTV dose (Figure 14–7b, dotted line), the maximum spinal cord dose is 3% above the prescription dose (Figure 14–7c, dotted line). Thus the prescription would have been limited to approximately 14 Gy, which is known to be insufficient. For the plan with an inhomogeneous target dose distribution (solid lines, Figures 14–7b, 14–7c), the maximum spinal cord dose was only 58% of the prescription. Furthermore, this plan allowed 12.5 cc (~78%) of the GTV to be treated to at least 24 Gy with less than 0.1 cc below 15 Gy. There is a growing body of evidence that such a dose distribution can give the patient effective palliation at the treated disease site.

A DVH is an efficient quantitative summary of the 3-D dose distribution, but it lacks geographic information. It can reveal that 1 cc of the PTV is underdosed by 20% but does not specify the location of the undertose. Isodose displays
provide detailed geographic information but one can miss seeing target underdoses or normal tissue hot spots if the display regions are limited. Also, it is impossible to assess accurately such quantities as mean dose or the fraction of target getting full prescription dose (\(V_{100}\)) from an isodose display. Useful DVH descriptors of target coverage include the maximum, mean, and minimum doses (\(D_{\text{max}}, D_{\text{mean}}, D_{\text{min}}\)); the minimum dose to the “hottest” p-percent of the target volume, \(D_p\) (\(D_{95} = 60\) Gy means that 95% of the target receives at least 60 Gy); \(V_q\) the percent target volume receiving at least \(q\%\) of prescription (\(V_{35} = 35\%\) means that 35% of the volume receives 20 Gy or more); and target uniformity indices such as the ratio of \((D_{05} - D_{95})\) to the prescription dose.

An increasingly popular index of target coverage is the tumor generalized equivalent uniform dose (gEUD or EUD). This is defined as the dose which, if delivered uniformly to the entire tumor, would give the same TCP as the actual dose distribution (Niemierko 1997, 1999; Yorke 2003; Wu et al. 2002a, 2005). There are two formulations of EUD for tumors. One explicitly uses a mechanistic dose-response model (Niemierko 1997), while the other is based on a purely phenomenological equation (Nieimerko 1999). This second version, which has been directly integrated into plan optimization algorithms, calculates gEUD as

\[
gEUD = \left( \sum v_i D_i^a \right)^{1/a},
\]

where \(v_i\) is the fraction of the structure volume in the dose bin centered on \(D_i\), and the sum is over the dose bins. For tumors, the parameter \(a\) is a large negative number (\(a \leq -10\)); the more negative it is, the more sensitive is gEUD to a cold

Figure 14–6. Two (a differential and an integral) dose-volume histograms (DVHs) for treating the target in a prostate plan, and another pair of DVHs for sparing spinal cord in a lung plan. (a) The gray line is the PTV differential DVH from the prostate IMRT plan, intending to treat the PTV with a prescription dose of 86.4 Gy. The dashed line is the ideal DVH for this case: the entire PTV volume is to be covered uniformly to the prescription dose; (b) the real and ideal integral DVH curves that correspond to the differential DVH curves for this plan. (c) Differential DVH for healthy spinal cord, from a plan to treat lung cancer. The [unattainable] ideal differential DVH for the normal tissue is a spike, corresponding to the total volume at zero dose; (d) the corresponding integral DVH.
14. TREATMENT PLAN OPTIMIZATION: AN ELUSIVE GOAL

14.3.2 Normal Tissues

Complications in healthy tissues

In attempting to avoid radiation-induced normal-tissue complications, a physician or a skilled planner must identify and segment the normal organs near the target on the planning scan, and evaluate the planned dose distributions. DVHs, graphic dose distribution displays, and NTCP models can all help in assessing normal tissue dose distributions.

Each potential complication has different symptoms, consequences, and dependences on dose distribution. Some, such as minor transient skin irritation, are “merely” annoying, and the risk of their occurrence is often not even estimated in treatment planning. Severe radiation pneumonitis or fibrosis, on the other hand, can have a great impact on quality of life; so, too, may xerostomia (drastic reduction in salivary flow) or persistent rectal bleeding. Finally, complications such as radiation myelitis (paralysis due to spinal cord injury), radiation-induced liver disease, and radiation pneumonitis can be lethal.

Treatments almost always hold the estimated risk of radiation myelitis below 1% (Schultheiss et al. 1995), but high rates of xerostomia (>50% of patients) may have to be tolerated in the interests of cure (Yeh et al. 2005). As cancer survival improves, moreover, there are more observations of significant complications that arise years after treatment (Miller et al. 2005; Raj et al. 2005; Suit et al. 2007), and future treatment planners will be pressed to reduce their incidence.

Because of all of this, it is clearly imperative that the risks and degrees of severity of the various possible complications that the patient may have to tolerate be discussed in depth by the patient and physician together, as an essential and early part of the planning process.

Incorporating complication probabilities into treatment planning

NTCP dose-response is assumed to follow a general sigmoidal curve, as in Figure 14–3. For non-uniform irradiation, the x-axis is often used to indicate a dosimetric measure such as mean dose or $D_{0.5}$. The NTCP depends not only upon the dose distribution but also on the dose fractionation and nondosimetric factors such as chemotherapy, co-morbid...
conditions, and age. The advent of 3DCRT led to a great increase in dose distribution information and a growing number of publications that document correlations between clinical complications and dose distribution. References to many of these publications can be found in reviews in *Seminars in Radiation Oncology*, volume 11, issue 3 (2001) and in Yorke (2003). Most clinical information relates to photon and electron external beam treatments with doses per fraction of 1.8 to 2.0 Gy and may not be valid for much larger or smaller prescriptions or other treatment modalities.

For some dose-limiting complications, including the feared radiation myelitis and other neurological complications (Schultheiss et al. 1995), statistically significant correlation is found between the NTCP and the highest doses to the organ, with negligible dependence on other aspects of the DVH. These complications are sometimes attributed to a serial tissue architecture (Wolbarst 1984; Withers et al. 1988) where radiation damage to a single functional subunit of the organ causes the complication. To limit these complications, the planner restricts $D_{\text{max}}$ or $D_p$, where $p$ is a small percent of the volume (1% to 5% or 1 cc); Figure 14–7 shows a case where the maximum cord dose dominated the plan. Conventionally fractionated treatment plans, at 1.8 to 2.0 Gy/fraction, are often designed to keep the maximum spinal cord dose below 45 to 50 Gy.

Other complications (xerostomia, radiation pneumonitis, radiation-induced liver disease) have a strong volume dependence and are sometimes modeled as a collection of functional subunits working in parallel so that the organ can tolerate some loss (Wolbarst et al. 1980, 1982; Withers et al. 1988). There is clinical evidence that the mean organ dose is a dosimetric predictor of these complications (Eisbruch et al. 2001; Dawson et al. 2001; Seppenwoolde et al. 2001, 2003). For example, average parotid gland doses above 26 Gy are correlated with an increased incidence of permanent xerostomia. Finally, there are complications (esophagitis, severe rectal bleeding, cardiac complications) for which volumes at high and intermediate doses and geometric features of the dose distribution correlate with NTCP (Maguire et al. 1999; Jackson 2001; Gagliardi et al. 2001).

Several NTCP mathematical models aim to relate specific complications to the entire dose distribution rather than discrete parts of the DVH (Wolbarst et al. 1980, 1982). One that is often used is the phenomenological Lyman model (Lyman 1985; Lyman and Wolbarst 1987, 1989; Kutcher et al. 1991). The Lyman NTCP is a sigmoidal increasing function of an EUD that is calculated from the DVH according to equation (14.1), with $a > 0$ (the Lyman model volume-dependence parameter $n$ is $1/a$). The higher the gEUD, the greater is the predicted NTCP though additional parameters and an additional equation are needed to calculate NTCP. For $a = 1$, the gEUD equals the mean dose; for large positive $a$ (e.g., $a = 10$), the gEUD is dominated by the highest dose part of the DVH, and would be applicable to serial model complications. Further complexity can be added by calculating gEUD from biologically equivalent DVHs, in which each dose bin is modified to account for cellular response by using a radiobiological model such as the LQ model.

There are also mechanistic models that attempt to incorporate both organ structure and intrinsic radiosensitivity using the LQ model. These include the serial model for complications that depend primarily on the high dose part of the DVH (Withers et al. 1988; Niemierko and Goitein 1991), the parallel or critical volume model for complications with a strong volume dependence (Withers et al. 1988; Yorke et al. 1993; Jackson et al. 1993; Niemierko and Goitein 1993b), and a semiphenomenological combination of these two, the serial-parallel model (Källman et al. 1992). Model parameters are chosen so the resulting NTCP predictions fit available clinical data. But biology is full of surprises, and extrapolation of models to new situations is problematic. Whether very approximate mechanistic models, EUD, or pure dose/dose-volume constraints are best for plan optimization and evaluation remains a controversial issue.

### 14.4 Plan Optimization Methods

A planner must attempt to account for all the goals of the treatment, including the prescription dose, general treatment technique, desired target dose homogeneity, constraints on normal organs at risk (OAR) and beam energy. Normal tissue constraints include maximum doses (e.g., spinal cord $D_{\text{max}} \leq 50$ Gy), dose-volume limits (rectum $V_{75} \leq 30\%$, gEUD limits (lung gEUD $\leq 20$ Gy), or model-based requirements.

In forward planning, the planner strives to reach these goals by trial-and-error choice of beam directions and weights. While inverse planning for 3DCRT would be helpful for some cases, it is rarely used because many planners are accustomed to forward planning for 3DCRT and, consequently, most planning systems only provide inverse planning for IMRT. But inverse planning is mandatory for most IMRT cases, as it is very difficult to generate patient-specific intensity modulations manually for a busy practice.

Reviews and descriptions of the current status, history, and progress of formal plan optimization methods and many references are given in Bortfeld (2003, 2006), the 2001 report of the Intensity-Modulated Radiation Therapy Collaborative Working Group (ICWG 2001), two dedicated issues of the journal *Medical Dosimetry* (volume 26, issues 1 and 2 [2001]), and in Chui and Spirou (2001), Oelfke and Bortfeld (2001), Oelfke et al. (2006), DeNeve et al. (2006a,b), and Mohan and Bortfeld (2006). We give only a brief overview of inverse planning methods below.
14.4.1 Objective Functions

Many inverse planning algorithms minimize or maximize an objective function (also called a score or cost function), which is a mathematical expression by which the planner conveys the treatment goals to the computer program. Below we will assume that the optimum treatment plan corresponds to the score function’s global minimum.

Usually the algorithm discretizes the patient anatomy, as defined on the planning scan, into voxels a few millimeters on a side, and breaks each beam into rays a few square millimeters in cross-section area. A very simple example of a dose-based objective function that is minimized by a uniform dose $D_{\text{Rx}}$ in the target (with no other considerations) is

$$F_{\text{uniform}} = \sum_i \left( D_i - D_{\text{Rx}} \right)^2,$$  \hspace{1cm} (14.2a)

$$D_i = \sum_j D_{i,j},$$  \hspace{1cm} (14.2b)

where $D_i$ is the total contribution of all the rays, $j$, to the $i$th voxel, and where $D_{i,j}$ is the dose contribution of the $j$th ray to the $i$th voxel of the target; the sums are over all rays and target voxels. $F_{\text{uniform}}$ has its minimum value of zero when the target dose is completely uniform.

A term that is often employed to prevent a dose above $D_{\text{max}}$, from occurring in either a target or a normal structure is

$$F_{\text{max,structure}} = \sum_i \left( D_i - D_{\text{max}} \right)^2 \text{Heaviside}(D_i - D_{\text{max}}),$$  \hspace{1cm} (14.3)

In equation (14.3), the sum runs over all the voxels in the structure, and $\text{Heaviside}$ is the Heaviside function, so that only voxels with total dose $D_i$ exceeding $D_{\text{max}}$ contribute to the sum. A similar form with weighting term $H(D_{\text{min}} - D_i)$ imposes a minimum dose constraint; this is used for targets, but not OARs, where low doses are desirable. Dose-volume constraints (denoted by the subscript “DV”) use a similar sum where $H$ picks up doses above the constraining value while the algorithm works to reduce dose only to enough of the points so that the constraint is satisfied (Spirou and Chui 1998). Quadratic objective functions are most common but other powers have also been implemented (Vineberg et al. 1998). Some of these methods are described in Agren et al. (1990), Wu et al. (2002a), Jones and Hoban (2002), Thieke et al. (2003), and Yang and Xing (2005).

There are at least two gEUD-based types of objective functions. In Jones and Hoban (2002), the parameter $a$ is specified for each structure of interest, as is the desired target $gEUD$, $gEUD_{\text{0,target}}$, and the maximum accepted $gEUD$ for each OAR, $gEUD_{\text{0,OAR}}$. The objective function is composed of quadratic terms, similar to equation (14.3):

$$F = w_{\text{tag}} \left( gEUD_{\text{tag}} - gEUD_{\text{0,tag}} \right)^2 H\left( gEUD_{\text{0,tag}} - gEUD_{\text{tag}} \right) + \sum_{\text{OAR}} w_{\text{OAR}} \left( gEUD_{\text{OAR}} - gEUD_{\text{0,OAR}} \right)^2 H\left( gEUD_{\text{0,OAR}} - gEUD_{\text{OAR}} \right).$$  \hspace{1cm} (14.5)

The $gEUD$ is calculated each time beamlet intensities change and the parameter $a$ must be specified for each optimization structure.

The approach of Wu et al. (2002a) is to maximize an objective function that is the logarithm of the product of logistic functions of $gEUD$. The function for each optimization structure is

$$f_{\text{structure}} = \frac{1}{n} \left[ 1 + \left( \frac{gEUD_{\text{0,structure}}}{gEUD_{\text{structure}}} \right)^n \right].$$  \hspace{1cm} (14.6)

Here $n$ is made positive if the structure is a target ($f$ is small for a plan that underdoses the target) and negative if the structure is an OAR ($f$ is small if the OAR is overdosed); different targets or OARs have different $n$’s. The basic objective function is

$$F = \prod_{\text{structure}} f_{\text{structure}},$$  \hspace{1cm} (14.7a)
but what is usually maximized is the natural logarithm of $F$, which is a sum of terms:

$$\ln F = \sum_{\text{structure}} \ln f_{\text{structure}}.$$  \hspace{1cm} (14.7b)

The $n$’s and the EUD,$n$’s are used to steer the optimization, in a manner similar to the use of the weights and dose constraints with quadratic objective functions. For EUD-based methods, it is necessary to define a virtual normal tissue that coincides with the target to prevent excessive dose in the target region, which might harm normal tissues that coexist with tumor cells.

Some early attempts at treatment plan optimization maximized the uncomplicated control objective function, $P_+$, the probability of controlling the tumor without normal tissue complications. It is defined as

$$P_+ = \text{TCP} - \prod_{\text{OAR}} \text{NTCP}_{\text{OAR}} + \delta \prod_{\text{OAR}} \text{NTCP}_{\text{OAR}} (1 - \text{TCP}),$$  \hspace{1cm} (14.8)

where $\delta$ is the fraction of patients for whom tumor control and normal tissue injury are statistically independent. If injury and control are independent, then $\delta$ is 1.0, and $P_+$ becomes TCP(1 - $\prod_{\text{OAR}} \text{NTCP}_{\text{OAR}}$). $P_+$ with modifications to account for the relative importance of different types of complications is not much used in recent optimization work, though it is still a metric of plan quality.

Other terms can be included in objective functions to address practical issues such as reducing unnecessary intensity fluctuations that can interfere with accurate delivery. Special methods are also required to prevent high intensities, which may damage skin, when target volumes are within the buildup region, and to provide “skin flash” to allow for setup error of superficial targets (Chui and Spirou 2001).

The treatment plan reached by the objective function extremum is not guaranteed to satisfy a constraint that is a clinical deal-breaker, such as the maximum cord dose. Thus it would be desirable to have optimization methods that do not leave these non-negotiable quantities to compromise, but rather impose hard constraints that limit the search space within which the algorithm is allowed to modify the intensities and evaluate the objective function for less important treatment goals (Spalding et al. 2007).

14.4.2 Inverse Planning Algorithms

Inverse planning algorithms are reviewed in the references cited at the beginning of this section. Only a brief discussion is provided here.

Many inverse planning algorithms divide each beam into rays (also called bixels or beamlets). The number of rays per beam is proportional to the cross-section area of the beam. Many implementations save time by precalculating the dose $a_{i,j}$ to each patient voxel, $i$, from each unit intensity ray, $j$, and storing the contributions to point $i$ in a vector $a_i$. The vector dimension equals the number of rays. During optimization, the beamlet intensities change. These intensities are stored in another vector, $x = \{x_j\}$, whose dimension also equals the number of rays. The dose to the $i^{th}$ voxel for a general intensity distribution is the dot product

$$D_i = x \cdot a_i.$$  \hspace{1cm} (14.9)

By systematically varying the individual beamlet intensities (also called ray weights), the algorithm seeks the minimum of $F$. Negative intensities are nonphysical and are not permitted.

Optimization algorithms may be deterministic or stochastic. Most deterministic techniques use numerical methods, such as the conjugate gradient method, to navigate downhill toward an objective function minimum. These methods are not guaranteed to find the global minimum, though this may not cause a clinically inferior plan (Wu and Mohan 2002). They are fast (CPU time for the Memorial Sloan-Kettering Cancer Center (MSKCC) planning system running on a Pentium-4, 3.06 GHz PC with 2 GB of RAM is between 0.07 sec and 1 sec per iteration, depending on complexity of the case) and are implemented on many commercial planning systems.

The most frequently used stochastic method is simulated annealing (Webb 1992). As the name implies, it derives from statistical mechanics simulations of the approach of a material system to thermal equilibrium. Historically, this was among the first optimization techniques used in radiation therapy treatment planning and it is implemented on some commercial systems. At each iteration, random changes are made in the ray weights and the objective function is evaluated. If it decreases, the change is accepted. If it increases, the change is accepted with a probability of $\exp(-\Delta F/kT)$, where $\Delta F$ is the change in score function, $k$ is Boltzmann’s constant, and $T$ is a “temperature” parameter that can be adjusted to improve the algorithm efficiency. In general, simulated annealing is much slower than the deterministic methods. Since the process accepts some changes that do not decrease the score function, in theory it avoids getting trapped in a local minimum; in practice, one can never be certain that the solution is truly the global minimum.

Optimization with linear programming has been attempted but is not widely used (Rosen et al. 1991). A recent study describes the use of mixed integer programming to optimize both beam directions and intensities (Lee et al. 2006).

Aperture-based optimization is a quite different approach that is implemented on some commercial planning systems. This method does not dissect the beams into beamlets, but designs a number of discrete apertures in beam’s-eye view
For example, if, in a BEV projection, a single OAR intersects the PTV, the planner or the algorithm defines one aperture that includes the entire PTV and one or more others that include only the PTV minus the OAR. The optimization algorithm can choose the aperture weights and may also fine-tune each aperture shape to minimize the objective function. A variant of this method for breast treatment, where the entire breast is the CTV, uses the traditional breast tangent beam arrangement shown in Figure 14–8a. The user selects a few (≤5) apertures based on the dose distribution calculated for unmodulated fields (Figure 14–8b), and either the planner or an algorithm chooses weights to improve dose uniformity (Kestin et al. 2000). Since there are many fewer apertures than beamlets (≤5 apertures versus hundreds of beamlets per beam), these intensity distributions are more intuitively understood when viewed against the BEV projection of the patient’s anatomy. The weighting algorithms run quickly.

### 14.5 Optimization Is Not Magic!

Anyone doing treatment plan optimization soon realizes that there are practical limitations. For one thing, the absolute minimum of $F$ should occur mathematically with the dose distribution satisfying all the imposed constraints. It is easy, however, to choose sets of constraints that are physically impossible to satisfy simultaneously. A trivial example occurs if the planner requests a uniform target dose to a PTV with an embedded serial-type normal tissue, and the desired dose is higher than the maximum allowed normal tissue dose. But such situations arise also when the normal tissue is outside of, but too close to (≤5 to 8 mm), the PTV (Hunt et al. 2002). In these cases, an optimization algorithm may approach a minimum in the stipulated number of iterations but, depending on specifics of the score function, the resulting plan may be unacceptable. The planner is left to decide what changes in the plan, or perhaps in the objective function, might produce better results.

Also, an optimization algorithm can, at best, only account for treatment goals that are specified in the objective function and its inputs. The resulting dose distributions must be carefully checked to be sure there are no hot spots in a “nonspecific” normal tissue or cold spots in an inadequately defined target.

To save time, some algorithms restrict the precalculated doses (the vectors $a_i$) to beamlets that pass near the $i^{th}$ voxel, and ignore small but non-zero contributions from more distant rays. And most optimization algorithms ignore realistic dosimetric effects of the delivery hardware (e.g., radiation transmission through MLC leaves). The common clinical practice of doing physical dosimetry in a phantom prior to a complex patient treatment sometimes reveals a delivered dose distribution that is quite different from that predicted by the optimization algorithm (Figure 14–2). The planner and the physician then face a difficult judgment on what to do next.

Finally, as will be seen in an example below, the choice of beam angle and number of beams can be important for nonrotational forms of IMRT. Optimizing beam angles can

---

**Figure 14-8.** Traditional and field-within-field treatments of the breast. [a] The breast tangent beam arrangement applied to a left breast, shown in the transverse CT slice through isocenter. The central axis and field edges are shown for each beam. The central axes meet at isocenter (the “×”). They are slightly angled relative to each other so that their posterior edges form a plane. The anterior field edges extend a few centimeters beyond the breast surface, providing a “flash” region so that breathing motion or setup error cannot easily move breast tissue outside the fields. [b] Fields from the aperture optimization (or field-within-field) technique applied to breast tangents. For each beam direction, there are three treatments, delivered in sequence: there are three apertures for each beam, and each cones down to boost the volume covered by its predecessor. The top panels illustrate the three beams delivered from one direction, the bottom panels from the other. The posterior borders of the two largest apertures (leftmost panel in each row) are defined by the MLC leaves [jagged line], while the other borders are defined by the linac jaws [solid lines]. Additional dose is subsequently delivered to the posterior chest wall by the other mostly MLC-defined apertures, to compensate for increased attenuation in this thicker region.
be a very difficult problem (Lee et al. 2006), and most optimization procedures leave this job to the planner’s expertise and intuition.

These features conspire to blur the distinction between forward and inverse planning. The planner tweaks maximum or minimum doses, dose-volume combinations, penalties, or $g_{EUD}$'s rather than wedges and beam weights, but it is tweaking all the same.

14.5.1 Steps in Treatment Planning: A Summary

The physician and physicist/planner define and contour all structures of interest. The treatment goals are defined, either by a protocol or in a patient-specific fashion. The treatment machine/beam energy is assigned; this is often done by therapy staff, not the planner. But hereafter, the choices are made primarily by the treatment planner:

- Decide whether to use IMRT or 3DCRT.
- For treatment with discrete linac gantry angle methods (3DCRT; step-and-shoot and dynamic-MLC IMRT), choose the number and directions of beams. This is unnecessary for tomotherapy.
- For 3DCRT, design the blocking, choose candidate standard wedges.
- For inverse-planned IMRT or 3DCRT, input the treatment goals according to the rules of the optimization algorithm and run the algorithm. If the goals are inadequately specified, a poor plan may result.
- Calculate the doses.
- For forward planning, choose beam weights. For all plan types, normalize the plan so that the correct dose is delivered to the prescription point or prescription isodose line, following department or protocol policy.
- For both 3DCRT and IMRT, the planner and the physician evaluate the plan by examining graphic isodose distributions, DVHs, and other figures of merit (TCP, NTCP, EUD). If the dose distribution is unacceptable or if either one feels it can be substantially improved, the planner recycles through the above steps, usually making smaller changes first (e.g., beam weights for 3DCRT) until either the intended goals are met or they are changed to acknowledge what is reasonably achievable.

For example, a physician may ask to treat a tumor uniformly to a dose of 60 Gy, and require that a nearby critical organ get a maximum dose of 45 Gy. The planner may be unable to satisfy both requirements for a prescription dose above 50 Gy, and that may be true also for other, more experienced colleagues. In such a situation, it is necessary to make some compromises: The physician may lower the prescribed dose; she may accept the inhomogeneous target dose, as in Figure 14–7b; she may define a cone-down sub-volume of the target which can be treated to 60 Gy without jeopardizing the normal tissue, and then deliver a lower dose to the rest of the target; or she may allow the normal tissue limits to be exceeded, arguing that the cost of not controlling the tumor justifies the risk.

The bottom line is that figures of merit, as well as the planner’s tools, skill, persistence, and available time and the physician’s critical input all determine the “optimal” plan for a given case. There is no universal standard.

Figure 14–9 summarizes the planning process.

14.5.2 Case Study

Suppose that a physician requests a plan to treat a medium-sized, non–small-cell lung cancer (GTV volume 100 cc) to a dose of 60 Gy at 2 Gy per fraction with 6 MV photons. This exercise, based on a real clinical case, demonstrates how the plan was developed on the in-house planning system at MSKCC, which is described in Mohan et al. (1988) and in Spirou and Chui (1998).

The GTV to PTV margin is 1 cm in all directions; the PTV volume is 394 cc. On the planning scan, approximately half of the PTV is occupied by normal lung, but the physician wishes to give the entire PTV the full prescription dose because tumor cells might be anywhere within it.

The achievable prescription dose is limited by the spinal cord and lungs. The maximum spinal cord dose is held below 50 Gy to avoid even a slight risk of radiation myelitis. To keep the estimated risk of severe radiation pneumonitis low as well, the lung NTCP calculated according to the Lyman-Kutcher-Burman (LKB) model is maintained at $\leq 25\%$ (Lyman 1985; Burman et al. 1991; Kutcher et al. 1991), implying need for a mean lung dose below 20 to 21 Gy. The planner also checks for excessive “hot spots” within normal tissue, though this is less of an issue for the relatively low prescription dose. If normal tissue constraints cannot be achieved, the physician will reduce the prescription.

Unfortunately, this patient has small lungs (volume ~1490 cc) and the central tumor location puts the spinal cord at risk, as is apparent in the coronal view of Figure 14–10a. Nonetheless, much of the PTV is in the mediastinum rather than the lung itself, and only ~9% of lung is actually within the PTV, so the theoretical ideal dose distribution (60 Gy to PTV, 0 Gy elsewhere) would easily meet the NTCP criteria, with a mean lung dose of 5.4 Gy. The problems come from entrance and exit doses—the physics of photon beams.

In the 2-D era, the “parallel opposed AP-PA” beam arrangement was employed frequently (Figure 14–10b). The beams are incident from directly above and below, and each
beam gives equal dose to the isocenter, which is at mid-depth. But for a prescription dose of 46 Gy, this arrangement gives a high maximum spinal cord dose of 50 Gy, while the PTV $D_{95}$ is 45 Gy. The target dose is quite inhomogeneous, with relatively high dose in nonspecific tissues such as the soft tissue of the chest wall (Figures 14–10b and 14–10c). Lung NTCP is 24%.

Two different 3DCRT plans were tried, the 3-field and 4-field beam arrangements of Figures 14–11a and 14–11b. The beams are shaped with a static MLC to conform to the PTV in the BEV (Figure 14–11c), and the dose distribution is further refined with standard wedges. The planner manually adjusted beam weights. But the highest PTV $D_{95}$ allowed by the lung NTCP limit is 45.8 Gy, for the 3-field plan.

Two different IMRT plans were also generated using an algorithm that minimizes a score function similar to equation (14.4) to impose dose and dose-volume constraints (to suppress hot spots) on the PTV, cord, lungs, and dummy structures. For a four-beam IMRT plan using the same beam directions as the four-beam 3DCRT plan, lung NTCP limited the PTV $D_{95}$ to 53.8 Gy (Figures 14–12a and 14–12c). For a nine-equispaced–beam plan with beam directions following a proposed “class solution” for IMRT lung plans (Liu et al. 2004), the lung NTCP restricted the PTV $D_{95}$ to $\leq$59 Gy (Figures 14–12b and 14–12c). With more beams there are more degrees of freedom for satisfying constraints, which is one rationale for tomotherapy and arc-based IMRT. On the other hand, depending on the delivery method, the time for delivering nine beams rather than two to four beams may be a practical problem.

The lung DVHs are compared in Figure 14–12d. For this case, the 4-field 3DCRT plan and both IMRT plans exposed more normal lung to doses in the 10 to 20 Gy range. There is concern that this might have undesirable clinical consequences (Seppenwoolde et al. 2003; Allen et al. 2006). However, the IMRT plans had more “conformal” PTV dose distributions (integral DVHs more like step functions) than the 3DCRT plans, Figure 14–12c, contradicting an urban legend that IMRT plans inherently give less uniform target coverage than 3DCRT. Cord doses were below 50 Gy in all cases.

In the end, normal tissue constraints prevented any plan from reaching the goal of 60 Gy to the target, although the IMRT plans came closer than the others. This patient might
have benefited from a patient-specific margin design method, if that would reduce the PTV (van Herk 2004). But could other planners/optimization algorithms/methods have done better? We don’t know.

14.6 Dose Delivery Methods

The dose delivery system influences the ability to optimize and to deliver a treatment plan. Concise descriptions and further references on dose delivery systems for 3DCRT are found in Court and Chin (2006) and Khan (2003). There are many detailed descriptions of IMRT delivery systems, among them the AAPM 2003 Summer School Proceedings (Palta and Mackie 2003); Medical Dosimetry, volume 26, issues 1 and 2 (2001); Webb (2006), and Mackie (2006). Only a brief description of the major IMRT methods, in relation to their influence on treatment plan optimization, is given here.

14.6.1 Beams at Discrete Angles on Conventional Medical Linacs

Computer-controlled MLCs on medical linacs were first designed as convenient substitutes for customized shielding blocks. The motion of each strongly attenuating leaf is under individual computer control. Thus an MLC can define apertures of almost any desired shape, eliminating most of the need for manufacturing, carrying, and securing heavy attenuating blocks. It was soon appreciated that the computer-controlled leaves could expose different portions of the field for different amounts of time (different numbers of monitor units), thus producing the intensity variations needed for IMRT. For example, to create the intensity valley that runs down the middle of the field in Figure 14–13, fewer MU are delivered when the valley is exposed than when the hills are.

For DMLC, at each selected gantry angle the leaves sweep unidirectionally and continuously, with varying speeds, from one side of the field to the other to form the intensity pattern generated by the optimization algorithm, while the MU are delivered at a constant rate (Spirou and Chui 1994). For example, to form the low-intensity valley in Figure 14–13, the leaves, and the open area between them, move quickly in that region so that radiation is transmitted for only a short time.

For SMLC, the intensity pattern at each beam direction is built up through a set of discrete subapertures. The beam is

**Figure 14–10.** Treatment of a medium-sized, non–small-cell lung cancer (GTV = 100 cc) with 6 MV photons in a patient with relatively small lungs. The clinical goal was to treat the PTV to 60 Gy. None of the plans reached this goal. (a) Coronal slice through isocenter. The small x’s outline the PTV, the smaller dots outline the GTV. (b) The transverse plane through isocenter, with the dose distribution for parallel-opposed beams, the central axes of which are indicated. One beam is incident vertically from above the patient, and the other vertically from below. The inner thick-lined structure enclosing the isocenter (at the x) is the GTV, the outer one the PTV. The thinner contours are isodose levels. The highest dose regions are within the 5500 cGy line surrounding the beam axes, and two small high-dose regions near the right anterior and posterior beam entrances. The high-dose contours do not conform well to the PTV. (c) Integral absolute dose (cGy), % volume DVHs for PTV (solid curve), spinal cord (small dashes), and lung (gray) for the plan of Figure 14–10b.
on (“shoots”) for a set number of MU while the leaves define one static subaperture. The beam is then off as the leaves “step” to the next shape, where additional MU are delivered, and so on (Siochi 1999). For both DMLC and SMLC, the design of the leaf motion sequence is an intimate part of the planning process, though often it is performed in the background by the treatment planning computer and is not under the planner’s conscious control. DMLC requires less delivery time but more MU than SMLC, but it places greater mechanical and quality assurance demands on the MLC. For further discussions of the relative merits of these two techniques, see Chui et al. (2001).

While custom metal compensators can be fabricated to produce intensity-modulated fields, this “low-tech” method is less convenient than MLC-based methods, and it is seldom used.

14.6.2 Accuray CyberKnife®

First developed for intracranial radiosurgery treatment, the CyberKnife® (Accuray Inc., Sunnyvale, CA) is a 6 MV linear accelerator mounted on a robotic arm. The linac is provided with collimators to deliver small circular fields of different sizes, which approximate the “beamlets” that are theoretical constructs in some optimization algorithms. A much wider range of beam directions is available to this system than to a conventional linac, which must avoid collisions between the bulky treatment head and couch. The planning system determines the intensity of each circular beamlet to optimize the cumulative dose distribution. Image guidance, via computer-processed images from ceiling-mounted x-ray units and floor-mounted detectors, is an important feature of this system (Adler et al. 1999).
14.6.3 Rotational Delivery Methods

Early in the history of IMRT, it was thought that using many gantry angles was highly advantageous, especially for centrally located and concave target volumes. Increasing the number of beam directions does provide more degrees of freedom, but several studies agree that there is only marginal advantage to plans with more than seven to nine beams (Bortfeld 2003; Söderström and Brahme 1995). And in practice, increased delivery time becomes a disadvantage. But time-efficient rotational delivery methods, which eliminate or reduce the need to select beam directions, have been developed. Although it is difficult to make fair comparisons between treatment plans developed on different planning systems, by different individuals, using different techniques, published treatment plans produced by rotational methods are at least comparable to those produced by fixed-beam methods.

Two forms of rotational IMRT—serial and helical tomotherapy—use a narrowly collimated, intensity-modulated fan beam that rotates around the patient who, during the treatment, moves relative to the plane of rotation. Each approach is implemented within a package that includes the computer programs for treatment planning and MLC leaf sequencing, and for control of the radiation source rotation, MLC leaf motion and couch translation. For delivery with
14. TREATMENT PLAN OPTIMIZATION: AN ELUSIVE GOAL

acquire MV helical CT images of the patient in treatment position, giving a major impetus to the now-burgeoning field of image-guided radiation therapy. The system is thoroughly described and many further references are given in Mackie et al. (2003) and Mackie (2006).

Intensity-modulated arc therapy (IMAT) is a third form of rotational IMRT (Yu 1995; Shepard et al. 2003). It is delivered by a conventional linac, capable of running in arc mode and equipped with a dynamic MLC—both are routine features of modern linacs. In the current implementation, the planner must choose the number of arcs and their angular ranges, which is akin to picking beam direction.

14.7 Major Disease Sites Treated with ‘Optimized’ Plans

Both 3DCRT and IMRT are used for radiation therapy of almost every type of cancer. Plan optimization is now often used for many common disease sites including breast, prostate, lung, and head and neck cancers. General information about each of these disease sites from a radiation oncology perspective is found in medical textbooks such as Perez et al. (2004b) and Leibel and Phillips (2004). The National Cancer Institute (NCI) website is another great source of information; relevant URLs are:

- http://www.cancer.gov/cancertopics/types/lung; and

Recent reports in the literature attribute clinical advantages at several major disease sites to the use of plan optimization techniques, especially in combination with IMRT. These studies find that treatments according to optimized plans do not compromise local control and also lead to reduced normal tissue toxicity and improved overall quality of life. For breast treatment, see Freedman et al. (2006), and Donovan et al. (2007); for nasopharyngeal cancers, see Lin et al. (2003), Wei and Sham (2005), Fang et al. (2007), and Lee et al. (2007); and for general head-and-neck cancer, see Graff et al. (2007).

14.8 Plan Optimization for Prostate Cancer

Prostate cancer is the most common male cancer in the United States, excluding non-melanoma skin cancer. Many patients present with disease that is localized to the prostate gland and are in otherwise healthy middle or old age. The presence of prostate cancer cells can be detected by a blood test for prostate-specific antigen (PSA). After treatment, PSA-relapse (biochemical relapse) can precede clinical
symptoms by years. Until recently, prostatectomy was major surgery with high morbidity, making a non-invasive curative treatment very desirable.

Since prostate cancer often has multiple small foci dispersed throughout the gland, the entire prostate (~50–60 cc) is the CTV. The usual prescription dose per fraction is 1.8 to 2.0 Gy for 30 to 48 treatments. Variable bladder and bowel contents can lead to unpredictable daily prostate positional changes exceeding 0.5 cm (Pollack and Price 2003; Buuyounouski et al. 2006). Usually the CTV is expanded by ~1 cm to account for setup error and organ motion, resulting in an approximately 100 to 150 cc PTV. The small, centrally located target is nestled among radiosensitive critical normal tissues, most important of which are the rectum and bladder (Figure 14–14). The femoral heads, the urethra (passing through the middle of the gland), and sometimes bowel are also of concern. Until the mid-1990s, radiation-induced rectal complications limited external beam treatment doses to 65 to 70 Gy. PSA-relapse-free survival at 5 years ranged from ~80% for the lowest risk patients to ~30% for the most locally advanced patients treated with radiation alone (Shipley et al. 1999; Zagars et al. 1999). As institutions attempted 3DCRT treatment to higher doses, it became clear that exposing the full length of rectum abutting the PTV to doses above 74 to 75 Gy produced unacceptable levels of severe rectal bleeding (Jackson 2001).

The central location of the prostate lends itself well to beam directions approximately equispaced over a full 360 degrees and to rotational treatments. Common field arrangements for prostate 3DCRT are the “4-field box” and a 6-field arrangement (Figures 14–15a and 14–15b). Each aperture conforms to the shape of the PTV in the BEV, with an extra ~5–10 mm margin to the aperture edge to account for beam penumbra. Common IMRT beam arrangements use five to nine symmetrically placed fields (Figure 14–15c).

With 3DCRT, PTV coverage is excellent; but even trials with prescription doses of approximately 76 Gy with rectal blocking for the last 5 fractions resulted in a 12% rate of serious rectal bleeding, and bleeding rates with higher prescriptions were unacceptably high. With IMRT, prescription dose could be escalated first to 81 Gy and recently to 86.4 Gy with minimal (<5%) severe rectal bleeding and urinary toxicity (Zelefsky et al. 2006). To achieve this result, the score function and the evaluation process include strict dose-volume constraints on the rectal wall. For the plan shown as an example, the constraints were based on analysis of radiation-induced rectal bleeding (Jackson 2001; Jackson et al. 2001) and require that the rectal wall $V_{75.6\text{ Gy}} \leq 30\%$ and $V_{47\text{ Gy}} \leq 53\%$.

PSA-relapse-free survival improves as the prescription dose is increased. For example, Zelefsky et al. (2006) report that PSA-relapse-free survival 8 years after IMRT treatment to 81 Gy is 85% to 90% for low-risk, 76% to 79% for intermediate-risk patients, and 61% to 69% for those at high risk; the variation in survival depends on how the risk group is defined. A 5-field beam arrangement, similar to Figure 14–15c and delivered with the DMLC technique, was used for these patients. Figures 14–15a, 14–15b, and 14–15c show the relative isodose distributions in the transverse plane through isocenter for 4-field and 6-field 3DCRT plans and a 5-field IMRT plan delivered to a prone patient. The IMRT plan was optimized to deliver a total treatment of 86.4 Gy (48 fractions). The 100% line represents prescription dose. Note how the 100% and 90% isodose lines are concave around the rectum in the IMRT plan while, for the 4-field and 6-field plans, these isodose lines encompass about half the rectal wall. Figure 14–13 shows the intensity distribution generated by the optimization algorithm for the posterior field of Figure 14–15c. The low-intensity region at the center helps protect the rectum. Figures 14–16a and 14–16b show the relative PTV and rectal wall DVHs for this patient’s 6-field 3DCRT and IMRT plans. The IMRT plan has less homogeneous target coverage ($D_{\text{min}} \sim 70\%$, $D_{\text{max}} \sim 110\%$, $V_{100\%} \sim 84\%$) than the 3DCRT plan ($D_{\text{min}} \sim 94\%$, $D_{\text{max}} \sim 106\%$, $V_{100\%} \sim 96\%$), but the 3DCRT plan distributes the dose in the rectal wall very differently. The dose inhomogeneity of the IMRT plan is caused by the dose-volume constraints that protect the rectal wall. Had the IMRT plan been optimized to deliver a lower prescription dose (e.g., 70 Gy).
14. TREATMENT PLAN OPTIMIZATION: AN ELUSIVE GOAL

Treatment planners and radiation oncologists are often confronted by ambiguous alternatives, as in the following example. Figure 14–17 shows the absolute dose-% volume histograms for PTV if the IMRT plan prescription dose is 86.4 Gy and the 6-field prescription is 77.4 Gy. At that dose, the 3DCRT plan would lead to quite high rectal NTCP. The IMRT DVH (solid line) lies slightly below that of the 3DCRT (dotted line) for doses between approximately 72 Gy and 77 Gy. This means that, despite the higher prescription dose, a small part of the PTV (approximately 2%) gets a lower dose from the IMRT plan than it would from the 3DCRT plan. On the other hand, with 3DCRT, the mean PTV dose is 79.9 Gy and \( D_{95} \) is 74.4 Gy while with IMRT, the mean PTV dose is 87.6 Gy and \( D_{95} \) is 81.2 Gy. From equation (14.1), the IMRT plan has higher gEUD if the parameter \( a > -23 \); for more negative values, the effect of the cold spot dominates the nominal increase in prescription dose. Introducing the LQ model would add at least three very uncertain radiobiological parameters. As an additional confounding factor, the DVH is for the PTV, not the tumor foci, which are at unknown locations inside the gland, invisible to practical, non-invasive imaging techniques at this time. And despite immobilizing devices to maintain a reproducible position of the pelvis, daily position changes of the prostate confound the question of whether the PTV DVH is the correct one to analyze (Levegrün et al. 2001).

Which treatment is closer to the optimal: the IMRT with a prescription of 86.4 Gy or the 3DCRT with a prescription of 78 Gy? The answer is not clear to the practical treatment planner. Perhaps in the future, clinical research will give us a better handle on prostate tumor response. There is strong evidence that the IMRT rectal wall dose distribution greatly reduces the incidence of very unpleasant complications, and it has been found (Pollack and Price 2003; Zelefsky et al. 2006) that escalating prescription dose from approximately...
70 Gy to approximately 80 Gy improves PSA-relapse-free survival, especially in intermediate-risk patients. Prostate cancer recurrence manifests over long time-scales (>5 years), so benefits of escalation to 86.4 Gy are not yet in the literature. Dose escalation above a prescription of 86.4 Gy has not been attempted due to further deterioration of target coverage. Rather, techniques such as hypofractionation (dose per fraction > 1.8 to 2.0 Gy), image guidance to reduce motion effects, and multimodality treatments such as combination of external beam and brachytherapy (Zaider et al. 2007) are being explored to further improve outcomes.

**14.9 Future Directions in Treatment Plan Optimization**

Despite great progress in designing and delivering treatment plans, there is room for improvements. Some areas which need further development are:

*Target definition and localization*
Treatment planners can do a pretty good job of hitting a target known in *silico*, but the true target location is often uncertain, even after examination with MRI, PET/CT fusion, etc. Biological and other new imaging modalities may prove capable of finding cancers, perhaps picking up on subtle cellular differences. At the same time, both radiographic (image guidance) and nonradiographic target localization at treatment will bring the delivered dose closer to where it needs to go.

*True medical goals*
Normal tissue constraints derived from 2-D-era studies (Emami et al. 1991) may needlessly limit planning options. Systematic efforts are needed to generate and to maintain an up-to-date set of consensus recommendations based on outcomes studies of the 3DCRT and IMRT eras. Similar considerations apply to understanding the benefits of target dose homogeneity, the effects of dose-per-fraction variations for both targets and normal tissues, and other such issues.

*Physics considerations*
Accounting accurately for tissue inhomogeneity and subtle radiological aspects of the delivery hardware in the planning process will reduce or eliminate the necessity for laborious patient-specific pretreatment measurements. And the effect of dose calculation and organ motion on the DVHs under analysis should be understood at least approximately.
Optimizing the modality

Although we have focused on plan optimization for external photon beam treatments, similar conceptual processes apply to proton and heavy ion therapy and to brachytherapy. Today, modality optimization is commonly limited to combinations of photon energies and electron fields. Future combinations may extend to brachytherapy (Zaider et al. 2007), protons, and modalities other than ionizing radiation, such as photodynamic therapy and systemic therapies.

Optimizing the optimization

Present-day plan optimization can be very time-consuming. Often the process terminates when the planner runs out of time. In part because current score functions cannot fully convey the treatment goals to current optimization algorithms, plan optimization involves much more art and manual adjustment than it might.

Bortfeld and colleagues (Bortfeld et al. 2003, 2006) propose finding Pareto optimal solutions where one criterion (e.g., target dose uniformity) cannot be improved without producing a worse result in another (e.g., a normal tissue limit). Once a computer generates the Pareto front, the planner can navigate through the solutions to seek the choice that best suits the treatment goals. Other suggestions for more efficient, better optimization have been proposed by Lee et al. (2006), Spalding et al. (2007), and Lu et al. (2007).

14.10 Conclusion

Treatment plan optimization is a complex and still imperfect process. It requires an understanding of the medical treatment goals, efficient and accurate algorithms, and knowledgeable treatment planners. To move a plan from the virtual world to reality also requires accurate dose calculations, a range of advanced technology delivery systems, and the utmost human care in carrying out the treatments. Early news about clinical outcomes for patients treated with “optimized” plans during the last decade encourages physicians and the medical physics community to continue their efforts.

14.11 References


