apy boost (Grigsby et al. 2001a). Figure 11.9 shows a color-wash dose distribution for delivery of 60 Gy to a PET-positive lymph node and 50 Gy to the PALN bed. The red-colored region corresponds to 60-Gy dose and the blue area represents doses between 50 and 60 Gy. Esthappan et al. (2004) described in detail IMRT treatment-planning concerns for this delivery technique.

Levivier et al. (2000) investigated potential advantages of PET-guided gamma knife stereotactic radiosurgery. They were able to acquire acceptable stereotactic image registration accuracy with PET images; however, they concluded that areas of increased FDG uptake were located predominantly within the MRI-defined tumor volumes and that PET provided little additional information. Gross et al. (1998) and Nuutinen et al. (2000) also concluded that PET contributes relatively small amount of additional information to MRI data for treatment of brain tumors. From these manuscripts it appears that PET may provide useful information regarding boost volumes and it may allow differentiation of residual tumor better than MRI.

Matic et al. (2002) and Malyapa et al. (2002) described FDG-PET based 3D brachytherapy treatment planning for treatment of cervical cancer. The proposed treatment technique was based on a postulate that FDG-PET images may allow a more accurate delineation of target volumes for brachytherapy gynecologic (GYN) implants. Both studies evaluated the feasibility of using PET as the sole source of target, normal structure, and applicator delineation for intracavitary GYN implant treatment planning. Standard Fletcher-Suit brachytherapy tandem and colpostat applicators which were used for radiation delivery can cause artifacts on CT and MR imaging and these imaging modalities require specialized applicators. In the proposed technique, after insertion of the applicators in the operating room, the patients were taken to a PET scanner where 555 MBq (15 mCi) of $^{18}$F-FDG was intravenously administered. Forty-five minutes later, three localization tubes containing $^{18}$F-FDG were inserted into the source afterloading compartments of the tandem and colpostat. A whole-pelvis scan was performed and the images were transferred to a commercial brachytherapy 3DTP system. A Foley catheter was also inserted into the urinary bladder while the patient was in the operating room. The regions of radioactivity in the three applicator tubes on PET images were used to identify applicator locations. Tumor volume, bladder, and rectum were also contoured on PET images. Figure 11.10 shows a PET image with three applicators, bladder, and rectum. From the contours, a 3D geometry of the implant, tumor volumes, and critical structures was generated. Created treatment plans included dose-volume histograms (DVH) and 3D dose distribution displays.

Malyapa et al. (2002) compared PET-based GYN brachytherapy plans with conventional 2D treatment plans for 11 patients. They concluded that FDG-PET provides reliable estimate of the cervical cancer volume and 3D spatial relationship of the tumor to the tandem and ovoid applicators. There were no significant differences in the calculated doses at the ICRU-38 defined point A, bladder, and rectum points between the conventional and proposed PET-based treatment-planning techniques. The maximum bladder and rectal doses determined from the FDG-PET based dose-volume histograms were found to be higher than those obtained using conventional treatment planning. The PET-based treatment planning revealed that the ICRU reference point A isodose line did not cover all the tumors. The minimum dose to the tumor volume defined by FDG-PET ranged from 50 to 475 cGy for treatment plans designed to deliver 650 cGy to point A and exhibited an inverse correlation with tumor volume. The authors proposed that this technique has the potential for improving iso-
dose tumor coverage for patients with cervical cancer while sparing critical structures.

Better understanding of tumor volume location with respect to the applicator could allow optimization of source position and dwell times to conform delivered doses to tumor volume. Figure 11.11a shows a conventional GYN brachytherapy dose distribution derived from PET images. This dose distribution is designed to deliver approximately 650 cGy in a single implant to point A using $^{192}$Ir high dose rate source. If the same integrated reference air kerma strength (IRAK) from Fig. 11.11a is rearranged and optimized to conform to the PET defined tumor volume, dose distribution in Fig. 11.11b is achieved. IRAK is a product of the source air-kerma strength (cGy·cm$^2$·h$^{-1}$) and the total dwell time Grigsby et al. (2001b). Dose conformality to the tumor in Fig. 11.11b is much better than the one in Fig. 11.11a, whereas the critical structure doses are comparable. A clinical trial is needed to explore the potential benefits of better dose conformity in GYN brachytherapy.

11.5 Discussion

There are numerous clinical trials in which PET, especially FDG-PET, has been shown to significantly enhance the accuracy of staging for patients who are candidates for surgical treatment and in whom the imaging results can be validated by pathology. The numbers of such studies for patients who are candidates for radiotherapy are few and there are even fewer studies which attempt to validate the accuracy of tumor volumes as outlined on PET images and the impact of revised tumor volumes on clinical outcomes. Without a more substantial clinical validation of PET-instigated changes in the management of radiotherapy patients, any modifications in radiotherapy treatment intent and/or modification of target volumes based on PET information must be approached with caution. Hicks and MacManus (2003) bring attention to many issues regarding radiotherapy treatment changes based on PET imaging and concerns about implementation of this technology in the patient management process. These authors note that in many cases being evaluated for curative radiotherapy, the same factors that precluded surgical therapy also prevent pathological confirmation of PET findings. Also important is their observation that PET imaging can be used as a complementary, rather than a competing, diagnostic tool, and that the complimentary use of CT and FDG-PET has been shown to provide higher accuracy than either of the imaging modalities alone for staging of NSCLC (Steinert et al. 1997; Pieterman et al. 2000), for example. This is due to the fact that CT may help overcome spatial resolution limitations of PET imaging. Other studies have also demonstrated limitations of PET imaging and argued for a complementary role in patient management (Poncelet et al. 2001; Osman et al. 2003). Using FDG-PET imaging as a complimentary tool in the management of radiotherapy patients is a very reasonable approach, especially given the fact that this is a relatively new imaging modality for many radiation oncology practices and that there is a relatively small amount of published data on the use of PET in radiotherapy. One of the major concerns is that false-positive FDG-PET results may deny some patients potentially curative treatment. This is an especially important concern since the majority of studies report increased number of upstaged patients due to PET information. The majority of oncological PET scans are acquired with FDG which is not a very tumor-specific substance. The FDG can, for example, be accumulated.
in benign lesions with elevated glucose metabolism, or there may be an increased accumulation in lymph nodes due to inflammatory reaction rather than tumor presence, resulting in possibly false-positive findings. The FDG uptake can be increased due to the effects of radiotherapy or chemotherapy, further complicating the evaluation of tumor location, shape, extent, and function. In order to accurately evaluate PET images, it is essential to understand various physiological uptake patterns of FDG for different cancer sites, and also to be aware of possibilities for false-positive and false-negative findings (Strauss 1996; Cook et al. 1996). Numerous investigators have reported superiority of FDG-PET for staging and detection of many tumor sites. Inevitably, the majority of these reports also warn of false-positive findings. Additionally, there are many studies which report possible sources of false-positive findings for many cancer sites. The number of these reports far surpasses the number of papers describing the use of PET for radiotherapy treatment-planning purposes. The correlation of FDG uptake with anatomic abnormalities detected by CT or MRI provides a valuable insight about functional and anatomical properties of a tumor. The PET/CT scanner and the availability of registered images simplifies, to an extent, the validation of PET findings.

11.6 Conclusion

One great opportunity for an overall improvement of radiation oncology is better understanding of tumors through biological imaging. Biological imaging has been shown to better characterize the extent of disease than anatomical imaging and also to better characterize individual tumor properties. Enhanced understanding of individual tumors can improve selection of the most appropriate therapy and better definition of target volumes. Improved target volumes can utilize the full potential of IMRT delivery. Biological imaging can also allow evaluation of tumor response and possibly modifications in therapy plan, if the initial therapy is deemed not effective.

There are numerous biological imaging techniques under development which will allow improvement of radiotherapy; however, the main modality that currently allows imaging beyond anatomy is FDG-PET. It is well established that PET imaging can significantly affect management of cancer patients, including those that are candidates for radiotherapy. FDG-PET has been shown to improve staging is several cancer sites. This improvement allows for more appropriate course of therapy for a significant number of patients who would otherwise potentially undergo futile radical therapy. This results in better quality of life for these patients and also health care costs reduction. Due to these benefits, the utilization of PET in oncology will only continue to grow.

It has been shown that PET can affect target volumes and result in delivery of escalated and more conformal dose distributions. It is unclear whether these modifications will result in improved tumor control and/or reduced side effects. Studies demonstrating positive impact of PET imaging in radiotherapy treatment planning process and meaningful improvement in target definition are needed to justify modification of conventional treatment volumes based on PET data.

The $^{18}$F-FDG is not a tumor-specific agent, there are numerous normal process and benign disease which can affect finding of PET imaging. Understanding of these process and potential for errors is imperative for implementation of FDG-PET imaging in radiotherapy treatment planning. The PET imaging will not benefit all patients; therefore, it is important to have appropriate selection process and timing of scans.

Other PET agents, many under development, have a potential to aid in achievement of some of the goals for this imaging modality in radiotherapy. The PET/CT combined scanner simplifies this process and improves access to PET images for many radiation oncology departments. Performance requirements for radiotherapy imaging are well defined and need to be applied to PET when used for treatment planning. Appropriate quality assurance program should also be implemented to avoid dosimetric and localization errors.

References


12 Patient Positioning in Radiotherapy Using Optical-Guided 3D Ultrasound Techniques

Wolfgang A. Tomé, Sanford L. Meeks, Nigel P. Orton, Lionel G. Bouchet, and Mark A. Ritter

CONTENTS

12.1 Introduction 151
12.2 Optical Tracking 152
12.3 Optically Guided 3D Ultrasound 152
12.4 Mathematics of Optical Tracking and Ultrasound Calibration 154
12.4.1 Mathematics of Optical Tracking 154
12.4.2 Mathematics of Optical-Guided 3D Ultrasound Calibration 155
12.4.2.1 Singular Value Decomposition 156
12.4.2.2 Horn’s Algorithm 156
12.5 Commissioning and Quality Assurance of 3D Ultrasound-Guided Systems 157
12.6 Clinical Applications 159
References 163

12.1 Introduction

Over the past decade, virtual simulation has become the standard of care for planning the majority of external beam radiotherapy treatments. There are many obvious technical advantages of virtual simulation over conventional radiotherapy simulation. Target identification is more accurate, because the target may be identified directly on CT and/or co-identified on MRI, functional imaging scans, and PET scans. Furthermore, the spatial relationship of the tumor to organs at risk is more easily appreciated in virtual simulation. This allows the treatment planner to tailor the beam orientations, field sizes, and field shapes to conformally avoid these organs at risk as far as possible. Further gains in conformal avoidance of organs at risk can usually be achieved when inverse planning is added to the treatment-planning arsenal. Virtual simulation has provided the framework for 3D conformal therapy and intensity modulated radiotherapy.

While virtual simulation provides a significant improvement over conventional radiotherapy, the actual delivery of such virtually designed treatment plans has been limited by the accuracy of the fiducial systems traditionally chosen in radiotherapy. Optically guided radiotherapy systems have the potential for improving the precision of treatment delivery by providing more robust fiducial systems. The ability of optically guided systems to accurately position internal targets with respect to the linear accelerator isocenter and to provide real-time patient tracking enables one at least theoretically to significantly reduce the amount of normal tissue included in the total irradiated volume by decreasing treatment field margins. Implementation and correct use of such systems present new challenges for the clinical physicist, and it is important that one thoroughly understands the strengths and weaknesses of such systems.

Furthermore, high-precision radiotherapy outside the cranium is challenging, because the target position can shift relative to bony anatomy between the time of initial image acquisition for virtual simulation and the time of the actual delivery of treatment. Our group has recently developed and clinically implemented an optically guided 3D ultrasound system for high-precision radiation delivery. The purpose of this chapter is to describe (a) optical ultrasound-guided radiotherapy, (b) the underlying mathematics that drive optical guidance, (c) the quality assurance techniques for such systems, and (d) the clinical use of optically guided 3D ultrasound patient positioning systems.
12.2 Optical Tracking

Tracking is the process of measuring the location of instruments, anatomical structures, and/or landmarks in 3D space and in relationship to each other. Various technologies have been tested for determining an object’s location, including mechanical, magnetic, acoustic, inertial, and optical position sensors. Most of these technologies have been tested for medical use in either image-guided surgery or image-guided tracking in radiation therapy.

Optical tracking systems use infrared light to determine a target’s position. The target may either be active or passive. The most common active targets are infrared light emitting diodes (IRLED). Passive targets are generally spheres or disks coated with a highly reflective material that reflects infrared light from an external source. Various detectors can be used to determine the positions of an optical target; however, charged couple device (CCD) cameras are used most often. The CCD cameras are simply a collection of light sensitive cells, or pixels, arranged in either a one- or two-dimensional array. When light strikes a CCD cell, electron production is proportional to the intensity of the light incident on the cell; thus, a 2D CCD array provides a 2D digital “image” of the target, with brighter pixels in the array corresponding to a higher light intensity and darker pixels corresponding to lower light intensity. This digital image can then be analyzed to determine the pixel with the highest light intensity. Each camera in a 2D CCD array determines a ray in 3D. If two 2D CCDs are used in the camera of an optical tracking system, the intersection of the two 2D rays emanating form the CCDs determines a line in 3D space, whereas if three 2D CCDs are used, the intersection of all three 2D rays determines a point in 3D space.

12.3 Optically Guided 3D Ultrasound

In extracranial radiotherapy, soft tissue targets can move relative to bony anatomy between the time of image acquisition for treatment planning and the time of the actual treatment delivery; therefore, real-time imaging of internal anatomy is required in extracranial radiotherapy if one would like to accurately localize a target at the time of treatment delivery. We have developed and tested an optical-guided system for 3D ultrasound guidance (BOUCHET et al. 2001; BOUCHET et al. 2002; MEERK et al. 2003; RYKEN et al. 2001; TOME et al. 2002) that is commercially available under the trade name SonArray (ZMed, Inc., Ashland, Mass.). Ultrasound was chosen because it is an inexpensive, yet flexible and high-resolution imaging modality that can easily be adapted for use in a radiation therapy treatment room. Systems have been developed that rely on 2D ultrasound probes attached to mechanical tracking systems, and these have proved effective for improving the precision of patient localization for prostate radiotherapy (TROCCAZ et al. 1995; LATTANZI et al. 1999; LATTANZI et al. 2000). The interpretation of 2D ultrasound images, however, is difficult and can be highly dependent on the skill and expertise of the operator in manipulating the transducer and mentally transforming the 2D images into a 3D tissue structure.

In principle, the use of 3D ultrasound can help overcome this limitation, but the relatively high cost of true 3D ultrasound devices has prohibited their use in radiotherapy for target localization on a routine basis; however, the development of our optical-guided 3D ultrasound target localization system has decreased the cost of such a system to a level that is comparable to commercially available 2D ultrasound target localization systems, and has therefore led to a more widespread use of reconstructed 3D Ultrasound for target localization in radiotherapy. The 3D ultrasound data sets are generated through optically tracking the acquisition of free-hand 2D ultrasound images. The operator holds the ultrasound probe and manipulates it over the anatomical region of interest. The raw 2D ultrasound images are transferred to a workstation using a standard video link. During acquisition, the position and angulation of each 2D ultrasound image is determined form the position and angulation of the 2D ultrasound probe by optically tracking a rigidly attached active fiducial array that has four infrared light-emitting diodes (IRLEDs; see Fig. 12.1). The position of each ultrasound pixel can therefore be determined, and an ultrasound volume can be reconstructed by coupling the position information obtained through optical tracking with the raw ultrasound data (see Figs. 12.2, 12.3).

Therefore, in addition to building the 3D image volume, optical guidance is also used to determine the absolute position of the ultrasound volume in the treatment room coordinate system. Because the relative positions of the ultrasound volume and the ultrasound probe are fixed, the knowledge of the probe position in the treatment room coordinate system at the time of image acquisition is sufficient to determine the position of the image volume relative to the linear accelerator isocenter. The relative position of the image and probe is determined during a calibration.
procedure (BOUCHET et al. 2001). This calibration procedure employs an optically tracked ultrasound phantom that contains 39 echoic wires at 13 depths in an anechoic medium. Because the phantom is optically tracked, the room position of each wire is known very accurately. The ultrasound image coordinate of each wire is then determined by collecting multiple images of the phantom. From this the ultrasound image to treatment room coordinate system transformation can be easily determined. The ultrasound image to ultrasound probe coordinate system transformation is then obtained by multiplying from the left the ultrasound image to treatment room coordinate system transformation with the inverse of ultrasound probe to treatment room coordinate system transformation. Once one has established the ultrasound image to ultrasound probe coordinate system transformation one is free to move the ultrasound probe anywhere in the treatment room coordinate system established by the optical tracking system.

Fig. 12.1. The ultrasound probe is optically tracked using an array that consists of four infrared light-emitting diodes, which is rigidly attached to the probe. Using optical tracking a 3D ultrasound volume can be obtained and referenced to the optical tracking system origin, which is typically the linear accelerator isocenter.

Fig. 12.2. Freehand ultrasound acquisition is used to acquire ultrasound data of an echoic spherical target at any orientation by either sliding or arcing the probe across the region of interest. One acquires arbitrary 2D ultrasound images until enough images have been acquired to fill a 3D matrix covering the volume of interest. The sagittal view (upper right hand corner) through the volume of interest shows the operator when a sufficient number of 2D ultrasound images have been acquired.
12.4 Mathematics of Optical Tracking and Ultrasound Calibration

12.4.1 Mathematics of Optical Tracking

In optical guidance for radiation therapy, we determine the image coordinates of the passive markers and the desired isocenter location from the virtual simulation CT scan, and we determine the room coordinates of the passive markers relative to the machine isocenter from optical tracking. The mathematics required for optical guidance simply entails the determination of the relationship between these two sets of points. Let us denote the desired marker coordinates from the CT images and the optically measured room coordinates of the markers by the following two $4 \times 1$ column matrices $\vec{p}_i = (x_i, y_i, z_i, 1)^T$ and $\vec{p}_R = (x_R, y_R, z_R, 1)^T$, respectively. Since it is unlikely that $\vec{p}_i$ is equal to $\vec{p}_R$, one must mathematically determine the relationship between these two matrices, and hence, determine the rotational and translational misalignment of the patient at the time of treatment relative to the time of virtual simulation.

The geometric transformation that relates $\vec{p}_i$ to $\vec{p}_R$ is
given by a 4×4 matrix $E_{I→R}$, which is an element of the Euclidian group in three dimensions E(3):

$$\tilde{p}_k = E_{I→R} \tilde{p}_i.$$  \hspace{1cm} (12.1)

It is well known that E(3) is the semidirect product of the group of rotations SO(3) and the group of linear translations T; therefore, we can write Eq. 12.1 also as:

$$\tilde{p}_k = \hat{R} \tilde{p}_i + \hat{T},$$  \hspace{1cm} (12.2)

where $\hat{R}$ is a 3×3 rotation matrix, $\hat{T}$ is a 3×1 translation matrix, and $\tilde{p}_i$ and $\tilde{p}_k$ are now considered as vectors in 3D space.

Only three non-colinear points with known positions in both room and image coordinates are needed to solve this equation; however, using more points reduces the statistical noise and increases the accuracy in determining the transformation matrices (Yang et al. 1999). We use a minimum of four fiducial markers in all of our fiducial arrays. Assuming that N points are available, $\hat{R}$ and $\hat{T}$ are the solution of the following least-square fit equation:

$$\varepsilon^2 = \sum_{k=1}^{N} \left\| \tilde{p}_{R,k} - (\hat{R} \tilde{p}_{I,k} + \hat{T}) \right\|^2.$$  \hspace{1cm} (12.3)

It is useful to change Eq.(12.3) by referring all points to the centroids $\tilde{p}_{R,C}$ and $\tilde{p}_{I,C}$ of the sets of points in room and image space using the centroid coincidence theorem (Yang et al. 1999). The centroids are determined using the geometric averages of the points in both coordinate systems:

$$\tilde{p}_{R,C} = \frac{1}{N} \sum_{k=1}^{N} \tilde{p}_{R,k},$$  \hspace{1cm} and  \hspace{1cm} (12.4)

$$\tilde{p}_{I,C} = \frac{1}{N} \sum_{k=1}^{N} \tilde{p}_{I,k}.$$  \hspace{1cm} (12.5)

Equation 12.3 can therefore be written in terms of $\tilde{p}_{R,k}' = \tilde{p}_{R,k} - \tilde{p}_{R,C}$ and $\tilde{p}_{I,k}' = \tilde{p}_{I,k} - \tilde{p}_{I,C}$:

$$\varepsilon^2 = \sum_{k=1}^{N} \left\| \tilde{p}_{R,k}' - (\hat{R} \tilde{p}_{I,k} + \hat{T}) \right\|^2,$$  \hspace{1cm} or  \hspace{1cm} (12.6)

$$\varepsilon^2 = \sum_{k=1}^{N} \left\| \tilde{p}_{R,k}' - \hat{R} \tilde{p}_{I,k}' \right\|^2 - 2(\tilde{p}_{R,C}' \cdot \hat{R} \tilde{p}_{I,C}' - \hat{T}).$$

$$\sum_{k=1}^{N} \left( \tilde{p}_{R,k}' \cdot \hat{R} \tilde{p}_{I,k}' \right) + N \left\| \tilde{p}_{R,C}' \cdot \hat{R} \tilde{p}_{I,C}' - \hat{T} \right\|^2.$$  \hspace{1cm} (12.7)

Since

$$\sum_{k=1}^{N} \left( \tilde{p}_{R,k}' \cdot \hat{R} \tilde{p}_{I,k}' \right) = N \left( \tilde{p}_{R,C}' \cdot \hat{R} \tilde{p}_{I,C}' \right) = 0,$$  \hspace{1cm} (12.8)

$$\sum_{k=1}^{N} \left( \tilde{p}_{R,k}' \cdot \hat{R} \tilde{p}_{I,k}' \right) = R \sum_{k=1}^{N} \left( \tilde{p}_{I,k}' \right) - N \left( \hat{R} \tilde{p}_{I,C}' \right) = 0.$$  \hspace{1cm} (12.9)

From Eqs. (12.8) and (12.9) we find that the middle term in Eq. 12.7 is equal to zero, and therefore Eq. 12.7 simply becomes:

$$\varepsilon^2 = \sum_{k=1}^{N} \left\| \tilde{p}_{R,k}' - \hat{R} \tilde{p}_{I,k}' \right\|^2 + N \left\| \tilde{p}_{R,C}' \cdot \hat{R} \tilde{p}_{I,C}' - \hat{T} \right\|^2.$$  \hspace{1cm} (12.10)

As one can see from Eq. (12.10) the translation minimizing the above equation is simply the difference vector of the image centroid and the rotated room centroid:

$$\hat{T} = \tilde{p}_{I,C}' \cdot \hat{R} \tilde{p}_{I,C}.$$  \hspace{1cm} (12.11)

Consequently, to determine the rotation, we have to minimize:

$$\varepsilon^2 = \sum_{k=1}^{N} \left\| \tilde{p}_{R,k}' - \hat{R} \tilde{p}_{I,k}' \right\|^2.$$  \hspace{1cm} (12.12)

A multitude of algorithms can be used to determine the best-fit rotation $\hat{R}$ from Eq. (12.12). Since this is a minimization problem, iterative optimization algorithms can be used to find the best-fit rotation. Several different optimization algorithms have been used to solve the patient orientation problem in stereotactic radiotherapy, including simulated annealing and various downhill algorithms such as the downhill simplex and the Hooke and Jeeves pattern search algorithm (Yang et al. 1999; Shoup and Mistree 1987). The solution space for the stereotactic radiotherapy application has been shown to be relatively flat, and downhill methods have proven fast and effective for stereotactic radiotherapy (Yang et al. 1999).

12.4.2 Mathematics of Optical-Guided 3D Ultrasound Calibration

The geometric transformation required for ultrasound calibration is ultrasound image space to treatment room space, and the end result is still minimization of Eq. (12.12). While downhill optimization algorithms are sufficient for solution of the absolute orientation as applied to stereotactic
radiotherapy, additional noise from ultrasound imaging decreases the reliability of these simple algorithms for determining the ultrasound calibration matrix (Bouchet et al. 2001). Below are brief discussions of two closed-form solutions algorithms (the singular value decomposition algorithm and Horn’s algorithm using quaternions) to the absolute orientation problem.

12.4.2.1 Singular Value Decomposition

The minimum of Eq. 12.12 is achieved when for all \( k \in \{1, \ldots, N\} \), we have \( \bar{p}_{k,k} = \hat{R}_{k,k} \). This can be rewritten in terms of a simple matrix equation:

\[
\begin{bmatrix}
X_{1,1} & X_{1,2} & \cdots & X_{1,N} \\
Y_{1,1} & Y_{1,2} & \cdots & Y_{1,N} \\
Z_{1,1} & Z_{1,2} & \cdots & Z_{1,N} \\
\end{bmatrix}
= \hat{R}
\begin{bmatrix}
X_{1,1} & X_{1,2} & \cdots & X_{1,N} \\
Y_{1,1} & Y_{1,2} & \cdots & Y_{1,N} \\
Z_{1,1} & Z_{1,2} & \cdots & Z_{1,N} \\
\end{bmatrix}
\tag{12.13}
\]

where \((X, Y, Z)\) represent the Cartesian coordinates of the points \( p \). Equation (12.13) represents an overdetermined linear set of equations that can be solved by using the singular value decomposition theorem. This theorem yields the decomposition of the resulting \( N \times 3 \) matrix into a product matrices of the form \( U W V^T \) where \( U, V \) are orthogonal matrices and \( W \) is a diagonal matrix whose diagonal elements are either positive or zero. After decomposition, the solution of Eq. (12.13) can be determined by inverting the corresponding orthogonal matrices. This closed-form solution gives the rotation that minimizes the least-square problem stated in Eq. (12.12).

12.4.2.2 Horn’s Algorithm

Another algorithm that can be used to solve Eq. (12.12) is the closed-form solution presented by Horn using quaternion theory (Horn 1987; Bouchet et al. 2001). Quaternions are an extension of complex numbers. Instead of just \( i \), one has three different numbers that are all square roots of \(-1\) labeled \( i, j, \) and \( k \), that fulfill the following relationship:

\[ i * j = \mathbb{E}_{ijk} k, \]

where \( \mathbb{E}_{ijk} \) is the totally antisymmetric tensor. The conjugate and the magnitude of a quaternion are formed in much the same way as the complex conjugate and magnitude:

\[
\hat{q} = q_0 + iq_x + jq_y + kq_z, \\
\hat{q}' = q_0 - iq_x - jq_y - kq_z.
\]

\[
\| \hat{q} \| = \sqrt{q_0^2 + q_x^2 + q_y^2 + q_z^2} = 1
\tag{12.14}
\]

In what follows we only between will and deal with unit quaternions, i.e., quaternions for which \( \| \hat{q} \| = 1 \). For unit quaternions the inverse of a quaternion is equal to its conjugate. Since unit quaternions form a representation of the special unitary group \( SU(2) \) one can express any 3D rotation \( R \) in terms of unit quaternions using the following isomorphism:

\[
\hat{r}_i = R \hat{r}_0 \rightarrow \hat{r}_i = \hat{q} * \hat{r}_0 * \hat{q}',
\tag{12.15}
\]

Expanding Eq. (12.12) and using the fact that rotations are isomorphisms, i.e., leave the norm unchanged, one finds:

\[
\varepsilon^2_R = \sum_{k=1}^{N} \| \bar{p}_{k,k} \|^2 + \sum_{k=1}^{N} \| \hat{R} \bar{p}_{k,k} \|^2 - 2 \sum_{k=1}^{N} \hat{R} \bar{p}_{k,k} \cdot \hat{R} \bar{p}_{k,k}' \tag{12.16}
\]

Therefore, minimizing Eq. (12.16) corresponds to maximizing the last term in this expression since the first two terms of this expression are positively definite:

\[
\xi = \sum_{k=1}^{N} \hat{R} \bar{p}_{k,k} \cdot \hat{R} \bar{p}_{k,k}' = \sum_{k=1}^{N} \hat{R} \bar{p}_{k,k} \cdot \bar{p}_{k,k}'
\tag{12.17}
\]

which can be written in terms of quaternions using the isomorphism (Eq. (12.15)) as follows:

\[
\xi = \sum_{k=1}^{N} \hat{q} * \hat{p}_{k,k} * \hat{q}' * \hat{p}_{k,k}'.
\tag{12.18}
\]

Using the fact that there exists a matrix representation of quaternion multiplication, Eq. (12.18) can be rewritten as (cf. Horn (1987)):

\[
\xi = qNq^T
\tag{12.19}
\]

where \( N = \sum_{k=1}^{N} N_k \) is a \( 4 \times 4 \) matrix formed from the coordinates of the points \( \bar{p}_{k,k} \) and \( \bar{p}_{k,k}' \) and the unit quaternion \( q \) is represented as a row vector of the form \( q = (q_x, q_y, q_z, q_w) \). Horn (1987) has shown that the unit quaternion maximizing Eq. (12.19) is the eigenvector corresponding to the largest positive eigenvalue of the matrix \( N \). Since \( N \) is a \( 4 \times 4 \) matrix, this corresponds to finding the roots of the fourth-degree characteristic polynomial of \( N \). This unit quaternion corresponds to the closed-form solution of the rotation minimizing the least-square problem stated in Eq. (12.12).
12.5 Commissioning and Quality Assurance of 3D Ultrasound-Guided Systems

To ensure accurate localization, all possible errors in the imaging, patient localization, and treatment delivery processes must be systematically analyzed (Bouchet et al. 2002; Tomé et al. 2002). Outlined below are test procedures necessary to meet the quality assurance challenges presented by an optically guided 3D ultrasound system for real-time patient localization. While all tests were performed using the SonArray system, the general philosophy and procedures are applicable to all systems utilizing this technology. Determination of absolute localization accuracy requires the user to establish a consistent stereotactic, or 3D, coordinate system in both the treatment planning system and the treatment vault. While we chose to establish this coordinate system through optical guidance, it can also be done using mechanical means, as in conventional stereotactic procedures, or other telemetry technologies that are commercially available. Regardless of the methodology utilized, it is imperative that acceptance tests be performed prior to clinical use of the system to ensure that the image-guided system allows for safe, controlled, and efficient delivery of both conventional and intensity-modulated radiotherapy.

For our testing, we used a specially designed ultrasound phantom (Fig. 12.4). This phantom consists of 12 echoic spheres embedded in a tissue-equivalent non-echoic medium (with speed of sound 1470 m s\(^{-1}\)), with a passive infrared fiducial array attached to it. The spheres are arranged at five different nominal depths: one at 30 mm; two at 50 mm; and the remaining nine arranged in groups of three at nominal depths of 70, 100, and 130 mm. A CT scan (0.49\(\times\)0.49\(\times\)2.0-mm resolution) of the phantom was acquired. Using the fiducial array for optical tracking, the same stereotactic coordinate system was established in the treatment planning system and in the treatment room, relative to which the positions of each sphere are known within imaging uncertainty. Each of the spheres were localized in the Pinnacle treatment planning system and its coordinates were transferred to the SonArray system as intended treatment isocenters. In order to reproduce the exact position of the ultrasound phantom at the time of the CT, a 2D couch mount has been employed in the treatment room. This couch mount has two orthogonal rotational axes (spin, tilt). Together with the four degrees of freedom of the treatment couch, three orthogonal translations (anteroposterior, lateral, vertical), and one rotational degree of freedom, this allows reproduction of the position of the phantom to within a predefined error tolerance. Each target sphere was positioned at the treatment machine isocenter using the following method: first the isocenter corresponding to the target sphere chosen for ultrasound localization was selected on the control computer. The phantom was then moved using optical tracking and the couch controls until the RMS error between actual and desired position of the target sphere was less than 0.2 mm. Once a target sphere had been positioned at the treatment machine isocenter using optical tracking, the ultrasound probe was fixed on top of the phantom. A 3D ultrasound volume of each sphere was acquired using optical tracking as described above. The 3D-ultrasound-based position of the target sphere was determined by finding the center of the sphere in the axial, sagittal, and coronal planes using a circle tool placed on each of the three orthogonal ultrasound views. The target localization accuracy of the 3D-ultrasound optically guided system was thus determined by comparing the experimentally determined position of each sphere to its predicted position from the treatment planning system.

Fig. 12.4. A specifically designed ultrasound phantom that has a number of echoic spheres embedded at different depths in a non-echoic medium. This phantom can be optically tracked to determine the coordinates of the spheres in the treatment room, which enables it to be used to test an ultrasound target localization system.
Our data show that the localization error does not depend on the target depth or the ultrasound focal depth used. Table 12.1 demonstrates representative values for such a test of optically guided 3D-ultrasound target localization for a 15-cm depth ultrasound probe format, which is the typical depth one employs in ultrasound localizations of the prostate. Similar to optical guidance testing described previously, optically guided 3D-ultrasound localization should be able to localize a well-defined internal target to within the inherent imaging uncertainty; however, localization errors in each of the spatial dimensions may exceed the predicted localization error due to finite image pixel size.

Other experiments can be performed using anthropomorphic phantoms. As an example we describe below an experiment using a specially designed prostate phantom. This prostate phantom consists of three layers. The first layer contains a model of the bladder, the second a model of the prostate and urethra, and the third a model for the rectum in the form of a long cylinder. The bladder, prostate, urethra, rectum, and the background material are made of Zerdine with different echogenicity closely mimicking sound absorption and speed properties characteristic for these anatomical structures. A CT scan (0.49×0.49×2.0-mm resolution) of the phantom with a passive infrared fiducial array attached was acquired. Using optical tracking as described previously, the positions of each organ can be determined in both CT image space and treatment room space within imaging uncertainty. In the CT data set four distinct regions of interest (ROIs) – bladder, prostate, urethra, and rectum – were segmented. A treatment isocenter was chosen in the treatment planning system, and the planning CT, isocenter, and segmented ROIs were transferred to the control computer of the 3D-ultrasound target localization system. The phantom was then set up on the treatment couch as shown in Fig. 12.5.

Again, optical tracking can be used to align the planned treatment isocenter with the treatment machine isocenter within 0.2 mm RMS error. The fiducial array attached to the phantom is then covered and a separate fiducial array is attached to the treatment couch, and the position of this second array is recorded using the optical positioning sensor system. In this way a fixed “bony anatomy” is established about which introduced internal organ motion can be simulated as follows: first the fiducial array attached to treatment couch is covered and the fiducial array attached to the phantom is uncovered. Then the phantom is shifted from its starting position ±5 mm in the lateral and/or superior/inferior directions using a 2D translation table. These shifts are measured to within ±0.2 mm RMS error using the optical positioning sensor system. Once the shifts have been made, the fiducial array attached to the phantom is again covered and the fiducial array attached to the treatment couch is uncovered. Since the treatment couch position remains fixed, the “bony anatomy” is not changed, even though the phantom has been physically moved; hence, through the alternate use of two fiducial arrays one is able to introduce accurate apparent organ motion while maintaining a fixed “bony anatomy.” This apparent organ motion can then be detected using optically guided 3D ultrasound imaging by employing the optically guided 3D ultrasound target localization system.

<table>
<thead>
<tr>
<th>Depth (mm)</th>
<th>Anteroposterior distance (mm)</th>
<th>Lateral distance (mm)</th>
<th>Axial distance (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.8±0.6</td>
<td>0.5±0.3</td>
<td>1.0±1.7</td>
</tr>
<tr>
<td>50</td>
<td>0.1±0.7</td>
<td>0.6±0.8</td>
<td>1.3±1.6</td>
</tr>
<tr>
<td>70</td>
<td>-0.2±0.6</td>
<td>1.0±0.9</td>
<td>0.1±1.5</td>
</tr>
<tr>
<td>100</td>
<td>0.2±0.8</td>
<td>1.2±0.9</td>
<td>0.3±1.8</td>
</tr>
<tr>
<td>130</td>
<td>0.3±0.6</td>
<td>1.1±1.0</td>
<td>0.2±1.5</td>
</tr>
</tbody>
</table>

Fig. 12.5. Experimental setup of the prostate phantom in the treatment room for determining unknown induced motion in a target structure using an ultrasound target localization system.
tem in clinical mode. This experiment gives a lower limit of the localization accuracy of the optically guided 3D ultrasound target localization system in daily clinical use.

The measurements for each translation of the phantom were repeated non-consecutively five times. Once an organ shift had been measured using optic guided 3D-ultrasound localization, the phantom was translated to a new position and a new 3D-ultrasound localization was performed. This procedure was followed until the required number of ultrasound localizations of each phantom shift had been obtained. Repeated performance of this experiment shows that one is able to localize an internal structure to within the inherent imaging uncertainty (cf. Table 5 of Tomé et al. 2002). Again, localization errors in each spatial dimension may exceed the predicted localization error due to finite image pixel size.

In the preceding paragraphs we have described two methods for quantitative testing of the accuracy of an optically guided 3D ultrasound target localization system. While many qualitative tests exist for testing these systems, we believe that it is extremely important for the physicist to perform commissioning tests of the system that quantify the errors locally. For this reason we have summarized these tests in Table 12.2 together with their suggested frequency. As mentioned in the introductory paragraph of this section, this requires the user to establish some local standard coordinate system in the treatment room. We have chosen to establish this through optical tracking, but such a standard can also be established using alternative mechanisms.

### 12.6 Clinical Applications

We have used daily transabdominal 3D ultrasound localization for conformal and intensity-modulated radiotherapy of the prostate employing a rectal balloon (Patel et al. 2003), whereas Chinnaiyan et al. (2003) have employed daily transabdominal 3D ultrasound localization in the treatment of prostate cancer in the post-operative adjuvant or salvage setting to improve the reproducibility of coverage of the intended volumes and to enhance conformal avoidance of adjacent normal structures. They studied 16 consecutive patients who received external beam radiotherapy and underwent daily localization using an optically guided 3D-ultrasound target localization system. Six of these patients were treated in a post-prostatectomy setting, either adjuvantly or for salvage, while the remaining 10 with intact prostates were treated definitively. During 3D ultrasound localization the bladder base was used as the primary localization structure for the post-prostatectomy patients (see Fig. 12.6), whereas for patients treated definitively the prostate was the primary reference structure (see Fig. 12.7). They found that the ultrasound-based displacements were not statistically different between the two groups of patients, and therefore concluded that daily transabdominal 3D-ultrasound localization is a clinically feasible method of correcting for internal organ motion in the post-prostatectomy setting.

In addition, this system has been used for patient localization for patients undergoing extracranial radiosurgery for a variety of abdominal, paraspinal,

<table>
<thead>
<tr>
<th>Initially (perform tests in sequence)</th>
<th>Daily before first treatment</th>
<th>Annually and with each software update</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verify camera calibration</td>
<td>Verify camera calibration</td>
<td>Verify camera calibration</td>
</tr>
<tr>
<td>Verify ultrasound calibration</td>
<td>Verify ultrasound calibration</td>
<td>Verify ultrasound calibration</td>
</tr>
<tr>
<td>Known target test (verify entire process on sphere phantom)</td>
<td>Known target test (verify entire process on sphere phantom)</td>
<td>Acquire new CT image set</td>
</tr>
<tr>
<td>Acquire new CT image set</td>
<td>Load into TPS and create contours and isocenter</td>
<td>Load into TPS and create contours and isocenter</td>
</tr>
<tr>
<td>Send to US system</td>
<td>Send to US system</td>
<td>Send to US system</td>
</tr>
<tr>
<td>Verify isocenter and contours</td>
<td>Verify isocenter and contours</td>
<td>Localize spheres</td>
</tr>
<tr>
<td>Localize spheres</td>
<td>Unknown target test</td>
<td>Unknown target test</td>
</tr>
<tr>
<td>Unknown target test (verify entire process on prostate phantom as outlined above)</td>
<td>Unknown target test (verify entire process on prostate phantom as outlined above)</td>
<td></td>
</tr>
</tbody>
</table>
and pelvic lesions (Ryken et al. 2001; Meeks et al. 2003). In general, the clinical use of optical-guided 3D ultrasound image guidance proceeds as follows: prior to CT scanning, the patient is immobilized using a custom vacuum cushion as is commonly used in radiation therapy (Vac-Loc, Med-Tec, Inc., Orange City, Iowa). The CT is acquired with the patient immobilized in the same position that will be used during the radiotherapy treatment in order to maintain a generally consistent position of mobile anatomy. The CT images are transferred to the treatment planning system, where the tumor volume and normal structures of interest are delineated. A treatment plan is then designed to conform the prescription dose closely to the planning target volume (PTV) while minimizing the dose to the nearby normal structures, using either 3D conformal radiotherapy or intensity modulated radiotherapy.

On each day of the treatment, the patient is placed in the same immobilization cushion that was used during CT scanning. The patient is initially set up relative to isocenter using conventional laser alignment. A 3D ultrasound volume is then acquired and reconstructed on the workstation. The target volume and critical structure outlines, as delineated on the planning CT scans, are overlaid on the acquired ultrasound volume in relation to the linear accelerator isocenter. The contours determined from the CT scans are then manipulated until they align with the anatomic structures on the ultrasound images (cf. Fig. 12.8). The amount of movement required to align the contours with the ultrasound images determines the magnitude of the target misregistration with iso-
center based on conventional setup techniques. The target is then placed at the isocenter by tracking a fiducial array attached to the treatment couch, which allows precise translation from the initial treatment room laser setup position to the 3D ultrasound determined setup position. Once the shifts indicated by optically guided 3D ultrasound target localization system have been made, treatment proceeds as planned.

However, the question becomes: What is the dosimetric impact on the treatment plan due to shifting the patient daily using 3D ultrasound target localization? ORTON and Tomé (2004) have studied the effects of such daily shifts on dose distributions in the prostate PTV, rectal wall, and bladder wall for intensity-modulated radiotherapy for a ten-patient cohort. The shifts in their study were based on daily ultrasound imaging using an optical-guided 3D-ultrasound target localization system; however, their results are applicable to daily shifts derived using other methods. To investigate how these shifts affect dose distributions and predicted outcomes, they generated treatment plans for three cases: (a) the initial preplan, which represents the ideal case in which no shifts are necessary; (b) a postplan incorporating each day’s actual shifts; and (c) a postplan in which no shifts were made but the internal organs were moved by the amounts indicated by daily 3D US imaging. They found that when daily shifts were made, doses to the target, rectal wall, and bladder wall are virtually identical to those in the preplan; however, when the indicated shifts were not carried out dose distributions degraded as shown in Fig. 12.9. For a typical patient, PTV-EUDs are 99.7% of the preplan PTV-EUD for the postplan with shifts and 92.7% of the preplan PTV-EUD for the postplan without shifts.

![Graphical interface which allows the operator to correlate the treatment-planning CT data and the reconstructed 3D ultrasound volumes to each other. Contours of the target structures and the organs at risk generated during the treatment-planning process are overlaid onto the 3D ultrasound data set. The operator can shift the contours in all three views until a satisfactory match between contours and depicted ultrasound anatomy is achieved.](image-url)
In their study an evenly spaced seven-beam arrangement around the patient was used, so differences in SSD and depth for each beam were largely offset by beams entering on the opposite side of the patient. It is important to note that other beam geometries might not yield such good results. Also, daily differences in organ shapes were not modeled in their study. Nonetheless, this study illustrates that delivery of IMRT without adequate target localization can induce cold spots in the target that can potentially compromise local tumor control probability, especially if smaller PTV margins are used (Tomé and Fowler 2002).

Whereas ultrasound localization can provide high-precision target localization, it is important to note that the clinical use of ultrasound images in radiation therapy requires training and skill from the user. It is unlikely that users with no training in the interpretation of ultrasound images will generate results that improve target localization accuracy. To illustrate this, we have undertaken a pilot study at the University of Iowa to determine the impact of training in the interpretation of 3D ultrasound sound images on the localization of the prostate. In this pilot study nine different users independent of each other retrospectively registered 15 patient data sets. Four of the users had approximately 1 year of 3D ultrasound interpretation and localization experience, whereas the other five users had been trained in the use of 3D ultrasound localization system but had no experience in ultrasound interpretation. Results of this pilot study are shown in Table 12.3 and indicate that the standard deviation among untrained users is of the order of the average required shift. Among those trained in 3D ultrasound interpretation, however, the results are consistent and indicate that significant increases in target localization using daily 3D ultrasound localization can be obtained.

![Theoretical composite treatment plan](image1)

![Composite Treatment plan in which the indicated shifts were applied](image2)

![Composite Treatment plan in which the indicated organ shifts were not applied](image3)

**Fig. 12.9a–c.** a The ideal isodose distribution that would be delivered if internal organ motion and setup variation were absent. b Isodose distribution that is delivered when one corrects for organ motion and setup variation daily using 3D ultrasound imaging. c Isodose distribution that would result if one did not correct for these variations in setup. The 76 Gy (green), 70 Gy (sky blue), 65 Gy (purple), 54 Gy (dark green), and 44 Gy (dark purple) isodose lines are shown.
Table 12.3 User variability in prostate localization using ultrasound guidance. AP anteroposterior

<table>
<thead>
<tr>
<th></th>
<th>AP (mm)</th>
<th>Lateral (mm)</th>
<th>Axial (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average shift</td>
<td>3.4</td>
<td>2.7</td>
<td>4.5</td>
</tr>
<tr>
<td>Standard deviation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(four trained users)</td>
<td>1.2</td>
<td>0.9</td>
<td>1.4</td>
</tr>
<tr>
<td>Standard deviation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(five untrained users)</td>
<td>3.6</td>
<td>1.5</td>
<td>2.9</td>
</tr>
</tbody>
</table>

References

3D Treatment Planning for Conformal Radiotherapy
13 Definition of Target Volume and Organs at Risk.

Biological Target Volume

Anca-Ligia Grosu, Lisa D. Sprague, and Michael Molls

CONTENTS

13.1 Introduction 167
13.2 Definition of Target Volume 168
13.2.1 Gross Target Volume 168
13.2.2 Clinical Target Volume 168
13.2.3 Internal Target Volume 169
13.2.4 Planning Target Volume 169
13.2.5 Treated Volume 169
13.2.6 Irradiated Volume 169
13.3 Definition of Organs at Risk 169
13.4 New Concepts in Target Volume Definition: Biological Target Volume 170
13.4.1 Where is the Tumour Located and Where Are the (Macroscopic) Tumour Margins? 170
13.4.1.1 Lung Cancer 170
13.4.1.2 Head and Neck Cancer 172
13.4.1.3 Prostate Cancer 172
13.4.1.4 Brain Gliomas 172
13.4.2 Which Relevant Biological Properties of the Tumour Could Represent an Appropriate Biological Target for Radiation Therapy? 173
13.4.3 What is the Biological Tumour Response to Therapy? 174
13.5 Conclusion 175
References 176

13.1 Introduction

The definition of the target volume and of critical structures is a crucial and complex process in three-dimensional conformal radiation therapy. In a planning system, based either on computed topography (CT) or magnetic resonance imaging (MRI), the radiation oncologist is required to outline the target volume to be irradiated, and the organs of risk to be spared, because of possible side effects. In this process, a multitude of information has to be taken into consideration: the results of radiological and clinical investigations; tumour staging; surgical and histo-pathological reports; other additional treatments such as chemotherapy; immune therapy; the patient’s history; the anatomy of the region to be irradiated; and the acceptance of the patient concerning radiation treatment. But also the technique used for irradiation, including the patient’s positioning and fixation, are of major importance. As a consequence, the complexity of the process when defining the target volume requires sound clinical judgement and knowledge from the radiation oncologist.

Three-dimensional conformal treatment planning in radiation oncology is based on radiological imaging, CT and MRI. These investigative techniques show the anatomical structures with a high accuracy. Both CT and MRI image tumour tissue by taking advantage of the differences in tissue density (or signal intensity in MRI), contrast enhancement or the ability to accumulate water (oedema). But these signals are not specific to tumour tissue alone. They can also be observed after a trauma or surgery, or inflammatory or vascular disease. This can hamper diagnosis and delineation of a tumour, and represents the most important limitation of these traditional radiological investigation techniques.

Biological imaging visualises biological pathways. Positron emission tomography (PET) and single photon computed emission tomography (SPECT) are characterised by visualising a tumour using radioactive tracers. These tracers have a higher affinity for tumorous tissue in comparison with normal tissue. Magnetic resonance spectroscopy (MRS) imaging provides information on the biological activity of a tumour based on the levels of cellular metabolites; therefore, the information obtained from different imaging modalities is in general complementary by nature.

This chapter discusses the standard concepts defined in the ICRU report 50 (ICRU 50 1993) and the ICRU report 62 (ICRU 62 1999). Additionally, we discuss the impact of biological investigations in target volume delineation.
13.2 Definition of Target Volume

Both ICRU report 50 from 1993 and ICRU report 62 from 1999 standardised the nomenclature used for three-dimensional conformal treatment planning and thus gave the community of radiation oncologists a consistent language and guidelines for image-based target volume delineation. The following terms were defined: gross tumour volume (GTV); the clinical target volume (CTV); the internal target volume (ITV); the planning target volume (PTV); the treated volume; and the irradiated volume (Fig. 13.1).

Fig. 13.1. Concepts used in target volume definition for radiation treatment.

13.2.1 Gross Target Volume

The gross tumour volume (GTV) is the macroscopic (gross) extent of the tumour as determined by radiological and clinical investigations (palpation, inspection). The GTV-primary (GTV-P) defines the area of the primary tumour and GTV-nodal (GTV-N) the macroscopically involved lymph nodes. The GTV is obtained by summarising the area outlined by the radiation oncologist in each section, multiplied by the thickness of each section. The extension of the GTV is of major importance for the treatment strategy: in many cases the gross tumour tissue is irradiated with higher doses, as it is encompassed within the boost volume. The GTV can represent the total volume of the primary tumour, the macroscopic residual tumour tissue after partial tumour resection or the region of recurrence after either surgical, radiation or chemotherapeutic treatment.

The delineation of the GTV is usually based on data obtained from CT and MRI. In general, tumour tissue shows contrast enhancement, has a different density (CT) or intensity (MRI) compared with the normal tissue and is surrounded by perifocal oedema. Based on these radiological characteristics, the radiation oncologist has to outline on each section the areas of gross tumour tissue. The demarcation of the GTV needs profound radiological knowledge. The technique used for the CT or MRI investigation and the window and level settings have to be appropriate for the anatomical region. The volume of the gross tumour tissue visualised on CT or MRI should correspond to the volume of the actual macroscopic tumour extension.

A major problem when delineating residual tumour tissue in surgically treated patients is the recognition and differentiation of changes caused by surgery itself. But also contrast enhancement, oedema and hyper- or hypodensity (intensity) of the tissue can cause difficulties, ultimately resulting in inaccurate GTV delineation.

Sometimes it is extremely difficult or even impossible to delineate the gross tumour mass when using conventional radiological investigation techniques such as CT and MRI. An example is the visualisation of prostate cancer with current imaging methods (CT, MRI, sonography). In this case, for three-dimensional treatment planning, the radiation oncologist delineates the prostate encompassing the GTV and the clinical target volume (CTV).

An accurate delineation of tumour tissue should focus the irradiation dose on the GTV and spare surrounding normal tissue. New investigative techniques, such as positron emission tomography PET, SPECT and MRS, visualise tumour tissue with a higher specificity. It seems likely that these techniques will help in the future to delineate tumour tissue with higher precision. The possibility to integrate these “biological” investigative techniques in GTV definition is discussed in the section “Biological Target Volume”.

13.2.2 Clinical Target Volume

The GTV, together with the surrounding microscopic tumour infiltration, constitutes the primary clinical target volume CTV (CTV-P). It is important to mention that the definition of the CTV-P also includes the tumour bed, which has to be irradiated after a complete macroscopic tumour resection, in both R0 (complete microscopic resection) and R1 (microscopic residual tumour on the margin of the tumour bed) situation. Moreover, for CTV-P definition, anatomical tumour characteristics have to be considered, such as the likelihood of perineural and perivascular extension or tumour spread along anatomical borders. As the margins
between CTV and GTV are not homogenous, they have to be adjusted to the probable microscopic tumour spread. The CTV-nodal (CTV-N) defines the assumed microscopic lymphatic tumour spread, which has also to be included in the radiation treatment planning.

13.2.3 Internal Target Volume

The internal target volume (ITV) is a term introduced by the ICRU report 62. The ITV encompasses the GTV/CTV plus internal margins to the GTV/CTV, caused by possible physiological movements of organs and tumour, due to respiration, pulsation, filling of the rectum, or variations of tumour size and shape, etc. It is defined in relation to internal reference points, most often rigid bone structures, in an internal patient coordinate system. Observation of internal margins is difficult, in many cases even impossible. Examples for reducing internal margins are the fixation of the rectum with a balloon during irradiation of the prostate or body fixation and administration of oxygen to reduce respiration movements during stereotactic radiotherapy.

13.2.4 Planning Target Volume

The planning target volume (PTV) incorporates the GTV/CTV plus margins due to uncertainties of patient setup and beam adjustment; therefore, these margins consider the inaccuracy in the geometrical location of the GTV/CTV in the irradiated space, due to variations in patient positioning during radiotherapy and organ motility. The PTV can be compared with an envelope and has to be treated with the same irradiation dose as the GTV/CTV. Movements of the GTV/CTV within this envelope should not change the delivered radiation dose to the GTV and CTV. The setup margins are not uniform. The radiation oncologist should define the margins together with the radiation physicist, taking into consideration the possible inaccuracy in patient and beam positioning. Significant reduction of setup margins can be obtained by applying patient fixation and re-positioning techniques, as used, for example, in stereotactic radiotherapy. Beam positioning uncertainties should be subject of a specific quality control program of each treatment unit. Consequently, the volume of normal tissue surrounding the tumour and included in high irradiation dose areas can be considerably reduced and the dose applied to the GTV/CTV can be escalated.

13.2.5 Treated Volume

The treated volume is the volume of tumour and surrounding normal tissue included in an isodose surface representing the irradiation dose proposed for the treatment. As a rule this corresponds to the 95% isodose. Ideally, the treated volume should correspond to the PTV; however, in many cases the treated volume exceeds the PTV. The coherence of an irradiation plan can be described/illustrated as the correlation between PTV and treated volume.

13.2.6 Irradiated Volume

The irradiated volume is a volume included in an isodose surface with a possible biological impact on the normal tissue encompassed in this volume; therefore, the irradiated volume depends on the selected isodose curve and the normal tissue surrounding the tumour. The choice of an isodose surface depends on the end point defined for possible side effects in normal tissue. For grade-III and grade-IV side effects, the isodoses will correspond to higher doses, whereas low-dose areas are significant for the risk of carcinogenesis.

13.3 Definition of Organs at Risk

Organs at risk (ICRU report 50), also known as critical structures, are anatomical structures with important functional properties located in the vicinity of the target volume. They have to be considered in the treatment planning, since irradiation can cause pathological changes in normal tissue, with irreversible functional consequences. The critical structures must be outlined in the planning process and the dose applied to these areas has to be quantified by either visualising the isodose distribution or by means of dose-volume histograms (DVH). It is essential to include the tolerance dose of the organs at risk in the treatment strategy.

The ICRU report 62 furthermore added the term “planning organs at risk volume” (PRV). This term
takes into account that the organs at risk are in the majority mobile structures; therefore, a surrounding margin is added to the organs at risk in order to compensate for geometric uncertainties.

With regard to histopathological properties of tissue, organs at risk can be classified as serial, parallel and serial-parallel (Källmann et al. 1992; Jackson and Kutcher 1993). Serial organs can lose their complete functionality even if only a small volume of the organ receives a dose above the tolerance limit. Best known examples are spinal cord, optical nerve and optical chiasm. In contrast, parallel organs are damaged only if a larger volume is included in the irradiation region, e.g. lung and kidney; however, in many cases these two models are combined in a serial-parallel organ configuration. Side effects occur due to various pathological mechanisms, size of the irradiated volume and the maximal dose applied to the organ. A classical example for a serial-parallel organ is the heart: the coronary arteries are a parallel and the myocardium a serial organ.

The organs at risk can be located at a distance from the PTV, close to the PTV or incorporated within the tumour tissue and thus be in the PTV. These three situations have to be considered in the treatment planning. Organs with a low tolerance to irradiation (lens, gonads) have to be outlined even if they are not located in the immediate vicinity of the PTV. For organs situated close to the PTV a plan with a very deep dose fall towards the critical structures has to be designed. If the organs at risk are encompassed within the PTV, it is crucial to achieve a homogeneous dose distribution, and to consider the $D_{\text{max}}$ and the isodose distribution in the PTV.

Thus far, biological imaging has not yet been integrated into radiation treatment planning; however, several trials indicate that biological imaging could have a significant impact on the development of new treatment strategies in radiation oncology. These trials show that PET, SPECT or MRS might be helpful in obtaining more specific answers to the three essential questions given in the sections that follow.

13.4.1 Where is the Tumour Located and Where Are the (Macroscopic) Tumour Margins?

The co-registration of biological and anatomical imaging techniques seems to improve the delineation of viable tumour tissue compared with CT or MRI alone. Biological imaging enables the definition of a target volume on the basis of biological processes. For some cases this results in clearer differentiation between tumour and normal tissue as compared with CT and MRI alone; however, anatomical imaging will continue to be the basis of treatment planning because of its higher resolution. Incorporation of biological imaging into the tumour volume definition should only be done in tumours with increased sensitivity and specificity to biological investigation techniques.

We discuss three clinical entities: lung cancer, head and neck cancer, prostate cancer and brain gliomas as biological imaging proved to be helpful in target volume delineation.

13.4.1.1 Lung Cancer

The role of fluoro-deoxyglucose (FDG)-PET in radiation treatment planning of lung cancer has been thoroughly investigated in a total of 415 patients in 11 trials (Table 13.1; Grosu et al. 2005c). These investigations compared FDG-PET data with CT data. The CT/PET image fusion methods were used in five trials. One study used the integrated PET/CT system. All these studies suggested that FDG-PET adds essential information to the CT with significant consequences on GTV, CTV and PTV definition. The percentage of cases presenting significant changes in tumour volume after the integration of FDG-PET investigation in the radiation treatment planning ranged from 21 to 100%. Ten studies pointed out the significant implications of FDG-PET in staging lymph node involvement. These findings are supported by data in the literature, showing an advantage of FDG-PET over CT in lung cancer.
Table 13.1 Impact of FDG-PET in gross target volume (GTV) and planning target volume (PTV) delineation in lung cancer.

<table>
<thead>
<tr>
<th>Reference</th>
<th>No. of patients</th>
<th>Change of GTV and PTV post-PET (n)</th>
<th>Increase of GTV and PTV using PET</th>
<th>Decrease of GTV and PTV using PET</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herbert et al. (1996)</td>
<td>20</td>
<td>GTV: 7 of 20 (35%)</td>
<td>GTV 3 of 20 (15%)</td>
<td>GTV 4 of 20 (20%)</td>
<td>PET may be useful for delineation of lung cancer</td>
</tr>
<tr>
<td>Kiffer et al. (1998)</td>
<td>15</td>
<td>GTV: 7 of 15 (47%); PTV: 4 of 15 (27%)</td>
<td></td>
<td></td>
<td>PET defects positive lymph nodes, not useful in tumour delineation</td>
</tr>
<tr>
<td>Munley et al. (1999)</td>
<td>35</td>
<td>PTV: 12 of 35 (34%)</td>
<td>PTV: 12 of 35 (34%)</td>
<td></td>
<td>PET target smaller than CT not evaluated</td>
</tr>
<tr>
<td>Nestle et al. (1999)</td>
<td>34, stage IIIA–IV</td>
<td>Change of field size (cm²) in 12 (35%); median 19.3%</td>
<td>Increase of field size 9 (26%)</td>
<td>Decrease of field size: 3 (9%)</td>
<td>Change of field size in patients with dys- or atelectasis</td>
</tr>
<tr>
<td>Vanuytsel et al. (2000)</td>
<td>73 (N+), stage IIA–IIIB</td>
<td>GTV: 45 of 73 (62%)</td>
<td>GTV: 16 of 73 (22%) 11 patients with pathology; 1 patient unnecessary; 4 patients insufficient</td>
<td>GTV: 29 of 73 (40%); 25 patients pathology; 1 patient inappropriate; 3 patients insufficient</td>
<td>PET data vs pathology: 36 (49%)=pathology; 2 (3%) inappropriate; 7 (10%) insufficient</td>
</tr>
<tr>
<td>MacManus et al. (2001)</td>
<td>153 stage IA–IIIB, unresectable candidates for radical RT after conventional staging</td>
<td>GTV: 22 of 102 (21%)</td>
<td>GTV: 22 of 102 (21%); inclusion of structures previously considered uninvolved by tumour</td>
<td>GTV: 16 of 102 (15%); exclusion of atelectasis and lymph nodes</td>
<td>Post-PET stage but not pre-PET stage was significantly associated with survival</td>
</tr>
<tr>
<td>Giraud et al. (2001)</td>
<td>12</td>
<td>GTV, PTV: 5 of 12 (42%)</td>
<td></td>
<td></td>
<td>4 of 12 lymph nodes; 1 of 12 atelectasis and distant meta</td>
</tr>
<tr>
<td>Mah et al. (2002)</td>
<td>30, stage IIA–IIIB</td>
<td>GTV: 5 of 23 (22%); FDG-avid lymph nodes</td>
<td>PTV: 30–76% of cases (varied between the three physicians)</td>
<td>PTV: 24–70% of cases (varied between the three physicians)</td>
<td>Addition of PET does lower physician variation in PTV delineation; PET-significant alterations to patient management and PTV</td>
</tr>
<tr>
<td>Erdi et al. (2002)</td>
<td>11, N+</td>
<td>PTV: 11 of 11 (100%)</td>
<td>PTV: 7 of 11 (36%); 19% (5–46%) cc; detection of lymph nodes</td>
<td>PTV: 4 of 11 18% (2–48%) cc; exclusion of atelectasis trimming the target volume to spare critical structure</td>
<td>PET improves GTV and PTV definition</td>
</tr>
<tr>
<td>Ciernik et al. (2003)</td>
<td>6</td>
<td>GTV: 100%</td>
<td>GTV: 1 of 6 (17%)</td>
<td>GTV: 4 of 6 (67%)</td>
<td>PET/CT improves GTV delineation</td>
</tr>
<tr>
<td>Bradley et al. (2004)</td>
<td>26</td>
<td>PTV: 14 of 24 (58%)</td>
<td>11 of 24 (46%); 10 lymph nodes, 1 tumour</td>
<td>3 of 24 (12%); tumour vs atelectasis</td>
<td>PET improves diagnosis of lymph nodes and atelectasis</td>
</tr>
</tbody>
</table>
172

A.-L. Grosu et al.

staging. The PET may help to differentiate between viable tumour tissue and associated atelectasis; however, it shows limitations when secondary inflammation is present. As GTV delineation is a fundamental step in radiation treatment planning, the use of combined CT/PET imaging for all dose escalation studies with either conformal radiotherapy or IMRT has been recommended by RTOG as a standard method in lung cancer (Chapman et al. 2003).

13.4.1.2
Head and Neck Cancer

A significant number of studies have demonstrated that FDG-PET can be superior to CT and MRI in the detection of lymph nodes metastases, the identification of unknown primary cancer and in the detection of viable tumour tissue after treatment.

Rahm et al. (1998) studied 34 patients with histologically confirmed squamous cell carcinoma of the head and neck by performing FDG-PET prior to treatment planning in addition to conventional staging procedures, including CT, MRI and ultrasound. The integration of FDG-PET in radiation treatment planning led to significant changes in the radiation field and dose in 9 of 22 (44%) patients with a primary tumour and in 7 of 12 (58%) patients with tumour recurrence. This was due mainly to the inclusion of lymph nodes metastases detected by FDG-PET. In a recent study, Nishioka et al. (2002) showed that the integration of FDG-PET in radiation treatment planning might also cause a reduction in the size of the radiation fields. The GTV for primary tumour was not altered by image fusion in 19 of 21 (90%) patients. Of the 9 patients with nasopharynx cancer, the GTV was enlarged by 49% in only 1 patient and decreased by 45% in 1 patient. In 15 of 21 (71%) patients the tumour-free FDG-PET detection allowed normal tissue to be spared. Mainly, parotid glands were spared and, thus, xerostomia was avoided. The authors concluded that the image fusion between FDG-PET and MRI/CT was useful in GTV, CTV and PTV determination, both for encompassing the whole tumour area in the irradiation field and for sparing of normal tissue. The integrated PET/CT was used for GTV and PTV definition in 12 patients with head and neck tumours and compared with CT alone. The GTV increased by 25% or more due to FDG-PET in 17% of these cases and was reduced by 25 in 33% of the patients. The corresponding change in PTV was approximately 20%; however, this study did not integrate MRI in the analysis, which has a higher sensitivity than CT (Ciernik et al. 2003).

In conclusion, the value of FDG-PET for radiation treatment planning is still under investigation. In some cases, FDG-PET visualised tumour infiltration better than CT or MRI alone. The FDG-PET could also play an important role in the definition of the boost volume for radiation therapy; however, the relatively high FDG uptake in inflammation areas could sometimes lead to false-positive results.

13.4.1.3
Prostate Cancer

Choline and citrate metabolism within cytosol and the extracellular space were investigated in prostate cancer using H-MRS. Clinical trials analysing tumour location and extent within the prostate, extra-capsular spread and cancer aggressiveness in pre-prostatectomy patients have indicated that the metabolic information provided by H-MRS combined with the anatomical information provided by MRI can significantly improve the tumour diagnosis and the outline of tumour extension (Mizowaki et al. 2002; Mueller-Lisse et al. 2001; Wefer et al. 2000).

Coakley et al. (2002) evaluated endorectal MRI and 3D MRS in 37 patients before prostatectomy and correlated the tumour volumes measured with MRI and H-MRS with the true tumour volume measured after prostatectomy. Measurements of tumour volume with MRI, MRS and a combination of both were all positively correlated with histopathological volume (Pearson's correlation coefficient of 0.49, 0.59 and 0.55, respectively), but only measurements with 3D MRS and a combination of MRI and MRS were statistically significant (p<0.05).

The integration of H-MRS in brachytherapy treatment planning for target volume definition in patients with organ-confined but aggressive prostatic cancer could improve the tumour control probability (Zaider et al. 2000).

13.4.1.4
Brain Gliomas

Although only preliminary data are available, the presented literature indicates that amino-acid PET, SPECT and H-MRS, in addition to conventional morphological imaging, are superior to the exclusive use of either MRI or CT in the visualisation of vital tumour extension in gliomas (Table 13.2).

Our group investigated the value of amino-acid PET and SPECT in GTV, PTV and boost volume (BV) definition for radiation treatment planning of brain gliomas. We demonstrated that I-123-alpha-methyltyrosine (IMT)-SPECT and L-(methyl-11C) methionine (MET)-PET offer significant additional infor-
mation concerning tumour extension in high-grade gliomas, compared with anatomical imaging (CT and MRI) alone (Grosu et al. 2000, 2002, 2003, 2005a,c). The MRS studies led to similar results (Pirzkall et al. 2001, 2004). A current analysis of an amino-acid SPECT- or PET-planned subgroup of patients with recurrent gliomas suggests that the integration of amino-acid PET or SPECT in target volume definition might contribute to an improved outcome (Fig. 13.2; Grosu et al. 2005b).

13.4.2 Which Relevant Biological Properties of the Tumour Could Represent an Appropriate Biological Target for Radiation Therapy?

Tumour hypoxia (Fig. 13.3) is an unfavourable prognostic indicator in cancer as it can be linked to aggressive growth, metastasis and poor response to treatment (Molls and Vaupel 2000; Molls 2001). In a clinical pilot study, Chao et al. (2001) demonstrated the feasibility of [60Cu]ATSM-PET guided radiotherapy planning in head and neck cancer patients; however, the tumour-to-background ratio in hypoxic tumour tissues did not dramatically differ from previous studies using [18F]FMISO and other nitroimidazole compounds. The authors reported the integration of the hypoxia tracer Cu-ATSM in radiotherapy treatment planning for patients with locally advanced head and neck cancer treated with IMRT. They developed a CT/PET fusion image based on external markers. The GTV outline was based on radiological and PET findings. Within the GTV, regions with a Cu-ATSM uptake twice that of the contralateral normal neck muscle were selected and outlined as hypoxic GTV (hGTV). The IMRT plan delivered 80 Gy in 35 fractions to the ATSM-enriched tumour subvolume (hGTV) and the GTV received simultaneously 70 Gy in 35 fractions, keeping the radiation dose to the parotid glands below 30 Gy.

More recently, Dehdashti et al. (2003) were the first to demonstrate a negative predictive value of enhanced Cu-ATSM uptake for the response to treatment in 14 cervical cancer patients. Rischin et al. (2001) used 18F-misonidazole scans to detect hypoxia in patients with T3/4, N2/3 head and neck tumours treated with tirapazamine, cisplatin and radiation therapy. Fourteen of 15 patients were hypoxia positive at the beginning of the treatment, but only one patient had detectable hypoxia at the end of radiochemotherapy.

By imaging either hypoxia (Rischin et al. 2001) with 18F-misonidazole or [60Cu]ATSM (Chao et al. 2001), angiogenesis with 18F-labelled RDG-containing glycopeptide and PET (Haubner et al. 2001), proliferation with fluorine-labelled thymidine analogue 3'-deoxy-3'-[18F]-fluorothymidine (FLT) and PET (Wagner et al. 2003), or apoptosis with a radio-labelled recombinant Annexin V and SPECT (Belhocine et al. 2004), different areas within a tumour can be identified and individually targeted. This approach, closely related to the IMRT technique, has been named “dose painting”. The IMRT combined with a treatment plan based on biological imaging could be used for individualised, i.e. customised, radiation therapy. The biological process visualised by the tracer needs to be specified; therefore, the selected BTV should be named after the respective tracer (i.e. BTV[FDG-PET], BTV[FAZA-PET]; Grosu et al. 2005c).

This novel treatment approach will generate a new set of problems and questions such as: What are the dynamics of the visualised biological processes? How

---

**Table 13.2 Impact of biological imaging: PET, SPECT and MRS in target volume delineation of gliomas. High G high-grade gliomas**

<table>
<thead>
<tr>
<th>Reference</th>
<th>No. of patients</th>
<th>Diagnosis</th>
<th>Biological imaging</th>
<th>Additional information to MRI and CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voges (1997)</td>
<td>46</td>
<td>Low+high G</td>
<td>MET-PET, FDG-PET</td>
<td>Yes</td>
</tr>
<tr>
<td>Julow (2000)</td>
<td>13</td>
<td>High G</td>
<td>MET-PET, SPECT ?</td>
<td>Yes</td>
</tr>
<tr>
<td>Gross (1998)</td>
<td>18</td>
<td>High G</td>
<td>FDG-PET</td>
<td>No</td>
</tr>
<tr>
<td>Nuutinen (2000)</td>
<td>11</td>
<td>Low G</td>
<td>MET-PET</td>
<td>Yes</td>
</tr>
<tr>
<td>Grosu (2000)</td>
<td>30</td>
<td>Low+high G</td>
<td>IMT-SPECT</td>
<td>Yes</td>
</tr>
<tr>
<td>Graves (2000)</td>
<td>36</td>
<td>High G</td>
<td>H-MRS</td>
<td>Yes</td>
</tr>
<tr>
<td>Pirzkall (2000)</td>
<td>30</td>
<td>High G</td>
<td>H-MRS</td>
<td>Yes</td>
</tr>
<tr>
<td>Pirzkall (2004)</td>
<td>30</td>
<td>High G</td>
<td>H-MRS</td>
<td>Yes</td>
</tr>
<tr>
<td>Grosu (2002)</td>
<td>66</td>
<td>High G</td>
<td>IMT-SPECT</td>
<td>Yes</td>
</tr>
<tr>
<td>Grosu (2005)</td>
<td>44</td>
<td>High G</td>
<td>MET-PET</td>
<td>Yes</td>
</tr>
</tbody>
</table>
many biological investigations are necessary during treatment and when? Which radiation treatment schedules have to be applied?

Prospective clinical trials and experimental studies should supply the answers to these questions.

13.4.3
What is the Biological Tumour Response to Therapy?

Biological imaging could be used to evaluate the response of a tumour to different therapeutic interventions. The usefulness of PET for monitoring patients treated with chemotherapy has been documented in several studies. Although evaluating the response after radiochemotherapy is often difficult due to treatment-induced inflammatory tissue changes, preliminary data for lung (Weber et al. 2003; MacManus et al. 2003; Choi 2002), oesophageal (Flamen et al. 2002) and cervical cancer (Grigsby et al. 2003) suggest that the decrease of FDG uptake after treatment correlates with histological tumour remission and longer survival; however, it still has to be clarified at which time points PET imaging should be performed, which tracer is to be used and which criteria can be used for the definition of a tumour response in PET. Theoretically, the integration of biological imaging in treatment monitoring could have significant consequences for future radiation treatment planning.
Definition of Target Volume and Organs at Risk. Biological Target Volume

A radiation oncologist has to outline the GTV, CTV, ITV and PTV and BTV. In this process, a lot of medical and technological aspects have to be considered. The criteria for GTV, CTV, etc. definition are often not exactly standardised, and this leads, in many cases, to variability between clinicians; however, exactly defined imaging criteria, imaging with high sensitivity and specificity for tumour tissue and special training could lead to a higher consensus in target volume delineation and, consequently, to lower differences between clinicians. It must be emphasised, however, that further verification studies and cost-benefit analyses are needed before biological target definition can become a stably integrated part of target volume definition.

The ICRU report 50 from 1993 and the ICRU report 62 from 1999 defining the anatomically based terms CTV, GTV and PTV still have to be considered as the gold standard in radiation treatment planning; however, further advances in technology concerning volume delineation during radiation therapy; however, this still has to be analysed in clinical studies.

In summary, biological imaging allows the visualisation of fundamental biological processes in malignancies. To date, the suggested superiority of biological strategies for treatment planning has not yet been sufficiently demonstrated; therefore, before the BTV definition can be generally recommended, further clinical studies based on integrated PET, SPECT or MRS need to be conducted. These studies must compare the outcome of biologically based treatment regimes with conventional treatment regimes.

**13.5 Conclusion**

Target volume definition is an interactive process. Based on radiological (and biological) imaging, the radiation oncologist has to outline the GTV, CTV, ITV and PTV and BTV. In this process, a lot of medical and technological aspects have to be considered. The criteria for GTV, CTV, etc. definition are often not exactly standardised, and this leads, in many cases, to variability between clinicians; however, exactly defined imaging criteria, imaging with high sensitivity and specificity for tumour tissue and special training could lead to a higher consensus in target volume delineation and, consequently, to lower differences between clinicians. It must be emphasised, however, that further verification studies and cost-benefit analyses are needed before biological target definition can become a stably integrated part of target volume definition.

The ICRU report 50 from 1993 and the ICRU report 62 from 1999 defining the anatomically based terms CTV, GTV and PTV must still be considered as the gold standard in radiation treatment planning; however, further advances in technology concerning
signal resolution and development of new tracers with higher sensitivity and specificity will induce a shift of paradigms away from the anatomically based target volume definition towards biologically based treatment strategies. New concepts and treatment strategies should be defined based on these new investigation methods, and the standards in radiation treatment planning – a continuous, evolutionary process – will have to integrate new imaging methods in an attempt to finally achieve the ultimate goal of cancer cure.

References

ICRU 50 (1993) Prescribing, recording and reporting photon beam therapy. ICRU report no. 50. ICRU, Bethesda, Maryland
ICRU 62 (1999) Prescribing, recording and reporting photon beam therapy. ICRU report no. 62 (supplement to ICRU report no. 50). ICRU, Bethesda, Maryland
Kiffer JD, Berlangieri SU, Scott AM et al. (1998) The contribution of 18F-fluoro-2-deoxyglucose positron emission...


14 Virtual Therapy Simulation

Rolf Bendl

CONTENTS

14.1 Introduction 179
14.2 Modelling 180
14.3 The Radiotherapy Planning Cycle 180
14.4 Graphical Tools for Virtual Simulation 180
14.4.1 Beam's Eye View: Selection of Irradiation Directions and Beam Portals 181
14.4.2 Observer’s View: Optimal Combination of Irradiation Directions 181
14.4.3 Multi-planar Reconstructions of Tomographic Image Series 183
14.4.4 Digital Reconstructed Radiographs 183
14.5 Tools for Evaluating Dose Distributions 184
14.5.1 Comparison of Treatment Plans 185
References 186

14.1 Introduction

Therapy simulation is widely used and has a long tradition in radiotherapy. A therapy simulator is an X-ray imaging device with the same geometric properties as the linear accelerator used for therapy. This device allows the acquisition of images with the radiation source located in the same position as during therapy. This way it is used to determine and verify patient setup. Furthermore, therapy simulators are used to determine irradiation directions.

Modern fast computer technology and dedicated software tools allow the simulation and examination of a large variety of problems of real life in a virtual reality. By means of those tools complex courses and tasks can be simulated and optimised in advance. It was an early insight in radiotherapy planning that those simulations can improve treatment planning and hopefully the treatment itself markedly; therefore, the core of all modern treatment planning systems are programs and algorithms which allow interactive simulation of the visible parts of the treatment, e.g. determination of irradiation directions and beam portals, as well as simulation of the physical dose deposition in the patient's tissue (Sherouse and Chaney 1991). Since the term “therapy simulation” has already been used, this kind of simulation is called “virtual therapy simulation”. Often “virtual therapy simulation” is also used as synonym for radiotherapy planning.

The goal of virtual therapy simulation is to find the optimal treatment plan for an individual patient by testing, evaluating and optimising alternative treatment strategies before treatment starts. In this chapter the most basic concepts and tools used for simulating the “visible” parts of treatment are explained. For more information about pre-calculation of dose distributions refer to Chap. 15.

The idea of conformal radiotherapy is to concentrate the therapeutic dose precisely to the shape of the tumour and to shield surrounding healthy tissue as best as possible. This way side effects can be reduced, while simultaneously the opportunity to escalate dose in the target volume can be increased which can result in a better tumour control probability.

One of the two most important techniques to concentrate dose to target volume is the precise adjustment of beam shapes to the shape of the tumour, to spare surrounding organs at risk. The second technique is the determination of treatment techniques which consist of series of different beams from different directions. These beams will superimpose in the target volume and sum up there to the desired dose level, whereas dose in regions which are hit only by one or a small number of beams can be kept below tissue-dependent tolerance levels. Virtual simulators basically supply interactive graphical tools to facilitate those complex tasks.
For virtual simulation a three-dimensional model of the patient’s anatomy is required. This model is based on at least one or more three-dimensional image series. Two-dimensional tomograms can be combined to form such a three-dimensional “image cube”.

Since modern dose calculation algorithms refer to “Hounsfield” units to calculate the dose deposition, at least one series acquired with a CT scanner is necessary, but consideration of other image modalities (MRI, PET, SPECT) can help in precise delineation of anatomy and in outlining of the target volume. Since CT imaging is a robust technique and images are considered to be geometrically correct, they are of particular relevance and treatment planning starts with the acquisition of a planning CT series. To use information of image modalities simultaneously, it is necessary to establish well-defined relationships between the volume elements of the different image sequences. That means they must be registered with the planning CT series.

Since all subsequent steps of therapy planning, optimisation and evaluation are based on that three-dimensional patient model, the delineation of anatomy and the definition of the target volume is one of the most crucial steps in treatment planning. While that step is time-consuming and difficult, image segmentation is an inevitable pre-requisite or component of radiotherapy planning, since treatment plan selection is based on quantitative evaluations of dose distributions and information on how much dose is applied to organs at risk and the target volume. Without delineation of therapy-relevant structures it would not be possible to generate answers to those questions. While delineation of normal anatomy can be supported by various interactive, semi-automatic or automatic segmentation tools, definition of target volumes cannot be automated reasonably, since target volumes do not include only visible tumour regions but surrounding suspicious areas and security margins as well to compensate for organ movement and uncertainties in patient positioning. Therefore, the target volume spans over image regions with various visually perceptible types of tissue and therefore automatic algorithms considering only image intensities normally fail. For more information about registration and segmentation refer to Chaps. 3–5.

Based on segmentation results it is possible to create a three-dimensional model of the individual anatomy which can be presented and utilized in various views during treatment planning.

In conventional forward treatment planning the definition of treatment parameters is an iterative, trial-and-error process. The therapist starts with the definition of one or more alternative treatment strategies, then he has to start dose calculations based on the defined strategies. Subsequently, the expected dose distributions can be analysed and compared. If plans do not meet specific individual constraints, the treatment parameters are modified until an acceptable plan is found (Fig. 14.1).

Fig. 14.1. Radiotherapy planning cycle

To supply a virtual environment which allows the definition of all treatment parameters and the examination of consequences to the expected dose distributions, modern three-dimensional planning systems offer a variety of different graphical tools. Besides the availability of three-dimensional images, the interactivity of these scenes is of particular relevance. The possibility of getting direct feedback on how defined settings will influence the treatment plan plays an important role and enables users to generate highly individualized treatment plans in a short time. Due to the extensive increase of graphical performance of currently available computer hardware, it is possible to display very complex scenes without any delay and this way the necessary interactivity can be assured.
14.4.1
Beam’s Eye View: Selection of Irradiation Directions and Beam Portals

The most important tool for selecting appropriate irradiation directions and for defining beam portals is the beam’s eye view (Fig. 14.2; Goitein et al. 1983). In the beam’s eye view the three-dimensional model of the patient’s anatomy is presented from the position of the radiation source. In regarding this scene it immediately becomes clear which structures are enclosed by the current beam, since the positions of the beam-limiting devices are integrated, too; therefore, an interactive beam’s eye view is the best tool to select appropriate irradiation directions.

Subsequently, the beam’s eye view can be used to adjust the shape of the beam to the shape of the target volume. Depending on the used radiation type, several possibilities for transcription of these definitions should be available. For example, irregular beam shaping in conformal radiotherapy with photons normally is done by using multi-leaf collimators; therefore, the defined leaf positions should be integrated in the beam’s eye view, too. On using electrons beam shaping is mostly done by applicators or individual blocks. Necessary tools for defining and displaying these components should also be available.

14.4.2
Observer’s View: Optimal Combination of Irradiation Directions

The observer’s view (Fig. 14.3) was designed to help treatment planners in combining several beams applied from different irradiation directions. These three-dimensional scenes show the same patient model as the beam’s eye view from an arbitrary point of view (Bendel et al. 1993). In addition, the shapes of all defined beams are integrated, since one of the most important criteria in combining beams is to minimize that sub-volume where the single beams overlap. In this scene the degree of overlap can easily be perceived. In addition, the observer’s view gives a fast impression of the complete treatment strategy and is therefore a good base for information exchange between the involved staff. Last, but not least, is the observer’s view a valuable tool for qualitative evaluation of dose distributions. By means of integrated dose information the overall quality of treatment plans can be examined and hot spots on organs at risk or cold spots on the target volume can be detected easily.

During definition of irradiation directions the therapist must consider whether the selected directions can be transferred to the existing treatment

Fig. 14.2. Beam’s eye view, three-dimensional patient model. Red: target volume; green: brain stem; violet: eyeballs. Left: a conventional rectangular beam portal which encloses a large volume of healthy tissue. Right: with multi-leaf collimators it is possible to form irregular-shaped portals and to spare healthy tissue.
unit. This aspect is of particular relevance, especially when using non-coplanar irradiation directions. Depending on the accelerator specifications and the selected accessories (multi-leaf collimators, applicators, etc.) it might be possible that some desired directions cannot be realised because the selected gantry and couch angles would lead to a collision of gantry and couch or even with the patient. These “collision zones” depend on the position of the target point in the patient and cannot be determined universally. Since the patient model consists only of that part of the patient viewed by the scanner during image acquisition, it is not possible to establish a reliable automatic collision detection procedure without supplementary information. An approved method is therefore the presentation of a three-dimensional scene of the treatment unit where the patient model is integrated correctly. In this scene, all degrees of freedom of the linear accelerator (Linac; Fig. 14.4) and Couch can be inspected visually. Since therapists know their patients, and are normally able to estimate the extension of a patient quite well, they can determine visually which directions can be used without any problems and which directions must be verified with the patient at the accelerator prior to the first treatment fraction.

Fig. 14.3a,b. Observer’s view. a The sub-volume where all three beams overlap can be identified easily. b Display of a 14 technique with 14 individually shaped beams, designed for stereotactic single dose therapy

Fig. 14.4. Linac view: presentation of treatment situation
14.4.3
Multi-planar Reconstructions of Tomographic Image Series

Besides artificial three-dimensional scenes, it is necessary to project selected irradiation directions and beam shapes onto the slices of the planning CT, too, to allow the examination of selected parameters on the originally acquired patient data. Of course, it is helpful to have not only projections on the initially acquired transversal slices but at least on multi-planar reconstructions in the other two main directions (sagittal, coronal). The selection of oblique sections might be helpful during examination of dose distributions. Sometimes an additional gain of information can be achieved by projecting beams on images of other modalities, but this is usually considered to be of minor priority.

14.4.4
Digital Reconstructed Radiographs

Generation and presentation of digital reconstructed radiographs (DRRs) is a virtual simulation of real treatment simulators or portal imaging devices. Based on the information in the acquired CT image series it is possible to calculate artificial X-Ray images from arbitrary directions comparable to those mentioned above (Galvin et al. 1995).

During X-Ray image acquisition, X-Rays are absorbed differentially according to tissue density. Structures with high density (bones) absorb more intensity of the ray than soft tissue, which means that the incident radiation on the film after having passed bony structures is low, resulting in a low optical density on the film (bright). Water, fat muscle tissue or air absorb only a small amount (or none) of the radiation. Rays running only through soft tissue keep a higher intensity which results in a high optical density on the film (dark or black).

To simulate these behaviours ray-tracing algorithms are used (see also Chap. 4.4.2). Starting from the position of the X-ray source, the algorithm traces a bundle of beams through an image cube, until a specified projection plane (i.e. the film position) is reached (Siddon 1985). Each ray passes through different elements of the data cube (volume elements or voxel). Depending on how the values of the voxels passed through contribute to the final value of the ray in the imaging plane, different kinds of images can be produced (Fig. 14.5).

If the absorption of a ray’s intensity is considered to be proportional to the density of a given voxel, the total absorption is proportional to the sum of the voxel densities along the ray. To generate DRRs, the voxel intensities must simply be summed (Fig. 14.5a). If an algorithm considers only the maximum voxel values along a ray and projects that value onto the image plane, maximum intensity projections (MIPs) can be generated. Applying such an algorithm to ap-

![Fig. 14.5a–c. Principle of ray tracing. a Digital reconstructed radiographs. b Surface rendering. c Maximum intensity projection](image-url)
propitious image cubes (CT or MR angiographies), the blood vessel system can be visualised (Fig. 14.5b). Adding some constraints to the ray-tracing algorithm, e.g. the ray tracing should finish when a voxel with a particular intensity arrives, the surface of anatomical structures can be detected and visualized. As a prerequisite, the intensity of those structures must have good contrast with respect to the surrounding tissue. In this way, the patient’s surface and bony structures can easily be visualised (Fig. 14.5c). Images generated with surface-rendering methods from the position of the radiation source show the same geometrical relations as the beam’s eye view images explained above.

For radiotherapy, DRRs and MIPs are especially important. The MIPs can be used for detecting malformations of blood vessels (AVMs). The DRRs can help during patient positioning and positioning control. Based on the specified treatment parameters and the known geometry of the X-ray imaging systems, DRRs can be pre-calculated. To verify patient positioning X-ray images can be acquired prior or during treatment. They can be compared to the calculated DRRs, deviations between both images can be calculated automatically and it is possible to compensate for displacements by correcting the treatment position to the planned one.

On comparing DRRs with portal images generated with the beam of the treatment unit it is necessary to take care of different imaging energies. The CT images are normally acquired with an imaging energy in the lower kV range, the energy used for irradiation of photons is in the range of 6–20 MV. Mega-voltage images usually show a much lower contrast, and it is difficult to perceive soft tissue structures. Due to those differences, a comparison of artificial images with verification images is often difficult, too. A solution could be the consideration of different mean image energies (Mohan and Chen 1985) already on generation of the DRRs.

Nonetheless, the possibility of examination and evaluation of dose distributions should be an integral part of a virtual simulator, since only the resulting dose distribution can answer the question, whether the selected treatment strategy considers all individual constraints sufficiently.

There is a large variety on methods to display dose information. By integrating dose distributions into the three-dimensional patient model, presented in the observer’s view, a fast impression about the global dose distribution can be achieved. This integration can be done with different techniques. Figure 14.6 shows two examples. On the left side dose information is integrated as iso-dose ribbons, which means that sub-volume which receives a dose higher than a given level is enclosed by some yellow ribbons. This way a planner can immediately verify whether the target volume is enclosed completely by the therapeutic dose level. On the right side, the surface dose is displayed. That means the surfaces of all defined structures are coloured according to the expected dose. Black and blue areas do not receive any or only a small amount of dose; in red areas the dose is equal to the desired therapeutic dose level. This way, hot spots on organs at risk (here, for example, on the chiasm and on the right optical nerve) or cold spots in the target volume (here in the topmost region) can be detected very easily.

Of course variations of these techniques are common, too. For example, some treatment planning systems offer the possibility to display dose clouds as semi-transparent surfaces or are using more advanced display techniques such as fog.

While those 3D scenes give a very fast impression of dose distributions, they are not sufficient to get precise information about the dose distribution inside of an organ at risk or in the target volume, since the patient model is reduced to the surface of anatomical structures and dose clouds. Therefore, most of the available radiotherapy simulators and planning systems supply possibilities to project dose information onto the initially acquired images. Again, various different techniques are possible. The most common techniques are the display of dose distributions as iso-dose lines, and the integration of dose information using colour-wash techniques (see Fig. 14.8: comparison of concurrent treatment plans.). An important feature is the ability to present dose not only on the original CT image slices but on multi-planar reconstructions and oblique sections, too. The possibility to project dose not only on the planning CT but also on other available series of other imaging modalities is desirable.

14.5 Tools for Evaluating Dose Distributions

An optimal virtual simulator would not only present the result of geometrical changes (beam directions, shapes, etc.) immediately, but also the consequences on the expected dose distribution in real-time. Admittedly, the immediate feedback can be assured only on the geometrical aspects of treatment parameters. Up to now it is still not possible to display the consequences of changed settings on dose distributions in real-time on standard computer hardware.
Besides those qualitative evaluation tools, a quantitative analysis of dose distributions should be possible, since the complex shape of three-dimensional dose distributions makes it difficult to compare concurrent plans. Calculation of quantitative parameters enables objective measurements, allows the perception of under- and overdosed regions and supports the evaluation of the dose homogeneity.

Dose-volume histograms (DVHs; Fig. 14.7) are a simple and the most accepted way of displaying information about three-dimensional dose distributions. The DVHs are usually displayed as cumulative histograms, showing the fraction of the total volume of a particular structure receiving doses up to a given value (Chen et al. 1987; Drzymala et al. 1991). Since the information of the complex three-dimensional dose distribution is reduced to a group of simple two-dimensional graphs, comparison of concurrent plans is facilitated. The loss of spatial information can be compensated for by qualitative display techniques mentioned above.

The examination of additional statistical parameters, such as minimum, maximum and mean dose, variance, and number of voxels below or above certain levels can give additional information.

### 14.5.1 Comparison of Treatment Plans

During the radiotherapy planning cycle not only tools for defining a single treatment plan are needed, but tools for comparing alternative treatment strategies are desirable as well. During the iterative enhancement of treatment plans differences in dose distributions decrease, and often it is difficult to recognize resulting changes in dose distributions; therefore, a planning system or a virtual simulator should offer possibilities to compare dose distributions of concurrent treatment plans and special display techniques to enhance small deviations in dose distributions. Figure 14.8 shows an example how differences in dose distributions can be presented and emphasised. In the upper row a dose distribution is shown...
generated based on beam configuration displayed in Fig. 14.8, left. In the second row a dose distribution is displayed of a plan where a fourth beam was added. The third row presents the difference dose. Regions where the first one shows a higher dose are coloured in red and yellow, and those regions where the second plan lets to higher doses are coloured in blue and green. Besides these qualitative comparisons, methods for quantitative analysis, e.g. simultaneous display of DVHs and generation of difference DVHs, are favourable.

References


Fig. 14.8 Comparison of concurrent treatment plans. Upper row: colour-wash display of dose distribution of plan based on beam configuration of Fig. 14.3a. Centre row: dose of an alternative plan with four beams. Bottom row: display of difference dose
15 Dose Calculation Algorithms

Uwe Oelfke and Christian Scholz

15.1 Introduction

The accurate and fast calculation of a 3D dose distribution within the patient is one of the most central procedures in modern radiation oncology. It creates the only reliable and verifiable link between the chosen treatment parameters, and the observed clinical outcome for a specified treatment technique, i.e., the prescribed dose level for the tumor, the number of therapeutic beams, their angles of incidence, and a set of intensity amplitudes – obtained by a careful treatment plan optimization – result in a distribution of absorbed dose which is the primary physical quantity available for an analysis of the achieved clinical effects of this specific treatment. The twofold application of dose calculation algorithms in radiation oncology practice, firstly for the plan optimization in the treatment planning process, and secondly for the retrospective analysis of the correlation between treatment parameters and clinical outcome, defines two mutually conflicting goals of the respective dose algorithms. Firstly, the dose calculation has to be fast such that the treatment planning process can be completed in clinically acceptable time frames, and secondly, the result of the dose calculation has to be sufficiently accurate so that the establishment of correlations between delivered dose and clinical effects remains reliable and meaningful. The conflict between “high speed” and “high accuracy” is one of the crucial challenges for the development of modern dose calculation algorithms.

The accuracy of the dose calculation algorithms becomes a problem only for very heterogeneous tissues, where a very detailed modeling of the energy transport in the patient is required. The prediction of the dose around air cavities, e.g., often encountered for tumors adjacent to the paranasal sinuses or for solid tumors embedded in lung tissue, is intricate and time-consuming. Almost all new developments related to dose algorithms are specifically concentrating on these or equivalent areas of tissue heterogeneities, whereas for the majority of clinical cases with almost homogeneous tissues existing, simple calculation methods can be reliably applied.

Naturally, dose algorithms for high-energy photon beams were first developed for the ultimate “homogeneous” patient – a patient completely consisting of water (Schoknecht 1967). Measurements of a set of generic dose functions, e.g., tissue air ratios, tissue phantom ratios, output factors, and off-axis ratios, are measured in a water phantom for a set of regular treatment fields under reference conditions. The dose within a patient is then calculated by extrapolating these measurements to the specific chosen treatment fields and by the application of various correction algorithms, e.g., for the inclusion of missing tissues at the patient surface or the approximate consideration

U. Oelfke, PhD
Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
C. Scholz, PhD
Siemens Medical Solutions, Oncology Care Systems, Hans-Bunter-Str. 19, 69123 Heidelberg, Germany
of tissue heterogeneities. These phenomenological “correction-based methods,” which rely almost completely on a set of measurements, are very fast and have the further advantage of not needing to distinguish between the radiation field provided by the linear accelerator – often labeled as the phase space of the respective linac – and the subsequent energy transport by photons and electrons in the patient (see Fig. 15.1).

Model-based algorithms in their various implementations constitute the standard algorithms provided by currently commercially available treatment planning systems. The simplest form, the so-called pencil-beam algorithm, is still the standard and fastest dose engine (Hogstrom et al. 1981; Schoknecht and Khatib 1982; Mackie et al. 1985; Mohan et al. 1986; Bortfeld et al. 1993). More sophisticated and accurate are the superposition algorithms (Mackie et al. 1985; Mohan et al. 1986; Ahnesjö et al. 1987; Mackie et al. 1988; Ahnesjö 1989; Scholz et al. 2003b), which are also becoming more widespread; however, as mentioned above, model-based algorithms still rely on approximations and only partly describe the physical processes involved in the microscopic absorption of the energy delivered by the radiation field. The most sophisticated approach to include almost all known physical features about the microscopic radiation–tissue interactions is the Monte Carlo approach. Usually, a Monte Carlo dose calculation consists of two independent components: (a) the simulation of the already mentioned “linac phase space” based on the geometrical design of the treatment head for the considered machine and a few characteristic parameters for the electron beam before it impinges on the “bremsstrahlungstarget” of the linac; the respective radiation field serves as input for (b) the simulation of the energy absorption and transport within the patient; however, the implementation of these sophisticated algorithms and their verification is non-trivial and often relies on expert knowledge. A more detailed description of the involved principles is provided in section 4.1.4.

In the following section we focus on the description of the most prominent dose calculation models implemented in commercially available treatment planning systems, i.e., the concepts of pencil-beam and superposition algorithms is presented. Furthermore, the application of these models for different phantoms and clinical cases, especially for the optimization of IMRT cases with significant tissue inhomogeneities, is discussed.

15.2 Model-Based Algorithms

In this section we briefly review the most prominent dose calculation concepts which include some explicit models for the relevant energy transport in the patient. The absorption of energy in the patient, i.e., the accumulation of dose, for these methods is sepa-
rated into several steps. Before the energy absorption process is considered itself, one has to model the radiation output of the considered treatment machine. This is often accomplished by a simple model for the “primary energy fluence” of the photons emerging from the linear accelerator. The respective models are calibrated to measured dose data for simple treatment fields in water and are therefore not independent of the models employed for the actual energy absorption physics. The determined energy fluence of the primary photons serves as input for the subsequent calculation of the energy absorption and transport within the patient. Firstly, the absorption of the primary photons is considered and expressed by the “total energy released per unit mass” (TERMA; Ahnesjö et al. 1987). Then, the transport of this energy via secondary electrons and photons is accounted for by the introduction of specific “dose kernels.” The three mentioned components – primary photon fluence, TERMA, and dose kernels – are discussed below and are then employed in various forms for dose calculations in homogeneous and inhomogeneous media.

15.3 Modeling of the Primary Photon Fluence

The radiation field delivered by a linear accelerator is a very complex mixture of primary photons, scattered photons, and electrons, of which the following physical properties have to be known in principle for an accurate dose calculation: the particle’s energy spectrum, their spectrum of velocity directions, and their lateral distribution or fluence (number of particles per area) in a defined plane perpendicular to the central beam axis. The simplest approximation of this phase space of the linear accelerator (see Fig. 15.1) is provided by modeling an effective fluence of primary photons, based on a few phenomenological parameters, which are calibrated to simple measurements.

A common assumption of the respective models is that the energy spectrum of the primary photons factorizes with the remaining phase space, i.e., the energy spectrum is independent of the lateral location of the photons with respect to the beam axis. There are two basic methods how these energy spectra are derived for clinical practice. Both of them require to different extent a calibration to measured dose data in water. Firstly, there is the approach to calculate the whole phase space of the emerging primary photon beam via Monte Carlo simulations (see Chap. 16; Mohan and Chui 1985; Rogers et al. 1995), i.e., the process of creating photons from a narrow electron beam impinging on the “bremsstrahlung” target of the linear accelerator and its subsequent attenuation and scattering in the treatment head is modeled from first physical principles. Provided that the geometrical location and physical properties of all relevant components in the treatment head are adequately known, these calculations depend only on a few parameters characterizing the initial electron beam of the linac, e.g., the average energy of the electrons, the variance of the energy spectrum, and the angular divergence of the electron beam. The complete phase space of the primary particles can then be used to derive an average energy spectrum of the respective photons and electrons. More details about these procedures are given in Chap. 16.

A more practical-motivated method relies on a direct comparison of measured depth dose data and an energy weighted sum of pre-calculated depth dose curves (Scholz et al. 2003b; Ahnesjö and Aspradakis 1999). Each energy bin $E_i$ of the actual energy spectrum is assumed to contribute a depth dose $D(E_i, z)$ to the measured values $D_{\text{spectrum}}(z)$ of a simple treatment field. The elementary, mono-energetic depth dose curves $D(E_i, z)$ are, for instance, also derived by Monte Carlo simulations. Mostly, a simple least-squares analysis is employed to derive weighting factors $\alpha_i$ such that $D_{\text{spectrum}}(z) = \sum_i \alpha_i D(E_i, z)$, i.e., an energy spectrum is generated that reproduces the measured data (Scholz et al. 2003b; Altschuler et al. 1992). Admittedly, this phenomenological procedure only selects one possible energy spectrum of the photon beam which is compatible with the given experimental depth dose data. Moreover, it only accounts implicitly for any electron contamination of the photon beam. As an example, we show an effective energy spectrum for the primary photons delivered by a Siemens Primus in Fig. 15.2.

In addition to the energy spectrum, the spatial distribution of the primary fluence also has to be modeled for the application in model-based algorithms. At least the following effects should be accounted for by the overall algorithm: (a) the spatial shape of the primary fluence distribution including so-called beam horns; (b) the broadening of the lateral penumbra through a spatially extended photon source; (c) the collimator scattering as described by the collimator scatter factor (CSF); and (d) the attenuation of the fluence inside the medium. As starting point for the modeling of the primary fluence one often uses the measured fluence in air for the largest aperture of interest (see Fig. 15.3). For the respective modification of this initial fluence, required by the aspects (b)–(d),
15.4 Dose Calculation in Homogeneous Media

15.4.1 TERMA

We first consider the dose deposition of a mono-energetic, infinitely narrow photon beam in z-direction of energy \( E \) and initial fluence \( \Phi \) in a homogeneous medium, which is naturally chosen to be water for all applications in radiotherapy. The energy fluence \( \Psi \) of the primary photons is to a first approximation determined by the linear photon absorption coefficient \( \mu(E) \) in water, i.e., at the interaction point of the primary photons \( \bar{r} \) one gets:

\[
\Psi(\bar{r}) = \Phi(\bar{r}, 0) E e^{-\mu(E) z}
\]

where \( \bar{r} \) denotes the coordinates perpendicular to the beam direction. The rate of the primary photon interactions in the medium determines the TERMA \( T(\bar{r}) \), i.e., the total energy per unit mass released by a radiation field interacting with a medium of density \( \rho \) at a certain point \( \bar{r} \):

\[
T(\bar{r}) = \frac{\mu}{\rho} \Psi(\bar{r})
\]

This locally released energy of the radiation field is subsequently available for a further transport emerging from the interaction point \( \bar{r} \), which is usually described by the concept of dose kernels.

15.4.2 Dose Kernels: Point-Spread Kernel and Pencil Beam

It is common to use two elementary dose kernels for model-based algorithms (Mohan et al. 1986; Bortfeld et al. 1993; Mackie et al. 1988; Ahnesjö 1989). The most elementary kernel \( k(\bar{r}, \bar{r}', E) \), the so-called point-spread kernel, indicates the distribution of absorbed energy in water at the coordinate \( \bar{r} \) which is created by interactions of primary photons of energy \( E \) at the coordinate \( \bar{r}' \) (see Fig. 15.4). The elemental mono-energetic dose deposition kernels can be taken from Monte Carlo simulations, e.g., from Mackie et al. (1988).

The second class of dose kernels and most commonly used in current treatment planning systems is the pencil beam. A pencil-beam kernel is obtained through the integration of all point-spread kernels along an infinite ray of photons in the medium as indicated in Fig. 15.5. It is evident that the pencil-beam kernel uses the more condensed information about the dose in water along the central kernel axis, i.e., it provides a coarser sampling of the physical processes than the point-spread kernel and it is therefore harder to adapt the dose calculations based on pencil-beam kernels to regions with intricate tissue inhomogeneities. On the other hand, pencil-beam kernels bear the obvious advantage of reduced dose calculation times. The first pencil-beam-type dose calcula-
Dose Calculation Algorithms

15.4.3 Superposition and Convolution Algorithms

Finally, the two components of TERMA and dose kernels are combined to achieve the accurate calculation of absorbed dose. The most general approach of model-based dose calculation algorithms is the superposition method (Mackie et al. 1985; Ahnesjö 1989; Scholz et al. 2003b). It generates the dose delivered at a point \( \vec{r} \) by superimposing the dose contributions from all dose kernels \( k(\vec{r}, \vec{r}', E) \) of the defined energy spectrum originating from all primary interaction points \( \vec{r}' \) and weighs their contributions with the respective TERMA, i.e.,

\[
D(\vec{r}) = \int dE' \int d\vec{r}' \ T(\vec{r}, E') \ k(\vec{r}, \vec{r}', E').
\]

The superposition approach certainly is a too sophisticated method to be applied for a dose calculation in homogenous media, but it can be used very well for dose calculations in regions of interest with tissue inhomogeneities.

A reduction of the computational effort required for the superposition approach, is achieved, if one assumes that the shape of the dose kernel is translational invariant. Then, the kernel \( k(\vec{r}, \vec{r}', E) \) is only a function of the distance between the interaction \( \vec{r}' \) point and the coordinate \( \vec{r} \) where the dose is measured such that the superposition formula reduces to the well known convolution approach, i.e.,

\[
D(\vec{r}) = \int dE' \int d\vec{r}' \ T(\vec{r}, E') \ k(\vec{r} - \vec{r}', E').
\]

The calculation times of the convolution algorithm are reduced dramatically compared with the more accurate superposition methods.

The application of a pencil-beam kernel, usually employed with convolution algorithms, leads to a further substantial reduction of calculation times, e.g., with a simple single value decomposition of the kernel the dose calculation can be reduced to a few 2D convolutions (Bortfeld et al. 1993), so that a 3D dose calculation for a conventional treatment plan can be accomplished in seconds.

15.5 Accounting for Tissue Inhomogeneities

In order to calculate the dose in regions with tissue inhomogeneities, the dose calculation has to account for the variations of electron densities derived from...
CT scans. The electron densities influence the dose calculation in two aspects:

1. The local TERMA depends on the path of the primary photons through the patient to their interaction point.
2. The energy distribution around the primary interaction point described by the dose kernels is also influenced by variations of the respective electron densities. The accurate calculation of the TERMA is accomplished by ray tracing the pathway of a photon to its interaction point. For this process the absorption rate of the photons along a considered trajectory is scaled by the ratio of the electron densities of the encountered media to the electron density of water. More important and more difficult to deal with is the influence of electron density variations on the dose kernels. This problem and some related practical aspects are briefly discussed below.

15.5.1 Pencil Beam and Pathlength Scaling

The pencil beam is a dose kernel describing the 3D dose distribution of an infinitely narrow mono-energetic photon beam in water. The individual interaction points of the photons are all assumed to be on the central axis of the pencil beam. Here tissue inhomogeneities are accounted for by the same pathlength scaling with relative electron densities between tissue and water as applied for the calculation of the TERMA. The values of the pencil-beam kernel in water are used according to the radiological pathlength calculated along the central axis of the pencil beam. Electron variations perpendicular to the pencil-beam axis are not accounted for, i.e., this inhomogeneity correction assumes a slab geometry of tissue inhomogeneities, which are represented by the values of the electron densities along the pencil-beam axis.

15.5.2 Density Scaling of Point-Spread Kernels

For the superposition algorithm the most accurate sampling of the primary interaction points within the patient is required. Consequently, this method also applies the most accurate technique of inhomogeneity corrections for the dose kernel. According to O’Connor's theorem, MOHAN et al (1986) devised a density scaling method which is applied directly to the point-spread kernel. Basically, it is assumed that the energy transport through the kernel from the interaction point $\mathbf{r}$ to the dose point $\mathbf{r}'$ occurs along a straight line, i.e., within the kernel one also introduces an “internal” ray tracing. Along each internal ray the contribution of the kernel is scaled with the average electron density encountered along the line connecting $\mathbf{r}$ and $\mathbf{r}'$. This leads to the so-called density-scaled dose deposition kernels.

As an example we show in Fig. 15.6 a hypothetical dose kernel in water together with a density-scaled kernel including some tissue inhomogeneities. It can be clearly seen that the kernel extends further from the interaction point if the internal energy transport encounters a medium of an electron smaller than water on its way to the dose deposition point. The dimensions of the scaled dose kernel are shrinking in comparison with the original dose kernel in water if higher electron densities represent an additional obstacle for the energy transport within the dose kernel.

15.5.3 Collapsed Cone and Kernel Tilting

The problem of the superposition approach are its long computation times, which may still take hours for a complicated IMRT case even if performed with state-of-the-art hardware technology (SCHULZE 1995); therefore, various approximations have been suggested for accelerating the dose calculation with the superposition technique. One method introduced by AHNESJÖ (1989) is the collapsed cone-beam tech-
nique which refers to a specific internal sampling of the dose kernels. Furthermore, it seems to be important that the axis of the point-spread kernels be aligned with the original photon rays employed for the determination of the TERMA within the patient (Ahnesjö 1989; Scholz et al. 2003b; Sharpe and Battista 1993). The neglect of the required kernel tilting saves considerable calculation time; however, it also can introduce significant dose errors (Sharpe and Battista 1993; Liu et al. 1997).

Another effect which might have to be considered is the hardening of the photon-energy spectrum. Works of Liu et al. (1997) and Metcalfe et al. (1990) showed that these effects are usually minimal in routine clinical practice.

15.5.4 A Simple Example: Doses at Cork–Water Interfaces

In order to demonstrate the sensitivity of the various types of dose algorithms to inhomogeneous media, we show the results of calculations performed for a simple phantom geometry. A 4-cm slab of cork was placed at a depth of 6 cm into a water phantom. The irradiation geometry of a 6-MV photon field was fixed as a 3×3 cm² open field with a source-to-skin distance SSD=95 cm. This field size was chosen since the measurements could still be performed reliably and because a significant effect on the dose patterns was anticipated. The 3D dose distributions for this simple geometry were calculated with three different algorithms: a pencil beam, a superposition, and a Monte Carlo dose calculation algorithm (Scholz 2004). All applied algorithms were proven to give equivalent results for 3D dose distributions in water for various regular and irregularly shaped fields.

Firstly, we show the result of the 2D lateral dose distributions on a central slice of the phantom in Fig. 15.7. It is clearly indicated in Fig. 15.7a that the pencil beam almost completely misses the effect of lateral scattering within the cork slab. The result of the superposition algorithm, displayed in Fig. 15.7b, accounts for most of the lateral scattering of the secondary electrons as indicated by the extended 10% isodose line in cork. The effect of energy transport through secondary electrons is even more pronounced if the results of the Monte Carlo calculation, shown in Fig. 15.7c, are considered.
These results are consistent with the respective depth-dose curves shown in Fig. 15.7d. The enhanced lateral scattering of secondary electrons in the low-density medium reduces the dose values on the central-beam axis, an effect which is almost completely missed by the pencil-beam calculation resulting in a severe overestimation of the respective dose of up to 12% of the maximum dose. This discrepancy is substantially reduced by the superposition algorithm. Only the Monte Carlo calculation was able to reproduce the experimental data with an accuracy of 2%, which was the estimated accuracy of the measurement.

While these simple phantom experiments may reveal some generic features of the discussed dose calculation algorithms, it is not clear to what extent the various algorithms generate dose differences which are of clinical relevance.

In the final section of this review we address this issue briefly by considering the influence of different dose calculation engines on IMRT dose optimization for clinical cases with abundant severe tissue inhomogeneities.

15.6 Dose Calculations and IMRT Optimization

The role of advanced high-energy photon dose calculations for iterative IMRT treatment planning is still under investigation. Since conventional algorithms, such as pencil-beam methods with poor consideration of inhomogeneities, could produce substantial deviations in media different to water, the quality of intensity-modulated treatment plans generated through those dose calculation methods was recently reviewed by different groups (Jeraj et al. 2002; Siebers et al. 2001, 2002; Scholz et al. 2003a).

Particularly two aspects are important referring to dose calculation in IMRT: firstly, one has to assess the accuracy of the respective algorithm as required for the determination and evaluation of the final dose pattern originating from a given set of fluence amplitudes. Deviations from a reference dose calculation can be called systematic errors. Secondly, the influence of the dose algorithm in the optimization process itself should be analyzed, since deviations in the dose calculation induces differences in bixel intensities in the fluence maps, which is usually referred to as convergence error. In order to demonstrate the nature of these two errors, we consider one of the most sensitive clinical cases with respect to dose calculations, the irradiation of small, solid lesions embedded in lung tissue.

15.6.1 The Systematic Error

As mentioned previously, the systematic error indicates the difference in dose which arises from the application of different dose calculation algorithms for a given set of fluence matrices. A solid lung tumor of about 30-mm diameter located in the left lung is totally surrounded by low-density lung tissue. It is irradiated with five coplanar, intensity-modulated beams. As indicated by the dose distribution on a transversal CT slice of the patient in Fig. 15.8a, the original optimization based on the pencil-beam dose calculation seems to cover well the PTV by the prescribed dose of 63 Gy (100% isodose); however, the recalculation with the superposition algorithm shown in Fig. 15.8b reveals that only a small volume of the target is encompassed by the 60 Gy isodose (95% isodose) and that a severe underdosage of about 13 Gy in mean dose is observed for the PTV.

The systematic error induced by the pencil-beam algorithm demonstrates that this dose calculation method severely overestimates the dose inside the tumor. For the considered case the respective underdosage of the tumor shown by the superposition algorithm is based on the fact that for high-energy photon fields the range of many secondary electrons is larger than the radius of the tumor volume shown above; thus, there is not enough material to absorb the total number of secondary electrons inside the target region. The remaining electrons simply are scattered into the low-density lung tissue. These findings are also well represented by the dose-volume histogram shown in Fig. 15.8d.

15.6.2 The Convergence Error

The convergence error is defined by the dose difference which arises from the application of two different fluence matrices – obtained by inverse planning with two different dose algorithms – for which the same reference dose calculation is applied. For our example we compare the dose distributions in Fig. 15.8b, where the pencil-beam algorithm was used for the generation of the fluence matrices, and the dose distribution in Fig. 15.8c, where the superposition algorithm was employed for the optimization. Once the fluence matrices were established, the final dose distributions were obtained by a calculation with the superposition algorithm. As clearly indicated by the fluence matrices shown in Fig. 15.9, the optimiza-
tion with the superposition algorithm compensates the obvious underdosage of the PTV created by the fluence generated with the pencil-beam method. The fluence matrix on the right-hand side of Fig. 15.9, obtained with the superposition algorithm, enhances the bixel weights at the periphery of the projected tumor such that an adequate dose coverage of the PTV is guaranteed. This fact is also well reflected by the respective values of the dose-volume histogram shown in Fig. 15.8d.

**15.7 Conclusion**

Dose calculation algorithms play a central role for the clinical practice of radiation therapy. They form
the basis for any treatment plan optimization, a feature which becomes increasingly important with the development of complex treatment techniques such as IMRT. The role of highly accurate and therefore mostly time-consuming dose algorithms, such as superposition algorithms or Monte Carlo simulations, in clinical radiation therapy is still under investigation. Their increased accuracy offers substantial advantages for clinical cases which involve intricate tissue inhomogeneities.

Even if in many radiotherapy centers the treatment plans are still based on the pencil-beam method, its general applicability to inhomogeneous clinical cases has to be questioned. On the other hand, inside quite homogeneous regions, as in the central head region or the abdomen, the pencil beam generates dose distributions with excellent precision and provides the best trade-off between accuracy and calculation times.

In the case of severe tissue inhomogeneities the superposition method produces dose distributions which fairly cover the target region, even if minor differences are observed in comparison with Monte Carlo calculations. These offer the best prediction of the deposited dose inside arbitrary tissue types. Some Monte Carlo-based programs already offer computation times comparable to those of superposition algorithms, and therefore their applicability in clinical practice will probably further increase for a small and special class of clinical cases.

References

16 Monte Carlo Dose Calculation for Treatment Planning

Matthias Fippel

CONTENTS

16.1 Introduction 197
16.2 Concept of Monte Carlo Integration 198
16.3 Photon Transport Simulation 199
16.4 Electron Transport Simulation 199
16.5 Statistical Noise and Variance Reduction Techniques 200
16.6 Material Parameters 201
16.7 Beam Modelling 202
16.8 Monte Carlo Dose Engines 203
16.9 Monte Carlo Dose Calculation for Inverse Planning and IMRT 204
16.10 Monte Carlo Dose Calculation for Ion Beam Therapy 205
References 205

16.1 Introduction

To calculate the dose in human tissue caused by ionizing radiation one has to solve a complex equation, the so-called transport equation. This equation is different for each patient and is also dependent on the treatment conditions, i.e. field size, field shape, energy of the radiation, beam direction, behaviour of the irradiation device, etc. There are two classes of calculation methods to solve the transport equation, direct and indirect methods. Indirect methods (see Chap. 15) start with a known solution of the transport equation, e.g. a measured dose distribution in water. This distribution is then corrected to take into account the beam configuration and the tissue inhomogeneities of the patient. Direct methods calculate dose distributions numerically by solving the transport equation explicitly. The patient geometry is modelled by a three-dimensional distribution of tissue types based on a set of CT images. The particle fluence of photons and electrons at the patient’s surface is taken into account by a model of the treatment device, e.g. a model of the linear accelerator head. Direct methods are pencil-beam (see Chap. 15), collapsed cone (see Chap. 15) and Monte Carlo (MC) algorithms.

Pencil-beam and collapsed-cone algorithms are different implementations of the superposition/convolution technique. They are based on various approximations and simplifications, such as:

- The modelling of the treatment device by simple sources, e.g. point sources or parallel sources
- The modelling of electron transport on straight lines
- The neglect or simplified modelling of the lateral density scaling in the case of pencil-beam algorithms (slab approximation)
- The discretisation of the point-spread function (energy kernel) in the case of collapsed-cone algorithms
- The neglect or approximate representation of the energy kernel tilt angle in order to reduce computation time
- The neglect or estimation of effects such as depth hardening and off-axis softening

Usually, these techniques provide results within seconds. On the other hand, dose calculation can be incorrect especially in the head, neck and thorax regions because of the approximations used. As an example, the dose distribution of a 12-MeV electron beam with bolus in a patient with mammary-gland carcinoma is calculated with Monte Carlo and the pencil-beam algorithm of a commercial treatment planning system (see Fig. 16.1). The MC-generated dose distribution (Fig. 16.1a) shows the influence of low density tissue (lung) on the range of the electrons; however, the electron range is underestimated if a pencil-beam algorithm is used for dose calculation. This effect is demonstrated clearly by the cor-
Fig. 16.1a,b. Dose distribution of a 12-MeV electron beam with bolus in a patient with mammary-gland carcinoma. a Isoline representation of the dose distribution calculated with Monte Carlo. b Dose vs depth in the patient calculated with Monte Carlo (crosses) and the pencil-beam algorithm of a commercial treatment planning system (line).

responding depth dose curves in Fig. 16.1b. As a consequence, the pencil-beam algorithm underestimates the dose in lung. Also the heart can be affected if it is close to the target volume. It is a common objective of medical physics to achieve an accuracy of better than \( \pm 5\% \) for the delivery of dose. But this can be realised only if the dose calculation accuracy is better than \( \pm 2\% \); therefore, in the future, Monte Carlo algorithms will have a clear preference compared with all other methods of dose calculation.

16.2 Concept of Monte Carlo Integration

Monte Carlo methods are older than 200 years. They are used for solving mathematical problems, e.g. for numerical integration of functions. Figure 16.2 shows a simple demonstration example. The integral of \( f(x) \) in the interval \([a,b]\) must be solved to calculate the area embedded by function \( f(x) \), the \( x \)-axis and the interval limits \( a \) and \( b \). If function \( f(x) \) is complicated, this cannot be performed using integration rules. But it can be solved numerically by Monte Carlo integration. For this purpose we choose a random number \( \xi \) uniformly distributed in interval \([a,b]\).

Random numbers are provided by computer algorithms called random number generators. These random numbers are not really random because for a given state of the random number generator the series of random numbers produced by the generator is pre-defined. However, a high-quality random number generator should produce uncorrelated numbers; therefore, these numbers are often called pseudorandom numbers.

The function value \( f(\xi) \) for random number \( \xi \) multiplied by the length of the interval \( b-a \) provides a first and a very rough estimate of the real integral (see Fig. 16.2, right). But this can be performed again and again. We sample a large number of random numbers, calculate the area of the corresponding rectangles and calculate the mean. This mean value converges to the real integral (or real area) in the limit of infinite random number samples. In reality we cannot sample infinite random numbers, i.e. we stop if the desired accuracy is achieved.

This method of integration is not very efficient for one-dimensional integrals, i.e. for integrals of functions with only one variable as in the example of Fig. 16.2. It is also not very efficient for two-dimensional or three-dimensional integrals. There are more efficient algorithms such as Gaussian Quadrature. Monte Carlo integration becomes important if the dimensionality of the problem is very large or if the dimensionality approaches infinity. Especially this is the case if the dose distribution in a patient caused by ionizing radiation is calculated. For that purpose the so-called transport equation must be solved. The
transport equation is different for each patient. It also depends on treatment conditions such as beam directions, field sizes, energy, etc. But it can be solved numerically using MC methods by simulating a large number of so-called particle histories. In this context a particle history is given by the path of one photon or one electron entering the calculation geometry until the whole energy is absorbed or until the particle and all secondary particles have left the volume of interest (see Fig. 16.3). On this path, energy and momentum of the particles can change and secondary particles can arise because of interaction processes with tissue molecules. The features of these processes are determined by probability distributions given by the total and differential interaction cross sections. Sampling process parameters randomly from these probability distributions forms the basis of MC simulations in radiation physics.

![Fig. 16.3. Example of a photon history in the Monte Carlo simulation of radiation transport. A photon p (upper solid line) enters the region of interest and is transported to the first Compton interaction site. There, the photon is scattered and an atom is ionised. The Compton electron e⁻ (dashed line) leaves the interaction site and is transported until it has lost its energy. The scattered photon can undergo further interactions, e.g. a pair production process (lower photon). Each particle (primary or secondary) must be simulated until the whole energy is absorbed or until it has left the geometric region of interest.](image)

**16.3 Photon Transport Simulation**

Photon transport is an excellent example to demonstrate how MC simulations in radiation physics work. Given a photon at a definite location with given momentum and energy, the first step is to sample the free path lengths until the next interaction site. The probability distribution \( f(s) \) of these path lengths is given by the exponential attenuation law:

\[
f(s) = \exp(-\mu s),
\]

with \( \mu \) being the linear attenuation coefficient. The photon path lengths can be sampled from this distribution using a uniformly distributed random number \( \xi_1 \) from interval \([0,1]\) and the relation:

\[
s = -\frac{1}{\mu} \ln(\xi_1).
\]

Using this path length the photon can be tracked to the interaction site taking into account different materials with different attenuation coefficients \( \mu \) in each voxel of the calculation grid. In the energy range of radiation therapy \( \mu \) is calculated as a sum of three relevant contributions:

\[
\mu = \mu_{\text{photo}} + \mu_{\text{Compton}} + \mu_{\text{pair}}
\]

with \( \mu_{\text{photo}}, \mu_{\text{Compton}} \) and \( \mu_{\text{pair}} \) being the linear interaction coefficients or total cross sections for photoelectric absorption, Compton scatter and pair production, respectively. These parameters are different for photons of different energy. They also depend on the atomic composition of the material, i.e. they change from voxel to voxel. The interaction coefficients must be calculated from the Hounsfield units of the CT image set. A second random number \( \xi_2 \) from interval \([0,\mu]\) can be used to sample the interaction type. We simulate a photoelectric absorption if \( \xi_2 \) is less than \( \mu_{\text{photo}} \). We simulate a Compton interaction if \( \xi_2 \) is larger than \( \mu_{\text{photo}} \) but less than the sum of \( \mu_{\text{photo}} \) plus \( \mu_{\text{Compton}} \); otherwise, we simulate a pair production process. The parameters of secondary particles after the chosen interaction, such as energy and direction, can be sampled using further random numbers and the corresponding differential cross sections for that interaction type. The formulas of the probability distributions are more complex compared with the formulas above but the sampling principle remains the same. Secondary particles are simulated like the primary particle, i.e. their transport starts with sampling the free path length to the next interaction site. The procedure continues as in the case of primary particles. In each voxel the absorbed energy must be determined and accumulated. Later this leads to the dose distribution. The particle history ends if the photon leaves the calculation matrix or if its energy drops below a minimum energy.

**16.4 Electron Transport Simulation**

Theoretically, electrons (primary or secondary) could be simulated like photons; however, there is a big difference between photons and electrons. The mean
free path lengths of photons between two interaction sites in human tissue is of the order of several centimetres; therefore, the procedure described in section 16.3 works very well. Electrons, on the other hand, undergo millions of interaction processes along their path through the patient; therefore it would be too time-consuming to simulate electrons like photons. Fortunately, most of the electron interactions are elastic or semi-elastic, i.e. without or with small transfer of energy; therefore, electron interactions can be classified into two groups: hard interactions with large energy transfer and soft collisions without or with small electron energy loss. Hard interactions can be simulated explicitly, and soft collisions can be simulated by continuous electron energy transfer to the surrounding tissue. This technique, called condensed history technique, was introduced by BERGER (1963). Direction changes of the electron due to elastic collisions are modelled using a multiple scattering theory; therefore the multiple scattering properties of the different media must be known. Also the electron stopping power is an important material parameter to calculate the correct amount of energy transferred from the electron to the medium. The electron stops if its residual range becomes smaller than the voxel size or the spatial resolution.

To classify electron interactions into hard interactions and soft collisions, an arbitrary parameter must be introduced. Typically this parameter is given by a threshold energy. The collision is called soft if the energy transfer in a collision is smaller than this threshold energy; otherwise, it is called hard. During hard interactions, secondary electrons (also called delta electrons) or "bremsstrahlung" photons can be released. In a Monte Carlo calculation these secondary particles are simulated like primary particles, again with the capability to emit further secondary particles. Because the threshold energy is an arbitrary parameter, the final result, e.g. the dose distribution, must not depend on the value of this parameter.

There are additional arbitrary parameters. One of the most important of these parameters is the maximum length of a condensed history step. This parameter is necessary to minimize the error introduced by the condensed history technique. The path of electrons between two hard interactions in condensed history simulations is modelled by a straight line. This is in contradiction to nature, because electron paths are curved; therefore, the charged particle path length (or range) can be overestimated in the simulation. Most Monte Carlo algorithms calculate path length corrections to avoid this overestimation. Nevertheless, the simulation results can be wrong, if the electron path length is too large. A further problem of the condensed history technique arises if the electron crosses a boundary of different media. Also here artefacts can be avoided by restricting the length of the condensed electron steps; therefore, the maximum step length must not be too large. On the other hand, it should not be too small. In this case the simulation time can be too long. Meanwhile there are sophisticated condensed history algorithms (KAWRAKOW 2000a) with almost no artefacts. With these algorithms step-size restrictions are no longer necessary.

The discussion of these problems is included here to show that the use of Monte Carlo techniques without care is no guarantee for correctness. It also shows that Monte Carlo is not absolutely equivalent to nature. There are a lot of potential sources for errors. But on the other hand, only Monte Carlo algorithms have the potential to be as accurate as possible. This is not the case for all other types of dose calculation.

16.5 Statistical Noise and Variance Reduction Techniques

The statistical variance or noise in each voxel is determined by the number of simulated particle histories. The more histories are simulated, the smaller is the statistical variance. The number of histories must be increased by a factor of four to reduce the statistical variance by a factor of two, i.e. the calculation time increases by a factor of four. For an accurate dose distribution the statistical variance should be smaller than 2% relative to the maximum dose of the distribution. This accuracy can be achieved by simulating a corresponding number of histories; however, this may lead to long calculation times. Often, this time is not available in clinical practise; therefore, other techniques and tricks are required to reduce the variance. These variance reduction techniques do not require additional computer time. During past years many such techniques have been developed and tested such as interaction forcing, cross-section enhancement, woodcock tracing, initial-ray tracing, particle splitting, Russian roulette, correlated sampling the use of quasi-random numbers, history repetition, continuous boundary crossing, range rejection, and Super Monte Carlo or Macro Monte Carlo (KAWRAKOW and FIPPEL 2000). Furthermore, approximations can be implemented to reduce the variance. Examples are the Kerma approximation for photons with an energy smaller than a cut-off value or the continu-
ous slowing approximation for electrons below some minimum energy. All the additional parameters must be chosen carefully to assure correct results within the allowed accuracy limitations.

Another possibility to reduce the variance are denoising techniques and filters. These techniques are known from diagnostic imaging. During the image measurement process the detector signal is superimposed by noise. This noise can be removed using digital filters. Similar techniques can also be used to smooth a MC-generated 3D dose distribution. Known techniques are, for example, Savitzky-Golay filters (Kawrakow 2002), wavelet threshold denoising (Deasy et al. 2002), anisotropic diffusion (Miao et al. 2003) and iterative reduction of noise methods (Fippel and Nüsslin 2003); however, there is a problem with all these methods: they can introduce systematic bias and the amount of this systematic error is difficult to estimate. Therefore, these techniques should be used with care. They should be part of a MC treatment planning system. They can shorten the planning time especially in the intermediate stage of the planning process; however, the final clinical decision should not be based on smoothed dose distributions, i.e. after the plan optimisation, the whole dose distribution should be recomputed without denoising filters.

Unfortunately, it is impossible to remove all statistical fluctuations from the dose distributions, because it is impossible to simulate an infinite number of histories. In an isodose line representation of the dose distribution this may lead to jagged isodose lines (see Fig. 16.4). This can be problematic because it is not clear in advance whether this is caused just by noise or by a real physical effect; therefore, we have to learn how to deal with these fluctuations. Generally, in air the statistical variance is larger compared with soft tissue; thus, we can assume a jagged isodose line in air is probably caused just by noise (Fig. 16.4).

The Monte Carlo simulation time also depends on parameters such as energy, field size and voxel size. Increasing the energy in the case of electron beams leads to larger ranges of the electrons; therefore, more time for tracing these electrons is required. In the case of photon beams the calculation time only slightly depends on the nominal energy. If we increase the field size, more particle histories must be simulated, i.e. the calculation time is approximately proportional to the field area for both, electron and photon beams. Also the voxel size is a sensitive parameter to control the calculation speed. With larger voxels the simulation can be accelerated considerably. On the other hand, spatial resolution is lost. For the intermediate planning stage larger voxels can be very useful. The final dose distribution should be calculated with high spatial precision; therefore, the planning system should permit the change of voxel sizes.

16.6 Material Parameters

For an accurate simulation of photon and electron transport through human tissue the interaction cross sections must be known. For photons we have to know mainly the interaction coefficients for photoelectric absorption, Compton scatter and pair production. For electrons we have to know the collision and radiation stopping powers as well as the scattering power in each voxel. These parameters can be calculated if the atomic composition is known; however, in general we only know the CT number for the corresponding volume element.

In a straight forward approach it is possible to divide the whole range of CT numbers (Hounsfield units) into intervals with each interval corresponding to a definite material or tissue type. For these materials the cross sections can be calculated and tabulated taking into account their atomic composition; thus, the cross sections can be determined in each voxel based on the CT number information. A major problem of this approach is that a large cross-section data basis for each material is needed. If the number of materials is too small, the errors can be significant.
Another approach obviates the use of pre-defined materials. Instead, the cross sections are calculated directly from a 3D distribution of a single material parameter like mass density. This means that a precise calibration of the CT scanner must exist to map each CT number into the corresponding mass density. In general, this mapping is different for different CT scanners because the CT number is dependent on the features of the scanner such as its energy. Mass density, on the other hand, is a pure material parameter. If the model of the patient is given by a 3D distribution of mass densities, then the cross sections can be calculated directly from the mass density without knowledge of the atomic composition. This is possible because human tissue from lung to bone behaves similarly in terms of electromagnetic interactions (see Fig. 16.5). Here for all materials from ICRU report 46 (ICRU 1992) the electron density ratio to water normalized by the mass density is plotted (crosses) vs the mass density. Knowledge of the electron density is important especially to determine the Compton interaction cross section. Figure 16.5 shows that most of the materials can be modelled by a functional dependence on mass density (solid line). The few outliers are materials such as urinary stone or gallstone. Similar dependencies exist for all the other interaction coefficients; thus, all interaction parameters for a material with given mass density can be determined by scaling the corresponding cross sections in water using the known correction factors.

All methods to determine the interaction parameters have limitations because they strongly depend on the accuracy of the CT calibration. Especially artefacts in the CT image can destroy the cross-section information. For example, artefacts caused by metal implants can make the whole image set unusable for an accurate dose calculation; therefore, more research will be necessary to solve these problems. This can be achieved by getting more information from the imaging process, e.g. by using additional modalities such as MRI or PET.

16.7 Beam Modelling

The dose calculation accuracy is also influenced by the quality of the treatment head model. Ideally this model should be realised by a Monte Carlo simulation of the whole linac head geometry (see Fig. 16.6) using a program code system such as BEAM (ROGERS et al. 1995). For that purpose a geometric model of all linac head components (target, primary collimator, flattening filter, collimators, etc.) must be created. Then photon and electron transport can be simulated through the whole geometry taking into account the correct material compositions. Unfortunately, this is still time-consuming. Variance reduction techniques, fast computer systems or clusters can speed up these simulations considerably; however, too much time is spent to simulate the transport of unnecessary particles, because they are later absorbed by the collimating system. A possible solution of the calculation time problem is the use of phase space files. In these files parameters are stored such as position, direction, energy, or charge of all particles hitting a plane below the collimating system. This plane is called phase space plane (see Fig. 16.6). During the beam commissioning process a phase space file containing a huge number of particles is produced. Later this plane (and the file) can be used as source for the Monte Carlo transport simulation through the patient. In this manner, no computation time is required during the treatment planning process for tracking the particles through the linac head geometry.

A serious problem in this approach is the lack of information about the properties of the electrons hitting the target. This electron source is characterized by a shape, an angular distribution and an energy spectrum. The modelling of the electron source is

![Fig. 16.5. Electron density ratio to water normalised by the mass density as function of the mass density for all materials from ICRU report 46 (crosses). This ratio provides the correction factor \( f_C(\rho) \) to calculate the Compton interaction cross section. It is also possible to use a fit function (solid line) instead of the data points. Similar relationships exist for other photon and electron interaction processes.](image-url)
complicated because position, direction and energy of the electrons are correlated, i.e. the phase space is given by a five-dimensional function. Furthermore, it is impossible or at least difficult to measure these parameters; therefore, a fitting procedure is required to adjust the electron source parameters using measured dose distributions in water and air. A whole treatment head simulation including a dose calculation in water and/or air with high statistical accuracy must be started each time a source parameter has changed; thus, a large number of simulations is necessary, i.e. models of this kind are inconvenient for the clinical practise.

A more practical option is to determine the features of the photon sources in the target and the photon and electron sources below the target (filter) using measurements (FIPPEL et al. 2003). Photons mainly have their origin in the target (primary photons) or in the flattening filter (secondary or scatter photons). The main source for electron contamination is also the filter. These sources are more complex compared with the primary electron source of the linear accelerator; however, primary and secondary photons directly influence the detector signal, if the detector (e.g. an ionisation chamber) is placed in air and if it is equipped with a small built-up cap. In this case the detector signal is proportional to the photon fluence and the parameters of the photon sources can be adjusted analytically without the need of extensive Monte Carlo simulations. Also information from the treatment head geometry can be used for these models, e.g. the positions of the target and the flattening filter. The features of the electron contamination source as well as the photon energy spectrum can be determined using measured depth dose distributions in water; however, this type of model makes sense only for the upper part of the linac head, i.e. for the part independent on the patient geometry and the treatment plan. Especially the beam collimating devices, such as the multi-leaf collimator, should be integrated directly into the Monte Carlo simulation (FIPPEL 2004); thus, it is possible to model effects such as inter-leaf leakage, inter-leaf transmission, tongue-and-groove effects or the influence of rounded leaf ends. Also dynamic treatment modalities can be simulated easily using the Monte Carlo method. This includes dynamic multi-leaf collimators or dynamic gantry rotations. In contrast to conventional dose calculation, these dynamic techniques can be simulated by a continuous movement of the corresponding device, e.g. a real moving leaf or a continuous gantry rotation. Sometimes this is called 4D Monte Carlo. Superposition dose calculation algorithms on the other hand require some discretisation of dynamic processes leading to drastically increasing calculation times. With Monte Carlo no additional calculation time is needed.

Similar approaches exist to model the treatment head of clinical electron beams. Here especially the influence of the electron applicator and the beam cut-out is important.

16.8 Monte Carlo Dose Engines

Many efforts have been made during the past years to develop Monte Carlo dose engines for treatment planning. One option is to adapt general-purpose Monte Carlo codes, such as EGS4/EGSnrc (NELSON et al. 1985; KAWRAKOW 2000a), MCNP (BRIESMEISTER 1997), PENELope (BARO et al. 1995) or GEANT4 (AGOSTINELLI et al. 2003), by interfacing them to treatment planning systems. The implementation of
additional variance reduction techniques (see section 16.5) leads to calculation times acceptable for clinical routine. Examples of this approach are the Macro Monte Carlo code (Neuenschwander and Born 1992) for electron beams as well as the Super Monte Carlo technique (Keall and Hoban 1996) and MCDOSE (Ma et al. 2002), both for photon and electron beams. A second option is the development of entirely new Monte Carlo codes especially designed for radiotherapy purposes. Examples are the Voxel Monte Carlo (VMC) algorithm (Kawrakow et al. 1996) for electron beams, its extensions for photon beams XVMC (Fippel 1999) and VMC++ (Kawrakow 2000b), the DPM code (Sempau et al. 2000) and the PEREGRINE system (Hartmann Siantar et al. 2001).

These new techniques allow dose calculations within minutes for photon beams and within seconds for electron beams using computer clusters (e.g. PEREGRINE) or ordinary personal computers (XVMC, VMC++). All treatment planning companies are working on the implementation of Monte Carlo for dose calculation. Some systems are already available. In the future all conventional dose algorithms will be replaced by Monte Carlo.

16.9 Monte Carlo Dose Calculation for Inverse Planning and IMRT

The quality of dose calculation is of special importance for inverse treatment planning (see Chap. 17) in IMRT (see Chap. 24). In contrast to conventional radiotherapy, with IMRT techniques larger dose values can be achieved in the tumour volume compared with the organs at risk. This leads to large dose gradients if these organs are close to the tumour. The situation is even more complex if these structures are located in an environment with density inhomogeneities consisting of soft tissue, bone and air cavities. This is the case especially in the head, neck and thorax regions. Here the influence of photon Compton scatter and lateral electron flow is not negligible. Conventional dose calculation algorithms (see Chap. 15) often fail to predict the dose in the neighbourhood of air cavities. The effect of secondary electron transport is energy dependent. For small energies (like 6 mV) the range of these electrons is of the order of 1 cm in water. For large energies (15 or 18 mV) this range is larger (up to 5 cm); therefore, dose errors can be large if the secondary electron transport is not modelled correctly. In inverse planning the dose error translates into a corresponding convergence error, i.e. the dose distribution will be as required, but the fluence distribution is wrong. In other words, the fluence distribution determined during the optimisation process will in reality lead to a dose distribution different to the calculation and different to the requirements from the objective function.

A further problem in IMRT arises with the modelling of the beam delivery system, i.e. mainly with the multi-leaf collimator or the compensator. These devices have a significant influence on the dose calculation (see section 16.7). Many IMRT planning systems optimise photon fluence weights. The MLC sequences are generated subsequent to the optimisation process. To ensure correct dose distributions a dose re-calculation must be performed based on the generated MLC segments. Often Monte Carlo algorithms are used for the purpose of dose verification. A treatment plan re-optimisation may be necessary if the result of the dose verification is significantly different from the optimised dose distribution.

A better alternative is the integration of the MLC modelling and sequencing (or the modelling and calculation of Compensator shapes) into the optimisation loop (Alber et al. 2003). This means that after the optimisation of the fluence weights the MLC sequencing process is started. Then the dose distribution for each MLC segment can be calculated separately using Monte Carlo taking into account full head scatter, MLC leakage, MLC transmission and tongue-and-groove effects. Now the optimisation process is initiated again with the aim to improve the total dose distribution by adjusting the weights of each segment. Using this approach it is also possible to change MLC contours, i.e. to change leaf positions. For this purpose the dose distribution for the corresponding segment must be re-calculated. In this manner there is no disagreement between the optimised dose distribution and the final calculated dose distribution.

A real clinical benefit from Monte Carlo can be expected if it is used in conjunction with IMRT and inverse planning. On the other hand, there are additional problems in inverse planning if we use Monte Carlo for dose calculation instead of analytical techniques. One problem is caused by the statistical noise. It has been shown theoretically (Kawrakow 2004) that the statistical uncertainty of the dose in each voxel causes a systematic uncertainty of the objective function. Fortunately, this systematic deviation can be calculated and taken into account during the optimisation. Another problem comes from the ran-
dom number generator. The final dose distribution changes slightly if we change the initial state (also called seed) of the random number generator. But this can lead to different solutions of the optimisation problem if the objective function provides a large variety of optimum treatment plans. In this case we have to accept that there is more than one reasonable result. A third problem can arise from voxels of air belonging to the planning target volume. Generally, this is no problem for conventional dose calculation such as pencil-beam dose calculation because these algorithms have the tendency to overestimate the dose in air. Using Monte Carlo the dose in air is calculated correctly, i.e. it is significantly smaller than the dose in the surrounding tissue. If these voxels are part of the target volume, the optimisation algorithm tries to compensate for this apparent underdoseage. However, clinically this compensation is unnecessary; therefore, we should avoid air in the planning target volume wherever possible. This may be difficult if the planning target volume is given by the clinical target volume with a margin to take into account setup and motion errors. It will be impossible to avoid air in the planning target volume if air cavities are close to the clinical target volume. In these cases it is useful to introduce weight factors with small weights in the objective function for regions with air in the target volume.

16.10 Monte Carlo Dose Calculation for Ion Beam Therapy

The electromagnetic interaction properties of protons and heavier ions are similar to electrons; however, because of their larger mass, ions show less lateral scattering effects. Only in the Bragg-Peak region is the beam broadened because of multiple elastic scattering; therefore, dose calculation can be performed accurately and efficiently using pencil-beam algorithms. On the other hand, nuclear interaction processes have a significant influence on dose and radio-biological effectiveness. Within the framework of pencil beam dose calculation these processes can be taken into account only in an approximate manner. Especially the dependence of the nuclear interaction cross sections on the tissue composition is difficult to model. Using Monte Carlo this is much easier and more precise. The most important requirements are accurate cross sections and nuclear interaction models. If this is available, all kinds of particles produced in nuclear reactions can be simulated and tracked through the patient anatomy in a manner similar to electrons and photons.

References

17 Optimization of Treatment Plans, Inverse Planning

Thomas Bortfeld and Christian Thieke

17.1 Introduction

Computerized optimization of treatment plans has been proposed in radiation therapy since computers became available to hospitals in the 1960s (Hope et al. 1967; Bahr et al. 1968; Redpath et al. 1975; McDonald and Rubin 1977). The first prototypes were presented at that time but did not find their way into clinical practice. The main reason was that those early approaches were basically restricted to the optimization of beam weights, sometimes in combination with wedges, which have a limited potential to shape dose distributions in 3D. Optimization of beam directions in fully 3D was not really possible with computers of that time.

The situation changed drastically with the invention of intensity-modulated radiation therapy (see Chap. 23) by Brahme and colleagues in 1982, along with exponential increase of computer power. The IMRT has a much greater potential to shape complex spatial dose distributions. Interestingly, the first approaches to calculate intensity maps for IMRT were not based on optimization techniques. They instead inverted the underlying integral equation analytically, which means that the spatial dose distribution was prescribed, and the beam intensity distribution that would precisely yield this dose distribution was calculated. To find an exact solution to this inverse problem, several assumptions and approximations had to be made, and solutions could only be found for very simple symmetrical cases. Nevertheless, based on this inversion idea, IMRT planning is often still called inverse planning. Other early attempts at solving the inverse problem of radiation therapy planning used deconvolution techniques that were borrowed from image processing (Brahme et al. 1982; Holmes et al. 1991). Again, the solutions were always approximations.

A true solution of the inverse problem cannot be found in any but the simplest hypothetical cases, because the beam intensities can never become negative and therefore we can never subtract dose. Mathematically, the problem to generate a desired dose distribution with rays of radiation of variable intensity is equivalent to the problem of creating an arbitrary drawing using uniform straight lines drawn with a pencil and ruler on paper. This problem was mathematically analyzed by Birkhoff (1940). He found that the problem can only be solved if an eraser is also available. Unfortunately, we do not have a physical dose eraser at hand in radiation therapy.

Therefore, other mathematical methods were soon applied to this problem. One of them is called feasibility search (Censor et al. 1988; Starkschall et al. 2000). Here the idea is that, even though there is no
exact solution to the inverse problem that will yield the ideal dose distribution to the tumor with no dose to normal structures, we may still be able to find a plan that fulfills all clinical and physical constraints. This approach can be very useful if the constraints are carefully chosen: we do not want to make it too easy to fulfill the constraints because that may result in suboptimal plans. We also do not want to over-constrain the problem, because then there will be no solution at all. Feasibility search could be particularly successful for clinical cases that do not vary too much between individuals, such as prostate cancer. Here it is possible to find class solutions of constraints that will yield good treatment plans for many patients.

Presently, the most commonly used approach in IMRT planning is the optimization approach, in which one tries to find the best physically and technically possible treatment plan with respect to given physical and clinical criteria. Among the first published optimization techniques were those by Webb (1989) and Bortfeld et al. (1990). It was shown that with the optimization approach complex dose distributions could be generated with a limited number (less than ten) of beam directions (Bortfeld et al. 1990). In most other previous approaches the use of an unrealistically large number of beams was assumed. This paved the way for implementation of IMRT at conventional linear accelerators equipped with a multileaf collimator. Even though we focus on IMRT optimization, many of the concepts that we discuss in this chapter can be applied to general radiotherapy optimization with other modalities.

We first discuss the principles of current optimization approaches in IMRT. In section 3 we discuss the clinical aspects of IMRT optimization, and in section 4 we give an outlook on future developments in this field.

17.2 Optimization Principles

Optimization is mathematically defined as the maximization or minimization of a score (a number) under certain constraints. The scoring function (also called “objective function”), which yields the score, is a function of all variables that need to be optimized. In general mathematical optimization the problem is to find the set of variables that yield the maximum (or minimum) score while fulfilling all constraints. Specifically, in radiotherapy optimization the problem is to find the treatment plan with all its associated variables (e.g., beam intensities and angles) that yield the best scoring treatment for an individual patient. In IMRT the number of variables (the intensities for all beam elements for all beams) can be in the thousands, so we are typically dealing with a highly dimensional optimization problem; however, it is important to note that the score is a single number. The issues that are related with this are discussed later, but it is immediately clear that it is a formidable task to characterize the goodness of a radiation treatment plan with a single number.

The basics of optimization in radiotherapy have been discussed in many recent reviews and book chapters (Bortfeld 1999; IMRT 2001; Bortfeld 2003; Censor 2003), so we are brief on this subject.

17.2.1 Optimization Criteria

The definition of the objective function and constraints (for which we use the combined term “optimization criteria” in the sequel) is arguably the most challenging part of IMRT optimization, and requires a lot of thought. Optimization criteria that are not clinically relevant or not complete can lead to treatment plans that are mathematically optimal but clinically useless. Mathematical optimality is not a guarantee for a good treatment plan at all. In the definition of the objective function one has to translate the clinical experience gained over the past 100 years or so into mathematical terms, which in turn needs to be translated into computer language. The key question is: What characterizes the “best” treatment plan? Clinicians often find it difficult to formulate a complete, unique, quantitative set of optimization (scoring) criteria for radiotherapy planning, even though they feel capable of ranking individually prepared plans (Langer et al. 2003). This is somewhat different from other applications of optimization, e.g., in economics, in which the objective is often quite clear (e.g., minimize the cost).

It may seem obvious to define the optimization criteria in terms of treatment outcome, say, maximize the probability of curing the patient without exceeding tolerance limits of side effects in normal tissues. The problem is that there exists no commonly accepted and validated model that translates the physical dose distribution into normal tissue complication probabilities (NTCP) or tumor control probabilities (TCP), let alone cure rates. In 1968 Alexander Solzhenitsyn wrote in his novel Cancer Ward: “There was no formula for calculating the right intensity of
The most common approach in practical radiotherapy optimization at present is therefore to use optimization criteria that are defined in terms of the physical dose distribution. In many cases these criteria are not absolutely fixed, but are adjusted by the treatment planner in a “human iteration loop,” as the optimization progresses. The clinical aspects of this are explained in detail in the subsequent section. Suffice it to say that optimization using these physical criteria is not expected to yield the best clinical plan in one step. What is more important is that these criteria provide good “steerability” of the plan: If, after the first optimization run, the plan is not satisfactory from a clinical point of view, an adjustment of the criteria should allow to steer the plan in the desired direction in a straightforward manner using only a few of the “human iteration loops.”

We now discuss optimization criteria that are used in various commercial IMRT optimization systems. The physical dose criteria discussed below can be used as either constraints or as objectives. In the constraint formulation one merely requests that the criteria stay within certain limits. If they are defined as (part of) the “objective function,” the goal is to maximize or minimize these quantities.

17.2.1.1 Physical Dose Criteria

A widely used criterion is the quadratic deviation of the actual dose from a prescribed dose level $D_{\text{min}}$ (in the target volume) or tolerance dose level $D_{\text{max}}$ (in both the target volume and critical structures), which is to be minimized. Different weight factors (also called penalty factors or importance factors) can be used for underdose and overdose. The objective function (“costlet”) for a particular organ is then:

$$F(\hat{b}) = \sum_{i=1}^{N} \left( u [D_{\text{min}} - d_i(\hat{b})]^2 + w [d_i(\hat{b}) - D_{\text{max}}]^2 \right),$$

where $u$ and $w$ are the weight factors for under- and overdose, respectively. $d_i(\hat{b})$ is the dose at voxel $i$ as a function of the beam element (bixel) intensities and other variables (e.g., beam angles) combined into the argument vector $\hat{b}$. $D_{\text{min}}$ and $D_{\text{max}}$ are the prescribed minimum and maximum doses that should ideally not be exceeded. Note that $D_{\text{min}}$ and $D_{\text{max}}$ are not the minimum and maximum doses of the actual dose distribution $d_i(\hat{b})$. One often sets $D_{\text{min}} = D_{\text{max}} = \text{prescribed dose in the target volume}$. In critical structures the first term is omitted by setting $u=0$ (or equivalently, setting the minimum dose constraint to 0). $[x]_+$ stands for $x$ if $x > 0$ and 0 otherwise. This quadratic deviation approach has been criticized because it does not sufficiently penalize dose cold spots in the target volume or hot spots in critical structures. For example, if the dose in the target volume becomes zero in only one voxel, $F$ may still be in an acceptable range, as long as most voxels receive the prescription dose. Higher-order measures of the deviation (such as differences taken to the power of 4, 6, ...) have been suggested (Fraass 2002) but are not common in clinical IMRT optimization.

One way to enforce dose homogeneity is to put hard constraints on the actual minimum and maximum dose in the target volume (as opposed to the “soft” deviation approach above). For example, an optimization constraint might require to keep the target dose within -5% and +7% of the prescription dose, as recommended by the International Commission on Radiation Units and Measurements (ICRU 1993).

A possible optimization objective is to maximize the minimum dose value in the target volume (Langer et al. 1996). Maximum dose limitations (but, of course, not minimum dose limitations) make sense in critical structures as well. In structures with a serial organization, such as the spinal cord, the complication is correlated with the maximum dose. For example, in the spinal cord the maximum dose should be limited to, for example, 45 Gy.

The mean dose is a useful descriptor of the dose effect in the target volume as long as deviations from the mean are not too big (Brahme 1984; Levegrun et al. 2001). The mean dose is also a useful clinical parameter in critical structures that are organized in parallel, such as the lung (Kwa et al. 1998). It should be noted, however, that different regions in the lung have different radiation sensitivities. The mean dose overall can therefore only be a very coarse predictor of side effects. It should also be emphasized that different kinds of complication can occur in the same organ, e.g., in the lung, such that different criteria may have to be used for one organ in the optimization.

Another common method to represent dose effects in parallel critical structures is to use dose-volume histogram (DVH) constraints (Langer and Leong 1989; Langer et al. 1990; Niemierko 1992; Bortfeld et al. 1997; Spirou and Chui 1998; Gustafsson and Langer 2000), which take into consideration the volume dependence to some degree. The DVH constraints can be formulated as “no more than $V_{\text{max}}$% of...”
the volume should receive more than a dose of $D_{\text{max}}$”. They can be visualized as a barrier with a corner at the point $D_{\text{max}}, V_{\text{max}}$ on the DVH plot. The use of multiple DVH constraints may also be indicated in some cases (Carcal et al. 1997), and DVH constraints can be used for the target volume as well.

One general advantage of physical criteria, such as dose-volume criteria, is that, because they are simply and clearly defined, they can be used easily in clinical protocols. Such IMRT protocols using a combination of various physical dose criteria have been developed and distributed by the Radiation Therapy Oncology Group (RTOG, www.rtog.org). A well-known example is their protocol 022 for IMRT of oropharyngeal cancer.

A potential mathematical problem with dose-volume constraints is that they are non-convex (Deasy 1997). As a consequence of this, optimization based on DVH constraints may get trapped in local minima; however, it has been shown that this is mainly a theoretical problem, which seems to be of little relevance in practical optimization (Wu and Mohan 2002; Llacer et al. 2003).

The final physical dose criterion that we discuss is the equivalent uniform dose (EUD). It is defined as the uniform dose that would create the same biological effect in a specific organ as the actual non-uniform dose distribution (Brahme 1984; Niemierko 1997). Because it involves the biological effect, it is often interpreted as a biological rather than physical dose criterion; however, according to a more recent definition by Niemierko (1999) the EUD is simply the generalized mean ($a$-norm) of the physical dose distribution:

$$
\text{EUD} = \left( \sum v_i \cdot d_i^a \right)^{1/a}
$$

where $v_i$ is the volume of voxel $i$ divided by the total volume of the organ. With this definition and with $a = 1$ the EUD is in fact the mean dose and with $a = \infty$ it equals the maximum dose. Of course, the value of $a$ is organ specific. Values between 1 and infinity represent organs with a mixture between parallel and serial structures. It is interesting to note that the value of $a$ can be derived from the well-known power-law relationship of the tolerance dose $TD$ as a function of the relative treated volume $v$:

$$
TD(v) = \frac{TD(1)}{v^a}.
$$

One finds that $a = 1/n$.

For target volumes, the value of $a$ is negative. The beauty of this approach is its simplicity and general-
in subsection 2.1.1 using organ-specific weight factors $u_k$ and $w_j$. According to this definition the optimum treatment is the one with the smallest overall deviation from the prescription. In the plan steering process (the “human iteration loop”) it is not a priori clear by how much to change the weights to achieve the desired effect on the dose distribution. Suitable weight factors have to be determined by trial and error, which can be quite time-consuming and is discussed further below. Some researchers argued that optimized “inverse” IMRT planning is just replacing the conventional manual trial-and-error search to find the best beam parameters by a manual trial-and-error search for suitable weight factors and constraints. While there may be some truth to this statement, it is also true that IMRT allows one to deliver highly conformal concave dose distributions that are impossible to achieve with conventional 3D techniques.

17.2.2 Variables to Be Optimized

After determining the objective function and the set of constraints, the next question is about the variables to be optimized. Even with the most powerful modern computers it is impossible to optimize all treatment parameters (variables). In practice, a subset of the treatment parameters, including, for example, the beam angles, need to be manually pre-selected.

17.2.2.1 Intensity Maps

In IMRT the main variables to be optimized are obviously the intensity maps for each beam. Each beam is typically subdivided into beam elements (bixels) of 5×5 to 10×10 mm². The intensity (fluence) for each of the bixels is optimized. The total number of bixels for all beams is typically of the order of 1000–10,000. Because there is no way to deliver intensity-modulated photon beams directly with a linac, the intensity maps are then converted to a series of multileaf collimator (MLC) shapes (segments) in a second almost independent step, which is called leaf sequencing. Of course, there has to be some link between optimization and sequencing. For example, the optimizer must know the leaf width of the MLC and should use that as the bixel size in one dimension. More thorough approaches for the consideration of delivery constraints in intensity map optimization have also been suggested (Cho and Marks 2000; Alber and Nüsslin 2001).

Besides the common two-step approach, it has also been suggested to avoid the intermediate step of using intensity maps altogether and directly optimize MLC shapes (apertures) and their weights. This approach has been suggested by DeNeve et al. (1996), among others. The MLC shapes can be determined manually based on the geometry (anatomy) of the problem, or they can be directly optimized together with the weights of the segments. The latter method has recently been published (Shepard et al. 2002). The direct optimization of MLC shapes and weights is mathematically a difficult, non-convex problem.

17.2.2.2 Number of Beams

The question of how many intensity-modulated beams should be used is highly relevant for the practical delivery of IMRT. As a consequence of the analogy between image reconstruction in computed tomography (CT) and inverse radiotherapy planning, the early theoretical approaches to inverse planning assumed a very high number of coplanar beams (Brahme et al. 1982; Cormack and Cormack 1987). Later it was recognized that quite acceptable results can also be achieved with a moderate number of beams. In fact, some publications claim that one generally does not need more than three intensity-modulated beams to obtain results that can hardly be improved any more (Söderström and Brahme 1995). These issues have been discussed frequently in the literature (Brahme 1993; Brahme 1994; Mackie et al. 1994; Mohan and Ling 1995; Mohan and Wang 1996; Söderström and Brahme 1996).

In principle, it is clear that the higher the number of beams, the higher is the dose conformation potential; however, the incremental improvement of the conformity of the total dose distribution diminishes as more beams are added. The real question about the “optimum” number of beams in IMRT is therefore: What is the number of beams beyond which one does not see any practically relevant improvement of the treatment plan? Several authors have found independently that it is hardly ever necessary to use much more than ten intensity-modulated beams to achieve results that are close to optimum (Bortfeld et al. 1990; Webb 1992; Söderström and Brahme 1995; Stein et al. 1997).

On the other extreme, with a number of beams as small as three or four, the conformity of the resulting dose distribution is considerably reduced when compared with a plan incorporating ten beams; however, thanks to intensity modulation, it is still
possible to imprint the desired shape on at least one isodose curve, e.g., the 80% isodose. This finding is related to image reconstruction of homogeneous objects using only three or four projections (Natterer 1986); hence, in cases where the tolerance of the critical structures is not too low as compared with the required target dose, one may get away with very few intensity-modulated beams (Söderström and Brahme 1995).

17.2.2.3

Beam Angles

The question about the optimum beam angles is related to the number of beams. Clearly, if very many beams (more than ten) are used, they can be placed at evenly spaced angular intervals and there will be no need to optimize beam orientations. Even with a moderate number of beams of the order of seven or nine, one may often use evenly spaced beams without compromising the dose distribution (Bortfeld and Schlegel 1993); however, it has been shown that this is case dependent, and complex cases, such as head and neck, sometimes benefit from beam orientation optimization even for nine or more beams (Pugachev et al. 2001). With very few beams, such as four or less, very careful and time-consuming optimization of beam orientations is always essential; otherwise, one will not be able to achieve acceptable results.

It should be noted that optimal orientations of intensity-modulated beams are generally different from those of uniform beams: in IMRT it is not generally necessary and often not even advantageous to avoid beam directions through organs at risk, because these can be spared by reducing the intensity for the corresponding rays (Stein et al. 1997). Also, for similar reasons, non-coplanar beams are rarely used in IMRT. Another point worth mentioning is that parallel opposed beams should be avoided in IMRT, such that, for evenly spaced beams, the number of beams should be odd. The reason for this is that a parallel opposed beam adds much less beam-shaping potential than a slightly angled beam. Clearly, if the attenuation were zero, parallel-opposed beams would be completely useless.

17.2.2.4

Number of Intensity Levels

Most IMRT planning methods assume a continuous modulation of the intensity. Several investigations have shown that promising results can be achieved with step-like beam profiles as well (Bortfeld et al. 1994; Gustafsson et al. 1994; DeNeve et al. 1996). In fact, using a moderate number of stair steps with five to seven “intensity levels” in each beam profile, the results are almost as good as with continuous modulation (Keller-Reichenbecher et al. 1998). Consequently, it is not necessary to go to a fully dynamic treatment mode to perform IMRT with an MLC. The IMRT can instead be realized in a “step-and-shoot” mode, i.e., by the successive delivery of a number of static MLC-shaped beam segments from each direction of incidence. The total number of beam segments or “subfields” to be delivered in this way is of the order of 100. Modern treatment machines can deliver such a sequence of subfields automatically and quickly. A comparison of the features of dynamic vs step and shoot IMRT has been published by Chui et al. (2001).

17.2.2.5

Beam Energy

The choice of the beam energy is less critical in IMRT than in conventional radiotherapy. In fact, it was suggested that very low energies around 1 MeV or less suffice in IMRT. The reason is once again that in IMRT one tends to spread the beams more evenly around the patient, and the depth-dose fall-off is therefore not very relevant.

17.2.3

Optimization Algorithms

The objective function as a function of the variables to be optimized, in combination with the optimization constraints, defines the optimization problem. As for its solution, many mathematical algorithms have been developed to solve optimization problems of various kinds. The algorithm of choice depends on the type of the problem (linear/non-linear, convex/non-convex). We cannot describe these algorithms in any detail in this chapter. The interested reader is referred to other reviews (Bortfeld 1999; Shepard et al. 1999).

The most common approach in commercial IMRT optimization systems is the non-linear (often quadratic) problem formulation. The algorithm used for finding the solution is usually a variant of the “gradient” technique. It converges rapidly and can be applied to a wide range of optimization problems; however, in the non-linear case one does not usually let the algorithm converge to the numerical optimum,
but stops after an acceptable number of iteration steps. Statements about the proximity to the true optimal solution are more difficult to make than in the linear case, or impossible.

There is no doubt that in general optimization the biggest experience exists in the field of linear optimization algorithms ("linear programming"). Even though many of the IMRT objectives and constraints discussed above are non-linear, a linear model or its variants can approximate the problem sufficiently well to yield useful solutions (Langer et al. 1996; Romeijn et al. 2003). An advantage of the linear problem formulation is that the optimality of the solution can be proven.

17.3 Inverse Planning in Practice

17.3.1 General Approach

Most planning systems used in clinical practice presently were not designed from ground up for intensity-modulated radiotherapy. Instead, the IMRT optimization engine is usually implemented as a separate program that communicates with the conventional three-dimensional forward treatment planning system (TPS). Since many steps of the complete planning process are the same for both forward and inverse planned treatments, this is not necessarily a disadvantage.

A general approach to radiotherapy planning, including IMRT, can be formulated as follows:

Step 1. The contours of target structures and organs at risk are outlined in the TPS (using the appropriate imaging modalities such as CT, MRI, and PET).

Step 2. In some cases, at this stage it might be unclear whether conventional unmodulated fields are sufficient or whether IMRT is necessary to achieve a satisfying dose distribution. In these cases one should first try to find a good treatment plan with open fields by varying number, direction, and weight of the beams and their individual shape defined by the MLC leaves. E.g., for intracranial meningiomas a 3D-planned, conventional treatment with uniform fields usually is sufficient. But some meningiomas are located directly next to several critical structures, such as the optical nerves and the brain stem, so it may turn out that curative target doses with open fields would lead to unacceptable doses in those organs at risk, and IMRT should be used instead.

There are also clinical cases where even complex shaped targets can safely be treated with conventional techniques by skillfully combining different radiation modalities (photons and electrons), wedges, individual blocks, and non-coplanar beam angles (Yajnik et al. 2003); however, the planning and delivery of those treatments might be even more time-consuming than an IMRT treatment with a comparable dose distribution. So even from an economic point of view there can be an indication for IMRT. But of course, in general IMRT is more time-consuming and more expensive than conventional radiotherapy and should only be applied if a clinical benefit for the patient can be expected.

Step 3. From here on we assume that open fields did not lead to a satisfying dose distribution or the geometry of the planning problem obviously is too complex for open field treatment. The IMRT module has to be started from within the TPS. It imports the CT data, the beam configuration, and the organ contours from the TPS.

Step 4. Now the parameters needed for the optimization have to be defined, namely the dose prescription and weight factors for each structure. Figure 17.1 shows the input window of the inverse planning program KonRad (distributed by Siemens OCS). Please note that a priority is assigned to structures which have some overlap with other structures. In the region of overlap the optimization engine is using the settings of the structure with the higher priority.

Step 5. Then the optimization is carried out. Depending on the specific case and on the program used, this process takes from less than a minute up to hours. In the end the inverse planning program presents the resulting treatment plan. It is evaluated with regard to the dose distribution in the target structures and the organs at risk using DVHs and full 3D dose information in form of isodoses or color-wash displays. If the actual treatment plan is not acceptable in one or more parts, the optimization has to be restarted with modified start parameters. The steering parameters are both the dose constraints and the weight factors. In the case of step-and-shoot delivery, the planner also has to check the number of segments of the plan. Too many segments can lead to unacceptable delivery times, making it necessary to reduce the number of intensity levels or to activate profile smoothing.

Hunt et al. (2002) have systematically investigated the interplay of the optimization parameters for a simple test case consisting of a concave-shaped target structure (PTV) around an organ at risk (normal tissue, NT) shown in Fig. 17.2a. The evaluation crite-
rion for the PTV was the Uniformity Index defined as maximum dose divided by 95% dose level, and for the NT the maximum dose was used. Figure 17.2b shows the impact of the NT dose constraint onto the PTV: especially for small distances between PTV and NT, a demanding NT dose constraint can greatly deteriorate the target dose. Note that, in this figure, a large PTV uniformity index corresponds with a highly non-uniform dose. In Fig. 17.2c, the influence of the weight factors is shown: for different NT dose constraints from 10 to 70% of the prescribed target dose, the associated weight factor was varied from 10 to 1000. The arrowhead in Fig. 17.2c indicates the direction of increasing weight factors: higher weight factors led to worse PTV dose uniformity in all cases, but only for the 10% dose constraint did they also improve the dose to the normal tissue. Of course, the results depend on the specific inverse planning system, but it is a general observation in inverse planning that the sensitivity of a solution, i.e., the impact of changes to the optimization parameters onto the optimization result, greatly varies and can only be found out by trial and error.

Based on the analysis of the above test case and their experience in clinical practice, Hunt et al. (2002) formulated a strategy for inverse planning (Fig. 17.3).
Depending on PTV dose uniformity or normal tissue dose, they propose different changes to the constraint and penalty settings. In clinical practice there will be modifications to this scheme, depending on the planning system actually used, but it gives a good general overview over the parameters used for steering the optimization result.

Step 6. Once an acceptable IMRT plan has been found, the treatment plan can be exported to the delivery machine for verification and application purposes. Depending on the accuracy of the dose calculation used for the IMRT optimization, it might be necessary to recalculate the dose in the TPS.

### 17.3.2 Clinical Example

In this section an IMRT treatment of a head and neck tumor is demonstrated. The patient, a 63-year-old woman, was diagnosed with lymphoepithelioma originating from the left eustachian tube. The macroscopic tumor spread, as seen in the MRI image, was 3×1.5×2 cm³. No positive lymph nodes or distant metastases were found. Concomitant to the radiotherapy, the patient received chemotherapy consisting of 5-FU and Cisplatin.

On the left side of Fig. 17.4, the structures of this case are shown as they appear in the observer’s view of the TPS. Two different target structures were defined: the uninvolved lymph-node-bearing tissue ("Target") that was electively treated because the tumor is known to have a high risk of lymphatic dissemination; and the macroscopic tumor volume ("GTV"). Both parotid glands are very close to the target contour. Also the eyes with optic nerves and chiasm, the brain stem, the spinal cord, the esophagus, and both lungs are in immediate vicinity to the target volume and have therefore been considered in the IMRT optimization.

The right side of Fig. 17.4 shows an exemplary CT slice of the final IMRT plan obtained with the inverse planning program KonRad. One can see an additional contour called "Target + 10 mm." This contour was defined solely for better control of the dose gradient between the high-dose area and the normal tissue during the optimization (such pseudostructures can be used whenever an IMRT plan shall be altered in circumscribed areas of the irradiated volume). The median dose to the GTV was normalized to 66 Gy, and the target volume received a median dose of 54 Gy. Treatment was delivered in 30 fractions, resulting in a single fraction dose of 2.2 Gy to the GTV and 1.8 Gy to the Target. This technique is also called “simultaneously integrated boost (SIB).” The SIB is a unique feature of IMRT treatments which simplifies treatment and potentially leads to higher local control rates (MOHAN et al. 2000; LAUVE et al. 2004). Another outstanding feature of IMRT treatments for head and neck tumors is the sparing of salivary glands. In the example given here, the mean dose to the right parotid could be restricted to 22 Gy (the left parotid gland, located near the macroscopic tumor, receives a mean dose of 37.4 Gy). One year after treatment, the patient still does not suffer from xerostomia, an inevitable side effect of conventional head-and-neck
radiotherapy that can greatly reduce the quality of life. Especially in cases with a very good long-term prognosis, like in this example, such considerations are important. Also the doses to all other organs at risk in this case could be kept well below the tolerance level.

17.4 Multicriteria Optimization in Inverse Planning

In the previous sections we saw that inverse planned IMRT is already in use in clinical practice; however, it also became clear that “inverse planning” as done today leaves a lot to be desired, and a lot of research effort is still put into this field. Presently, inverse planning is still a recursive approach that might include several optimization runs and can therefore be quite time-consuming. Setting the optimization parameters is non-intuitive. The penalty factors are artificial parameters that do not have a clinical meaning, and even the constraints, which should stand for the tolerance doses of the particular organ at risk, are often used as mere steering parameters when they cannot be fulfilled in every voxel of the structure anyway. The impact of changes to the parameters cannot be seen a priori. Instead, a new optimization run is needed, making changes to the plan non-interactive. The planning might stop when a certain time limit has been reached, not necessarily when the best treatment plan for the patient has been found.

A new optimization paradigm, the multicriteria optimization, is a promising concept to overcome these problems. In the following we present the joint effort of the Massachusetts General Hospital (Boston, Mass, USA), the Fraunhofer Institute for Industrial Mathematics (Kaiserslautern, Germany), and the German Cancer Research Center (Heidelberg, Germany; Küfer et al. 2000; Bortfeld et al. 2002; Küfer et al. 2003; Thieke 2003).

To motivate multicriteria optimization, it is helpful to step back and look at the original optimization problem: There is a target structure which should be treated with full dose, and there are organs at risk nearby that should receive as little dose as possible. But even with IMRT, due to the physical properties of photon beams (namely, the depth-dose curve with an exponential tail and the penumbra), it is impossible to deliver the ideal dose distribution, which is 100% in the target and 0% everywhere else; therefore, every treatment plan is a certain compromise between these conflicting goals. Mathematically speaking, the optimization problem does not have a single solution where all criteria are at their individual optimum at the same time. This is exactly the definition of a multicriteria optimization problem. Instead of a single optimum, there is a whole solution space of optimal compromises, also called Pareto optimal solutions. A compromise is Pareto optimal when it cannot be improved in one criterion without worsening at least on other criterion. The whole set of Pareto optimal solutions is called the Pareto front.

Fig. 17.4. Organ contours and dose distribution of the IMRT head and neck case
The result of multicriteria optimization for inverse planning is therefore no longer a single treatment plan. Instead, a whole database of Pareto optimal treatment plans is generated. Each plan represents a certain clinical compromise, e.g., plan A of the database might have a lower dose in an organ at risk than plan B. But since every plan is Pareto optimal, plan A will also deliver more dose to another organ at risk and/or less dose to the target than plan B. Which compromise is the best for the patient is a clinical decision that cannot be made by the computer. Different clinicians might even choose different plans because they prefer different balances between tumor dose and normal tissue sparing. The decision for a particular plan is made by interactively browsing the database of plans with an intuitive graphical interface.

In order to rank different dose distributions in the target and organs at risk, a costlet function has to be defined separately for each structure. Options are the dose-based and dose-volume-based functions, as well as the equivalent uniform dose (EUD) discussed above, or models of tumor control probability (TCP) and normal tissue complication probability (NTCP). In multicriteria optimization, these costlets are evaluated separately; they are not simply added up to yield a single score. We propose the use of EUD because it is clinically meaningful, is still in the dose domain and therefore familiar to the treatment planner, and is differentiable and convex which allows for efficient optimization algorithms.

Inverse planning for radiotherapy using the multicriteria optimization program differs significantly from the current trial-and-error approach. In a first step, the database of Pareto optimal plans, based on the individual patient geometry and contours, is generated. This can take up to several hours, which is not a big issue since no human interaction is required during the calculation. In the second step, the graphical user interface (Fig. 17.5) comes up. The planner browses through the database interactively using this interface.

![User interface of the interactive navigator (developed by ITWM, Kaiserslautern)](image-url)
The user interface is divided into two main windows: the left panel shows the part used for the navigation. Here the complete database is visualized by a star with one axis for each structure of the treatment plan. The gray area shows all EUD values that can be reached for each organ. This way the complete planning horizon becomes apparent at first sight. The planner does not have to find out by trial and error what dose might be achievable in a certain critical structure, instead he just looks at the extreme values on the EUD navigation axes. In addition, the current treatment plan is represented by a polygon connecting the EUD values of all structures. The right panel visualizes the current plan as dose volume histogram and as 3D dose distribution in frontal, sagittal, and transverse projection. All plans are normalized to the same target EUD, and the planning horizon can be changed by the global normalization factor.

At the beginning, the system suggests a starting solution. If one or more aspects of the plan are not desirable, e.g., the EUD in one critical structure is too high, the planner can immediately go to another solution in which that specific criterion is fulfilled. This is done by grabbing the marker at the EUD bar of the particular organ and dragging it in the desired direction. Instantaneously, the system finds the corresponding solution in its database and updates the information for the other structures and in the DVH and isodose windows.

In our opinion the advantages of this concept are threefold:

• Artificial weight factors, which have no clinical meaning, are avoided. The whole concept is based on dose-like values, which are amenable to a clinical interpretation.
• Unnecessarily high doses in some of the critical structures, which can occur in constrained optimization, are avoided by definition of the Pareto optimal solution.
• Plan tuning can be done interactively using “knobs” that have a clinical meaning. It is easy to do a sensitivity analysis and determine the dependency of, for example, the target EUD on any of the critical-structure EUDs.

Multicriteria optimization for radiotherapy planning as shown in this chapter is still under development, but when it becomes available in clinical routine it may change the way radiotherapy planning is done in the future.

17.5 Conclusion

Optimized inverse planning can yield superior treatment plans, especially in complex situations with convex–concave target volumes and nearby critical structures; however, the optimization criteria must be carefully chosen. Determining appropriate optimization criteria is not straightforward and requires some trial and error in a “human iteration loop.” Using current commercial inverse planning systems this process can be quite time-consuming. Experienced treatment planners know how to steer an IMRT plan in the desired direction by appropriately changing the optimization criteria. Also, class solutions can help to avoid or reduce the “human iteration loop” in cases that do not vary too much between individuals, such as prostate treatments, because optimization criteria can be re-used. Nevertheless, plan optimization leaves something to be desired. The main problem is that it may not be possible to come up with a quantitative, complete optimization formulation for radiotherapy planning in the near future; however, an achievable alternative is to design optimization systems that let the physicians exercise their experienced clinical judgment or intuition in the most direct interactive way. Therefore, some future developments aim at a more interactive approach towards inverse planning. Multicriteria optimization and navigating a treatment plan database have been described as promising approaches in this context.

References

Birkhoff G (1940) On drawings composed of uniform straight lines. J Math 19:3


Thieke C (2003) Multicriteria optimization in inverse radiation therapy planning, Doctor of Natural Sciences dissertation, Combined Faculties for the Natural Sciences and for Mathematics, Ruperto-Carola University of Heidelberg


18 Biological Models in Treatment Planning

Christian P. Karger

CONTENTS

18.1 Basic Parameters of Radiation Tolerance 221
18.1.1 Tumour Control and Normal Tissue Complications 221
18.1.2 Determination of Dose-Response Curves 221
18.1.3 Implicit Dependencies of Dose-Response Curves 222
18.2 Aims of Biological Models 222
18.3 Fractionation Regime and Treatment Time 223
18.3.1 Historical Models 223
18.3.2 The Linear-Quadratic Model 223
18.3.3 Extensions to the Linear-Quadratic Model 224
18.4 NTCP Models 225
18.4.1 The Lyman-Kutcher Model 225
18.4.2 The Critical Element Model 226
18.4.3 The Critical Volume Model 227
18.5 TCP Models 228
18.6 RBE of High-LET Radiation 229
18.6.1 Protons 229
18.6.2 Heavy Charged Particles 230
18.6.3 Neutrons 231
18.7 Clinical Relevance of Biological Models 231
18.7.1 The Linear-Quadratic Model 231
18.7.2 NTCP and TCP Models 231
18.7.3 RBE of High-LET Radiation 232
18.8 Conclusion 233
References 233

18.1 Basic Parameters of Radiation Tolerance

18.1.1 Tumour Control and Normal Tissue Complications

The aim of radiotherapy is to give sufficient dose to the tumour to achieve local control without introducing severe complications in the surrounding normal tissue. These conflicting aims can be quantitatively described by dose-response curves for the tumour and normal tissue, respectively (Fig. 18.1). With increasing dose to the tumour, the tumour control probability (TCP) also increases. Dose escalation, however, also rises the normal tissue complication probability (NTCP), which frequently is the limiting factor in clinical situations. In the region between both curves (denoted as “therapeutic window”), the probability of tumour control without normal tissue complications reaches a maximum at the optimum dose $D_{opt}$. If type and probability of the related complications are not acceptable, however, this optimum dose may not be feasible to be applied in clinical situations and the probability for tumour control will therefore be even lower.

Fig. 18.1 Dose-response curve for tumour control probability (TCP) and normal tissue complication probability (NTCP). The maximum probability for tumour control without normal tissue complications (dashed line) is reached at the optimum dose ($D_{opt}$).

18.1.2 Determination of Dose-Response Curves

In experimental or clinical situations, dose-response data are obtained in terms of incidence rates $x/n$ ($x$ out of $n$ subjects show the selected endpoint) at several dose levels. As incidence rates are binomi-
ally distributed and show a large spread for a small number of subjects, \( n \), an analytical curve is adjusted to the data using a maximum likelihood fit. Although several parameterisations may be used to describe dose-response curves mathematically (Källman et al. 1992; Niemierko and Goitein 1991), the following one is frequently used because of its simplicity:

\[
P(D) = \frac{1}{1 + \left( \frac{D_{50}}{D} \right)^k}
\]

(1)

\( P(D) \) gives the expectation value of the probability that the selected end point occurs at the dose level \( D \). The parameter \( D_{50} \) (often referred to as tolerance dose, \( TD_{50} \), or equivalent dose, \( ED_{50} \)) is the dose at which an effect probability of 50% is expected. The parameter \( k \) is related to the slope of the dose-response curve.

Most statistical software packages supply the logistic formula for parameterisation of dose-response curves, which is not equivalent to Eq. (1). The parameters \( D_{50} \) and \( k \) of Eq. (1) as well as their standard errors, however, may be calculated from the fit of the logistic function, if \( \ln(D) \) instead of the dose \( D \) is used as independent variable (Karger and Hartmann 2001).

Tolerance doses for normal tissue complications may be determined with reasonable accuracy in animal experiments. The question of whether these data also apply to humans, however, remains an intrinsic problem, and tolerance doses from animal experiments are generally not believed to be directly transferable to humans. In humans, on the other hand, the parameter \( D_{50} \) usually cannot be determined as such high complication rates are usually prevented by clinical experience. For clinical applications, quantities such as \( D_5 \) (dose leading to 5% complication probability) are more relevant. From such low complication probabilities, however, it is nearly impossible to determine the slope of the curve.

Emami et al. (1991) published a compilation of data for \( D_5 \) and \( D_{50} \) for a variety of normal tissues. Although the authors stressed the limited accuracy of these data, no significant improvements in knowledge was achieved up to now and most articles still refer to these data.

18.1.3 Implicit Dependencies of Dose-Response Curves

Figure 18.1 suggests that the effect probability is solely a function of the radiation dose. The dose-response curves, however, implicitly depend on several biological and physical parameters. At first, the biological end point for the tissue response has to be specified, including the time after radiation used for follow-up. For normal tissue complications, the dose-response curves will be different for early and late effects and even within one type of complication the curves may differ, depending on the exact end point definition and the method of investigation. In the same way, the curve for the tumour may depend on the definition of tumour control, e.g. whether local control or remission of the tumour is regarded.

In the clinical situation, relevant end points have to be selected, e.g. radiological tumour control and early or late complications which are definitely to be prevented.

Besides the definition of the biological end point, the dose-response curves are strongly influenced by various treatment parameters, most of them being of physical nature. The most important parameters are given by the time pattern of the applied dose, the volume of irradiated normal tissue and the radiation quality used for treatment. Since the very beginning of radiotherapy treatments, it was the aim of many investigations to optimise the physical treatment parameters to improve clinical outcome, i.e. to open the “therapeutic window” between the dose-response curves for tumour control and normal tissue complications (Fig. 18.1).

18.2 Aims of Biological Models

The aim of biological models is to predict the radiation response of biological systems. While early approaches focused on modelling the radiation response for different fractionation schemes, newer developments attempt to model effect probabilities (TCP and NTCP), their volume dependence and the relative biological efficiency (RBE) of radiation with high linear energy transfer (LET).

In treatment planning, biological models may be applied with different intentions:

1. Transfer of one treatment regime to a biologically iso-effective new regime or new radiation modality without predicting absolute values for TCP or NTCP.
2. Calculation of TCP or NTCP values to compare either competing treatment plans for an individual patient or different treatment techniques for a specified clinical application. In this case, the TCP/
NTCT values are not expected to be completely correct in their absolute values, but it is believed that they can be used as rationale to prefer one treatment plan (or technique) over another.

3. Prediction of absolute TCP or NTCP values for individual patients.

4. Integration of TCP/NTCP models into the cost function of the dose optimisation algorithm to generate biologically optimised treatment plans.

18.3 Fractionation Regime and Treatment Time

Among the various influence factors on TCP and NTCP four factors have always been considered to be most important. These factors are denoted as the four “R”s of radiotherapy (Withers 1992; Thames et al. 1989) and include repair of sublethal cell damage, repopulation of tumour cells, redistribution of cells over different cell cycle phases with different radio-sensitivity and reoxygenation of radio-resistant hypoxic tumours after beginning of the radiotherapy course. Repair and repopulation have been identified as the most important factors with respect to radiation response and were therefore subject of early models.

18.3.1 Historical Models

Repair and repopulation are related to the fractionation regime, i.e. to the dose per fraction, the number of fractions and the overall treatment time. Historically, several models have been developed on the basis of skin data (Barendsen 1982; Ulmer 1986) to describe iso-effect relations for different treatment regimes. Examples for these models are the nominal standard dose (NSD) model and its derivates denoted as the partial tolerance (PT), the time, dose, fractionation (TDF) and the cumulative radiation effect (CRE) model. These models, however, were criticized for several reasons (Barendsen 1982; Fowler 1984, 1989, 1992; Thames 1982). Using this model, the survival fraction of cells irradiated with a single dose \(d\) is described by

\[
SF = e^{-\alpha d - \beta d^2},
\]

(2).

\(\alpha\) and \(\beta\) are parameters measuring the amount of lethal and sub-lethal cell damage, respectively. In a logarithmic representation, the survival curve shows an initial linear decrease at low doses followed by a "shoulder" for which the bending is determined by the ratio \(\alpha/\beta\) (Fig. 18.2).

18.3.2 The Linear-Quadratic Model

The linear-quadratic model was introduced to replace the former iso-effect relations for different fractionation regimes (Barendsen 1982; Fowler 1984, 1989, 1992; Withers 1992). Using this model, the survival fraction of cells irradiated with a single dose \(d\) is described by

\[
SF = e^{-\alpha d - \beta d^2},
\]

(2).

\(\alpha\) and \(\beta\) are parameters measuring the amount of lethal and sub-lethal cell damage, respectively. In a logarithmic representation, the survival curve shows an initial linear decrease at low doses followed by a "shoulder" for which the bending is determined by the ratio \(\alpha/\beta\) (Fig. 18.2).

Fig. 18.2. Cell survival curves for tumour and late-reacting normal tissue. The curves are displayed for single and fractionated treatment, respectively. For single doses the curves for tumour and normal cells intersect each other, leading to lower survival fractions for the normal cells. For a fractionated treatment, however, the survival fraction for tumour cells always remains below the one for normal cells. A fractionated treatment therefore spares late responding normal tissue.
The smaller \( \alpha/\beta \) is, the more pronounced the shoulder of the survival curve will be. If the time between two fractions exceeds the minimum of about 6 h (Fowler 1989), the sub-lethal damage can be repaired and the shape of the cell survival curve will be reproduced for the subsequent fraction (Fig. 18.2). After \( n \) fractions of equal doses, \( d \), the survival fraction will then be \( e^{-n(ad + \beta d^2)} \). As the survival level is determined by \( n(ad + \beta d^2) \), a biologically effective dose (also termed as extrapolated response dose, ED) may be defined (Fowler 1989, 1992; Prasad 1992) by the equation

\[
BED = nd(1 + \frac{d}{\alpha/\beta}).
\] (3)

Two fractionation regimes having the same BED are considered to be iso-effective. As a consequence, one fractionation regime may be converted to another iso-effective regime using the relation

\[
n_1d_1 = n_2d_2, \quad \frac{\alpha/\beta + d_1}{\alpha/\beta + d_2}
\] (4)

\( n_1, n_2, \) and \( d_1, d_2 \) are the number of fractions and the doses per fraction for both regimes, respectively. It is a great advantage of the linear-quadratic model that only the ratio \( \alpha/\beta \), rather than the absolute values of \( \alpha \) and \( \beta \) are required in Eq. (4).

Values for \( \alpha/\beta \) can be measured for in vitro as well as in vivo systems. Although the same cell type may be involved, the \( \alpha/\beta \)-values may be different as cells in intact tissue are not expected to respond independently from their physiological environment. As in vivo settings do not allow to determine the fraction of surviving cells, \( \alpha/\beta \) has to be measured by equating the BED of two different fractionation regimes according to Eq. (3). The required iso-effective total doses may be obtained from the dose-response curve of the respective regime (e.g. the tolerance doses \( D_{50} \)). The resulting equation can then be resolved for \( \alpha/\beta \). Alternatively, the inverse of the total dose \( (nd)^{-1} \) may be plotted against the dose per fraction \( d \) (referred to as Douglas-Fowler plot) for more than two iso-effective fractionation regimes (Barendsen 1982; Fowler 1984; Thames 1982). According to Eq. (3), \( \alpha/\beta \) may then be determined from a linear regression to the data.

The value for \( \alpha/\beta \) does not only depend on the irradiated type of tissue, but also on the considered biological end point. Values for \( \alpha/\beta \) are given in the literature for various tissues and end points (Barendsen 1982; Fowler 1984, 1989; Thames et al. 1989; Withers 1992). Typical \( \alpha/\beta \)-values are around 3 Gy for late reactions and around 10 Gy for early reactions. Most tumours show \( \alpha/\beta \)-values of 10 Gy or more. As a consequence, late reacting tissue can be spared by fractionated treatment relative to the tumour response, while the fractionation effect is small or may even be neglected for early reacting tissues. For these tissues the overall treatment time is more important.

Although attempts have been made to derive Eq. (2) as an approximation from mechanistic considerations (Gilbert 1980), the linear-quadratic model is mostly regarded as an empirical parameterisation. The linear-quadratic model has been validated in the dose range of about 2–8 Gy per fraction (Fowler 1984, 1989; Withers 1986, 1992). At higher doses some experimental cell survival curves asymptotically approach to a purely exponential shape and deviations to the linear-quadratic model have been seen. This behaviour may be explained by multi-hit killing from accumulation of multiple sub-lethal events (Withers 1992). Although there are other parameterisations for cell survival curves (e.g. the two-component model), which account for the purely exponential shape at high doses, the linear-quadratic model is usually preferred since the derived iso-effect relation of Eq. (4) depends only on the single parameter \( \alpha/\beta \) (Withers 1992). Because of the limitation of the linear-quadratic model at high doses, Eq. (4) should not be used to transfer a fractionated treatment to an iso-effective single dose treatment, although it has previously been used as approximation (Larson et al. 1993; Prasad 1992; Flickinger et al. 1990).

### 18.3.3 Extensions to the Linear-Quadratic Model

Several extensions have been developed for the linear-quadratic model to account for additional dependencies of the radiation response. As repopulation may play an important role for early reacting tissues, a time factor has been introduced:

\[
SF = e^{(\alpha d + \beta d^2) \gamma (T - T_k)}
\] (5),

where \( \gamma \) is related to the average doubling time \( T_p \) of the cells by \( \gamma = \ln 2/T_p \) and \( T_k \) is the kick-off time until proliferation starts (Fowler 1989, 1992). For late effects the time factor can be neglected and in this case Eq. (5) reduces to Eq. (2). Equation (5) leads to a modified expression for the BED given by

\[
BED = nd(1 + \frac{d}{\alpha/\beta}) \gamma \frac{(T - T_k)}{\alpha}.
\] (6)
The apparent disadvantage of this extended model is that the absolute values of $\alpha$ is required which is not always available with sufficient accuracy. In addition, the values of the two parameters $\gamma$ and $T_F$ have to be known.

As there is some evidence that $\alpha$ and $\beta$ depend on the cell cycle phase, the linear-quadratic model was extended to account for this heterogeneity (Schultheiss 1987). To do so, three additional parameters are introduced, describing the spread and the correlation of the parameters $\alpha$ and $\beta$. In an other approach it was attempted to give a complete description for repair, repopulation, redistribution and re-oxygenation (Brenner 1995). This model also needs five input parameters.

A special situation is given by the application of brachytherapy techniques, since the irradiation is performed either continuously over several days or in a fractionated fashion, where the time between the fractions is in the order of one hour. In both cases, the sub-lethal damage may not completely be repaired. Several investigations have been performed either continuously over several days or in a fractionated fashion, where the time between the fractions is in the order of one hour. In both cases, the sub-lethal damage may not completely be repaired.

In Eq. (9) the abbreviations $x = e^{-T_F/t_0}$ and $y = e^{-\Delta t/t_0}$ have been used. For a standard fractionation regime with $T_F << t_0$ and $\Delta t >> t_0$, $G$ reduces to $1/N$ and Eq. (7) then will be reduced to $SF = e^{-mud + vD^2}$, which is the product of the survival fractions of $n$ biologically independent fractions.

### 18.4 NTCP Models

The NTCP models aim to describe the complication probability in normal tissues in terms of dose-response curves. As there is extensive evidence, that the radiation response of normal tissue depends on the amount of irradiated normal tissue (Burman et al. 1991; Cohen 1982; Emami et al. 1991; Flickinger et al. 1990; Schultheiss 1983; Withers et al. 1988), the irradiated volume is included as an additional important parameter. The extent of the volume effect is dependent on the architecture of the respective tissue and several models have been proposed. While some of them are only of phenomenological nature, others include more basic bio-statistical principles.

#### 18.4.1 The Lyman-Kutcher Model

A four-parameter model was proposed by Lyman (Lyman 1985). In this model, the complication probability $P(D, V)$ for a uniform irradiation of a normal tissue volume $V$ with a dose $D$ is given by (Burman et al. 1991; Burman 2002; Lyman 1985; Kutcher 1996):

$$P(D, V) = \frac{1}{\sqrt{2\pi}} \int e^{-x^2} dx$$

where $x = \frac{D}{TD_{50}(V)} - 1$.

$$TD_{50}(V) = TD_{50}(1) \cdot V^n$$

The four parameters of the model are given by $TD_{50}$, $m$, $n$ and $V_{ref}$, which have to be adjusted to
clinical data for each tissue type using a specified biological end point. \(TD_{50}(v)\) is the tolerance dose for the fractional volume \(v\), \(m\) is related to the slope of the dose-response curve, \(n\) describes the volume effect and \(V_{\text{ref}}\) is the reference volume to which the fractional volume refers to. \(V_{\text{ref}}\) may be chosen as the whole organ or as a part of it. Equation (10c) relates the tolerance doses of the partial volume \(v\) to that of the reference volume (\(v=1\)).

Emami et al. (1991) published tolerance doses for various tissues and fractional volumes which were derived from a literature search and from clinical experience. The authors considered the uncertainty of these tolerance doses to be rather high. Subsequently, the model parameters of Eq. (10) were adjusted to fit these tolerance data (Burman et al. 1991). As there were only few tolerance doses for each organ, the parameters were fit “by eye” rather than by using statistical methods; therefore, and due to the uncertainty of the underlying data, the derived model parameters have to be treated with great caution. Although some of the tolerance doses were refined later (Burman 2002), most of them have remained unchanged up to now.

In clinical practice, the normal tissue will not be uniformly irradiated as assumed by the Lyman model; therefore, the model was extended by introducing histogram-reduction algorithms, which transform the multi-step dose volume histogram obtained for a specific treatment plan into a biologically iso-effective single-step histogram, i.e. a non-uniform irradiation is transformed in an biologically iso-effective uniform irradiation. Two different types of reduction algorithms have been proposed which lead to similar although not identical NTCP-values:

The first one (Lyman and Wolbarst 1987, 1989) replaces the two rightmost bins (at doses \(D_n\) and \(D_{n-1}\), and volumes \(V_n\) and \(V_{n-1}\)) of the cumulative histogram by a single bin at dose \(D'_{n-1}\) and Volume \(V_{n-1}\). The dose \(D'_{n-1}\) is calculated such that the new histogram has the same NTCP according to Eq. (10). This procedure is iterated until a single-step histogram is achieved which corresponds to a homogeneous irradiation of the reference volume (\(v=1\)) with a dose \(D_1\) for which the Lyman model can directly be applied.

The second algorithm (Kutcher and Burman 1989; Kutcher et al. 1991) transforms the initial multi-step histogram (having the maximum dose \(D_{\text{max}}\)) to a biologically iso-effective single-step histogram with an effective volume \(V_{\text{eff}}\) at the dose \(D'_{\text{max}}\). For this approach, a volume effect according to Eq. (10c) is assumed. The single-step histogram then corresponds to a homogeneous irradiation of the fractional volume \(v_{\text{eff}}=V_{\text{eff}}/V_{\text{ref}}\) and the NTCP is then calculated by Eq. (10).

18.4.2

The Critical Element Model

The critical element model (Niemierko and Goitein 1991; Schulteiss 1983; Wolbarst 1984) assumes that an organ consists of a number of identical functional subunits (FSU; Withers et al. 1988), each of them responding independently to radiation. The term “critical element” means that it is additionally assumed that a complication occurs, if a single FSU is inactivated (Niemierko and Goitein 1991). The critical element model is expected to describe the radiation response for organs such as spinal cord, brain or bowel.

If \(P(D,v)\) is the complication probability that a dose \(D\) to the fractional volume \(v\) will produce a complication, \(1-P(D,v)\) is the probability that no complication occurs. If a whole organ consisting of \(N\) equal-sized compartments (each of volume \(v=1/N\)) is uniformly irradiated with a dose \(D\), the probability that the organ escapes injury \(P(D,1)\) is given by the product of the probabilities that each sub-volume escapes injury. \(P(D,v)\) can then be expressed by (Schulteiss 1983):

\[
P(D,v) = 1 - [1 - P(D,1)]^v
\]

Equation (11) may easily be generalized to non-uniform dose distributions \([D]\) by replacing the right side by \(1 - \prod_i [1 - P(D_{i,1})]^{v_i}\).

As the size of the product is strongly affected by the smallest factor, the size of the complication probability \(P(D,v)\) is governed by large values of \(P(D_{i,1})\), i.e. by the highest doses of \([D]\).

It follows from Eq. (11) that the dose-response curve for any partial volume irradiation can be calculated if the dose-response curve for the whole organ is known. No specific dose-response model has to be assumed.

It is a characteristic feature of the critical element model that the dose-volume iso-effect curve is determined solely by the slope parameter \((k\text{ in Eq. (1) and } m\text{ in Eq. (10b)}, \text{ respectively})\) of the dose-response curve for the whole organ (Niemierko and Goitein 1991). In contrast to this, the volume dependence in the Lyman model (Eq. (10c)) uses the additional parameter \(n\), which can be selected independently from the slope parameter \(m\). That means that in general,
the Lyman model describes a tissue architecture different from the one of the critical element model. In the approximation of \( P(D,1) < 1 \) (i.e. small doses, \( D << D_{50} \)), however, Eq. (11) yields \( P(D,v) = v \cdot P(D,1) \). If in addition the dose-response model of Eq. (1) is taken in the same approximation one obtains \( P(D,1) = \left( D / D_t \right)^n \). From these two relations, a dose-volume iso-effect relation can be derived (Schulteiss 1983):

\[
D_t = v^{-\frac{1}{n}} D
\]  

(12)

This relation is of the same structure as Eq. (10c). This means that the Lyman model is able to describe the dose-volume relation of tissue with critical element structure only for small complication probabilities (Niemierko and Goitein 1991). It also has to be pointed out that the histogram reduction methods (Lyman and Wolbarst 1987, 1989; Kutcher and Burman 1989; Kutcher et al. 1991) of the Lyman model implicitly make use of Eq. (10c) which is in general not valid for tissues with critical element structure. In this case the algorithms have to be adapted according to Niemierko and Goitein (1991).

The two dose-volume relations of Eqs. (11) and (12) were tested in animal experiments for the spinal cord and the brain, which both are considered to be of the critical element architecture (Niemierko and Goitein 1991; Schulteiss 1983). As expected, Eq. (11) was found to give a better description of the data. The critical element model was also used to calculate the complication probabilities and dose-volume iso-effect relations for radiosurgery treatments of the brain (Lax and Karlsson 1996; Flickinger 1989; Flickinger et al. 1990).

Although the publication of Schulteiss (1983) does not explicitly use the term “FSU”, the described model comprises all characteristic features of the critical element model. A more theoretical approach was presented by Wolbarst (1984), using the dose-response curve of a single FSU as starting point to model the radiation response of the entire organ.

**18.4.3 The Critical Volume Model**

The critical volume model describes tissues, where the FSUs of an organ are assumed to be arranged in a parallel fashion (Jackson et al. 1993; Niemierko and Goitein 1993a; Wolbarst et al. 1982; Yorke et al. 1993). In contrast to the critical element model, an inactivation of a single FSU will not lead to a complication in the organ as the organ function will be maintained by the remaining FSUs. If more than a critical number of FSUs will be inactivated, however, a complication will occur. This especially means that the organ tolerates any dose as long as the number of affected FSUs is below this threshold. The critical volume model is expected to describe the complication probabilities of organs such as the lung, kidney, liver, or parotid glands.

If an organ is assumed to consist of \( N \) parallel organized and independently responding FSUs, the probability that more than \( M \) FSUs are inactivated by an uniform dose \( D \) is given by (Niemierko and Goitein 1993a):

\[
P(D) = \sum_{k=M}^{N} \binom{N}{k} P_{FSU}^k (1 - P_{FSU})^{N-k},
\]  

(13)

where \( P_{FSU} \) is the dose-dependent probability for inactivating a single FSU. Equation 13 may also be generalized for inhomogeneous dose distributions (Niemierko and Goitein 1993a; Jackson et al. 1993; Yorke et al. 1993) leading to the concept of integral responding tissues (Wolbarst et al. 1982). As a consequence, the radiation response of parallel organized tissues should be governed by the mean rather than by the maximum doses as found for tissues of critical element structure.

For the special case of \( M=0 \), the critical volume model reduces to the critical element model (see previous section). In this case Eq. (13) can be written as:

\[
P(D) = (1 - P_{FSU})^N.
\]  

(14)

Another special case is given by \( M=N \), which means that all FSUs have to be inactivated to produce a complication. In this case, Eq. (13) yields:

\[
P(D) = P_{FSU}^N.
\]  

(15)

An example for this situation is a tumour where the FSU is identified with a single clonogenic cell. \( P_{FSU} \) then is the probability of inactivating one cell and \( P(D) \) is the probability of controlling the tumour, i.e. that all \( N \) clonogenic cells of the tumour are inactivated; therefore, with respect to the end-point tumour control, the critical volume model can be applied. With respect to the end-point tumour recurrence, however, tumours behave according to the critical element model. This can be seen by substituting the survival probability of the FSU by \( S_{FSU} = 1 - P_{FSU} \) and the recurrence probability for the tumour \( S(D) = 1 - P(D) \) into Eq. 15 which then results in an expression, which is formally identical to Eq. 14:
$S(D) = 1 - (1 - S_{FSU})^N. \quad (16)$

For larger values of N and M, the binomial distribution in Eq. 13 may be approximated by a normal distribution leading to (Niemierko and Goitein 1993a):

$$P(D) \approx \frac{1}{\sigma_{FSU} \sqrt{2\pi}} e^{-\frac{(x-NP_{FSU})^2}{2\sigma_{FSU}^2}} \quad (17a)$$

$$\sigma_{FSU} = \sqrt{NP_{FSU}(1-P_{FSU})} \quad (17b)$$

Evaluation of Eq. (17) requires the parameters N, M and the dose-response model $P_{FSU}(D)$ for a single FSU. As the dose-response of a single FSU is unlikely to be measurable, it is derived by basic statistical and biological considerations. Assuming that a FSU is composed of L clonogenic cells, and that the FSU will be able to regenerate if at least one clonogenic cell survives, the response curve is given by

$$P_{FSU}(D) = (1-SF^n)^L = (1-e^{-\alpha n d + \beta d})^L. \quad (18)$$

$SF$ is the surviving fraction according to the linear-quadratic model (Eq. (2)), characterized by the cell-specific parameters $\alpha$ and $\beta$. $n$ is the number of fractions, each of dose d ($D=nd$). A time factor for repopulation may be added, if necessary.

The five parameters of the model are now given by L, M, N, $\alpha$ and $\beta$. Using realistic values for these parameters, however, results in dose-response curves that are much steeper than those observed in patients (Yorke et al 1993; Niemierko and Goitein 1993a). The reason for this is considered to be the variation in radio-sensitivity among the patient population as well as the spread in sensitivity of the FSUs within an organ of an individual patient. This variation can be considered by replacing $P_{FSU}(D)$ by an averaged dose-response curve defined by

$$\bar{P}_{FSU}(D) = \int G^{ind}_{\alpha \beta L} P_{FSU}(D) d\alpha d\beta dL, \quad (19)$$

where $G^{ind}_{\alpha \beta L}$ is the (e.g. normal) distribution of the parameters $\alpha$, $\beta$ and L of the FSUs of the organ of an individual patient. The population-based complication probability $P^{pop}(D)$ is then obtained by averaging the NTCP values for the individual patient calculated by Eq. (17)

$$P^{pop}(D) = \int G^{pop}_{\pi \beta L M N} P^{ind}(D) d\pi d\beta dL dM dN, \quad (20)$$

where $G^{pop}_{\pi \beta L M N}$ now is the (e.g. normal) distribution of the mean values $\bar{\pi}$, $\bar{\beta}$ and $\bar{L}$ for an individual organ as well as of the inter-patient variation of the parameters M and N. A major disadvantage of this averaging procedure is, however, that several additional parameters (the width of the distributions $G^{ind}_{\alpha \beta L}$ and $G^{pop}_{\pi \beta L M N}$) have to be introduced to the model, which makes it more difficult to apply the model to clinical data than the phenomenological model of Lyman.

### 18.5 TCP Models

The situation for TCP models is much more complicated than for NTCP, since tumour response is influenced by various dynamically changing factors. While the radio-sensitivity of normal tissues within an individual patient may be considered to be constant in time, the sensitivity of tumours strongly depends on factors such as oxygen status and the amount of angiogenesis. Moreover, these conditions can be different for different parts of the tumour and may furthermore change, even in the relatively short time of the radiotherapy course.

Several TCP models have been proposed in the literature (Borkenstein et al. 2004; Niemierko and Goitein 1993b; Roberts and Hendry 1998; Sanchez and Nahum 1999; Webb and Nahum 1993). All models are based on the assumption that the tumour consists of a number of non-interacting clonogenic cells, which respond independently to irradiation. Local tumour control is achieved if all clonogenic cells are inactivated, i.e. if none of them survive (Niemierko 1998). This means that tumours behave according to a special case of the critical volume model (Eq. (15) with respect to the end-point cell inactivation) or according to the critical element model (Eq. (16)), with respect to the end-point cell survival.

The probability that k out of N clonogenic cells survive radiation is given by binomial statistics, which may be approximated for large numbers of N and constant values of N-SF by Poisson statistics:

$$TCP(D) = \left(\frac{N}{k}\right) SF^k (1-SF)^{N-k} = \frac{(N-SF)^k}{k!} e^{N-SF} \quad (21a)$$

$SF$ is the expectation value for the dose dependent survival probability, which may be identified as the survival fraction calculated by the linear-quadratic model (Eq. (2)). Achieving none surviving clonogenic cells means $k=0$, which results in
This means that the dose-response curve for tumours is determined by the number of clonogenic cells as well as by their parameters $\alpha$ and $\beta$ referring to the linear-quadratic model. Using $N = \rho \cdot V$, where $\rho$ is the density of the clonogenic cells and $V$ the volume of the tumour, it follows that a higher cell density as well as a larger tumour volume require a higher dose to arrive at the same TCP value. Equation (21) may also be extended to non-uniform distributions of dose and clonogenic cell density (Webb and Nahum 1993).

As most tumours show $\alpha/\beta$-ratios of 10 or more, the $\beta$-term in Eq. (21) may be neglected for small doses. This approximation has been applied, for example, by Webb and Nahum (1993) and Sanchez-Nieto and Nahum (1999). On the other hand, it may be necessary to include a time factor similar to Eq. (5) to account for proliferation of the clonogenic cells for a fractionated treatment (Roberts and Hendry 1998).

The intra-tumour as well as the inter-patient variation of tissue parameters is considered to be larger for tumours than for normal tissues; therefore, most models attempt to introduce this heterogeneity into the TCP models by averaging over the most important parameters, e.g. over the cell sensitivity $\alpha$, the repopulation factor $\gamma$ and cell number $N$ (Roberts and Hendry 1998). This is done in a similar fashion as for the previously described NTCP models and requires additional parameters for the widths of the parameter distributions.

A special situation arises for hypoxic tumours, in which the radio-sensitivity of cells is significantly reduced. This effect may in principal be described by introducing the oxygen enhancement ratio defined by

$$OER = \frac{D_{\text{hypox}}}{D},$$

where $D$ and $D_{\text{hypox}}$ are biologically iso-effective doses for well-oxygenated and hypoxic tumours. Typical values of the OER vary between 1 (well oxygenated) and 3 (strongly hypoxic). The OER may then be introduced into TCP models by replacing the dose $D$ by $D_{\text{hypox}}/OER$, which increases doses for the same TCP.

Hypoxia, however, is a very complex phenomenon which results from an interaction between tumour growth and tumour angiogenesis. As a consequence, the OAR may not only be different for different parts of the tumour, but it may change strongly even within a fractionated radiotherapy course. Since a detailed modelling of the OER also requires a model for tumour angiogenesis, most TCP models do not consider hypoxia.

An alternative approach to model TCP is the numerical simulation of tumour growth and response to radiation using Monte Carlo techniques (Borkenstein et al. 2004). With this approach, it is possible to include angiogenesis and to describe the dynamic behaviour of tumours on a statistical basis. Furthermore, the model can easily be extended, e.g. by introducing a cell cycle dependence of the radio-sensitivity.

### 18.6 RBE of High-LET Radiation

It is a general finding that high-LET radiation needs lower doses to produce the same biological effect as compared with photon radiation (Kraft 2000). This is due to the differences in the microscopic dose distribution between particles and photons. As clinical experience in radiotherapy almost completely relies on data from photon therapy, the relative biological efficiency (RBE) has to be introduced:

$$RBE = \frac{D_{\text{photon}}}{D_{\text{particle}}},$$

where $D_{\text{photon}}$ and $D_{\text{particle}}$ are biologically iso-effective doses for photons and high-LET radiation, respectively. The biological effective dose of a particle irradiation is then given as the product of the physical dose multiplied by the RBE. The unit of the biological effective dose is GyE (Gray equivalent) or CGE (Cobalt Gray equivalent), if $^{60}$Co-radiation is used as reference beam quality.

The RBE is a complex quantity (Wambersie and Menzel 1993) and rises with increasing LET and decreasing dose. Moreover, the RBE depends on the particle type and energy, on the biological system (e.g. cell or tissue type) as well as on the biological end point (e.g. early vs late effects). Similar to the parameters $\alpha$ and $\beta$ of the linear-quadratic model, one has to distinguish experimental RBE values for a specific cell type and RBE values for a clinical setting.

#### 18.6.1 Protons

Compared with heavier charged particles, the LET of proton radiation is low and as a result the RBE is
only slightly increased with respect to $^{60}$Co-radiation. Yashkin et al. (1995), for instance, published values of 1.07–1.10 for the plateau region and 1.07–1.14 in the spread-out Bragg-peak (SOBP). A compilation of measured RBE data for proton beams is given by Paganetti (2003). As the variation of the RBE with depth is small, a fixed RBE value of 1.1 is currently used for all clinical applications of proton beams, i.e. the potential difference of the RBE between the plateau region and the SOBP is neglected.

There is, however, experimental evidence that the RBE increases significantly in the last few millimetres of the SOBP, where the LET reaches its maximum (Kraft 2000; Paganetti 2003); therefore, models have been developed to calculate the RBE in the SOBP as a function of depth (Paganetti 2003). These so-called track structure models assume a radial dose distribution around the track of a particle. The response of the biological system is then calculated from the overlap of this microscopic dose distribution with the biological target using the known radiation response of the system for photon irradiation at equal dose. An alternative approach is to analytically model the dose-averaged LET for a SOBP (Wilkens and Oelfke 2003) and to describe the RBE as a function of this quantity.

As the fixed RBE value of 1.1 was found to be appropriate in proton therapy, and since the uncertainty of theoretical RBE values is larger than the RBE variation within the target volume, a modified RBE concept is not considered to be warranted (Paganetti 2003); therefore, the described models have not been clinically applied yet.

### 18.6.2 Heavy Charged Particles

The only ions heavier than protons, which are currently applied to patients are carbon ions (Kanai et al. 1999; Kraft 2000). With increasing ion charge, the LET and hence the RBE is increasing. While proton therapy operates with a fixed RBE value, this is not possible for heavier ions since the RBE varies not only between the plateau and the SOBP region but also within the SOBP. A RBE model is therefore necessary to achieve a homogeneous biological effect in the target volume.

Two different approaches are currently applied, depending on whether a passive beam-shaping technique with a fixed modulation depth (Kanai et al. 1999) or a scanned beam with active energy variation (Haberer et al. 1993) is used.

### RBE as Function of LET

For passive beam-shaping techniques, a modulator wheel is used to spread out the Bragg peak to the extension of the tumour in depth. As a consequence, several mono-energetic beams with different LET values contribute to the dose at a certain point. For this mixed beam situation, the RBE can be calculated according to Eq. (23) using iso-effective doses from the cell survival curves for photons and carbon ions, respectively (Kanai et al. 1997, 1999). To do so, the parameters $\alpha$ and $\beta$ of the linear-quadratic model have to be known for both beam qualities. As $\alpha$ and $\beta$ are LET-dependent, average values for the mixed carbon ion beam have to be determined by (Kanai et al. 1997):

\[
\alpha_{\text{mix}} = \sum_i \frac{d_i}{D} \alpha_i \quad (24a)
\]

\[
\sqrt{\beta_{\text{mix}}} = \sum_i \frac{d_i}{D} \sqrt{\beta_i} \quad (24b),
\]

where $d_i/D$ is the dose fraction of the $i$-th mono-energetic beam at a specified depth. $\alpha_i$ and $\beta_i$ are the parameters of the linear-quadratic model for the $i$-th mono-energetic beam, which are determined experimentally as a function of LET. As the depth dose and hence the LET distribution is fixed by the selected modulator wheel, the RBE distribution is also fixed over the cross-section of the tumour.

The transfer the RBE of cell lines into a clinical RBE was based on clinical experience with neutron therapy. As a carbon ion beam at a dose-averaged LET of 65 keV/µm was found to be biologically equivalent to a previously applied neutron beam, the RBE distribution calculated from the cell line was normalized to the clinical neutron-RBE at the depth of this LET value (Kanai et al. 1999).

### The Local Effect Model

In contrast to passive beam-shaping techniques, a scanned beam with active energy variation (Haberer et al. 1993) allows arbitrary-shaped depth dose distributions which moreover may have varying modulation depths over the tumour cross section. As a result the LET and hence the RBE may be different for each point within the treatment field.

The local effect model has been developed to calculate the cell survival after charged particle irradiation based on the survival curve for photon irradiations (Kraft et al. 1999; Kraft 2000; Scholz
and Kraft 1994; Scholz et al. 1997). The following considerations are used for the calculation (Scholz et al. 1997):

1. The survival curve for photon irradiation is described by an equation similar to Eq. (2). As the survival fraction may be needed for high doses, Eq. (2) should be modified to behave purely exponentially above a certain dose.

2. The shape of the radial dose distribution around a particle track has to be assumed.

3. The size of the target structure, e.g. the cell nucleus, has to be specified.

4. Then the distribution of the particle tracks over the target structure is simulated using a Monte Carlo technique. As some of the incident particles undergo fragmentation, the beam also contains a certain fraction of lighter ions, which increases with depth. The spatial distribution of these fragments over the target structures has also to be simulated; therefore, a fragmentation model is necessary to obtain the amount of fragments and their energy distribution at a certain depth.

5. At a specified point in the target structure, the local dose, \( d \), may result from contributions of different particle tracks. The probability for the whole target structure to survive may then be obtained by integration over the volume, \( V \), of the structure:

\[
\ln \left( \frac{SF_{\text{cell}}}{SF_x} \right) = \int \ln \left( \frac{SF_x(d)}{SF_{\text{cell}}} \right) \frac{dV}{V}. \tag{25}
\]

\( SF_x \) is the survival fraction for photon irradiation evaluated for the local dose \( d \). \( SF_{\text{cell}} \) is the survival probability of a selected cell. As the local dose, \( d \), is a stochastic quantity, this is also the case for \( SF_{\text{cell}} \). The expectation value of the survival fraction may then be obtained by averaging \( SF_{\text{cell}} \) over a large number of cells.

From these considerations, the survival curve for charged particles can be calculated. The RBE is then derived from the survival curves for photons and ions using the doses at the same survival level (Eq. (23)).

The integration in Eq. (25) implicitly assumes a critical element architecture of the target structure, i.e. an inactivation of any sub-volume of the target structure inactivates the respective cell. The mathematical structure of Eq. (25) is the same as the one for the complication probability after irradiation of an organ with an inhomogeneous dose distribution \( \{D\} \) (generalisation of Eq. (11)). This can be seen if the survival fractions are replaced by the effect probabilities (\( P=1-SF \)).

To apply the local effect model for treatment planning (Krämer and Scholz 2000), the RBE for cells lines have to be transferred into clinical RBE values. This is done by replacing the \( \alpha/\beta \)-ratio for cell lines by the \( \alpha/\beta \)-ratio for clinical end points (Krämer et al. 2003). Although the local effect model contains additional biological parameters, the \( \alpha/\beta \)-ratio is considered to be most important for the calculation of the RBE.

18.6.3 Neutrons

Previous clinical application of fast neutron therapy was purely based on experimental RBE values for various biological systems (Wambersie and Menzel 1993). No sophisticated biological models were used. Recently, the local effect model described above was also applied to predict RBE values in cell experiments (Scholz et al. 1997). In addition, models based on micro-dosimetry and Monte Carlo techniques have been developed for application in boron neutron capture therapy (Van Vliet-Vroegindeweij et al. 2001; Zamenhof et al. 1996).

18.7 Clinical Relevance of Biological Models

Biological models may in principle play an important role in clinical applications as they claim to predict the radiation response in patients. Due to the intrinsic uncertainty of the involved model parameters, however, there remains some uncertainty in these predictions.

18.7.1 The Linear-Quadratic Model

The linear-quadratic model has been found to be appropriate for in vitro experiments, animal experiments as well as for clinical applications, as long as moderate doses per fraction are applied. As the relations between different iso-effective treatment regimes depend only on the \( \alpha/\beta \)-ratio, the in vitro value may be replaced by an in vivo value. Nevertheless, conclusions for clinical applications have to be drawn very carefully. If extensions to the linear-quadratic model are applied, the situation is more complicated as additional parameters (e.g. \( \alpha \) and the repopulation
parameter \( t_0 \) have to be known in their absolute values, which may be difficult for an in vivo setting. Predictions of these extensions have to be treated with ever greater caution.

### 18.7.2 NTCP and TCP Models

The clinical application of NTCP models has significantly improved the understanding of the volume dependence of normal tissue response to radiotherapy. For clinical applications, mainly the phenomenological model of Lyman has been applied using the tolerance data provided by Emami et al. (1991) and the fit parameters of Burman et al. (1991). Although the uncertainty of these data was already stressed by the authors, these data are still used as reference in most of the recent literature. Since these tolerance data have been published, almost no attempt was made to refine this data base and it is not likely that this situation will improve in the near future. One reason for this may be the fact that the Lyman model requires input data from uniform partial-volume irradiations, which were less frequently applied with the upcoming of 3D-conformal radiotherapy. If this is the actual reason, the historical data published by Emami et al. (1991) may be the best data which can be achieved. Prediction of absolute NTCP values are therefore problematic, and the use of such absolute NTCP values as only criteria for clinical decisions is currently not warranted.

Although extensions to the Lyman model have been proposed, they are usually not applied to clinical data. As these extensions need additional biological parameters, the uncertainty of all parameters in the model increases and it is not expected that the description of clinical data will be improved.

The NTCP models are frequently applied to comparative planning studies and it is argued that this is justified as the models are known to give a correct qualitative description of the radiation response and only the ranking of NTCP values is considered. Although this kind of application contains weaker demands to the models, one major problem persists: the uncertainty of the predictions due to the uncertainty of the model parameters is mostly not specified quantitatively and the question arises as to whether a difference in NTCP values for different treatment plans (or techniques) may be regarded as significant.

The NTCP models based on more radiobiological principles may be applied for improving the principal understanding of normal tissue response to radiation. As these models contain more parameters than the Lyman model, they are usually not applied to clinical data.

In principal, the restrictions to NTCP models apply also to TCP models. For TCP models, however, the situation is even more complicated, since parameters such as proliferation, oxygenation and angiogenesis are much more heterogeneous for tumours than for normal tissues. Moreover, as these parameters may change under radiotherapy, the clinical application of TCP models is difficult. Nevertheless, the models may be applied to improve the understanding of the tumour response to radiation and its interaction with accompanying influence factors.

Although TCP/NTCP models have been developed and implemented into the cost functions of dose optimisation algorithms (Brahme 2001), clinically applied treatment plans continue to be optimised in terms of physical dose because of the intrinsic uncertainties of the biological models. Amols et al. (1997) discussed an approach to optimise treatment plans on the basis of NTCP- and TCP predictions and some additional parameters describing the risk acceptance of individual patients and physicians. This integrated approach to biological plan optimisation appears to be far away from application in clinical reality.

### 18.7.3 RBE of High-LET Radiation

The RBE models for high-LET radiation take an exceptional position in the field of biological modelling as most clinical experience is based on photon therapy. For the application of high-LET radiation, the RBE must be considered somehow to make use of this experience. In this context, TCP models are the only biological models which are routinely applied in clinical practice and it is out of discussion that the application is necessary for the optimisation of treatment plans.

The RBE variations for protons are small and there are currently no clinical indications that a more detailed RBE model than the constant factor of 1.1 is needed. For heavy ions, however, the RBE varies between much larger values and moreover depends on several physical as well as biological parameters. Current models describe the main characteristics of these dependencies and enable the safe application of heavy ion therapy. Similar to other biological models, intrinsic uncertainties are involved in the predictions of these models. This uncertainty has to be kept in mind, if the model is introduced into clinical applica-
tion. This especially means that the prescribed dose has to be selected very carefully and the full potential of heavy ions has to be determined in dose escalation studies as it is currently done at the HIMAC facility in Japan (Tsujii et al. 2002).

18.8 Conclusion

Several biological models have been developed. Although these models give a correct description of the main characteristics of the radiation response, great caution has to be taken if these models are to be applied to patients.

While the linear-quadratic model provides a good description of experimental settings, a larger uncertainty is involved in the prediction of iso-effects for clinical applications. The more advanced NTCP and TCP models should only be applied for relative, rather than absolute, predictions of effect probabilities. When using relative values, the uncertainty of the predictions should be considered to decide whether a detected difference is really significant. As TCP/NTCP models are currently not completely validated, integration of these models into the cost function of the dose optimisation algorithm is not warranted. Whether it is possible to arrive at fully biologically optimised treatment plans for photon therapy has to be investigated by further research.

In this context, the clinical application of heavy charged particles plays an exceptional role as biological optimisation is routinely performed and an adequate RBE model is an essential prerequisite. The applied RBE model may still contain some degree of uncertainty which has to be considered carefully at treatment plan assessment and dose prescription.

References

Dale RG (1986) The application of the linear-quadratic model to fractionated radiotherapy when there is incomplete normal tissue recovery between fractions, and possible implication for treatments involving multiple fractions per day. Br J Radiol 59:919–927
Kanai T, Furusawa Y, Fukutsu K et al. (1997) Irradiation of mixed beam and design of spread-out Bragg peak for heavy-ion radiotherapy. Radiat Res 147:78–85
Niemierko A (1998) Radiobiological models of tissue response to radiation in treatment planning systems. Tumori 84:140–143
19 2D and 3D Planning in Brachytherapy

Dimos Baltas and Nikolaos Zamboglou

CONTENTS
19.1 General 237
19.2 2D Treatment Planning 237
19.3 3D Treatment Planning 238
19.3.1 Anatomy Localization 239
19.3.2 Catheter Localization 239
19.3.3 Dose Calculation 242
19.3.3.1 The TG 43 Dosimetry Protocol 243
19.3.3.2 The Analytical Dose Calculation Model 245
19.3.4 Dose Optimization 246
19.3.4.1 Optimization Objectives 247
19.3.4.2 Multiobjective Optimization 248
19.3.4.3 Optimization Algorithms 248
19.3.5 Dose Evaluation 250
References 252

19.1 General

In contrast to external beam radiotherapy, the treatment planning procedure in brachytherapy includes an additional and specific component, namely the identification and reconstruction of the radiation emitters, the radioactive sources for permanent implants itself, or of the catheters and applicators used for temporary implants and other type of brachytherapy applications.

This means that although the dosimetric properties and kernels of the sources used are known, their actual position in the patients body has to be firstly defined/reconstructed. This is specific precondition establishes the calculation of the dose distribution possible.

The above step is indirectly considered in the external beam radiotherapy planning procedure through patient positioning and alignment respective to treatment machine gantry.

The brachytherapy treatment planning procedure consists generally of the following steps:

- Definition of the planning target volume (PTV) and organs at risk (OARs)
- Reconstruction of the implanted sources or catheters and applicators
- Calculation and optimization of the dose distribution
- Evaluation of the dose distribution

All above steps can be realized using a technology adequate for the aims of the therapy. As a result of this, all mentioned components can be approached using 2D representations and documentations for simplified applications or using 3D imaging techniques such as CT, MR and US (Baltas et al. 1994; Baltas et al. 1999; ICRU 1997).

In the era of 3D conformal radiation therapy, brachytherapy treatments have proved to be adequate competitors or alternatives to the 3D conformal external beam treatments, especially in the age of intensity modulation technology. This can be only achieved when modern imaging tools are used for guidance and navigation for the realization of a brachytherapy implant, as well as for the treatment planning procedure itself.

When 3D methods are applied for all the above steps or components of the treatment planning procedure, then we can characterize this as 3D treatment planning.

19.2 2D Treatment Planning

Here projectional imaging methods, such as X-ray fluoroscopy or radiographs, are used for verifying and documenting the placement of usually a sin-
gle catheter or applicator. This is the case for the “standard treatments” using simple standard applicators as in the case of a cylinder applicator for the postoperative intracavitary brachytherapy of corpus uteri carcinomas (Krieger et al. 1996; Baltas et al. 1999).

The treatment delivery itself is then based on pre-existing standard plans with isodose distribution documentation. Due to the rigidity of such kind of applicators, the main item/challenge here is to check the correct placement of the catheter in the patient.

The 2D treatment planning procedure is mainly applicator oriented/based. When the placement of the applicator is validated using simple X-rays and is found to be at the adequate position, then the dose delivery to the anatomy around the applicator can be assumed as appropriate for such kind of simple geometries and catheter/source configurations.

19.3 3D Treatment Planning

Here the target and organ at risk localization as well as the catheter reconstruction are based on 3D methods using modern imaging modalities. The same is valid for the dose calculation and evaluation.

A common procedure, at least in the past for gynaecological and other intracavitary applications, was based on two or more X-ray films, which are mainly used for the 3D reconstruction of the used catheters or applicators. For the intracavitary brachytherapy of the primary cervix carcinoma a set of discrete anatomical points has been and is continuously being used for documenting the dose distribution to the patient anatomy. These points have been selected in a way that they can be identified on X-ray films when a specific geometry is applied (ICRU 1985; Herbort et al. 1993). This method of reconstruction is called projectional reconstruction method (PRM; Tsalpatouros et al. 1997; Baltas et al. 1997; Baltas et al. 2000). Due to the fact that PRM is of limited practicability with reference to the definition of anatomical volumes such as PTV and OARs, PRM can be considered an intermediate, 2.5D, treatment planning method, where the catheters and the dose calculations are realized in the 3D space but only a limited correlation of this distribution to the anatomy can be achieved.

Figure 19.1 demonstrates the two localization radiographs used for the treatment planning of a brachytherapy cervix implant using the ring applicator.

Due to the missing correlation between anatomy and dosimetry when using conventional X-ray radiographs, it is presently common to use 3D sectional imaging such as CT, MR or ultrasound (US) for treatment planning purposes. This is becoming increasingly more popular and tends to replace the traditional methods, at least in the western world. In fact, the establishment of brachytherapy as first-choice treatment for early stages of prostate cancer, where US imaging for the pre- and intraoperative planning...
and needle insertion, as well as the CT imaging for the post-planning, are mandatory for an effective and safe treatment, gave rise to developments in the field of imaging-based treatment planning which is also of benefit for all other brachytherapy applications.


In addition CT, MR and US are currently used for guidance during needle insertion, offering through this a high degree of safety and intra-implantation approval of the needle position relative to the anatomy (Zamboglou et al. 1998; Kolotas et al. 1999a; Kolotas et al. 1999b; Kolotas et al. 2000).

Table 19.1 presents an overview of the different imaging modalities regarding their role and possibilities for treatment planning in brachytherapy. Herein the different steps of the 3D imaging-based treatment planning in modern brachytherapy is addressed in detail.

### 19.3.1 Anatomy Localization

One or more imaging modalities can be included for the delineation of the patient’s anatomy, GTV, CTV, PTV and organs at risk (OARs) that have to be considered either for the preparation of the implant (pre-planning) or for the planning of brachytherapy delivery when all catheters are already placed (post-planning).

Here the standard tools, known as the external beam planning systems, are also used for effective and accurate 3D delineation of tissues and organs. Figures 19.2–19.4 demonstrate the tissue delineation for MRI-based pre-planning, 3D US-based intraoperative pre-planning and CT-based post-planning of a prostate monotherapy implant, respectively.

For a more accurate delineation especially of localization in soft tissues, such as gynaecological tumours, brain tumours and perhaps prostate, MRI imaging (pre-application) can be considered fused with CT or US imaging used for the implantation procedure itself.

### 19.3.2 Catheter Localization

The greatest benefit when using 3D imaging modalities such as CT, MRI or US for the localization and reconstruction of catheters is that there is no need of identifying and matching of catheters describing markers or points on two different projections, as is the case for the PRM method (Tsalpatouros et al. 1997; Milickovic et al. 2000a; Milickovic et al. 2000b). The PRM methods are man-power intensive and require the use of special X-ray visible markers that are placed within the catheters in order to make them visible for the reconstruction.

The available technology enables effective and fast catheter reconstruction using CT imaging and currently US imaging using automatic reconstruction tools (Milickovic et al. 2000a; Giannouli et al. 2000). Such a kind of technology makes the reconstruction procedure user independent and increases the reliability of the brachytherapy method.

Figures 19.5 and 19.6 demonstrate the results of the 3D US-based and CT-based reconstruction, respectively, of the realized implant with 19 catheters for the prostate cancer case shown in Figs. 19.2–19.4.

The 3D US imaging has been used for the intraoperative, live, planning and irradiation, whereas the CT imaging has been used for the treatment planning of the second fraction with that implant.

The in-plane resolution of the US imaging is as high as some tenths of a millimetre, whereas for the CT imaging the in-plane resolution is about 0.5 mm. The limited resolution when using CT imaging in the sagittal and coronal planes results from the inter-slice

---

<table>
<thead>
<tr>
<th>Imaging modality</th>
<th>Anatomy</th>
<th>Catheter/applicators</th>
<th>Availability</th>
<th>Speed</th>
<th>Live/interactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional X-ray</td>
<td>+</td>
<td>++++++</td>
<td>+++++</td>
<td>++++</td>
<td>++++</td>
</tr>
<tr>
<td>CT</td>
<td>++++</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>MRI</td>
<td>+++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>US</td>
<td>+++</td>
<td>+++</td>
<td>++++</td>
<td>++++</td>
<td>++++</td>
</tr>
<tr>
<td>3D US</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
<td>++++</td>
<td>+++</td>
</tr>
</tbody>
</table>
Fig. 19.2. Example of an anatomy delineation for the pre-planning of a high dose rate (HDR) monotherapy implant using MR imaging (T2-weighted imaging) and the SWIFT treatment planning system (Nucletron B.V., Veenendaal, The Netherlands). 
Upper left: coronal view. Lower left: axial image. Lower right: sagittal view. Upper right: 3D view of the planning target volume (PTV; red), urethra (yellow) and rectum (purple).

Fig. 19.3. Example of an anatomy delineation for the intraoperative pre-planning for the HDR monotherapy implant of the prostate cancer case of Fig. 2, using 3D US imaging and the SWIFT treatment planning system (Nucletron B.V., Veenendaal, The Netherlands). Upper left: coronal view. Lower left: axial image. Lower right: sagittal view. Upper right: 3D view of the PTV (red), urethra (yellow) and rectum (purple). The benefit of using US imaging for identifying the apical prostate limits is clearly demonstrated in the sagittal view. This 3D imaging and anatomy model was used for the creation of an intraoperative pre-plan for the catheter placement.
Fig. 19.4. Example of the anatomy delineation for the post-planning of an HDR monotherapy implant for the second brachytherapy fraction and for the prostate cancer case of Fig. 19.2, using CT imaging and the SWIFT treatment planning system (Nucletron B.V., Veenendaal, The Netherlands). Upper left: coronal view. Lower left: axial image. Lower right: sagittal view. Upper right: 3D view of the PTV (red), urethra (yellow) and rectum (purple). Contrast media has been used for an adequate visualization of the bladder and the urethra. The difficulty of identifying the apexal prostate limits when using CT imaging is demonstrated in the sagittal view. The 19 implanted catheters (black holes or curves) are also clearly identified on all images.

Fig. 19.5. Catheter reconstruction for the intraoperative live planning of the HDR monotherapy implant of the prostate cancer case of Fig. 19.2, using 3D US imaging. Upper left: coronal view. Lower left: axial image. Lower right: sagittal view. Upper right: 3D view of the PTV (red), urethra (yellow) and rectum (purple), and of the catheters (yellow lines) with the automatically selected appropriate source steps in these (red circles). All 19 implanted catheters (white surfaces and curves) are also clearly identified on all images.
distance. In the example of Figs. 19.4 and 19.6 this was 3.0 mm. Generally, the total accuracy that can be achieved with CT or MR imaging is half the slice thickness with the precondition that slice thickness equals the inter-slice distance (no gap). It is recommended to use for the reconstruction of not straight (metallic) catheters a slice thickness and inter-slice distance of 3 mm, achieving in this way an accuracy as high as 1.5 mm. When using axial MR imaging, then usually the longitudinal image distance is 5.0 mm resulting in an accuracy of 2.5 mm, which makes MRI for catheter reconstruction in several cases of limited interest. This can be overcome if non-axial MRI can be incorporated in the reconstruction procedure.

In contrast to CT, in 3D US imaging the inter-plane distance is as low as 1.0 mm (Figs. 19.3, 19.5), resulting thus in an accuracy better than 1.0 mm (actually of ca. 0.5 mm). Another benefit of US imaging for the reconstruction of catheters is the possibility to combine 3D volume reconstruction with live 2D imaging, offering in this way the possibility of an interactive reconstruction.

19.3.3 Dose Calculation

Although several national protocols exist for the dose calculation around brachytherapy sources, the protocol proposed and established by the American Association of Physicists in Medicine (AAPM), Task Group 43, and published in 1995 (Nath et al. 1995), has been widely accepted and builds the standard protocol that the majority of vendors of treatment planning systems in brachytherapy are following. Even if this was primarily focused to low dose rate (LDR) sources (in the original publication it was explicitly mentioned that high activity sources and iridium wires were beyond the scope of that report), the TG 43 formalism has been widely used and virtually internationally accepted also for high dose rate (HDR) iridium sources used in remote afterloading systems.

The TG 43 formalism is a consistent, and a simple to implement, formalism based on a small number...
of parameters/quantities that can be easily extracted from Monte Carlo (MC) calculated dose rate distributions around the sources in a water-equivalent medium.

The basic concept of the TG 43 dosimetry protocol is to derive dosimetry parameters for calculating dose rates or dose values directly from measured or MC calculated dose distributions around the sources in water or water-equivalent medium. This increases the accuracy of the calculations to be carried out in the clinic, which are always for water medium and not in free space. Furthermore, this method avoids the use of any term of activity (apparent or contained) that has led to significant discrepancies in the past.

19.3.3.1 The TG 43 Dosimetry Protocol

Figure 19.7 summarizes the geometry and coordinate definitions used in the TG 43 dosimetry protocol.

The dose rate $D(r, \theta)$ at a point $P$ around a source having cylindrical coordinates $(r, \theta)$ relative to the source coordinate system is given according to that protocol by:

$$D(r, \theta) = S_K \cdot \Lambda \cdot \frac{G(r, \theta)}{G(r_0, \theta_0)} \cdot g(r) \cdot F(r, \theta)$$  \hspace{1cm} (1)

where $S_K$ is the air kerma strength of the source, $\Lambda$ is the dose rate constant, $G(r, \theta)$ is the geometry function, $g(r)$ is the radial dose function and $F(r, \theta)$ is the anisotropy function.

Air Kerma Strength

The air kerma strength ($S_k$) replaces the previous commonly used quantity apparent activity $A_{app}$ and describes the strength of the brachytherapy source. $S_k$ is defined as the product of air kerma rate in free space at the distance of calibration of the source $d \cdot \bar{K}(d)$ and the square of that distance, $d^2$:

$$S_k = \bar{K}(d) \cdot d^2$$  \hspace{1cm} (2)

The calibration must be performed at a distance, $d$, defined along the transverse bisector of the source, $r=d$ and $\theta=\pi/2$ in Fig. 19.7, that is large enough so that the source can be considered as a point source. For direct measurements of $\bar{K}(d)$ using an in-air setup the possible attenuation of the radiation in air has to be considered. According to the above definition, $S_k$ accounts also for the scattering and attenuation of the radiation which occurred in the source core and source encapsulation. The reference calibration distance for $\bar{K}(d)$ is common use is 1 m. TG 43 recommends as a unit for reporting the air kerma strength $S_k$ of a source, $\mu$Gy.m$^2$.h$^{-1}$, and denotes this by the symbol U: $1 \text{ U} = 1 \mu$Gy.m$^2$.h$^{-1}$ = 1 cGy.cm$^2$.h$^{-1}$

Dose Rate Constant

The dose rate constant $\Lambda$ is defined according to:

$$\Lambda = \frac{D(r_0, \theta_0)}{S_k}$$ \hspace{1cm} (3)

where $r_0=1.0$ cm and $\theta_0=\pi/2$. For the definition of the polar coordinates $r$ and $\theta$ see Fig. 7.

Geometry Function $G(r, \theta)$

$G(r, \theta)$ is the geometry function describing the effect of the spatial distribution of the activity in the source volume on the dose distribution and is given by:

$$G(r, \theta) = \frac{\int_{\text{source}} \rho(\vec{r}) \cdot dV}{\int_{\text{source}} \rho(\vec{r}) \cdot dV}$$ \hspace{1cm} (4)
where \( \rho(r') \) is the activity per unit volume at a point \( r' \) inside the source and \( dV' \) is an infinitesimal volume element located at the same position. This factor reduces to:

\[
G(r, q) = G(r) = \frac{1}{r^2} \text{ for a point source} \tag{5}
\]

\[
G(r, \theta) = \frac{\theta_1 - \theta_2}{L \cdot r \cdot \sin \theta} \text{ for a finite line source} \tag{6}
\]

Here \( L \) is the active length of the source and the angles \( \theta_1 \) and \( \theta_2 \) are illustrated in Fig. 19.7.

**Reference Point of Dose Calculations**

The reference point is that for the formalism chosen to be the point lying on the transverse bisector of the source at a distance of 1 cm from its centre; expressed in polar coordinates as defined in Fig. 19.7, i.e. \( (r_0, \theta_0) = (1 \text{ cm}, \pi/2) \).

**Radial Dose Function**

\[
g(r) = \frac{G(r, \theta_0)}{G(r, \theta)} \cdot \frac{\bar{D}(r, \theta_0)}{\bar{D}(r, \theta)} \tag{7}
\]

**Anisotropy Function**

Finally, \( F(r, \theta) \) is the anisotropy function defined as:

\[
F(r, \theta) = \frac{G(r, \theta_0)}{G(r, \theta)} \cdot \frac{\bar{D}(r, \theta)}{\bar{D}(r, \theta_0)} \tag{8}
\]

**Anisotropy Factor**

Because of the difficulty in determining the orientation of the implanted seeds, post-implant dosimetry for low dose rate permanent implants is based on the point source approximation using the anisotropy factor \( \varphi_{an}(r) \):

\[
\varphi_{an}(r) = \left[ \frac{1}{2 \cdot \bar{D}(r, \theta_0)} \right] \cdot \int_0^\pi \bar{D}(r, \theta) \cdot \sin \theta \cdot d\theta =
\]

\[
\left[ \frac{1}{2 \cdot G(r, \theta_0)} \right] \cdot \int_0^\pi F(r, \theta) \cdot G(r, \theta) \cdot \sin \theta \cdot d\theta \tag{9}
\]

This is uncommon for the case of HDR iridium-192-based brachytherapy where the TG 43 formalism as given in Eq. (1) with the line source approximation described in Eq. (6) is used.

**Anisotropy Constant**

Using a \( 1/r^2 \) weighted-average of anisotropy factors, for \( r > 1 \text{ cm} \), the distance independent anisotropy factor \( \varphi_{an} \) is calculated using the equation:

\[
\varphi_{an} = \frac{\sum \varphi_{an}(r)}{\sum 1/r^2} \tag{10}
\]

In the literature the TG 43 parameter values and functions for the common used seeds or HDR iridium sources can be found.

Recently AAPM has updated the TG 43 protocol for low-activity seeds (Rivard et al. 2004), where it is recommended to use separately radial dose functions \( g(r) \) and anisotropy functions \( F(r, \theta) \) as well as anisotropy factors \( \varphi_{an}(r) \) and anisotropy constants \( \varphi_{an} \) in dependence on the geometry factor is used; point (see Eq. (5)) or line (see Eq. (6)) source approximation. This report contains the corresponding tables for all factors and functions for the most common seeds according to the new formulation.

Although the TG 43 formalism offers a stable platform for calculation of dose or dose rate distributions in brachytherapy, it can be easily seen from Eq. (1) that the TG 43 formalism is actually a 2D model.

Tissue inhomogeneities and bounded geometries are not considered by this formalism. The effects of the presence of inhomogeneities and the variable dimensions of patient-specific anatomy are ignored. The MC simulation would be the only accurate solution to the aforementioned deficiencies based on actual patient anatomical data. That is, however, still too time-consuming to be incorporated in a clinical environment in spite of promising correlated simulation techniques (Hedtjärn et al. 2002); therefore, kernel superposition methods (Williamson and Baker 1991; Carlsson and Ahnesjö 2000; Carlsson Tedgren and Ahnesjö 2003) and analytical models (Williamson et al. 1993; Kirov and Williamson 1997; Daskalov et al. 1998) have been employed and tested in a variety of geometries, thus opening the way of handling tissue and shielding material inhomogeneities and bounded patient geometries.

A simpler analytical dosimetry model (Anagnostopoulos et al. 2003) based on the primary and scatter separation technique (Russell and Ahnesjö 1996; Williamson 1996) was published and evaluated in patient-equivalent phantom geometries (Pantelis et al. 2004; Anagnostopoulos et al. 2004). The kernel superposition as well as the analytical dose calculation models announce in this
way the future of 2.5 and real 3D dose calculation in brachytherapy.

In the following a short description of this recently proposed simple analytical dose calculation model that has been shown to describe adequately the dose distribution in inhomogeneous tissue environments is given.

### 19.3.3.2
#### The Analytical Dose Calculation Model

According to the analytical dose rate calculation formalism proposed in the work of Anagnostopoulos et al. (2003) the dose rate per unit air kerma strength, $S_K$, in a homogeneous tissue medium surrounding a real $^{192}$Ir source can be calculated using the following equation (Anagnostopoulos et al. 2003; Pantelis et al. 2004):

$$
\frac{D_{\text{medium}}(r, \theta)}{S_K} = \frac{\left(\mu_{\text{en}}/\rho\right)_{\text{air}}}{\mu_{\text{medium}}} \cdot e^{-\mu_{\text{medium}}(1 + \text{SPR}_{\text{water}}(\rho_{\text{medium}} r))G(r, \theta)F(r, \theta)}
$$

(11)

where $r$ is the radial distance, $(\mu_{\text{en}}/\rho)_{\text{air}}$ is the effective mass energy absorption coefficient ratio of the medium of interest to air, $\mu_{\text{medium}}$ is the effective linear attenuation coefficient of the medium, $\text{SPR}_{\text{water}}$ is the scatter to primary dose rate ratio calculated in water medium and $\mu_{\text{medium}}$ is the density of the medium.

The effective mass energy absorption coefficient ratio, $(\mu_{\text{en}}/\rho)_{\text{air}}$, and the effective linear attenuation coefficient, $\mu_{\text{medium}}$, are calculated by weighting over the primary $^{192}$Ir photon spectrum, while the scatter to primary dose rate ratios for water medium (Russell and Ahnesjö 1996; Williamson 1996) is calculated using the polynomial fitted function (Anagnostopoulos et al. 2003; Pantelis et al. 2004):

$$\text{SPR}_{\text{water}}(\rho r) = 0.123 (\rho r) + 0.005 (\rho r)^2
$$

(12)

that can accurately calculate (within 1%) the $\text{SPR}_{\text{water}}$ values for density-scaled distances of $\rho r \leq 10$ g cm$^{-2}$. For the general application of Eq. (11) for every homogeneous medium, changing from homogeneous water to a different homogeneous medium would necessitate MC calculated $\text{SPR}_{\text{medium}}(r)$ results thus reducing the versatility of an analytical model; however, due to the range of the $^{192}$Ir energies and the consequent predominance of incoherent scattering (Anagnostopoulos et al. 2003), $\text{SPR}_{\text{medium}}(r)$ results for tissue materials are in good agreement with that of water when plotted vs distance scaled for the corresponding density (i.e. in units of grams per square centimetre). This is shown in Fig. 19.8 where MC calculated (Briesmeister 2000) $\text{SPR}_{\text{bone}}(r)$ ratios of cortical bone are also plotted vs distance from a point $^{192}$Ir source multiplied by the corresponding material density (1.92 g/cm$^3$ for bone) and an overall good agreement within 1–5% may be observed (Anagnostopoulos et al. 2003).

In this equation, $G(r, \theta)$ is the geometry function of the source accounting for the spatial distribution of radioactivity. $F(r, \theta)$ is the anisotropy function accounting for the anisotropy of dose distribution around the source.

In order to account for the presence of different inhomogeneous materials along the path connecting the source and a dose point in a patient anatomy-equivalent phantom, Eq. (11) was generalized to:

$$
\frac{D_{\text{medium}}(r, \theta)}{S_K} = \frac{\left(\mu_{\text{en}}/\rho\right)_{\text{air}}}{\mu_{\text{medium}}} \cdot e^{-\sum_i \mu_{\text{medium}}^i(1 + \text{SPR}_{\text{water}}^i(\rho_{\text{medium}}^i r))G(r, \theta)F(r, \theta)}
$$

(13)

where $i$ is the index of every material transversed along the connecting path of the source point to

![Fig. 19.8. Scatter to primary (SPR) dose rate ratio results for unbounded, homogenous water and bone phantoms calculated with MC simulations are plotted vs density-scaled distance, $\rho r$, in units of grams per square centimetre. In the same figure a polynomial fit on water SPR values $\text{SPR}(\rho r) = 0.123(\rho r) + 0.005 (\rho r)^2$ is also presented](image)
the dose calculation point. The SPR in water for the scaled distance is parameterized according to Eq. (12), where \( r \) is the sum of the mass density scaled path lengths inside the inhomogeneities along the radius connecting the source with the dose calculation point.

The effect of patient inhomogeneities surrounding the oesophagus on the dosimetry planning of an upper thoracic oesophageal \(^{192}\)Ir HDR brachytherapy treatment was studied (Anagnostopoulos et al. 2004) and the analytical dose calculation model of Eq. (13) was found to correct for the presence of tissue inhomogeneities as it is evident in Fig. 19.9, where dose calculations with the analytical model are compared with corresponding results from the MCNPX Monte Carlo code (Hendricks et al. 2002) as well as with corresponding calculations by a contemporary treatment planning system software featuring a full TG-43 dose calculation algorithm (PLATO BPS v. 14.2.4, Nucletron B.V, The Netherlands) in terms of isodose contours. The presence of patient inhomogeneities had no effect on the delivery of the planned dose distribution to the PTV; however, regarding the OARs, the common practice of current treatment planning systems to consider the patient geometry as a homogeneous water medium leads to a dose over-estimation of up to 13% to the spinal cord and an underestimation of up to 15% to the bone of sternum (Anagnostopoulos et al. 2004). These discrepancies correspond to the dose region of about 5–10% of the prescribed dose and are only significant in case that brachytherapy is used as a boost to external beam therapy.

### 19.3.4 Dose Optimization

The objectives of brachytherapy treatment planning are to deliver a sufficiently high dose in the cancerous tissue and to protect the surrounding normal tissue (NT) and OARs from excessive radiation. The problem is to determine the position and number of source dwell positions (SDPs), number of catheters and the dwell times, such that the obtained dose distribution is as close as possible to the desired dose distribution. Additionally, the stability of solutions can be considered with respect to possible movements of the SDPs. The planning includes clinical constraints such as a realistic range of catheters as

![Fig. 19.9a,b. Percentage isodose contours calculated with the PLATO BPS v. 14.2.4 (- - -), the Monte Carlo (· · · ·) and the analytical model (—) in the inhomogeneous patient-equivalent geometry of an upper thoracic oesophageal \(^{192}\)Ir HDR brachytherapy treatment (Anagnostopoulos et al. 2004). The 100% isodose contour encompasses the cylindrical-shaped oesophageal PTV and is not altered due to the presence of the surrounding tissue inhomogeneities. The results in (a) are plotted on the central transversal plane (z=0 cm, adjacent to the central CT slice), whereas in (b) the same results are plotted on the sagittal plane containing the catheter inserted inside the oesophagus (x=0 cm)](image)
well as their positions and orientations. The determination of an optimal number of catheters is a very important aspect of treatment planning, as a reduction of the number of catheters simplifies the treatment plan in terms of time and complexity. It also reduces the possibility of treatment errors and is less invasive for the patient.

As analytical solutions cannot be determined, the solution is obtained by inverse optimization or inverse planning. The term “inverse planning” is used considering this as the opposite of the forward problem, i.e., the determination of the dose distribution for a specific set of SDPs and dwell times. If the positions and number of catheters and the SDPs are given after the implantation of the catheters, we term the process “post-planning”. Then, the optimization process to obtain an optimal dose distribution is called “dose optimization”. Dose optimization can be considered as a special type of inverse optimization where the positions and number of catheters and the SDPs are fixed.

Inverse planning has to consider many objectives and is thus a multiobjective (MO) optimization problem (Miettinen 1999). We have a set of competing objectives. Increasing the dose in the PTV will increase the dose outside the PTV and in the OARs. A trade-off between the objectives exists as we never have a situation in which all the objectives can be in the best possible way satisfied simultaneously. One solution of this MO problem is to convert it into a specific single objective (SO) problem by combining the various objective functions with different weights into a single objective function. Optimization and analysis of the solutions are repeated with different sets of weights until a satisfactory solution is obtained as the optimal weights are a priori unknown. In MO optimization a representative set of all possible so-called non-dominated solutions is obtained and the best solution is selected from this set. The optimization and decision processes are decoupled. The set provides a coherent global view of the trade-offs between the objectives necessary to select the best possible solution, whereas the SO approach is a trial-and-error method in which optimization and decision processes are coupled.

19.3.4.1 Optimization Objectives

An ideal dose function \( D(r) \) with a constant dose equal to the prescription dose, \( D_{\text{ref}} \), inside the PTV and 0 outside is physically impossible since radiation cannot be confined to the PTV only as some part of the radiation has to traverse the OARs and the surrounding NT. Out of all possible dose distributions the problem is to obtain an optimal dose distribution without any a priori knowledge of the physical restrictions. Optimality requires quantifying the quality of a dose distribution. A natural measure quantifying the similarity of a dose distribution at \( N \) sampling points with dose values, \( d_i \), to the corresponding optimal dose values, \( d_i^* \), is a distance measure. A common measure is the \( L_p \) norm:

\[
L_p = \left( \sum_{i=1}^{N} (d_i - d_i^*)^p \right)^{1/p}
\]

For \( p=2 \), i.e. \( L_2 \), we have the Euclidean distance.

The treatment planning problem is transformed into an optimization problem by introducing as an objective the minimization of the distance between the ideal and the achievable dose distribution. These objectives can be expressed in general by the objective functions \( f_L(x) \) and \( f_H(x) \):

\[
f_L(x) = \frac{1}{N} \sum_{i=1}^{N} \Theta(D_L - d_i(x))(D_L - d_i(x))^p,
\]

\[
f_H(x) = \frac{1}{N} \sum_{i=1}^{N} \Theta(d_i(x)-D_H)(d_i(x)-D_H))^p
\]

where \( d_i(x) \) is the dose at the \( i \)th sampling point that depends on parameters \( x \) such as dwell times, \( p \) is a parameter defining the distance norm, \( N \) the number of sampling points, \( D_L \) and \( D_H \) the low and high dose limits; these are used if dose values above \( D_L \) and below \( D_H \) are to be ignored expressed by the step function \( \Theta(x) \).

The difference between various dosimetric based objective functions is the norm used for defining the distance between the ideal and actual dose distribution, on how the violation is penalized and what dose normalization is applied. For \( p=2 \) we have the quadratic-type or variance-like set of objective functions. Specific objectives of this type were used by Millickovic et al. (2002) including an objective for the dose distribution of sampling points on the PTV surface that results in an objective value that is correlated with the so-called conformity index used by Lahanas et al. (1999) directly as an objective. The objective functions require that the SDPs are all inside the PTV. In the case of SDPs outside the PTV additional or modified objective functions are required. For \( p=1 \), a linear form, results were presented by Lessard and Pouliot (2001). For \( p=0 \) (Lahanas et al. 2003a) we have DVH-based objectives as the DVH value at the dose, \( D_H \), is given by

\[
DVH(D_H) = \frac{100}{N} \sum_{i=1}^{N} \Theta(d_i - D_H)
\]

2D and 3D Planning in Brachytherapy 247
The benefit in this case is that the objective values are easier to interpret than for other objective functions, although different dose distributions could produce the same objective values, as the dose distributions are only required to satisfy some integral properties. In this case gradient-based optimization algorithms cannot be used for dose optimization.

Dose-volume histogram specifying constraints for a clinically acceptable dose distribution can be included in the optimization (constraint dose optimization). Such constraints could specify upper bounds for the fraction of the volume of a region that can accept a dose larger than a specific level, or a lower bound for the fraction that should have a dose at least larger than a specific value.

There are physical limitations of what dose distributions can be obtained for a specific number of catheters and number of SDPs. The solutions obtained by inverse planning depend also on the used set of objective functions. The closer the dose distributions of these solutions are to the physically possible optimal solution, the better the set of objective functions is.

19.3.4.2 Multiobjective Optimization

For MO optimization with M objectives we have a vector objective function \( f = (f_1(x), \ldots, f_M(x)) \). In general, some of the individual objectives will be in conflict with others, and some will have to be minimized while others are maximized. The MO optimization problem can now be defined as the problem to find the vector \( x = (x_1, x_2, \ldots, x_N) \), i.e. solution which optimizes the vector function \( f \). Normally, we never have a situation in which all the \( f_j(x) \) values are optimal for a common point \( x \). We therefore have to establish certain criteria to determine what would be considered an optimal solution. One interpretation of the term optimum in MO optimization is the Pareto optimum (Miettinen 1999).

A solution \( x_1 \) dominates a solution \( x_2 \) if and only if the two following conditions are true:
1. \( x_1 \) is no worse than \( x_2 \) in all objectives, i.e. \( f_j(x_1) \leq f_j(x_2) \) \( \forall j=1,\ldots,M \).
2. \( x_1 \) is strictly better than \( x_2 \) in at least one objective, i.e. \( f_j(x_1) < f_j(x_2) \) for at least one \( j \in \{1,\ldots,M\} \).

We assume, without loss of generality, that this is a minimization problem. \( x \) is said to be non-dominated by \( x_2 \) or \( x_1 \) is non-inferior to \( x_2 \) and \( x_2 \) is dominated by \( x_1 \). Among a set of solutions \( P \), the non-dominated set of solutions \( P' \) are those that are not dominated by any other member of the set \( P \). When the set \( P \) is the entire feasible search space then the set \( P' \) is called the “global Pareto optimal set”. If for every member \( x \) of a set \( P \) there exists no solution in the neighbourhood of \( x \), then the solutions of \( P \) form a local Pareto optimal set. The image \( f(x) \) of the Pareto optimal set is called the “Pareto front”. The Pareto optimal set is defined in the parameter space, whereas the Pareto front is defined in the objective space.

19.3.4.3 Optimization Algorithms

The optimization algorithm used in brachytherapy planning depends on the selected set of objective functions. In the presence of local function minima deterministic algorithms may not work well. For variance-based objectives gradient-based optimization algorithms guided by gradient information can be used for post-planning and the solutions obtained are global optimal (Lahanas et al. 2003b). It is also possible to use special calculation methods (Lahanas and Baltas 2003) to perform a fast MO optimization in which the optimization is repeated with different uniform distributed sets of weights until a representative set of non-dominated solutions is obtained. For MO inverse planning with other objectives functions which consider the problem of the optimal position and number of catheters to be used for specific MO hybrid evolutionary algorithms combine the power of efficiency of deterministic algorithms with the parallel character nature of evolutionary algorithms with the aim of obtaining fast a representative set of Pareto optimal solutions.

Evolutionary algorithm (EA) is a collective term for all variants of probabilistic optimization algorithms that are inspired by Darwinian evolution. A genetic algorithm (GA) is a variant of EA, which, in analogy to the biological DNA alphabet, operates on strings which are usually bit strings of constant length. The string corresponds to the genotype of the individual. Usually, a GA has the following components:
- A representation of potential solutions to the problem
- A method to create an initial population of potential solutions
- An evaluation function that plays the role of the environment, rating solutions in terms of their fitness (expressed by a fitness function, in principle an objective function)
- Genetic operators that alter the composition of the population members of the next generation.
• Values of various parameters that the genetic algorithm uses (population size, probabilities of applying genetic operators, etc.)

In contrast to the canonical GA with bit-encoded parameters, the genome of real-coded GA consists of real-valued object parameters, i.e. evolution operates on the natural representation. Selection in EA is based on the fitness. Generally, it is determined on the basis of the objective value(s) of the individual in comparison with all other individuals in the selection pool. Elitism is a method that guarantees that the best ever found solution would always survive the evolutionary process. Crossover operators allow the parameter space to be searched initially sufficiently in large steps. During the evolution the search is limited around the current parameter values with increasing accuracy. Mutation operators are used for a search usually in the local neighbourhood of the parent solution.

For MO optimization we have a class of MO evolutionary algorithms (MOEAs) that use mainly dominance-based selection mechanisms. The MOEAs are designed to maintain a diverse Pareto front and can guide the population towards only important parts of the Pareto front.

An implementation of a GA begins with a population of (typical random) chromosomes. One then evaluates in such a way that those chromosomes which represent a better solution to a problem are given more chances to reproduce than those chromosomes which are poorer solutions. The goodness of a solution is typically defined with respect to the current population defined by a fitness function. Figure 19.10 shows the principal steps for GAs. For MOEAs, except the different selection method, some algorithm-specific additional steps are included.

The MOEAs can be more effective than multistart single objective optimization algorithms. The search space is explored in a single optimization run. More powerful are combinations of deterministic and EA algorithms including specific-problem knowledge that allows the reduction of the search space. Only in the past years has the MO character of the brachytherapy planning problem been recognized. The approach enables to obtain better solutions as the alternatives are known (see Fig. 19.11).

Fig. 19.10. Principal steps of a genetic algorithm

1. Initialize population chromosome values.
2. Assign fitness for each individual.
3. Select individuals for reproduction, dominance based for MO optimization and fitness based for SO optimization.
4. Perform crossover between random selected pairs with probability $p_C$.
5. Perform mutation with probability $p_M$.
6. Stop. If maximum generation is reached or any other stopping criterion is satisfied, see point 2.

Fig. 19.11a–c. Example of a Pareto front obtained with MO optimization with 231 representative Pareto optimal solutions for a prostate implant, for three variance-based objectives, $f_S$, $f_V$ and $f_{\text{urethra}}$ for the conformity and homogeneity within the PTV and for the organs at risk (OAR) urethra, respectively (see also Milickovic et al. 2002). The three 2D projections of the 3D front are shown. a Conformity–homogeneity Pareto front. b Conformity–OAR Pareto front. c Homogeneity–OAR Pareto front. The selected solution based on the trade-off between PTV dose coverage and protection of the urethra is marked with red
Figure 19.12 demonstrates the results of a multi-objective optimization for a prostate monotherapy implant using the variance-based objectives: conformity and homogeneity for PTV; OAR urethra with a dose limit (critical dose, $D_{1\%}$, value for the $f_1$ objective in Eq. (15)) of 125%; and OAR rectum with a dose limit (critical dose, $D_{1\%}$, value for the $f_1$ objective in Eq. (15)) of 85% of the reference dose (100%).

The multiobjectivity of the anatomy-based optimization and the need of decision tools is clearly demonstrated in Fig. 19.12.

19.3.5 Dose Evaluation

There have been several concepts and parameters defined and proposed for the evaluation of the 3D dose distribution in brachytherapy.

ICRU report 58 (ICRU 1997) offers an extended summary of the classical parameters that could be used for evaluating the dose distribution, but it is mainly focused on the system-based treatment planning in interstitial brachytherapy. This report, on the
other hand, introduces for the first time anatomy-oriented parameters for evaluation as well as the volume definitions as already known in the external beam treatment planning: gross tumour volume (GTV); clinical target volume (CTV); planning target volume (PTV); and treated volume (TV).

In this report it is recommended to use the minimum target dose (MTD) defined as the minimum dose at the periphery of CTV and the mean central dose (MCD) that is taken as the arithmetic mean of the local minimum doses between sources or catheters in the central plane for reporting and evaluating. The latest is according to the Paris system of dosimetry in interstitial low dose rate (LDR) brachytherapy. Finally, the high dose volume, defined as the volumes encompassed by the isodose corresponding to 150% of the MCD, and the low dose volume, defined as the volume within the CTV, encompassed by the 90% isodose value, are considered.

In the past, efforts have been made for establishing some parameters or figures to describe the homogeneity of the dose distribution in brachytherapy; all of these have been based on the absence of an anatomical dose calculation space. In other words, these efforts have considered the dose distribution simply around the catheters. The method that found a wide application, at least in the field of low-dose-rate brachytherapy, is that of natural dose volume histogram (NDVH) introduced by Anderson (Anderson 1986).

Currently published recommendations (Ash et al. 2000; Nag et al. 1999; Nag et al. 2001; Pötter et al. 2002) proposed DVH-based parameters (cumulative DVHs) for the evaluation and documentation of the dose distribution (see below).

**PTV-Oriented Parameters**

**D100**: the dose that covers 100% of the PTV volume, which is exactly the MTD proposed by ICRU report 58, if we assume that CTV equals PTV for brachytherapy.

**D90**: the dose that covers 90% of the PTV volume

**V100**: the percentage of the PTV volume that has received at least the prescribed dose, which is set to 100%.

**V150**: the volume, normally the PTV volume, that has received at least 150% of the prescribed dose.

The definition of these parameters is graphically shown in Fig. 19.13. Statistical values, such as the mean dose value (Dmean) and the standard deviation of the dose value distribution in the PTV, can also be used.

**OAR-Oriented Parameters**

For the OARs there are not widely established dosimetric parameters. The only exception are the values considered to be representative for the irradiation of bladder and rectum for the primary intracavitary brachytherapy of the cervix carcinoma as proposed by ICRU 38 (ICRU 1985; Pötter et al. 2002). Pötter et al. (2002) propose to use the maximum dose for the OARs, at least for the case of primary intracavitary brachytherapy of the cervix carcinoma, where the maximum doses are considered to be the doses received in a volume of at least 2 and 5 cm³ for bladder and rectum, respectively.

Furthermore, the D10, defined as the highest dose covering 10% of the OAR volume, is commonly used for the interstitial brachytherapy of prostate cancer for documenting the dose distribution in the related OARs urethra, rectum and bladder.

**The Conformal Index**

Baltas et al. (1998) has introduced a utility function as a measure of the implant quality, the conformal index (COIN), which has been later expanded to include OARs (Milickovic et al. 2002). The COIN takes into account patient anatomy, both of the PTV, surrounding normal tissue (NT) and OARs. The COIN for the reference dose value, D_ref (prescribed dose), is defined as:

\[
COIN = c_1 \cdot c_2 \cdot c_3
\]
\[ c_1 = \frac{\text{PTV}_{\text{ref}}}{\text{PTV}} \quad c_2 = \frac{\text{PTV}_{\text{ref}}}{V_{\text{ref}}} \]

\[ c_3 = \prod_{i=1}^{N_{\text{OAR}}} \left( 1 - \frac{V^{i}_{\text{OAR}}(D > D^{i}_{\text{crit}})}{V^{i}_{\text{OAR}}} \right) \]

where the coefficient \( c_1 \) is the fraction of the PTV (\( \text{PTV}_{\text{ref}} \)) that receives dose values of at least \( D_{\text{ref}} \). The coefficient \( c_2 \) is the fraction of the reference isodose volume \( V_{\text{ref}} \) that is within the PTV (see also Fig. 19.14).

\( V^{i}_{\text{OAR}} \) is the volume of the \( i \)th OAR and \( V^{i}_{\text{OAR}}(D > D^{i}_{\text{crit}}) \) is the volume of the \( i \)th OAR that receives a dose that exceeds the critical dose value \( D^{i}_{\text{crit}} \) defined for that OAR. The product in the equation for \( c_3 \) is calculated for all \( N_{\text{OAR}} \) OARs included in the treatment planning.

In the case where an OAR receives a dose, \( D \), above the critical value defined for that structure, the conformity index will be reduced by a fraction that is proportional to the volume that exceeds this limit. The ideal situation is \( \text{COIN} = c_1 = c_2 = c_3 = 1 \).

The COIN assumes in this form that the PTV, the OARs and the surrounding normal tissue are of the same importance.

When the COIN value is calculated for every dose value, \( D \), according to Eq. (17), then the conformity distribution or the COIN histogram is calculated.

Figure 19.15 demonstrates the conformity distribution for the implant shown in Fig. 19.12 and for the solution having the maximum COIN value. A good implant is that where the maximum COIN value is observed exactly at or near the reference dose \( V_{\text{ref}} \) (100%).

![Fig. 19.14](image1) volumes necessary for computation of the conformal index COIN

![Fig. 19.15a,b](image2) Evaluation results for the implant shown in Fig. 19.12. Here the solution with the highest maximum COIN value is selected. The critical dose values or dose limits for the OARs urethra und rectum were 125 and 85% of the reference dose (100%). a The cumulative DVHs for PTV, and the OARs urethra and rectum. b The conformity distribution (COIN distribution) calculated according to Eq. (17) and based on the above DVHs including both OARs demonstrates a COIN value of 0.90 for the 100% reference isodose line. The maximum COIN value is 0.91 and it is observed for the 98% isodose value that is very close to the reference dose for that implant and treatment plan selected.

### References


Ash D, Flynn A, Battermann J, de Reijke T, Lavagnini P, Blank
compliant analytical dosimetry model in clinical $^{192}$Ir HDR brachytherapy treatment planning and assessment of the significance of source position and catheter reconstruction uncertainties. Phys Med Biol 49:55–67


New Treatment Techniques
20 Beam Delivery in 3D Conformal Radiotherapy Using Multi-Leaf Collimators

W. Schlegel, K.H. Grosser, P. Häring, and B. Rhein

20.1 Conformal Treatment Techniques

Conformal radiation therapy (CRT) was introduced in the early 1960s by radiation oncologist S. Takahashi, who came up with many ideas of how to concentrate the dose to the target volume using various forms of axial transverse tomography and rotating multi-leaf collimators (MLC; Takahashi 1965).

Three-dimensional conformal radiotherapy (3D CRT) is an extension of CRT by the inclusion of 3D treatment planning and can be considered to be one of the most important advances in treating patients with malignant disease. It is performed in nearly all modern radiotherapy units.

The goal of 3D CRT is the delivery of a high radiation dose which is precisely conformed to the target volume while keeping normal tissue complications at a minimum.

The preconditions which have to be fulfilled in order to achieve conformal dose distributions are discussed in the preceding chapters (Chaps. 2–12) on imaging and treatment planning (Chaps. 13–19). In summary, it can be said that first of all detailed diagnostic imaging information has to be available from a variety of sources including conventional X-ray imaging, CT, MRI and PET in order to be able to define the target volume and the organs at risk with sufficient accuracy. Furthermore, a 3D computerized treatment planning system and an exact and reproducible patient positioning system have to be used. If these boundary conditions are fulfilled, conformal irradiation techniques can be used optimally.

In general, the attainable dose conformity in conventional conformal radiotherapy depends on the boundary conditions described in Table 20.1. As is seen from this table, there are many approaches to conformal therapy using sophisticated irradiation techniques with multiple isocentric beam irradiations, irregularly shaped fields (either using cerrobend blocks or MLCs), and computer-controlled dynamic techniques.

An important step in 3D CRT was the introduction of irregularly shaped irradiation fields, made of metal blocks from alloys with low melting points (in radiotherapy often called “cerrobend” blocks). Individually shaped irregular fields realized by cerrobend blocking turned out to be time-consuming and expensive; therefore, great progress in conformal radiotherapy was achieved by the development and application of MLCs. The dose distributions achieved with MLCs turned out to be equivalent to conformal blocks; however, the cost of conformal radiation therapy could be minimized and the flexibility significantly enhanced by using the new MLC technology (Adams et al. 1999; Foroudi et al. 2000).

The MLCs are beam-shaping devices that consist of two opposing banks of attenuating leaves, each of which can be positioned independently. The leaves can either be moved manually or driven by motors to such positions that, seen from the “beam’s eye view” of the irradiation source, the collimator opening corresponds to the shape of the tumor (Fig. 20.1). The leaf settings are usually defined within the virtual

CONTENTS

20.1 Conformal Treatment Techniques 257
20.2 Multi-Leaf Collimators 258
20.2.1 Geometrical and Mechanical Properties 258
20.2.2 Physical Properties 262
20.2.2.1 Focusing Properties and Penumbra 262
20.2.2.2 Interleaf Leakage 263
20.2.2.3 Leaf Transmission 263
20.2.3 Operating Modes 265
20.3 Commercial MLCs 265
20.3.1 Linac-Integrated MLCs 265
20.3.2 Accessory-Type MLCs 265
20.4 The Limits of Conventional Conformal Radiation Therapy 266
References 265

There were different other designs of MLCs with up to six leaf banks (Topolnjak et al. 2004), which, however, up to now have not played an important role in the clinical practice of 3D CRT.

---

### Table 20.1 Irradiation techniques for conformal radiotherapy, MLC multi-leaf collimator

<table>
<thead>
<tr>
<th>Physical parameter</th>
<th>Impact on conformity</th>
<th>Scorecard</th>
<th>Drawbacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of beam incidents</td>
<td>Better conformity can be achieved with a higher number of irradiating directions</td>
<td>++</td>
<td>Higher complexity, longer planning time, normal tissue dose becomes possibly larger, longer irradiation time</td>
</tr>
<tr>
<td>Optimization of beam directions</td>
<td>Higher conformity possible</td>
<td>++</td>
<td>Higher complexity, longer planning time; in case of non-coplanar beam directions: longer treatment time</td>
</tr>
<tr>
<td>Optimization of beam energy (photons)</td>
<td>Higher conformity possible</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Application of an MLC</td>
<td>Higher conformity possible</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>Width of the MLC leaves</td>
<td>Small leaf width enables a better field adjustment and therefore better conformity</td>
<td>++</td>
<td>Higher complexity, longer planning time</td>
</tr>
<tr>
<td>More than one target point at the same time</td>
<td>In cases where more than one target volume is treated simultaneously, conformity might be higher</td>
<td>+</td>
<td>Higher complexity, longer planning time, sometimes a lower homogeneity</td>
</tr>
<tr>
<td>Moving beam irradiations; computer-controlled dynamic radiotherapy</td>
<td>Depends on shape of target volume</td>
<td>+</td>
<td>Longer irradiation time, complex verification and quality assurance</td>
</tr>
</tbody>
</table>

---

### 20.2 Multi-Leaf Collimators

Multi-leaf collimators permit the quick and flexible adjustment of the irradiation fields to the tumor shape and the shape of the organs at risk. Though already proposed by Takahashi in 1960, it took nearly 25 years before the first commercial computer controlled MLCs appeared on the market. This was due to the fact that MLCs are mechanical devices with high mechanical complexity, and they have to fulfill very rigid technical, dosimetric, and safety constraints. Detailed reviews of the history and performance of MLCs for 3D CRT are given by Webb (1993, 1997, 2000). The use of MLCs for static or dynamic IMRT is discussed in more detail in another work by Webb (2005).

This chapter describes briefly the general design and performance of the MLCs as they are currently being used in routine applications for 3D CRT.

---

### 20.2.1 Geometrical and Mechanical Properties

The most important technical parameters (Fig. 20.2) which characterize the performance of an MLC are mechanical and geometrical properties such as:

1. The maximum field size
2. The leaf width
3. Maximum overtravel
4. Interdigitation
5. Configuration of the MLC with respect to the collimator jaws

For MLCs which are used for IMRT, other important parameters are also the minimum and maximum leaf speed and the precision of leaf positioning. Other aspects are of course the complexity of calibration and the overall efforts for maintenance.

![Multi-leaf collimator with the most relevant mechanical parameters](image)

**Maximum Field Size**

Two kinds of MLCs are employed presently: those for medium-sized and large fields of up to 40×40 cm², which are implemented in the gantry of linacs; and “add-on-MLCs” for small field sizes (often called mini- or micro-MLCs) which can be attached to the accessory holder of the treatment head, and, for example, used in conjunction with stereotactic conformal radiotherapy. Mini- and micro-MLCs have characteristic maximum field sizes of about 10×10 cm². Maximum field sizes depend for some MLCs (see Tables 20.2, 20.3) from the maximum overtravel: when the maximum overtravel is used, maximum field size will become smaller, because the whole leaf banks have to be shifted in order to achieve complete overtravel.

**Leaf Width**

MLCs integrated into the linac head. Computer-controlled MLCs integrated into the head of the accelerator usually have a spatial resolution of 0.5–1 cm in the isocenter plane, perpendicular to the leaf-motion direction, and a positioning accuracy in the range of 1 mm in the direction of the motion.

The leaf width (measured in the isocenter plane) should be adapted to the size and complexity of the target volumes. Maybe an effective leaf width of 10 mm is completely sufficient in case of prostate cancer; however, in the case of a small target volume located around the spinal cord, 10 mm is too large! A leaf width of 5 mm is presently considered to be a good compromise.

**Table 20.2 Commercial integrated MLCs**

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Product name</th>
<th>Leaf width at isocenter (mm)</th>
<th>Midline overtravel (cm)</th>
<th>No. of leaves</th>
<th>Maximum field size (cm²)</th>
<th>Focusing properties</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elekta-1</td>
<td>Integrated MLC</td>
<td>10</td>
<td>12.5</td>
<td>40×2</td>
<td>40×40</td>
<td>Single focusing</td>
<td></td>
</tr>
<tr>
<td>Elekta-2</td>
<td>Beam modulator</td>
<td>4</td>
<td>11</td>
<td>40×2</td>
<td>16×22</td>
<td>Single focusing</td>
<td></td>
</tr>
<tr>
<td>Siemens-1</td>
<td>3D MLC</td>
<td>10</td>
<td>10</td>
<td>29×2</td>
<td>40×40</td>
<td>Double focusing</td>
<td></td>
</tr>
<tr>
<td>Siemens-2</td>
<td>Optifocus</td>
<td>10</td>
<td>10</td>
<td>41×2</td>
<td>40×40</td>
<td>Double focusing</td>
<td></td>
</tr>
<tr>
<td>Siemens-3</td>
<td>160 MLC</td>
<td>5</td>
<td>20ᵃ</td>
<td>80×2</td>
<td>40×40</td>
<td>Single focusing</td>
<td>Announced for 2006</td>
</tr>
<tr>
<td>Varian-1</td>
<td>Millennium MLC-52</td>
<td>10</td>
<td>20ᵃ</td>
<td>26×2</td>
<td>26×40</td>
<td>Single focusing</td>
<td></td>
</tr>
<tr>
<td>Varian-2</td>
<td>Millennium MLC-80</td>
<td>10</td>
<td>20ᵃ</td>
<td>40×2</td>
<td>40×40</td>
<td>Single focusing</td>
<td></td>
</tr>
<tr>
<td>Varian-3</td>
<td>Millennium MLC-120</td>
<td>Central 20 cm of field: 5 mm; outer 20 cm of field: 10 mm</td>
<td>20ᵃ</td>
<td>60×2</td>
<td>40×40</td>
<td>Single focusing</td>
<td></td>
</tr>
</tbody>
</table>

ᵃRequires movement of the complete leaf bank and leads to reduced maximum field sizes
ᵇThe Siemens 3D MLC consists of 2×27 inner leaves with 1-cm leaf width and two outer leaves with 6.5-cm leaf width
Some of the commercially available MLCs have sections with various leaf widths. The central section of the leaf bank has a higher spatial resolution than the outer sections.

Examples of integrated MLCs are shown in Fig. 20.3.

**Accessory-Type MLCs.** The need for conformal, homogeneous dose distributions in connection with stereotactic radiotherapy and radiosurgery treatments was the motivation for the development of high-resolution MLCs for small field sizes. These collimators are attachable to the accessory holder of the linac (Schlegel et al. 1993, 1997; Debus et al. 1997).

The leaf resolution is in the range of 1.5–4 mm (see Table 20.2).

As examples for accessory MLCs used in stereotactic treatments, Fig. 20.4 shows the manual and the computer-controlled micro-MLCs developed at DKFZ (Heidelberg, Germany; Schlegel et al. 1992, 1993, 1997).

**The Optimum Leaf Width of a MLC.** In general, it seems evident, that the higher the spatial resolution of the MLC, the better the quality of the resulting dose distributions formed with such a MLC. This has empirically been shown using clinical treatment planning examples for irregularly shaped target volumes (Föller et al. 1998; Nill 2002). There is, however, a definitive limit given as a physical constraint in principle: for a MLC with the penumbra p (=distance between the 20 and 80% isodose produced by the leaf edge) a leaf width finer than p/2 does not lead to further improvement in the dose distribution. This was concluded by Bortfeld et al. (2000) according to sampling considerations. For a 6-MV beam, for instance, the optimum leaf width of a stereotactic add-

![Fig. 20.3 a Integrated MLC with a leaf width of 1 cm at the isocenter (the 3D-MLC from Siemens; see Table 20.2). b New-generation MLC: the 160 MLC from Siemens (see Table 20.2) on MLC with a penumbra of approximately 3 mm therefore is in the range of 1.5–2 mm. An integrated MLC is much closer positioned to the target and has a penumbra of 8–10 mm. The optimum leaf width is thus in the range of about 5 mm.**

Table 20.3 Commercial add-on MLCs (mini- and micro-MLCs)

<table>
<thead>
<tr>
<th>Company</th>
<th>BrainLAB (m3)</th>
<th>Radionics</th>
<th>Siemens (MRC) μ-MLC</th>
<th>Siemens (MRC) Moduleaf</th>
<th>3D Line (Wellhöfer)</th>
<th>Direx AccuLeaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of leaf pairs</td>
<td>26</td>
<td>31</td>
<td>40</td>
<td>40</td>
<td>24</td>
<td>36</td>
</tr>
<tr>
<td>Field size (cm²)</td>
<td>10x10</td>
<td>10x12</td>
<td>7.3x6.4</td>
<td>12x10</td>
<td>11x10</td>
<td>11x10</td>
</tr>
<tr>
<td>Overcenter travel (cm)</td>
<td>5</td>
<td>No data</td>
<td>1.4</td>
<td>5.5</td>
<td>2.5</td>
<td>No data</td>
</tr>
<tr>
<td>Leaf width (mm)</td>
<td>3.0–5.5</td>
<td>4.0</td>
<td>1.6</td>
<td>2.5</td>
<td>4.5</td>
<td>No data</td>
</tr>
<tr>
<td>Leaf transmission (%)</td>
<td>&lt;4</td>
<td>&lt;2</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>0.5</td>
<td>&lt;2</td>
</tr>
<tr>
<td>Maximum speed (cm/s)</td>
<td>1.5</td>
<td>2.5</td>
<td>1.5</td>
<td>3</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>Clearance to isocenter (cm)</td>
<td>31</td>
<td>35</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td>Total weight (kg)</td>
<td>31</td>
<td>35</td>
<td>38</td>
<td>39.7</td>
<td>35</td>
<td>27</td>
</tr>
<tr>
<td>Geometric design</td>
<td>Single focused</td>
<td>Single focused</td>
<td>Parallel</td>
<td>Single focused</td>
<td>Double focused</td>
<td>Two sets of leaf pairs at 90°</td>
</tr>
</tbody>
</table>
Maximum Overtravel

The overtravel characterizes how far a leaf can be moved over the midline of the MLC (Fig. 20.2). A large overtravel is important for very complexly shaped target volumes, but even more for the production of intensity-modulated fields in IMRT. Large overtravel is a mechanical challenge, because very long leaves are needed, which may lead to big weights and mechanical guiding problems. Overtravel distances for commercial collimators are listed in Tables 20.2 and 20.3. It has to be recognized that complete overtravel can only be realized in some collimators by moving a whole leaf bank (which is the case in all Varian MLCs and in the 160-MLC from Siemens). Moving the whole leaf bank of course reduces the maximum field size.

Interdigitation

In some cases one leaf cannot pass an adjacent opposing leaf without collision (Fig. 20.5); thus, fields designed without considering such constraints cannot be delivered with such an MLC. This is generally not such an important issue for conventional conformal radiotherapy, but for IMRT applications, where many small and often complexly shaped segments have to be delivered, such leaf “interdigitation” is often required.

MLC configuration in the treatment head.

MLC configurations with respect to the rectangular collimator jaws may be the following:

1. Total replacement of the upper jaws
2. Total replacement of the lower jaws
3. Tertiary collimator configuration

The three main vendors of integrated MLCs have chosen different configurations leading to different performances, especially for dosimetric properties as leakage and penumbra. The configurations are illustrated in Fig. 20.6. (Accessory-type MLCs are always used in the tertiary configuration, of course.)
20.2.2 Physical Properties

20.2.2.1 Focusing Properties and Penumbra

The penumbra is an important design feature of a beam defining device. In order to obtain a steep dose gradient between the target volume and healthy tissue, the penumbra has to be as small as possible.

First of all, penumbra depends on the position of the collimator relative to the source and the patient's surface and on the diameter of the source. As a rule, in order to obtain a small penumbra, the source diameter has to be as small as possible (2–3 mm in modern linacs) and the distance between the source and the collimator as large as possible. On the other hand, the clearance between the patient and the irradiation head should be as large as possible in order to have the full flexibility both for using accessories (block trays, wedges, or accessory MLCs) and to apply non-coplanar beams. In that sense, a compromise has to be made between penumbra and clearance.

Secondly, the penumbra also depends on the collimator edges. In an MLC, in order to produce a small penumbra, the edges of the leaves must always be directed towards the source, independent of the leaf position. This property is called “focusing.”

**Focusing Perpendicular to the Leaf Motion Direction**

Good focusing properties are reached by trapezoid leaf cross sections which causes focusing in the direction perpendicular to the leaf motion (Fig. 20.7a).

**Focusing in the Direction of the Leaf Motion**

Focusing in the leaf direction can be obtained by moving the leaves on a circular path or by rotating the leaf edges (see Fig. 20.7c, d; Pastyr et al. 2001). Both solutions are connected with engineering problems. That is why in most modern MLCs curved edges are being used which also give a reasonable penumbra (Fig. 20.7b). In the case of curved leaves penumbra is, however, not completely independent of the position of a leaf (Butson et al. 2003). The penumbra variation has to be implemented into the treatment planning systems. It also may complicate the delivery especially of small off-center segments in IMRT.

---

**Fig. 20.7a-d.** Focusing properties of MLCs: the leaves have trapezoid cross sections to perform focusing perpendicular to the direction of the leaf motion (a). Focusing in the direction of leaf motion can either be realized by leaves traveling on a circular path (b), rounded leaf edges (c), or rotating leaf edges (d).
The three main vendors of linear accelerators (Siemens, Elekta, and Varian) have implemented their MLCs at different distances from the source, and they also have different edge designs of the MLC leaves (Table 20.2). This leads to different performances and has to be considered when purchasing a new linear accelerator.

From a dosimetric point of view, penumbra is normally specified as the distance between the 20 and 80% isodose line. If, in the worst case, the leaf movement has a direction of 45° to the isodose lines, the penumbra will become larger for leaves with bigger leaf width, because of the ribble which superimposes to the isodose lines (Fig. 20.8). Figure 20.9 shows dosimetric film measurements for this 45° situation and a physical leaf width of 1, 2, and 3 mm. It can easily be recognized that penumbra is increasing with leaf width.

20.2.2.2 Interleaf Leakage

In order to avoid friction, there has to be small gap of about 0.1 mm between the leaves. This gap causes leakage radiation, which has to be minimized below a level of about 4%. This is especially a problem when the leaves have a trapezoid cross section for beam focusing (Fig. 20.10b). To suppress interleaf leakage, the leaves are manufactured using a tongue-and-groove design (Fig. 20.10c). Another trick to reduce interleaf leakage is to slant the whole arrangement of the leaves with respect to the direction of the divergent rays (Fig. 20.10d).

Interleaf leakage cannot be avoided completely by any of the abovementioned leaf designs. Figure 20.11 shows dosimetric film measurements with the typical spikes caused by interleaf leakage.

20.2.2.3 Leaf Transmission

When high energy X-rays have to collimated, there is always a small fraction of X-rays which will penetrate through the jaws or leaves (Fig. 20.12). That is why high-Z material such as tungsten has to be used for the jaws or leaves. For tungsten, the thickness of the material has still to be in the range of 8–10 cm in order to reduce transmission below 1%.

In general, the fraction of intensity transmitted through the collimators is higher in IMRT step-and-shoot treatments than in conventional treatments because the treatment volume is irradiated with more field components to reach the prescribed dose level.
At a rough guess, the transmitted intensity is twice the original physical level. If the transmission is, for example, 2%, the maximum of the transmitted intensity mounts up to 4% at some positions. Especially for MLCs used in IMRT, transmission as well as interleaf leakage should therefore be kept as low as possible.

Restrictions for leakage radiation of MLCs are given in IEC (1998): If the MLC is covered by rectangular jaws, which are automatically adjusted to the MLC shape, leakage radiation must be below 5% of an open 10×10-cm² field; otherwise, maximum leakage should be less than 2% and average leakage less than 0.5%.
20.2.3 Operating Modes

There are in principle two different operating modes for MLCs for 3D CRT, depending on whether the leaves are moving when the beam is on (dynamic mode) or the beam is shut off (static mode).

Although modern MLCs in principle have dynamic properties, they are still commonly applied in static treatment techniques. The real potential of MLCs in 3D CRT will be demonstrated in the future, when, for example, dynamic arc treatment techniques with MLCs will be more widely accepted and implemented to perform dynamic field shaping.

The two modes (static mode=step-and-shoot technique; dynamic mode=sliding-window technique) currently play a bigger role in IMRT delivery (see Chap. 23). In IMRT dynamic delivery is in many cases more time efficient than step and shoot, but the step-and-shoot technique is less complex and quality assurance may be easier to perform. CHUI et al. (2001) compared the delivery of IMRT by dynamic and static techniques.

20.3 Commercial MLCs

20.3.1 Linac-Integrated MLCs

The major manufacturers of commercial MLCs integrated into the irradiation head of a linear accelerator are the companies Elekta, Siemens, and Varian. GALVIN (1999) provides a useful review with tables of MLC properties. HUQ et al. (2002) compared the MLCs of all three manufacturers using precisely the same criteria and experimental methods. The result of this investigation was that there is no clear superiority of one vendor compared with the others: the different designs and configurations of the MLCs are leading to a balance of advantages and disadvantages and there was no clear superiority of one MLC compared with the others. An updated list of MLC characteristics is given in Table 20.2.

20.3.2 Accessory-Type MLCs

There are a couple of companies manufacturing and distributing accessory MLCs, which are especially suited to treat small target volumes in conjunction with stereotactic irradiation techniques. BORTFELD et al. (1999) have provided an overview on the characteristics and performances of these mini- and micro-MLCs. The specifications of these add-on collimators are summarized in Table 20.3. In summary, MLCs of this type are closer positioned to the patient’s surface and have smaller leaf widths. As a consequence, mini- and micro-MLCs have much smaller penumbra and produce dose distributions with higher conformity; however, they are restricted to the treatment of small target volumes.

20.4 The Limits of Conventional Conformal Radiation Therapy

Complex-shaped target volumes close to radio-sensitive organs remain a challenge for the radiotherapist. With MLCs and conventional irradiation techniques conformal and homogeneous dose distributions cannot be obtained in all cases. This is particularly true for concave-shaped target volumes. In Chaps. 17 and 23 (Inverse planning and IMRT) of this book it is shown that as the result of a superposition of several irradiation segments with homogeneous intensity, a concave-shaped dose distribution is produced. The MLCs to produce these IMRT field segments have to fulfill very special criteria concerning leaf speed, accuracy, and reproducibility of leaf positioning, overtravel, transmission, and leakage. The specific requirements of MLCs in IMRT are discussed and analyzed in much more detail by SCHLEGEL and MAHR (2000) and WEBB (2005).

References


IEC (1998) IEC 601-2-1, 2nd edn
21 Stereotactic Radiotherapy/Radiosurgery

Anca-Ligia Grosu, Peter Kneschaurek, and Wolfgang Schlegel

21.1 Introduction

Stereotactic radiotherapy dates back more than 50 years; however, this form of treatment has entered the domain of radiation oncology only in the past 10–15 years. Initially an exotic technique, stereotactic radiotherapy has become an established, widespread treatment approach, characterized by the delivery of the irradiation with a very high geometrical precision. The method is especially used for benign and malignant cranial tumors but has lately been adapted to the body.

Stereotactic coordinates are used in the methodology of different irradiation techniques: implantation of seeds; Bragg-Peak irradiation; irradiation with gamma knife (Co60) or with linear accelerator (LINAC or X-Knife). This chapter focuses on the description of the technique of stereotactic teletherapy (percutaneous stereotactic radiotherapy) with LINAC, also known as stereotactic convergent beam irradiation. This is the most widespread irradiation technique using stereotactic coordinates.

21.2 Definition

Stereotaxy (stereo + taxis – Greek, orientation in space) is a method which defines a point in the patient's body by using an external three-dimensional coordinate system which is rigidly attached to the patient. Stereotactic radiotherapy uses this technique to position a target reference point, defined in the tumor, in the isocenter of the radiation machine (LINAC, gamma knife, etc.) with a high accuracy. This results in a highly precise delivery of the radiation dose to an exactly defined target (tumor) volume. The aim is to encompass the target volume in the high-dose area and, by means of a steep dose gradient, to spare the surrounding normal tissue. This allows the definition of the planning target volume (PTV) without, or with only a very small (1–2 mm), safety margin around the gross tumor volume (GTV) or the clinical target volume (CTV). Stereotactic radiotherapy is performed in various treatment techniques, including gamma-knife units using 60 Co photons, heavy charged particles, or neutron beams units and modified LINAC units.

Treatment can be administered in a single fraction (radiosurgery, RS) or as multiple fractions (stereotactic fractionated radiotherapy, SFR). Some authors use the term stereotactic radiotherapy for the fraction-
ated delivery of the radiation; however, we consider it as a general term for all the irradiation treatments which are characterized by the integration of stereotactic coordinates in the treatment planning and delivery, including the RS and the SFR.

The intention of RS is to produce enough cell kill within the target volume in a single fraction in order to eradicate the tumor. This single high irradiation dose can produce considerable side effects in normal tissue located close to the tumor or within the target volume. The SFR combines the precision of target localization and dose application of RS with the radiobiological advantage of fractionated radiotherapy, i.e., breaking the total dose into smaller parts and thus allowing repair of DNA damage in normal tissue during the time between fractions. Time intervals of more than 6 h between fractions can significantly reduce the risk of side effects in normal tissue (HALL and BRENNER 1993; SCHLEGEL et al. 1993).

### 21.3 Historical Background

The first one to combine stereotactic methodology – which up to that point was used only in neurosurgery – with radiation therapy was the Swedish neurosurgeon Lars Leksell. He had the idea that X-ray could be used to treat certain neurological functional diseases instead of stereotactically guided needles used in neurosurgery. Leksell performed the first treatment in 1951, at the Karolinska Institute, and called the new therapy approach radiosurgery (RS). The patient was fixed in a stereotactic ring and irradiated with a precisely guided roentgen tube using 200-kV Röntgenbremsstrahlung (LEKSELL 1951). Bragg-peak studies with protons were begun in Uppsala, Boston, and Berkeley (KJELLBERG et al. 1968; LARSSON et al. 1958; LAWRENCE et al. 1962). In Berkeley Bragg-peak RS using helium ion beams was also developed (LYMAN et al. 1977). Leksell continued his work and built the first isotope radiation machine, in 1968, the gamma knife.

The gamma-knife unit consists of 201 Co-60 sources arranged on the surface of a hemispherical shell, each aiming at an isocenter in a uniform distance of 40 cm. Each source is collimated by a fixed primary collimator and subsequently by a helmet, which consists of 201 collimators to which the patient and frame are attached and which is docked with the primary collimator array at the time of treatment. The helmets define an approximately spherical dose distribution at the isocenter with nominal diameters of 4, 8, 14, or 18 mm. The physical point of interest inside the patient is positioned at the isocenter of this distribution before treatment, and, if required, a sequence of treatment at different isocenters, with possibly different diameter helmets, are used to produce a conformal dose distribution. Multiple isocenters treatments are required when the target shape is irregular. By combining several spherical dose distributions of smaller diameter in the appropriate location, it is possible to create a sum of isodose distributions similar to the target volume. As a result the dose to the target itself is inhomogeneous, with larger number of isocenters leading to higher mean target doses. To date, the radiobiological consequences of dose inhomogeneity in tumor tissue cannot be evaluated yet, but the advantages of dose escalation in the tumor and dose reduction in the normal tissue are undisputed (LINDQUIST et al. 1995; VERHEY and SMITH 1995).

The stereotactic radiation therapy with LINAC started in the early 1980s: the Swedish physicist Larsson proposed to use the LINAC instead Co 60 or protons (LARSSON et al. 1974). The first published reports on clinical use of LINAC came from Buenos Aires (BETTI and DERECHINSKY 1983), Heidelberg (HARTMANN et al. 1985), and Vicenza (COLOMBO et al. 1985), the authors of which all had developed the concept to deliver a single-dose radiation with convergent non-coplanar dynamic irradiation. Many clinical investigators made use of this technology all over the world in a very short time, showing, in comparison with the gamma-knife RS, comparable clinical results (ENGENHART et al. 1989, 1992; MEHTA et al. 1995; BECKER et al. 1996; DEBUS et al. 1999).

Further progress was made in LINAC stereotactic radiotherapy due to the development of non-invasive, replaceable head-fixation systems, which allow the implementation of the dose fractionation (SCHLEGEL et al. 1992, 1993; STÄRKE et al. 1997). Another important development is the micro-multileaf collimator which permits the adoption of the irradiated area to irregular-shaped target volumes (SCHLEGEL et al. 1997) and the extension of the stereotactic irradiation technique to the head and neck (GADEMANN et al. 1993) and the body (HERFARTH et al. 2000).

### 21.4 Stereotactic Coordinates

The key idea of all stereotactic radiation therapy techniques is the use of stereotactic coordinates to
define the volume which has to be irradiated three dimensionally; therefore, target volume and anatomical structures are localized in the space defined by the stereotactic coordinates system. Computed tomography (CT) or magnetic resonance imaging (MRI) are used for defining the anatomical structures which are of crucial importance for radiation therapy. The stereotactic localizer which is imaged simultaneously allows transfer of the image coordinates into the stereotactic coordinates system. During the planning of the radiation therapy a target point in the stereotactic space is defined. Before onset of radiation the patient is positioned in such a way that this target point is placed in the isocenter of the radiation machine with the help of the stereotactic positioning system.

In general, the stereotactic coordinates are a cartesian three-dimensional coordinates system attached to the stereotactic frame in a rigid relationship. The origin of the stereotactic coordinates system is generally in the center of the volume defined by the stereotactic frame: the x and y axes correspond to the lateral and frontal side of the frame and the z axis to the cranio-caudal direction (Fig. 21.1).

### 21.5 Method of Stereotactic Irradiation Treatment

The main steps in the planning and delivering of stereotactic irradiation treatment are:
1. Rigid application of the stereotactic frame to the patient
2. Imaging (CT, MRI, angiography) of the patient with the frame and localizer attached to the frame
3. Treatment planning
4. Positioning of the patient for the stereotactic radiation therapy
5. Delivery of the irradiation
6. Quality assurance

#### 21.5.1 Stereotactic Frame

Stereotactic radiotherapy is based on the rigid connection of the stereotactic frame to the patient during CT, MRI, and angiography imaging (Figs. 21.2, 21.3). The stereotactic frame is the base for the fixation of the other stereotactic elements (localizer and positioner) and for the definition of the origin (point 0) of

---

Fig. 21.1 The stereotactic coordinates define three-dimensionally the space which has to be irradiated

Fig. 21.2 Head-ring fixation. A Carbon fiber posts. B, E Artifact-free fixation pins with ceramic tips. C Robust and lightweight frame. D Two torque wrenches for pressure adjustment of pins

Fig. 21.3 Non-invasive repeat head fixation with the mask system and upper jaw support
the stereotactic coordinates. The constant geometrical relationship between the stereotactic frame and anatomical structures, including the PTV, is realized by fixation of the frame to the patient. During the whole treatment procedure, from the performance of the stereotactic imaging to the delivery of the irradiation treatment, the stereotactic frame must not be removed from the patient. In case of relocatable frames it must be assured that the position of the patient is exactly the same relative to the frame after reapplication of the relocatable frame.

For the treatment of cranial lesions by RS the frame system is neurosurgically fixed onto the patient’s skull (Fig. 21.2). For SFR the head is fixed non-invasively in a relocatable thermoplastic mask attached on the stereotactic frame (Fig. 21.3).

There are different stereotactic frame systems described in detail in the literature: the BRW system (Brown 1979; Heilbrun et al. 1983); the CRW system (Couldwell and Apuzzo 1990; Spiegelmann and Friedman 1991); the Leksell system (Leksell and Jernberg 1980; Leksell et al. 1985); the Leibinger-Fischer system (Riechert and Mundinger 1955; Sturm et al. 1983); and the BrainLAB system (Stärk et al. 1997; Auer et al. 2002). Each system is different with regard to material of the stereotactic frame, design, and connection with the localizer and positioner and accuracy of repositioning (Kortmann et al. 1999).

21.5.2 Imaging for Stereotactic Irradiation Treatment

Imaging is used in stereotactic radiotherapy for: (a) localization and positioning; (b) definition of target volume and organs at risk; and (c) calculation and 3D representation of the isodose distribution.

Most stereotactic systems use CT for localization. During the CT investigation the localizer is attached to the frame (Fig. 21.4). The localizer is a box with CT-compatible fiducial markers on each plane, which are visualized on CT on each scan; thus, the localizer defines the link between the stereotactic coordinates and the imaging coordinates, so that for any point in the imaging the 3D stereotactic coordinates can be determined. The stereotactic frame, the patient fixation system, and the localizer form a fix unit. They have to be compatible with the radiological investigations and offer an accurate visualization of the tumor and of the critical structures, without artifacts. The use of MRI for localization and positioning needs a high homogeneity of the magnetic field to avoid spatial distortions artifacts, which could disturb the geometrical correlation between the stereotactic coordinates system and imaging coordinates system.

Non-invasive imaging techniques are a central component of treatment planning in radiation oncology. The information gained from different imaging modalities is usually of a complementary nature: (a)
MRI describes the anatomical structures of soft tissue with a high accuracy; (b) CT is important for the delineation of bone and soft tissue; (c) positron emission tomography (PET) and single photon computed emission tomography (SPECT) offer additional information about tumor extension and biology; and (d) angiography is essential for the visualization of the arterio-venous malformations; thus, the definition of tumor extension and critical structures is characterized by the correct integration of multiple different investigational tools, by using specialized image fusion software (Figs. 21.5, 21.6; Grosu et al. 2003).

The calculation and 3D representation of the isodose distribution is discussed in the next chapter.
21.5.3
Treatment Planning

21.5.3.1
Definition of Target Volume and Organs at Risk

The delineation of the target volume is a complex interactive process in which all the information of the imaging tools and the clinical information (operation, histopathology, other treatment approaches, etc.) are considered. The tumor-specific morphology, the growth pattern of the tumor, and the anatomical relationship to the normal tissue are essential parameters in defining the target volume.

Of major importance for the stereotactic radiation therapy is the delineation of the organs at risk. All the organs at risk which may get significant dose have to be delineated.

21.5.3.2
Definition of the Stereotactic Target Point

The target point is the point in the target volume that must be positioned with exact precision in the isocenter of the LINAC. The position of the target point can be defined interactively. One or more target points can be defined. In stereotactic planning programs the coordinates of the target points are related in such way that the resulting dose distribution meets the clinical requirements. The planning system outputs the position of these points in stereotactic coordinate. Prior to therapy, these coordinates will be used to correctly position the patient. This is performed with a positioner, a device attached to the stereotactic frame, which allows the connection of the stereotactic coordinate system to the room coordinate system, where the isocenter of the treatment device is defined (Fig. 21.4).

21.5.3.3
Planning of the Radiation Technique

The stereotactic radiation is characterized by a very steep dose fall-off on the margin of the target volume. The steep dose gradient is achieved by the use of appropriate collimators and a multitude of radiation directions.

Stereotactic Collimators. Tertiary stereotactic collimators for circular or oval target volumes are attached to the tray holder of the LINAC. The diameter of the irradiated area is defined by the size of the circular collimators and varies usually between 1 and 35 mm (Fig. 21.7). For irregular target volumes different individual apertures may be used, but the production and the use thereof is very cumbersome.

Only recently have micro-multileaf collimators become available (Fig. 21.8). The beam shape can be selected by computer or by hand. In this way the contours of the irradiation field can be adjusted individually to the tumor shape. Micro-multileaf collimators, in comparison with the traditional multi-leaf collimators, have the advantage of a decreased leaf width and therefore optimized the resolution (between 1 and 3 mm). Computer-controlled micro-multileaf

Fig. 21.7 Set of eight conical collimators

Fig. 21.8 Micro-multileaf collimator; 3-mm fine leaves at center; 4.5-mm intermediate leaves; and 5.5-mm outside leaves. Maximum field size is 10×10 cm
collimators offer the possibility of a dynamic field adjustment during irradiation and permit the implementation of the intensity-modulated radiotherapy (IMRT; Bortfeld et al. 1994a, b) and of the field shaping by dynamic arcs (Solberg et al. 2001).

Convergent Radiation Techniques. The radiation techniques are in general isocentric and implemented by using a rotational technique (using circular collimators or dynamic fields) or a static-field technique; both can be combined with an isocentric table rotation. In the rotational technique (Fig. 21.5) usually five to ten radiation arcs are used. The size and the angle between the arcs are variable and are responsible for the conformal isodose distribution. The stereotactic irradiation with the micro-multileaf collimator is done with multiple static irradiation fields (usually 6–12 fields; Fig. 21.6).

The following parameters can be defined interactively in the process of radiation planning: the number and position of the target points; the number of the radiation arcs and static fields and their shape; the position of the gantry and radiation table; and the radiation dose in the target point for each field or arc. By combining these parameters the radiation plan is developed.

21.5.3.4
3D Dose Calculation

The stereotactic radiation techniques are composed of complex rotation or multi-field irradiation. During the stereotactic radiation therapy planning the dose has to be calculated three-dimensionally on the basis of CT images. Most of the planning systems use CT images for the calculation of the correct dose. The planning software converts the Hounsfield number of the CT data into an electron density. Some planning software programs use MRI information only, by considering homogenous soft tissue density for the calculation of the dose. In comparison with 3D planning systems, the stereotactic radiation therapy can use simple, but clearly quicker, dose-calculation algorithms because no large-density inhomogeneities are in the brain. This simplification does not influence the precision of the dose calculation due to the use of a high number of subfields.

21.5.3.5
Dose Specification

The stereotactic dose distribution is defined as absolute or normalized dose distribution. The prescribed dose, \( D_0 \), is in this case the isodose surface which is intended to completely encompass the PTV. The minimal dose, \( D_{\text{min}} \), and the maximal dose, \( D_{\text{max}} \), in the PTV have to be specified as well. In the radiation plan, based on ICRU 50, different volumes have to be considered as well: PTV, treated volume, as well as the percentage of the target volume which will be irradiated with a dose higher than \( D_0 \). In addition to the information on the target volume, the maximal dose in the area of risk structures has to be defined as well.

21.5.3.6
Visualization of the Dose Distribution

The decision for the best radiation plan is made after evaluation of the dose distribution based on the isodose curves (Figs. 21.5, 21.6), dose volume histograms, conformity index, or mathematical models for the normal tissue complication probability and tumor control probability, similar to the conventional 3D radiation. The definitive decision for the best treatment plan must be made by the physician, using clinical judgment, after the rigorous evaluation of the dose distribution in the complete 3D data base.

21.5.4
Positioning

The positioning of the patient on the LINAC is done by using a stereotactic positioner (Fig. 21.4). This instrument allows to project the coordinates of the target point onto orthogonal planes attached to the stereotactic frame. By the use of this projected target point, the patient can be positioned in a way that the target point and the isocenter of the LINAC overlap exactly. The position of the isocenter is indicated by a room-based laser positioning system.

21.5.5
Delivery of the Radiation

After positioning the patient, the target instrument (positioner) is removed and the radiation can start. It is important to know that the mechanical stability and precision of the LINAC has to be much better than in the conventional radiation treatment. The most important requirement for the use of the isocentric LINAC for RS and stereotactic radiation therapy is the accuracy of the isocenter: under ideal conditions the axis of the gantry rotation, the central axis of the beams and the rotation axis of the rotation table
in one point, the isocenter. In practical terms these requirements cannot be achieved. In general, it is acceptable that the three axes – gantry rotation axis, central axis, and table rotation axis – meet in a sphere which coincides with the isocenter and has a diameter of approximately 1 mm. The amount of inaccuracy in the LINAC isocenter is similar to the imprecision of target volume and target point definition with modern radiological techniques. Despite the high weight of the LINAC and of the patient table, these mechanical requirements can be achieved. They must be constantly controlled during regular quality-control procedures.

21.5.6 Quality Assurance

The essential requirement for the clinical use of the LINAC is quality control based on well-defined protocols (REPORT OF TASK GROUP 42, 1995, DIN 6827-1, 2001). These quality requirements are different from those needed in conventional radiotherapy (ICRU REPORT 50, 1984). The quality-assurance protocols address the precision of the target volume and target point with CT, MRI, PET and angiography, the dosimetry, the planning of the irradiation, and especially with the calibration of the absolute dose and of the dose application. For the quality-assurance assessment proper phantoms and specialized dosimetric instruments must be available. Excellent documentation of quality-control requirements have been published by Hartman (1995).

21.6 Future Developments

The stereotactic radiation therapy originated from RS and has been further developed into a fractionated stereotactic radiation therapy. The use of this method at present is still limited predominantly to central nervous system lesions. The use of stereotactic radiation therapy for other regions of the body (paravertebral tumors as well as in tumors of the liver, lung, and pelvis) is under development. The first clinical results are promising (Herfarth et al. 2000; Zimmermann et al. 2004).

The combination of the stereotactic radiation therapy of the LINAC with IMRT opens new perspectives for those entities where exact conformal and high doses must be delivered (Khoo et al. 1999).

The first analysis of RS with dynamic field shaping technique in comparison with conformal static beams and multi-isocentric non-coplanar circular arcs showed that the dynamic-arc technique combines simple planning, short treatment times, dose homogeneity within the target, and rapid dose fall-off in normal tissue (Solberg et al. 2001; Stärk et al. 2004).

A new method under development is robot-assisted RS. The LINAC in this device is mounted on a robotic arm with 6 degrees of freedom. This CyberKnife at this time point can only be used for static radiation, but principally it can be developed for dynamic fields as well (Yu et al. 2004; Gerszten et al. 2004).

In past years progress has been made in the field of frameless stereotactic radiation therapy. This method aims at delivering stereotactic radiation therapy without the stereotactic frame and thus does not use an invasive fixation system, but is still required to deliver the radiation dose with a precision of 1 mm. Frameless stereotactic therapy originated in neurosurgery. For neuronavigation internal and external markers are used for positioning the patient with stereoscopic video cameras and X-ray machines. Despite the high potential of frameless stereotactic treatments, the stereotactic frame will remain the standard. In stereotactic radiation therapy the frame is not only used for target localization but also for the exact positioning of the patient on the LINAC and for patient immobilization during the treatment.

21.7 Conclusion

Both RS and SFS have gained eminent positions in radiation oncology and have become established modalities in the treatment of cranial lesions. Most leading radiation departments offer this technique and their numbers have grown significantly in the past decade.

The LINAC RS and gamma knife RS are equivalent techniques; however, technological and physical differences between these two methods have led to some confusion. Considering the RS, comparative clinical studies have documented that both therapeutic methodologies can be used with similar results. In comparison with gamma knife, the use of LINAC technology offers the possibility of dose fractionation, which has substantial clinical implications. The quality control of the complex LINAC is higher than...
of gamma knife and requires a specialized team of medical physicists and radiation oncologists. On the other hand, it is undisputed that stereotactic radiation therapy with isocentric LINAC has a high potential for further developments. Examples in this direction are the introduction of computer-guided micro-multileaf collimators which allows the delivery of a conformal dose distribution with only one isocenter, using static fields or dynamic arcs and the implementation of the stereotactic intensity-modulated radiotherapy. These new technologies amplify substantially the potential of the stereotactic modality.

References


Brown RA (1979) Stereotactic headframe for use with CT body scanners. Invest Radiol 14:300–304


DIN 6827-1, Protokollierung bei der medizinischen Anwendung ionisierender Strahlung. Teil I: Therapie mit Elektronenbeschleuniger sowie Röntgen- und Gammasbestrahlungseinrichtungen, Deutsches Institut für Industrie-Normung, 2001


Report of Task Group 42 (1995) Radiation Therapy Committee, AAPM Report no. 54, Schell MC (Chairman), Bova FJ, Larson DA et al., published for the American Association of Physicists in Medicine by the American Institute of Physics
22 Extracranial Stereotactic Radiation Therapy

Klaus K. Herfarth

CONTENTS

22.1 History 277
22.2 Systems and Setup 277
22.3 Organ Motion 278
22.4 Clinical Results 279
22.4.1 Lung 279
22.4.2 Liver 282
22.4.3 Spinal and Paraspinal Tumors 285
22.5 Conclusion 286
References 286

22.1 History

The use of a 3D stereotactic system for radiation therapy was first described by Leksell with the development of the gamma knife (Leksell 1951). The head was invasively attached to a stereotactic frame. Numerous cobalt sources could be individually focused to the isocenter allowing a precise application of a focused high dose with a deep dose gradient to the surrounding tissues. With the development of linear accelerators, the stereotactic system was adapted with the serial application of collimated beams to the specified isocenter (Hartmann et al. 1985). Again, only targets in the head could be treated when the head was fixed with a stereotactic ring. Several approaches using non-invasive fixation methods in a stereotactic environment made a fractionated stereotactic radiation treatment in the head possible. New systems of immobilization have had to be developed for stereotactic treatments of extracranial targets. Two major uncertainties have to be taken into account: a reliable fixation of the body is more difficult to achieve than of the head. On the other hand, organ motion has a major impact on the accuracy of the stereotactic system in some of the extracranial targets.

The Karolinska group around Blomgren and Lax published the first extracranial stereotactic system, the stereotactic body frame (SFB), in 1994 (Lax et al. 1994). One year later, Hamilton and Lulu (1995), from the University of Tucson, Arizona, presented a different prototype of extracranial stereotactic system, which was originally developed at the German Cancer Research Center.

Several systems for extracranial stereotactic treatment, the accuracy, advantages, and disadvantages are discussed in the following section. Published clinical data of stereotactic treatments of extracranial targets are summarized later in this chapter.

22.2 Systems and Setup

There are several systems available for stereotactic immobilization of the body stem. For some of these systems, data about the positioning accuracy has been published.

The stereotactic body frame (SFB, Elekta) consists construction of plastic and wood over a length of approximately 1 m. Indicators for calculation of the stereotactic coordinates are implemented in the side walls. The construction of the longitudinal scale consists of seven indicators, on which multiples of 100 mm can be read. Additional 45° indicators allow the exact calculation of the longitudinal coordinate. An additional ruler is placed over the patient connecting the side walls. This ruler is used to determine the correct setup in the transversal plane. The patient is positioned in a vacuum pillow within the stereotactic frame. The published accuracy of the system is 5–7 mm in the transversal plane and below up to 10 mm deviation in the longitudinal direction (Lax et al. 1994). Also, maximal deviations of up to ten in lateral directions were reported (Wulf et al. 2000).
Another system with published accuracy was developed at the German Cancer Research Center and has been available through Leibinger (Freiburg, Germany). This frame was originally developed for invasive fixation of the spine for extracranial radiosurgery of spinal tumors (Hamilton and Lulu 1995; Hamilton et al. 1995; Hamilton et al. 1996); however, the system is very flexible and allows also the patient positioning in a vacuum pillow or a whole-body cast (Herfarth et al. 2000b, c; Lohr et al. 1999). A metal arch is rigidly mounted on a carbon-fiber-compound base which covers the full patient. Three V-shaped indicators (lateral and anterior) are attached to the arch, and the posterior indicator is mounted on the base plate. Metal wires in the indicators are visible on the CT scan and are used for stereotactic target-point calculations. The indicators can be moved in defined steps along the longitudinal axis of the base plate. The accuracy of the system is comparable to the Elekta frame if the vacuum pillow is used. The mean achievable repositioning error is below 5 mm in the transversal plane and maximal 10 mm in cranio-caudal direction; however, a verification CT scan was performed before therapy and the positioning was optimized, if necessary. A correction of the positioning was performed in 60% of the cases if the accuracy was more than 5 mm in any direction (Herfarth et al. 2000b).

For fractionated stereotactic radiation therapy of paraspinal or pelvic targets a whole-body cast was developed and adopted to the Heidelberg frame. The preparation of the cast is very time intensive. On the other hand, the achievable accuracy allows a very precise fractionated therapy. The mean deviation of the bony structures was 1.4, 1.6, and 0.9 mm in the transversal plane of the thoracic spine, the lumbar spine, and the pelvis, respectively (Herfarth et al. 2000c; Lohr et al. 1999). This system is used in Heidelberg for stereotactic fractionated treatment of paraspinal tumors, prostate cancer, and for pelvic chordomas or chordrosarcomas (Herfarth et al. 2000c; Lohr et al. 1999; Milker-Zabel et al. 2003; Thilmann et al. 2002).

A third system (BodyFIX, Medical Intelligence) has been evaluated by Fuss et al. (2004). It consists of a base plate with variable sizes of a vacuum cushions and a clear plastic foil covering the patient's body. The cushion is modeled using an additional vacuum between the patient's front and a plastic foil. An arch-like attachment can be affixed to the base plate providing CT-, MR- and PET-visible fiducials. The reported accuracy in 109 control CT studies were below 3 mm in all three dimensions. Deviations of larger than 5 mm were found in 29% which resulted in a significant loss of target coverage in these cases. Another significant influence on target coverage had the patient's body mass index (BMI). Obese patients had a significantly higher probability of larger setup deviations with following loss of target-volume coverage (Fuss et al. 2004).

Summarizing the published setup data, a high accuracy in patients repositioning can be achieved; however, control CT scans are strongly recommended before high-dose radiation therapy to ensure correct patient position. This is mandatory with respect to the small safety margins since a certain percentage of patients show deviations of more than 5 mm during repositioning which makes corrections of the target coordinates or the positioning necessary.

A frameless stereotactic radiation therapy system is the Cyberknife Image-Guided Radiosurgery System (Accuray, Sunnyvale, Calif.). The system consists of a lightweight 6-mV linear accelerator mounted on a robotic arm. Targeting is achieved by real-time imaging and patient movement tracking. The system works for cranial and extracranial targets. Instead of a fixed frame, it uses internal radiographic markers. The accuracy of the system had been measured using a head phantom. The accuracy was 1.1±0.3 mm for a CT slice thickness of 1.25 mm (Chang et al. 2003). Several isocenters have to be targeted for larger or complex-shaped lesions since the system works only with round collimators. This results in treatment times up to 6 h under those circumstances (Koong et al. 2004).

### 22.3 Organ Motion

Organ motion plays a major role in the stereotactic radiation treatment of extracranial targets. There are several approaches to deal with organ motion which can be separated in two major groups: (a) minimizing the tumor motion and treatment, calculation of the volume of movement and treatment of this volume; and (b) target gating or target chasing. Both approaches have advantages and disadvantages. While the gating is more precise in the dose than the first approach, treatment of the minimized movement area allows much shorter treatment times. Some of the approaches are discussed in more detail.

#### Abdominal Compression

The easiest way to minimize organ breathing motion in the upper abdomen and in the thoracic space
is by using abdominal compression. The movement in cranio-caudal direction can be limited to a mean of 7 mm (Herfarth et al. 2000b; Lax et al. 1994). The volume of tumor movement can be calculated using multislice CT (Hof et al. 2003a). Hof et al. calculated the movement for 15 tumors in the lung. The mean movement in cranio-caudal direction was 5.1±2.4 mm. The mean movement laterally or in anterior-posterior direction was 2.6±1.4 and 3.1±1.5 mm, respectively (Hof et al. 2003a).

**Breath-Hold Technique**

Several papers have discussed the advantages of a breath-hold technique in radiation therapy of lung tumors (Hanley et al. 1999a; Mageras and Yorke 2004; Onishi et al. 2003; Rosenzweig et al. 2000). An advantage of the deep-inspiration breath-hold technique (DIBH) is the reduced lung density in the radiated area which results in a significantly reduced mean lung dose and the volume treated in the high-dose range (Hanley et al. 1999b). The patients have to be trained to get the best reproducibility. Hanley et al. (1999b) described an inter-breath-hold reproducibility of 2.5±1.6 mm determined from diaphragm position. Onishi et al. (2003) described their best results with a reproducibility of 2.2, 1.4, and 1.3 mm in cranio-caudal, anterior-posterior, and lateral axis, respectively. The length of each breath-hold is estimated to be 12–16 s; therefore, the delivery of high doses during a fraction remains time-consuming. Even after training sessions, not every patient is suitable for this procedure. Patients tolerated 10–13 breath-holds per session during training exercises (Hanley et al. 1999a).

**Active Breathing Control**

Active breathing control (ABC) allows to monitor the breathing cycle and perform a gated radiation therapy. An ABC setup to temporarily immobilize the patient’s breathing has been described by Wong et al. (1999). The patient’s inspiration and expiration paths are monitored. Again, a patient training is necessary to obtain the best reproducibility. The inter-fraction mobility of the diaphragm has been reported to be 4±3.3 mm on average (Wong et al. 1999). Dawson et al. (2001) look for the reproducibility of an ABC technique in the treatment of liver tumors. While the intra-fraction reproducibility of a liver marker was 2.5 mm (compared with skeleton markers), the average interfraction offset showed much higher values (5.2 mm). This indicated a change of diaphragm and liver microcoil position relative to the skeleton over the course of treatment with repeat breath-holds at the same phase of the respiratory cycle. Other gating techniques use the motion of the body surface to gate through the breathing cycle (Minohara et al. 2000).

As mentioned above, the described methods result in longer treatment times than radiation therapy without gating. Treatment times for a single-dose radiation therapy of lung tumors were 1–2 h when respiratory gating was used, whereas it lasted only 30 min without gating (Hara et al. 2002). The desirability of the gating procedures was investigated by Starkschall et al. (2004): patients with larger tumor volumes did not benefit from gating. Gating seemed to be advantageous for patients whose tumors are <100 ml and for whom the center of the tumor exhibited significant motion.

**Jet Ventilation**

The respiratory movement of the lung can be actively controlled by artificial respiration under anesthesia. Jet ventilation is a technique that minimizes the movement of the lung by a pulsating gas flow (Warner et al. 1988). Studies in dogs showed a remaining target motion of less than 3 mm (Yin et al. 2001). Our own unpublished experience with jet ventilation shows an undetectable organ motion under a ventilation frequency of 200 Hz. The remaining target motion is completely visualized in the planning CT images; however, this method also has major disadvantages. It is a personal and time-consuming procedure since the patient needs a deep general anesthesia. It has also been shown that the position of a lung tumor might differ between two jet-ventilation sessions; therefore, before stereotactic treatment, the correct tumor position has to be determined again and coordinates might have to be adapted.

**22.4 Clinical Results**

**22.4.1 Lung**

Surgery of lung metastases can result in a significant rate of long-time survivors in selected patients. Five-year survival rates of 27–43% are published (Dienemann et al. 2004; International Registry of Lung Metastases 1997). Surgical resection of
earlyphase lung cancer results in a 50% 5-year survival of the treated patients. Survival rates for inoperable early-stage lung cancer patients are much worse after conventional radiation therapy with a median survival of 20.5 months (Zierhut et al. 2001). If higher radiation doses can be applied, the long-term results improve (Dosoretz et al. 1993). Stereotactic irradiation of lung tumors, metastases, and primary lung cancers offers the opportunity of dose escalation and, therefore, long-lasting local tumor control. The adjuvant radiation therapy of the mediastinum is not always necessary since recent publications have not shown an increase of mediastinal recurrences after treatment of radiologically affected areas only (Hayakawa et al. 1999). An example of a dose distribution for a stereotactic treatment of a stage-I non-small cell lung cancers (NSCLC) is shown in Fig. 22.1.

Stereotactic irradiation of lung tumors have been performed using a hypofractionated approach (radioablation) and a radiosurgical approach (single-dose irradiation). All published results are summarized in Table 22.1.

**Hypofractionation**

Blomgren et al. (1998) reported on 17 intrathoracic tumors (3 primary, 14 metastases) in 1998. The total dose ranged between 15 and 45 Gy on the surrounding 65% isodose given in one to three fractions. With a median follow-up of 8.2 months, there was only one recurrence of a colorectal metastasis.

Uematsu et al. (1998) treated 66 tumors (23 primary and 43 secondary) with total doses of 30–75 Gy (80% isodose) given in 5–15 fractions. Only 3% recurrences were observed with a median follow-up of 11 months. They reported only of minimal interstitial changes.

Nagata et al. (2002) treated 31 stage-I NSCLC and 9 metastases with 4 times 10–12 Gy to the isocenter. They observed no greater toxicity and a local control rate of 90% with a median follow-up of 19 months.

The University of Indiana started a dose escalation trial for peripheral primary NSCLC stage I. They started with 3×8 Gy to the surrounding 80% isodose. Three to five patients were treated in every cohort followed by a successive dose escalation of 2 Gy per fraction. The maximal tolerable dose was not reached after 3×20 Gy. Six of 37 patients experienced a local failure with a median follow-up of 15.2 months, all of whom had received doses of <18 Gy per fraction (Timmerman et al. 2003).

Stereotactic radiation therapy of lung tumors started with 3×10 Gy to the surrounding 65% isodose at the University of Würzburg (Wulf et al. 2001). The dose was escalated to 3×12.5 Gy or a radiosurgical approach of 1×26 Gy (surrounding 80% isodose) after five local recurrences had been observed. The toxicity was mild with only 3% pneumonitis requiring medication. Eighty-one percent of the patients were without any clinical side effect. The actuarial local control rate was 95% in stage-I NSCLC (n=20) and 80% for metastases (n=51) at 12 months with a median follow-up of 9.5 months. While five local recurrences were seen after 3×10 Gy, only one local failure was observed after the escalated dose (Wulf et al. 2004).

![Fig. 22.1 Dose distribution of a radiosurgical treatment of a stage-I non-small cell lung cancer (left: transversal plane; right: frontal plane); 30 Gy at the isocenter were applied using this plan. Close safety margins and a steep dose gradient are visible. The 80% isodose and the 50% isodose are labeled. CTV clinical target volume; PTV planning target volume](image-url)
Extracranial Stereotactic Radiation Therapy

The term “radiosurgery” implies a focused single-dose radiation therapy. Most of the stereotactic treatments in the brain were successfully performed using a radiosurgical approach (Chen et al. 1999; Pirzkall et al. 1998). Nakagawa et al. (2000) treated 24 lung tumors (23 metastases and 1 primary lung cancer) with single-dose therapy (18–25 Gy peripheral dose; some combined with fractionated conformal RT). There was only one local recurrence during a median follow-up of 8 months. Twelve tumors showed a complete response.

Even higher doses were applied by Hara et al. (2002): 23 pulmonary tumors (5 primary, 18 metastases) were treated with peripheral doses of 20–30 Gy. With a median follow-up of 13 months, 4 local recurrences were observed: 3 local failures when a dose of less than 30 Gy was given and 1 with 30 Gy.

As mentioned above, Wulf et al. (2004) also treated pulmonary targets using single-dose therapy with 26 Gy to the surrounding 80% isodose if the clinical target volume was below 25 ml. None of the 26 targets treated with this dose showed a local failure.

The radiosurgical treatment of lung tumors was started in 1997 at the University of Heidelberg. The applied dose was escalated from 1×14 Gy in the beginning to 1×28 Gy applied to the isocenter with the 80% isodose surrounding the PTV. The actuarial local tumor control rate of 48 metastases and 21 primary NSCLC was 77% after 12 months with a median follow-up of 10.7 months. The local tumor control rate was dependent on the applied dose: the actuarial local control rate was 90% after 12 months for all tumors treated with more than 24 Gy. There was no grade-III toxicity (Hof et al. 2004). An example of the follow-up examination of a colorectal cancer metastasis is shown in Fig. 22.2.

A large Japanese series that treated early-stage lung cancer was published in 2004 (Onishi et al. 2004a). Two hundred forty-five patients were stereotactically treated in different centers with varying fraction and normalization methods. The local recurrence rate was 14.5% in this retrospective series. Local control was

### Table 22.1 Published papers of stereotactic radiation therapy of lung tumors. OS overall survival, PD progressive disease, SD stable disease, PR partial remission, CR complete remission, LC local tumor control, BED biological effective dose

<table>
<thead>
<tr>
<th>Reference</th>
<th>Number</th>
<th>Median follow-up (months)</th>
<th>No. of fractions</th>
<th>Dose (normalization; Gy)</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blomgren et al. (1998)</td>
<td>17</td>
<td>8.2</td>
<td>1–3</td>
<td>15–45 (65%)</td>
<td>Median OS 11.3 months&lt;br&gt;1 × PD&lt;br&gt;3 × SD&lt;br&gt;7 × PR&lt;br&gt;6 × CR</td>
</tr>
<tr>
<td>Nakagawa et al. (2000)</td>
<td>22</td>
<td>10</td>
<td>1</td>
<td>18–25 (minimum)</td>
<td>Median OS 9.8 months&lt;br&gt;1 × PD&lt;br&gt;2 × NC&lt;br&gt;7 × PR&lt;br&gt;12 × CR</td>
</tr>
<tr>
<td>Nagata et al. (2002)</td>
<td>43</td>
<td>19</td>
<td>4</td>
<td>40–48 (isocenter)</td>
<td>76% PR&lt;br&gt;18% CR</td>
</tr>
<tr>
<td>Hara et al. (2002)</td>
<td>23</td>
<td>13</td>
<td>1</td>
<td>20–30 (isocenter)</td>
<td>13 months LC:&lt;br&gt;63% (&lt;30 Gy)&lt;br&gt;88% (&gt;30 Gy)</td>
</tr>
<tr>
<td>Hof et al. (2003b)</td>
<td>10</td>
<td>14.9</td>
<td>1</td>
<td>19–26 (isocenter)</td>
<td>2 × PD&lt;br&gt;LC 2 years 71%</td>
</tr>
<tr>
<td>Onishi et al. (2004b)</td>
<td>35</td>
<td>13</td>
<td>10</td>
<td>60 (minimum)</td>
<td>2-year OS 58%&lt;br&gt;2 × PD&lt;br&gt;25 × PR&lt;br&gt;8 × CR</td>
</tr>
<tr>
<td>Wulf et al. (2004)</td>
<td>61</td>
<td>11</td>
<td>1</td>
<td>26 (80%)</td>
<td>1 year actuarial LC:&lt;br&gt;92% (primary tumor)</td>
</tr>
<tr>
<td>Onishi et al. (2004a)</td>
<td>234</td>
<td>24</td>
<td>1–122</td>
<td>18–78 (isocenter)</td>
<td>Local progression 14.5%&lt;br&gt;BED &gt;100 Gy superior&lt;br&gt;BED &lt;100 Gy</td>
</tr>
</tbody>
</table>

**Radiosurgery**

The term “radiosurgery” implies a focused single-dose radiation therapy. Most of the stereotactic treatments in the brain were successfully performed using a radiosurgical approach (Chen et al. 1999; Pirzkall et al. 1998). Nakagawa et al. (2000) treated 24 lung tumors (23 metastases and 1 primary lung cancer) with single-dose therapy (18–25 Gy peripheral dose; some combined with fractionated conformal RT). There was only one local recurrence during a median follow-up of 8 months. Twelve tumors showed a complete response.

Even higher doses were applied by Hara et al. (2002): 23 pulmonary tumors (5 primary, 18 metastases) were treated with peripheral doses of 20–30 Gy. With a median follow-up of 13 months, 4 local recurrences were observed: 3 local failures when a dose of less than 30 Gy was given and 1 with 30 Gy.

As mentioned above, Wulf et al. (2004) also treated pulmonary targets using single-dose therapy with 26 Gy to the surrounding 80% isodose if the clinical target volume was below 25 ml. None of the 26 targets treated with this dose showed a local failure.

The radiosurgical treatment of lung tumors was started in 1997 at the University of Heidelberg. The applied dose was escalated from 1×14 Gy in the beginning to 1×28 Gy applied to the isocenter with the 80% isodose surrounding the PTV. The actuarial local tumor control rate of 48 metastases and 21 primary NSCLC was 77% after 12 months with a median follow-up of 10.7 months. The local tumor control rate was dependent on the applied dose: the actuarial local control rate was 90% after 12 months for all tumors treated with more than 24 Gy. There was no grade-III toxicity (Hof et al. 2004). An example of the follow-up examination of a colorectal cancer metastasis is shown in Fig. 22.2.

A large Japanese series that treated early-stage lung cancer was published in 2004 (Onishi et al. 2004a). Two hundred forty-five patients were stereotactically treated in different centers with varying fraction and normalization methods. The local recurrence rate was 14.5% in this retrospective series. Local control was
significantly better when a biological effective dose (BED) of more than 100 Gy was applied. Three-year survival was 88.4% in medically operable patients if treated with a BED >100 Gy (Onishi et al. 2004a).

Focal Lung Reaction

Radiation pneumonitis is a known phenomenon after conventional fractionated radiotherapy. The term “pneumonopathy” might be better since there is no infective process in place. The rate of clinical symptomatic pneumonopathy is dependent on the radiated volume and the dose. It usually occurs 1–3 months after completion of thoracic irradiation. Recovery from so-called pneumonitis usually occurs and a fibrosis follows and progresses in time. Symptoms may involve low-grade fever, congestion, cough, fullness in the chest, and also dyspnea and chest pain. Radiological changes involve a homogenous increase in radiodensity and patchy, discrete, or solid consolidation with varying time tables, appearing weeks to years after radiotherapy corresponding to the acute pneumonitis and the chronic fibrosis phases.

The same pattern of focal lung reaction have also been seen after stereotactic radiation therapy of lung tumors. Takeda et al. (2004) described ground-glass opacities in 18% of hypofractionated treated tumors and dense consolidation in 73%. We have found changes in 80% of all tumors treated with 24 Gy single dose or more (Herfarth et al. 2000a). On the other hand, the rate of clinically symptomatic radiation pneumonopathy is low due to the small volumes irradiated. The shape of the resulting consolidating fibrosis might change over years and should not be falsely diagnosed as local recurrence (Wulf et al. 2001). Many questions remain unanswered: Which factors influence the pattern and the occurrence of the pneumonopathy? Is there a threshold dose? Can we predict radiation pneumonitis using the mean lung dose? There are several trials underway which should give answers to these questions.

22.4.2 Liver

Hypofractionation

Blomgren et al. were the first authors to publish data about stereotactic irradiation of liver tumors. Their initial report in 1995 was followed by an update in 1998 (Blomgren et al. 1995, 1998). After having had negative experiences with single-dose therapy, which is discussed later, they mainly used hypofractionated radiotherapy. The fractionation and the overall time of treatment were very variable. The dose ranged from 2×8 to 3×15 Gy or 4×10 Gy. The dose was prescribed to the PTV compassing 65% isodose, which resulted in maximal total doses of 20–82 Gy. The treatment time varied between 3 and 44 days (Blomgren et al. 1998).

Blomgren et al. (1995, 1998) treated 20 primary intrahepatic cancers in 11 patients. The clinical target volume was median 22 cc with a range of 3–622 cc. With a mean follow-up of 12 months, they described no local failure; however, two cases of radiation-induced liver disease (RILD) with fatal ending were reported. Both patients had liver cirrhosis. The first patient presented with a 57-cc hepatocellular carcinoma (HCC) nodule due to hepatitis C and liver cirrhosis. The tumor was treated with 3×15 Gy to the periphery of the PTV. The patient developed as-

Fig. 22.2 Lung metastasis of a colorectal cancer before radiosurgical treatment (left) and 7 months after treatment (right)
cites 20 days after completion of the treatment and died the next month. The other patient had a 293-cc large HCC treated with $3 \times 10$ Gy to the periphery of the PTV. Also this patient developed non-tractable ascites in the first 6 weeks after treatment and died shortly after that. Unfortunately, there is no detailed information about the size of the liver, the degree of pretherapeutic liver impairment, or the mean liver dose; therefore, no definite conclusions about the risk assessment can be drawn from this published data. Apart from these fatal side effects, patients experienced nausea, fever, or chills for a few hours after radiosurgery.

Ten patients with 20 metastases were also treated at the Karolinska Institute using the hypofractionated stereotactic regimen. The median CTV was 24 cc with a range of 2–263 cc. Tumor response was evaluated after a mean follow-up time of 9.6 months. All tumors showed response to the therapy. One local recurrence was observed 6 months after therapy. Again, patients experienced nausea, fever, and chill a few hours after the procedure. These symptoms were handled with a prophylactic treatment with paracetamol and antiemetics later. One patient suffered from a hemorrhagic gastritis a few weeks after treatment. One third of the stomach wall had been exposed to 7 Gy for two treatment sessions. Parts of the duodenum were exposed to $4 \times 5$ Gy in another patient. This patient developed a duodenal ulcer, which was treated conservatively.

These early Stockholm data indicated the feasibility and the possible success rate of a hypofractionated stereotactic treatment of liver tumors. Unfortunately, no dose-volume-success constraints can be drawn from these data due to the wide range of the applied dose and different fractionation schemes. The Stockholm group has continued to treat patients with hepatic cancer with the stereotactic approach; however, new data have not been published.

Wulf et al. (2001), from the University of Würzburg (Germany) adopted parts of the Stockholm treatment approach. They treated 24 patients with liver tumors (one clear cell carcinoma and 23 metastases). The median clinical target volume was 50 cc with a minimum of 9 cc and a maximum of 512 cc. All but one patient were treated with $3 \times 10$ Gy to the 65% isodose at the periphery of the PTV. One patient was treated by $4 \times 7$ Gy also normalized to the periphery of the PTV. The reason for this other fractionation was a close proximity of the target to the esophagus. The crude local control was 83% with a mean follow-up of 9 months. The actuarial local control after 12 months was reported with 76%. The recurrences occurred 3, 8, 9, and 17 months after treatment. All were treated with $3 \times 10$ Gy. The failures of three of these targets occurred marginally. The median survival of these patients was calculated to 20 months. The morbidity of the treatment was low: only 7 of 24 patients reported side effects of grades 1 or 2 according to the WHO classification. The side effects were mostly related only to one fraction and included fever, chills, or pain with a typical onset a few hours after irradiation. Additionally, nausea and/or vomiting might occur at the same time. The symptoms ceased spontaneously or could successfully be treated with paracetamol or prednisolone. Only one patient showed longer-lasting fatigue, weakness, and loss of appetite.

**Radiosurgery**

Blomgren et al. (1998) also started with a single dose therapy for liver tumors. Six tumors in 5 patients were radiosurgically treated. The prescribed dose to the periphery of the PTV was median 15.5 Gy ranging from 7.7 to 30 Gy. No recurrences were observed during a median follow-up of 5 months; however, one patient died 2 days after treatment. These patients had a 229-cc large HCC in a cirrhotic liver. The tumor was treated with 30 Gy applied to the periphery of the PTV. The isocenter dose was 48 Gy. The patient already was icteric and showed signs of ascites at the time of treatment. The other four patients showed marginal recurrences during follow-up as is mentioned in a later paper by Blomgren et al. (1998). These two circumstances forced Blomgren et al. to abandon the radiosurgical approach for large liver tumors.

In 1997, a phase-I/II trial was initiated at the German Cancer Research Center in Heidelberg (Germany) proving the feasibility and the clinical outcome of a single-dose radiation therapy of liver tumors (Herfarth et al. 2001a). Inclusion criteria for the study were non-resectable tumors in the liver. The number of liver lesions was not to exceed three tumors (four, if two tumors <3 cm were close together). The size of a single lesion was not to exceed 6 cm, and none of the tumors could be immediately adjacent to parts of the gastrointestinal tract (distance >6 mm). Exclusion criteria were an insufficient liver function. Thirty-seven patients were included. A total of 60 tumors were radiosurgically treated on 40 occasions. The targets included 4 primary hepatic tumors and 56 metastases (mainly colorectal cancer or breast cancer). The median target size was 10 cc (range 1–132 cc). The dose was prescribed to the isocenter with the 80% isodose encompassing the PTV.
The dose was escalated from 14 to 26 Gy based on the liver dose in the dose-volume histogram. After initial dose escalation, an actuarial local tumor control of 81% at 18 months could be achieved with a mean follow-up of 9.5 months. All patients received a prophylactic dexamethasone medication before and after radiation therapy. The actuarial 2-year survival was 59%. Patients with curative treatment intention showed a significant longer survival (actuarial 87% at 2 years) than patients with additional extrahepatic tumor manifestations at the time of treatment (median survival 12 months; Herfarth et al. 2001b). An update of these study patients with a mean follow-up of 17 months was published in 2003 (Herfarth et al. 2003b). Two patients developed late local recurrences 4 years after therapy. The actuarial local control remained unchanged with 81% after 18 months. A follow-up series of a pancreatic cancer metastasis is shown in Fig. 22.3.

As described later in this chapter, a follow-up trial was initiated after these promising initial results. More patients had been radiosurgically treated according to the initial phase-II protocol until recruitment of the follow-up trial could be started. A combined total of 78 patients were treated until spring 2003. The mean follow-up has been 12 months and the actuarial local tumor control dropped to 72% at 12 months. Analysis of the increased failure rate revealed that patients with metastases of a colorectal cancer showed a significantly worse local tumor control than patients with other histologies (Herfarth and Debus 2003). Especially, all 11 patients who already had received chemotherapy using CPT-11 or oxaliplatin had shown local recurrences during the first 15 months after therapy. These recurrences were in-field and marginal recurrences; therefore, higher doses and/or larger safety margins should be used, especially if colorectal cancer metastases are treated.

Side effects of the treatment were minimal (Herfarth et al. 2001a). It included mild nausea or loss of appetite for 1–2 weeks in about one-third of the patients. A singultus was observed in 2 patients and 1 patient developed fever. There were no clinical signs of radiation-induced liver disease. On the other hand, a focal liver reaction occurred after radiation which is described more in detail later in this chapter.

The hypofractionated and the single-dose approach in the stereotactic radiation treatment of liver metastases is currently evaluated in a phase-III study. The StRaL trial (Stereotactic Radiation Therapy of Liver Metastases) is a prospective randomized multicenter trial that started patient recruitment in March 2003 with a planned enrollment of 276 patients over 5 years. Inclusion criteria are a maximum of three liver metastases which are surgically inoperable. The maximal size of the tumors is dependent on the number of targets: It is 5 cm for one target, 4 cm for two targets, and 3 cm for three targets. Primary study goal is the comparison of the local tumor control. Secondary goals are survival, morbidity, and quality of life. The study is designed to prove the equivalence of both treatment arms. Patients in arm A receive a single-dose radiation therapy of 28 Gy normalized to

![Fig. 22.3](image-url) Liver metastasis of a pancreatic cancer at A the time of treatment, B 6 weeks after , and C 5 months after radiosurgical treatment
the isocenter with the 80% isodose (22.4 Gy) encompassing the PTV. Patients in arm B receive a hypofractionated therapy with $3 \times 12.5$ Gy normalized to the 65% isodose (encompassing the PTV).

**Focal Liver Reaction**

The focal liver reaction after single-dose radiation therapy has been examined by the Heidelberg group. All patients who were followed using multiphasic CT scanning showed a sharply demarcated focal radiation reaction. Tumor and radiation reaction could be well differentiated in the portal-venous contrast-enhanced CT scans. Liver vessels run through the liver reaction and were not displaced, as is seen in cases of expanding tumor. A detailed evaluation and characterization of this focal radiation reaction in 36 of the Heidelberg patients was published in 2003 (Herfarth et al. 2003a).

The area of radiation reaction was hypodense in the majority of the non-enhanced CT scans. Three different types of appearance of the reaction could be defined based on the liver density in the portal-venous and the late phase after contrast agent administration:

1. Type-1 reaction: Hypodensity in portal-venous contrast phase, isodensity in the late contrast phase
2. Type-2 reaction: Hypodensity in portal-venous contrast phase, hyperdensity in the late contrast phase
3. Type-3 reaction: Isodensity/hyperdensity in portal-venous contrast phase, hyperdensity in the late contrast phase

The onset of the reaction was after median 1.8 months. While the type-1 or type-2 reactions usually showed up earlier, type-3 reactions appeared later than the other types. It was also seen that there was a shift of the appearance during follow-up towards type-3 appearances. In addition, the volume of the radiation reaction decreased with follow-up time. The most dramatic shrinkage was observed during the first months after appearance. This led to the speculation that the whole reaction goes through different radiological stages (type 1, 2, and 3 appearances). The histological basis of these stages was not determined since no biopsies were taken; however, others had reported a type-2 appearance after single-dose radiation therapy and it was histologically confirmed veno-occlusive disease (Willemart et al. 2000).

Based on reconstruction of the dose-volume histograms, the mean threshold dose was 13.7 Gy with a wide range between 8.9 and 19.2 Gy given in a single fraction. One reason for this large variance might be the fact that the volume strongly decreased between the initial detection and the further follow-up examinations. The examination might have not detected larger reaction volumes and, therefore, the calculated threshold doses might have been overestimated. This was sustained by the significant correlation between the threshold dose and the time of detection (correlation coefficient $r=0.709$). Apart of the time factor, other factors, which might influence the individual radiation sensitivity (e.g., additional toxic liver agents such as alcohol) might have been another reason for the variance. More data are needed to strengthen these threshold doses.

**22.4.3 Spinal and Paraspinal Tumors**

In opposite to the lung and the liver with a pronounced volume effect, the myelon is characterized by a serial function structure: the side effects are more influenced by the dose maximum than by the volume treated. The goal for a stereotactic treatment of spinal and paraspinal tumors is to reduce the total dose to the myelon. Figure 22.4 shows a stereotactic treatment plan which allows a dose escalation of paraspinal regions with a dose limitation to the spinal cord.
chord. Again, there have been different stereotactic approaches to achieve that goal: single-dose therapy and fractionation. What they have in common is a greater effort in positioning of the patient.

**Fractionated Stereotactic Treatment**

We treated 18 patients with 19 recurrent spinal metastases using a normofractionated stereotactic approach (Milker-Zabel et al. 2003). All patients had previous irradiation of the tumors with a median dose of 38 Gy. The median reirradiation dose was 39 Gy in 1.8- to 2-Gy fractions. The dose to the spinal chord was limited to less than 20 Gy in all patients. The positioning accuracy was provided using a whole-body cast which guarantees an optimal repositioning of the spine using a non-invasive method (Lohr et al. 1999). Of the patients, 95% showed radiological response after a median follow-up of 12.3 months. There were no signs of radiation-induced myelotoxicity (Milker-Zabel et al. 2003).

**Radiosurgery**

Hamilton et al. (1995) used an invasive fixation of the patient. The patient’s spine was rigidly fixed to the stereotactic frame while the patient was under general anesthesia. The first experience included 5 patients who already had normal fractionated radiation therapy to the spinal cord. They again stereotactically delivered 8–10 Gy to the recurrent tumor, which was near the myelon. None of the patients showed a spinal chord injury with a median follow-up of 6 months.

Gerszten et al. (2004) reported on the clinical experience in the radiosurgical treatment of 125 spinal lesions using the CyberKnife. Of the patients, 78% had received radiation treatment of treated region previously. The tumors were treated with a median dose of 14 Gy to 80% isodose line (range 12–20 Gy) and a maximal dose of 8 Gy was allowed to the spinal canal; however, the maximum dose to the edge of the spinal canal was 13 Gy in a single patient. With a median follow-up of 18 months, none of the patients experienced exacerbation of the symptoms or new neurological deficits. Nearly all patients who suffered from pain described an improvement.

Ryu et al. (2003) described 10 patients who received 10×2.5 Gy via standard external beam radiotherapy. The tumors were additionally treated with an image-based radiosurgical boost of 6–8 Gy, prescribed to the 90% isodose line. The maximum dose of radiation to the anterior edge of the spinal cord within a transverse section, on average, was 50% of the prescribed dose. There was no acute radiation toxicity detected clinically; however, the mean follow-up was only 6 months.

**22.5 Conclusion**

Stereotactic radiation treatment of extracranial targets shows promising initial results. The techniques are getting increasingly more specialized, especially in the treatment of small lung tumors; however, many questions remain unanswered. More experimental work and clinical trials are underway which should answer these question and should strengthen this promising approach.

**References**


23 X-IMRT
Simeon Nill, Ralf Hinderer, and Uwe Oelfke

CONTENTS
23.1 Introduction 289
23.2 MLC-based IMRT Delivery 289
  23.2.1 MLC Design 289
  23.2.1.1 IMRT-Relevant MLC Data 289
  23.2.2 Step-and-Shoot Dose Delivery 290
    23.2.2.1 The Step-and-Shoot Technique 290
    23.2.2.2 Leaf-Sequencing Algorithms 290
    23.2.2.3 Hardware Constraints: Matchlines 292
  23.2.3 Dynamic Dose Delivery 292
  23.2.4 Improved Spatial Resolution: Add-on MLCs 294
  23.3 Helical Tomotherapy Delivery 295
  23.4 Compensator-Based IMRT Delivery 296
  References 297

23.1 Introduction

The IMRT dose delivery techniques are the second cornerstone of X-IMRT in addition to the concepts of inverse treatment planning and optimization (discussed in Chap. 17). The result of an inverse treatment plan usually consists of ideal photon-fluence distributions for each selected beam angle. In this chapter we discuss on a very fundamental level the main concepts and strategies for the actual delivery of these fluence patterns to a patient. This very brief summary of some of the interesting techniques cannot be complete and very detailed. The interested reader who aims to explore the touched topics in more depth is referred to the standard literature, e.g., Webb (2001) or (Palta and Mackie 2003).

The main focus in this chapter is on the discussion of the most prominent dose delivery techniques with linac-integrated multi-leaf collimators (MLCs). The basic concepts employed for the two standard methods – the “step-and-shoot” approach and the “dynamic” dose delivery – are briefly reviewed before the new concept of “helical tomography” is introduced. Finally, a short section describes the dose delivery technique with individually designed compensators.

23.2 MLC-based IMRT Delivery

Most modern linear accelerators are equipped with MLCs, which were originally designed to replace the use of lead blocks to shield normal tissue during the treatment (Biggs et al. 1991; Boesecke et al. 1991; Jordan and Williams 1994). In the next section we first describe the main IMRT relevant characteristics of a typical MLC including its dosimetric properties. Then, various dose delivery techniques are briefly described. Based on the MLC computer-control system, two main dose delivery methods are distinguished: (a) the “step-and-shoot” approach; and (b) the “dynamic” dose delivery concept. A very detailed description of both techniques with references to the original papers can be found in the excellent book by Webb (2001). Besides these two basic IMRT delivery concepts with linac-integrated MLCs, we also briefly discuss some examples of “high-resolution” IMRT with add-on mini- or micro-MLCs.

23.2.1 MLC Design

23.2.1.1 IMRT-Relevant MLC Data

Multi-leaf collimators and their design and function for 3D conformal radiotherapy (3D CRT) are described in detail in Chap. 20, together with their geometrical and mechanical characteristics, see also Galvin et al. (19939).
To deliver intensity-modulated fields (IMRT) the following geometric characteristics are of most importance:

1. Leaf width. The leaf width (see Chap. 20.2.1) determines one dimension of the achievable spatial resolution of the fluence modulation. The spatial resolution of the fluence in the direction of the leaves’ movement is only restricted by the positioning accuracy of the MLC. For the delivery of IMRT fields a possible large overtravel range of the different leaves (10–15 cm) is often required to irradiate extended target volumes.

2. Maximum leaf speed. The leaf speed is especially important for the “dynamic” delivery of intensity-modulated fields. A typical leaf speed is of the order of 2–4 cm/s (Webb 2001; Hug 2002).

3. The maximum field size of the MLC. For IMRT applications, the maximum field size is not the same as the maximum open field which can be achieved, but it is mainly defined by the maximum overtravel.

4. Maximum overtravel. Maximum overtravel (see Chap. 20.2.1) is the distance of how far a leaf can move over the midline of the MLC.

5. Leakage and transmission. The leakage and transmission (see Chaps. 20.2.2.2, 20.2.2.3) play an important role specifically for the delivery of IMRT treatments because the dose delivery often requires a substantial number of monitor units (MU); these are applied where most areas of the treatment field are covered with closed leaves, i.e., the leakage radiation received by these areas is significantly enhanced.

6. Leaf-positioning accuracy. One of the most important properties of an MLC with respect to IMRT is the leaf-positioning accuracy. For IMRT delivery the leaf-positioning accuracy is even more important than for most conventional treatment techniques due to the large number of small field components for an average IMRT treatment. For a typical fluence grid with a resolution of 10 mm in both directions a leaf-positioning error of 1 mm has to be considered as a large error. This is due to the characteristic output factor curve with a very steep gradient for small field sizes. An error of 1 mm for the leaf position can cause a dose error of up to approximately 10% irradiating a 10×10-mm field and therefore the maximum leaf-positioning error should not be larger than 0.5 mm. It can also be concluded that the required leaf-positioning accuracy depends on the fluence grid resolution of the fluence maps.

23.2.2
Step-and-Shoot Dose Delivery

As result of the inverse planning process, described in Chap. 17, one receives the “ideal” fluence maps for each incident beam direction. In most treatment planning systems these ideal intensity maps have to be converted into field segments to be delivered with the MLC, a process that is called the “sequencing” of the fluence. In this section we shortly describe the “step-and-shoot” dose delivery method and discuss briefly the two basic sequencing techniques.

23.2.2.1
The Step-and-Shoot Technique

The “step-and-shoot” technique of IMRT dose delivery (Bortfeld et al. 1994) is a straightforward extension of the conventional multiple-field irradiation technique. The “step-and-shoot” approach superimposes the dose delivered by a number of irregularly shaped and partially overlapping treatment fields, often called subfields or segments. For each segment a well-defined number of monitor units is delivered. Then, the beam is turned off while the leaves of the MLC move to the positions required by the next IMRT segment. After the verification and record system (V&R) has validated the new leaf positions, the beam is turned on and the dose is delivered for this segment. This process is repeated for all segments per incident beam angle and all beam directions (see Fig. 23.1).

23.2.2.2
Leaf-Sequencing Algorithms

In order to describe the basic features of leaf-sequencing algorithms we introduce the following terms: a fluence map is defined on a 2D fixed grid covering the respective beam aperture. It is usually divided into discrete, quadratic elements called “bixels.” The fluence map assigns one beam intensity to each bixel. A 1D line of intensities for all bixels created by a single leaf pair is called a channel. The width of each channel coincides with the leaf width of the MLC. The goal of the sequencing process is to decompose the fluence maps into a number of field components or subfields.

The first step of the sequencing process is the stratification of the continuous fluence profiles provided by the inverse planning process, i.e., each line of intensity for a specific channel is forced to take on only a few discrete fluence levels (Fig. 23.2). For each
fluence level the MLC leaves shape and deliver a beam segment. In general, not all levels may be deliverable without an additional step. One problem is the appearance of spatial "holes" for a fluence pattern generated by the optimization process (e.g., two pyramids next to each other with zero fluence in between). For these cases each level has to be divided into separately deliverable subfields. This is one constraint which, in addition to other MLC hardware constraints, must be taken into account during the process of calculating the required leaf positions.

The two most prominent concepts are the "close-in" and "sweep" technique. In Fig. 23.3 we show an example of the "close-in" technique. For this example the fluence map was divided into three levels. The typical numbers of levels for clinical cases is in the range of five to ten. In the first step the left leaf is moved towards the first positive gradient of the fluence pattern while the right leaf is moving towards the first negative fluence gradient. This defines the first segment which is then delivered. Next, the two leaves move to the next positive and negative gradients of the respective fluence map to define the next subfield followed by its delivery. This procedure is repeated until the whole fluence map is delivered.

We briefly describe the basic algorithm of the "sweep" technique for the 1D example of a fluence distribution given in Fig. 23.4. For this fluence map, Fig. 23.2. Stratification of a continuous fluence distribution into a fluence map with a number of discrete levels

Fig. 23.1. The basic idea of the "step-and-shoot" approach is to deliver an intensity-modulated beam as a superposition of a set of irregularly shaped, partially overlapping field components

Fig. 23.3. Decomposition of a one-dimensional fluence map into deliverable segments using the "close-in" technique

Fig. 23.4. Decomposition of a one-dimensional fluence map into deliverable segments using the "sweep" technique