DEVELOPMENT AND APPLICATION OF EFFICIENT PORTAL IMAGING SOLUTIONS
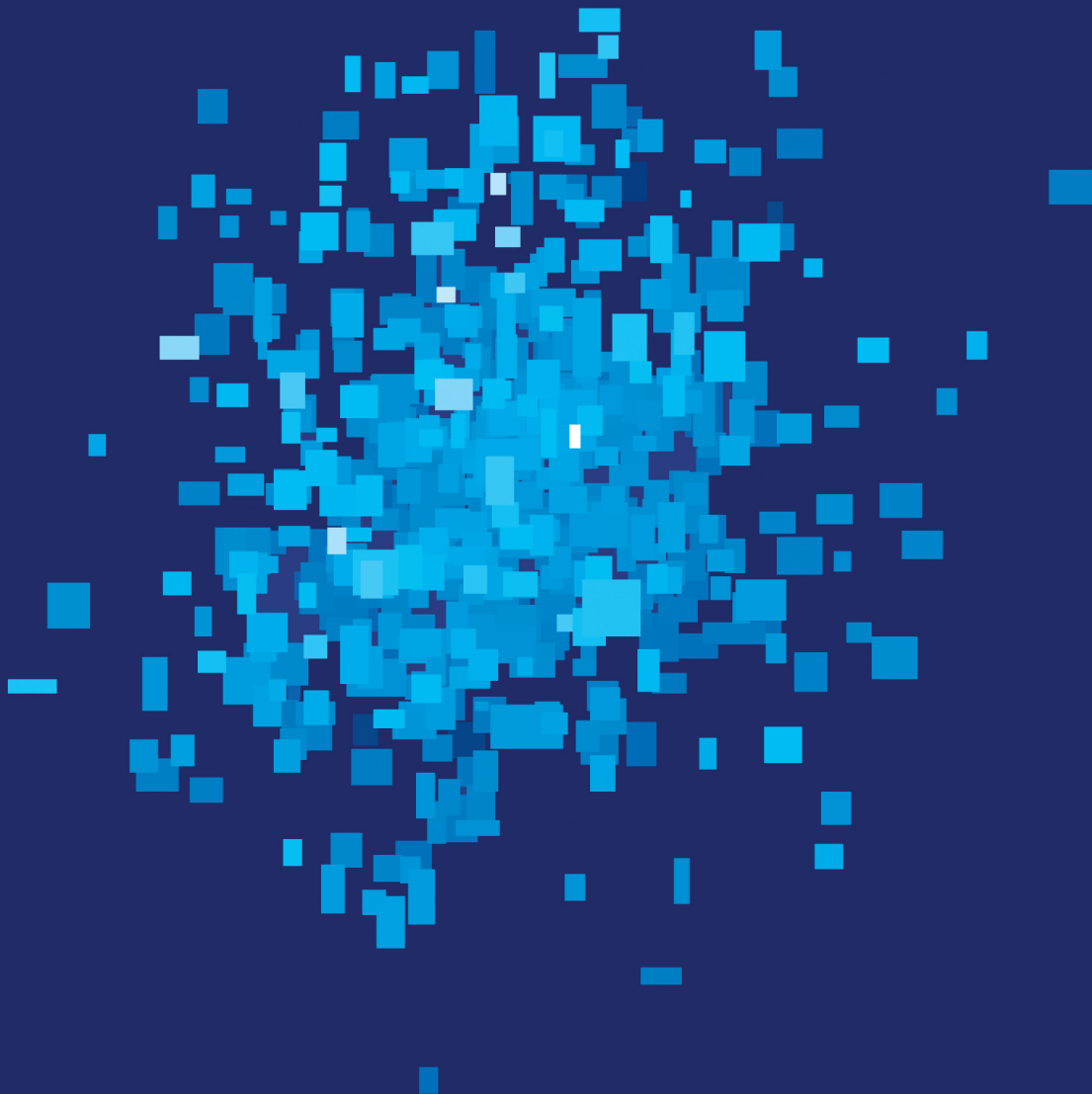
Hans de Boer

Development and Application of Efficient Portal Imaging Solutions

Cover: A representation of the distribution of displacements measured during treatment in the patient group described in chapters 6 and 7, in case of no set-up corrections. For each patient, the systematic displacement in the cranial-caudal direction (plotted horizontally) and anterior-posterior direction (plotted vertically) is indicated by a rectangle. The size of each side of this rectangle is a linear measure of the spread of the random displacements in the corresponding direction (but displayed on a different scale than the systematic displacements). Furthermore, the color of the rectangle reflects the magnitude of the systematic displacement in the left-right direction: darkest blue squares indicate maximum distance from the reader, whereas white squares indicate minimum distance.

Development and Application of Efficient Portal Imaging Solutions

Ontwikkeling en toepassing van efficiënte oplossingen bij het afbeelden van patiënten tijdens radiotherapie

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus

Prof.dr.ir. J.H. van Bemmel

en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
woensdag 5 februari 2003 om 13:45 uur
door

Johannes Cornelis Jacob de Boer

geboren te Leiden

Promotiecommissie

Promotor: Prof.dr. P.C. Levendag

Overige leden: Prof.dr.ir. C.J. Snijders
                     Prof.dr. B. Löwenberg
                     Prof.dr. E.P. Krenning

Copromotor: Dr. B.J.M. Heijmen

Er is niets dat we met volledige zekerheid weten,
maar alles met waarschijnlijkheid,
en er zijn sterke verschillen in de mate van waarschijnlijkheid.

Christiaan Huygens, 1673

Voor Mariëtte,
Niels en Veerle

Aan mijn ouders

Address for correspondence:
J.C.J. de Boer
Erasmus MC – Daniel den Hoed Oncology Center
Subdivision clinical physics
Groene Hilledijk 301
3075 EA Rotterdam
The Netherlands
E-mail: boer@kfih.azr.nl

Contents

1 Introduction

The central subject of this thesis is to derive clinically applicable methods to measure and improve the reproducibility of treatment delivery in radiotherapy by means of portal imaging. The most important criteria that such methods should meet, apart from being effective, is that (1) they are relatively simple to implement and (2) the additional workload required in daily practice is small. This approach was inspired by the observation that routine application of portal imaging in clinical practice, according to well-defined protocols, remains relatively rare.

Below, we first sketch the general aims and practice of radiotherapy. From this brief overview, a number of aspects become apparent that are essential to the work described in the following chapters. In particular, the meaning of systematic and random geometrical errors in radiotherapy is emphasised.

1.1 Radiotherapy

Surgery, radiotherapy and chemotherapy are currently the three predominant therapies in the treatment of cancer. Although a combination of these treatment modalities is often used to provide optimal treatment, here we will concentrate solely on the role of radiotherapy. In radiotherapy, high energy particles ('ionising radiation') are used to deliver damage to tumour cells by interaction of these particles with the cellular matter. Irreparable damage to the nuclear DNA may lead to direct cell kill through loss of vital cell functions or can induce cellular suicide ('apoptosis'). High energy particles may penetrate deeply into tissue. This property is exploited in external beam therapy, where beams of photons, electrons or other particles of MeV energies are generated outside the patient (usually by linear accelerators) and collimated into beams that enter the patient (Fig. 1.1). This approach introduces a number of fundamental problems, which not only generate most of present radiotherapy physics research but also constrain the way radiotherapy treatment can be delivered in practice.

Each clonogenic tumour cell should be killed by the delivered radiation dose to definitely eradicate malignant growth. However, tumours are usually detected only after they have reached macroscopic dimensions (10-100 grams). This implies that lethal radiation doses must be delivered to relatively large volumes of tissue. The external radiation beams, however, will deposit energy in normal tissue both before and beyond the target tumour volume, which may induce substantial cell kill in tissues vital to either life itself or the quality of life. The present approach to simultaneous minimisation of dose in these 'critical organs' and maximisation of dose in the target volume of tumour cells is by calculating optimal radiation beam entrance directions and energy fluence distributions within these beams. Taking such complex dose delivery techniques to the limit may in theory provide distributions of high dose confined almost exactly to the target volume, thereby enabling a high degree of normal tissue sparing at a large tumour dose. However, practical issues in the execution of radiotherapy treatment may introduce a significant difference between the calculated ('planned') distribution of dose and the actual distribution achieved. This thesis deals with a number of inaccuracies introduced by the use of a static geometrical model of the patient anatomy.
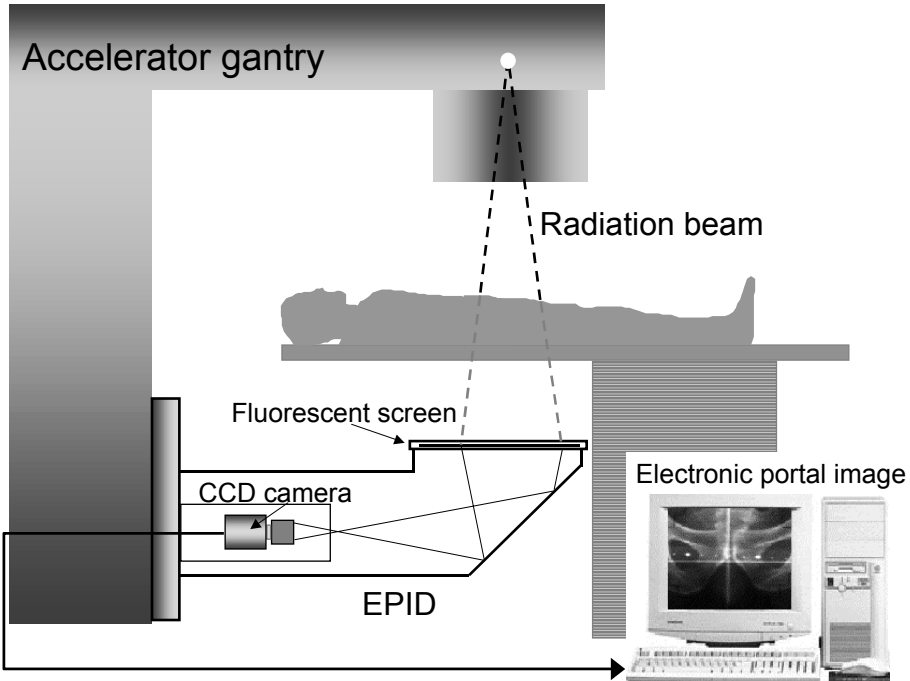
Figure 1.1 External beam radiotherapy configuration with EPID

1.2 Radiotherapy treatment chain

We will limit our discussion to patients treated on the basis of 3-dimensional (3D) anatomic information obtained from a computed tomography (CT) scan acquired with the patient in treatment position, but most of the important aspects are also valid when treatment planning is based on other diagnostic modalities. The volumes of the tumour to be treated and the critical organs are defined within the 3D CT dataset. The tumour volume is then usually extended by a geometrical margin to define the actual volume to be irradiated to a certain prescribed dose level. This margin must account for the geometrical uncertainty involved in the definition of the tumour as well as for the inaccuracies introduced in subsequent steps of the treatment chain. Next, the treatment planning system allows for the calculation and optimisation of the (3D) dose distribution achieved with a specific treatment technique. In so-called 'conformal radiotherapy,' the aim is to conform this planned dose distribution as closely as possible to the patient specific anatomy, using all practically available irradiation techniques. Conventionally, the patient is then positioned on a simulator to mimic the actual treatment on a linear accelerator. Using geometical information from the treatment plan established previously, reference marks are tattooed or drawn on the patient skin or on specific fixation devices (e.g. a head mask) based on the co-ordinate system in the simulator room. This co-ordinate system is usually defined by a set of laser beams generated by wall-mounted lasers in the latter room. Because a similar set of lasers is present in the treatment room, the reference marks enable a quick set-up of the patient at the treatment unit. Once the set-up at the simulator is considered satisfactory, reference radiographs are acquired which prescribe the daily set-up of the bony anatomy of the

patient in the treatment beams. With the advent of CT-simulators, this separate simulator step can be omitted by marking the patient directly after the planning CT scan has been obtained. From the CT data, so-called digitally reconstructed radiographs (DRRs) are calculated which can replace the simulator radiographs as reference images.

The prescribed radiation dose is usually not delivered in one session but instead divided into a number of fractions, with intervals of typically one day. There are various radiobiological reasons for such a fractionated dose delivery. For instance, normal tissues often exhibit more effective DNA damage repair than tumours. Hence, normal tissues may partly recover from such damage in the intervals between fractions, allowing for a larger total tumour dose.

The treatment chain sketched above introduces many possibilities for inaccuracies. Because we will concentrate on geometrical inaccuracies, we give an example of how small errors related to skin marks might emerge at the CT scanner, simulator and during treatment. On the CT scanner, temporary marks are placed on the patient's skin to define the co-ordinates of the patient in the planning CT scan using the laser system of the scanner. If these skin-marks are later used to position the patient on the simulator to mark the definitive patient set-up, inaccuracies occur because

(1) The co-ordinate system of the CT scanner and simulator may not exactly coincide

(2) The skin generally is a flexible structure, allowing for motion relative to the internal anatomy of interest

(3) Judgement of the alignment of skin marks on the lasers is usually by eye. Because multiple skin marks are involved, a 'reasonable fit' must be sought. However, as deviations of the posture of the patient relative to that during the CT scan may not always be uniquely identifiable, such a fit represents a compromise of unpredictable quality.

If, in the above, we replace simulator by treatment unit and (if a simulator is used) CT scanner by simulator, it is obvious that these three causes of inaccuracy are equally applicable at the treatment unit during each of the treatment fractions. Apart from the errors related to the use of skin marks, there exist many other sources of geometrical inaccuracy between treatment preparation and execution. For instance, due to dynamic processes such as breathing, both the treated tumour and normal tissues may exhibit their own specific motion on various time scales. Irrespective of the source of the geometrical deviations between planned and delivered treatment, a separation can be made into 'systematic' and 'random' inaccuracies. Because such a separation turns out to be crucial for treatment planning and optimisation purposes, we describe its application in patient set-up in the next section.

1.3 Systematic and random set-up errors

As in most patient set-up studies, below we will sometimes refer to the geometrical inaccuracies as errors, although strictly speaking this may not always be the most appropriate terminology. To understand the concept of systematic and random errors as applied in this thesis, it is sufficient to consider the simple situation in which the treatment simulation can be skipped, i.e. reference marks suited for patient set-up at the treatment unit are placed at the CT scanner (see section 1.2). Each set-up is executed with a finite accuracy, hence a displacement of the patient relative to the intended set-up used in the treatment planning occurs during each treatment day. In Fig. 1.2(a) these displacements are given for 2 directions for a treatment of 20 fractions. The *mean* displacement over all fractions, represented by the arrow 's' in Figure 1.2(a), is referred to as the *systematic set-up error*. The systematic set-up error represents the recurring difference between intended and

actual set-up. However, patient set-up is slightly different during each fraction and so the total set-up error in a given fraction is the sum of the systematic set-up error and an additional, fraction specific error (represented by the arrow 'r' in Fig. 1.2(a)). The latter error is referred to as *random set-up error*, because it represents the fluctuating, stochastic component of the set-up error during treatment. In Fig. 1.2(b), we present the frequency distribution of random set-up errors in the anterior-posterior (AP) direction, as measured in 600 patients treated for prostate cancer (containing data of over 6000 fractions). As shown in Fig. 1.2(b), this distribution is described well by a normal distribution. Hence, the random set-up errors are often characterised by the standard deviation (SD) of their frequency distribution. This SD is usually denoted by σ (see Fig. 1.2(b)).

The matter gets slightly confusing when we consider more than one patient. Then, the systematic set-up error also becomes a fluctuating quantity, but now the fluctuations are between patients and not between fractions. This is illustrated in Fig. 1.3(a), where systematic set-up errors for 10 patients are drawn as solid arrows. The stochastic nature of the systematic error roots in the fact that many of the processes that contribute to the net observed systematic error, such as the skin marking described in section 1.2, are non-reproducible. Hence, the systematic error observed for a patient would be different if the entire treatment chain was repeated. Consequently, one can also interpret the arrows in Fig.

(a)   (b)   AP random error (mm)
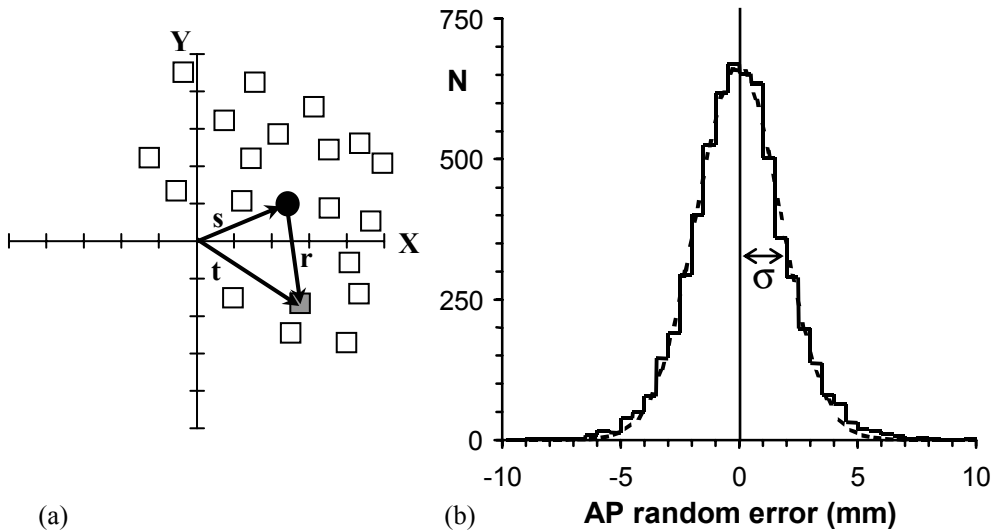
Figure 1.2. (a) Measured set-up errors for 20 fractions of a specific patient, in two directions (X,Y). Each square represents the set-up error of a particular fraction. The solid circle indicates the mean of these errors and the corresponding displacement (arrow 's') is the systematic set-up error of this patient. The total set-up error of any fraction (e.g. the greyed square and corresponding arrow 't') is the sum of the systematic error and the random set-up error (arrow 'r') of that fraction.

(b) Frequency distribution of random set-up errors as measured in 600 prostate ca. patients in the anterior-posterior direction (solid line), and a fitted normal distribution (dashed line). The SD (σ) is indicated, as well as the mean of the distribution (vertical line), which by definition equals 0.

1.3(a) as the systematic errors that could occur if the patient of Fig. 1.3(a) would complete the treatment chain 10 times, neglecting that some systematic errors may be patient specific. Note that certain components that make up the systematic set-up error are completely patient independent. For instance, if the lasers in the treatment room are mis-aligned by a constant amount from those used for placing skin marks, this constant deviation will be present in the systematic set-up error for each patient. The frequency distribution of the systematic errors in the AP direction, obtained for the same patient group as in Fig. 1.2(b), is depicted in Fig. 1.3(b). Again, the distribution is nearly normal, and is usually characterised by the mean systematic error μ and the standard deviation Σ (both indicated in Fig. 1.3(b)).

The description of set-up errors in terms of systematic and random errors turns out to be particularly useful when patient populations are studied. Appropriate pre-treatment quality assurance can render μ small, and so usually the important quantities are Σ and σ. Both SD's measure the spread of set-up errors around the ideal value 0 and margins used in treatment planning to cover for uncertainties in set-up accuracy (section 1.2) can therefore be expressed in these quantities. Note that the illustration of set-up errors given above pertains to translations, which is sufficient if one considers a single point is space. However, in realistic anatomy, rotations can also be of importance. In principle, these rotations can be described similar to translations, i.e. by the values (μ, Σ, σ). In practice, rotations are more problematic as they complicate the 2D analysis of images described below and furthermore the translations measured in various directions become mutually dependent in the presence of rotations. If these effects are neglected, one must make sure that rotations can be kept sufficiently small.

(a) (b)
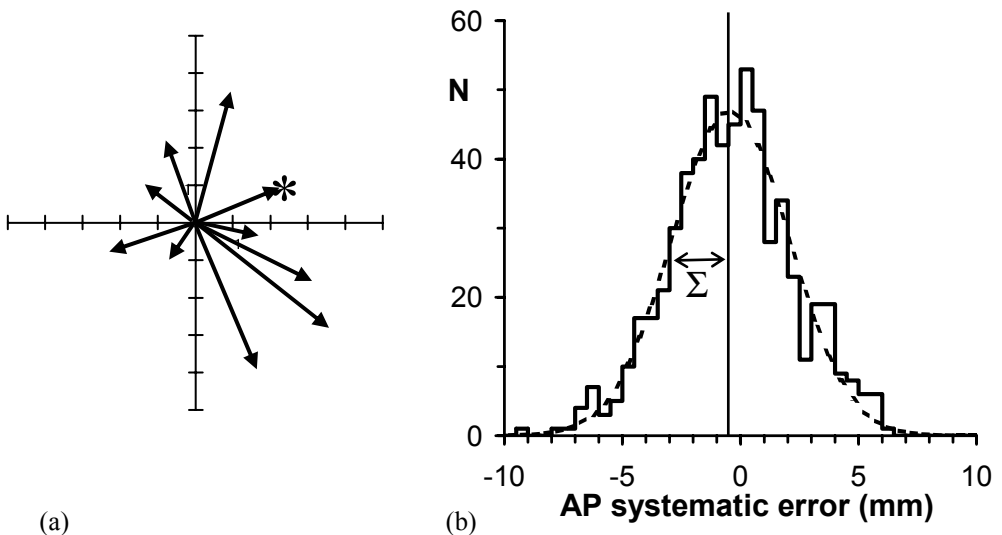
Figure 1.3. (a) Systematic set-up errors for 10 patients, represented by arrows. An asterisk marks the systematic error of the patient in Fig. 1.2 (a). (b) Frequency distribution of systematic set-up errors (solid line), measured in the patient group of Fig. 1.2(b), and a fitted normal distribution (dashed line). The mean of this distribution is indicated by the vertical line, which is displaced by an amount μ from 0, the SD equals Σ.

1.4 Portal imaging and its role in advanced radiotherapy treatment

All research presented below is related to the use of Electronic Portal Imaging Devices (EPIDs) in radiotherapy. An EPID utilises the radiation exiting the patient during treatment to create an image of the patient anatomy crossed by the radiation beam (Fig. 1.1). These *electronic portal images* (EPIs) are similar to X-ray radiographs, but because the photons in the treatment beam are of a much higher energy (typically a factor 100) and the source of the treatment beam is not specifically designed for imaging, portal images suffer from less contrast, signal to noise ratio and spatial resolution. Therefore, creating high quality EPIs remains a technological challenge and EPIDs are under continuing technological development. The EPIDs described in this work were developed in our clinic and are CCD camera based. This type of EPID uses a fluorescent screen similar to those used in combination with X-ray films, to convert the high-energy photons of the treatment beam into visible light. The light is captured by a (non-standard, high sensitivity) CCD camera and the ensuing images are processed and analysed by dedicated software. Based on the excellent performance of our original prototype (Rotterdam Advanced Portal Imaging Device or RAPID), a collaboration with Cablon Medical (Leusden, the Netherlands) was started in 1999 to develop the Theraview New Technology (TNT) system. Both prototype and commercial TNT systems were used in the studies comprising this thesis.

EPIs can be used to measure set-up errors relative to the intended set-up and hence improve the accuracy of that set-up. There are two different approaches to calculate and correct for measured set-up errors, i.e. *off-line* and *on-line*. In the on-line approach, an EPI is obtained using a small part of the fraction dose at the beginning of *each* fraction, the set-up deviation is measured (while the patient remains on the treatment couch) and corrected for immediately, after which the remaining dose is delivered. In the off-line approach, measured deviations are not used to correct the set-up immediately but instead an estimate of the systematic error is calculated from EPIs of various fractions and used for correction in subsequent fractions. The above definition implies that on-line corrections can reduce both random and systematic set-up errors whereas off-line corrections reduce only systematic errors.

Next to set-up corrections, EPIs can be used in transit dosimetry. This application requires that the intensities measured in the EPI can be translated to the radiation dose impingent on the detector with a high accuracy (~1 %). The CCD camera's we apply in our EPIDs are well suited for this task. The resultant *portal dose images* (PDIs) can be used to verify the dose delivered to the patient.

EPIs thus combine information on geometry and dosimetry during the actual treatment and therefore yield an ideal tool for quality assurance and improvement of treatment delivery. Particularly in conformal radiotherapy, where the demands on the set-up accuracy are high and the technique of dose delivery may involve complex dynamic processes, EPIDs have become a prerequisite for treatment verification.

1.5 Subjects treated in this thesis

The main aim of this thesis was to derive practical and effective methods to reduce set-up and dosimetric errors using portal imaging. Below, we outline the contents of the following chapters.

Chapter 2. Introduction to the TNT EPID and dosimetry applications
In this chapter, a technical description of our EPID is given and its suitability for dosimetry measurements is investigated. The advantage of the applied 'high elbow', in which the distance between the fluorescent screen and the mirror that directs the light towards the camera is larger than in conventional EPIDs, is described. A practical algorithm to translate EPIs into accurate PDIs is developed.

Chapter 3. Relation between geometrical uncertainties and treatment planning margins
In section 1.2, we mentioned that in treatment planning, margins are used to account for geometrical uncertainties such as set-up errors. In Chapter 3, the *coverage probability* (CP) model is developed to quantitatively investigate the consequences of geometrical uncertainties on the tumour dose distribution and to derive appropriate planning margins. If the statistical distribution of systematic and random set-up errors is known for a specific patient population, this model can be used to calculate the probability of under-dosage of a tumour treated with a certain planning margin. By constraining this probability, appropriate margins may be derived. These margins may be assessed in a patient specific manner, taking into account the exact dose distribution and patient anatomy. As an alternative, a rule of thumb can be formulated which expresses margins that are approximately correct for an entire patient population.

Chapters 4 and 5 Investigation of the systematic and random set-up errors in lung cancer patients
If the skin marks that are used for patient positioning at the treatment unit are placed on a simulator (section 1.2), the set-up error during the simulation session (relative to the intended set-up in the treatment plan), will be present in all subsequent treatment fractions. Hence simulator set-up errors are systematic errors. Furthermore, set-up corrections at the treatment unit are based on comparing portal images with reference images that should reflect the correct position of the patient anatomy with respect to the treatment field. Consequently, if the reference images are simulator radiographs, such set-up corrections cannot reduce the systematic errors introduced during simulation. In chapter 4, the magnitude of the simulator set-up errors was investigated by comparing simulator films with digitally reconstructed radiographs (DRRs). DRRs are calculated directly from the CT scan used in treatment planning and therefore represent the intended set-up exactly.

In chapter 5, we describe the application of a procedure to omit the simulation session, and mark patients only during the CT scan procedure. Together with the use of DRRs as reference images, this enables a true reduction of systematic set-up errors by use of EPI.

Chapters 6 and 7. A low workload method for accurate correction of systematic set-up errors.
In chapter 6 we describe a new (off-line) protocol to reduce systematic errors (i.e., reduce the width Σ of the distribution in Fig. 1.3(b)). This No Action Level (NAL) protocol presents a straightforward approach to executing set-up corrections, and aims at a major reduction in systematic errors at a minimal imaging and image analysis workload. A theoretical description is developed to calculate the impact of this protocol on systematic errors, which is validated with retrospective analyses of measured set-up data and Monte

Carlo simulations. All obtained results are compared to those achieved with the Shrinking Action Level (SAL) set-up protocol, as this is the most frequently applied protocol to date.

An extension of the NAL protocol, the so-called correction verification stage or COVER is introduced in chapter 7. COVER was designed to prevent the systematically erroneous execution of set-up corrections determined with the NAL protocol (e.g, a correction of 1 cm to the left is prescribed for a specific patient, but is consequently executed as a 1 cm shift to the right). Similar to the NAL approach, we investigated how the impact of wrongly executed set-up corrections could be maintained sufficiently small at a minimum workload. To do this, a detailed risk analysis was performed, based on both Monte Carlo simulations and analytical calculations to derive a general recipe for COVER measurements and analysis.

Chapter 8: Set-up accuracy improvement by portal imaging in head-and-neck cancer patients

Here, we describe a clinical study on the set-up accuracy of head-and-neck cancer patients. An off-line correction protocol (SAL) was applied to reduce the systematic set-up errors. In addition, the possible impact of time trends and results achievable with a NAL protocol in this population were investigated in detail.

Chapter 9: Problems in set-up analysis introduced by non-rigid body anatomy

Most of the work introduced above deals with correcting patient set-up based on translations of the patient using the treatment couch. This usually works fine because rotations in the volume of interest can be kept small. However, for larger treatment volumes, such rotations and in general changes in posture of the patient may become important. Most models that describe the impact of rotations on the translations observed in a 2D EPI projection are based on a rigid patient anatomy. However, the occurrence of rotations may be inconsistent with such a rigid anatomy. This problem is illustrated in Chapter 9 for cervix cancer patients treated with fields of up to 40 cm length.

2 Characterisation of a high elbow, fluoroscopic electronic portal imaging device for portal dosimetry

J.C.J. de Boer, B.J.M. Heijmen, K.L. Pasma and A.G. Visser

Abstract

Purpose: The application of a newly developed fluoroscopic (CCD-camera based) Electronic Portal Imaging Device (EPID) in portal dosimetry is investigated. A description of the EPID response to dose is presented in terms of stability, linearity and optical cross talk inside the mechanical structure.

Methods and Materials: The EPID has a relatively large distance (41 cm on-axis) between the fluorescent screen and the mirror (high elbow), which results in cross talk with properties quite different from that of the low elbow fluoroscopic EPIDs that have been studied in the literature. In contrast with low elbow systems, the maximum cross talk is observed for points of the fluorescent screen that have the *largest* distance to the mirror, which is explained from the geometry of the system. An algorithm to convert the images of the EPID into portal dose images (PDIs) is presented.

Results: The correction applied for cross talk is a position dependent additive operation on the EPID image pixel values, with a magnitude that depends on a calculated effective field width. Deconvolution with a point spread function, as applied for low elbow systems, is not required. For a 25 MV beam, EPID PDIs and ionisation chamber measurements in the EPID detector plane were obtained behind an anthropomorphic phantom and a homogeneous absorber for various field shapes. The difference in absolute dose between the EPID- and ionisation chamber measurements, averaged over the four test fields presented in this paper, was $0.1 \pm 0.5\%$ (1 SD) over the entire irradiation field, with no deviation larger than 2%.

2.1 Introduction

Accurate dosimetric treatment verification (within 1 standard deviation (SD) ~ 2%) with Electronic Portal Imaging Devices (EPIDs) has developed rapidly over the past years, and the first detailed patient studies now become available [30, 69, 100]. Recent studies combine the 2D dosimetric information present in EPID images with 3D CT data to derive information about deviations in patient anatomy from the transmission dose [69, 129]. Since EPID images can be used simultaneously for verification and improvement of patient set-up accuracy as well as dosimetric quality assurance [69, 142], their value in the safe delivery of complex conformal radiotherapy plans, particularly if dose escalation is involved, is obvious [43, 146].

Two approaches to dosimetric quality control during patient treatment (*in vivo* dosimetry) with EPIDs can be identified. The first approach, proposed by Leong [74] and Wong *et al* [142], rests on the prediction of a *portal dose image* (PDI) in the EPID detector plane, based on CT and photon beam data, which can be compared to the PDI derived from the EPID image [84, 98]. The potential advantage of the comparison of portal dose in the detector plane is that not only errors in the magnitude and distribution of the dose may be detected, but significant deviations between the patient anatomy during the planning CT scan and during the treatment may also become apparent [69]. The second approach involves back-projection of the EPID PDI to estimate the dose at a dosimetrically relevant point or plane in the patient [30, 100] or even the 3D dose distribution [50, 83, 148], generally assuming that the planning CT scan is consistent with the patient anatomy during treatment. The back-projected dose may then be compared to the dose predicted by the treatment planning system.

Any method of *absolute* dosimetric verification with EPIDs rests on the availability of reproducibly measurable PDIs. The two commercially available EPIDs which are considered suited to render such PDIs are the CCD-camera based fluoroscopic Philips SRI-100 EPID [52, 68, 99] and the Varian Portal Vision liquid-filled matrix ionisation chamber [42, 147]. The SRI-100 system was developed in collaboration with our institute and its dosimetric properties were characterised in detail by Heijmen *et al* [52] and Pasma *et al* [99]. They showed that EPID images can be converted into PDIs with an accuracy of 1 SD = 1% with respect to ionisation chamber measurements for clinically applied fields, including wedged and dynamic multileaf collimated fields. The use of a CCD camera (as opposed to a tube camera), although not favorable in terms of *signal-to-noise ratio* (SNR) due to its small optical coupling factor and relatively low quantum efficiency [2], is appropriate since it yields a long term response stability that meets the requirements of portal dosimetry [38, 43, 52].

Based on the experience with the SRI-100, we developed a new EPID equipped with up-to-date (commercially available and unmodified) components. Special care was taken in the selection of the CCD camera and its electronics, as these were the limiting factors for image SNR in the SRI-100 [2]. The EPID, to which the acronym RAPID (Rotterdam Advanced Portal Imaging Device) was assigned, currently employs a commercially available mechanical structure (TheraView Technology by Cablon Medical BV). This structure is available in a "low elbow" (on-axis distance of 23 cm between fluorescent screen and mirror) and in a "high elbow" version (corresponding distance of 41 cm).

The aim of this paper is twofold: first, we establish that the RAPID detector has the proper characteristics for practical portal dose measurements, i.e. (i) there exists a unique

relation between image pixel value and portal dose measured with an ionisation chamber which is independent of absorber thickness or distance from the EPID, and (ii) this relation is reproducible over both short (minutes to hours) and long (weeks to months) time scales. These investigations were performed with a high elbow system, but apply equally well to the low elbow version of the RAPID. Next, we propose and experimentally verify a method for conversion of EPID images to PDIs which deals specifically with the optical cross talk in the structure of a high elbow EPID. Due to the relatively large distance between screen and mirror, the cross talk of a high elbow system is expected to differ significantly from that measured by Pasma *et al* [99] for the (low elbow) SRI-100 EPID.

2.2 Methods and materials

All references below to position (x,y), pixel size and field size are in a plane parallel to the EPID fluorescent screen which contains the isocentre. The position coordinates x and y are measured in a right-handed system where the positive y direction points towards the gantry and the origin $(0,0)$ coincides with the isocentre. For clarity, we shall sometimes measure distances in the plane of the fluorescent screen and use the coordinates (x_{FS}, y_{FS}) which are equal to $1.5 \times (x,y)$ (see section 2.2.1.1).

2.2.1   Description of the EPID

2.2.1.1 EPID hardware

The RAPID detector consists of a fluorescent screen (FS), a front-surface mirror, a CCD camera, and a PC with a frame grabber. The FS, which is a layer of 150 mg/cm^2 Gd$_2$O$_2$S:Tb (Kodak Lanex Fast-B) coated onto a 2 mm thick brass build-up plate, and the mirror (tilted at an angle of 45° relative to the FS) are part of the standard TheraView structure [89]. We increased the build-up of the FS with a 1 mm thick stainless steel plate to bring the fluorescent layer near the depth of the maximum dose at 25 MV (see section 2.2.3). The on-axis distance between the fluorescent screen and the mirror for the high elbow version of the structure investigated in this paper is 41 cm, with a minimum separation of 22 cm near the edge (Δz in figure 2.1(a)).

The CCD camera (Adimec MX12) is placed in a horizontal position directed towards the mirror (figure 2.1(a)). It is shielded from photo-neutrons by 5 cm thick blocks of Boron-doped polyethylene above and beside the camera. A similar 15 cm thick block, which has a conically shaped opening for the lens, is placed in front of the camera. Inside this neutron shield is a photon shield consisting of a 2 cm thick lead plate above, and a plate of equal thickness but with a circular opening for the lens, in front of the camera.

The CCD contains 1024×512 active pixels of 7.5×15 µm^2 each. The quantum efficiency of the CCD peaks near the central wavelength of the FS (545 nm) at 30% (manufacturer specification). A Navitar High Speed lens with a focal length of 25 mm was selected (stopped at f/1.1) to ensure that the image of the FS area of 38×38 cm^2 ($\Delta y_{FS} \times \Delta y_{FS}$, Δy_{FS} as in figure 2.1(a)) just covers the entire CCD area.

The CCD has frame transfer read-out with a transfer time of 5 µs to the storage part of the chip, implying that, in principle, measurements with virtually no dead-time can be performed since the camera electronics allow a new acquisition to be started during the read-out of the storage part (which takes less time than the minimum frame integration time

of 40 ms). The camera is used in so-called 'trigger mode', in which the camera requires a trigger pulse to initiate a frame acquisition. In this mode, the integration time on the CCD can be selected from 40 to 1280 ms (in steps of 40 ms) by the RAPID software via a serial link to the PC. A printed circuit board (PCB) was developed to generate these trigger pulses in response to a signal from the camera which indicates that a frame has been completely acquired. The analog video signal of the camera is digitized by a 10 bit frame grabber (Matrox Pulsar), yielding images with a pixel value range of 0-1023 ADC units.

The vertical position of the detector, i.e. the source to detector distance (SDD), can be selected via a manual control over a range of ~ 40 cm, but for general clinical use it is kept fixed at 150 cm. All measurements described below were performed at SDD = 150 cm (maximum field size: 25×25 cm^2 at isocentre level) with the 25 MV scanning photon beam of a MM50 Racetrack Microtron (Scanditronix Medical AB) at a nominal dose rate D_r of 300 MU/min.

2.2.1.2 EPID software and image acquisition procedure

The host PC runs a dedicated windows application, which was developed in our hospital. The program supports a patient database, configurable image acquisition procedures, display enhancement and quantitative image analysis tools, EPID performance quality assurance procedures and fully integrated decision protocols for off-line and on-line patient set up corrections [116]. The image processing functions make use of the library of routines (Matrox Imaging Library) that comes with the frame grabber.

Before a new image acquisition is started, the CCD frame integration time (T_{int}) which yields an expected signal of 700 ADC units is calculated. The expected signal is predicted according to the model described by Althof *et al* [2]. The calculation aims at a maximum frame integration time while avoiding saturation of the signal. This approach minimises the contribution of the camera read-out noise on the final image [107]. In practice, signal levels are usually such that T_{int} is between 0.5 and 1.0 s. The total number of frames (n_f) to be acquired to obtain an EPID image for a selected exposure X (MU) is calculated from $n_f = X / [D_r (T_{int} + T_d)]$, where T_d is the dead-time per frame (48 ms) introduced by the PCB described in the previous section. For patient set-up verification, X can be much smaller than the total MU delivered in the treatment field. However, for *in-vivo* dosimetry, the X selected in the software is always 25% larger than the total MU to be delivered. As a result, the total image acquisition time becomes 25% larger than what would be required to integrate the delivered exposure at the nominal dose rate and so drops in effective dose rate of up to 25% (e.g. due to beam start up effects) will not lead to a stop of acquisition before the irradiation has stopped.

The frame integration time of the camera is set to the calculated T_{int} and a pre-exposure 'dark current' image is obtained with a number of frames equal to n_f or, if the exposure time $n_f (T_{int} + T_d)$ exceeds 5 s, the number of frames corresponding to an acquisition time $\leq$ 5 s. The frames are added real-time in a 32 bit deep frame buffer allocated in host RAM. The dark current image measures signal offsets that are mainly due to thermally generated charge on the CCD [2]. Next, the exposed image is obtained in a similar fashion. The addition of frames for the exposed image is started as soon as the average signal level in a frame increases by a software configurable amount (typically 5%). Given the 10 bit ADC resolution, the 32 bit frame buffer can store over 4×10^6 frames without saturation which allows for a maximum exposure of $\sim 10^7$ MU. The dark current image, scaled according to

the ratio of the acquisition times of the exposed and the dark current image, is subtracted from the exposed image.

To improve SNR, the contents of adjacent pairs of pixels in the horizontal direction are added, effectively yielding 512×512 pixels. For the SDD of 150 cm, this yields an effective pixel size of 0.46×0.46 mm^2 at isocentre level. The image is divided by n_f and stored to disk with 16 bit precision and an image header from which the acquisition settings can be retrieved. This 'raw' image can be used for portal dosimetry.

Generally, for patient set-up verification, the raw image is divided by an open field 'calibration image' (with floating point precision) to correct for fixed pattern noise due to position dependent EPID sensitivity variations [2, 67]. The resultant image is processed for optimal contrast resolution on the monitor and displayed. This processed image is also stored and may be used for off-line set-up analysis.

The EPID images that were used in this study were obtained with 150 monitor units (MU) exposure, unless mentioned otherwise. This is slightly larger than the exposures usually encountered in clinical practice (typically 50-100 MU) but ensures a highly reproducible accelerator output (1 SD < 0.2%) which is convenient for measurements (the dose measurements with the ionization chamber described below are not performed simultaneously with the EPID acquisition but a short time after).

2.2.2 *Response stability and linearity*

If a pixel value in the stored raw image is G^s, then the quantity of interest to dosimetry is

$$G \equiv n_f G^s (T_{int} + T_d) / T_{int}$$

which is the accumulated signal over all frames corrected for the dead time of the detector (48 ms, see previous section). The EPID response (G/D_p, where D_p is the portal dose) must meet a number of requirements to be suited for (high precision) *in vivo* dosimetry. Firstly, G/D_p must be sufficiently constant (1 SD $\approx$ 0.5%), both on a short time scale (seconds to minutes) and a long time scale (months), to permit a reliable detection of discrepancies in portal dose of $\sim$ 2%. Secondly, the dependency of the response on the frame integration time must be known so that the response can be predicted for any T_{int} value encountered in clinical practice. Images obtained with a T_{int} that optimises the SNR for set-up verification, may then be converted to PDIs. Finally, G/D_p should be independent of the actually delivered exposure, as verification of this exposure is one of the major goals of *in-vivo* dosimetry.

The *on-axis* pixel value G_c considered below is the average of $G(x,y)$ over a central region of 1.5×1.5 cm^2, which renders the impact of random image noise negligible: a typical clinical RAPID image obtained with a few MU exposure already has a SNR $\sim$ 100 (pixel value divided by random noise on that value, see Althof *et al*, [2]). Since the average is calculated over $\sim$ 900 pixels, the relative error on G_c due to random noise processes is < 0.05% at even the lowest exposures (1-2 MU).

2.2.2.1 *Response Stability*

The *short term* response stability was assessed by sequential acquisition of ten images with an interval shorter than one minute between acquisitions. This procedure was performed 3 times with a separation of approximately one month. Since the accelerator output is very

stable over the short measurement period of a single acquisition session (1 SD ≈ 0.2% [38]), the SD of G_c is a measure of the short term response variation.

The *long term* stability has been derived from results of the beam output quality assurance (QA) procedure which was described by Dirkx *et al* [38]. The QA procedure involves the daily acquisition of portal images of open 17×17 cm^2 fields for the gantry angles 0, 90, 180 and 270°. The exposure for each field is 150 MU. For each gantry angle, the day-to-day variation of G_c was assessed. The values of G_c reflect both accelerator output fluctuations and EPID sensitivity variations. Therefore, the variation in G_c presents an upper limit on the EPID long term sensitivity variations. In addition to the on-axis region, eight 1.5×1.5 cm^2 regions that were centred on the corners and midpoints of the sides of a central square of 12×12 cm^2 were analysed. The average pixel value in these eight regions reflects position dependent changes in the detector sensitivity as well as field flatness variations. For each day, the average pixel values in the off-axis regions were divided by the on-axis G_c value to correct for overall output or response variations. The SD of the day-to-day variation of this ratio was determined in all regions.

2.2.2.2 Response variation with CCD frame integration time Tint

We have investigated how G_c varies with T_{int} for fixed exposures of 50 MU. Images were acquired with a T_{int} that ranged from 280 to 1280 ms for a 10×10 cm^2 open field behind a 15 cm thick flat polystyrene absorber. This range of T_{int} encompasses the frame integration times encountered in clinical practice. A dedicated acquisition procedure, in which the user can select T_{int}, was applied. The exposure X set in the software was 70 MU, to ensure that the delivered 50 MU exposure would be acquired entirely. If the camera output is linear in the amount of optical photons received by the CCD for all T_{int}, and the input stages and ADC of the frame grabber are linear as well, then G_c should be independent of T_{int}.

2.2.2.3 Linearity of pixel value with exposure

We verified the linearity of the system with exposure as follows: images were obtained with fixed acquisition settings and a number of frames n_f sufficient to capture 200 MU at the nominal dose rate. The images were acquired for 20, 40, 60, 90, 120, and 150 MU. The lower bound of 20 MU was based on measurements of the Racetrack Microtron cGy/MU value, which showed that the SD on this value is < 0.2% for exposures > 20 MU (Maarten Dirkx, *priv. comm.*) Due to the frame transfer read-out of the CCD (section 2.2.1.1), the RAPID is an integrating detector (in contrast to a scanning system), and should not be sensitive to beam start-up effects or dose rate variations in general. In the above described measurement situation, one therefore expects that G_c is linear in the delivered MU, which implies constancy of G_c/D_p for a given absorber and treatment beam. We tested this linear relationship by calculating the SD of G_c/MU with respect to the mean value of G_c/MU.

2.2.3 Ionisation chamber measurements and the water equivalent EPID thickness

Ionisation chamber (IC) measurements were adopted as the standard for the assessment of the accuracy of PDIs derived from EPID images. Absolute portal dose measurements were

obtained with a PTW ionisation chamber (N31002) which was placed in the polystyrene miniphantom described by Pasma *et al* [99]. The IC was placed with its centre at a depth of 5 mm in the miniphantom (150.5 cm from the focus). On top of this 5 mm, the build-up could be varied using 5 mm thick polystyrene plates which had the same cross-section as the miniphantom. The phantom was scanned in the FS plane using an empty RFA-300 water phantom (Scanditronix Medical AB). Corresponding EPID images were converted to the same resolution by averaging pixel values in 0.5×0.5 cm^2 regions centred on the IC scan grid.

In general, a signal from the IC in the miniphantom will differ from the corresponding EPID signal due to the fact that (i) the EPID FS is made of non-water equivalent material, (ii) the geometry of the EPID FS is different from that of the miniphantom and (iii) the conversion of deposited energy to a measurable signal is different for the two systems. It is therefore essential that, for the range of clinically encountered patient thicknesses, air gaps between patient and EPID, and field sizes, the ionisation chamber signal is matched to the EPID signal as closely as possible. To achieve this, we follow the method applied by Pasma *et al* [99] which relies on the existence of an optimal build-up thickness of the IC. This optimal build-up yields the most accurate proportionality between EPID signals and absolute IC doses.

The optimal IC build-up thickness, $\hat{d}_{IC}$, which is effectively the water-equivalent thickness of the FS, is measured as follows. For a set of fields (Ω) and IC build-up thicknesses (d_{IC}) the on-axis response is measured for a set of homogeneous water-equivalent absorbers of thickness t with air-gap L to the EPID [98]. The variation in the response for a given field and build-up thickness is defined by

$$\sigma^2(\Omega, d_{IC}) = \frac{1}{N_{t,L}} \sum_{t,L} \left[\frac{G_c(t,L,\Omega)/D_p(t,L,\Omega;d_{IC})}{\left\langle G_c(t,L,\Omega)/D_p(t,L,\Omega;d_{IC}) \right\rangle_{t,L}} - 1 \right]^2 \tag{2.1}$$

where $<A>_B$ denotes the average of A over the range of B and $N_{t,L}$ is the number of (t,L) combinations for which measurements were performed. We define $\hat{d}_{IC}$ to be the value of d_{IC} at which $\sigma_T = \left\langle \sigma(\Omega, d_{IC}) \right\rangle_\Omega$ is minimal. To derive $\hat{d}_{IC}$, EPID images and IC measurements for various d_{IC} were obtained for $t = 10, 25, 40$ cm, $L = 25, 37.5, 50$ cm and square fields Ω with a field side of 5, 10 and 18 cm. As mentioned in section 2.2.1.1, the build-up of the EPID FS is close to the depth of maximum dose at 25 MV so that (i) the impact of electrons generated in the patient on the EPID signal is small and (ii) $\hat{d}_{IC}$ is well defined since the EPID signal will be a weak function of variations in the spectrum of the incident photon beam.

We derive in the Appendix that if $\hat{d}_{IC}$ renders σ_T suitably small, then under reasonable assumptions the image pixel value G at position (x,y) can be expressed by

$$G(x,y) = \left[\frac{G}{D_p} \right]_{ref} \int_{FS} dx' dy' D_p(x',y') S(x',y') K(x,y;x',y') . \tag{2.2}$$

The subscript *ref* refers to values measured in the reference situation, defined by: (x, y) = (0,0), $t = 25$ cm, $L = 37.5$ cm and Ω a 10×10 cm^2 symmetrical field. $S(x,y)$ is the relative sensitivity of the EPID at (x,y) and $K(x,y;x',y')$ is the optical cross talk kernel (more details

are in the Appendix). Note that G in equation (2.2) depends solely on D_p, and not explicitly on patient anatomy or set-up.

2.2.4   Effect of extra build-up layer on EPID spatial resolution

The thickness of the fluorescent screen build-up was increased by 1 mm of stainless steel. The effect this might have on the spatial resolution was investigated with the QC-3 line bar pattern phantom [118] which contains line pairs (lp) at spatial frequencies of 0.1, 0.2, 0.25, 0.45 and 0.75 lp/mm. The phantom was placed on top of the FS and images were obtained with and without the steel plate. The relative (square wave) modulation transfer function (RMTF) and its half value spatial frequency, f_{50}, were obtained according to the method described by Rajapakshe *et al* [106]. To estimate the accuracy of the f_{50} determination, the regions of interest (ROI) were shifted by ± 2 pixels (in steps of 1 pixel) in the x and y direction, the size of the ROI was varied by $\pm 20\%$ in steps of 10%, and f_{50} was calculated for each case.

2.2.5   An algorithm to convert high elbow EPID images into PDIs

Pasma *et al* [99] have published an algorithm to calculate PDIs from images obtained with the SRI-100 system. They assumed an approximately shift invariant kernel to cast equation (2.2) into the form

$$G(x,y) = \left[\frac{G}{D_p} \right]_{ref} \int_{FS} dx' dy' D_p(x',y') S_w(x',y') K(x-x', y-y') . \qquad (2.3)$$

In their description, the function S becomes slightly field size dependent ($S \rightarrow S_w$) to compensate for the fact that K (which is measured only *on-axis*) is not completely shift invariant. It therefore no longer represents the true *local* EPID sensitivity, since it also includes a contribution from optical cross talk.

For a high elbow EPID, the situation is quite different, as becomes apparent from figure 2.1. The EPID image may be decomposed into two components. The first component is the primary contribution, which consists of photons that are reflected once off the mirror and then reach the camera lens (the solid arrows, labelled 'p' in figure 2.1). For a point source in the FS plane, the primary contribution is made up by photons emitted within the solid angle subtended by the camera lens aperture as seen from that point source ($\sim 3 \times 10^{-4}$ sr). The secondary contribution, or cross talk, stems from photons reaching the lens after multiple reflections on the mirror and reflection or scatter from the (nearly white) fluorescent screen. We assume the inside of the EPID structure to be perfectly black and hereafter indicate by cross talk the light which is reflected from the mirror back to the FS and re-emitted from the screen to the mirror. Based on this assumption, and the fact that the plane $x = 0$ is perpendicular to the mirror, we can find the regions of the FS which may potentially yield cross talk by considering the projection of pathways of light onto the plane $x = 0$.

As figure 2.1(a) reveals, for a source point at $y = - 6$ cm ($y_{FS} = - 9$ cm) only a primary contribution exists. Since we are interested in dosimetric information on a scale of $\sim 0.5 \times 0.5$ cm^2, this implies $K(x, y = -6; x', y') \propto \delta(x-x', -6-y')$ with δ the Dirac delta distribution. However, for a source point at $y = 0$, cross talk may occur at $y > 0$, as indicated by figure 2.1(b).

The dotted line arrows in figure 2.1(b) are the rays that enclose the beam of emission (of opening angle θ_s) which could generate such cross talk. Note that the region in which cross talk is manifested is rather extended in the y direction (in this case, from $y = 3.3$ cm to $y = 12.7$ cm or equivalently $y_{FS} = 5$ cm to 19 cm).

Figure 2.1(c) reveals that for a source point at $y > 0$, similar cross talk is generated for an even larger range, namely all positive y and also at slightly negative y. Figure 2.1(d) depicts what happens when the x direction is entered into the problem. We have drawn the projection onto the x-z plane of a number of rays which originate at $x = y = + 6$ cm and
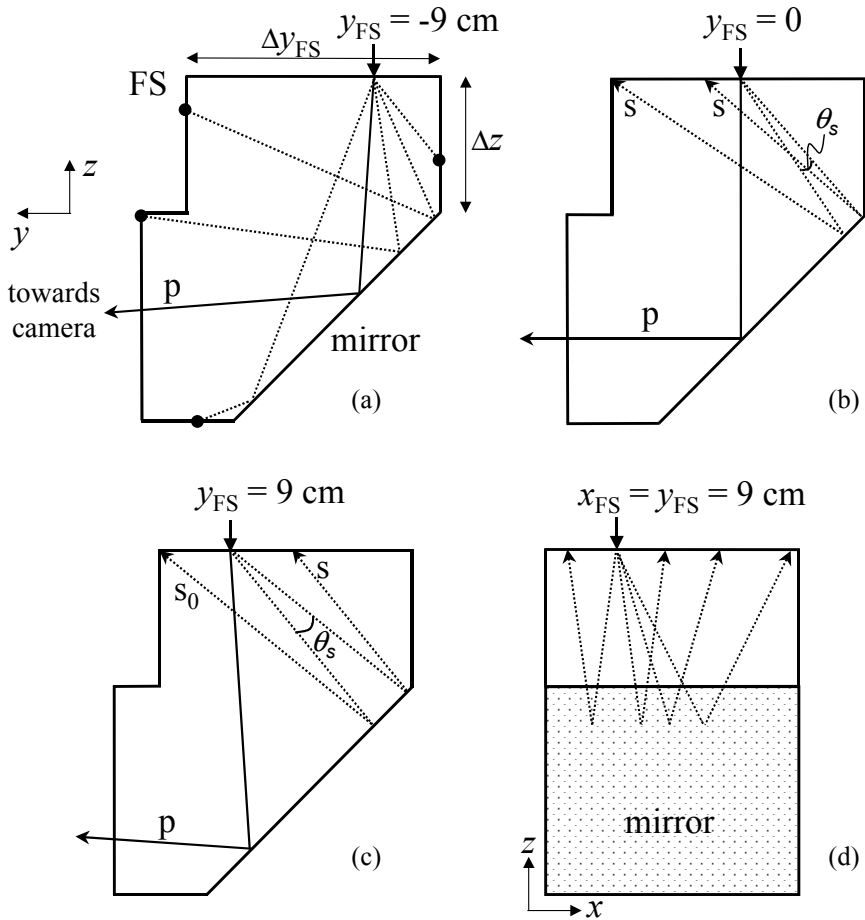
Figure 2.1. Schematic view of light propagation inside the high elbow structure. The drawings in (a)-(c) depict a (to scale) cross-section at constant x, and the optical photon paths (indicated with arrows) are perpendicular projections onto this plane. Solid lines represent the pathways that give a primary contribution to the image signal (indicated by 'p'), dashed lines represent secondary pathways. Secondary pathways that may give rise to cross talk in the image are indicated by 's'. The secondary pathways indicated in (b) and (c) are the extremes of the range that could possibly contribute to cross talk for a source point at the indicated y-position. In (d), a projection onto a $y = $ constant plane is given of rays which project in (c) as the line indicated by s_0.

which would all occur like the ray indicated by the label 's$_0$' in figure 2.1(c) when projected onto the y-z plane. From this figure, and the symmetry of the EPID, it is evident that if a source point at a certain position (x,y) generates a cross talk contribution at a point (x',y'), it will do so for all x and x'; in fact, all x are optically coupled to all x' for such a pair (y,y'). Therefore, the cross talk contribution will be a broad function in x' too.

From the geometry depicted in figure 2.1(a) one can easily derive that the FS region $A_{s,\mathrm{FS}}$ that may generate cross talk is the area within $\Delta y_{\mathrm{FS}}/2 < x_{\mathrm{FS}} < \Delta y_{\mathrm{FS}}/2$ (in fact, all x in the FS) and $y_{\mathrm{min,FS}} < y_{\mathrm{FS}} < \Delta y_{\mathrm{FS}}/2$ with $y_{\mathrm{min,FS}} = -\Delta y_{\mathrm{FS}}/2 +(\Delta z)^2/\Delta y_{\mathrm{FS}}$ as long as $\Delta y_{\mathrm{FS}} > \Delta z$ (the secondary contribution would completely vanish if $\Delta y_{\mathrm{FS}} < \Delta z$). Note that, due to the fact that the direction of each ray in figure 2.1 which adds to the cross talk may be reversed, the cross talk in the image can only appear in $A_{s,\mathrm{FS}}$.

Similarly, we may calculate the source region $A_{s,\mathrm{FS}}(y_{\mathrm{FS}})$ specific to cross talk at y_{FS}. For $y_{\mathrm{FS}} > y_{\mathrm{min,FS}}$, $A_{s,\mathrm{FS}}(y_{\mathrm{FS}})$ has a finite area, given by $\Delta y_{\mathrm{FS}}/2 < x_{\mathrm{FS}} < \Delta y_{\mathrm{FS}}/2$ and $y_{\mathrm{min,FS}}(y_{\mathrm{FS}}) < y_{\mathrm{FS}}$ $< \Delta y_{\mathrm{FS}}/2$ with $y_{\mathrm{min,FS}}(y_{\mathrm{FS}}) = -\Delta y_{\mathrm{FS}}/2 +(\Delta z)^2/(y_{\mathrm{FS}} + \Delta y_{\mathrm{FS}}/2)$. Again, because ray directions can be reversed, a source point at y_{FS} may *contribute* to the cross talk in, and only in, $A_{s,\mathrm{FS}}(y_{\mathrm{FS}})$, which implies that $A_{s,\mathrm{FS}}(y_{\mathrm{FS}})$ is also the source region for all multiple reflections between the mirror and the fluorescent screen that may contribute to cross talk at y_{FS}. By substition of $(x,y) = (x_{\mathrm{FS}},y_{\mathrm{FS}})/1.5$ in the above expressions, it is straightforward to obtain the source region in the (x,y)-coordinate system. Hereafter, we will only refer to (x,y) and source region $A_s(y)$ in the development of the algorithm.

The amplitude of the secondary contribution is expected to be small relative to the primary contribution, due to (i) the (relatively) large distance between screen and mirror, (ii) the forwardly peaked light distribution of the primary emission [143], (iii) the small solid angle in which the emission relevant to cross talk occurs (see θ_s in figure 2.1) and (iv) the small fraction of light re-emitted from the FS which will actually contribute to the observed cross talk (light which reflects from the FS will never reach the lens, so only the fraction which is re-emitted by the FS into the same small solid angle as the primary emission can be detected in the EPID image).

Based on this smallness of the cross talk we develop equation (2.2) to first order in K. The kernel may be separated into $K(x,y;x',y') = \delta(x-x',y-y') + \varepsilon(x,y;x',y')$, where the δ-function is due to the primary contribution, and ε describes the cross talk contribution.

Since $D_p = \dfrac{G}{S}\left[\dfrac{D_p}{G}\right]_{ref}$ + higher order terms in ε , the first order expansion of equation

(2.2) becomes

$$G(x,y) = \left[\frac{G}{D_p}\right]_{ref} S(x,y)D_p(x,y) + \int_{A_S(y)} dx' dy'\, \varepsilon(x,y;x',y')G(x',y') \qquad (2.4)$$

where the integral has now been confined to the source region $A_s(y)$ discussed above (the integral vanishes for $y < y_{\mathrm{min}}$, yielding $D_p = \dfrac{G}{S}\left[\dfrac{D_p}{G}\right]_{ref}$ for this region).

Suppose that the field width (the extent in the x-direction) at y', $w(y')$ is constant for the source region of y: $w(y') = w_0$ for $y' \in A_s(y)$. The number of source points which contribute to the cross talk at y is then proportional to the product of w_0 and the length of $A_s(y)$, $l_s(y) = [\Delta y_{\mathrm{FS}} -(\Delta z)^2/(1.5y+\Delta y_{\mathrm{FS}}/2)]/1.5$. Since all points in $A_s(y)$ contribute to the cross talk (albeit in slowly varying degrees, e.g. due to the emission profile, the optical pathlength and the

angle of incidence after reflection off the mirror), we assume that the amount of cross talk at y scales with the average pixel value $<G>(y)$ in $A_s(y)$:

$$\int_{A_S(y)} dx'\, dy'\, \varepsilon(x, y; x', y') G(x', y') = <G>(y) f(y; w_0) \tag{2.5}$$

with the cross talk distribution function $f \ll 1$. We expect that f increases with increasing y because $l_s(y)$ increases monotonically with y ($> y_{min}$). The quantity $<G>(y)$ is defined by

$$<G>(y) = \frac{\displaystyle\int_{A_S(y)} dx'\, dy'\, G(x', y')}{w_0 l_s(y)}. \tag{2.6}$$

For an arbitrary field shape, we assume that equation (2.5) still holds if we replace w_0 in equations (2.5) and (2.6) by

$$<w>(y) = \frac{\displaystyle\int_{y_{min}(y)}^{\Delta y/2} dy'\, w(y')}{\displaystyle\int_{y_{min}(y)}^{\Delta y/2} dy'\, H(w(y'))} \tag{2.7}$$

where H is the step function ($H(x) = 0$ if $x \le 0$, $H(x) = 1$ if $x > 0$). Using this approximation, we assume that the cross talk generated by the pixel value distribution $G(x',y')$ for $(x',y') \in A_s(y)$ is equal to the cross talk generated by homogeneous pixel value $<G>(y)$ in a rectangular "equivalent" field of width $<w>(y)$ and length $l_s(y)$. The definition of $<G>(y)$ (equation (2.6)) assures that the amount of detected light generated in the equivalent field is equal to that in the source region:

$$\int_{A_S(y)} dx'\, dy'\, G(x', y') = <w>(y)\, l_s(y)\, <G>(y).$$

With the aid of the expression in equation (2.5), we can now derive D_p from equation (2.4) by

$$D_p(x, y) = \left[\frac{D_p}{G}\right]_{ref} \frac{G(x, y) - <G>(y) f(y; <w>(y))}{S(x, y)}. \tag{2.8}$$

To calculate D_p from G we must obtain the sensitivity array S and the cross talk distribution functions f. We define $S_{w \times 20} = (G/D_p)/(G/D_p)_{ref}$ for a field of constant width $w(y) \equiv w$ and length of 20 cm, for $t = 25$ cm and $L = 37.5$ cm and measure the array $S_{20 \times 20}(x,y)\ \forall\ (x,y)$ and the profiles $S_{w \times 20}(0,y)\ \forall\ y$ and $w = 4, 8, 10, 12, 14, 16$ and 18 cm. To determine $S_{w \times 20}$, G is obtained from EPID images and D_p from IC measurements of the $w \times 20$ fields. Note that, with the field length of 20 cm at isocentre level used in these measurements, the *effective* $\Delta y_{FS} = 30$ cm (instead of 38 cm) yielding a value of $y_{min} = -4.5$ cm.

From equation (2.8) it is easy to derive that, to first order in f,

$$S_{w \times 20}(x, y) = S(x, y) \left[1 + \frac{<G_{w \times 20}>(y)}{G_{w \times 20}(x, y)} f(y; w)\right] \tag{2.9}$$

where $G_{w \times 20}$ is the pixel value distribution for the $w \times 20$ field and $<G_{w \times 20}>$ is calculated according to equation (2.6). Since, in our first order model, the cross talk will vanish for $<w> = 0,$ we derive S and f from

$$S(0,y) = \lim_{w \to 0} S_{w \times 20}(0,y) \quad \wedge \quad f(y;w) = \left[\frac{S_{w \times 20}(0,y)}{S(0,y)} - 1 \right] \frac{G_{w \times 20}(0,y)}{<G_{w \times 20}>(y)}. \quad (2.10)$$

The limit $w \to 0$ is obtained from linear extrapolation to $w = 0$ of $S_{w \times 20}(0,y)$ for each y. Finally, the local sensitivity array $S(x,y)$ may be calculated using equations (2.9) and (2.10):

$$S(x,y) = S_{20 \times 20}(x,y) \left[1 + \frac{<G_{20 \times 20}>(y)}{G_{20 \times 20}(x,y)} f(y;20) \right]^{-1}. \quad (2.11)$$

An arbitrary EPID image, $G(x,y)$, can now be converted into a PDI with the aid of equations (2.6-8). The function $f(y;<w>(y))$ to be applied in equation (2.8) is found from linear interpolation of $f(y;w)$ in w between the two measured field widths that enclose $<w>(y)$.

2.3 Results

2.3.1 Response Stability

Regarding the short term stability, the SD of G_c was 0.2% with respect to its mean for all 3 measurement sessions, comparable to the short term output fluctuations of the accelerator and sufficiently small for *in vivo* dosimetry.

The beam output QA measurements showed a long term variation of G_c over a period of 306 days of 0.5% (1 SD), independent of gantry angle. This variation includes the (long term) output fluctuations of the accelerator. Ionisation chamber output measurements over the same period indicate a similar fluctuation (1 SD $\approx$ 0.5%). In all 8 off-axis regions, the variation of G/G_c in the 306 day period was 0.2% (1 SD), and again no dependency on gantry angle was observed. These results are similar to those found by Dirkx *et al* [38] for the SRI-100 EPID. We conclude that the RAPID is sufficiently stable to perform *in-vivo* portal dose measurements without frequent recalibration.

2.3.2 Linearity with frame integration time

The results of the linearity measurements are displayed in figure 2.2. The reproducibility of the measurements was better than 0.2% (1 SD) for all 9 studied T_{int}, consistent with the reproducibility found in section 2.3.1. The value of G_c showed little variation with T_{int} : the SD of G_c was 0.16% with respect to the average (note that the value of G_c^s, which is a measure for the charge accumulated by the CCD per image frame, varied by over a factor 4). Therefore, the EPID response is independent of T_{int}.

2.3.3 Linearity with exposure

The value of G_c turned out to be perfectly linear in the delivered exposure over the entire exposure range 20-150 MU : 1SD = 0.17%, which is equal to the short term reproducibility of the accelerator output. Therefore, within the measurement accuracy, G_c/D_p is constant for

a given treatment situation. The smallness of the SD confirms that the SNR of G_c (which was obtained from a single image), is sufficiently large, even at small exposures, to render the random noise on G_c negligible (see section 2.2.2). To obtain clinically relevant information from PDIs, a pixel size of 0.5 cm at isocenter in both directions yields sufficient spatial resolution [69]. With the numbers quoted in section 2.2.2, we find that the random noise on pixel values $G(x,y)$ obtained at such a resolution $\sim 0.1\%$ in a clinical image of only 2 MU exposure. Combined with the above results on the linearity of G_c in the delivered exposure, we conclude that the number of MU used to generate a PDI presents no relevant limitation on its accuracy.

2.3.4   Ionisation chamber build-up

The optimal thickness $\hat{d}_{IC}$ of the IC build-up was found to be 3 cm of polystyrene, for which $\sigma_T = 0.5\%$ and all the measured residual errors (the term in brackets in the right-hand side of equation (2.1)) are less than 1%. The on-axis G/D_p ratios obtained for the 3 studied field sizes and for "isocentric" irradiation ($L + t/2 = 50$ cm) are depicted in figure 2.3. We conclude that the use of $\hat{d}_{IC}$ renders G/D_p sufficiently constant to be applicable to *in vivo* dosimetry. The extra 1 mm thick stainless steel plate mounted on top of the FS did not appreciably reduce f_{50}. From the QC-3 measurements we obtained $f_{50} = 0.217 \pm 0.002$ (1 SD) lp/mm for the standard screen, and $f_{50} = 0.213$ mm ± 0.002 lp/mm with the added steel
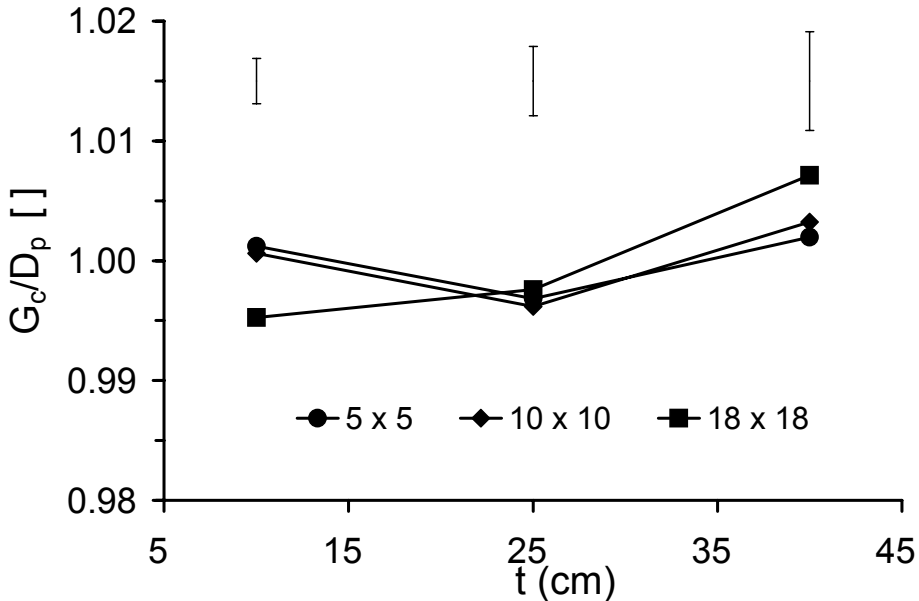
Figure 2.3. The ratio G_c/D_p for isocentric irradiation of absorbers of thickness t per field size. The curves are normalized to the mean per field size. The 2 SD wide error bars at the top of the graph were estimated from repeated G_c and D_p measurements.

plate. The actual reduction of f_{50} of 2% is therefore within the estimated accuracy of the experiment. The value of f_{50} is estimated from two sample points of the RMTF at relatively low spatial frequencies [106]. However, also at the higher frequencies 0.25 and 0.45 lp/mm the value of the RMTF measured with the extra steel plate was identical to the value measured with the standard screen. This finding is consistent with the fact that we could not separate by eye the corresponding QC-3 phantom images. Therefore, the expected impact of the steel plate on the quality of clinical images is negligible.

2.3.5   The EPID sensitivity array and the cross talk distribution functions

The profile $S(0,y)$, as derived from equation (2.10), is shown in figure 2.4(a). We found that to derive the limit $w \rightarrow 0$ of $S_{w \times 20}(0,y)$ a linear extrapolation to $w = 0$ is appropriate since the change in $S_{w,20}(0,y)$ with w is to a high accuracy linear in w for all y: we fitted a first order polynomial in w to $S_{w,20}(0,y)$ for each y, and found that the deviation over the full range of w between this fit and the measured values, averaged over y, had 1 SD = 0.2%.

   In figure 2.4(a) we also show profiles of $S_{w \times 20}$ at x = -8, 0 and 8 cm. For comparison purposes, the $S_{20 \times 20}$-profiles at x = -8 and 8 cm were normalized to have the same average as the $x = 0$ profile, as indicated by the prime in the figure legends. The small scale noise on the profiles is due to inhomogeneities in the light output of the FS, which are different for each x. The profile indicated by the crosses in figure 2.4(a), was obtained by fitting a second order polynomial to the average profile of $S_{20 \times 20}$ for the 3 displayed x values. From the difference of the latter profile with $S(0,y)$ we find that the cross talk manifests itself increasingly at positive y, as expected from the fact that $l_s(y)$ increases with y. We also see that the $S_{20 \times 20}$-profiles at x = -8 cm and +8 cm have the same large scale structure as the profile at $x = 0$. In fact, the SD of the $S_{20 \times 20}$-profiles with respect to the polynomial fit is 0.8%, 0.9% and 0.8% respectively for x = -8, 0 and 8 cm, reflecting that the effect of cross talk is hardly dependent on x, as was made plausible in section 2.2.5.

   The above results indicate that the approximation expressed by equation (2.5) is valid, and so we applied equation (2.10) to derive the cross talk distribution
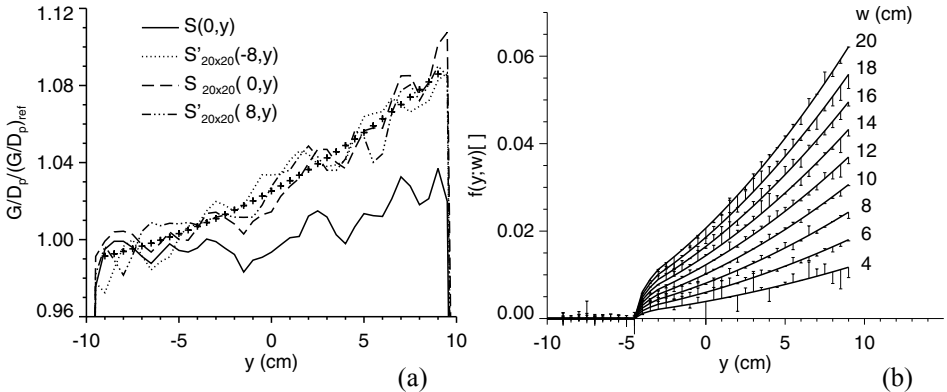
Figure 2.4. (a) Profiles along the y-axis of the sensitivity S and the array $S_{20 \times 20}$. Details are in the text. (b) The cross talk distribution functions $f(y;w)$. The vertical lines connect the actually measured points to the corresponding fitted curves.

functions $f(y;w)$ displayed in figure 2.4(b). The cross talk is indeed of small amplitude (always less than 6%), and so the first order approach developed in section 2.2.5 is appropriate. As predicted, there is negligible cross talk for $y < -4.5$ cm and in fact we put $f(y;w) \equiv 0$ for this range of y. For the other y-values, the solid curves in figure 2.4(b) were determined as follows. First, we observe that $f(y;w)$ is almost perfectly linear in w for all y, as expected from section 2.2.5 and the linearity in w of $S_{w \times 20}$ described above. This implies that $f(y;w)$ may be separated into variables according to $f(y;w) = w \times f_0(y)$, which defines the 'shape function' $f_0(y)$. Inserting this result into equation (2.10), and taking the first derivative with respect to w we obtain

$$f_0(y) = \frac{1}{S(0,y)} \lim_{w \to 0} \left[\frac{G_{w \times 20}(0,y)}{<G_{w \times 20}>(y)} \left(\frac{\partial S_{w \times 20}(0,y)}{\partial w} \right)_y \right].$$ (2.12)

The limit at the right-hand side of equation (2.12) was obtained by linear extrapolation to $w = 0$ of the term in square brackets (which is virtually constant in w). The solid lines in

(a)   (b)
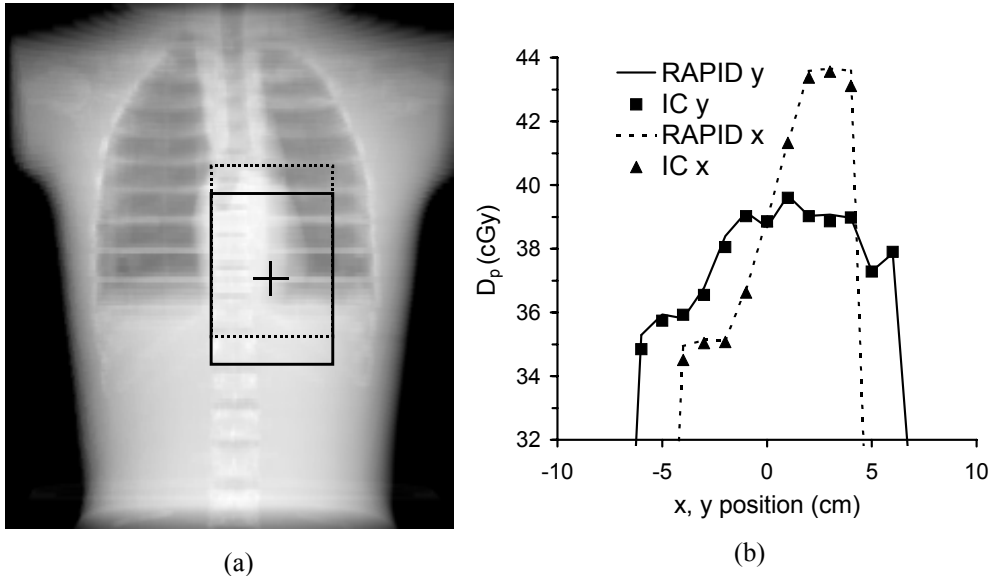
Figure 2.5. (a) Outlines of the 10×14 cm^2 fields used to assess the accuracy of the proposed method for conversion of EPID images into PDIs, projected onto a DRR of the Alderson phantom. The cross denotes the isocentre, the solid line the symmetrical field and the dashed line the asymmetrical field. (b) Absolute dose profiles for the RAPID PDI and for the scanning IC PDI for the symmetrical field in (a). (c) (*next page*) dPDI as derived from RAPID and IC PDIs for the symmetrical field in (a), expressed in %. Each square has an area of 1 cm^2, the IC measurement points being at the centre of the squares. The dashed line indicates the field edge, as extracted from the EPID image. The projected isocentre is at the centre of the square which has the inverted grey scale. (d) As (c), but now for the asymmetrical field displayed in (a).

figure 2.4(b) are equal to $wf_0(y)$. Figure 2.4(b) illustrates that the same shape function $f_0(y)$ describes $f(y;w)$ well for all w (the SD of the difference between the fit and the measured data is < 0.2% for each w, averaged over w it is 0.1%). These results confirm the assumptions yielding equation (2.12), which include the cross talk model presented in section 2.2.5. The curves $wf_0(y)$ were applied as $f(y;w)$ in subsequent analyses.

2.3.6   The accuracy of PDIs derived from EPID images

In figure 2.5(a), a DRR of the Alderson phantom is presented with the outlines of a symmetrical and asymmetrical field of 10×14 cm^2 in the thorax region. The asymmetrical field extends 9 cm superiorly ($y > 0$) of the isocentre and 5 cm inferiorly, and therefore covers the y range where the magnitude of the cross talk is largest.The magnitude of the portal dose variation inside the field can be judged from figure 2.5(b), where the absolute dose profiles along the x- and y-directions as derived with equation (2.8) from the measured EPID image are compared with the absolute dose obtained with the scanning IC for the symmetrical field. The differences in the absolute doses measured with RAPID and with the IC for the   symmetrical field may be appreciated from figure 2.5(c), where the difference of the PDI derived from the RAPID image and the PDI measured with the IC (hereafter 'difference PDI' or dPDI) is displayed in %. The dPDI values are $0.2 \pm 0.4\%$ (1 SD) over the field area. Note that this difference includes noise in the IC and EPID measurements. The dPDI for the asymmetrical field is shown in figure 2.5(d), which illustrates that also in this case the agreement over the field is satisfactory: $0.0 \pm 0.8\%$, with all deviations < 2%.The validity of equation (2.8) was evaluated for more extreme field shapes using a 25 cm thick polystyrene absorber with its centre of mass positioned at the

1.1	0.9	0.8	-0.1	-0.1	-0.3	0.5	0.6	1.5
0.8	0.1	0.4	0.3	0.4	0.6	0.5	0.3	1.3
0.7	0.5	0.1	-0.1	-0.1	-0.3	-0.1	-0.1	0.6
0.7	0.6	0.4	-0.2	0.1	-0.0	-0.1	-0.1	0.6
0.9	0.3	0.5	-0.1	0.1	-0.3	-0.3	0.0	0.6
0.7	0.1	0.2	0.0	-0.1	-0.3	-0.4	-0.4	0.5
0.5	0.2	0.1	-0.1	-0.1	-0.5	-0.1	-0.3	0.2
0.4	0.1	-0.1	-0.3	0.0	0.1	-0.1	-0.2	0.3
0.6	0.2	0.1	-0.1	0.3	0.2	0.1	0.1	0.8
0.7	0.4	0.1	0.3	0.3	0.1	0.3	0.1	0.9
0.2	-0.2	-0.2	-0.2	-0.3	-0.2	-0.1	-0.2	0.5
0.7	0.1	0.3	-0.3	-0.2	-0.2	-0.2	-0.3	0.4
1.1	0.6	0.1	0.3	0.4	0.5	0.3	0.1	1.1

2.5(c)

1.2	0.7	-0.1	-0.1	0.0	0.0	0.3	0.3	1.1
0.3	0.1	-0.4	-1.0	-1.1	-1.6	-0.6	-0.6	0.2
-0.2	-0.4	-0.4	-1.3	-1.4	-1.3	-0.3	-0.1	0.9
0.1	-0.7	-0.1	-0.5	-0.2	-0.1	0.5	0.4	1.3
-0.1	-0.0	-0.7	-1.0	-1.0	-1.1	-0.9	-0.6	0.2
-0.3	-0.0	-0.4	-0.9	-1.0	-0.7	-0.7	-0.6	0.2
0.5	0.0	0.0	-0.7	-0.5	-1.0	-1.0	-0.6	0.3
-0.1	-0.5	-0.0	-0.4	-0.4	-0.9	-1.1	-1.0	-0.1
0.7	0.0	-0.2	-0.4	-0.2	-0.5	-0.1	-0.3	0.2
0.9	0.3	-0.1	-0.3	0.2	0.1	-0.0	0.1	0.5
1.2	0.5	0.4	0.2	0.4	0.4	0.1	0.0	0.9
1.4	1.2	0.5	0.8	0.8	0.8	0.8	0.6	1.3
1.6	1.5	1.5	1.5	1.5	1.5	1.4	1.5	1.9

2.5(d)

isocentre. In figure 2.6(a), the dPDI of a diagonal multileaf collimator (MLC) field, which extended in the y-direction from -9 to +9 cm, is shown. The separation between opposing leaves was 10 cm. This particular field tests for the x-shift invariance of the correction $f(y;<w>(y))$, i.e. the lack of dependency on x expressed by the right-hand side of equation (2.5). From figure 2.6(a) we conclude that the cross talk is indeed largely x independent, since application of the same $f(y;<w>(y))$ for all x yields absolute dose differences of $-0.1 \pm 0.3\%$ over the entire field. Finally, the field shape used to generate the dPDI in figure 2.6(b) tests for the validity of the assumption that the cross talk for a field of varying width may be derived from that of an equivalent field width expressed by equation (2.7). In figure 2.6(b), the field width runs from $w = 18$ cm at the top ($y = 9$ cm) to $w = 5$ cm at the bottom ($y = -9$ cm.) Even for this somewhat extreme case, the accuracy of the EPID PDI is $-0.6 \pm 0.6\%$, with no deviation larger than 1.5%.

2.4 Discussion

An algorithm for the generation of absolute portal dose images has been developed and applied to images obtained with the RAPID high elbow fluoroscopic EPID for a 25 MV photon beam. Images of an anthropomorphic phantom and of field shapes which test the assumptions behind the algorithm for extreme cases yield PDIs that agree well with absolute portal dose measurements performed with an IC (mean deviation per field $\leq 0.6\%$, 1 SD over the field $\leq 0.8\%$). The maximum observed difference was 1.9% in a PDI pixel close to the field edge.

The PDI algorithm is based on a first order description of the optical cross talk (i.e. reflection and re-emission via the fluorescent screen) in the high elbow structure, which is applicable because the cross talk is of small magnitude: at worst 6%, in the region furthest

(a)   (b)

Figure 2.6. (a) dPDI for a diagonally shaped field, in %. This figure is similar to figure 2.5(c), however the squares now cover an area of 1.5×1.5 cm^2, reflecting the larger step between IC measurements for this case. For the bottom row (which would still fit inside the field) no IC measurements were available. (b) As (a), but for a triangular field shape.

away from the mirror, and on-axis $\leq 2\%$. Because the cross talk for the high elbow is to first order independent of the x coordinate (parallel to the mirror), it can be accounted for by a simple correction, which is derived from 8 portal dose *profiles* measured along the line $x=0$. However, although the high elbow cross talk is relatively small, it must be corrected for in high accuracy dosimetric applications. For example, if we apply $S_{20,20}$ instead of the sensitivity S and set $f \equiv 0$ in equation (2.8) (thus effectively neglecting cross talk), we obtain absolute dose differences of 1.6 ± 1.3 % (1 SD) for the case depicted in figure 2.5(d) with extremes of up to 4.4% at the most positive y values.

The situation is quite different from that of the low elbow SRI-100 system, where the on-axis cross talk for a 20×20 cm^2 field is $\sim 10\%$ and the cross talk increases towards the mirror [99]. The different magnitude and spatial behaviour of the cross talk in the high elbow system studied in this paper can be attributed to the larger distance between mirror and screen. With the high elbow PDI algorithm described in this paper, a single sensitivity array is sufficient for all field shapes, contrary to the algorithm developed for the SRI-100. The latter algorithm requires a set of sensitivity arrays for various field sizes, which each express both EPID sensitivity *and* the (small) cross talk contribution that remains after deconvolution with a kernel that was derived from on-axis measurements [99].

Since the actual photon detector of the system is a CCD, the RAPID response has a short and long term stability which equals that of the SRI-100 EPID [38]. The current data on the stability imply that the frequency of a calibration of the system for *in vivo* dosimetry can be as low as once a year.

The simplified electronics of the RAPID yields advantages for both low and high elbow systems. For the SRI-100 EPIDs investigated at our institute, it was found that the measured signal is nonlinear in the amount of light incident on a CCD pixel [99]. This phenomenon could be largely attributed to the nonlinearity of the complex camera and sampling electronics. For the RAPID, deviations from linearity of pixel value with signal intensity per frame (i.e. frame integration time) are negligible.

The use of the low elbow version of the RAPID for in vivo dosimetry is currently investigated, based on the algorithms developed for the SRI-100 EPID. In the future, we will also investigate whether we can apply a cross talk correction algorithm similar to the algorithm described in this paper to the low elbow RAPID to account for the spatial dependency of the cross talk in that EPID structure. In addition, we are investigating the properties of the high elbow RAPID for a 10 MV beam to expand the possibilities of clinical application of PDIs. The 10 MV beam requires a different optimal build-up for the IC [101], but little changes are expected in the PDI algorithm presented in this paper. In fact, we will attempt to apply the same cross talk distribution functions at both energies.

At present, we are interfacing the MU signal from the accelerator to the RAPID software. This feature, together with the fast image capture and image processing capabilities of the system, will enable on-line tracking of the leaf positions during dynamic MLC treatments [96] combined with a PDI measurement derived from the composite EPID image. With an EPID which can simultaneously check leaf (and block) positions and portal dose, deviations between measured and predicted PDIs can be separated into treatment machine errors and anatomy deviations with respect to the planning CT scan [69] during dynamic treatments.

Acknowledgements

Appendix: The relation between portal dose and fluoroscopic EPID image pixel value

We denote the portal dose measured with an IC with build-up d_{IC} in the plane of the FS behind a homogeneous absorber of thickness t, with the exit surface at a distance L from the FS and for a field shape Ω by $D_p(x,y,t,L,\Omega;d_{IC})$. Suppose that a fluoroscopic EPID could be constructed which was free of cross talk. Then, for the absorber described above, the image pixel value at position (x,y) can always be exactly expressed by

$$G(x,y,t,L,\Omega) = \left[\frac{G}{D_p(d_{IC})}\right]_{ref} D_p(x,y,t,L,\Omega;d_{IC})\,\Phi(x,y,t,L,\Omega;d_{IC})S_r(x,y)\,. \quad \text{(A2.1)}$$

The measurement situation for the reference response value ($[G/D_p(d_{IC})]_{ref}$) is described in section 2.2.3. The *relative screen emission* Φ describes the ratio of the light output of a *perfectly homogeneous* FS to D_p and is implicitly normalized to $[\Phi]_{ref} \equiv 1 \; \forall \; d_{IC}$ through $[G/D_p(d_{IC})]_{ref}$. The quantity Φ takes into account the differences between the FS and the IC as dose detectors which were mentioned in section 2.2.3: it describes the relative difference between the IC measurement and the process of energy conversion and optical photon transport and escape in the FS. This *interpretation* of Φ, together with the fact that we take the product of D_p and Φ in equation (A2.1) and not a convolution, reflects the assumption that the difference in secondary electron transport in the miniphantom-IC combination and the FS is negligible on the scale of $\sim 0.5 \times 0.5$ cm^2 we are interested in. The thickness and composition of the homogeneous FS are assumed to be equal to the actual thickness and composition at $(x,y)=(0,0)$.

$S_r(x,y)$ is the position dependent relative sensitivity ($S_r(0,0) \equiv 1$) which describes variations in the FS layer thickness, the spatial distribution of the phosphor grains as well as light distribution effects such as the nearly Lambertian emission profile [99, 143].

In reality, fluoroscopic EPIDs exhibit optical cross talk due to reflections and scatter inside the mechanical structure. If $K(x,y;x',y')$ is the optical cross talk kernel which takes into account the signal contribution of a light emitting point at (x',y') to a point at (x,y), then equation (A2.1) can be generalized to

$$G(x,y,t,L,\Omega) = \left[\frac{G}{D_p(d_{IC})}\right]_{ref} \int_{FS} dx'dy'[D_p(x',y',t,L,\Omega;d_{IC})$$

$$\times \Phi(x',y',t,L,\Omega;d_{IC})S_r(x',y')K(x,y;x',y')]. \quad \text{(A2.2)}$$

To simplify matters, we seek an 'optimal' IC build-up, $\hat{d}_{IC}$, which renders $\Phi \approx 1$ (to within 1%) for clinically relevant t, L and Ω (i.e., the light emitted by the FS is proportional to the IC reading). Based on equation (A2.2) we derive a method to arrive at $\hat{d}_{IC}$ which rests on a number of approximations.

For an EPID which renders portal images which can be used in clinical practice, K must be a function strongly peaked near $(x',y')=(x,y)$. Furthermore, for fluoroscopic EPIDs, the low-amplitude tails of K decay slowly [52] because the optical coupling between screen and mirror takes place over range which is comparable to the size of the screen (see figure 2.1).

Therefore, the on-axis value pixel value $G(0,0,t,l,\Omega)$ as defined by equation (A2.2), will be insensitive to off-axis variations in the spatial distribution of D_p. For this reason, *and* because the contribution of scatter to D_p is relatively small (< 10%) for clinically relevant t, L and Ω values at a FS-isocentre distance of 50 cm [99], we approximate D_p by $D_p = D_p(0,0,t,L,\Omega;d_{IC})\Delta(x,y,\Omega)$, for fields Ω that enclose the (x,y)-origin. The first 'effective transmission' term takes into account the *on-axis* change in primary transmission with t and Ω as well as the change in photon spectrum. The second or 'spatial' term

$$\Delta(x,y,\Omega) \equiv D_p(x,y,t_{ref},L_{ref},\Omega;d_{IC})/D_p(0,0,t_{ref},L_{ref},\Omega;d_{IC})$$

is the relative dose in the detector plane with respect to the beam-axis, and depends explicitly on Ω, e.g. due to the fact that $\Delta \approx 0$ for (x,y) outside Ω.

As t, L, and Ω vary, Φ can only change because of changes in the incident photon spectrum (which includes the spectrum of photon incidence directions). The *change* in beam hardness, relative to the *ref* situation, as t, L or Ω are varied with respect to their *ref* value, will be to first order independent of (x,y) [98]. An additional argument to use this approximation is that, just like for D_p, the integral in equation (A2.2) is insensitive to off-axis variations of Φ. We therefore rewrite Φ as

$$\Phi(x,y,t,L,\Omega;d_{IC}) = \Phi(0,0,t,L,\Omega;d_{IC})\psi(x,y;d_{IC}) \equiv \Phi_c(t,L,\Omega)\psi(x,y;d_{IC})$$

Using the separation of variables described above, we arrive at

$$\frac{G(0,0,t,L,\Omega)}{D_p(0,0,t,L,\Omega;d_{IC})} = \left[\frac{G}{D_p(d_{IC})}\right]_{ref} \Phi_c(t,L,\Omega;d_{IC})I(\Omega) \tag{A2.3}$$

with the definition

$$I(\Omega) = \int_{FS} dx\,dy\, \Delta(x,y,\Omega)\psi(x,y;d_{IC})S(x,y)K(0,0;x,y).$$

Constancy of the left hand side of equation (A2.3) with varying t, L for a fixed Ω implies constancy of Φ_c at that Ω. As a result, changes in the incident MeV photon spectrum with varying t and L have negligible impact on Φ_c. If a build-up $\hat{d}_{IC}$ exists for which this can be achieved for a clinically relevant range of Ω, spectral changes obviously have negligible impact on Φ_c and so we can write in good approximation $\Phi_c(t,L,\Omega;\hat{d}_{IC}) \equiv 1 \,\forall\, (t,L,\Omega)$.

Since the method described by Pasma *et al* [99] to derive $\hat{d}_{IC}$ is based on the minimization of variations in the left hand side of equation (A2.3), they implicitly assumed a separation of the variables (x,y) and (t,L). If, for instance, the convolution of the kernel with variations in the spatial distribution of $D_p\Phi$ with t leads to significant (~ 1%) on-axis changes in G with t, a suitable $\hat{d}_{IC}$ can not be found with their method. Also note that, for the high elbow EPID studied in this paper, the approximations leading to equation (A2.3) are expected to hold better than for low elbow systems due to the small on-axis cross talk in the high elbow.

Substitution of $\Phi_c \equiv 1$ into equation (A2.3) yields equation (2.2), where the fact was used that the explicit reference to t, L, Ω and $\hat{d}_{IC}$ can be dropped from the equation for constant Φ_c. The effective sensitivity S in equation (2.2) is defined by $S(x,y) \equiv \psi(x,y;\hat{d}_{IC})S_r(x,y)$.

3   Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability

J.C. Stroom, J.C.J. de Boer, H. Huizenga and A.G. Visser.

Abstract

Purpose: Following the ICRU-50 recommendations, geometrical uncertainties in tumor position during radiotherapy treatments are generally included in the treatment planning by adding a margin to the clinical target volume (CTV) to yield the planning target volume (PTV). We have developed a method for automatic calculation of this margin.
Methods and materials: Geometrical uncertainties of a specific patient group can normally be characterized by the standard deviation of the distribution of systematic deviations in the patient group (Σ) and by the average standard deviation of the distribution of random deviations (σ). The CTV of a patient to be planned can be represented in a 3D matrix in the treatment room coordinate system with voxel values one inside and zero outside the CTV. Convolution of this matrix with the appropriate probability distributions for translations and rotations yields a matrix with coverage probabilities (CPs) which is defined as the probability for each point to be covered by the CTV. The PTV can then be chosen as a volume corresponding to a certain iso-probability level. Separate calculations are performed for systematic and random deviations. Iso-probability volumes are selected in such a way that a high percentage of the CTV volume (on average > 99%) receives a high dose (> 95%). The consequences of systematic deviations on the dose distribution in the CTV can be estimated by calculation of dose histograms of the CP matrix for systematic deviations, resulting in a so-called dose probability histogram (DPH). A DPH represents the average dose volume histogram for all systematic deviations in the patient group. The consequences of random deviations can be calculated by convolution of the dose distribution with the probability distributions for random deviations. Using the convolved dose matrix in the DPH-calculation yields full information about the influence of geometrical uncertainties on the dose in the CTV.
Results: The model is demonstrated to be fast and accurate for a prostate, cervix, and lung cancer case. A CTV-to-PTV margin size which ensures at least 95% dose to (on average) 99% of the CTV, appears to be equal to about $2\Sigma+0.7\sigma$ for three all cases. Because rotational deviations are included the resulting margins can be anisotropic, as shown for the prostate cancer case. *Conclusion*: A method has been developed for calculation of CTV-to-PTV margins based on the assumption that the CTV should be adequately irradiated with a high probability.

3.1 Introduction

Geometrical uncertainties in radiotherapy treatments cause differences between intended and actually delivered dose distribution in the clinical target volume (CTV), as defined by the ICRU [62]. The uncertainties primarily consist of external set-up deviations and internal organ movement. Both deviations consist of a systematic component, i.e. the same for each fraction of the treatment, as well as a random component, i.e. varying from day to day. The size of the patient set-up deviations can be assessed by comparison of images acquired during the treatment (with megavoltage portal films or electronic portal imaging devices) with those of the intended treatment (simulator radiographs or digitally reconstructed radiographs generated by the planning system). By imaging several patients of a specific patient group regularly, the typical size of the systematic and the random positioning deviations for that group can be determined [15, 35], which may indirectly lead to improved set-up techniques and/or equipment. In principle, systematic deviations of an individual patient can be estimated during the first few fractions and couch corrections can be applied for subsequent irradiations [13, 17, 144]. This so-called "off-line protocol" reduces systematic deviations while random deviations remain unchanged. Both systematic and random set-up deviations can be reduced to negligible values if on-line corrections are applied [91, 130]. In this case, the patient position is verified at each fraction using a small number of monitor units. If necessary, the couch position is adjusted before the remaining dose is given. At the moment however, on-line correction procedures are too time consuming to be routinely used in clinical practice.

Internal organ motion is the movement of an organ relative to the bony structures. For instance, the prostate can move due to variations in bladder- and rectum filling. These movements can not be assessed directly by portal imaging since the tumor is generally not visible. By implantation of radio-opaque markers in or near the CTV the internal organ motion can be visualized which enables on-line positioning corrections [10, 36]. In other clinical studies repeated CT-scans have been acquired to get an indication of internal prostate movements [54, 109]. Intra-fraction movement of the tumor will add to the random deviation. Due to breathing and cardiac motion a tumor in thorax or abdomen can vary significantly in position in a matter of seconds [110]. Complex techniques like real-time couch movement, respiration gated irradiation or breathing control might limit the consequences of this variation [70, 88, 141].

Whatever is done to minimize the geometrical uncertainties, to some extent inaccuracies are unavoidable. Once the typical values for a specific group of patients are known they should be included in the treatment planning for individual patients from that group. Patient set-up deviations not only affect the dose in the tumor region, but in neighboring, possibly critical, organs as well. For random deviations, the effect of this deviation can be simulated by a convolution of the dose distributions with the distribution of movements in three dimensions. Several groups have implemented this for translational deviations [61, 111]. Systematic deviations are more of a problem since they are a priori not known for a specific patient and only the distribution of systematic deviations for the patient group can be determined. The effect of systematic deviations on the dose distribution is more significant than that of random deviations, hence relatively small systematic deviations should not be ignored. One possible way to deal with systematic deviations has been proposed by Goitein [46]. He suggested three parallel planning calculations, one with nominal set-up deviations and the others with extreme values, and allow only those plans for which all three dose distributions are acceptable in terms of tumor coverage and critical organ sparing. Recently this idea has been further developed by Mageras and colleagues, especially to include

internal organ motion in radiotherapy planning [80]. Killoran and colleagues simulate systematic and random uncertainties simultaneously by multiple Monte Carlo simulations which result in multiple dose volume histograms (DVHs) that are used for evaluation of the treatment plan [66].

The above techniques operate directly on the CTV and are more sophisticated than the conventional approach as proposed by the ICRU [62], i.e. using a planning target volume (PTV) which is defined as the CTV plus margins for all geometrical uncertainties. However, the practical application of the concept of PTV is not always clear. First of all, it is rather cumbersome to manually draw margins in three dimensions around an irregularly shaped tumor volume [128]. Therefore several groups have developed algorithms for automatic margin calculation, either multiple 2D [7] or fully 3D [18, 131]. Furthermore, the geometrical uncertainties can originate from rotations as well as translations. Rotational deviations will yield anisotropic margins, i.e. the size of the margin will vary depending on the position with respect to the axis of rotation. None of the aforementioned algorithms have incorporated this. Finally the exact margin size necessary to ascertain adequate coverage of the CTV depends on the kind of deviation (systematic or random) and on the dose distribution. How this must be taken into account has up till now not been specified.

We have developed a model that calculates the CTV-to-PTV margins step by step, based on clinically measured CTV position deviations and on the requirement that the dose distribution delivered to the CTV will satisfy the ICRU recommendations for dose homogeneity with a high probability. Internal organ motion as well as external set-up deviations, translations as well as rotations, and systematic as well as random deviations are included in the model. Once the dose distribution has been planned around the resulting PTV, the same algorithms can be used to calculate the influence of all geometrical uncertainties on the dose in the CTV and hence to verify the planned dose distribution. The use of the model will be demonstrated for a prostate, cervix, and lung cancer planning.

3.2 Methods and materials

3.2.1 Parameters required

Since the geometrical uncertainties of an individual patient which is to be planned are not known, measured data of a group of similar patients must be used. Clinical studies performed in our department and elsewhere have shown that translational deviations in patient positioning of a specific group can be approximated by normal distributions of systematic and random deviations in the three main directions, e.g. [15, 35]. For each patient in the study the average set-up deviation and the standard deviation (SD) of the distribution around that average is determined. The random variation σ characterizing a certain patient group is then defined as the SD of the day-to-day set-up positions, averaged over all patients in the group. The systematic variation Σ is defined as the SD of the distribution of average set-up deviations per patient in the group of patients. The overall mean deviation M is the average value over all fractions and all patients. If the reference set-up (during simulation) can be considered as a sample from the random distribution, M will be close to zero and Σ will be close to σ. Rotational deviations around the three main axes can in principle be described similarly. The study of internal organ motion can yield random and systematic deviations for translations and rotations as well. Hence a set of twelve (x/y/z * rot/trans * int/ext) standard deviations Σ and σ and six rotation axes are

necessary to describe all geometrical uncertainties of a specific patient group. These values are the input parameters of our method and do not only depend on the tumor sites, but also on set-up techniques and treatment protocols (and possibly even on more specific variables like tumor stage, patient weight, accelerator etc.).

The parameters are used to calculate margins around a CTV which is initially represented by a set of input contours as outlined in 2D CT-slices. The contour data points are used to determine a 3D volume in a cubic calculation grid in the treatment room (i.e. CT) coordinate system. The algorithm has been described before [131] and can be summarized as follows. For each slice, intersection points of contour lines with 2D grid lines are calculated to fill a 2D matrix with values equal to the fraction of the grid element that is enclosed by the contour. The slices are stacked with increasing slice position so that a 3D matrix $M_{CTV}(x,y,z)$ is created with values 1 inside the volume and 0 outside the volume (and between 0 and 1 at the edge). This matrix is used for subsequent calculations.

3.2.2 Coverage probability calculation using convolutions

Two separate, equivalent methods have been developed to calculate a matrix with coverage probabilities (CPs) which is defined as the probability for each point to be covered by the CTV and which will be used for PTV margin determination. They will be designated as the convolution method and the Monte Carlo method. The *convolution method* uses a straight-forward convolution of the CTV matrix with normal distributions describing the geometrical uncertainties. The effect of translational and rotational deviations is calculated separately. The normal distribution of mutually independent deviations in translations is given by:

$$N(x,y,z) = \frac{e^{-\frac{1}{2}\cdot\left(\left(\frac{x}{sd_x}\right)^2 + \left(\frac{y}{sd_y}\right)^2 + \left(\frac{z}{sd_z}\right)^2\right)}}{\sqrt{8\pi^3}\, sd_x \cdot sd_y \cdot sd_z} \tag{3.1}$$

with sd_x, sd_y, and sd_z the standard deviations of the distributions in the three main directions. The input matrix $M_{CTV}(x,y,z)$ is convolved numerically with the probability distribution:

$$M_{CP}(x,y,z) = \sum_{x'}\sum_{y'}\sum_{z'} M_{CTV}(x',y',z')\, N(x-x',y-y',z-z')\, \Delta x\, \Delta y\, \Delta z \tag{3.2}$$

with $\Delta x \Delta y \Delta z$ the voxel size. In the output matrix $M_{CP}(x,y,z)$ the original contents of the voxels is spread out according to the distribution of translations. The value in each voxel of this "coverage probability" matrix represents the probability of the voxel being covered by the CTV.

Inclusion of rotational positioning deviations is cumbersome in the orthogonal coordinate system. To be able to handle rotations around axes in the three main directions, an input matrix $M_{CTV}(x,y,z)$ is transformed to cylindrical coordinates $M_{CTV}(r,\theta,a)$ using bilinear interpolations, where a is x, y or z for rotations about x-, y-, or z-axis respectively. The center of the new matrix ($r = 0$) is taken to be at the (user-defined) position of the rotation axis. A one-dimensional distribution matrix $N(\theta)$ is defined similar to $N(x,y,z)$ and the convolution is performed:

$$M_{CP}(r,\theta,a) = \sum_{\theta'} M_{CTV}(r,\theta',a)\, N(\theta-\theta')\, \Delta\vartheta \tag{3.3}$$

with $\Delta\theta$ the bin size. Subsequently the output matrix $M_{CP}(r,\theta,a)$ is transformed back to orthogonal coordinates $M_{CP}(x,y,z)$. The transformation to a different coordinate system and back, on a discrete grid, will already smooth the CTV matrix, even without convolutions. To limit this effect, the pixel size in the cylindrical coordinate system is kept four times smaller than in the original. To minimize the number of cylindrical matrix elements and keep the element size approximately equal for all values of r, the number of angles θ increases with r, i.e. the (r,θ)-calculation grid is not rectangular but rather triangular in shape.

In case more than one rotational deviation is present, the above procedure is performed again using the output matrix of the first calculation as input for the second, etc. The order is arbitrary for distributions that are mutually independent and for the small rotational deviations ($< 10°$) that are usual in patient set-up and organ movement (as will be justified later on, see figure 3.4). In case both rotational and translational deviations must be considered, the rotations should be performed first.

3.2.3   Coverage probability calculation using a Monte Carlo approach

The *Monte Carlo method* simulates the fractionated radiotherapy treatment more directly by sampling the translations t_x, t_y, t_z and rotations r_x, r_y, r_z from their respective distributions. Subsequently all points in the CTV matrix (x,y,z) are moved to a new position (x',y',z') according to:

$$x' = x + t_x + x_{r,y}(\cos(r_y)-1) + x_{r,z}(\cos(r_z)-1) - y_{r,z}\sin(r_z) + z_{r,y}\sin(r_y)$$

$$y' = y + t_y + y_{r,z}(\cos(r_z)-1) + y_{r,x}(\cos(r_x)-1) - z_{r,x}\sin(r_x) + x_{r,z}\sin(r_z) \quad (3.4)$$

$$z' = z + t_z + z_{r,x}(\cos(r_x)-1) + z_{r,y}(\cos(r_y)-1) - x_{r,y}\sin(r_y) + y_{r,x}\sin(r_x)$$

where $x_{r,y} = (x-R_{x,y})$, $x_{r,z} = (x-R_{x,z})$, etc. the coordinates of the matrix point with respect to the x-, y-, and z-axes of rotation, respectively $(R_{y,x},R_{z,x})$, $(R_{x,y},R_{z,y})$, $(R_{x,z},R_{y,z})$. The input value at point (x,y,z) can be assigned to the eight grid points nearest to (x',y',z') using trilinear interpolations or assigned completely to the one nearest point which is faster but less accurate. Hence a new matrix $M_n(x,y,z)$ is calculated with n being the sample number. The procedure is repeated many times (> 1000) and the resulting matrices $M_n(x,y,z)$ are averaged over the number of samples to yield the final result $M_{CP}(x,y,z)$.

A difference between Monte Carlo and convolution method is the way in which rotational deviations are incorporated. The Monte Carlo method handles all rotations at once. The voxel displacement resulting from each rotation is calculated assuming the same starting position for all rotations, i.e. the input voxel position for the second rotation is not the output from the first. The computation of translations is similar for both methods; all three directions are handled simultaneously. Comparison of the methods will give an indication of their accuracy.

3.2.4   Interpretation of coverage probability

As mentioned before, the elements in the CP matrix M_{CP} (of which the values vary from 0 to 1) represent the probability of a fixed point in space to be actually covered by the CTV. For sufficiently large volumes (larger in diameter than about 2 SD of the distribution of

deviations), the CP value of a certain voxel also represents the probability that the volume border lies outside that voxel; i.e. if a point is covered by some part of the CTV, the CTV border must lie outside of that point (or exactly on it). Iso-coverage probability volumes are therefore logical candidates to define the PTV. Considering translational deviations in one dimension and given a selected coverage probability CP_s smaller than 0.5, the probability (P_o) that any point of the CTV volume is outside the PTV will then be equal to 2 CP_s. For translations in more dimensions and for rotations the relationship between CP_s or the margin size and P_o becomes less straightforward. A special case may however serve as an estimate for the general case.

That special case is a spherical CTV with an isotropic margin m to represent the PTV. The probability P_o that any point of a CTV with normally distributed translational deviations sd in all three dimensions will be outside the PTV can be calculated analytically:

$$P_o = \frac{2}{\sqrt{2\pi}} \int\limits_{r > m/sd} r^2 \, e^{-\frac{1}{2}r^2} \, dr \qquad (3.5)$$

with r being equal to $\sqrt{(x^2 + y^2 + z^2)}$. The integral can be solved using partial integration. For deviations in two dimensions a similar expression can be derived. In figure 3.1 P_o is displayed as a function of the margin in units of sd, for the one-, two-, and three-dimensional case. The coverage probability values CP_s that should be selected to obtain a margin $m(sd)$ are displayed along the upper horizontal axis. The relation between CP_s and $m(sd)$ is normally independent of CTV shape as long as the volume is sufficiently large (> 2 sd, as indicated before). For 3D translational deviations, the probability of the CTV being partly outside the PTV is considerably higher than for 1D deviations. For instance, a PTV margin of 1 sd (or CP_s = 16%) has a P_o-value of 32% for deviations in one dimension as opposed to 80% when deviations in all three dimensions are considered. However, not only the probability of CTV miss but also the extent of that miss is important. Even for large P_o-values, the fraction of the CTV outside the PTV is small and the dosimetrical consequences are limited.
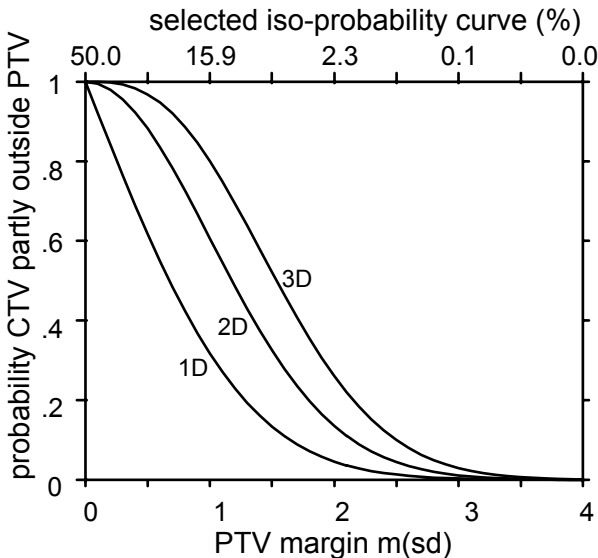
Figure 3.1 Theoretical curves for a spherical CTV with variations in translations with standard deviations sd. Rotational variations are zero. Indicated is the probability that the CTV is partly outside the PTV (P_o) for variations in one, two, and three dimensions with varying CTV-to-PTV margin size in units of sd. The corresponding selected iso-probability values (CP_s) are indicated on the upper horizontal axis.

3.2.5 Influence of systematic deviations on the CTV dose: dose probability histograms

To investigate the effect of the geometrical uncertainties on the dose distribution in the CTV, different approaches for systematic and random deviations are required. Systematic geometrical misses will cause underdosage of the same part of the CTV for every fraction of the treatment, whereas random deviations will cause underdosage in different parts of the CTV for each fraction. For *systematic* deviations, the CP matrix can be used to estimate the influence of systematic deviations on the DVH of the CTV. A normal cumulative DVH is constructed by summation of all CTV voxels that receive more than a certain dose, for all dose values D:

$$V(D) = \sum_{\vec{r} \in CTV} S_D(D'(\vec{r}))\Delta V = \Delta V \sum_{\vec{r} \in \Re_3} S_D(D'(\vec{r}))M_{CTV}(\vec{r})$$

$$\text{with } S_D(D'(\vec{r})) = \begin{array}{l} 1 \text{ for } D'(\vec{r}) \geq D \\ 0 \text{ for } D'(\vec{r}) < D \end{array} \tag{3.6}$$

with $D'(\vec{r})$ the dose value at position $\vec{r}$, ΔV the voxel volume, and $M_{CTV}(\vec{r})$ the previously defined CTV matrix. $\vec{r} \in CTV$ and $\vec{r} \in R_3$ are those positions which are member of the CTV and the whole 3D space, respectively. At the time of planning the systematical deviations in CTV position are unknown for a specific patient but the probability of the CTV being systematically at a different position is determined by the systematic variations Σ of translations and rotations for the patient group. The average DVH, taking all systematic deviations of the CTV position with respect to the dose distribution into account, can then be determined to be:

$$\langle V(D) \rangle = \Delta V \sum_{\vec{t} \in \Re_6} N(\vec{t}) \sum_{\vec{r} \in \Re_3} S_D(D'(\vec{r}))M_{\vec{t}(CTV)}(\vec{r})$$

$$= \Delta V \sum_{\vec{r} \in \Re_3} S_D(D'(\vec{r})) \sum_{\vec{t} \in \Re_6} N(\vec{t})M_{\vec{t}(CTV)}(\vec{r}) \tag{3.7}$$

$$= \Delta V \sum_{\vec{r} \in \Re_3} S_D(D'(\vec{r})) \, M_{CP}(\vec{r})$$

with $\vec{t}$ being a transformation (translations and rotations), $N(\vec{t})$ the probability for a certain transformation $\vec{t}$ (see Eq. 3.1), and $M_{\vec{t}(CTV)}$ the transformed CTV. In the last step of the derivation Eq. 3.2 is used. Hence instead of counting voxels receiving a dose $\geq D$, as for a normal DVH, the average of the DVH for all systematic deviations of the CTV is obtained by summation of the coverage probability values for each dose $\geq D$. Therefore the results of those calculations will be denoted as dose probability histograms (DPHs). It should be emphasized that the CP matrix M_{CP} in Eq. 3.7 is calculated using the *systematic* variations Σ. A dose histogram of a CP matrix for random (day-to-day) variations of one patient has no physical meaning since DVHs of different fractions should not be added; information about the position of the dose, which is essential when adding dose distributions of different days, is lost in a DVH.

In figure 3.2 a clinical example of DPH calculations is shown. For an arbitrary prostate cancer patient planned according to the ICRU dose specification rules (i.e. block margins are such that the 95% isodose volume encloses the PTV), DPHs of CTV have been calculated for six different 3D translational variations Σ. Naturally, the probability of CTV underdosage increases with increasing variations. Since the ICRU suggested a maximal

tumor underdosage of 5% [62], an additional dashed line is drawn to indicate the 95% dose. The DPH curves for the CTV now immediately indicate that the 95% isodose will enclose on average a large part of the CTV (> about 99%) as long as the CTV-PTV margin is at least 2Σ. This implies a high probability of the 95% isodose enclosing the whole of the PTV. For smaller margins there will be an increasing probability of underdosage of the CTV. The DPH as an average of the DVHs for all systematic deviations seems therefore a reliable tool to determine the margin size required to cover for these deviations for a specific treatment plan. Also indicated in figure 3.2 are similar curves for variation in rectum position. The same variations are assumed without any change in rectum shape. As expected, the maximum dose of the DPH increases as the variations increase; the probability that the rectum will be part of the higher dose regions will be higher. At the same time however, the lower dose volumes decrease with increasing variations. This is due to the fact that for isotropic movements in all directions, the probability of the rectum moving toward the higher dose regions is lower than for moving away from them.

3.2.6 Influence of random deviations on the CTV dose

The *random* deviations displace the CTV with respect to the dose distribution differently for each fraction of the treatment. This can be simulated by convolution of the dose distribution matrix with the probability distributions, as has been described before [14, 61, 111, 132]. The same algorithms as for the CP calculations can be applied. If the input file in Eq. 3.2 (M_{CTV}) is a dose distribution instead of the CTV, the output will be a dose distribution which is spread out locally as a result of random deviations. This distribution is the best estimate of the actually delivered distribution during the radiotherapy treatment. In general the higher iso-dose regions will decrease in size, while the lower iso-dose regions will increase. The extent of shrinkage of the 95% iso-dose volumes should give an indication for the size of the required margin [14]. If only the random *set-up* deviations are considered, one calculation will be sufficient to determine its effect on CTV and critical organs simultaneously. If the internal random deviations of the critical organs are different for different organs, separate convolutions of the combined distributions of random set-up and organ movement should be performed with the local dose matrix surrounding each organ. Compared to CP calculations, dose modifications are more accurate because gradients in dose matrices are considerably flatter than those in CTV matrices.

3.2.7 PTV margin determination

The goal of a PTV is to create a volume around which the 95% dose can be planned so that the CTV is adequately irradiated, which can be verified using DPHs (see next section). In principle one might find the correct PTV by trial and error but CP calculations can also be used to calculate a good PTV to start with. Because of the different effect of the systematic and random deviations on the CTV dose (Σ and σ cannot be added), the PTV is calculated in two steps.

For *systematic* geometrical uncertainties a high irradiation probability is obtained by choosing the margin according to a low iso-probability volume. From figure 3.2 a margin equal to about 2Σ would seem reasonable and from the upper horizontal axis in figure 3.1 can be deduced that this corresponds to iso-probability curves of about 2.5%. Hence the first step in the PTV calculation (PTV1) is the determination of the 2.5% iso-probability

volume of the CTV (i.e. the volume bounded by voxels having a CP equal to 2.5%), using the systematic variations Σ as input to the model. From figure 3.1 this implies that for a spherically symmetric situation and deviations in all directions, there still is at least a 28% chance of the a part of the CTV is outside PTV1 for all fractions. In other words, there is maximally a 72% probability of complete enclosure of the CTV by the PTV1 during the radiotherapy treatment. However, the DPH curves of the example in figure 3.2 indicate that the 95% isodose will enclose the CTV in practically all cases. The internal and external deviations can be handled simultaneously by adding the respective standard deviations in quadrature.

To cover for the remaining *random* deviations only a moderate increase in the margin will be necessary. The total random variation equals the quadratic sum of in- and external random variations. The procedure is similar to that for the systematic deviations, this time using PTV1 instead of the CTV as input volume. The difference is that random deviations do not affect the dose distribution in the tumor as much as systematic deviations. Therefore the margins can be smaller and the selected iso-probability volume can be higher. Bel *et al*. studied the dosimetric consequences of random translational variations σ for their prostate patients and concluded that a 0.7σ margin would be sufficient to keep the minimal CTV dose above 95% [14]. Based on those results we considered the 25% iso-probability volume appropriate for random deviations. This means that PTV1 will be partly outside the PTV in over 90% of the fractions (see figure 3.1). Once this final margin is added to PTV1, the calculation of the PTV is complete. Naturally, the choice of the iso-probability volume for either PTV1 or PTV margins is to some extent arbitrary and can be varied according to the individual preferences of the clinician. Eventually it should be based on quantitative models for the tumor control probability.
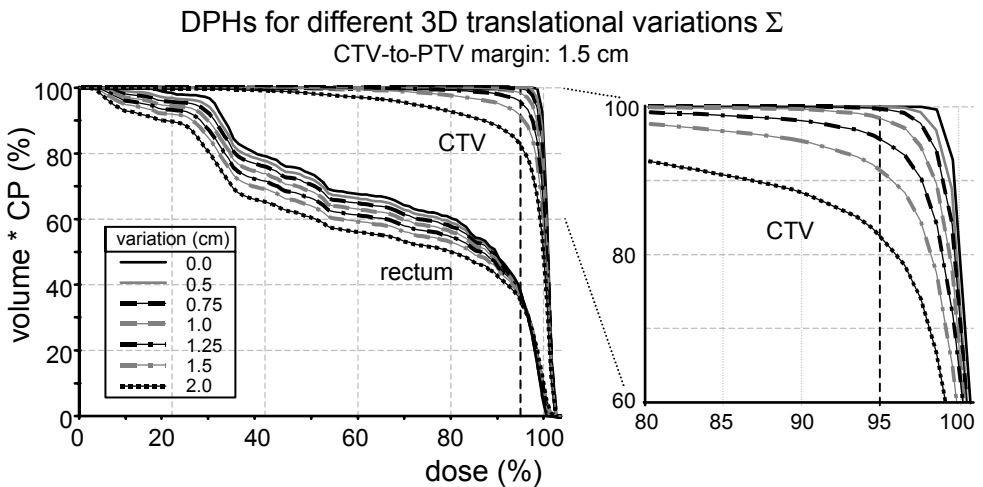
Figure 3.2 Example of dose probability histograms for six different systematic variations in translation (Σ) for a prostate case planned to conform to ICRU recommendations with a 1.5 cm CTV-to-PTV margin. With increasing variation in CTV position there will be a higher probability of underdosage. The rectum will on average also receive less dose as the variations increase, although the maximum dose increases. On the right the CTV curves near the 95% dose are magnified.

3.2.8   PTV margin verification

Normally, the patient will be planned with certain block margins around the PTV to account for the penumbra of the beam. The concepts described can be applied to evaluate and judge the effect of geometrical uncertainties on the dose distribution in the CTV. Firstly, to calculate the expected dose distribution actually delivered during the treatment series, the dose matrix is convolved with the distribution of random deviations. The 95% isodose volume will shrink but should still enclose the PTV1 volume. Secondly the systematic deviations are used to determine the DPH of the CTV (and possibly critical organs), using the dose modified for random deviations. This DPH indicates whether the dose distribution is adequate to irradiate the CTV, given the systematic ánd random deviations. If the patient has been planned correctly, the average CTV volume receiving > 95% dose must be high (e.g. > 99%). In case the probability of underdosage is too high, the 95% isodose volume is too tight around the CTV and either the PTV margin or the block margin is too small. The plan can be recalculated using lower iso-probability contours as PTV margin or larger block margins. In case the DPH is practically equal to the original DVH (i.e. systematic deviations have no influence on the dose in the CTV), the PTV margins might be too large and the plan should be recalculated using higher iso-probability values as PTV margins (or smaller block margins). In this iterative manner the size of the block-to-CTV margins is directly optimized for irradiation of a specific CTV and DPHs of CTV instead of DVHs of PTV are used to evaluate the planning.

3.2.9   Hardware

The method has been implemented using the C programming language on a HP 9000/712 (100 MHz) workstation. It runs as a separate application next to the CadPlan planning system[1] and hence uses CadPlan contour and dose files as input. The results are written back to CadPlan files for visualization and further planning. At the moment, the voxel sizes in the calculations are equal to those used in the planning system, which vary in practice from about 2 mm for CT pixel size to maximally 10 mm for the slice distances. The dose matrices normally have a resolution of 2.5 or 5 mm. Geometrical uncertainties smaller than about 1 mm (1 SD) are consequently not reflected in the calculations.

3.3 Results

3.3.1   Accuracy of the methods

To get an indication of the accuracy of the convolution and Monte Carlo method, results of the two methods are compared for a schematic example. Figure 3.3a depicts a 2D geometrical object which represents a target volume. The simulated random translational variations are different in horizontal and vertical direction and the coverage probabilities are calculated for both methods. The number of runs in the Monte Carlo method was 1500. The iso-probability contours of the different methods coincide well, only the 1% iso-probability curves deviate slightly. In figure 3.3b an additional variation in rotation is simulated. As expected, the area within the higher probability curves decreases while that in

---

[1] Varian-Dosetek

the lower increases. Due to the rotations the distance from one curve to the next becomes anisotropic. Near the upper right part of the object the lower iso-probability curves actually shrink compared to the curves in figure 3.3a; because of the rotations there is a lower probability that voxels in that area are enclosed by the object. The differences between the two methods become slightly larger, especially near the low probability curves. This is attributed to the limited number of runs of the Monte Carlo method which yields poor statistics in those regions.

The largest differences between the two CP calculation methods are expected for coverage probability calculations with large rotational deviations. Hence an exaggerated (and unrealistic) variation in rotations around the three orthogonal axes through a point in the lower left and cranial corner of the input matrix is calculated for a lung tumor CTV in the upper thorax. Figure 3.4 shows the 5% iso-probability curves for both methods in a transversal, sagittal, and frontal slice near the center of the CTV. The two contours do not overlap in all areas but the differences are small: maximally 2 mm (1 CT pixel) while the margin is on average 2 cm. In slices near the edges of the volume the maximum differences that are found are 4 mm which is adequate considering the irregularity of the Monte Carlo contours. The difference between the two 5% iso-probability volumes is 4 cc (1%). In clinical practice rotational deviations of extreme values are rare and the occurrence of more than one deviation simultaneously is unlikely. Consequently, this close agreement may serve as assurance that the calculation of rotational deviations is reliable for both methods.
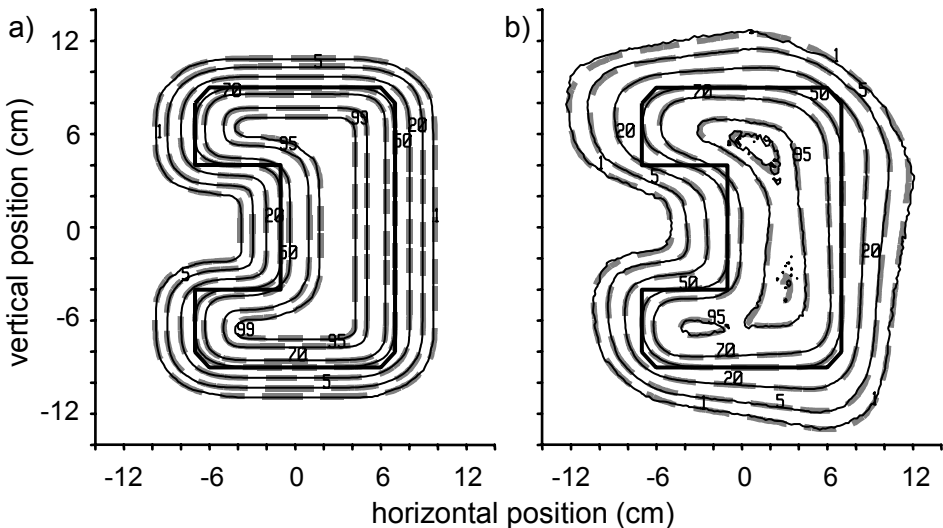
Figure 3.3 Comparison of convolution and Monte Carlo method. The schematic 2D CTV is the object outlined by the thick solid contour, the pixel size is 2 by 2 mm. A) A horizontal 12 mm (1 SD) and vertical 8 mm variation in translation has been simulated. The 1, 5, 20, 50, 70, 95, and 99% iso-probability curves of both methods are shown; the thin black lines represent the result of the Monte Carlo method (1500 runs), the thick grey dashes the convolution method. B) A 5 degrees (1 SD) variation in rotation is added. The rotation axis is at the lower left corner of the figure (-12, -12 cm). The iso-probability curves of the two methods still coincide well.
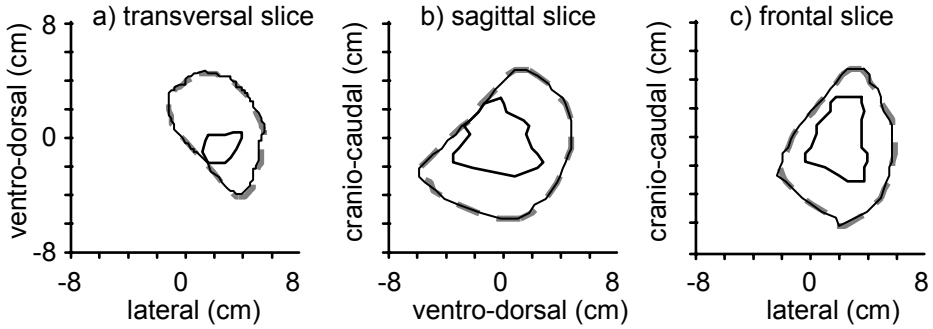
Figure 3.4 Comparison of the two methods for an extreme case of rotational deviations. A lung CTV (thick black lines) has been modified with rotational deviations of 10, 8, and 6 degrees (1 SD) around the three major axes in a point about 8 cm left, cranial, and below the CTV center. The 5% iso-probability curves for the convolution (thick grey dashes) and Monte Carlo method (thin black lines) are displayed in a transversal (a), sagittal (b), and frontal cut (c) through the CTV. Although the two methods handle the effect of rotational deviations differently, the results are quite similar.

Although Eq. 3.5 was derived for a spherically symmetric case, the results are a reasonable approximation for more general clinical situations. For translational deviations, computer simulations for a prostate CTV indicate that, even if the margins become anisotropic, the deviations from the values in figure 3.1 are small (< 5%). A restriction is that the CTV volume should be devoid of sharp edges, which is usually satisfied. Another limitation is that rotational deviations are not included in the calculations, hence Eq. 3.5 underestimates the P_o-value if rotations are present.

3.3.2   Clinical examples

In three clinical examples the stepwise PTV calculation and the verification procedure will be illustrated. The values for the geometrical uncertainties are taken from literature or have been measured in our institute. In all cases it is assumed that the application of an off-line protocol (as mentioned in the introduction) halves the systematic translational set-up deviations [13]. The values used are summarized in Table 3.1. All calculations are done with the convolution method and all rotation axes are through the center of the CTV unless specifically stated otherwise. CT slice distances are 5 mm and CT pixel size are about 2 x 2 mm. The grid size of the dose matrices is in all cases 2.5 mm in the plane of the CT slices and equal to the CT slice distance in the direction perpendicular to the CT planes. All three patients were planned with multi-leaf collimators. Block margins were such that the 95% isodose closely fitted the PTV and the dose homogeneity in the PTV satisfied the ICRU 50 recommendations (i.e. dose variation of maximally 95% - 107%).

3.3.3   Prostate cancer case

Due to variations in rectum and bladder filling, the range of the internal prostate movements is considerable. The values that are used in our calculations have been estimated from several studies available in the literature [10, 36, 54]. The lateral rotation axis was taken to be near the apex of the prostate as suggested by van Herk *et al*. [54]. The values for external set-up deviations are taken from routine portal imaging data of 228 patients treated at our clinic. The final PTV is constructed in several steps as shown in transversal and sagittal slices through the tumor (figure 3.5). The CTV has been outlined manually by a radiation oncologist. This volume is expanded with a margin to cover all systematic deviations; a CP matrix is calculated using the quadratically summed internal and external systematic variations (Table 3.1). PTV1 is taken to be the volume enclosed by the 2.5% iso-probability

variation	translation (mm, 1 SD)			rotation (degrees, 1 SD) /axis		
direction	LR	CC	AP	LR	CC	AP
Prostate						
σ_{ext}	2.0	2.0	2.0	1	1	1
Σ_{ext}	1.2	1.2	1.4	1	1	1
σ_{int}	1	2	2	4[*]	2	1
Σ_{int}	1	2	2	4[*]	2	1
Cervix						
σ_{ext}	3.5	4.0	3.5	3	1.5	2
Σ_{ext}	2.0	2.5	2.0	2	1	1
σ_{int}	1	1	1	1	1	1
Σ_{int}	1	1	1	1	1	1
Lung						
σ_{ext}	3.0	3.0	3.0	2	2	2
Σ_{ext}	2.0	2.0	2.0	1	1	1
σ_{int}	4	5	5	2	2	2
Σ_{int}	2	3	3	1	1	1

Table 3.1 Overview of geometrical uncertainties used as input for the model for three different tumor sites. Systematical (Σ) and random (σ) variations are indicated for translations as well as rotations and internal organ movement (int) as well as external set-up deviations (ext). All rotation axes are assumed to be at the center of the CTV except for those indicated by * which are assumed to be at the caudal apex of the CTV.

contours. To get the final PTV a subsequent margin is added to PTV1 from the calculated 25% iso-probability volume with the quadratically summed random variations as input for the CP calculations. The random deviations add only an extra 1 - 2 mm (which is close to the pixel size). In total, the margin around the CTV varies from minimally 6 in the caudal to maximally 13 mm in the cranial region of the PTV. The anisotropy is due to the significant rotation around the apex of the prostate.

    The interpretation of the contours in the two-dimensional slices is sometimes misleading due to the 3D aspect of the margins; they may appear too large in one slice due to the influence of a differently shaped tumor contour in the next slice as, is especially apparent in the (cranial) transversal slice. This effect is also visible in the cervix and lung cancer case that follow.

Figure 3.5 Prostate planning example of the stepwise PTV calculation in a cranial transversal slice (a), an isocentric slice (b), a caudal slice (c), and a sagittal reconstruction (d). The geometrical uncertainties are from Table 3.1. The CTV (grey curve) is extended with a margin to cover for the systematic deviations by selection of the 2.5% iso-probability volume of the CP matrix (PTV1, dashed curve). The final PTV (white curve) results from the 25% iso-probability volume of the CP matrix of random deviations applied to the PTV1. Especially in the sagittal slice the influence of the rotations around the apex is clear; the PTV margin in the caudal part is significantly smaller than that in the cranial part of the prostate. The critical organs (bladder and rectum) are depicted by the thin black curves.

A three-field technique was applied to plan the patient. To verify the dose distribution in CTV, rectum, and bladder, DVHs have been calculated for each volume and are shown in figure 3.6. Subsequently, the dose distributions around the volumes of interest have been convolved with the distributions of random deviations. For the CTV the values are directly obtained from Table 3.1, for bladder and rectum the internal random motion was estimated to be equal to that of the prostate without the rotation. Bladder and rectum volumes were assumed to be constant. Resulting DVHs show that the effect of random deviations on the CTV dose is negligible, whereas the bladder and rectum high dose volumes are somewhat reduced. Finally, the DPHs for the systematic deviations have been determined using the dose distributions modified for random uncertainties. Standard deviations are taken from Table 3.1 similar to the random deviations. The DPH of the CTV is different from the
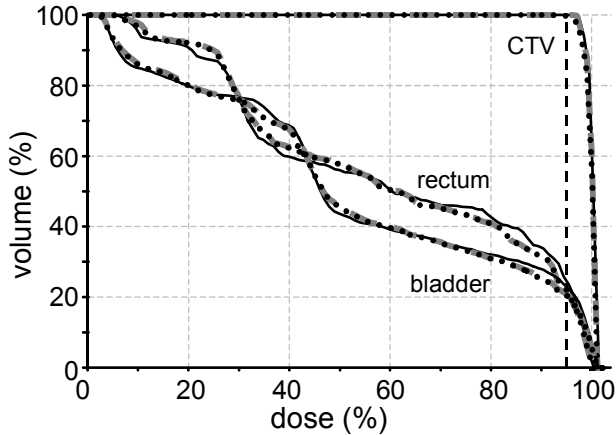
Figure 3.6 DVHs and DPHs for the prostate cancer patient in figure 3.5. The original DVHs of CTV, bladder and rectum are indicated by the thin black lines. DVHs for the dose distributions adjusted for random deviations are represented by the dashed grey curves. The effect of random deviations on the CTV dose is negligible, whereas the bladder and rectum high dose volumes decrease. The DPHs (dotted curves) for the systematic deviations have been determined using the dose distributions modified for random uncertainties. There is only a noticeable effect on the CTV but the curve does not intersect the 95% dose line (dashed) so the probability of underdosage of the CTV is negligible.

original DVH but the DPH-curve does not quite intersect the 95% dose line, i.e. the probability of underdosage of the CTV is negligible. The DPHs for the critical organs do not deviate from the DVHs.

3.3.4   Cervix cancer case

Compared to prostate cancer patients, external set-up deviations play a major role in the planning of (post-operative) gynecological cancer patients. Positioning accuracies of cervix cancer patients are also determined by studies in our own institute, one of which is described by Creutzberg *et al*. [35]. Also in contrast with the prostate case, the internal organ movement is expected to be relatively small considering the involved anatomy. In figure 3.7 four slices through the initially drawn CTV are depicted. The PTV1 margins due to systematic deviations are calculated similar to the prostate case and vary from 6 to 9 mm. They are mainly caused by the external deviations. The addition of a margin for random deviations completes the PTV calculation. The final CTV-to-PTV margin is then about 1 cm. An additional feature of margin calculation using coverage probabilities is the smoothness of the PTV surface. This especially manifests itself in the sagittal view in figure 3.7d; the inconsistencies in the CTV delineation disappear in the PTV.

   The patient was planned in prone position with a three field technique. The relatively large random positioning deviations are solely responsible for the random deviations. DVHs and DPHs for CTV, small bowel, and rectum have been calculated similar to the prostate case. The results are shown in figure 3.8. The effect of the random uncertainties on

Figure 3.7 Example of the PTV calculation for a cervix cancer plan. The geometrical uncertainties are from Table 3.1. The same procedure and line styles as for the prostate case (figure 3.5) are used except for the small bowel that is depicted instead of the bladder. Three transversal (a,b,c) and a sagittal slice (d) through the PTV are shown. The overall CTV-to-PTV margin becomes about 1 cm.

the rectum and small bowel is again a noticeable reduction of the high dose volume. The original DVH of the CTV indicates a less homogeneous dose distribution than for the prostate case. It is however hardly influenced by the systematic or random deviations and the final DPH proves that adequate margins have been applied.

3.3.5 Lung cancer case

A last example is a 3-field lung booster plan designed to spare the left lung (figure 3.9). Due to breathing and cardiac motion there is considerable internal tumor movement the magnitude of which has been estimated from previously published values [110]. Values for external set-up deviations are based on the preliminary results of a lung cancer patient positioning study recently conducted in our own institute. The manually outlined CTV is expanded with a 6 - 9 mm margin to cover internal and external systematic deviations. The random deviations require an additional 3 - 5 mm margin for the final PTV and the total margin becomes 10 - 13 mm.

Variations in position of the lungs and the spinal cord are assumed to consist only of set-up deviations. In the DVHs of CTV and right lung the random-deviation-adjusted dose
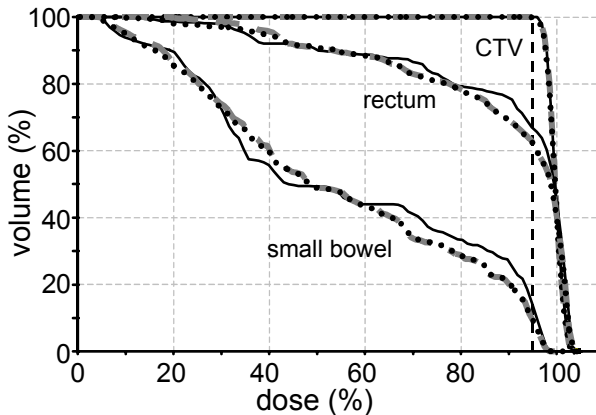
Figure 3.8 Dose distribution verification for the gynecological cancer patient in figure 3.7. DVHs and DPHs for CTV, small bowel, and rectum have been calculated similar to the prostate case of figure 3.6. The effect of the random uncertainties on the rectum and small bowel (dashed grey curve) is again a significant reduction of the high dose volume compared to the original DVH (thin black curve). The DPH curves (dotted) indicate that the CTV is hardly influenced by the systematic or random deviations.

distributions display similar differences with the original as the cervix cancer case (figure 3.10). However, the spinal cord is an exception. Since the beams in the three field plan are positioned (too) close to the spinal cord, there is a strong dose gradient just next to it causing an increase of spinal cord dose in the random-deviation-adjusted plan. To a lesser degree the same holds for the left lung. The DPHs for the critical organs are again practically equal to the DVHs, except for the spinal cord; due to the position of the beams the systematic movements of the spinal cord cause a (slight) increase in average spinal cord volume receiving high dose. The DPH of the CTV is different from the DVHs but the high probability of sufficient dose homogeneity indicates that the PTV and block margins were adequate.

3.3.6   Computer performance

The calculation speed depends on the selected method and input. The margin and DPH calculations normally take about 1-2 minutes with the convolution method. Generally the Monte Carlo method is about three to six times slower since a large number (> 1000) of samples have to be taken to obtain sufficient accuracy in the low probability regions. Therefore the Monte-Carlo method is only used for verification in case of questionable results. For both methods the computation time increases linearly with the size of the input and convolution matrices. Since the dose matrices are currently fixed to the standard CadPlan format (160 x 112 x number of slices), dose modifications are slower. (In principle only that part of the dose matrix that surrounds the volume of interest needs to be included in the calculations). Besides, almost all elements of the dose matrix are non-zero whereas volume matrices contain a substantial part of zero elements that are ignored in the calculations. The normal time required for dose modifications with the convolution method varies
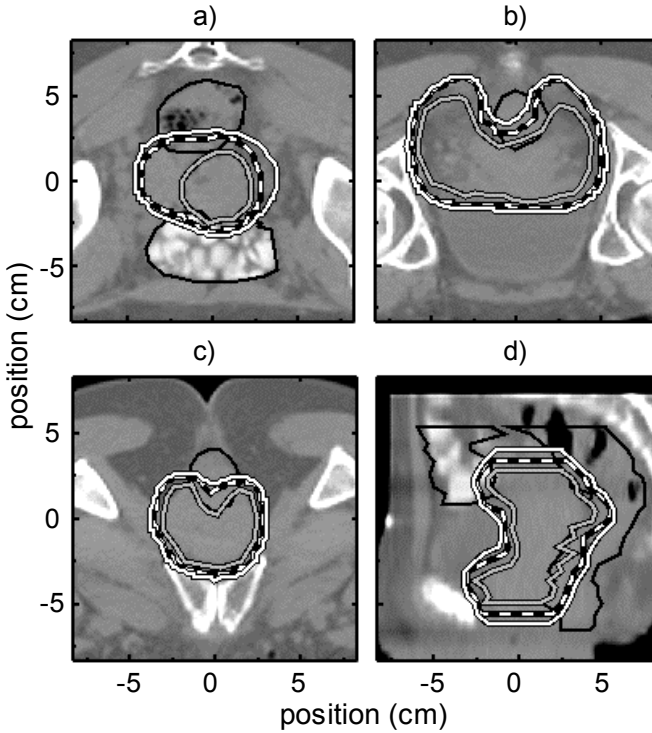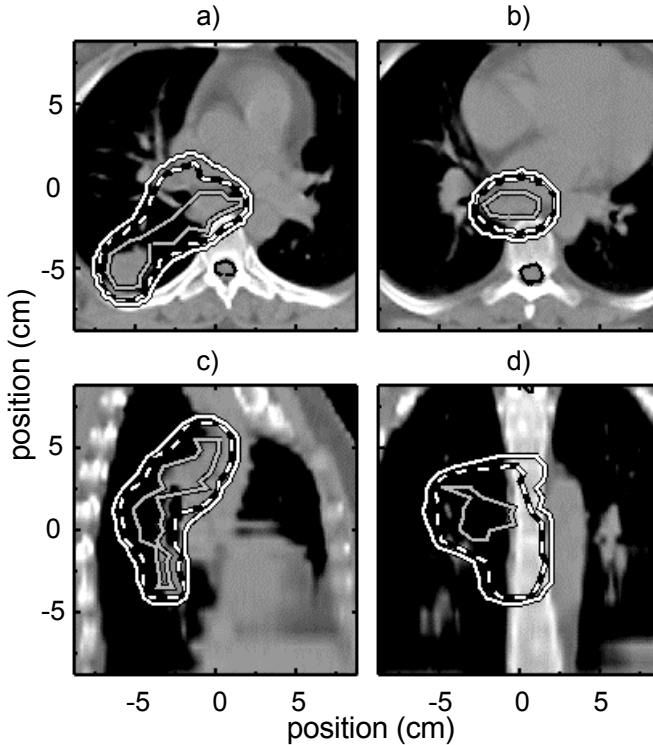
Figure 3.9 Example of the PTV calculation for a lung cancer plan. The geometrical uncertainties are from Table 3.1. The same procedure and line styles as for the prostate case are used (figure 3.5). Two transversal (a,b), a sagittal (c), and a frontal cut (d) through the PTV are shown. The overall CTV-to-PTV margin becomes about 10-13 mm. The critical organs are the two lungs and the spinal cord.

from two (without rotations) to about ten minutes which is a fraction of the time needed for volumetric dose calculations. Hence the method is sufficiently fast to be used in the iterative planning process.

3.4 Discussion

3.4.1 Margin calculations

The methods proposed in this paper will be particularly useful for designing PTV margins in case of new conformal therapy studies. Based on knowledge (or intelligent guesses) of a set of standard deviations describing all possible geometrical uncertainties of the CTV position, the CP values will give an indication where the CTV of an individual patient will be positioned over the course of treatment. However, repeated calculations within a specific patient group will normally yield equal margins for all patients independent of the shape of the CTV. For normally distributed deviations and in absence of significant rotational deviations, our choice of iso-probability volumes, which is based on the assumption that on
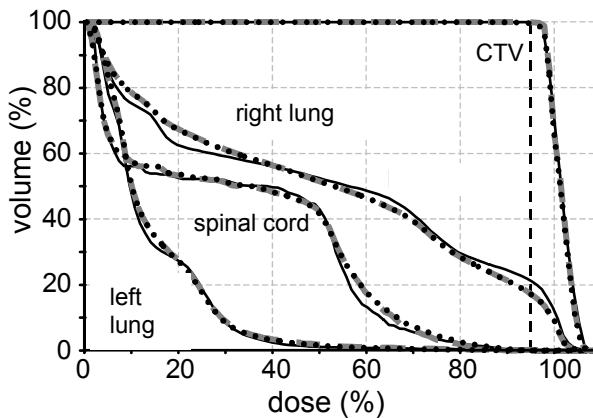
Figure 3.10 DPHs and DVHs of a plan which spares the left lung for the lung cancer patient from figure 3.9. The effect of the random uncertainties (dashed grey curve) on the CTV and right lung is similar to the prostate and cervix case. However, for spinal cord and left lung the effect of random deviations is a slight increase in the high dose volume. The DPH curves (dotted) indicate that the dose distribution in the CTV is influenced by the systematic deviations but the 95% isodose is still adequately placed around the PTV.

average a high percentage of the CTV volume (> 99%) should receive a high dose (> 95%), yields margins of about $2\Sigma + 0.7\sigma$. Consequently, the margins might also be applied directly to the CTV by straightforward CTV expansion algorithms [7, 18, 131], without having to perform the CP calculations each time.

One difference between rigid margin addition and the CP method might reveal itself at sharp edges. If no smoothing is performed, the former method will normally closely follow all irregularities in the CTV surface so all random deviations in the delineation of the CTV will be present in the PTV as well. The latter method will in itself tend to smooth the surface and can yield a slightly smaller PTV due to rounding of the corners as has been shown in figure 3.3 and figure 3.7d. Another difference will occur at small volumes or near small extensions of a volume. For volumes with diameters less than about 2 SD of the distribution of deviations, a CP value will no longer represent the probability that the volume will be partly outside of it. The actual probability will be larger. Therefore, choosing an iso-probability volume as new PTV will yield too small margins. Rigid margin addition is insensitive to the size of the input volume.

In clinical practice one might prefer a tighter CTV-to-PTV margin near a dose-limiting structure. In the model described in this paper, nearby critical organs are currently ignored in the PTV calculation. However, a critical organ might also be represented by a 3D volume matrix which can then be used to modify the CTV-to-PTV margins locally, as has been described in a previous paper about straightforward margin calculation [131]. The dose probability histograms can of course still be used and the influence of the local margin changes on the average dose volume histogram can immediately be visualized.

The described method assumes that all probability distributions of translations and rotations are mutually independent and of a gaussian nature. The mutual independence is however not always an accurate description of reality; for example, for internal prostate

movement a relation between rotations and translations has been found [54]. This might more easily be incorporated in the Monte Carlo method than in the convolution method. Instead of sampling from normal distributions, translations and rotations should be simultaneously obtained from a data base of prostate movements for a large group of patients [80]. This "brute force" method will require a large data base and many samples and will therefore be time consuming. The gaussian nature of the distributions has been verified repeatedly by studies on set-up position verification. At present too few studies have been performed to establish whether this is the case for internal organ movement as well or if differently shaped probability distributions might be more appropriate. The algorithm does however not depend on the type of probability distribution. Instead of using normal distributions for the convolutions, other distributions can be implemented as well. The output of the calculations will still be a 3D matrix with coverage probabilities which will still represent the probability that the CTV lies partly outside the corresponding voxel. Hence the same methods can be applied to obtain the PTV margins.

3.4.2   Dose modifications

Dose modifications that account for random deviations cause a shift of high dose to the lower dose regions. Consequently, the critical organs that receive a high dose in the original plan, like the rectum and bladder in the clinical examples, will receive less dose in the modified plan. Furthermore, from figure 3.2 is clear that the average DVH curve for systematic deviations (i.e. the DPH) can also be lower than the original, although the effect is limited for the clinical examples. This implies that the standard DVHs of the planning CT slightly overestimate the dose in those organs. For the critical organs that receive relatively low doses in the planning CT situation, the reverse is true. In the example for the lung cancer patient the organs that have specially been spared (i.e. the spinal cord and the left lung) will on average receive a somewhat higher dose due to the geometrical uncertainties.

A consequence of the application running outside the planning system is that dose inhomogeneity corrections are not included when calculating the dose distribution corrected for random deviations. The effect might be less significant for calculations in the pelvic area but the accuracy of planning in the thorax or neck region will probably be affected. A further simplification of the model is that the position of the beam with respect to the direction of the deviations is of no consequence. In reality the change in dose of a volume moving in the direction of the beam is determined by the inverse square law and the slope of the depth-dose curve, which is not incorporated in the algorithm. The errors will be relatively small [14], but, ideally, dose modifications should be performed for each beam separately which would be a time consuming procedure. In a multiple beam plan the errors are less prominent. Nonetheless, the effect of the set-up deviations might be slightly over estimated.

The application can also be utilized for a check of the delivered dose after the irradiation series have been completed. Once all fractions have been delivered and portal imaging has been applied, systematic and random set-up deviations can be calculated and used as input to the application to evaluate the actually delivered dose distribution. Thus it can be assessed whether the treatment has been performed correctly and if complications or a different local control probability might be expected. The same procedure can also be applied before the end of the treatment series to verify the situation at that moment and possibly alter the course of the treatment, e.g. change the PTV margin [144]. Even the

actual set-up deviations for every fraction might be entered separately and the conse-
quences of the treatment on the dose can be computed more precisely.

3.4.3   Dose probability histograms

The concept of DPH is especially useful to determine whether the calculated dose
distribution guarantees adequate CTV irradiation despite systematical uncertainties. Since
the dose has been planned conform the CTV, every variation in CTV position will
deteriorate the dose distribution in the CTV. Hence the spread in DVH curves due to
systematic deviations will be reflected in the average curve. The mean DVH will therefore
directly indicate the probability of CTV underdosage and hence the goodness of the plan in
this respect. For critical organs however, positioning deviation in one direction might
improve the DVH, movement in the opposite direction worsen the DVH, and the average
will be equal to the original. Hence instead of the average DVH for all systematical
deviations, the spread of the DVH curves would be more interesting [66,80]. This cannot
directly be calculated using DPHs but it is possible to extend the critical organ with a
margin for systematical deviations so that an indication of possible dose values in the organ
is obtained. This might for instance deteriorate the lung plan because extension of the spinal
cord volume might yield a significant probability of higher spinal cord dose than indicated
in figure 3.10.

   At first sight, there appears to be a discrepancy between figure 3.1 and the DPH
calculations in this paper. Figure 3.1 predicts that our choice of the 2.5% iso-probability
contour as PTV will for deviations in 3D result in the CTV being partly outside the PTV in
at least 28% of all patients. The average DVHs (or DPHs) in the clinical examples indicate
however that the probability of underdosage of the CTV (dose < 95%) is at most about a
few percent. This is because the extent of CTV volume outside the PTV is small. Using a
formula similar to Eq. 3.5, it can be calculated that the average distance that the CTV
border will exceed the PTV border will be about $0.5\ \Sigma$ for those 28% of patients. Since
clinical Σ-values range from about 2 to 4 mm, a 1 to 2 mm thick slice of the CTV will be
systematically outside the PTV for those patients. This is about equal to the size of one CT
pixel and since the 95% isodose volume does normally not enclose the PTV exactly, the
probability of underdosage is significantly smaller than the probability of the CTV being
outside the PTV.

   Since the tightness of the 95% isodose around the PTV will be dependent on the
treatment technique used (and on the beam characteristics), the required PTV margins to
guarantee a high probability of sufficient dose in the CTV despite systematical deviations
will be technique dependent as well. For instance, for conformal techniques slightly larger
PTVs are required than for conventional techniques using rectangular fields to obtain the
same DPHs. The same holds for the margin for random deviations. If that margin is based
on the shrinkage of the 95% isodose, this will also be technique dependent. For patients
groups with standard treatment techniques, the PTV and block margins can be optimized
using the DPHs of the dose distribution modified for random deviations. Once the margins
are standardized, the iteration process to obtain optimal PTV margins will be superfluous
for each patient separately.

3.5 Conclusion

A model has been developed which allows for the inclusion of geometrical uncertainties in the radiotherapy planning process. Required parameters are a set of twelve standard deviations describing the various uncertainties. The model calculates PTV margins based on the requirement that on average a large part of the CTV (> 99%) is irradiated with a high dose (> 95%). The size of adequate margins appears to be approximately equal to $2\Sigma+0.7\sigma$. Since rotational deviations are included, the margins can well be anisotropic. Once the patient is planned, the influence of the systematical deviations on the dose distribution in the CTV is determined by the average DVH for all systematic deviations, using so-called dose probability histograms. The influence of random deviations is determined by convolution of the dose distribution with the probability functions. In an iterative process of planning and verification of the CTV coverage, the CTV-to-PTV and block margins can be optimized for each patient separately. For standard planning techniques of specific patient groups, the margins can be standardized and the iterations omitted.

4 Analysis and reduction of 3D systematic and random setup errors during the simulation and treatment of lung cancer patients with CT-based external beam radiotherapy dose planning

J.C.J. de Boer, J.R. van Sörnsen de Koste, S. Senan, A.G. Visser and B.J.M. Heijmen

Abstract

Purpose: To determine the magnitude of the errors made in (i) the setup of patients with lung cancer on the simulator relative to their intended setup with respect to the planned treatment beams and (ii) in the setup of these patients on the treatment unit. To investigate how the systematic component of the latter errors can be reduced with an off-line decision protocol for setup corrections.

Methods and Materials: For 39 patients with CT planning, digitally reconstructed radiographs (DRRs) were calculated for anterior-posterior and lateral beams. Retrospectively, the position of the visible anatomy relative to the planned isocenter was compared with the corresponding position on the digitized simulator radiographs using contour match software. The setup accuracy at the treatment unit relative to the simulator setup was measured for 40 patients for at least 5 fractions per patient in two orthogonal beams with the aid of an Electronic Portal Imaging Device (EPID). Setup corrections were applied, based on an off-line decision protocol, with parameters derived from knowledge of the random setup errors in the studied patient group.

Results: The standard deviations (SD) of the *simulator* setup errors relative to the CT planning setup in the lateral, longitudinal and anterior-posterior directions were 4.0, 2.8 and 2.5 mm, respectively. The SD of rotations around the anterior-posterior axis was 1.6° and around the left-right axis 1.3°. The setup error at the *treatment* unit had a small *random* component in all three directions (1 SD = 2 mm). The *systematic* components were larger, particularly in the longitudinal direction (1 SD = 3.6 mm), but were reduced with the decision protocol to 1 SD < 2 mm with, on average, 0.6 setup correction per patient.

Conclusion: Setup errors at the simulator, which become systematic errors if the simulation defines the reference setup, were comparable to the systematic setup errors at the treatment unit in case no off-line protocol would have been applied. Hence, the omission of a separate simulation step can reduce systematic errors as efficiently as the application of an off-line correction protocol during treatment. The random errors were sufficiently small to make an off-line protocol feasible.

4.1 Introduction

Despite a dose of 65 Gy, the local complete remission rate in patients with non-small cell lung cancer (NSCLC) is only around 20% [34]. Unless the irradiated volumes are decreased, radiation dose escalation can result in unacceptable pulmonary and esophageal toxicity, particularly when concurrent chemo-radiotherapy is used [29]. Volume reduction may sometimes be achieved by omitting elective nodal irradiation [51, 108] but also by enabling the use of small planning margins. Such small margins would also reduce toxity at currently standard dose levels. In this paper, we investigate the setup inaccuracies of lung cancer patients and discuss the impact on planning margins.

Treatment planning in accordance with the recommendations of the ICRU 50 report requires the definition of a clinical target volume (CTV), which must encompass the gross tumor and subclinical disease and possibly involved lymph nodes. The CTV must be expanded to a planning target volume (PTV) by some geometrical margin. This margin must guarantee adequate coverage of the CTV during treatment and should therefore be based on knowledge of target movement with respect to the treatment fields, and machine accuracy (e.g. reproducibility of the block positions). However, the ICRU 50 report does not give a clear recommendation on how the CTV-to-PTV margin should be chosen.

Stroom *et al.* [125] have derived a general calculation method based on the dose coverage probability of the CTV to derive a 3D margin from known setup errors and internal organ movements, which may have any probability density function. They found that a clear distinction must be made between systematic and random errors, where a systematic error is the average setup error of the target volume over all fractions for a certain patient and the random errors are the inter-fraction variations. For gaussian error distributions, with a standard deviation (SD) Σ for the systematic errors and σ for the average random error, a CTV-to-PTV margin of $2\Sigma + 0.7\sigma$ seemed appropriate to guarantee that, on average, 99% of the CTV receives at least 95% of the prescribed dose (if the 95% isodose contour encompasses the PTV in the treatment planning). Their calculations incorporated the presence of small rotations (1 SD ~ 1°). A very similar result was found by Van Herk *et al.* [55], who did an analytical calculation for the simplified situation of a spherical target volume in an ideally conformal homogeneous dose. Both these results confirm the intuitive notion that systematic errors are more important than random errors in establishing planning margins.

Most setup studies in lung cancer patients are 2D [40, 112, 124, 145] or do not properly separate and quantify setup errors in terms of random and systematic components [20, 48, 113]. These studies usually pertain to small (< 20) numbers of patients [40, 48, 112, 113, 124]. In addition, none of these studies takes into account the setup errors that are made on the simulator (which have been shown to be of importance for other tumor sites [12, 76]). Lastly, the impact of a 3D off-line correction protocol on the systematic errors during treatment has not yet been studied in this patient group. As stressed above, accurate knowledge of both the systematic and the random setup inaccuracies (and how they can be reduced) is a prerequisite for calculation of optimal planning margins. Therefore, we measured the magnitude of both simulator and treatment systematic setup errors as well as random treatment setup errors in 3D in 40 lung cancer patients. In addition we report on the reduction of systematic treatment errors through an off-line setup verification protocol.

Geographical tumor misses are caused both by external setup errors and by internal movement of the target volume [40, 110]. In this study we focus on the external setup errors for a group of lung cancer patients with CT planning, consisting of

(1) The errors made by using simulator films as definition of the reference setup [12, 76]. The position of the patient anatomy in the reference setup (during which the treatment isocenter is marked on the patient and reference images are obtained) relative to the isocenter should be in agreement with the corresponding position in the CT treatment plan. However, in many institutions (including, until recently, our own) it is customary to mark the final beam setup at the simulator, after the planning has been performed. The definition of the final isocenter is based on visual inspection, and therefore may deviate from the intended CT plan isocenter. This simulator setup error results in a *systematic* error in the patient treatment. For prostate [12] and head and neck [76] irradiation, it was found that the simulator setup errors are comparable to systematic setup errors at the treatment unit. In these two cases, bony structures relevant to patient setup can be clearly identified, somewhat in contrast to the thorax region. Therefore, we expect that the simulator setup errors for lung cancer patients will be at least equal to, or larger than, systematic treatment unit setup errors. The simulator setup errors are however not well appreciated in the literature on setup accuracy of these patients. In fact, of the 7 setup studies we refer to above [20, 40, 48, 112, 113, 124, 145], 6 defined explicitly which type of setup reference image was used. Out of these 6 studies, 5 used simulator images based on a classical simulation [40, 48, 112, 113, 124] (i.e., without registering with the digitally reconstructed radiographs (DRRs)) and one study [20] applied both simulator films and DRRs.

(2) The setup errors at the treatment unit relative to the reference setup, which have a systematic and a random component. The systematic component may be reduced with portal imaging and setup corrections based on an off-line decision protocol [13, 116]. We have investigated both components and report the results of an off-line decision protocol.

4.2 Methods and Materials

4.2.1 Description of setup errors

We adopt the definitions introduced by Bijhold *et al*. [22] for systematic and random errors. The systematic setup error of a patient is the setup error averaged over all dose fractions. The group mean of this error is denoted by μ and the SD by Σ (i.e., the inter-patient mean setup variation). The random error of a patient is the standard deviation (SD) of the setup error from fraction to fraction (i.e., the inter-fraction variation). An appropriate average is taken to obtain the average random error for the patient population, denoted by σ. A specific type of setup error (e.g. a translation in a certain direction) is characterized by the set (μ, Σ, σ). Details on their calculation can be found in the *Appendix*. Note that if the lasers and light-field are proper indicators of the isocenter in both the reference situation (simulator or CT scan) and the treatment setup, μ should be close to zero, but Σ may still deviate significantly from zero due to interpatient variations (e.g. due to setup errors and patient movement during the CT scan or simulation, skin mobility and changes in body weight).

For measurement of the setup errors, we adopt the fixed coordinate system recommended by the International Electrotechnical Commission (IEC 61217). In this system, the *Y* axis coincides with the rotation axis of the gantry, the *Z* axis is perpendicular

to the floor and the *X* axis is perpendicular to both previous axes. For our study, the (*X,Y,Z*) directions coincide with respectively the lateral (LR), longitudinal and anterior-posterior (AP) directions. The setup errors for direction *D* (*D* = *X, Y,* or *Z*) are summarized by (μ_D, Σ_D, σ_D). In a similar way rotations are reported, but now indicated by *AP* and *LR* for rotation axes *Z* and *X* respectively.

Treatment setup errors measured relative to a reference image (in our study, a simulator radiograph) have been denoted by superscript *T* and setup errors in the definition of that reference have superscript *R*. Measurements on the treatment setup error for patients which were subject to an off-line correction protocol were used to obtain the setup errors that would have been found if no protocol had been applied, as described in the Appendix. The quantities that result from this calculation are embellished with a tilde symbol (e.g. $\widetilde{\Sigma}$).

4.2.2 Anatomical structures and template matching

In a previous paper [112], we performed a detailed analysis of the intra-fraction movement
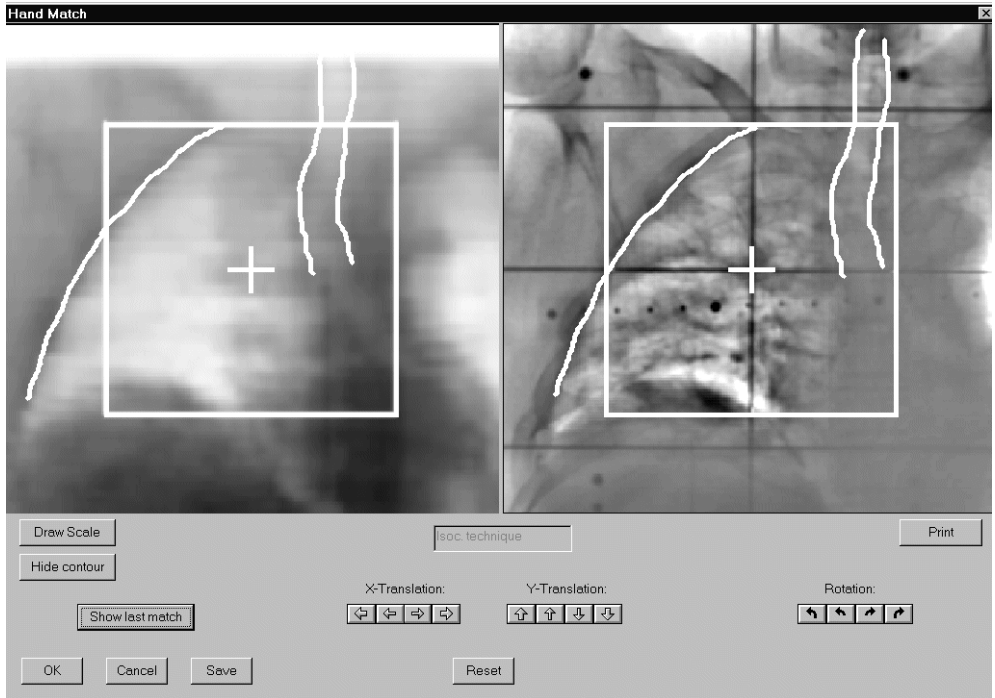
Figure 4.1 Screen shot of the RAPID template match. On the left is a DRR, on the right the corresponding simulator film. The contours (which were delineated on the right image and matched on the left) represent the thoracic wall and trachea. On the left, the cross indicates the actually planned isocenter, on the right the simulated isocenter. For both images, the square outlines a 10×10 cm^2 symmetrical field. As seen from the position of the anatomical contours on the simulator image with respect to the 10×10 cm^2 field outline, the simulated treatment field is shifted laterally by approximately 1 cm with respect to the planned position on the left.

of anatomical structures visible in AP portal images. Multiple portal images per treatment fraction were obtained, contoured, and matched with the first portal image of the corresponding fraction. Relevant structures that exhibited little motion were the thoracic wall and the trachea. For the thoracic wall, the average SD of intra-fraction movement in the X and Y directions was 0.8 respectively 0.6 mm and for the trachea 0.8 and 1.7 mm respectively. These SDs includes manual delineation and template matching errors, which are probably the origin of the relatively large SD for the trachea in the Y direction. If the two structures were matched simultaneously, the SD of the intra-fraction movement in both directions was 0.5 mm. We therefore used this structure combination in all the template matches referred to in this paper.

For lateral fields, we matched the vertebrae, sternum and occasionally the trachea. We expected the intra-fraction motion of these structures to be negligible, and a pilot study was performed to verify that expectation (see next section).

Fig. 4.1 shows a picture of the template match screen of the applied software of the RAPID Electronic Portal Imaging Device (EPID) described below. The contours are manually delineated on the reference image (right panel) and can be "picked up" and shifted over the image in the left panel using the mouse. In principle, the contours can also be rotated around a user-defined point, but the users were instructed to use this option with caution and only if they could not reach a satisfactory result without it.

4.2.3   *Intra-fraction movement in the Z direction as observed in lateral beams*

As mentioned, in a previous study we performed an analysis of intra-fraction movement in the longitudinal and lateral directions for AP treatment fields [112]. This analysis was performed with movie-loop images obtained with a *BeamView^{PLUS}* EPID (Siemens). A simultaneous analysis of the setup errors in the AP fields showed that the random inter-fraction errors were relatively small (1 SD = 2-3 mm). For the lateral fields it was hard to obtain movie-loop images with relevant structures, because the fields were usually too small. Therefore, we studied the intra-fractional movement of structures in the Z direction on the basis of inter-fraction variations.

We selected the vertebrae, trachea and sternum as structures for which we expected little motion due to breathing and heartbeat. Next, we obtained images of large rectangular lateral fields with a (Philips) SRI-100 EPID in 8 patients and studied the setup error in the Z direction using a (simultaneous) template match of the mentioned structures. The large portal imaging fields (PIFs) were necessary to capture these structures in the image. The exposure per PIF was 10 monitor units (MU) with a 10 MV beam at 200 MU/min. No more than 10 fractions per patient were imaged to keep the dose to healthy tissue at a level that was considered acceptable by the radiation oncologists. The PIF was taken into account during treatment planning and dose-volume histograms were generated with the maximally acceptable PIF dose. The mean PIF exposure time of 1.8 sec corresponds to about a third of the average breathing cycle period [70]. Therefore, if the mentioned anatomical structures exhibit motion with breathing, the measured random (day-to-day) setup error, σ_Z^T, presents an upper limit on the magnitude of these motions since it also includes random setup errors.

4.2.4   Simulator setup errors

Simulator setup errors were evaluated for 39 patients with NSCLC. Frontal and lateral laser lines at the CT scanner were marked with ink lines (21 patients) or tattoos (18 patients), as well as a longitudinal lead strip which was clearly visible in the CT scan. The patients were scanned (and treated) in the supine position with no fixation devices, except for an armrest at the top of the table in which the upper arms were positioned and a feet rest. The CT slice thickness and spacing were 5 mm. The CT data were transferred to the (Varian-DoseTek) CadPlan treatment planning system (TPS) to construct a 3D isocentric plan.

The resultant treatment portals were simulated and the radiation oncologist judged the accuracy of field placement on the basis of the simulator films, beam's-eye-view (BEV) plots and standard CT scout topograms (which suffer from incorrect beam divergence in the transversal plane and no divergence in the superior-inferior direction). No DRRs were used during the simulation, since the aim of our study was to investigate the accuracy of our routine simulation procedure, which did not involve the use of DRRs. At the start of simulation, each patient was positioned according to the CT markings using the simulator lasers. Next, a (pre-calculated) table translation was executed in agreement with the position of the definitive isocenter of the treatment plan. The radiographs obtained in this situation often led to additional table adjustments indicated by the clinician based on visual comparison with the BEV plots. Once the setup was judged satisfactory, the final isocenter position was marked on the patient using skin tattoos and long laser lines, and reference setup radiographs were obtained at gantry angles 0° and 90° or 270° (N.B. These angles may differ from the treatment beam angles).

Retrospectively, at the same gantry angles, DRRs were constructed from the 3D planning data. They were generated with in-house developed software, which had the benefit of a higher spatial resolution than the DRRs from the TPS. It was verified that the TPS DRRs coincided geometrically with the DRRs used in our analysis (difference < 0.5 mm). The setup difference between DRR and simulator film was measured using the anatomy and template match software described above (Fig. 4.1). These simulator setup errors were analyzed by calculating the mean error and standard deviation per translation direction and rotation axis (yielding, respectively, the systematic errors ($\mu_X^R, \mu_Y^R, \mu_Z^R, \mu_{AP}^R, \mu_{LR}^R$) and ($\Sigma_X^R, \Sigma_Y^R, \Sigma_Z^R, \Sigma_{AP}^R, \Sigma_{LR}^R$) in the terms explained above).

For 16 patients, the delineation of contours in AP and lateral digitized simulator films and the template matching procedure with the corresponding DRR was performed by two observers. We estimated the SD of the analysis error for a single observer, SD_{obs}, from the SD of the differences between the two observers (which equals $\sqrt{2} \times SD_{obs}$).

4.2.5   3D setup analysis at the treatment unit and off-line correction protocol

The setup errors at the treatment unit were studied in 40 patients who were irradiated with at least 20 fractions, of which at least 5 fractions were imaged and analyzed within the context of the correction protocol described below. When this setup study began, the awareness of the potentially large simulator setup errors was limited and so digitized simulator films were used as reference images, in agreement with our routine portal imaging practice at that time. Mainly due to logistic reasons, this group was not identical to the group studied for simulator setup errors, but did include 19 patients from that group. 10

MV PIF images were obtained with a fluoroscopic, CCD camera-based EPID developed in our hospital. With this RAPID (Rotterdam Advanced Portal Imaging Device) system, images of good quality can be obtained with only a few MU of exposure [26]. If the treatment field gantry angles did not deviate more than 20° from 0, 90 or 270°, the same angle would be used to obtain an enlarged PIF for setup verification. Within these 20° intervals, the anatomical structures that were identified to have negligible intra-fraction motion in the (exact) AP/PA or lateral portal images can still be considered stable structures. In addition, these structures can be easily identified in images in this gantry angle interval. The procedure also allowed for the generation of a 'double exposure' image in the RAPID software, an example of which is given in Fig. 4.2. If an appropriate gantry angle was not available, an extra (AP or lateral) beam direction for the PIF was added to the treatment plan. The exposure per PIF was 6 MU, the minimal dose at the clinically applied doserate for the MM50 Racetrack Microtron (Scanditronix Medical AB) at which this study was performed. 2 PIFs (with approximately orthogonal gantry angles) were given per imaged fraction. PIF images were template matched with the corresponding simulator films using the RAPID software (Fig. 4.1) and the previously mentioned anatomical structures.

An off-line (shrinking action-level) correction protocol, such as described by Bel *et al.* [17] for prostate cancer patients, was applied to reduce the systematic setup error. In short, the protocol tests the length of the *average* 3D setup error *vector* against an action level. In the first part of the protocol, the average vector at a certain fraction is calculated over all imaged fractions obtained up to and including that fraction (say n_{meas} fractions) whereas the
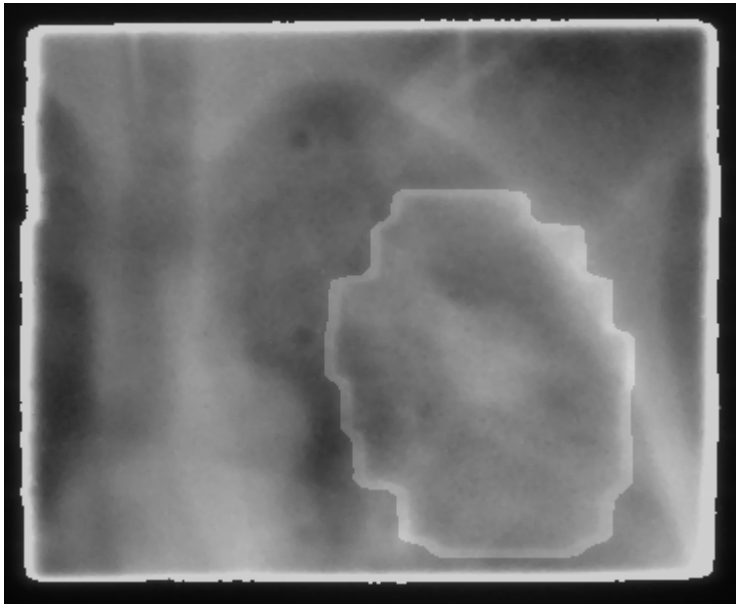
Figure 4.2 Overlay of an anterior treatment field on a corresponding 6 MU PIF image. Clearly, very few anatomical details may be discerned in the treatment field (although the primary tumor is visible in this specific case) and the PIF image is necessary to render important structures visible.

action level at that fraction becomes $\alpha / \sqrt{n_{meas}}$; α is a suitably selected 'initial' action-level which depends on the (known) random setup errors [13]. If the vector length exceeds the action level, a setup correction is performed in all subsequent fractions and the procedure is repeated with n_{meas} reset to 1. This process continues until n_{meas} $= N_{max}$ subsequent measurements have been obtained without the need for a correction, after which the first phase of the protocol is finished. In the second phase of the protocol we applied a weekly check measurement which was included in a running average of the last N_{max} measurements and tested against the fixed action level $\alpha_{final} = \alpha / \sqrt{N_{max}}$ [17]. If a correction is necessary in this second phase, the protocol is completely restarted. Effectively, the protocol should reduce the systematic 3D error vector length to below α_{final} in nearly all patients [13].

If random errors are large compared to systematic errors, the off-line protocol is no longer practical: a large initial action level must be selected to prevent frequent corrections due to the random errors, and so a large number of images must be acquired to reach a certain α_{final}. In a previous study [112], we found that the systematic setup errors Σ_X^T and Σ_Y^T as well as the random setup errors σ_X^T and σ_Y^T were 2-3 mm for a group of lung cancer patients with a setup technique identical to the group studied in this paper. Assuming the setup in the Z direction to be equally precise (confirmed by results presented in this paper), we found from Monte Carlo simulations similar to those described by Bel *et al.* [13] that for a correction protocol with $\alpha = 8$ mm, $N_{max} = 3$ ($\alpha_{final} = 4.6$ mm) applied to a 25 fraction treatment, with weekly check measurements, the average number of corrections per patient is approximately 1, with on average 9 imaged fractions. Such a correction protocol should typically halve Σ^T for each direction. The required *average* PIF dose to healthy tissue would be close to 1 Gy, which was considered acceptable (note that about half of the PIF dose is delivered with intervals of about 7 days).

Individual patients may, of course, have significantly larger random errors than the population mean, which manifests itself through multiple (inaccurate) corrections. To prevent invalid use of the off-line protocol, we accepted no more than 2 corrections per patient. If a third correction was necessary, the clinician was informed and the origin of the large random error was investigated

If the protocol generated a setup correction, in subsequent sessions the patient was first positioned using the isocenter demarcation applied at the simulator. Next, the table was translated by the technicians according to the values indicated by the protocol software. No new demarcation was applied, since in our experience this may lead to confusion, especially if multiple corrections are required. We measured that the accuracy of the table shift readout is better than 0.1 ± 0.3 mm, the effect of which we neglected in subsequent analyses.

4.3 Results

4.3.1 Intra-fraction movement in the Z direction

For the 8 patients in the pilot study, the random setup error in the Z-direction, as measured from the position of sternum, vertebrae and trachea, was $\sigma_Z^T = 2.2$ mm. This random error is as small as the random errors in the X-and Y-direction in a similar patient group [112],

which we obtained with structures that were proven to exhibit little intra-fraction movement. It is also equal to the random setup error we obtained for over 700 prostate cancer patients ($\sigma_Z^T = 2.0$ mm) at the same treatment unit [17, 73]. The latter patient group can be positioned very accurately and exhibits no indication for intra-fraction movement. In summary, our data confirm that the selected anatomical structures reveal negligible intra-fraction motion and may thus be used for accurate setup verification.

4.3.2   Simulator setup errors

We analyzed the results for the patient group which had CT lasers marked with ink respectively with tattoos separately and investigated differences between the population means and standard deviations with the *T*-test and *F*-test. At the p = 0.05 level, no significant difference was found for any translation direction or rotation axis. This result reflects the fact that, although the tattoos are slightly better visible than the ink demarcation at the time of simulation, the procedures described above to arrive at the final setup at the simulator were identical in both groups. We therefore combined the data from both groups to improve statistics (Table 4.1).

For each translation direction, the probability density function for simulator setup errors was found to be consistent with a normal distribution in a Shapiro-Wilk *W*-test (p > 0.05). The values of all means (μ^R) are consistent with 0 within the statistical error (p > 0.05), which indicates that there is no overall error in the definition of the coordinate system in the TPS relative to that on the simulator, nor in the delineation of the contours in the radiographs. However, large simulation setup errors do occur, ranging up to nearly 1 cm for all 3 directions.

The rotations around the AP axis could possibly lead to erroneously derived values of the setup error in the *Y* direction for the lateral fields, but since we only scored setup errors in the *Z* direction in the lateral portals, this did not influence our analysis. Similarly, the (smaller) rotations found in the lateral portals will, to some extent, project as translations in the *Y* direction in the AP fields. However, a graph of these rotations versus the *Y* setup errors actually found in the AP fields indicated no trend, which was confirmed by the negligible linear correlation coefficient of 0.01 (p > 0.95).

An interesting relation was found between the gender of the patients and the magnitude of simulator setup errors in the AP images. For the 13 female patients, $\Sigma_X^R = 5.3$ mm while for the 26 male patients $\Sigma_X^R = 3.2$ mm. A two-sided *F*-test yielded p = 0.03, indicating that this difference is statistically significant. A similar significant result was established for the rotations around the AP axis: for the female patients we measured $\Sigma_{AP}^R = 2.2°$ in contrast to $\Sigma_{AP}^R = 1.2°$ for the men (p = 0.01). For the *Y* and *Z* translations and the rotations around the LR axis no significant differences were observed.

The inter-observer variation SD$_{obs}$ was approximately 1 mm for each translation direction and approximately 0.5° for both rotation axes. One could subtract SD$_{obs}$ quadratically from Σ^R, to correct for the dispersion introduced by observer errors, which would reduce the tabulated values by no more than 10%. However, an even smaller correction is appropriate since the inter-observer dispersion in *delineation* of structures in the reference image contributes to both Σ^R and SD$_{obs}$, and this dispersion should be maintained in Σ^R since it represents a true systematic error for setup verification at the

treatment unit [22] and therefore must be included in margin calculations. For this reason we omitted a correction for inter-observer variations on the values presented in Table 4.1.

Table 4.1. Mean (μ) and SD (Σ) of the simulator setup errors for 39 patients. The entry *Max* indicates the absolute value of the most extreme deviation.

	X (mm)	Y (mm)	Z (mm)	AP (°)	RL (°)
μ^R	0.4	0.6	0.3	0.5	0.4
Σ^R	4.0	2.8	2.5	1.6	1.3
Max	9	8	8	5	4

4.3.3   3D setup analysis at the treatment unit

Of the 41 patients entered into the study, one had to be excluded from the off-line protocol because of a large random error in the *Y*-direction ($\sigma_Y^T = 6.1$ mm), which led to multiple corrections without significant improvement in the setup accuracy. This patient had trouble with keeping a constant position, specifically of the arms and upper body, due to pain complaints associated with previous surgery. For the other 40 patients, no irregularities were observed.

On average, about 8 fractions per patient needed to be imaged and 600 portal images were analyzed. The average number of required setup corrections per patient was 0.6. Only one patient required 2 setup corrections. These numbers are close to the results of the numerical simulation described in the methods section. The measured setup errors are summarized in Table 4.2. We find random errors of about 2 mm, whereas the systematic errors that would have occurred if no correction protocol had been applied are somewhat larger. As expected, the off-line correction protocol roughly halves the systematic errors and reduces the length of the systematic error vector to below 4.6 mm for all patients whereas without the protocol the distribution tail would have extended beyond 1 cm (Fig. 4.3).

Table 4.2 Setup errors at the treatment unit for 40 patients.   The correction protocol influences only the systematic errors in the translation directions, not the random errors.

		X (mm)	Y (mm)	Z (mm)	AP(°)	RL (°)
No protocol	$\widetilde{\mu}^T$	1.6	0.0	-0.7	0.0	-0.4
	$\widetilde{\Sigma}^T$	3.2	3.6	1.7	1.1	1.0
	σ^T	2.0	2.1	1.8	1.0	1.0
With protocol	μ^T	0.9	0.0	-0.5		
	Σ^T	1.3	1.6	0.9		

The observed systematic errors before correction are anisotropically distributed, i.e. the systematic error $\widetilde{\Sigma}_Z^T$ is about half the value of $\widetilde{\Sigma}_X^T$ and $\widetilde{\Sigma}_Y^T$. An F-test confirms the significance of this difference (p < 0.01). The random error is also smallest in the Z-direction, although here the difference with X and Y is not statistically significant. The small setup errors in the Z-direction may be rooted in the fact that the isocenter-table distance is used as the final indicator for setup in that direction and not the position of the patient tattoos with respect to the laser lines. The smallness of Σ_Z^T and σ_Z^T confirms our conclusion on the stability of the anatomical structures selected for template matching for the lateral beams.

The value of σ^T in Table 4.2 is a population average of measured patient specific random error SDs σ_i^T, as defined in the Appendix. Suppose that the true random error SD (which we could measure if we had an infinite number of setups) is 2 mm in each patient (close to the mean values given in Table 4.2). In that case, the inter-patient variation on the measured value of σ_i^T would be purely due to the finite statistics of the 7-8 measured setups and the SD of this variation would be ~ 0.6 mm [123]. The measured SDs of the inter-patient variation in σ_i^T are 0.8, 0.9 and 0.6 mm for the X, Y and Z directions respectively. These values are close to the value of 0.6 mm derived above, indicating that
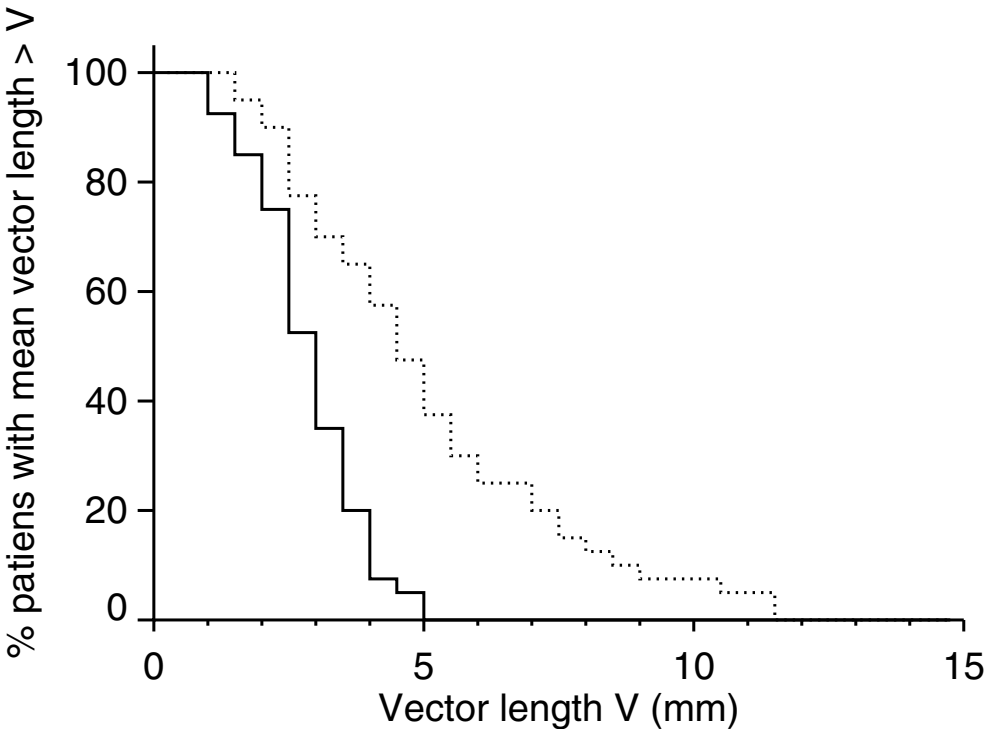
Figure 4.3 Cumulative distribution of the length of the systematic displacement vector for 40 patients treated with an off-line correction protocol (solid line). The dashed line indicates the distribution if this protocol had not been applied.

the inter-patient variation in σ_i^T is small. The slightly larger spread of random errors for the X and Y direction may be related to an inter-patient variation in the physical ability or experienced (dis-) comfort when holding the cross bar above their head. Despite this spread, the number of corrections was not dramatically larger for the 5 patients with the largest random errors in the X-Y -plane ($\sqrt{\sigma_{i,X}^2 + \sigma_{i,Y}^2}$): 0.8 correction per patient versus 0.6 on average.

Similar to the setup at the simulator, at the treatment unit female patients (14 in the studied group) seemed to be positioned less accurately than male patients (26 patients): they required, on average, 0.9 corrections whereas for the male patients 0.4 corrections were sufficient. Investigating this difference in setup correction frequency, we found that it originates in larger values of the systematic error before correction, $\widetilde{S}^T$, in the Y direction for the female patients (see appendix for a formal definition of $\widetilde{S}^T$). A Kolmogorov-Smirnov test on $|\widetilde{S}^T|$ in that direction confirms a significant difference between genders (p < 0.01) as is illustrated in Fig. 4.4 (an F-test on $\widetilde{\Sigma}_Y^T$ is not valid, since the distribution for the men is inconsistent with a normal distribution according to the Shapiro-Wilk test). For the other systematic errors, no significant difference in distribution between the men and women was observed. The random errors showed no difference in distribution between male and female patients either.
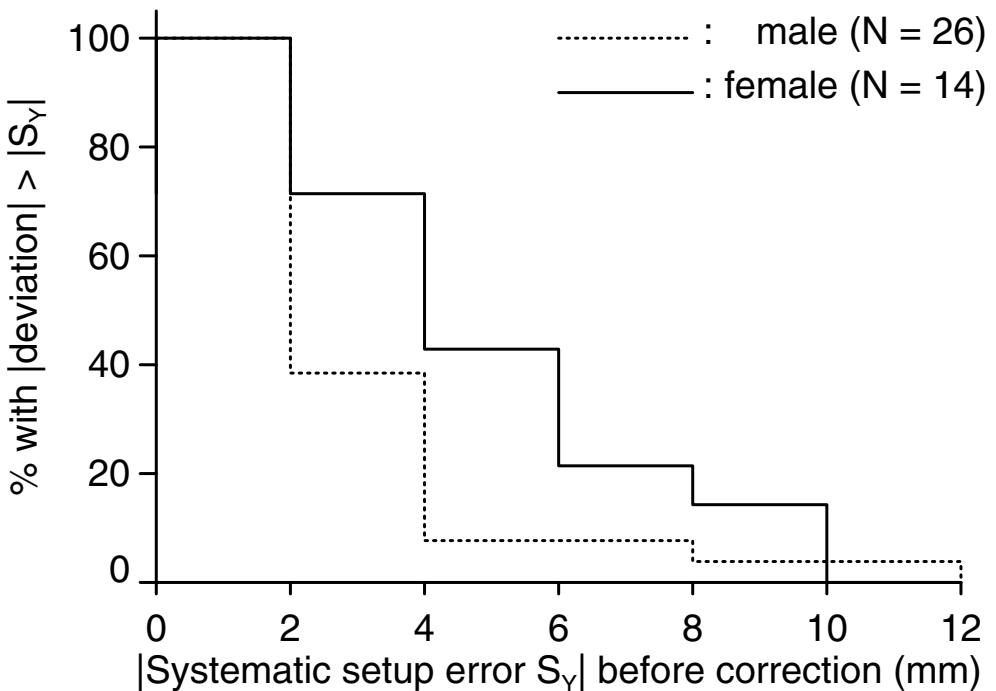
Figure 4.4 Cumulative distribution of $|\widetilde{S}^T|$ in the longitudinal direction, for male and female patients.

4.3.4   Relation between simulator and treatment unit systematic setup errors

If patient *i* has a simulation to mark the final field setup based on CT planning, then the *actual* systematic setup error during treatment with respect to the CT plan (S_i) is the sum of the simulator setup error (S_i^R) and the mean treatment setup error with respect to the simulator setup ($\widetilde{S}_i^T$ in case no setup corrections are performed during treatment; see Appendix). To calculate the SD of the actual systematic setup error from the SDs for the two types of errors given in Tables 4.1 and 4.2 (S_i^R respectively $\widetilde{S}_i^T$), we must establish to what extent they are correlated, i.e. do large positioning inaccuracies at the simulator lead to large systematic positioning inaccuracies at the treatment unit ? This requires a group of patients for which S_i^R and $\widetilde{S}_i^T$ have been measured separately.

   The overlap of the patient group studied for simulator setup errors (39 patients) and the group studied for treatment setup errors (40 patients) consisted of 19 patients. In this group, we searched for linear correlations between $\widetilde{S}^T$ and S^R and between their magnitudes $|\widetilde{S}^T|$ and $|S^R|$. There were no significant correlations for any of the directions (*X,Y,Z*). In fact, the covariance cov($\widetilde{S}^T$, S^R) was less than 0.4 for all 3 directions. Therefore, these two types of setup errors appear to be uncorrelated. From the Appendix and the numbers in Tables 4.1 and 4.2, we find that the impact of the small covariances on the SD of the total systematic setup errors is negligible. In the specified group of 19 patients we may calculate the actual systematic setup error exactly as $S_i^R + \widetilde{S}_i^T$ and find that the SDs of these actual errors are ($\widetilde{\Sigma}_X$, $\widetilde{\Sigma}_Y$, $\widetilde{\Sigma}_Z$) = (5,4,3) mm. These values are consistent with the numbers for $\widetilde{\Sigma}^T$ and Σ^R in Tables 4.1 and 4.2 and cov($\widetilde{S}^T$, S^R) = 0, again confirming that the two types of setup errors are uncorrelated.

   In the same 19 patients, both S^R and $\widetilde{S}^T$ have a distribution consistent with a normal distribution. An *F*-test on $\widetilde{\Sigma}^T$ versus Σ^R indicated no significant difference for any of the 3 directions. Hence, we conclude that the setup error at the simulator with respect to the setup at the CT scanner is distributed similar to the systematic setup error at the treatment unit with respect to the simulation setup.

4.4 Discussion

The purpose of investigating the possible displacements of a target volume during treatment relative to the planned situation, is to define appropriate planning margins as well as to identify efficient methods to ensure or even reduce those margins. These margins should account for both internal target volume displacements and external setup errors.
   The internal displacement for thoracic tumors is mainly due to breathing and cardiac activity. An estimate of the internal tumor displacements can be obtained from Ekberg *et al*. [40], who studied tumor mobility in all 3 directions in 20 patients with the aid of fluoroscopy. They found that, for tumor locations spread over left and right lung and all lobes, the average SD of the tumor displacement in the *X*, *Y* and *Z* directions was respectively 1.4, 2.6 and 1.3 mm during normal respiration. However, Ross *et al*.[110] described how the extent of motion measured with ultra-fast CT was found to be strongly dependent on the proximity to the heart/aorta or diaphragm. In fact, they describe a

displacement due to respiration of the order of 1 cm for lower lobe lesions. Indeed, Ekberg *et al*. also observed large variations in motion amplitude, ranging up to 12 mm. Apart from these large variations in amplitude and frequency with tumor location, the probability distribution function of these motions strongly deviates from a gaussian distribution [78]. Therefore, the SD of these displacements is a poor representation of the distribution function and it is not clear whether in such a case the expression for margins given in the Introduction ($M = 2\Sigma + 0.7\sigma$ which was derived for gaussian distributions [125]) can be applied. In addition, in this expression systematic and random displacements are separated, but this separation is not trivial in the case of internal movements. Particularly the systematic errors introduced by internal movement during the CT scan are not straightforward to describe since these movements occur on time scales of the order of the slice scan times of a CT scanner [8].

A margin calculation scheme that fully copes with internal organ motion and the associated problems sketched above, is beyond the scope of this paper. We therefore restrict our discussion to external setup errors. Hence, the margins derived below only reflect the impact of various treatment protocols on external setup accuracy. Their validity in clinical practice is strictly limited to volumes that are known to exhibit negligible internal motion, for instance tumors associated with the chest wall [110] or mediastinal lymphnodes. Note that due to our definition of matching anatomy, external setup accuracy in this discussion means accuracy of isocenter placement with respect to the chest wall and vertebrae.

In Table 4.3 we give the calculated margins in the *X*, *Y* and *Z* directions for 5 simulation and setup verification methods, based on the numbers we have found for the setup errors (Tables 4.1 and 4.2). In all 5 cases, we assumed that the patients had a CT based treatment plan. Margins were calculated following the methods described by Stroom *et al*. [125]: $M = 2\Sigma + 0.7\sigma^T$, where Σ is defined below for each method. From the first to the last case, the external setup accuracy increases. In order of increasing external setup accuracy, the 5 methods are:

1. The beam setup according to the treatment plan is demarcated at the simulator using BEV plots and planograms (as in the simulator setup study described in this paper) after which treatment takes place without a correction protocol. We have quadratically added the systematic simulation and treatment setup errors to arrive at the actual systematic setup error: $\Sigma = \sqrt{(\Sigma^R)^2 + (\widetilde{\Sigma}^T)^2}$ (see previous section).

2. As case (1), but now the 3D off-line setup correction protocol described in this paper is applied with digitized simulator films as reference images, and so $\Sigma = \sqrt{(\Sigma^R)^2 + (\Sigma^T)^2}$.

3. After the CT scan and treatment planning, no separate simulation is performed. The simulator step may be omitted e.g. by marking a temporary isocenter directly at the time of scanning. Setup to the final isocenter (from the 3D plan) may be achieved by performing a fixed table shift relative to the temporary isocenter during treatment or by making the temporary isocenter the definitive isocenter by using asymmetric fields. In this case, $\Sigma = \widetilde{\Sigma}^T$.

4. As case (3), but now including the effect of the off-line protocol with DRRs as reference images ($\Sigma = \Sigma^T$). We currently apply this protocol for lung cancer patients using high quality DRRs generated with CT-Simulator software (Picker AcQSim) for a CT slice distance of 2.5 mm.

5. As (4), but with an on-line instead of an off-line correction protocol. For the residual errors of the on-line correction protocol, we have adopted values from van der Steene *et al*. [124] who measured such errors in 16 patients in the AP plane for an action level of 2 mm per direction. We adopted the values of the systematic and random residual errors from Table 4.3 in their paper for Σ and σ^T, and assumed that the residual errors in the X direction are also applicable to the Z direction.

Table 4.3 CTV-PTV extension margins M in mm for external setup variations only. Margins were calculated from the expression $M = 2\Sigma + 0.7\sigma$ for the 3 setup directions (X,Y,Z). For a detailed description, see the discussion section.

Simulation/verification method	M_X	M_Y	M_Z
1: reference = sim film, no off-line protocol	7	12	11
2: reference = sim film, with off-line protocol	7	10	8
3: reference = CT scan, no off-line protocol	5	8	9
4: reference = CT scan, with off-line protocol	4	4	5
5: reference = CT scan, with on-line protocol	3	3	3

Table 4.3 shows that, for the group of lung cancer patients we studied, the large systematic simulator setup errors require large (~ 1 cm) CTV-PTV margins in each of the 3 setup directions (method 1). Although the application of an off-line correction protocol based on simulator images reduces these margins slightly (method 2), a similar result can be obtained by just skipping the separate simulator step and not using any correction protocol during treatment (method 3). The applied shrinking action-level off-line protocol efficiently reduces systematic errors and hence margins, only if DRRs are used as reference images (method 4, compare with method 2). It should be noted that the margins calculated for method 4 are still appropriate in case of a separate simulation step to mark the (approximate) isocenter position, as long as the off-line setup verification protocol is performed based on DRRs.

The 5th method represents what would happen if one would switch to an on-line protocol (imaging and possibly correcting during every fraction), based on DRRs. The margin reduction relative to an off-line protocol of ~ 1-2 mm in each direction must be balanced against the higher workload. Apart from the significant imaging and analysis workload, Van der Steene *et al*. [124] reported a correction frequency of 85% of the setups (however, their patient group seemed to suffer from larger systematic and random errors than presented in this paper). In addition, they could use the treatment portals to judge patient setup, which is impossible in the case of small thoracic fields (Fig. 4.2). When applying a 3D on-line protocol for a treatment with small fields, a large number of PIFs must be acquired to render the anatomy visible (minimally the number of fractions $\times$ 2). To keep the PIF dose to uninvolved tissue below ~ 1 Gy for a 30 fraction regime, images must be obtained with an exposure < 2 MU. If one also wishes to obtain check images to verify the corrections, the exposure must be even smaller: ~ 1 MU if 50% of the portals require a setup correction. This is near the limit of what may be achieved with video based EPIDs

[124], and it poses a challenge to scanning liquid-filled ionization chamber systems [53] since the required read-out time at doserates $\geq$ 200MU/min would be $\leq$ 0.3 sec. But apart from the PIF dose, note again that the margins given in Table 4.3 are only for external setup variations. Once the external systematic setup errors are down to < 2 mm (1 SD), like in method (4), these errors will no longer be accuracy limiting in most cases due to possibly large systematic errors related to internal organ movement [8] and tumor delineation inaccuracies [114]. If the latter two error sources are not well controlled, it is debatable whether the high degree of external setup precision that can be achieved with on-line setup corrections is relevant to treatment planning margins.

An unexpected result of our study was that systematic setup errors were significantly larger for women than for men, both at the simulator and treatment unit. However, this difference occurred for the lateral direction during simulation and in the cranial-caudal direction during treatment. The larger systematic errors in the cranial-caudal direction could be a result of the fact that the frontal tattoos on female patients cannot always be put exactly at the isocenter position in which case "direction tattoos" (which are placed on the laser lines, but at a small known distance from the isocenter) are used and the best longitudinal position is estimated from both lateral tattoos and the frontal tattoo during treatment. The reason behind the larger errors in the lateral direction during simulation for women is hitherto unexplained. Nevertheless, the benefit of omitting the simulator step together with the application of an off-line correction protocol based on DRRs is that such differences between gender would not influence the margins: the simulator setup error would vanish in all patients, while the off-line protocol ensures the treatment error is controlled. Indeed, we found no difference in the distribution of systematic treatment setup errors between men and women once the protocol was applied and so the distribution function given in Fig. 4.3 holds for both women and men. This approach fails if a patient has a much larger random error than the population mean, which we observed in one patient. For such a case, a transition to an on-line correction protocol could be considered if margin control is essential.

4.5 Conclusions

We have investigated the systematic and random external setup errors during the irradiation of lung cancer patients. The magnitude of the systematic errors at the simulator and the treatment unit is comparable if no protocol is used to reduce the latter errors (Tables 4.1 and 4.2), indicating that reduction of the former is no less important than reduction of the latter. Therefore, an off-line treatment correction protocol based on simulator films becomes inefficient (compare methods 2 and 3 in Table 4.3). This leads to an unnecessary loss of accuracy, since although no immobilization device was used, the random errors are sufficiently small (~ 2 mm) to reduce the systematic treatment errors significantly with an off-line protocol (Table 4.2). Such a protocol should therefore be based on DRRs, unless the simulation procedure is changed to better represent the planned treatment setup [76]. We have currently adopted the approach to mark the definitive isocenter directly after the CT scan on a Picker CT-Simulator, after which the (asymmetrical) treatment fields are planned while maintaining that isocenter. Thus, the simulation step and its associated errors are omitted altogether.

Acknowledgements

Appendix: Calculation of systematic and random errors

Consider a group of N_p patients with $N_{f,i}$ imaged treatment fractions for patient i. The measured setup error during *treatment* (indicated by T) for patient i in fraction j is ε_{ij}^T. In case of no setup corrections, the systematic error (i.e. the mean displacement.) of patient i is estimated by $S_i^T = N_{f,i}^{-1} \sum_{j=1}^{N_{f,i}} \varepsilon_{ij}^T$. The *population* mean setup error is calculated as

$$\mu^T = N_p^{-1} \sum_{i=1}^{N_p} S_i^T \, .$$

If one or more corrections took place, a time-weighed average of the systematic errors from correction to correction is obtained. Suppose that patient i has N_c (≥ 1) setup corrections, detected in fractions $f_1, f_2, ..., f_{N_c}$ (i.e., correction c, detected in fraction f_c, will be applied from fraction f_c+1 onwards). Hereafter, we omit subscription of the fractions f with patient index i to improve readability. Let the last measured fraction be f_l. If we define $f_0 = 0$ and $f_{N_c+1} = f_l$, then we calculate the systematic error for patient i by

$$S_i^T = f_l^{-1} \sum_{c=1}^{N_c+1} (f_c - f_{c-1}) S_{c,i}^T \tag{A4.1a}$$

where

$$S_{c,i}^T = N_{m,c}^{-1} \sum_{j \in \, <f_{c-1}, f_c]} \varepsilon_{ij}^T \tag{A4.1b}$$

is the estimated systematic error after correction c-1 has taken place but before correction c is applied and $N_{m,c}$ is the number of measurements in the corresponding fraction interval (in general, $N_{m,c} < f_c - f_{c-1}$ since not all fractions are measured).

Systematic errors that *would have occurred* if no corrections had been applied were calculated in accordance with Bel *et al.* [17]: if, prior to measurement j, n corrections c have been detected with magnitudes δ_c, the "uncorrected" setup error $\widetilde{\varepsilon}_{ij}^T$ is calculated from

$$\widetilde{\varepsilon}_{ij}^T = \varepsilon_{ij}^T - \delta_n \, , \text{ with } \delta_n = \sum_{c=1}^{n} \delta_c \, .$$

Similar to the definition of S_i^T, we define the systematic error of patient i for the "uncorrected" setup by $\widetilde{S}_i^T = N_{f,i}^{-1} \sum_{j=1}^{N_{f,i}} \widetilde{\varepsilon}_{ij}^T$ (note that in this case no time-weighed average is necessary, because if no setup corrections would have taken place, each measurement has an equal weight) and $\widetilde{\mu}^T = N_p^{-1} \sum_{i=1}^{N_p} \widetilde{S}_i^T$.

The calculation of $\widetilde{\varepsilon}_{ij}^T$ enables us to derive the random (inter-fraction) positioning error from data obtained with a correction protocol. We assume that the random errors are not affected by the corrections, which is reasonable since the corrections are applied as table shifts *after* the patient is positioned using skin marks. The random errors introduced by setting up on the skin marks and performing the template match do not depend on the table shift. Inaccuracies in the table shift are neglected, as motivated in the Methods section. We may then estimate the random error for patient i, σ_i^T, from

$$\sigma_i^T = \left[(N_{f,i} - 1)^{-1} \sum_{j=1}^{N_{f,i}} (\widetilde{\varepsilon}_{ij} - \widetilde{S}_i^T)^2 \right]^{\frac{1}{2}} \tag{A4.2}$$

The *population* random error, σ^T, was calculated from $\sigma^T = \left[N_p^{-1} \sum_{i=1}^{N_p} (\sigma_i^T)^2 \right]^{\frac{1}{2}}$ [123].

The *population* systematic error without a correction protocol is defined as the SD of the patient systematic errors if no correction had been applied:

$$\widetilde{\Sigma}_{bias}^T = \left[(N_p - 1)^{-1} \sum_{i=1}^{N_p} (\widetilde{S}_i^T - \widetilde{\mu}^T)^2 \right]^{\frac{1}{2}} \tag{A4.3}$$

Due to the rather limited number of observed fractions per patient (≤ 10), this estimator is slightly biased and will lead to an overestimation of the systematic errors due to the presence of random errors. With a little algebra, it can be derived that an unbiased estimator of the variance of the systematic errors is

$$(\widetilde{\Sigma}^T)^2 = (\widetilde{\Sigma}_{bias}^T)^2 - N_p^{-1} \sum_{i=1}^{N_p} N_{f,i}^{-1} (\sigma_i^T)^2 \tag{A4.4}$$

and this expression was used to calculate $\widetilde{\Sigma}^T$. In a similar fashion, Σ^T was calculated as the SD of the systematic errors after correction (S_i^T). Since each S_i^T is a weighed mean (Eq. A4.1a), the correction term for random errors is slightly more complicated, but performing the algebra it can be shown that the correction in Eq. (A4.4) is still a good approximation and we applied this same expression (note however, that this term is small).

The error ε_{ij}^T derived from a portal image measures the setup error relative to the reference image. It is this error which is usually reported in papers on positioning accuracy. However, for patients with CT planning we should measure errors with respect to the planning CT scan. This is the case if the reference image is a DRR. When the reference is a simulator film, then the systematic errors are underestimated. In this case, the actual setup error is the sum of ε_{ij}^T and the error in the reference setup for patient i, S_i^R, i.e. the setup error during simulation. If we adopt the above definitions for μ and Σ with T replaced by R and $\sigma^R \equiv 0$ (since we use one reference image per treatment portal) then the actual (μ, Σ, σ) are given by $(\mu^R + \mu^T, \sqrt{(\Sigma^T)^2 + (\Sigma^R)^2 + 2\mathrm{cov}(S^T, S^R)}, \sigma^R)$ where the covariance $\mathrm{cov}(S^T, S^R)$ accounts for potential interdependencies of S^T and S^R.

5 Procedures for high precision setup verification and correction of lung cancer patients using CT-simulation and digitally reconstructed radiographs (DRR)

J.R. van Sörnsen de Koste, J.C.J. de Boer, R.H. Schuchhard-Schipper, S. Senan and B.J.M. Heijmen

Abstract

Purpose: In a recent study, large systematic setup errors were detected in patients with lung cancer when a conventional simulation procedure was used to define and mark the treatment isocenter. In the present study, we describe a procedure to omit the session at a conventional simulator to remove simulation errors entirely. Isocenter definition and verification was performed at a CT-simulator and digitally reconstructed radiographs (DRRs) were used for setup verification and correction at the treatment unit.

Methods and Materials: A CT-simulation protocol was developed, in which radio-opaque markers were used to verify the coincidence of the isocenter marked on the patients' skin with the isocenter defined in the planning CT-scan. This protocol was evaluated for twenty patients. Subsequently, electronic portal images were acquired at the treatment unit. The three-dimensional (3D) setup error was established from a template match of the appropriate anatomy visible in two orthogonal beams with the corresponding anatomy in DRRs. An off-line setup correction protocol was applied to reduce systematic setup errors.

Results: For all patients, the skin marks defined the planning CT-scan isocenter to within ± 1.5 mm in each of the three main directions. Random setup errors at the treatment unit were 1.8, 2.0, and 1.9 mm (1 SD) for the lateral (x), the superior-inferior (y), and the anterior-posterior (z) directions, respectively. With the use of the correction protocol, the systematic errors for x, y and z were 1.5, 1.5 and 1.3 mm (1 SD).

Conclusions: As the distributions of treatment setup errors measured against DRRs obtained in our CT-simulation were equal to previously obtained distributions measured against simulator films, conventional simulation can be omitted and DRRs are well-suited for setup verification. By adopting our CT-simulation procedure the large systematic simulation setup errors, that remain hidden if a conventional simulation is performed, can be avoided.

5.1 Introduction

The local tumor control achieved with external beam radiotherapy (RT) for unresectable non-small cell lung cancer (NSCLC) patients still remains poor, with a complete pathological response achieved in only 20% of the patients treated to 65Gy [33]. Data from patients treated with 3D conformal radiotherapy suggest that a dose of at least 84Gy may be required in order to increase the probability of local control to over 50% [81]. Higher radiation doses may increase the risk of radiation pneumonitis and the aim of this present study was to evaluate steps to reduce the Clinical Target Volume to Planning Target Volume (CTV-to-PTV) planning margins for lung cancer patients, in order to reduce the irradiation of healthy lung tissue.

Nowadays, fast treatment planning systems (TPS) enable the generation of 3D treatment plans. In order to ensure dose delivery according to the plan, Electronic Portal Imaging Devices (EPID) are used to verify the patient setup [24, 27, 28, 72, 105, 112, 127]. With the use of an off-line setup correction protocol, the systematic and random setup inaccuracies of lung cancer patients relative to the reference setup (as defined by simulator films) are within 2 mm, 1 Standard Deviation (1 SD) [28, 112]. However, de Boer et al. [28] showed that large deviations (up to 1 cm) may occur between simulator films and digitally reconstructed radiographs (DRRs) for these same patients. Therefore, the setup at the simulator (where skin marks for treatment setup were placed and corresponding reference setup films were obtained) deviated significantly from the setup prescribed by the treatment plan. For a group of 39 lung cancer patients, the standard deviations of these simulator setup errors were 4.0 mm, 2.8 mm and 2.5 mm for the three orthogonal directions. Consequently, if marks placed during simulation are used during treatment setup, the simulator setup errors will cause large *systematic* (i.e., recurring each fraction) setup errors relative to the intended treatment plan [28]. These errors will persist if treatment setup verification and correction is based on simulator films, which is regular clinical practice. Similar findings, though smaller in magnitude, were found for prostate cancer patients [12].

In a previous study [28], we analyzed the effect of an off-line setup correction protocol on the setup accuracy in lung cancer patients. That study was performed using simulator films as reference images because (i) high quality DRRs were not available on a routine basis, (ii) our CT scanner did not allow for easy virtual simulation and most importantly (iii) the awareness of the size of simulator setup errors grew during the course of the study. Since then, an advanced CT-simulator has become available, which provides excellent DRRs and allows for isocenter definition and demarcation directly after the planning CT has been obtained. Therefore, omission of a conventional simulation session became feasible and simulator setup errors could potentially be eradicated. In the present work, we study a method to achieve that goal, consisting of the following steps:

1. A CT-simulator procedure was developed to ensure accurate geometrical alignment of the planning CT-scan coordinate system with tattoos placed on the patients' skin during the same CT session. The CT and treatment planning procedures ensured that the same tattoos could be used for setup at the treatment unit, and so no conventional simulation session was necessary.
2. At the treatment unit, setup verification and correction was based entirely on DRRs generated from the planning CT-scan.

If the systematic and random setup errors thus obtained at the treatment unit are as small as when measured against simulation films, this would imply that conventional simulation is

unnecessary in lung cancer patients and consequently that simulation setup errors, an important source of systematic setup errors in these patients, can be avoided.

5.2 Materials and methods

5.2.1 Patients

Twenty patients participated in the evaluation of our procedure. Of these, sixteen had non-small cell lung cancer (NSCLC), three had small cell lung cancer (SCLC), and one patient had a primary carcinoma of the esophagus. All patients were irradiated with a 3-field, coplanar isocentric technique. A total dose of 50-70 Gy was delivered in 20 to 35 fractions. Treatments were performed at a Scanditronix MM50 Racetrack Microtron. At both the CT-simulator and treatment unit, patients were immobilized in the supine treatment position using an arm base rest device (Sinmed BV, The Netherlands) and knee and foot brace devices.

5.2.2 Protocol for acquisition of the planning CT-scan, and definition and verification of the isocenter at the CT-simulator

Acquisition of the planning CT-scan and subsequent treatment isocenter definition and verification were performed at a Picker CT-simulator (PQ5000 system). The patient was positioned and immobilized as during treatment (see above). First, using the lateral and anterior lasers, lines were drawn on the patients' skin. These lines were used for visual inspection of the patient setup on the couch throughout the procedures at the CT-simulator. An anterior topogram was then made to establish the longitudinal scan area, followed by acquisition of the planning CT scan. Spiral CT-scanning was used with a table speed of 2 mm/sec and a revolution time of 1 sec. Next, while the patient remained on the scanner couch, transversal slices were reconstructed with a thickness of 3 mm, the distance between reconstructed slices was 2.5 mm (slice index = 2.5 mm). The radiation oncologist defined the treatment isocenter in the planning CT scan using the Picker virtual simulation software by selecting a position in the clinical target volume (CTV) close to its expected center of mass, followed by a movement of the couch and the top laser such that the treatment isocenter was defined by the laser lines. Using these laser lines, the isocenter was demarcated on the patients' skin with ink markings.

Subsequently, as a final check of the coincidence of the isocenter definition on the patients' skin (to be used for setup of the patient at the treatment unit) with the isocenter as defined in the planning scan (the basis for the design of the treatment plan), the following procedure was performed. A 1 mm diameter radio-opaque marker was put on the two lateral and the anterior ink markings. Three CT slices were then obtained, one at the level of the markers, and one on both sides, at a distance of 2.5 mm from the central slice. Next, the CT-simulator software was used to verify that the isocenter cross-hairs, as defined by the radiation oncologist in the previously acquired planning CT-scan (see above), did indeed agree with the positions of the radio-opaque markers in the central verification slice. Moreover, the central slice in the planning CT-scan was visually compared with the CT-slice containing the markers. Finally, tattoos were placed on the patients' skin at the positions of the ink markings/radio-opaque markers for a definitive demarcation of the

isocenter. The overall procedure at the CT-simulator required approximately twenty minutes.

During subsequent treatment planning, the CTV was delineated and it was found that its center of mass was in general close to the isocenter defined above. Furthermore, this isocenter was always located within the CTV volume and could in all cases be maintained for treatment.

5.2.3   The reference for setup verification and correction

For each patient, two portals were used for setup verification. The first beam had an angle of incidence that deviated not more than ± 15º from the vertical direction, the other beam not more than ± 15º from the horizontal direction. In previous studies [28, 112] we identified anatomical structures that (i) exhibited negligible intra-fraction movement when projected in a transmission image (1 SD = 0.5 mm, including delineation errors) and (ii) could be delineated with small inter-observer variation (1 SD < 1 mm). In the present study we used these structures for setup definition. For the anterior portal we took the combination of the projections of the thoracic wall and the trachea. In the lateral field, the ventral side of the spine vertebrae, the sternum and sometimes the trachea were used (Fig. 5.1).

The portals to be used for setup verification were selected at the TPS (Cadplan, Varian-Dosetek, Finland). In order to include the above-mentioned anatomical structures, verification portals were always larger than the corresponding, overlapping treatment fields. Moreover, the nearly orthogonal beam arrangement as described above could not always be realized using only the treatment beam directions. In those cases, setup verification required an extra, dedicated portal with a different angle of incidence.

The definitions of the established verification fields were electronically sent to the CT-simulator software for generation of DRRs with a resolution of 0.5 x 0.5 mm^2 at isocenter
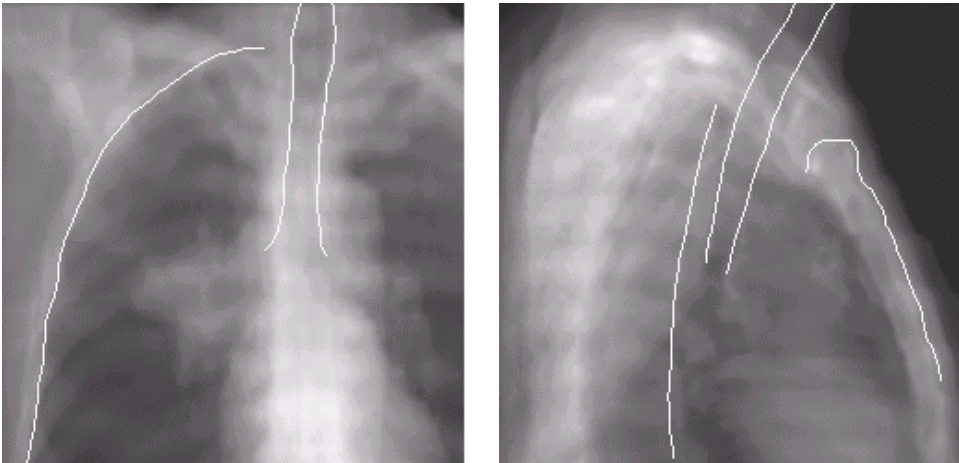
Figure 5.1 Examples of the used anterior (left panel) and lateral (right panel) DRRs. The white lines are the stable anatomical structures used as reference for setup verification (section 5.2.3).

distance. The spatial resolution of the planning CT-scans used for calculation of DRRs was 0.94 x 0.94 x 2.5mm^3. The DRRs were loaded into the EPID software (next section) and the selected anatomical structures were manually delineated according to delineation prescriptions (i.e. example images with marked reference structures) presented by this software.

5.2.4 Patient setup verification and correction

The patient setup during treatment was verified using a CCD-camera based EPID (Theraview-NT, manufactured by Cablon Medical, The Netherlands). This system has been described in detail elsewhere [26, 28, 90].

At the treatment unit, image acquisition with the EPID, image analysis, establishment of the appropriate setup corrections, and execution of patient setup corrections were performed by the technicians, using an unambiguous protocol integrated in the EPID software.

Prior to the image analysis by a technician, the field edge of an acquired portal image was automatically extracted and matched with the prescribed field as defined in the planning system, in order to define a common co-ordinate system (this match also constituted an independent verification of the delivered treatment field shape). Then, the anatomical structures as delineated in the corresponding DRR (previous section) were mapped onto the EPID image and shifted with the mouse to establish the optimal match
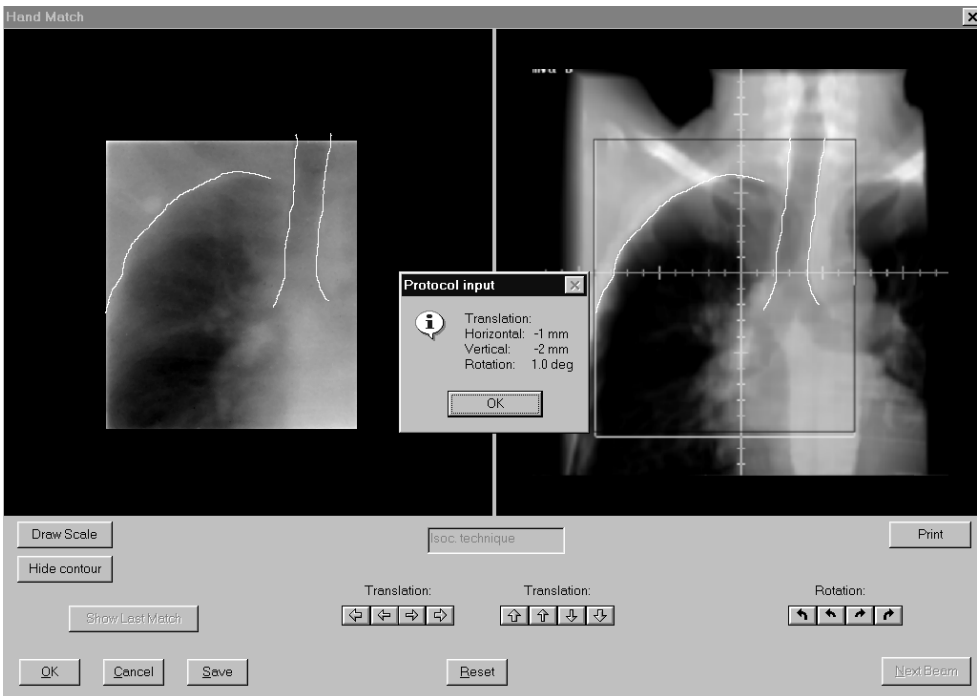
Figure 5.2 Overview of the match screen. The patient setup accuracy is shown here in terms of deviations between the AP portal image (left image) and the corresponding reference DRR (right image).

(template matching; Fig. 5.2). To improve the match, the technicians were allowed to also rotate the mapped reference structures around the isocenter, but were instructed to use this option with caution and only if they could not reach a satisfactory result without it.

Image analysis was performed after completion of the daily treatment fraction, i.e. off-line. The established anatomy translations in the two orthogonal verification portals of a patient determined the 3-dimensional (3D) patient setup error for that day. This setup error was used to drive a 3D shrinking action-level protocol [13, 17, 27, 28] to decide on the setup correction to be applied in the *remaining* treatment fractions. This protocol, which is integrated in the EPID software, aims at a reduction of systematic setup errors. The applied protocol had an initial action level $\alpha = 8$ mm and 3 subsequent measurements in the first stage ($N_{max} = 3$). The final 3D systematic error obtained with this protocol should be smaller than $\alpha / \sqrt{N_{max}} = 4.6$ mm. Details can be found in [28].

For easy and safe execution of setup corrections, the required couch translations as calculated by the EPID software were expressed in the treatment unit co-ordinate system. This software option is available for any treatment unit.

For each treatment fraction, the patient was first positioned such that the laser lines coincided with the tattoos on the patients' skin, as placed at the CT-simulator (section 5.2.2). Then, the required 3D setup correction was performed by shifting the treatment couch accordingly, followed by irradiation. We verified that these corrections could be performed accurately (1 SD < 0.5 mm), and with negligible increases in treatment time [28].

For each patient, a separate treatment plan was made to estimate the extra dose delivery due to the use of enlarged or additional portals for setup verification (section 5.2.3). The verification portals were acquired with 6 MU, which was the minimum allowed exposure at the treatment unit at the clinical doserate. The additional dose delivery due to application of the verification portals in 10 fractions was calculated, as this was estimated to be the numbers of fractions required to complete a lung setup protocol [28]. Typically, the extra-added dose in 10 fractions was about 1 Gy at isocenter level. The corresponding dose distribution was added to the cumulative planned dose distribution, and dose-volume-histograms of this summed distribution were evaluated by the radiation oncologist.

5.3 Results

The observed setup variations are summarized in Table 5.1. All setup errors were obtained from matching EPID images with the corresponding DRRs. For each patient, the magnitude of the random setup errors was characterized by the SD of the day-to-day (i.e. interfraction) variations in the setup errors. The average SD of the population was calculated as described in [28] to yield σ, which is a representative measure of the random setup errors. The values of σ thus found were 1.8, 2.0, and 1.9 mm for the lateral (x), the superior-inferior (y) and the anterior-posterior (z) directions, respectively. Furthermore, for each patient the systematic setup error was calculated as the mean of the fraction-specific setup errors observed in that patient over the course of treatment. The distribution of systematic setup errors in the patient group was summarized through μ, which is the mean of the systematic errors, and Σ, defined as the SD of those same systematic errors. After application of the correction protocol, we found that $\mu \leq 0.6$ mm for each of the three directions whereas Σ was only 1.5, 1.5 and 1.3 mm in the x, y and z directions respectively. The average number of setup corrections per patient was 0.8.

Using previously described methods [28], we calculated that the systematic errors Σ that would have been measured if no setup corrections had been applied, were 2.5, 3.4 and 2.4 mm (1 SD) in the x, y and z directions respectively while the mean systematic error μ remained < 1 mm in each direction. These relatively small systematic errors obtainable without setup corrections are consistent with our observations at the CT simulator: for all patients in this study, the radio-opaque markers used during CT-simulation to verify the accuracy of the isocenter demarcation on the patients' skin, corresponded to within ± 1.5 mm with the isocenter in the planning CT-scan as defined by the radiation oncologist. The last two columns in Table 5.1 show that the observed rotational setup variations were small. These values are independent of the application of the protocol, as no attempt was made to correct for rotations.

Table 5.1 Patient setup errors observed at the treatment unit. The x-, y-, and z-axes correspond to the lateral, the superior-inferior, and the anterior-posterior directions. AP refers to rotations around the anterior-posterior axis, and RL to rotations around the lateral axis. The applied correction protocol only influences the systematic errors in the x-, y- and z-directions, not the random errors, nor the rotations.

		X (mm)	Y (mm)	Z (mm)	AP(°)	RL (°)
No protocol	$\widetilde{\mu}^{T}$	-0.3	0.0	-0.9	0.2	0.1
	$\widetilde{\Sigma}^{T}$	2.5	3.4	2.4	0.9	0.6
	σ^{T}	1.8	2.0	1.9	1.3	1.0
With protocol	μ^{T}	0.0	-0.6	-0.3		
	Σ^{T}	1.5	1.5	1.3		

5.4 Discussion

We previously studied the setup of 40 lung cancer patients, using digitized simulator films as reference images instead of DRRs [28]. In that simulator-film-based study, the final isocenter demarcation on the patients' skin was performed at the conventional simulator, using standard CT scout topograms and beam's-eye-view (BEV) plots, as derived from the planning CT-scan and the treatment plan. The random setup errors (σ) thus obtained were 2.0, 2.1, and 1.8 mm (1SD). The systematic setup deviations (Σ) obtained with the application of the correction protocol were 1.3, 1.6, and 0.9 mm, and they would have been 3.2, 3.6 and 1.7 mm (1SD) if no protocol had been applied. These results are consistent with the numbers in Table 5.1, which were measured relative to DRRs. Particularly the random setup errors are almost identical, and the small differences between the values of Σ are statistically insignificant ($p > 0.05$, F-test). This close correspondence is an important result, because no patients from the previous study were included in the current study. We therefore conclude that (1) setup errors may be measured equally accurate relative to DRRs as to simulator films and (2) the isocenter definition procedure at the CT scanner is accurate

(otherwise, larger systematic errors would be expected). Therefore, the simulator setup errors can be avoided entirely and DRRs constructed as described in section 5.2 can (and should) replace simulator films as reference images. This way, the setup accuracy as perceived at the treatment unit will correspond to the true setup accuracy (apart from observer related errors).

Stroom et al. [125] have derived that the required (CTV-to-PTV) planning margin M to account for geometrical uncertainties may be calculated with

$$M = 2\Sigma + 0.7\sigma \qquad (5.1)$$

with Σ the overall standard deviation describing all *systematic* uncertainties, and σ the standard deviation describing all day-to-day variations in the treatment geometry. In case there are different independent sources i of geometrical uncertainties, each described by separate standard deviations Σ_i and σ_i, the overall standard deviations Σ and σ, used for calculation of the margin M with equation (5.1) are

$$\Sigma = \sqrt{\sum_i \left(\Sigma_i^2\right)} \quad \text{and} \quad \sigma = \sqrt{\sum_i \left(\sigma_i^2\right)} \qquad (5.2)$$

Generally, Σ and σ have contributions from patient setup inaccuracies and from uncertainties in the internal anatomy (organ motion) of the patients.

Both for simulator films and DRRs, the Σ for setup at the treatment unit should be included in the calculation of the overall Σ (equation (5.2)). For setup verification based on simulator films, the standard deviations describing the simulator setup errors, denoted as Σ^R, should also be included in Σ. The Σ^R-values found by de Boer et al. [28], $\Sigma^R = (4.0, 2.8, 2.5)$ mm for the (x,y,z) directions, are substantially larger than the treatment setup Σ that are obtained with a setup verification and correction protocol (Table 5.1). In equation (5.2), standard deviatons Σ_i, are quadratically added. Therefore, with the procedure applied in that study for reference definition with simulator films, the contribution of setup errors to the margin M would be dominated by the Σ^R. With the use of DRRs as described in the current paper, the Σ^R are zero.

To describe the internal displacement of thoracic tumors due to breathing and cardiac activity in terms of margins is a complex subject, beyond the scope of this paper. Although intrafraction excursions of tumors may range up to 2 cm, particularly when located close to the heart and aorta or diaphragm [110], in many cases the displacements are much smaller. For instance, Ekberg *et al*. [40] studied tumor mobility with the aid of fluoroscopy in a group of 20 patients with tumor locations spread over left and right lung and all 3 lobes and found that the average SD of the tumor displacement in the (x,y,z) directions was respectively 1.4, 2.6 and 1.3 mm during normal respiration. Recently, these results were confirmed in a 3D study of markers implanted in or near lung tumors [115]. These markers were imaged using a fluoroscopic real-time tumor tracking system. A time analysis of marker movements in 20 patients revealed major displacements of up to 2 cm in the cranial-caudal (z) direction for lower lobe tumors not attached to rigid structures, but significantly smaller movements (typical amplitudes 1-3 mm) for the other (x,y) directions as well as for the z-direction when attached to a rigid structure such as the chest wall. When heartbeat was separated from breathing motion by Fourier analysis, amplitudes of 1-2 mm (max. 4 mm) at frequencies > 1 Hz were found. The systematic errors introduced by internal motion will in general be smaller than the amplitudes given above, as these amplitudes describe the maximum systematic errors that could occur. Furthermore, due to the finite revolution time

of a typical CT scanner of ~ 1 sec the high frequency heartbeat motions are partly averaged out. Therefore, for many tumors the systematic simulator setup errors, with the SD values of 3-4 mm mentioned above (yielding extremes of 1 cm in each direction) are relatively large compared to the systematic errors induced by internal motion, particularly when considered in 3D. Consequently, techniques to reduce internal motions are only effective if simulator setup errors can be removed.

Nevertheless, for lung tumors, internal organ motion must of course be taken into account when defining the PTV. To avoid the problem of translating internal motion to margins in a practical way, we have investigated the use of multiple "slow" planning CT-scans (CT revolution period 4 seconds) for lung cancer patients to construct a composite target volume that should account for the respiratory and cardiac motions [71, 121]. Assuming that with this technique internal motion effects are indeed largely accounted for, we have below estimated how the treatment volumes of the ten patients in the study of Lagerwaard et al. [71] would increase with the use of simulator films as the reference for setup verification instead of DRRs. To account for subclinical disease and setup uncertainty, the composite target volumes were automatically extended [125, 131] with a 3D margin. The margin component related with subclinical disease was assumed to be 5 mm. The contribution of setup uncertainties was calculated with equation (5.1) as described above. It follows that, for planning margin calculations, the Σ values with application of the setup correction protocol (Table 5.1) are the most important data. As can be seen from the last row in this Table, these standard deviations are all small. Therefore, using the data mentioned in Table 5.1, the margins related to setup uncertainties are 4, 5, and 4 mm, for the x, y and z-axis, respectively. In case setup verification for these patients would have been based on simulator films, acquired with the procedure described in de Boer et al. [28], the simulator setup errors become the predominant factor, yielding margins of 10, 8 and 7 mm, respectively. For the ten patients in the study described by Lagerwaard et al. [71], setup verification based on simulator films instead of DRRs would have resulted in an average increase in the PTV of 54.4 % ± 9.8 % (1SD).

5.5 Conclusions

We have developed and applied a procedure to obtain reproducible and accurate setups of lung cancer patients at the treatment unit, thereby reducing the contribution of setup uncertainties to required planning margins. *Systematic* setup errors were reduced in a two-step approach. First, acquisition of the planning CT-scan and definition and verification of the isocenter were performed at a CT-simulator, thereby removing the conventional separate simulation procedure. As a result, the (systematic) errors between the intended setup defined by the treatment plan and the setup obtained at the simulator were removed. Second, portal images were acquired with an EPID, and an off-line correction protocol was used to reduce the systematic setup errors. Setup verification was based on anatomical reference structures in DRRs. The standard deviations Σ, describing the remaining systematic uncertainties in the setup were only 1.3-1.5 mm for the three main directions. Day-to-day variations in the patient setup were approximately 2 mm (1 SD). These values were identical to previous setup measurements performed relative to simulator films, indicating that DRRs obtained with commercially available software can replace simulator films in setup evaluation. Consequently, conventional simulation is not required for setup

definition or verification in lung cancer patients and should therefore be omitted to avoid systematic simulation errors.

Acknowledgements

6 A protocol for the reduction of systematic patient setup errors with minimal portal imaging workload

J.C.J. de Boer and B.J.M. Heijmen

Abstract

Purpose: The evaluation of a new off-line patient setup correction protocol which minimizes the required number of portal images, and a comparison with currently applied protocols.

Methods and Materials: We compared two types of off-line protocols: (i) the widely applied shrinking action level (SAL) protocol, in which the setup error, averaged over the measured treatment fractions is compared with a threshold that decreases with the number of measurements to decide if a correction is necessary and (ii) a new 'no action level' (NAL) protocol, which simply calculates the mean setup error over a fixed number of fractions, and always corrects for it. The performance of the protocols was evaluated by applying them to (i) a database of measured setup errors from 600 prostate patients (with on average 10 imaged fractions/patient) and (ii) Monte Carlo generated setup error distributions for various values of the population systematic and random errors.

Results: The NAL protocol achieved a significantly higher accuracy than the SAL protocol for a similar workload in terms of image acquisition and analysis, as well as in setup corrections. The SAL protocol required approximately 3 times more images than the NAL protocol to obtain the same reduction of systematic errors. Application of the NAL protocol to measured setup errors confirmed its efficacy in systematic error reduction in a real patient population.

Conclusion: The NAL protocol performed much more efficiently than the SAL protocol for both actually measured and Monte Carlo simulated setup data. The resulting decrease in required portal images does not only reduce workload, but also dose to healthy tissue if dedicated large fields are required for portal imaging (double exposure).

6.1 Introduction

Information on the setup of the patient anatomy with respect to the treatment beams as derived from portal images may be used in a variety of ways to improve the accuracy of that setup [116]. Whatever approach is adopted, it is important to distinguish between systematic (occurring each fraction) and random (different for each fraction) patient setup errors [22]. In fact, detailed calculations based on dose coverage of the clinical target volume (CTV) show that if the standard deviations (SD) of the systematic errors (SD = Σ) and of the random errors (SD = σ) are known, then the margin M to expand the CTV to a safe (in terms of low probability of underdosage) planning target volume may be expressed by $M = 2\Sigma + 0.7\sigma$ [125]. This expression for M confirms that planning margins can be reduced predominantly through reduction of systematic errors.

In this paper, we concentrate on patient setup verification protocols that aim at reducing the systematic errors. These 'off-line' verification protocols (OVPs) take into account measurements of the previous and current fractions, and generate a decision (to correct or not, and if so, by how much) for the setup of *subsequent* fractions [116]. The OVP is only efficient if the number of measurements required for a sufficiently accurate estimate of the systematic error is significantly smaller than the total number of treatment fractions. The off-line approach excludes the possibility to correct for the daily random setup errors. The latter can only be achieved through an 'on-line' protocol: in each fraction, an image is obtained with a few accelerator monitor units exposure and analysis is performed immediately after image acquisition. If necessary, a correction is directly applied before the remaining fraction dose is delivered [91, 124, 130].

Traditionally, off-line setup corrections are based on images obtained in a single (often first-day) fraction (see e.g. [20, 49, 113, 120, 133, 134]). These images are sometimes only judged by eye without clearly defined quantitative criteria for corrective actions [20, 113, 133], and both the decision criteria as well as the scheme of follow-up measurements vary appreciably from clinic to clinic. A more rational and well-defined approach was introduced by Bel *et al.* [13], who proposed the 'shrinking action level' (SAL) protocol. To date, the SAL protocol is the only OVP based on quantitative measurements which has been consistently applied for large patient groups in various hospitals [13, 17, 28, 103, 138, 139]. This OVP has also been used in our clinic for pelvic, thoracic and head-and-neck treatments, in over 2000 patients. Clinical introduction of EPIDs together with the SAL protocol has been essential in transferring the routine decision making process on patient setup correction from radiation oncologists to technologists. Thereby, accurate positioning verification and correction became available to a large patient population. However, although the application of the SAL protocol is well accepted in our clinic, we have also ran into a number of limitations. We found that the SAL OVP required quite a large number of imaged fractions (typically 10 per patient [17, 28]) to reduce the systematic dispersion Σ by 50%. While this result is consistent with numerical simulations, it is much larger than one would expect from the finally reached accuracy and the magnitude of the random setup errors ($\sigma \approx \Sigma$). The large workload thus generated by the SAL OVP has inhibited routine application of the protocol on most of our accelerators. In addition, particularly due to the growing application of conformal treatment techniques, portal images must often be acquired with dedicated portal imaging fields (PIFs, usually larger than the treatment fields and sometimes at different gantry angles) to render useful anatomy visible. Such fields imply additional dose to healthy tissue, and it is therefore essential to keep the number of

acquired images as small as possible. In our high dose conformal treatments, the cumulative dose of the PIFs is taken into account in the treatment planning to verify that for organs at risk the tolerance dose is not exceeded. Because of the above reasons, we started investigating an OVP which requires as few portal images as possible to reach a specified final distribution of systematic errors, and which preferably reaches such a distribution for a fixed number of exposures per patient so that the dose to non-target volumes is fully predictable (this is not the case for the SAL protocol).

The problems with the SAL protocol largely arise from the use of action levels. Such action levels introduce a sub-population of patients who do not get a correction, i.e. patients for which the estimate of the systematic error is sufficiently small. This would be considered an advantage in case correction of the patient setup would significantly contribute to the overall workload introduced by the OVP [13]. However, for treatment couches with accurate digital read-out or even remote software control, such corrections can be accurately executed within seconds and the workload is largely determined by the number of images to be acquired and analyzed. In fact, already in the prostate patient setup study of Bel *et al.* [17] it was reported that in two institutions who applied setup corrections by table translation (i.e., the Dr. Bernard Verbeeten Institute in Tilburg and our hospital) the execution of the setup corrections hardly prolonged treatment time. An alternative approach to executing setup corrections was proposed by Yan *et al.* [145, 146], who used a multileaf collimator (MLC) to correct for systematic setup errors. Although this procedure involved reshaping the MLC via an MLC preparation system and subsequent data transfer to the accelerator, this method assures no increase in daily workload at the accelerator.

Therefore, the alternative OVP we propose applies no action level (NAL), i.e. every patient has one setup correction and for each patient, the same number of images is used to estimate the systematic setup error. In this paper, we compare NAL and SAL protocols with respect to both the achieved patient setup accuracy and implied workload.

6.2 Methods and Materials

6.2.1 Definition of relevant quantities

All definitions below are described more precisely in the Appendix. In this section we only briefly introduce the most relevant quantities and terms.

For a patient p and for treatment fraction $f \in F = \{1, 2, ..., N_f\}$ the set-up deviation with respect to the reference set-up (usually defined by simulator films or digitally reconstructed radiographs) along a certain direction is $x_{p,f}$. We can now write $x_{p,f} = S_p + \Delta_{p,f}$ where S_p is the systematic error and $\Delta_{p,f}$ the random (fraction-to-fraction) error for patient p. The effective systematic error $S_{p,\text{eff}} = < x_{p,f} >_F$, (the brackets indicate an *average* over the fractions, see Appendix) differs slightly from S_p because of the random setup errors and the finite number of fractions N_f. The random error of p is characterized by σ_p, which is the SD over f of $\Delta_{p,f}$ [123].

If the applied setup corrections are $c_{p,f}$, the actual setup deviations become $x_{p,f}^{\text{act}} = x_{p,f} + c_{p,f}$. Since the corrections $c_{p,f}$ are known for each patient, we can obtain

the deviations without correction ($x_{p,f}$) directly from the measured deviations ($x_{p,f}^{\text{act}}$) apart from inaccuracies in the performed corrections and measurement inaccuracies. Below and in the Appendix we argue why these errors are small, but general expressions including all terms are given in the Appendix.

Bijhold *et al*. [22] introduced the use of the symbols Σ and σ to denote SDs of respectively systematic and random deviations, and we adopt their convention. We describe the setup characteristics of a patient population P through the *mean systematic error* μ, the *systematic dispersion* Σ and the *mean random error* σ_r, defined by

$$\mu = \left\langle S_p \right\rangle_P, \quad \Sigma = \text{SD}_P(S_p), \quad \sigma_r = \sqrt{\left\langle \sigma_p^2 \right\rangle_P} \tag{6.1},$$

where the subscript P indicates that these quantities are calculated over the patient population.

In practice, the dispersion of systematic errors is slightly larger than the dispersion defined in Eq. (6.1) due to random setup errors, and the *effective* systematic dispersion becomes

$$\Sigma_{\text{eff}} = \text{SD}_P(S_{p,\text{eff}}) = \sqrt{\Sigma^2 + \sigma_r^2 N_f^{-1}} \tag{6.2}.$$

Usually, $N_f \geq 30$, so $\Sigma_{\text{eff}} \approx \Sigma$. In the Appendix, we give complete analytical expressions for these dispersions if the NAL protocol (defined below) is applied.

In agreement with our clinical experience [17, 28], we will assume that $\mu = 0$, i.e. there is no net discrepancy between the intended treatment setups and the actual setups. Although OVPs will also reduce μ if non-zero [13], a mean discrepancy is the result of a systematic procedure error in the radiotherapy treatment chain and should be resolved independent of the OVP. In contrast, the systematic and random setup errors may always occur, even in case of a perfect mutual alignment of diagnostic scanners, simulator and accelerator and therefore require portal imaging to correct for.

The definitions given above pertain to any direction. Specific directions (X_1, X_2, X_3) will be indicated by subscripts 1, 2 or 3.

6.2.2 Setup correction procedure

To properly predict the results of an OVP, it is important to describe how setup corrections are performed. In our clinic, in each fraction the patient is initially positioned according to the tattoos and marked laser lines established in the treatment preparation phase. Next, if a setup correction is required, a table shift is performed using the digital read-out of the treatment couch. If multiple corrections are detected, as is possible with the SAL protocol, the last correction is added to the sum of the previous corrections to give the net correction with respect to the initial setup. In this setup procedure, no new lines or tattoos are applied, which greatly reduces the potential for errors due to misinterpretation of patient marks. Furthermore, this procedure allows for retrospective calculation of setup errors that would have been obtained without setup corrections (see *Definition of relevant quantities*), since the accuracy with which the patient is initially positioned using tattoos is independent of whether a correction is subsequently applied or not. The precision of such a retrospective calculation is therefore only limited by the accuracy with which the prescribed couch shifts were performed, as will be discussed below.

6.2.3   Verification protocols

6.2.3.1 The shrinking action level (SAL) protocol

The properties of the SAL protocol were investigated in detail, yet in only one dimension, by Bel *et al*. [13] using Monte Carlo simulations. A 3D version of the SAL protocol was applied in a multi-center study on the impact of this OVP on the setup accuracy of prostate cancer patients [17]. The protocol in its originally proposed form has two parameters, the initial action level α (mm) and the maximum number of subsequent measurements without a correction, $N_{\max}$. From measurement $N = 1$ onwards, the setup error is averaged over the last N measurements for each measured direction, yielding a mean error vector V_N of length d_N. This length is tested against the action level $\alpha_N = \alpha / \sqrt{N}$. The decreasing action level α_N reflects the increasing accuracy with which the systematic error is measured by d_N. If $d_N > \alpha_N$, a setup correction equal to $-V_N$ is applied in the following treatment fractions and the protocol is restarted completely with N set to 1. If, for $N_{\max}$ subsequent measurements, the condition $d_N \leq \alpha_N$ applies, the protocol has finished. We denote the number of imaged fractions per patient, averaged over the entire patient population, by N_{meas}. For any SAL protocol that significantly reduces systematic errors (i.e., yields corrections for most patients), N_{meas} is appreciably larger than $N_{\max}$.

   The procedure described above is sometimes referred to as the first stage of the protocol [17] to contrast it with a second stage. The second stage is aimed at detecting errors which grow with time during the course of treatment and various sites have adopted different second stage protocols [17, 138]. In this paper we will only deal with the first stage of the protocol.

   Due to the iterative nature of the SAL protocol, its results and workload cannot be predicted analytically but only through numerical simulations [13]. Such simulations require *a priori* knowledge of the distributions of systematic and random errors. As described above, the protocol allows for multiple correction detections per patient. A frequent occurence of multiple corrections points at an inefficient estimation of the systematic errors and inappropriate SAL parameters.

6.2.3.2 The no action level (NAL) protocol

The NAL protocol we propose here is very simple to conceive. It has only one parameter, which is the (fixed) number of fractions to be measured per patient (N_m, $<< N_f$). The average setup vector over the first N_m fractions (V_{N_m}) is determined, and a setup correction equal to $-V_{N_m}$ is applied in all subsequent fractions, irrespective of its length d_{Nm}. Since the averaged error is the unbiased estimator of the true systematic error with minimal variance [123] it is the optimal setup correction based on N_m measurements for a single patient. (In the discussion section, the impact of damping of the setup correction vectors based on population statistics, is described.)

   In the Appendix we show that, if the random errors are isotropically distributed ($\sigma_{r,i} \equiv \sigma_r$, $i = 1$, 2, 3), then, for $N_f >> N_m$ the NAL protocol yields a cumulative

probability distribution of the mean 3D setup error over all N_f treatment fractions, d_{N_f}, given by

$$P(d_{N_f} > x\Sigma_{\text{eff,NAL}}) = \sqrt{\frac{2}{\pi}} x \exp(-\tfrac{1}{2} x^2) + 1 - \text{erf}(\tfrac{1}{\sqrt{2}} x), \text{ with } \Sigma_{\text{eff,NAL}} \approx \frac{\sigma_r}{\sqrt{N_m}} \quad (6.3).$$

$\Sigma_{\text{eff,NAL}}$ is the SD of the effective systematic errors in the NAL protocol.

Therefore, the final distribution of systematic errors is basically independent of the initial distribution of systematic errors, and determined only by the SD of the random errors and the number of measurements N_m per patient from which the systematic error of a patient is estimated. The expression for $\Sigma_{\text{eff,NAL}}$ in Eq. (6.3) is intuitively plausible since it equals the random measurement error on the estimate of the systematic error.

The NAL protocol requires that each patient is corrected, since we apply no action level. Hence, its application requires a treatment table with which such corrections can be executed sufficiently accurate and which is preferably actively controlled from the accelerator record and verify system. With such an auto setup system, the potential reduction in workload of the SAL protocol due to the fact that not all patients are corrected vanishes. On our Scanditronix MM50 accelerator, we are already performing setup corrections in this manner (taking only a few seconds per patient) and treatment machines from various manufacturers offer this possibility.

To take into account the finite accuracy of the treatment couch and setup measurements, one could modify the NAL protocol by not performing corrections below a certain threshold, say 1 mm. Such a criterion would usually be met along a single coordinate axis only. For the other axes, a correction would still be applied. We will briefly discuss the (small) impact of a correction threshold. Note that for the NAL protocol, $N_{\text{meas}} = N_m$.

6.2.4 Retrospective application to measured setup data for 600 prostate cancer patients

During 4 years, we obtained setup data for a group of 720 prostate cancer patients. The 3D setup error vector was measured in each imaged fraction using one AP beam and two lateral beams. All these patients were treated with a 3D SAL protocol with either $(\alpha, N_{\text{max}}) = (9$ mm, 4) or (8 mm, 3). After the first stage of the protocol was completed, a second stage based on weekly checks was started: measurements were obtained once a week and a mean setup error vector $V_{N_{\text{max}}}$ over the last N_{max} measurements (including measurements of the first stage if necessary) was calculated. The length of $V_{N_{\text{max}}}$ was tested against the fixed action level $\alpha/\sqrt{N_{\text{max}}}$; if larger, a correction $-V_{N_{\text{max}}}$ was performed in subsequent fractions while the protocol was completely restarted. The combined first and second stages resulted in on average 9.5 imaged fractions per patient, which were distributed over the entire treatment period of about 7 weeks. N_{max} was either 3 or 4, and the number of weeks during which the second stage was applicable at least 6, but images were not always obtained during the last week. As a result, minimally 8 fractions should have been imaged, also for patients without a setup correction. Therefore, we selected patients with at least 8 imaged fractions, yielding a total of 600 patients with on average 10.1 imaged fractions per patient. With the methods described above, we calculated the setup errors that would have

occurred if no protocol had been applied and so derived the population setup distribution given in Table 6.1.

The above-described database of patient setup errors without correction protocol has been used for retrospective application of various 3D SAL (first stage) and NAL protocols. After each simulation, the systematic errors were calculated based on the assumptions that (1) the measurements used for the OVP are obtained in the first subsequent fractions and (2) each patient is treated in 30 fractions. For instance, for the $N_m = 3$ NAL protocol, the mean setup error of the first 3 measurements is weighed by 3/30 whereas the mean error of subsequent measurements after the NAL correction (on average 7 per patient) is weighed by 27/30. Hence, the effective systematic error per patient ($S_{p,\text{eff}}$) is estimated from the imaged fractions by taking into account the treatment time before and after correction. Based on Table 6.1 and Eq. (6.3), we estimate that a NAL protocol with $N_m = 3$ would reduce the initial systematic errors by $\sim 50\%$ and we selected this protocol as a baseline for comparison.

Table 6.1 Set-up deviation parameters (mm) in case no setup corrections would have been applied, for a group of 600 prostate cancer patients. Σ was derived from Σ_{eff} using Eq. (6.2) with $N_f = N_{\text{meas}} = 10.1$.

	X (lateral)	Y (Sup-Inf)	Z (Ant-Post)
μ	0.6	1.0	0.4
Σ	2.2	2.2	2.5
σ_r	2.1	1.8	2.0

6.2.5 Monte Carlo simulations

Studies of OVPs that use actually measured patient setup data, as described in the previous section, include the inter-patient variation of the random error, time trends, and possible covariance terms between various variables. However, such simulations require large input databases of a few hundred patients to properly represent all this information, and these are often not available. An additional problem is that in general not all fractions are measured (e.g. for the prostate patients described above, $\sim 30\%$ was covered) which renders the database useless for the simulation of certain SAL protocols (specifically for large values of N_{max} or small values of α which would generate many multiple corrections). Hence, to fully assess the effects of SAL OVPs, we have performed Monte Carlo simulations.

We developed a 3D Monte Carlo (MC) simulation program with a graphical user interface using IDL software (Research Systems Inc). Input to the program is, for each direction: the population parameters (μ, σ_r, Σ), the SD of the correction inaccuracy (σ_{table}) and a threshold for correction: a detected deviation below the threshold will not be corrected. Further input consists of selection of the type of OVP (which includes SAL and NAL), the OVP parameters, whether it will be followed by a second stage and if so, definition of that second stage, the number of treatment fractions per patient N_f and the

number of patients per simulation (10^4 in all following results). For each fraction, the setup is simulated, irrespective of whether the OVP prescribes that it should be imaged or not, so that the proper statistics can be calculated. In the results presented below, we have only simulated the first stage of the SAL OVP. In the simulations, the covariance between the stochastic variables is assumed to be 0, i.e. the covariance matrix is diagonal.

6.2.6   Overall verification of Monte Carlo simulations

Although we checked many details of our MC simulations in 1, 2 and 3D, we present here only an overall comparison with previously published results. Bel et al. [17] performed 3D simulations for setup inaccuracies measured in 3 different institutions. We took the input for our simulations from their Tables 1 and 2 and confined ourselves to simulating the first stage of the protocol. The three SAL protocols involved have parameters $(\alpha, N_{max}) = (6,2)$, (9,4) and (10,4) for patient populations described by $(\Sigma, \sigma_r) = (2,2)$, (3,2) and (3, 2.5) respectively. The outcomes of our simulations were compared to the data in Table 3 in [17]. Apart from these verifications for the SAL protocol, we verified the MC simulations for the NAL protocol described in the next section with the analytical expressions given above and in the Appendix.

6.2.7   Final setup accuracy and workload of NAL and SAL protocols

In a NAL protocol, for each patient the number of imaged fractions equals the protocol parameter N_m. Per value of N_m, we have determined the SAL parameter combinations (α, N_{max}) which yield, on average, an identical imaging and analysis workload as the corresponding NAL protocol: $N_{meas} = N_m$. This investigation was performed using the previously described MC simulations. We then selected those parameter combinations which yield a minimal $\Sigma_{eff,SAL}$: the SD of the effective systematic errors ($S_{p,eff}$) after the SAL protocol was applied (the definition is analogous to the definition of $\Sigma_{eff,NAL}$ in the Appendix). In this way, we determined the optimal SAL protocol (in the sense of minimal systematic dispersion) for a given portal imaging workload N_{meas}. A SAL protocol was considered clinically acceptable if the fraction of patients with 3 or more detected corrections (F_{3+}) did not exceed 10%. We posed this constraint since (i) if F_{3+} is large, the value of α is too small, yielding inaccurate corrections largely based on random errors [13] and (ii) a large fraction of multiple corrections is clinically unacceptable: 3 or more corrections tends to indicate a large random setup error, possibly due to a faulty setup demarcation (skin marks) for a particular patient; patients with more than 2 corrections are therefore re-simulated in our hospital. Nevertheless, although SAL protocols that generate $F_{3+} > 10\%$ may be clinically unacceptable, we will also give "optimal" results that may be obtained with such unconstrained SAL protocols.

The above described simulations were performed for a patient population with $\Sigma = \sigma_r$ (close to what is measured in most populations [17, 28, 103, 138, 139], see also Table 6.1), $\mu = 0$ and $\sigma_{table} = 0$ in each direction. For each patient, 30 treatment fractions were simulated. The optimized SAL protocols were also applied to a simulated population with $\Sigma = 2\sigma_r$ to study the sensitivity to population parameters.

6.2.8   Time behavior of systematic setup errors in NAL and SAL protocols

To gain insight in how NAL and SAL protocols reduce systematic errors from fraction to fraction, we used MC simulations to calculate the dispersion of systematic errors at each fraction. Since in a simulation we know the true systematic errors S_p as well as the applied setup corrections $c_{p,f}$, we can simply obtain the true systematic error for each fraction as $S_{p,f} \equiv S_p + c_{p,f}$ (as $c_{p,1} = 0$ for any off-line protocol, it follows that $S_{p,1} \equiv S_p$). This allows the calculation of the systematic dispersion over the patient population per fraction: $\Sigma_f \equiv \mathrm{SD}_P(S_{p,f})$. The decrease of Σ_f with fraction number f reflects how systematic errors are reduced with time by an OVP.

6.2.9   Correction inaccuracies

We have found that setup corrections by treatment couch shifts can be performed with a high accuracy: 1 SD = σ_{table} = 0.3 mm for the couch of our Scanditronix Racetrack Microtron [28]. Van den Heuvel *et al.* [58] described a Siemens ZIV digital read-out couch with remote control which could be positioned with a precision of 1 mm. For a remotely controlled Siemens ZXT table, an accuracy of 0.6 mm (1 SD) in each direction was found [16]. Even if the table accuracy is worse, since the inaccuracies in the table shifts are random, they are not very important to the final systematic errors (see Appendix). However, for the SAL protocol, repositioning inaccuracies may become of importance to the workload since they may generate new corrections.

We have simulated the effect of inaccuracies in the application of corrections by adding in each direction a component that is sampled for each fraction from a normal distribution with SD = σ_r. Although this is a (too) large value for errors on corrections performed by table translations (typically $\sigma_r \geq 2$ mm, whereas $\sigma_{\text{table}} < 1$ mm), it will help identify potential problems due to (random) errors in applied patient setup corrections.

6.3 Results

6.3.1   Retrospective application to measured setup data

The cumulative distribution of d_{Nf} (the length of the 3D setup error averaged over all fractions) if no corrections had been applied (i.e., $d_{Nf} = \sqrt{S_{p,\text{eff},1}^2 + S_{p,\text{eff},2}^2 + S_{p,\text{eff},3}^2}$) is given in Fig. 6.1. The tail extends to 1 cm and the fraction of patients with mean 3D errors > 5 mm is 27%, very close to the numbers reported in a multi-center study of a similar population [17]. The corresponding distribution for the NAL protocol with N_m = 3 is also depicted in Fig. 6.1. Obviously, the NAL protocol is very effective in reducing the systematic errors: with only 3 measurements per patient, a value of d_{Nf} > 5 mm would occur in 2% of the patients, which is an upper limit as we point out below.

Additionally, Fig. 6.1 shows the predicted analytical cumulative distribution based on Eq. (6.3). To properly compare the analytical distribution with results based on 10 measured fractions per patient, we have added a term to $\Sigma_{\text{eff,NAL}}$ which takes into account that, although we properly weigh the measured data with time, we have on average only 7 measurements available after the correction. This is only a sample of the 27 setups that would normally occur in a 30 fraction regime. Hence, the measured $\Sigma_{\text{eff,NAL}}^2$ is increased on average by $\sim \sigma_r^2 / 7$ with respect to the true $\Sigma_{\text{eff,NAL}}^2$ given in Eq. (A6.4) in the Appendix. With this small correction, the resulting analytical distribution in Fig. 6.1 agrees well with the distribution based on measured patient data.

A number of conclusions may be drawn from this agreement. First, it confirms the derivation in the Appendix that the population *mean* random error, σ_r (Table 6.1), is sufficient for analytical calculation of the distribution of the final systematic errors in the NAL protocol, despite considerable inter-patient variation in σ_p (ranging from 0.4 to 5 mm for each direction in the studied population). Second, the analytical curve does not take
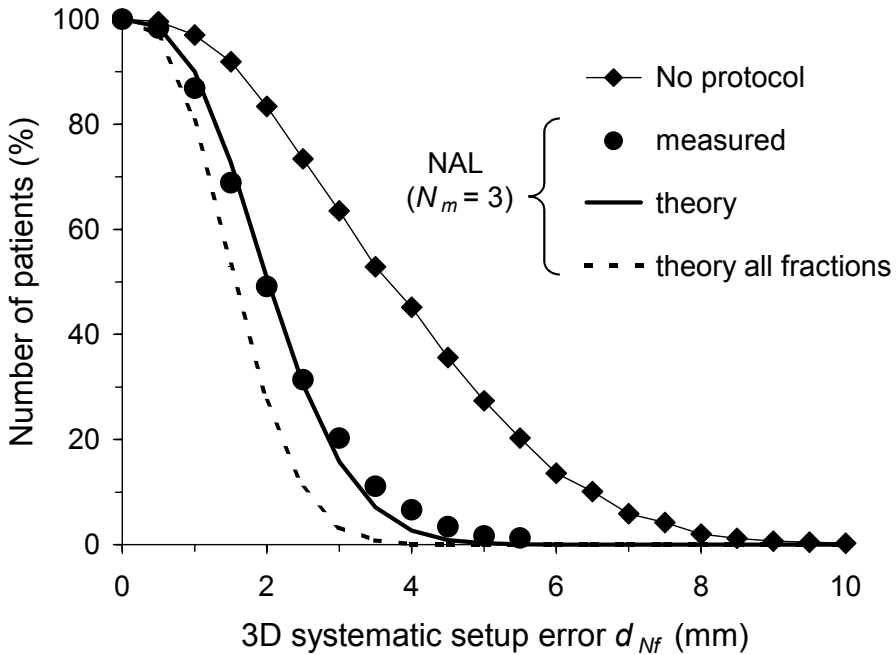

Figure 6.1 The cumulative distribution of the 3D systematic setup errors d_{Nf}, obtained by retrospective application of the NAL OVP with $N_m = 3$ to measured setup data of 600 prostate cancer patients, and the corresponding distribution in case of no OVP. Curves marked "theory" were calculated with the analytical expressions given in the Appendix.

into account explicitly the possible transient effects in the systematic setup error of the patient [17, 41, 49, 145]. If such time trends do occur they will increase the values of σ_r in Table 6.1 (which is the dominant factor in $\Sigma_{\text{eff,NAL}}$) with respect to the true random dispersion, but it has been shown that such an increase is very small particularly for prostate patients [17]. We observe from Fig. 6.1 that for the studied patient group, the (potential) influence of time trends on the retrospectively derived cumulative distribution does not introduce significant deviations with respect to our analytical prediction. From these observations, we conclude that in the studied (large) patient population, the NAL procedure without a second stage is already sufficient to reduce systematic errors significantly and this reduction is well described by analytical expressions. On that basis, we may calculate the true distribution of the effective systematic displacements as would have been measured in 30 fractions (Eq. (6.3) with $\Sigma_{\text{eff,NAL}}$ from Eq. (A6.4)) to arrive at the leftmost curve in Fig. 6.1. The impact of the fact that a sample of 10 fractions has been obtained in the measured data is non-negligible (in fact, the fraction of patients with $d_{Nf} >$ 5 mm in this 'true' distribution is $< 1\%$), hence we have continued with MC simulations of all treatment fractions.

6.3.2   Overall verification of Monte Carlo simulations

Bel *et al*. [17] quote the number of mean 3D deviations (i.e., d_{Nf}) > 5 mm, the average number of measurements per patient (i.e., N_{meas}) and the average number of detected corrections per patient for the first stage of 3 SAL OVPs simulated in 3D. For all 3 protocols, we found identical results with our simulations to within the accuracy with which they present their results in Table 3. Particularly the fraction of patients with $d_{Nf} >$ 5 mm after the first stage of the SAL OVP is sensitive to all aspects of the simulation and these coincide to within 0.1%. The simulation fluctuations on this number are of the order of 0.1% for 10^4 simulated patients, hence our results are consistent with the previously published results.
The agreement between the MC simulation and the analytical expression for $\Sigma_{\text{eff,NAL}}$ derived in the Appendix can be judged from Fig. 6.2. The mean absolute deviation between analytical prediction and simulation is 0.3%, on average, which equals $\sim 10^{-3}\sigma_r$. In Fig. 6.3 we give an analytically predicted and simulated cumulative distribution for a NAL protocol with $N_m = 3$ (see next section). The agreement between the two distributions is within 0.4%, consistent with the statistical fluctuations of $\sim 0.2\%$ (1 SD).

6.3.3   Final setup accuracy and workload of SAL and NAL protocols

The optimal (i.e., minimising $\Sigma_{\text{eff,SAL}}$) SAL parameters per workload value N_{meas} are given in Table 6.2. In Fig. 6.2, the corresponding values of $\Sigma_{\text{eff,SAL}}$ are plotted together with the values of $\Sigma_{\text{eff,NAL}}$ for an identical workload. All values are expressed in σ_r so that the results are independent of the specific value of σ_r. Obviously, for a given workload, the final systematic errors obtained with the NAL protocol are significantly

smaller than for the SAL protocol. For instance, if one aims at halving the effective systematic dispersion in each direction, a NAL protocol with $N_m = N_{meas} = 3$ will suffice, whereas for SAL one requires nearly 3 times as many measurements ($N_{meas} = 8\text{-}9$). Both NAL and SAL curves become flat and even turn upwards again for large values of N_{meas}. This is because in case of 'late' corrections (after a significant number of fractions have already been given without a correction) the pre-correction systematic errors start to dominate the effective systematic dispersion (see Eq. (A6.4)).

The optimised SAL protocols of Fig. 6.2 and Table 6.2 were constrained by $F_{3+} \leq 0.1$. Our simulations showed that, if one includes protocols with $F_{3+} > 0.1$, the reduction of $\Sigma_{eff,SAL}$ with respect to the constrained protocols of Table 6.2 is on average 6% (maximally ~ 10%), which is clinically negligible, whereas the average value of F_{3+} increases to 0.25. Moreover, we found that the outcome of such protocols in terms of workload is extremely sensitive to the pre-correction distribution of set-up errors and will thus be hard to predict in practice. Therefore, the SAL protocols used to construct Fig. 6.2 represent optimal protocols not only in terms of final accuracy but also regarding multiple correction detection and robustness.
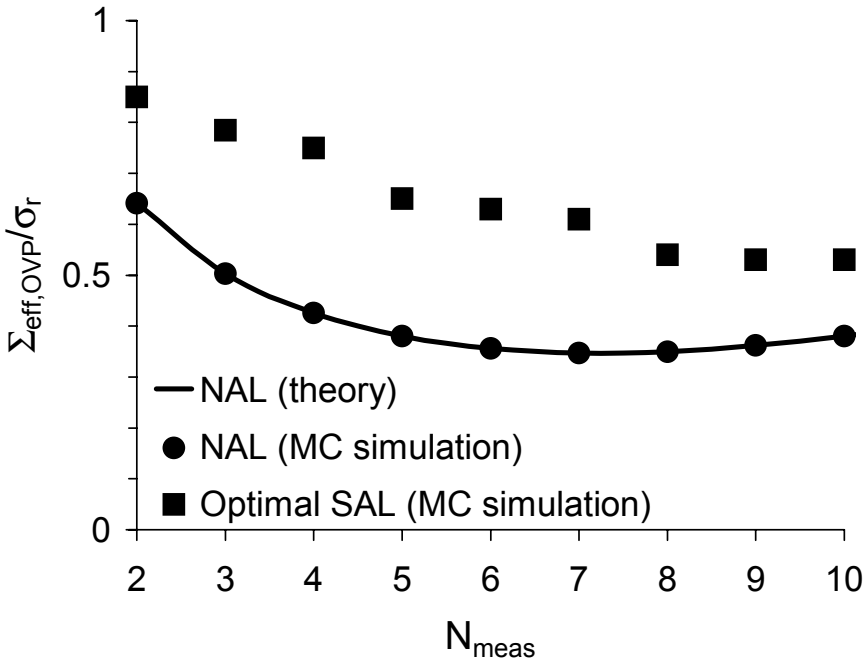
Figure 6.2 The effective systematic dispersion after application of the NAL and optimal SAL OVPs, both denoted by $\Sigma_{eff,OVP}$, as a function of the mean number of imaged fractions per patient N_{meas}, expressed in the SD of the random setup errors σ_r. The results were obtained with Monte Carlo (MC) simulations, but for the NAL OVP the analytical result is also depicted (solid line).
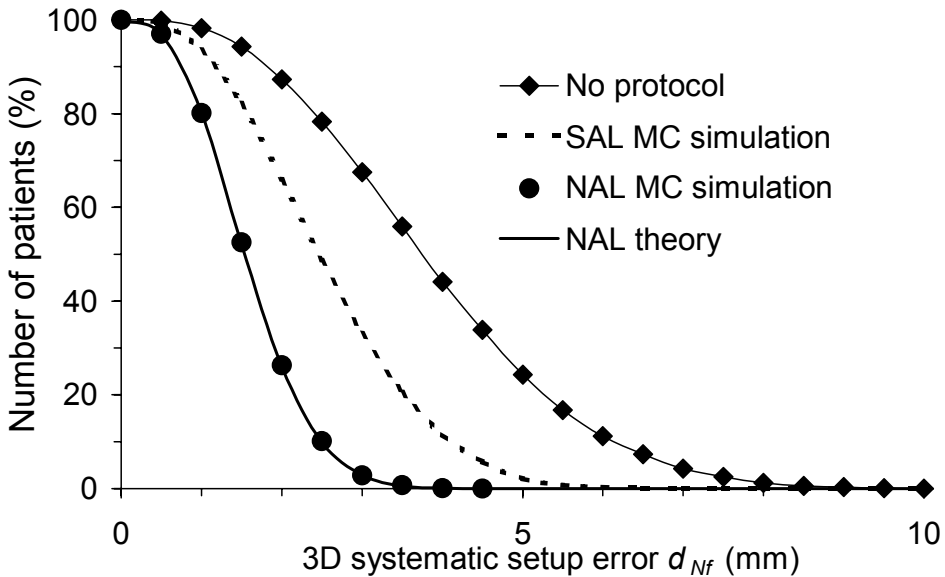
Figure 6.3 The cumulative distribution of the 3D systematic setup errors d_{Nf}, obtained from MC simulations (with population characteristics from Table 6.1) for the NAL and optimal SAL OVP for $N_{meas} = 3$: $(\alpha, N_{max}) = (6.2, 2)$. For the NAL OVP, the analytical result is also given (solid line).

Table 6.2 3D SAL protocol parameters (α, N_{max}) that yield a minimal $\Sigma_{eff,SAL}$ for a given average number of imaged fractions per patient, N_{meas}. These results were obtained for a population with $\Sigma = \sigma_r$. The initial action levels α are expressed in σ_r. The fraction of patients with at least one setup correction is denoted by F_{1+}.

N_{meas}	2	3	4	5	6	7	8	9	10
α/σ_r	2.3	3.1	3.7	3.1	3.6	4.0	3.5	3.8	4.3
N_{max}	1	2	3	3	4	5	5	6	7
F_{1+}	0.5	0.4	0.4	0.5	0.5	0.5	0.6	0.6	0.6

We investigated the impact of setting a correction threshold of 1 mm in the NAL protocol (i.e. corrections < 1 mm are not executed) and found a negligible impact on $\Sigma_{eff,NAL}$ for $\Sigma = \sigma_r = 2$ mm. The number of corrected patients was reduced by only a few % (because the setup was unaltered only if the correction was less than 1 mm in *each* direction).

If we apply the NAL and SAL protocols used in Fig. 6.2 to a population with $\Sigma = 2\sigma_r$, the values of $\Sigma_{\mathrm{eff,NAL}}$ and $\Sigma_{\mathrm{eff,SAL}}$ are practically unaltered (deviations of a few %), consistent with published results [13]. However, while N_{meas} for the NAL OVP remains the same by definition, for the SAL protocols N_{meas} increases by approximately 1 fraction/patient for all tabulated SAL parameter combinations. Moreover, the fraction of corrected patients increases to 90% in all cases, approaching the 100% correction percentage of the NAL protocol.

We have applied the NAL and optimal SAL OVP with $N_{\mathrm{meas}} = 3$ of Table 6.2 to a simulated population described by the parameters given in Table 6.1, assuming 30 treatment fractions. For the SAL protocol, this implies $(\alpha, N_{\mathrm{max}}) = (6.2, 2)$, very close to the $(\alpha, N_{\mathrm{max}}) = (6.0, 2)$ for which clinical results obtained in a group of prostate cancer patients have been published [17]. The resultant cumulative distributions of 3D systematic setup errors are given in Fig. 6.3. As expected from the above results, the actual value of N_{meas} for the SAL protocol becomes 3.3 instead of 3 due to the fact that Σ is slightly larger than σ_r for each direction (Table 6.1). Again, the gain in final accuracy by application of the NAL instead of the SAL OVP is evident.

6.3.4   Time behaviour of SAL and NAL protocols

For both the NAL $N_m = 3$ protocol and the optimal SAL protocol with $N_{\mathrm{meas}} = 3$ from Table 6.2, the systematic dispersion Σ_f for the first 15 treatment fractions is given in Fig. 6.4. As expected for the NAL protocol, Σ_f is constant up to $f = N_m$, drops sharply in the next fraction and then remains constant at the value Σ_{NAL} given in the Appendix. The SAL protocol yields a different picture. Although some patients are corrected already after 1 fraction ($f = 1$), the effect on Σ_f for the second fraction is zero. In the Appendix, we show that setup corrections equal to the setup error measured in one single fraction have no impact on the post-correction distribution of systematic errors if the pre-correction distribution of systematic and random errors is identical. This result holds irrespective of the action level applied or the dimension of the setup error vector tested, which is confirmed in Fig. 6.4.

After the second fraction, a gradual decay of Σ_f follows for the SAL protocol but it remains significantly larger than the corresponding value of the NAL protocol; after the 3rd fraction hardly any reduction is seen, which can be understood from the fact that $N_{\mathrm{max}} = 2$.

In Fig. 6.4 we also show Σ_f for the optimised SAL protocol with $N_{\mathrm{meas}} = 8$. This SAL protocol approximately reaches $\Sigma_{\mathrm{eff,SAL}} \approx \Sigma_{\mathrm{eff,NAL}}$ for the NAL protocol with $N_{\mathrm{meas}} = 3$ (Fig. 6.2). The decay of Σ_f is slow compared to that of the NAL protocol, from which the increased number of required measurements to reach the same final accuracy can be understood.
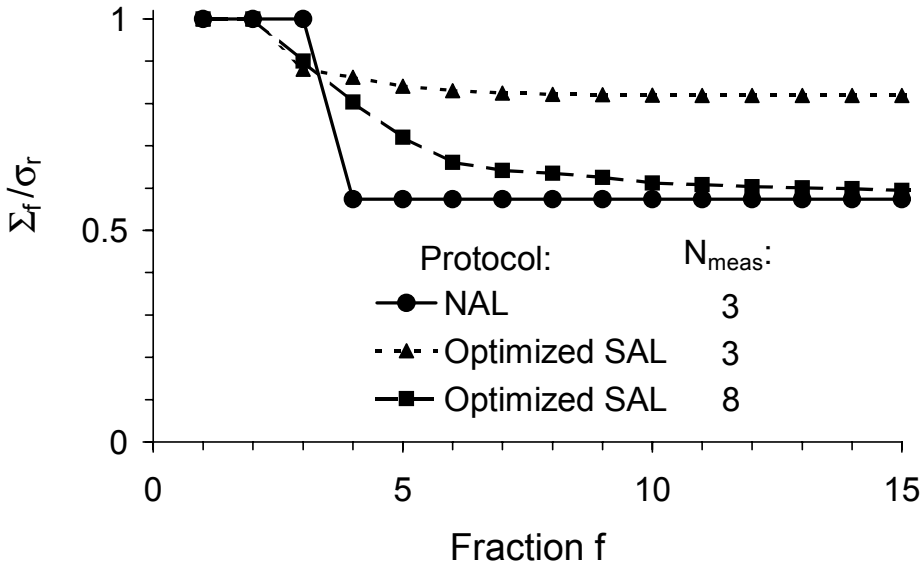
Figure 6.4 The SD of the systematic errors in fraction f, Σ_f (expressed in σ_r) as obtained from MC simulations (with $\Sigma_1 = \sigma_r$), for three OVPs.

6.3.5 Systematic setup errors in the SAL protocol for corrected and uncorrected patients

A significant fraction of the patients is not corrected in a SAL protocol (1- F_{1+}, Table 6.2), which was considered one of its attractive features [13]. Although one may argue that these patients already have acceptable final systematic errors, the information obtained in the N_{max} measurements is not applied to improve their setup accuracy, resulting in larger effective systematic dispersions than for the corrected patients. In Fig. 6.5 we show $\Sigma_{eff,SAL}$ as in Fig. 6.2 but now separated into the subgroups of patients who did not get a correction and patients who did. Obviously, the setup accuracy in the uncorrected group is worse than in the corrected group, particularly with increasing overall accuracy. Since the fraction of corrected setups hardly increases with increasing N_{meas} (Table 6.2) the gain in accuracy obtained with a larger number of measurements as presented in Fig. 6.2 may be attributed to increasingly accurate corrections for larger N_{meas}. This is also reflected by the stronger decrease of $\Sigma_{eff,SAL}$ with N_{meas} for corrected than for uncorrected setups.

6.3.6 Correction inaccuracies

The inclusion of large random errors in the couch translations had a negligible impact on the final systematic errors of the NAL protocols, which follows from Eq. (A6.3) in the
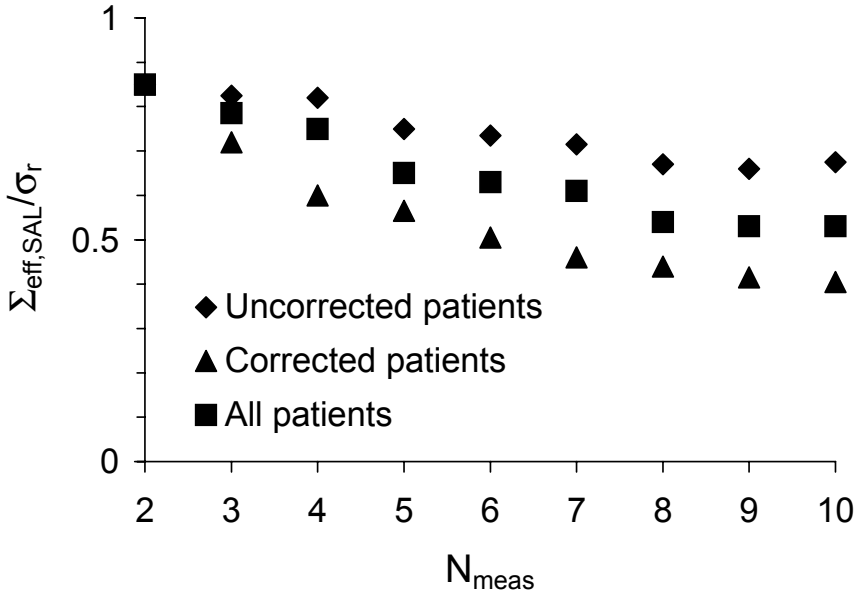
Figure 6.5 The effective systematic dispersions in the SAL protocol, $\Sigma_{\text{eff,SAL}}$ (expressed in σ_r) for the optimal SAL protocols of Table 6.2. The values labeled "all patients" (squares) are identical to those in Fig. 2, but here they are separated into the values of the subgroups of patients with (triangles) and without (diamonds) setup correction.

Appendix. The number of measurements per patient N_{meas} remained of course unaltered, whereas $\Sigma_{\text{eff,NAL}}$ was increased by less than 6% for $N_{\text{max}} \leq 5$.

For the SAL protocols with (α, N_{max}) from Table 6.2, we again found that the final systematic dispersion is hardly affected by correction errors: $\Sigma_{\text{eff,SAL}}$ increased on average by 3%, ranging up to 8%, consistent with published results for a 1D protocol [13]. However, N_{meas} and hence the workload increased by 22 ± 7% (the fluctuation is over the (α, N_{max}) parameter sets of Table 6.2). This is reflected by the fraction of triple and higher correction detections, which increased from $F_{3+} < 10\%$ to 21 ± 6%. Therefore, the optimal SAL protocols tend to become clinically unacceptable for large inaccuracies in the applied table translations.

6.4 Discussion

Based on our clinical experience that the workload introduced by OVPs is mainly related to the average number of images required per patient and not so much by the fraction of patients with setup corrections, we have proposed the NAL protocol and compared it with

the much applied SAL protocol. The presented results clearly indicate that the NAL protocol reaches a target final distribution of systematic errors with much less workload than the SAL protocol (Fig. 6.2 and 6.3).

The three main causes for the relatively high SAL workload are:

(1) A correction based on N measurements will yield a post-correction dispersion of systematic errors $\approx \sigma_r / \sqrt{N}$ (Appendix). Since N is fixed for each patient in the NAL protocol ($N = N_{\text{meas}}$), the systematic dispersion for this OVP $\approx \sigma_r / \sqrt{N_{\text{meas}}}$. For the SAL protocols listed in Table 6.2, MC simulations showed that for patients who had their last setup correction based on N measurements, again the post-correction systematic dispersion $\approx \sigma_r / \sqrt{N}$. For this OVP, $N_{\text{meas}} > N_{\text{max}} \geq N$, and therefore the overall systematic dispersion will always be larger than $\sigma_r / \sqrt{N_{\text{meas}}}$, even in the subgroup of corrected patients. Moreover, the inaccuracy of setup corrections based on too few measurements are the major cause of the multiple corrections and associated workload of the SAL OVP. For example, MC simulations of the protocol in Table 6.2 with N_{meas} = 4 showed that, of the patients who had a first setup correction based on 1 measurement, 68% required additional corrections. For patients who had a first setup correction based on 3 measurements, only 11% had additional corrections.

(2) The SAL protocol is completely restarted after a correction, and all measured pre-correction information is discarded. Since the magnitude of the performed corrections is known, there is little support for this approach if corrections can be performed accurately through table translations.

(3) A large fraction of patients is not corrected, as a result of the action level. Hence a significant amount of the measured information is not applied for setup improvement yielding less than optimal results (Fig. 6.5).

The first cause described above may be slightly alleviated through a correction damping scheme [45, 117]. With the aid of *a priori* information on the distribution of systematic and random errors in the population, one may calculate a damping factor which yields an improved (i.e. the maximum likelihood) estimate of the true systematic error for a patient [45]. However, corrections for systematic errors based on only one or two measurements, as occur in a SAL protocol, can never be quite accurate and hence only small improvements from damping have been reported. In a 1D simulation study of the SAL protocol with damping, the number of corrections per patient decreased by only 0.1 without any clinically significant improvement in post-correction setup accuracy [45]. In another study, such a protocol was retrospectively applied to setup data obtained from tangential breast fields and similarly small effects were observed [103]. Because these studies were 1D, we repeated the simulations in 3D with an optimal damping scheme [45]. For the SAL protocols of Table 6.2, we found that, on average N_{meas} decreased by 8%, and the systematic dispersion by only 1%. For none of the protocols a clear benefit was found. Although damping will slightly reduce the number of measurements in SAL protocols for patients with large random errors, the other two causes for the relatively high workload of the SAL protocol pointed out above remain.

An additional problem with SAL protocols is the requirement that knowledge of σ_r must be accurate since deviating values may rapidly increase workload or decrease final accuracy [13, 130]. We found a similar workload effect for variations in Σ , specifically for the fraction of corrected patients, which is considered of importance for SAL protocols. In contrast, the workload of the NAL OVP is completely predictable and its final accuracy for

a given number of measurements better than for the SAL protocol, even if σ_r is poorly known.

A psychological advantage of the SAL protocol is that large measured setup errors may be corrected directly after the first measured fraction. As is shown analytically in the Appendix, such corrections are not useful when Σ is comparable to σ_r, even if very large action levels are applied to this first measurement (see also Fig. 6.4). Of course, one is tempted to correct specifically large setup errors detected at the first fraction. However, especially for a highly significant error (outside the range one would expect for a given population), its origin should be investigated before a correction is applied. Nevertheless, if one wishes to correct immediately for very large setup errors (which can be useful if Σ is significantly larger than σ_r), the NAL protocol can be modified by introducing one (large) action level in the first fraction. If the setup error is smaller than this level, the NAL protocol runs unaltered. Otherwise, a correction (preferably with damping [45], because the correction is based on a single measurement) is performed for the next fractions and the standard NAL protocol continues by performing setup measurements for the remaining N_m-1 fractions. Next, the applied setup correction is subtracted from the last N_m-1 measurements before averaging all N_m measurements to arrive at an improved estimate of the systematic setup error. The corresponding improved setup correction is applied from fraction N_m+1 onwards. This approach allows for immediately correcting particularly large initial setup errors without increasing the workload of the NAL protocol.

Finally, a discussion on off-line setup corrections is incomplete without addressing transitions in the systematic setup errors with time. Both the NAL and SAL protocol focus on removing initial (large) systematic setup errors, and do not take into account possible time dependent effects in these errors. Their use is therefore limited to patient populations which do not exhibit large setup shifts over time. The published studies on transient effects in systematic setup errors of prostate patients have focussed on gradual systematic shifts of the patient with time, the so-called time trends and in fact are limited to linear time trends [17, 41, 49, 145]. For a linear time trend, the model of patient setup errors described in Methods and Materials should be extended with a shift linear in time. For each patient p, this shift has a certain slope s_p expressed in mm/fraction, and so at fraction f the setup error becomes $x_{p,f} = S_p + s_p \times f + \Delta_{p,f}$. In some studies, this slope was estimated by applying a least squares fit, after which the significance of the estimated slope (i.e., does it differ significantly from 0) can be tested with a Student's t-test [123].

However, if the estimated slope for each of the three setup directions is tested separately at the $p = 0.05$ significance level, the significance level for finding any linear trend in a patient is actually $1-(1-p)^3 = 0.14$. This result implies that, for a patient population without any trends, the applied statistical test would yield a significant trend ($p < 0.05$ in any direction) in on average 14% of the patients. Hence, patients with seemingly significant trends will be detected regularly with such a test. For instance, in the study by El-Gayed *et al.* [41], 3 out of 10 patients showed a significant trend at $p < 0.05$ in a single direction, which implies 3 significant detections in 30 (10 patients $\times$ 3 directions per patient) series of setup errors. If one supposes that no time trends occur (the true slope $s_p = 0$ for each patient), we can calculate from the binomial probability distribution [123] with N = 30, p=0.05, the probability of detecting 3 or more significant trends in 30 such series. The result is 10%, which implies that this is not a rare event. Therefore, [41] does not provide

unambiguous evidence on the frequent occurrence of time trends, although one particularly significant trend was observed (p < 10^{-5}). In a similar analysis, Hanley *et al.* [49] detected significant linear trends in 10 cases, when studying 50 patients (i.e., 150 series of setup data). With the calculation method described above, we find that the probability of detecting 10 or more significant trends in 150 setup series in which no trends actually occurred is 17%. This relatively large probability, and the fact that all the p-values in [49] were ≥ 0.02, indicate that the statistical evidence for time trends in their population is limited. In the multi-center study of Bel *et al.* [17], a frequent occurrence of time trends is described in the Appendix. Their Table A1 shows that in more than 50% of the patients, significant trends (p < 0.05) were observed in each of the three participating institutions. However, despite this regular occurrence of time trends the expected number of images and setup corrections for this patient group were consistent with simulations that did not incorporate such time trends. In addition, the SD of the interpatient variation of the estimated s_p values was presented in [17]: for each patient, a slope s_p was estimated from a linear regression fit, and the SD of these slopes was calculated. The SDs thus derived were on average about 0.08 mm/fraction. To assess the meaning of this value, we have calculated the value that would be expected for this SD if the true $s_p = 0$ for all patients, i.e.

for a population without time trends. Then, each estimated slope will still differ slightly from zero, due to the random setup fluctuations and the finite number of measurements available per patient. Using the magnitude of the random errors (σ_r) and the number and distribution in time of the measurements described in [17], we can estimate the SD of these estimated slopes from the linear regression equations [123] and arrive at a value of approximately 0.08 mm/fraction. Because this statistical lower limit on the SD of the estimated slopes is so close to the SDs actually derived in [17], the simultaneous frequent detection of significant (i.e. relatively large) time trends in the same patient group is hard to understand. Therefore, the results in [17] on time trends seem to be non-conclusive.

An off-line protocol suited to take into account time dependent effects is the adaptive treatment therapy (ART) [145, 146]. Apart from correcting systematic setup errors, ART adapts treatment plan margins once the setup errors have been estimated with sufficient confidence. On average, the procedure involves 6 imaged fractions, ranging up to 9 [146]. Follow up imaging is required to properly monitor time trends. Hence ART requires more imaging workload than the NAL protocol (next to the workload generated by repetitive treatment plan optimization and implementation at the accelerator), at the gain of patient specific treatment optimization. Limited data on time trends have yet been obtained with this protocol. In a retrospective study [145], time trends were detected in particular in pelvis patients. Since the latter patients also suffered from large random and systematic setup errors, the relative importance of these trends is not clear. In a prospective ART study [146], a linear time trend was detected in a single patient while in two other patients trends were detected but these seemed to alternate. Since no significance levels nor magnitudes were given for these trends, their impact on setup accuracy within the context of a NAL or SAL protocol is difficult to assess.

In summary, the results discussed above describe evidence of linear time trends, but do not provide a clear picture of the true incidence and magnitudes of such trends. Regarding our own measurements, we previously stated that the agreement between the predicted results for the NAL protocol and the retrospective application to measured patient setup data (which include weekly measurements obtained over the full treatment period) is satisfactory without taking into account time trends explicitly in our prediction (Fig. 6.1).

Because of this result, and the fact that the correction of time trends is not well possible with either the SAL or NAL protocols, we considered a detailed analysis on time trends beyond the scope of this paper. Nevertheless, because time trends are expected to occur in some patients, we have started investigations on an extension of the NAL protocol which involves an effective use of follow up measurements for time trend detection.

6.5 Conclusions

If off-line setup corrections are performed through treatment couch translations, the fraction of patients that must be corrected when applying an OVP is not important to the workload. Hence, the number of fractions to be imaged and analyzed to reach a specified reduction of systematic errors becomes the discriminating factor between OVPs. The NAL protocol we have proposed significantly decreases this number of fractions with respect to the much applied SAL protocol, while not compromising on the reduction of systematic errors. In addition, the workload for the NAL OVP is independent of fluctuations in population parameters and the distribution of the final systematic errors obtained can be derived analytically from knowledge of the random errors. These properties also render the NAL protocol highly suitable for setup verification with large dedicated portal imaging fields, since it both minimizes dose to healthy tissue and renders that dose completely predictable.

Appendix: Analytical derivation of the systematic dispersion in the NAL protocol

We first consider a population P of patients p that were treated in N_f fractions f without setup corrections. The patient setup errors may then be written as $x_{p,f} = S_p + \Delta_{p,f}$ where S_p is the systematic setup error for patient p and $\Delta_{p,f}$ the random error in fraction $f \in F = \{1,..., N_f\}$. The *effective* systematic error is defined as the actual systematic error for a given treatment, i.e. the mean over the treatment fractions F: $S_{p,\text{eff}} = <x_{p,f}>_F \equiv N_f^{-1} \sum_{f=1}^{N_f} x_{p,f}$. Its value deviates from S_p due to the random deviations $\Delta_{p,f}$ and the finite number of fractions N_f.

Suppose that we want to calculate the SD of $S_{p,\text{eff}}$ over the patient population P, which we denote by $\Sigma_{\text{eff}} \equiv \text{SD}_P(S_{p,\text{eff}})$. Using the definition of the SD, we can write $\Sigma_{\text{eff}}^2 = E_P(S_{p,\text{eff}}^2) - E_P^2(S_{p,\text{eff}})$, where E_P denotes the statistical expectation value (i.e. the population mean [123]) over the population P. However, we cannot calculate such expectation values directly, since as mentioned above, $S_{p,\text{eff}}$ not only depends on patient properties but also on the particular treatment. We therefore use the statistical concept of sample means [123]: we calculate the expectation value of any function of $x_{p,f}$, as the mean of its sampling distribution. We must then define the possible samples of $\{x_{p,f}\}$ for a single patient, since in practice only one sample of N_f fractions exists.

If we could treat the same patient p repeatedly under identical circumstances we would obtain various realizations γ of the treatment and corresponding samples $\{x_{p,f}\}^\gamma$ (each consisting of N_f values). The distribution of the random fraction-to-fraction fluctuations of $x_{p,f}$ is determined by the properties of p, i.e. the underlying probability density function is the same for each sample. Then, a function $\Phi_p(\{x_{p,f}\})$ (e.g., $\Phi_p = S_{p,\text{eff}}$) will have a different value Φ_p^γ for each sample $\{x_{p,f}\}^\gamma$, hence we refer to the values of such functions for a particular treatment as *effective* values. The fluctuation of Φ_p^γ from sample to sample (and hence the sampling distribution of Φ_p) is completely defined by p. The mean of Φ_p^γ over all samples $\gamma \in \Gamma$ is denoted by $E_\Gamma(\Phi_p)$, and only depends on p. Below, the expectation value of an effective value Φ_p over the patient population domain, $E_P(\Phi_p)$, is calculated from

$$E_P(\Phi_p) \equiv E_P(E_\Gamma(\Phi_p)) \tag{A6.1}.$$

We can now precisely describe the properties of the systematic and random errors. We define the mean systematic error by $E_P(S_p) = \mu$ and its SD by $\text{SD}_P(S_p) = \Sigma$. The random errors satisfy $E_\Gamma(\Delta_{p,f}) = 0 \ \forall f$ and we assume that they are independent: $E_\Gamma(\Delta_{p,f} \Delta_{p,f'}) = \delta_{f,f'} \sigma_p^2$ with $\delta_{f,f'}$ the Kronecker delta function. Note that each patient

can have a different SD of random errors, σ_p, so we do not assume a homogeneous patient population.

Using the above definitions, we now calculate Σ_{eff} since its derivation is illustrative for a more complex derivation below.

Using $S_{p,\text{eff}} = S_p + N_f^{-1} \sum_{f=1}^{N_f} \Delta_{p,f}$, we immediately find the mean of $S_{p,\text{eff}}$:

$$E_P(S_{p,\text{eff}}) = E_P(E_\Gamma(S_{p,\text{eff}})) = E_P(S_p) + N_f^{-1}\sum_{f=1}^{N_f} E_P(E_\Gamma(\Delta_{p,f})) = E_P(S_p) = \mu .$$

Its second moment becomes

$$E_P(S_{p,\text{eff}}^2) = E_P(S_p^2) + 2N_f^{-1}\sum_{f=1}^{N_f} E_P(S_p E_\Gamma(\Delta_{p,f})) + N_f^{-2}\sum_{f=1}^{N_f}\sum_{f'=1}^{N_f} E_P(E_\Gamma(\Delta_{p,f},\Delta_{p,f'}))$$

$$= \quad E_P(S_p^2) + N_f^{-1} E_P(\sigma_p^2). \quad \text{Using} \quad \Sigma^2 = E_P(S_p^2) - \mu^2, \quad \text{we} \quad \text{finally} \quad \text{obtain}$$

$\Sigma_{\text{eff}}^2 = \Sigma^2 + N_f^{-1}\sigma_r^2$ where $\sigma_r^2 \equiv E_P(\sigma_p^2)$ is the population-averaged variance of the random errors. We have thus formalized the expressions given in Eq. (6.1) and (6.2) in the text.

If we *measure* the setup errors $x_{p,f}$ in portal images, we introduce additional inaccuracies. Systematic inaccuracies ε_p stem from errors in the definition of reference setup [12, 28, 76] (including errors in the definition of the structures used in image registration) and systematic errors that may occur during image registration [22]. We characterize them by $E_P(\varepsilon_p) = 0, SD_P^2(\varepsilon_p) = \Sigma_{\text{ref}}^2$. The image registration process will also introduce random measurement inaccuracies $\omega_{p,f}$ and we assume that $E_\Gamma(\omega_{p,f}) = 0$ (no net registration error) and $E_\Gamma(\omega_{p,f}\omega_{p,f'}) = \delta_{f,f'}\sigma_{p,\text{reg}}^2$ (random registration errors are uncorrelated, and their magnitude can be patient dependent). With these definitions, the *measured* setup error in the case of *no* setup corrections becomes $x_{p,f}^m = x_{p,f} + \varepsilon_p + \omega_{p,f}$.

In the NAL protocol, we derive a single setup correction c_p based on N_m measurements: $c_p = -\sum_{f=1}^{N_m} x_{p,f}^m / N_m$, which is applied from fraction $N_m + 1$ onwards. The setup corrections are executed with a finite accuracy. Throughout this paper, we have assumed that corrections are performed with treatment table shifts and so the magnitude of the correction errors, $\tau_{p,f}$, is patient independent: $E_\Gamma(\tau_{p,f}\tau_{p,f'}) = \delta_{f,f'}\sigma_{\text{table}}^2$. Furthermore, we assume that the net error of the table shifts is zero: $E_\Gamma(\tau_{p,f}) = 0$. The actual corrections become $c_{p,f} = c_p + \tau_{p,f}$ with $c_{p,f} = 0$ for $f \leq N_m$. With this definition of $c_{p,f}$, we can write the measured setup error in case of the NAL protocol as $x_{p,f}^m = x_{p,f} + \varepsilon_p + \omega_{p,f} + c_{p,f}$. The corresponding *actual* setup errors are $x_{p,f}^{\text{act}} = x_{p,f} + c_{p,f}$. This leads to an effective systematic setup error for the NAL protocol, $S_{p,\text{eff,NAL}}$, and corresponding expectation value:

$$S_{p,\text{eff,NAL}} = <x_{p,f}^{\text{act}}>_F , \quad E_P(S_{p,\text{eff,NAL}}) = \frac{N_m}{N_f}\mu \tag{A6.2}$$

Hence, the mean effective systematic error is reduced by a factor N_m / N_f through the NAL protocol.

The SD of the effective systematic setup errors in the NAL protocol ($\Sigma_{\text{eff,NAL}}$) can now be calculated from the definition of $S_{p,\text{eff,NAL}}$ in Eq.(A6.2), analogous to the calculation of Σ_{eff} given above. Performing the algebra we obtain, after re-arranging terms,

$$\Sigma_{\text{eff,NAL}}^2 = r^2\left(\Sigma^2 + \frac{\sigma_r^2}{N_f}\right) + (1-r)^3\,\frac{\sigma_r^2}{N_m} + (1-r)^2\left(\Sigma_{\text{ref}}^2 + \frac{\sigma_{\text{reg}}^2}{N_m}\right) + (1-r)\frac{\sigma_{\text{table}}^2}{N_f} \qquad \text{(A6.3)}$$

where $r = N_m / N_f$, and $\sigma_{\text{reg}}^2 = E_P(\sigma_{p,\text{reg}}^2)$, i.e. the mean registration variance. The first term in brackets in equation (A6.3) is the systematic variance without corrections (Σ_{eff}^2) we derived above. Obviously, if one waits with correcting until the treatment is over, $r = 1$ and $\Sigma_{\text{eff,NAL}}^2 = \Sigma_{\text{eff}}^2$. The second term is the residual systematic dispersion due to the inaccuracy of the estimate of S_p because of random setup errors. The third term and fourth term stem from respectively the measurement and table translation inaccuracies.

We now simplify Eq. (A6.3), by observing that errors in the reference definition should and can be reduced to a minimum in the pre-treatment phase [76, 122, 134] and we may neglect them. Furthermore, in general the random image registration error $\sigma_{\text{reg}} \leq 1$ mm [17, 49, 124, 130] while $\sigma_r \geq 2$ mm and $\Sigma \geq \sigma_r$ [17, 22, 28, 103, 130, 138], so that the impact of σ_{reg} on $\Sigma_{\text{eff,NAL}}$ is also small. Finally, the SD of table translation errors can be very small [16, 28, 58] and the $1/N_f$ multiplier in Eq. (A6.3) further reduces this error. We therefore neglect the registration and correction errors terms, and only retain the patient setup errors to arrive at

$$\Sigma_{\text{eff,NAL}}^2 = r^2\left(\Sigma^2 + \frac{\sigma_r^2}{N_f}\right) + (1-r)^3\,\frac{\sigma_r^2}{N_m} \qquad \text{(A6.4)}.$$

The expression in Eq. (A6.4) for $\Sigma_{\text{eff,NAL}}$ was used in the calculation of the curve for the NAL protocol in Fig. 6.2.

In the case of setup corrections, we can define the systematic error of patient p in fraction f: $S_{p,f} = S_p + c_{p,f}$ (here, we again neglect the random correction errors $\tau_{p,f}$ in $c_{p,f}$). In the text, we study the time behavior of the systematic errors by calculating $\Sigma_f = \text{SD}_P(S_{p,f})$. For the NAL protocol, it is easy to see that $\Sigma_f = \sigma_r / \sqrt{N_m} \equiv \Sigma_{\text{NAL}}$ for $f > N_m$. With this expression, we can rewrite Eq. (A6.4) as

$$\Sigma_{\text{eff,NAL}}^2 = r^2 \Sigma_{\text{eff}}^2 + (1-r)^3 \Sigma_{\text{NAL}}^2 \qquad \text{(A6.5)}.$$

Note that for some parameter choices, $\Sigma_{\text{eff,NAL}}^2 < \Sigma_{\text{NAL}}^2$, i.e. the dispersion of the *effective* systematic errors over the entire treatment can be (slightly) smaller than the dispersion of the *actual* systematic errors after correction. However, for typical fractionation schemes with $N_f \geq 30$ and NAL protocols with $N_m < 5$, (A6.5) can be approximated by $\Sigma_{\text{eff,NAL}} \approx \Sigma_{\text{NAL}}$.

Cumulative distribution of the 3D systematic setup error in the NAL protocol

The expression (A6.3) for $\Sigma_{\text{eff,NAL}}$ can be applied to each direction i of patient translation to yield $\Sigma_{\text{eff,NAL},i}$. If we assume that the setup and measurement errors have normal distributions, so will $S_{p,\text{eff,NAL},i}$. Hence, the cumulative probability distribution of the length d_{Nf} of the 3D systematic error vector, $V_{Nf} \equiv (S_{p,\text{eff,NAL},1}, S_{p,\text{eff,NAL},2}, S_{p,\text{eff,NAL},3})^T$, is given by

$$P(d_{Nf} > d_0) = \kappa \iiint\limits_{\sqrt{x_1^2 + x_2^2 + x_3^2} > d_0} dx_1\, dx_2\, dx_3\; e^{-\frac{1}{2} \sum\limits_{i=1}^{3} \left(\frac{x_i}{\Sigma_{\text{eff,NAL,i}}} \right)^2} \tag{A6.6}$$

with $\kappa = (2\pi)^{-3/2} \prod_{i=1}^{3} \Sigma_{\text{eff,NAL},i}^{-1}$

We now assume that the NAL protocol results in isotropic effective systematic dispersions: $\Sigma_{\text{eff,NAL},i} \equiv \Sigma_{\text{iso}} \forall i$. In Eq. (A6.4) and below, we show that this condition is approximately fulfilled if the random setup errors are isotropic: $\sigma_{r,i} \equiv \sigma_{\text{iso}} \forall i$. Under this commonly occurring condition (e.g. Table 6.1), we can use spherical coordinates to cast equation (A6.6) into

$$P(d_{Nf} > d_0) = \sqrt{\frac{2}{\pi}} \int\limits_{d_0/\Sigma_{\text{iso}}}^{\infty} dr\, r^2 e^{-\frac{r^2}{2}} \tag{A6.7}$$

which leads to

$$P(d_{Nf} > r\Sigma_{\text{iso}}) = \sqrt{\frac{2}{\pi}} r\, e^{-\frac{r^2}{2}} + 1 - \text{erf}(\tfrac{1}{\sqrt{2}} r) \tag{A6.8}$$

where "erf" represents the standard error function. Eq. (A6.8) implies that $(d_{Nf}/\Sigma_{\text{iso}})^2$ has a χ^2 probability distribution with 3 degrees of freedom [123]. This result can be understood from the normality of each of the components of the systematic error vector.

The analytical prediction in Eq. (A6.8) is also quite accurate for anisotropic setup deviations. In Fig. 6.6 we show the result of MC simulations for setup distributions with 50% anisotropy in the initial systematic and random errors: $\Sigma_i = \sigma_{r,i} \equiv 0.5(1 + 0.5i)\sigma_{\text{iso}} (i = 1,2,3)$ and compare it with the result of Eq. (A6.8) with $\Sigma_{\text{iso}} = \sqrt{\frac{1}{3} \sum_{i=1}^{3} \Sigma_{\text{eff,NAL},i}^2}$. The difference in the cumulative distributions between the MC simulation and analytical approximation is everywhere less than 2% (2-3 SD of the statistical noise on the simulated data) and on average only 0.2%. Therefore, even for significant anisotropy, expression (A6.8) yields an accurate estimate of the final cumulative distribution of the 3D systematic errors.

Net effect of setup corrections based on a single measurement

Below we show that a setup correction based on a single (e.g. first-day) measurement has no impact on the distribution of systematic setup errors if the systematic and random setup errors have similar distributions. For clarity, we treat the 1D case explicitly, but the proof can be easily extended to higher dimensions, as will be shown. We will drop some of the subscripts of previously introduced variables if they are not strictly necessary.

Let the probability density function (PDF) of the systematic setup errors S be $p_s(S)$ and of the random setup errors Δ be $p_r(\Delta)$. We assume that the systematic and random errors have an identical PDF and that this PDF is an *even* function:

$$p_s(x) = p_r(x) \equiv p(x), \qquad p(x) = p(-x), \qquad \int_{-\infty}^{\infty} p(x)dx = 1 \qquad (A6.9).$$

A setup correction based on a single measurement of the actual setup error, $X = S + \Delta$, would in general depend on the magnitude of X. In fact, for action level α, if $|X| > \alpha$, we would apply a setup correction equal to $-X$ (assuming we don't use a priori information about the population distributions [45] resulting in a post-correction systematic error of $-\Delta$, otherwise the systematic error remains S. We now derive the PDF $p'(S')$ of the post-correction systematic errors S'.

The probability that the post-correction systematic error is in the infinitesimal interval $(S', S' + dS')$ is $p'(S')dS'$. This probability consists of two contributing probabilities. The first contribution is from uncorrected patients for which, as pointed out above, $S' = S$. Hence, the corresponding probability p'_{uncor} equals the probability that the systematic error before correction is in the interval $(S', S' + dS')$ times the probability that the total setup error is not corrected for, given this systematic error. Hence,

$$p'_{\text{uncor}}(S')dS' = p_s(S')dS' P(|S+\Delta| < \alpha \quad |S = S') = p_s(S')dS' \int_{|S'+\Delta| < \alpha} p_r(\Delta)d\Delta \qquad (A6.10),$$

where the standard notation for the conditional probability $P(A|B)$ denotes the probability that A occurs given B. The second contribution to p' comes from patients whose setup was corrected. For these patients, $S' = -\Delta$. Therefore, the corresponding probability p'_{cor} equals the probability that the random setup error is in the interval $(-S', -S' - dS')$ times the probability that the total setup error is not corrected for:

$$p'_{\text{cor}}(S')dS' = p_r(-S')dS' P(|S+\Delta| > \alpha \quad |\Delta = -S') = p_r(-S')dS' \int_{|S-S'| > \alpha} p_s(S)dS \quad (A6.11).$$

Eqs. (A6.10) and (A6.11) hold in general, irrespective of the PDFs. If we now apply the PDF properties in Eq. (A6.9), we obtain

$$p'(S') = p'_{\text{uncor}}(S') + p'_{\text{cor}}(S') = p(S')\left(\int_{|x+S'| < \alpha} p(x)\,dx + \int_{|x-S'| > \alpha} p(x)\,dx \right) = p(S') \qquad (A6.12),$$

i.e. the PDFs of the post-correction systematic errors is identical to the initial PDF. Hence, the distribution of the systematic errors is unaffected by the corrections, irrespective of the magnitude of the action level α.

We sketch the extension of the proof to multiple dimensions. For a single measurement in an N-dimensional space, the error vector becomes $(X_1,...,X_N) = (S_1 + \Delta_1,..., S_N + \Delta_N)$. The PDFs of the systematic and random setup errors are $p_s(S_1,...,S_N)$ respectively $p_r(\Delta_1,...,\Delta_N)$. As before, these PDFs are assumed to be identical: $p_s(x_1,...,x_N) = p_r(x_1,...,x_N) \equiv p(x_1,...,x_N)$. We concentrate on the PDF of the systematic error along a single axis, for which we arbitrarily select the first axis. We define its PDF by

$$p_1(S_1) \equiv \int_{R_{N-1}} p(S_1, S_2,..., S_N)dS_2...dS_N \tag{A6.13},$$

where the integration is over the entire range of $(S_2,...,S_N)$. Furthermore, we assume that the PDF is an even function along this axis, $p(x_1,...,x_N) = p(-x_1,...,x_N)$. Next, we extend the criterion for setup corrections, $|X| > \alpha$ for 1 dimension, to N dimensions. We consider two generally applied types of criteria. The first type is a combined test on the components of the error vector and the criterion for correction is $f(|X_1|,...,|X_N|) > \alpha$ where f is a function of N variables. For instance, in the SAL protocol one applies $f = \sqrt{X_1^2 +...+ X_N^2}$ to evaluate the first measurement. The second type tests
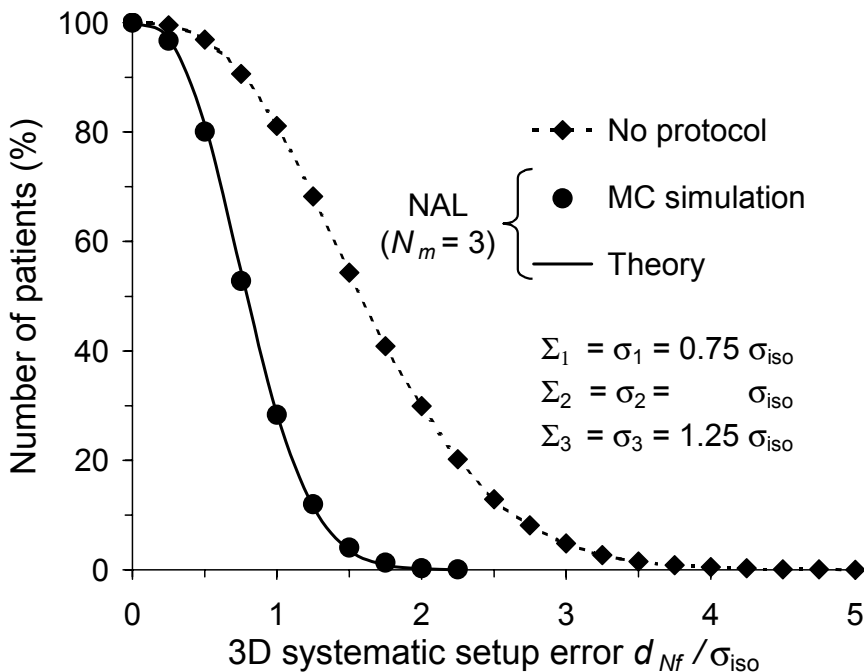
Figure 6.6 Cumulative distributions of the 3D systematic setup errors d_{Nf}, for a patient population with an anisotropic distribution of systematic and random setup errors. The SDs of the initial setup errors are indicated in the figure. The result from a MC simulation of the NAL OVP (closed circles) is given together with the analytical approximation from the Appendix (solid line).

individual vector components: the correction criterion is $|X_i| > \alpha_i$ for any $i \in [1,...,N]$, e.g. test if the setup error measured in a single fraction exceeds 5 mm in any direction. For both types of criteria, we can express the PDF $p_1'(S_1')$ of the post-correction systematic error along the first axis, S_1', analogous to Eq. (A6.10) and (A6.11). Because these criteria only depend on the *absolute* values of the setup errors, we obtain

$$p_1'(S_1') = p_1(S_1') \qquad\qquad (A6.14).$$

Because the first axis was arbitrarily chosen, Eq. (A6.14) implies that the PDF of systematic setup errors is unaltered by the corrections for any direction along which the PDF is an even function.

7 A NEW APPROACH TO OFF-LINE SETUP CORRECTIONS: COMBINING SAFETY WITH MINIMUM WORKLOAD

J.C.J. de Boer and B.J.M. Heijmen

Abstract

Purpose: Off-line patient setup correction protocols based on electronic portal images are an effective tool to reduce systematic patient setup errors. Recently, we have introduced the No Action Level (NAL) protocol which establishes a significant error reduction at a very small workload. However, this protocol did not include an explicit verification of the applied setup corrections. Systematic mistakes in the execution of setup corrections (e.g. a setup correction is always executed in the +X direction whereas a correction in the –X direction was prescribed) may introduce large systematic setup errors (irrespective of the setup protocol) and may seriously impair treatment outcome. We have therefore extended the NAL protocol with a correction verification (COVER) stage, solely aimed at detecting such mistakes.

Methods and Materials: In short, COVER tests the magnitude of the post-correction setup error in each relevant direction. If these residue errors are below the acceptance threshold T, no more EPI are required and the protocol has finished. If not, the origin of this result should be investigated; if no obvious mistakes are present, the procedure is repeated for one more treatment fraction. If the residue setup errors are confirmed to be larger than T, the entire protocol is restarted. Using both Monte Carlo simulations and analytical calculations, we performed a risk analysis and evaluated the workload for various choices of T.

Results: A threshold $T = 3 \times \sigma_r$, where σ_r is the mean standard deviation of the random setup errors, ensured that (1) COVER introduces only a small additional workload (1.05 measurement per patient, while the absolute minimum is 1.0) and (2) serious correction mistakes are detected with high probability. Even if setup corrections are wrongly applied in each patient (worst case scenario), COVER ensures that the final distribution of systematic errors is not wider than the pre-correction distribution of systematic errors; for realistic frequencies of correction mistakes (<< 1 per patient) this distribution becomes much more narrow.

Conclusion: The combination of NAL and COVER provides a highly efficient as well as safe method to reduce systematic setup errors.

7.1 Introduction

Due to the increasing complexity of radiotherapy techniques, the use of electronic portal images (EPIs) for verification of both geometrical and dosimetrical aspects of external beam radiotherapy treatment is a rapidly expanding field of research [72]. EPIs present simultaneous information on both patient anatomy and exit dose distributions during treatment in digital format. Therefore, electronic portal imaging devices (EPIDs) have been applied to various quality assurance related topics, ranging from verification of basic Linac properties such as beam output [38] and light field congruence [77] to dosimetric verification of dynamically delivered fluence distributions [95, 97], compensator design and verification [44, 101] and patient transit dosimetry [100]. The flat panel a-Si EPIDs now commercially available may improve image quality over conventional devices at significantly lower exposures per image [6], even allowing for acquisition of CT scans at the treatment unit [64]. Despite the active development of these promising techniques, the basic application of EPIDs in patient setup verification is only slowly growing, despite routine use in a number of centers [17, 28, 56, 75, 146]. Recently, AAPM Task Group 58 presented a report providing basic clinical and physics information aimed at lowering the barrier for clinical EPID use [56]. Part of this barrier is the workload generated by EPI-based setup correction protocols (SCPs) in relation to their benefits. To facilitate use of SCPs in our clinic, we have developed a low workload SCP which is highly effective in reducing systematic setup errors [24]. To promote clinical use of this SCP, we have added a verification stage to determine if setup corrections are properly executed. The design of an efficient verification stage, and its effect in combination with an SCP, is the topic of this paper.

Many approaches to verify and correct patient setup based on EPID images have been proposed. They can be separated into "off-line" and "on-line" SCPs [56, 116]. This classification refers to the moment at which setup corrections are performed. For off-line SCPs [13, 24, 145] the information from EPIs obtained in a certain treatment fraction is not applied to improve the setup of that fraction but, if necessary, a setup correction is performed in subsequent fractions. In contrast, on-line SCPs [91] require that EPIs are obtained in each fraction with a small percentage of the fraction dose, followed by image analysis and setup correction before the remaining fraction dose is delivered. The result of these two SCP types can be related to the division of setup inaccuracies into "systematic" and "random" errors [22]. The systematic setup error of a patient is the difference between the mean setup during treatment and the intended setup defined by the treatment plan. A random error is an incidental error specific to the setup at a certain treatment fraction, which is present in addition to the systematic error. It follows that off-line SCPs correct for systematic errors: the setup corrections do not take into account the random errors in the fractions in which they are applied. Off-line SCPs therefore aim at estimating the systematic error from a limited set of EPI measurements by filtering out the effect of random errors. In contrast, on-line SCPs may correct both systematic and random setup errors, at the cost of daily imaging and the need to perform accurate image analysis and setup corrections under time pressure.

To determine how random and systematic setup errors impact the dose distribution, Stroom *et al.* [125] analyzed the geometrical margin between the clinical target volume (CTV) [62] and planning target volume (PTV) in a statistical approach. They found that, if the standard deviation (SD) of the systematic setup errors in a patient population is denoted

by Σ and the SD of random errors by σ, a margin of $2\Sigma + 0.7\sigma$ yields an average coverage of 99% of the CTV with at least 95% of the prescribed dose for patients in that population. Hence, the reduction of systematic errors established by off-line SCPs is an efficient method to reduce the required planning margins. In addition, a much smaller number of fractions need to be imaged and analyzed in off-line than in on-line SCPs [24, 56]. Therefore, although on-line SCPs are useful in some cases, off-line SCPs are an obvious choice if workload is critical.

Because off-line SCPs may still introduce considerable workload when applied in many patients, we investigated the efficiency of off-line corrections in previous work [24]. The most widely applied SCP to date is the shrinking action-level (SAL) protocol [13], developed and clinically tested in a multi-site collaboration in the Netherlands [17]. In our institute, the setup of over two thousand patients has been verified and corrected using this protocol [17, 24, 27, 28, 93]. We found that careful EPI analysis is the major contributing factor to the SCP workload per patient, not the acquisition of the images or the execution of setup corrections [17, 24]. This is largely due to the fact that 3D setup evaluation requires multiple images per fraction to be analyzed whereas setup corrections can be executed rapidly and even automatically by treatment couch translations after the regular patient setup procedure [16, 130]. Hereafter, we will refer to all workload related to EPI acquisition and analysis as EPI workload.

Because clinical implementations of the SAL SCP required analysis of a considerable amount of EPIs per patient (for instance, $\sim$ 20-30/patient for prostate treatments) [24], we developed a new SCP to use EPI information more efficiently. This "no action-level" (or NAL) SCP enables the same reduction of systematic setup errors as the SAL SCP with approximately 3 times less images [24] and a corresponding reduction in EPI workload. This is achieved by establishing sufficiently accurate setup corrections very early in treatment for each patient [24]. Apart from a reduction in workload, the small number of imaged fractions required by NAL has the advantage that if dedicated fields (e.g. as in double exposures) are necessary for setup verification [27, 28], the cumulative dose to healthy tissue from such fields can be kept minimal. Furthermore, for the NAL SCP this dose is fully predictable *a priori*, in contrast to the SAL SCP [13].

Prior to implementation of the NAL SCP in our clinic, we developed the procedure described in this paper to detect errors in the execution of the required setup corrections. If a *systematic* mistake (recurring each fraction) is made in the execution of setup corrections (for instance because a number is copied with the wrong sign onto a form) the resulting systematic setup error may increase considerably, with possibly dramatic consequences. Even though quality assurance procedures should guarantee a small incidence of such mistakes, they may still occur and the potentially large impact on treatment outcome warrants a correction verification stage. At present, little attention has been given to explicit correction verification and particularly to the effect of *systematically* wrongly applied corrections. In contrast, the effect of *random* inaccuracies in the setup corrections has been studied [13, 24, 49] but the impact of such inaccuracies can be shown to be small [24]. We have therefore developed a correction verification stage to detect systematic mistakes in the execution of setup corrections. This stage consists of a test on the measured post-correction setup error. Because false detections of mistakes will incur most of the workload if corrections are properly executed, the test should have a high sensitivity to severe correction mistakes while maintaining a small false detection rate. Although the proposed verification stage is designed with the NAL SCP in mind, it can in principle be applied to any off-line SCP.

7.2 Methods and Materials

7.2.1 Descriptive model of setup errors and setup corrections

We will adopt the description of setup errors established in previous work. Below we give a short overview of the relevant quantities; details on how we measure and derive these quantities can be found elsewhere [17, 24, 28]. Because of the relevance to the procedures developed in this paper, we will pay some attention to the distribution of random setup errors in the patient population.

Consider a patient p who is treated in fractions $f \in F = \{1, 2, ..., N_f\}$. The setup error with respect to the reference setup along axis x is $x_{p,f}$. We describe the setup errors by $x_{p,f} = S_p + \Delta_{p,f}$ where S_p is defined as the systematic error (the recurring component) and $\Delta_{p,f}$ the random error (fluctuating from fraction to fraction) for patient p. Because the number of treatment fractions is finite, we define the *effective* systematic error [24] as the actual mean error over the treatment fractions:

$$S_{p,\text{eff}} = < x_{p,f} >_F \equiv N_f^{-1} \sum_{f=1}^{N_f} x_{p,f} \,. \tag{7.1}$$

Hence, $S_{p,\text{eff}}$ tends to S_p if N_f becomes large. The random errors $\Delta_{p,f}$ of patient p are characterized by their SD, denoted by σ_p (note that σ_p may be different in each patient.)

Bijhold *et al*. [22] introduced the use of the symbols Σ and σ to denote SDs of respectively systematic and random deviations measured in a patient population and we adopt their convention. The probability density functions of setup errors occurring in a patient population P can then be approximately described by the *mean systematic error μ*, the *systematic dispersion Σ* and the *mean random dispersion σ_r*, defined by

$$\mu = \left\langle S_p \right\rangle_P, \quad \Sigma = \text{SD}_P(S_p), \quad \sigma_r = \sqrt{\left\langle \sigma_p^2 \right\rangle_P} \tag{7.2}$$

where $\left\langle A \right\rangle_B$ is the mean of quantity A calculated over domain B and $\text{SD}_B(A)$ the corresponding standard deviation. The expression for σ_r yields a mean random error magnitude, useful if the inter-patient variation in the values of σ_p is small [17, 28]. However, non-negligible inter-patient variation in σ_p does occur and for our purpose we require a more detailed description of random errors. We assume that the probability distribution of σ_p in the patient population is normal, a choice validated in the Appendix. Consequently, the distribution is determined by the mean value of σ_p and its SD, denoted by respectively σ_{mp} and $\delta\sigma_p$. In the Appendix, we describe how we obtain these values from measured patient data. Although the normal distribution allows for a positive probability for the nonsensical values $\sigma_p < 0$, the parameters obtained in the Appendix ensure that this probability is negligible. For such a distribution, σ_r approaches $\sqrt{\sigma_{mp}^2 + \delta\sigma_p^2}$ (Eq. (7.2)) if calculated for a sufficiently large data-set. Despite this detailed description of σ_p, for theoretical purposes an analysis for a population without inter-

patient variation in random errors ($\delta\sigma_p = 0$) is often useful. Such a hypothetical patient population will be referred to as *iso-random population*.

Based on the above, the dispersion of the effective systematic errors if no setup corrections have been applied can be derived [28]

$$\Sigma_{\text{eff}} = \text{SD}_P(S_{p,\text{eff}}) = \sqrt{\Sigma^2 + \sigma_r^2 N_f^{-1}} \ . \tag{7.3}$$

In many studies Σ was found to be comparable to σ_r, [17, 22, 24, 27, 28, 49, 75, 93, 125, 130] so that for treatments consisting of more than 10 dose fractions, $\Sigma_{\text{eff}} \approx \Sigma$.

An off-line setup SCP will generate setup corrections $c_{p,f}(\{x_{p,f'}\})$ for fraction f based on measured setup deviations $\{x_{p,f'}\}$ in previous fractions $f' < f$. To take into account errors in the execution of setup corrections, we write $c_{p,f} = c_{p,f}^{\text{SCP}} + \tau_{p,f}$ where $c_{p,f}^{\text{SCP}}$ is the correction as calculated by the SCP, and $\tau_{p,f}$ is the inaccuracy in the execution of this correction. The setup deviations that occur when a SCP is applied ($x_{p,f}^{\text{SCP}}$) can now be described in terms of the setup deviations $x_{p,f}$ that would have occurred without setup corrections:

$$x_{p,f}^{\text{SCP}} = x_{p,f} + c_{p,f} \tag{7.4}$$

The above definitions apply to each relevant setup direction. We assume that there is no correlation between the setup errors in different directions. An extended description of measured and actual setup errors, explicitly taking into account image registration inaccuracies and random inaccuracies in the execution of setup corrections, was given elsewhere [24].

The way setup corrections are executed is essential to the current analysis. We assume the method clinically applied in our hospital, which can be summarized as follows. In each fraction the patient is initially positioned according to the skin tattoos or other demarcations established in the treatment preparation phase. Next, if a setup correction is required, a couch shift is performed either manually using the digital read-out of the treatment couch or automatically from the Linac couch-control software [130]. This way, the accuracy with which patients are positioned initially is the same as if no setup corrections would be applied. Therefore, the setup correction can be regarded as an additive operation, which has an accuracy which depends on the treatment couch precision and not on the patient specific setup. If only random inaccuracies in the execution of treatment couch shifts occur, $\tau_{p,f}$ can be characterized by the SD of these inaccuracies, hereafter referred to as σ_{couch} .

7.2.2   *The no action level (NAL) setup correction protocol*

The NAL protocol represents the simplest form of off-line SCP. We will describe this protocol below in some detail, as its application is the motivation behind our correction verification procedure. We will therefore evaluate the effect of the latter procedure in combination with the NAL SCP.

As described in the Introduction, an off-line SCP must filter out the effect of random errors to estimate the systematic error and generate a correction for subsequent fractions. The obvious method is to average measured setup errors, and the NAL SCP implements this method in a straightforward way. For each patient, setup errors are measured during the

first N_m fractions, N_m being a fixed number typically < 5 [24]. The average setup error over the first N_m fractions is determined, and a setup correction equal to minus this error is applied in all subsequent fractions ($f > N_m$), irrespective of its magnitude (see comment below). The protocol is then finished, and thus the number of imaged fractions per patient always equals N_m. In terms introduced in the previous section, the NAL SCP is described by $c_{p,f}^{\mathrm{SCP}} = 0$ for $f \leq N_m$ and $c_{p,f}^{\mathrm{SCP}} = c_p^{\mathrm{NAL}}$ for $f > N_m$ where

$$c_p^{\mathrm{NAL}} = -N_m^{-1} \sum_{f=1}^{N_m} x_{p,f} = -S_p - N_m^{-1} \sum_{f=1}^{N_m} \Delta_{p,f} \equiv -S_p + \delta_p . \tag{7.5}$$

As a result, the *residue systematic error* after correction equals δ_p, and the SD of these errors over the patient population P equals [24]

$$\Sigma_{\mathrm{res,NAL}} = \sigma_r / \sqrt{N_m} . \tag{7.6}$$

Equation (7.6) expresses a fundamental property of off-line SCPs: the initial systematic setup error is replaced by the inaccuracy in the estimate of that systematic error, [45] in this case due to random setup errors in the fractions of measurement.

To calculate the net effect of the protocol, both pre-correction and post-correction setups must be taken into account. When the NAL SCP is applied, the dispersion of the effective systematic errors (Eq. (7.1)) becomes

$$\Sigma_{\mathrm{eff,NAL}}^2 = r^2 \left(\Sigma^2 + \frac{\sigma_r^2}{N_f} \right) + (1-r)^3 \frac{\sigma_r^2}{N_m} \tag{7.7}$$

where $r = N_m / N_f$ [24]. As discussed below, a clinically realistic value is $r \approx 0.1$, in which case $\Sigma_{\mathrm{eff,NAL}}$ tends to the expression for the post-correction dispersion of Eq. (7.6).

From Monte Carlo simulations, analytical calculations and retrospective application to large databases of measured patient setup data [24, 27], we found that the dispersions given by Eq. (7.6) and (7.7) are quite small compared to those obtained with the SAL SCP with a similar average number of imaged fractions per patient. The reasons behind this relative inefficiency of the SAL SCP have been treated elsewhere [24]. Here we only state that the NAL SCP can reduce systematic setup errors with typically 3 times less measurements per patient than the SAL SCP. Appropriate values of N_m are typically in the range 2-4 [24, 27]. Hence EPI analysis workload as well as dose to healthy tissue in dedicated setup verification fields can be reduced accordingly.

Equations (7.5-7) hold if setup corrections are executed properly. If a recurrent error is made in the execution of the setup corrections, the obtained result will usually be worse. Therefore, a separate correction verification stage was developed, described in the next section.

7.2.3 The correction verification (COVER) procedure

Hereafter, we will refer to the average number of fractions that must be imaged per patient as the required number of measurements or N_{meas}. We assume that the number of EPIs that must be acquired and analyzed per imaged fraction is fixed, and hence the EPI workload will be proportional to N_{meas}. If correction verification (hereafter, COVER) is applied after the NAL protocol, N_{meas} will be at least $N_m + 1$, irrespective of the details of such a

verification. We aim to develop a COVER procedure which increases this minimum imaging workload by as little as possible but will still detect 'significant' correction errors with a high probability. We will focus on detection of systematic (recurrent) errors in the execution of corrections since random correction errors have a small impact on the final systematic setup errors [24]. Hereafter, recurrent errors in the execution of corrections are referred to as *correction mistakes*. The applied COVER approach may be summarized as follows.

In the first fraction in which the setup correction generated by the SCP is applied, the setup error is measured. Our strategy is that treatment may continue without further verification if this setup error is sufficiently small. To decide if the correction is acceptable, the setup error in this first post-correction measurement is subjected to a test, hereafter referred to as Test 1. Test 1 must alert the operators with a high probability if an actual correction mistake occurred, so that its origin can be identified and corrected for in subsequent fractions. If Test 1 indicates the presence of a mistake, but no specific cause can be identified, the setup error is measured again in the next fraction (the same setup correction is still applied). The setup error measured in this fraction is tested again (Test 2) to verify if the result of Test 1 reproduces in this fraction. If so, the SCP is restarted; otherwise the setup correction is accepted and will be applied in next fractions without further verification.

If Test 1 does not detect an actual correction mistake, we will refer to such an event as a "Type II error" in accordance with hypothesis testing conventions [123]. Because Test 1 is based on a single measurement, a small frequency of Type II errors implies it will generate some false alerts or "Type I errors". Hence, Test 2 must discriminate between a true mistake or a Type I error in Test 1.

7.2.3.1 *First verification measurement*

If, for a patient p, the first correction verification measurement is applied in fraction $v1$ (referring to 'verification 1'; generally $v1 = N_m + 1$), the setup error in that fraction ($x_{p,v1}^{\text{cor}}$) may be expressed as

$$x_{p,v1}^{\text{cor}} = S_p + \Delta_{p,v1} + c_{p,v1} = \Delta_{p,v1} + \delta_p + \tau_{p,v1} . \tag{7.8}$$

The second equality follows from the expression $c_{p,v1} = c_p^{\text{NAL}} + \tau_{p,v1}$, where $\tau_{p,v1}$ is the error in the executed correction (section 7.2.1), combined with Eq. (7.5). Therefore, $x_{p,v1}^{\text{cor}}$ is the sum of (1) the random setup error $\Delta_{p,v1}$ in fraction $v1$, (2) the correction execution error $\tau_{p,v1}$ in that fraction and (3) the average random setup error δ_p in the N_m fractions that determine c_p^{NAL}. In case the correction execution errors would be limited to random inaccuracies in treatment couch shifts, the SD describing the dispersion in $x_{p,v1}^{\text{cor}}$ becomes

$$\sigma_{p,v1} = \sqrt{(1 + N_m^{-1})\sigma_p^2 + \sigma_{\text{couch}}^2} \; . \tag{7.9}$$

To verify if the correction was properly executed, we test if the measured value of $x_{p,v1}^{\text{cor}}$ is likely given the above SD. As zero hypothesis H_0 we assume that no correction mistake occurred. Test 1 is defined by

$$\text{Accept } H_0 \text{ if} |x_{p,v1}^{\text{cor}}| < T . \tag{7.10}$$

where T is a threshold for acceptance. Test 1 must be applied to each setup direction separately and therefore in principle 3 different values of T may be involved. In this paper, we will assume T is constant for the patient population (for a patient specific choice of T, see Discussion). The main topic of this paper is to find an appropriate expression for T in the population averaged SD of the random errors.

If H_0 is not accepted by Test 1, it is not definitely rejected either because $x_{p,v1}^{\text{cor}} > T$ may occur due to random setup errors. Therefore, if the acceptance condition expressed in Eq. (7.10) is not fulfilled, an alert for a *possible* correction mistake is given to the operators by the COVER protocol software. In the next section we will develop the actions required in case of such an alert.

7.2.3.2 Second verification measurement and completion of COVER

If Test 1 generates an alert, the operators will search for a potential correction mistake (for specific actions, see section 7.4). If an actual mistake is involved, it will in general be identified and solved, i.e. it will not recur in subsequent fractions. EPIs are then obtained for the first next fraction to verify if the correction is indeed executed properly, and therefore Test 1 will be re-applied to the setup errors measured in these EPIs.

However, operators may also fail to find the cause of an alert. In this case, we can perform another verification measurement in the next fraction $v2 = v1 + 1$ and apply another test (Test 2) to accept or definitely reject H_0. To specify this test, we first summarize the possible reasons of alerts without traceable causes:

(1) A Type I error occurred (H_0 was not accepted while true) due to random setup errors.

(2) A *random* correction error other than a couch translation inaccuracy occurred. For example, a setup correction of –4 mm is required in a specific direction, but the operator executes a +4 mm couch shift. This may be an incidental error, not recurring in subsequent fractions. The magnitude and behavior of this type of error is hard to predict, but formally it may be regarded a Type I error because it does not involve a *systematic* correction mistake. In principle, such errors can be eliminated if treatment couch shifts are executed via the treatment unit control software [130].

(3) A systematic correction mistake is actually present, but its origin is not obvious.

If the SCP was properly introduced and tested, the incidence of correction mistakes should be very small. The subset of mistakes that are detected in Test 1 but subsequently can not be solved will be even smaller. Nevertheless, reason (3) should lead to an alert generated by Test 2, particularly if a large correction mistake is involved. Based on these considerations, we choose Test 2 to be similar to Test 1:

$$\text{Reject } H_0 \text{ if} |x_{p,v2}^{\text{cor}}| > T \tag{7.11}$$

where $x_{p,v2}^{\text{cor}}$ is the setup error measured in fraction $v2$. Hence, to definitely reject H_0, Test 2 requires that the result of Test 1 is reproduced. If more than one setup direction is involved, Test 2 is restricted to those setup directions for which H_0 was not accepted in Test 1. If H_0 is accepted in Test 2 for all involved setup directions, COVER has ended and no more verification measurements are required. However, if H_0 is rejected for any

direction, the presence of a correction mistake has been confirmed. Since the origin of this mistake is unknown, we propose a restart of the entire procedure, including off-line SCP. Such a restart is reasonable because if, after careful searching, no origin of a correction mistake was found, no actual mistake may be present. Instead, the prescribed setup correction itself may be inaccurate, not because of the regular random setup inaccuracies (Eq.(7.5)) but due to an actual mistake in either the setup or the setup error measurement that occurred during one or more of the pre-correction setups. If this mistake is not reproduced in fractions $v1$ and $v2$, both Test 1 and 2 may reject H_0 despite correct execution of the prescribed correction. In this sense, COVER also tests for pre-correction setup and image analysis mistakes. In addition, a small but unavoidable fraction of the protocol restarts will be due to random setup errors leading to Type I errors in both Tests 1 and 2 (see next section).

In Fig. 7.1, we depict the combination of off-line SCP and COVER in a flowchart. The off-line SCP has been left unspecified in this chart. The only constraint on the SCP is that it
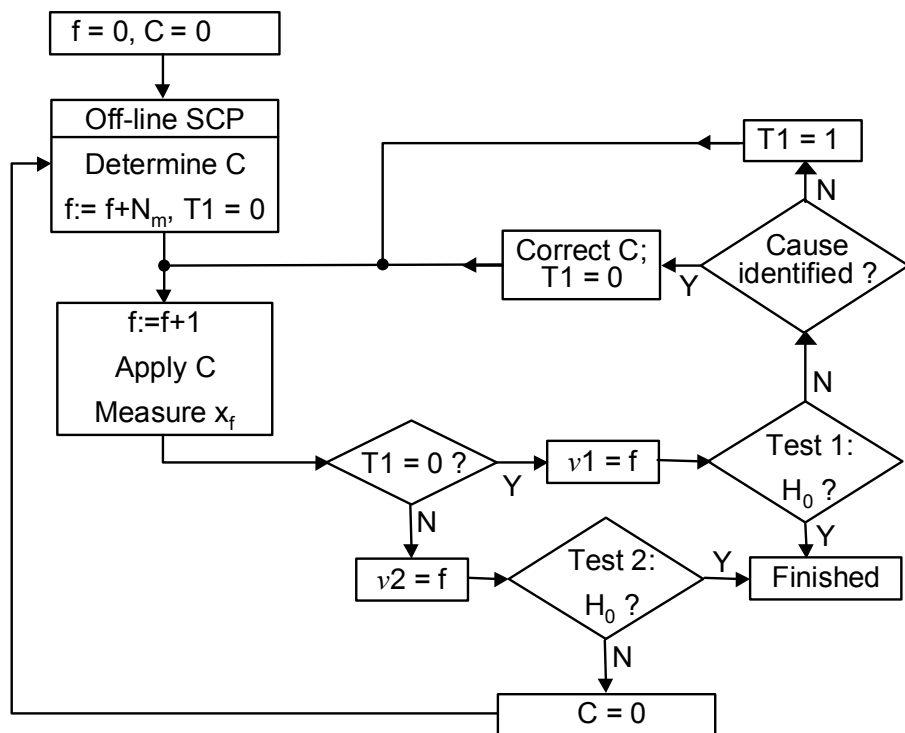
Figure 7.1 Flowchart of the off-line setup correction protocol (SCP) and the correction verification (COVER) stage. Treatment fractions are numbered by f, the setup error measured in that fraction is x_f, and the setup correction generated by the SCP is C. The fraction to which Test 1 respectively Test 2 are applied are indicated by $v1$ respectively $v2$ and T1 is a help variable which indicates if Test 1 is to be applied (if T1 = 0). Test 1 and Test 2 verify whether hypothesis H_0 (setup correction correctly executed) can be accepted. Further details are in the text.

should generate a unique setup correction after N_m fractions, but in principle N_m may be different for each patient. Below, we study COVER together with the NAL protocol and the full procedure will be referred to as NAL+COVER.

7.2.4 Verification of NAL+COVER with Monte Carlo simulations

In order to assess the performance of the COVER stage, we must revert to Monte Carlo (MC) simulations. Although the incidence of Type I errors may be evaluated analytically in some cases (Appendix), the impact of COVER in the presence of correction mistakes can only be properly evaluated using MC simulations. We have therefore extended the MC software originally designed for investigation of off-line SCPs [24] with the possibility to introduce correction mistakes and to verify correction executions with the described COVER procedure. Below, we describe the input of these simulations and the method of simulation.

The population setup accuracy is defined by the distribution parameters $(\mu, \Sigma, \sigma_{mp}, \delta\sigma_p)$ that would hold if no setup corrections are applied (section 7.2.1). The random inaccuracies in table couch translations are specified by σ_{couch}. Optionally, the population distribution of linear time trends [17, 49], i.e. gradual shifts in the patient setup, can be specified through the SD of the average shift/fraction. The default distribution parameters for our simulations were obtained from a setup study we performed for a prostate patient population of 600 patients [24]. The setup of all patients was verified with a SAL SCP, and the setup errors that would have occurred without SCP were derived using Eq. (7.4). The resultant systematic errors were nearly isotropically distributed and so we applied the mean of the values measured in the 3 setup directions to each direction in our simulations. Furthermore, the magnitude of the mean systematic error μ was ≤ 1 mm in each direction. Averaged over the 3 directions, $\mu = 0.7$ mm, which was used for each direction. The measured distributions of random errors were also isotropic, reflected in nearly direction independent values of σ_r [24]. In the Appendix, we describe how σ_{mp} and $\delta\sigma_p$, which determine the random error distribution in this population, were estimated from the measured distribution of σ_p. We did not include linear time trends into our simulations, as we have shown that their net effect on the distribution of setup errors was negligible in this population [24]. In Table 7.1, the values for the distribution parameters used in the simulations are given. We will refer to the patient population characterized by these parameters as the *clinical population* as opposed to iso-random populations (section 7.2.1).

In the MC simulations, setup errors $x_{p,f} = S_p + \Delta_{p,f}$ are generated for each patient p and fraction f in each setup direction. First, S_p and σ_p are drawn from normal distributions, respectively $N(\mu, \Sigma)$ and $N(\sigma_{mp}, \delta\sigma_p)$ ($N(a,b)$ denotes the normal distribution function with mean $= a$ and SD $= b$) for each patient. Next, $\Delta_{p,f}$ is drawn from $N(0, \sigma_p)$ for each fraction and the values $x_{p,f} = S_p + \Delta_{p,f}$ are processed by the off-line SCP to generate setup corrections $c_{p,f}^{\text{SCP}}$ for subsequent fractions ("off-line SCP" block in Fig. 7.1). For those fractions in which an actual setup error is applied, the random couch

translation error $\tau_{p,f}$ is drawn from a normal distribution $N(0, \sigma_{\text{couch}})$ after which the simulated setup error $x_{p,f}^{\text{SCP}}$ is calculated according to Eq. (7.4) ("Apply C" block in Fig. 7.1). This process is sufficient to simulate the impact of off-line SCPs on setup error distributions [17, 24] . In this paper, we simulated the NAL SCP with $N_m = 3$ imaged fractions per patient, because this protocol ensures an acceptable distribution of systematic errors ($\Sigma_{\text{eff,NAL}} = 1$ mm) for the clinical population discussed above [24]. Furthermore, a recent analysis of the improvement of tumor control probabilities in prostate patients after off-line setup corrections confirmed that a NAL SCP with $N_m = 3$ is a sensible choice for populations with $\Sigma \approx \sigma$ [4].

Table 7.1 Setup error distribution parameters (mm) for our clinical population of 600 prostate cancer patients, if no setup corrections are applied. These parameters were applied in each of the three setup directions (isotropic distribution of errors) in the MC simulations.

μ	Σ	σ_{mp}	$\delta\sigma_p$	σ_{couch}
0.7	2.3	1.9	0.4	0.5

We have extended the MC simulation with the possibility of correction mistakes occurring in a sub-population of the patients. The fraction of patients with correction mistakes is defined by the mistake incidence ε, which can be selected between 0 and 1. For this sub-population, a correction mistake is introduced from the first fraction in which $|c_{p,f}^{\text{SCP}}| > 0$ (the "Apply C" block in Fig. 7.1). The type of simulated mistakes can be specified. We investigated 3 types, which are listed in Table 7.2 together with corresponding expressions for $c_{p,f}$. All three types may occur if the numerical values of a setup correction generated by software have to be copied manually. Furthermore, mistake types M2 and M3 can occur if the definition of axes (X,Y,Z) in the off-line SCP software is different from operator interpretation of those axes, or from their actual orientation due to a software configuration error (section 7.2.5).

If the COVER stage is simulated, the acceptance threshold T must be entered. If H_0 is rejected in Test 1 while a mistake was actually simulated, the user can select if the cause of this mistake is identified and corrected by the (simulated) operators or not. If so, the "Cause Identified" test in Fig.7.1 always yields "Yes" and the mistake is not present in subsequent fractions; if not, the mistake will reoccur unless it is detected (H_0 rejected) in Test 2. In the latter case, the protocol restarts completely (Fig. 7.1), but correction mistakes will no longer be simulated in subsequently executed setup corrections (i.e., correction mistakes are only simulated the first time the SCP generates a setup correction).

For each MC run, a treatment of 30 fractions per patient was simulated in at least 10000 patients. For small ε, the evaluation of the effects introduced by correction mistakes may suffer from statistical noise. The following method was used to decrease such noise. If the asymptotic value (number of patients $\to \infty$) of a quantity Q for a specific value of ε is $Q(\varepsilon)$, some Q satisfy $Q(\varepsilon) = (1-\varepsilon)Q(0) + \varepsilon Q(1)$. This way, MC results obtained with

10000 or more patients for $Q(0)$ and $Q(1)$ can be used to obtain $Q(\varepsilon)$ for small ε with negligible statistical noise. An example of such a Q is the cumulative distribution of 3D systematic errors described below.

Table 7.2 Correction mistakes introduced in the MC simulations. In this table, the subscripts p and f of the corrections c have been dropped. Instead, the corrections have been subscripted with the direction in which they are applied. c^{SCP} is the correction prescribed by the correction protocol, whereas c is the applied correction (apart from an additional random correction error).

Type	Description	Applied corrections
M1	Fixed offset of Ω mm in X direction	$c_X = c_X^{SCP} + \Omega$ $c_Y = c_Y^{SCP}$ $c_Z = c_Z^{SCP}$
M2	Sign of correction in X and Z directions reversed	$c_X = -c_X^{SCP}$ $c_Y = c_Y^{SCP}$ $c_Z = -c_Z^{SCP}$
M3	Cyclic mutation of corrections in X,Y,Z directions	$c_X = c_Y^{SCP}$ $c_Y = c_Z^{SCP}$ $c_Z = c_X^{SCP}$

7.2.5   Strategy to establish the acceptance threshold T

A Type I error during Test 1 will be referred to as a "partial" Type I error, since H_0 may still be accepted in Test 2. However, if a subsequent Type I error occurs during Test 2 the entire protocol will be restarted unnecessarily. Such a repetition of Type I errors will be referred to as a "complete" Type I error. Because both partial and complete Type I errors lead to increased workload while decreasing setup accuracy, their frequency should be as small as possible. The frequencies of Type I and Type II errors (f_I respectively f_{II}) are defined as the probability that such errors occur in an arbitrary patient and are calculated as the number of Type I (II) errors divided by the number of patients. It is obvious that f_I will decrease with increasing T. The sensitivity to correction mistakes (1-f_{II}) will also decrease with increasing T. COVER should therefore simultaneously reduce frequencies f_I and f_{II} to acceptable levels. We determined an appropriate value of T in four steps:

Type I error frequencies were studied in a population in which no correction mistakes occur ($\varepsilon = 0$) for a range of T. Because, as argued previously and in point 3 below, the incidence of correction mistakes is expected to be $<< 1$, these frequencies and ensuing workload are representative of what is expected in reality. In the Appendix, we derive

closed form expressions for f_I and the resulting workload associated with COVER if $\varepsilon = 0$.

The influence of the threshold T on the frequency of Type II errors was studied in a population with mistake incidence $\varepsilon = 1$ using MC simulations. For each patient with a Type II error, the correction mistake will be present in all subsequent treatment fractions. Using MC simulations, we could determine f_{II} sorted according to the magnitude of the systematic errors they introduce. However, to evaluate the net result of all Type II errors, we evaluated the cumulative distribution of systematic 3D displacements in the patient population [17]. The 3D systematic displacement of patient p is defined by $d_{N_f} = (S_{p,\text{eff},X}^2 + S_{p,\text{eff},Y}^2 + S_{p,\text{eff},Z}^2)^{1/2}$, where the displacement vector components can be obtained from Eq. (7.1). The following criterion for the cumulative distribution of d_{N_f} was applied to select appropriate thresholds T:

If the NAL protocol is applied to the clinical population (Table 7.1) and correction mistakes occur in *each* patient ($\varepsilon = 1$), the fraction of patients with $d_{N_f} > 5$ mm (hereafter $p_{>5}$) after application of the COVER stage should be no larger than the corresponding fraction in case no setup corrections would have been applied. This constraint should be satisfied for each of the three mistake types in Table 7.2.

The value of $p_{>5}$ is chosen because it measures the quality of the systematic error distribution in a population of patients [17, 24]. If the above criterion holds for $\varepsilon = 1$, it will hold for any $\varepsilon < 1$ if systematic errors are indeed reduced by correct application of the SCP. Therefore, a COVER procedure that satisfies this criterion ensures that the application of setup corrections will result in a distribution of systematic setup errors equal to or better than that obtained without setup corrections. Because our criterion is very conservative (based on 100% incidence of correction mistakes), it is not clear *a priori* whether it can be met for a small value of f_I (step 1 above).

For a threshold T that meets the above criterion, the setup accuracy obtained with NAL+COVER for a realistic value of ε was evaluated. To estimate the expected magnitude of ε in a clinical situation, we evaluated correction mistakes detected in over 2000 patients treated with a SAL SCP in our hospital. Two instances were identified, both of type M2. These mistakes were due to wrong input to the off-line SCP program, i.e. operator errors. They were detected because the SAL SCP yielded increasingly large setup corrections after multiple iterations of the protocol (see Discussion). Although our current software design precludes this specific mistake to occur, software or machine configuration errors may always lead to similar mistakes. Based on the above, the value of ε measured in our institute is about 0.001. On the one hand, this is a lower limit because we can not be sure that each mistake was properly identified. On the other hand, because one origin of M2 mistakes has been removed by a different software implementation, the current value of ε may be smaller. As a conservative estimate of the time averaged value of ε in a situation where insufficient attention is given to the appropriate execution of setup corrections, we chose $\varepsilon = 0.05$. The robustness of the workload and mistake detection efficiency of COVER against deviations between the assumed magnitudes of random setup errors (on which the value of T will be based) and the actual magnitudes was determined.
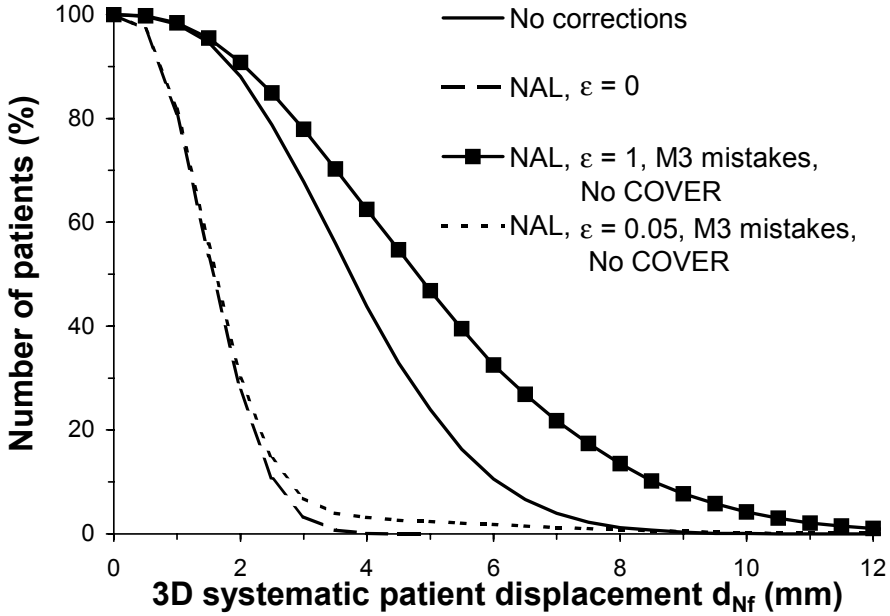
Figure 7.2 Cumulative distributions of the 3D systematic patient displacements, d_{Nf}, after 30 treatment fractions for the prostate patient population ('clinical population') described in the text, without any setup corrections (solid curve) and after application of the NAL protocol with 3 measurements per patient (long dash). The effect of M3 type correction mistakes in all patients and in 5% of the patients in case no COVER stage is applied is illustrated by the solid curve with squares and the short-dashed curve, respectively.

7.3 Results

7.3.1 The increase of systematic errors due to correction mistakes

In Figure 7.2, we display cumulative distributions of the systematic displacements in 3D, d_{Nf} (section 7.2.5) as obtained from MC simulations using the parameters in Table 7.1 (i.e., for the clinical population). If no setup corrections are performed, $p_{>5}$ is 24%, consistent with earlier results obtained in prostate patients [17, 24]. If the NAL protocol with 3 measurements per patient is applied and no correction mistakes occur ($\varepsilon = 0$) this fraction is reduced to 0%. The corresponding value of Σ_{eff} is 1.0 mm in each setup direction. In contrast, if a correction mistake of type M3 occurs in all patients ($\varepsilon = 1$) and no correction verification is applied, the resultant cumulative distribution displayed in Fig. 7.2 is clearly worse than in case of no setup corrections; $p_{>5}$ has doubled to 47% and the distribution tail is extended. The corresponding SD of systematic errors is $\Sigma_{\text{eff}} = 3.3$ mm (in each of the 3 directions). Of course, the persistence of a situation with $\varepsilon = 1$ would

require that a systematic procedural error (e.g. in software configuration) goes unnoticed during SCP commissioning together with a lack of quality assurance on SCP results. Such a situation should not occur if SCPs are conscientiously applied, but this case does emphasize the importance of COVER (see also Discussion).

The distribution obtained for the NAL protocol for a mistake incidence $\varepsilon = 0.05$ is also depicted in Fig. 7.2. Obviously, the overall shape of the cumulative distribution remains the same, but the tail is raised. In fact, this distribution is simply the weighed mean of the curves with $\varepsilon = 0$ (weight 0.95) and $\varepsilon = 1$ (weight 0.05) (section 7.2.4). In this case, $p_{>5} =$ 2.5% and the tail of the distribution extends up to 1 cm.

7.3.2   *Type I errors and EPI workload in NAL+COVER*

The frequency of Type I errors was studied in a simulated population with no correction mistakes ($\varepsilon = 0$). First, we studied iso-random populations ($\delta\sigma_{mp} = 0$, section 7.2.1) for $\sigma_{mp} = 1.5$, 2.0, 2.5 and 3.0 mm while $\sigma_{couch} = 0.5$ mm (note that for iso-random populations, σ_r in Eq. (7.2) equals σ_{mp}). Both MC simulations and evaluations of the closed form expressions given in the Appendix were performed. Frequencies of partial and complete Type I errors (section 7.2.5) are shown in Fig. 7.3a and b, respectively. The required number of measurements per patient (N_{meas}) for the complete NAL+COVER procedure is given in Fig. 7.3c. Obviously, there is good agreement between the MC simulated and analytical results, which confirms proper implementation of the NAL+COVER procedure in our MC software. From the derivation in the Appendix we obtain that the Type I error frequencies in Fig. 7.3 are independent of the initial distribution of systematic errors, a result we confirmed by MC simulations for various values of Σ. Therefore, we conclude that knowledge of the random setup errors in the patient population is sufficient to predict the workload from the NAL+COVER procedure.
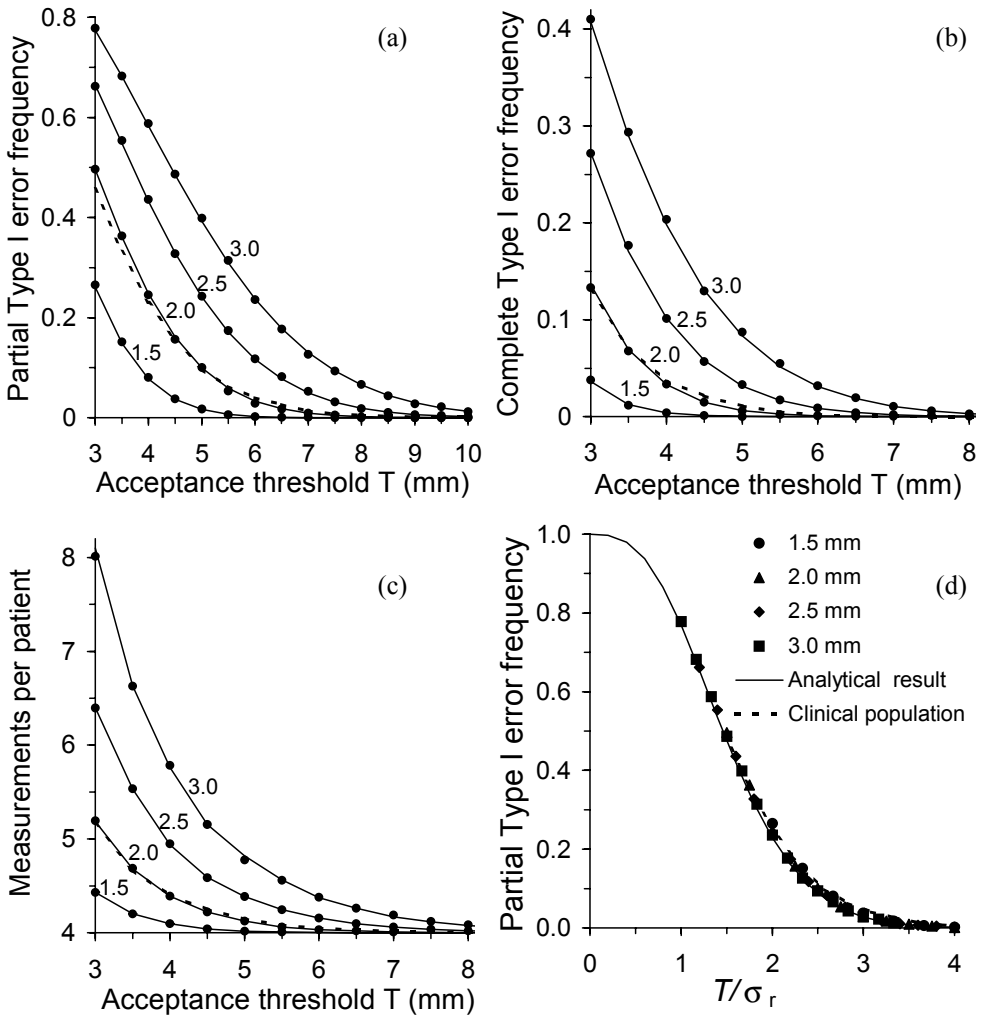
Figure 7.3 (a) Frequency of partial Type I errors (false alert in Test 1) as a function of the acceptance threshold T of the COVER stage following a NAL protocol with $N_m = 3$. Results obtained from MC simulations (closed circles) and analytical expressions (solid curves) for iso-random populations with $\sigma_{mp} = 1.5$, 2.0, 2.5 and 3.0 mm (curves are labeled with σ_{mp} value) and $\sigma_{couch} = 0.5$ mm. The dashed line gives the MC result for the clinical population. (b) As (a), but now for complete Type I errors (false alert in Test 1 and 2). (c) As (a) and (b), depicting the number of EPI measurements per patient (N_{meas}). Because the minimum number of measurements is 4 (3 for NAL and 1 for COVER) irrespective of T, the vertical axis starts at this value. (d) Same as (a), but now the frequencies in (a) are plotted as a function of T/σ_r, which equals T/σ_{mp} for iso-random populations. The MC results for such populations are indicated by closed symbols. The dashed line gives the MC result for the clinical population. The solid line was obtained analytically for $\sigma_{couch} = 0$.

We also performed MC simulations for the clinical population (dashed lines in Fig. 7.3), in which $\sigma_r \approx \sigma_{mp} = 1.9$ mm (Table 7.1). The results slightly deviate from the iso-random population with $\sigma_{mp} = 2.0$ mm, particularly the slope of the curves becomes less steep due to the presence of a distribution of σ_p values. Nevertheless, these differences are small and unimportant in clinical practice. Note that for iso-random populations with $\sigma_{couch} = 0$, the curves in Fig. 7.3 would scale in σ_{mp} (Appendix), i.e. if T / σ_{mp} is chosen as variable for the horizontal axis, all curves collapse into one. Since $\sigma_{couch} = 0.5$ mm makes a relatively small contribution to the SDs that determine the shape of those curves (see e.g. Eq.(7.9)), this is to a high degree also the case for the results in Fig. 7.3. For instance, in Fig. 7.3d the MC results of Fig. 7.3a are redisplayed as a function of T / σ_{mp}, together with the analytical prediction for $\sigma_{couch} = 0$. Obviously, a single curve is sufficient to determine the behavior for any of the studied σ_{mp} values. Again, the results for the clinical population differ little from the iso-random population with $\sigma_{mp} = 2.0$ mm.

Because partial Type I errors alert operators to a possible correction mistake while such a mistake is in fact not present, a high frequency of such errors will unnecessarily lead to a significant increase in workload. Therefore, the frequency of partial Type I errors should be small, say < 5%. Such a small frequency will also stimulate a high level of operator alertness to real correction mistakes. Then, for a practical protocol in the clinical population, T should be approximately 6 mm or larger and in general T should be at least $3 \times \sigma_r$ (Fig. 7.3a and d). For such T, the frequency of restarts of the protocol due to complete Type I errors is ≤ 0.002 (Fig. 7.3b), implicating that in less than 1 in 500 patients the protocol will be unnecessarily restarted. Hence, Test 2 sufficiently filters the partial Type I errors. For $T = 3 \times \sigma_r$, N_{meas} will not exceed 4.05 (Fig. 7.3c) in a population without correction mistakes, only slightly more than the minimum value of 4; the EPI workload from the COVER stage would thus be very small (1.05 additional measurement per patient). In the next section, we will verify if the choice $T = 3 \times \sigma_r$ is suited for the detection of significant correction mistakes.

7.3.3   Detection of correction mistakes

To assess the performance of COVER in detection of real setup correction mistakes, the NAL+COVER procedure was simulated for the clinical population and for all three types of mistake listed in Table 7.2. For a specific mistake type, the corresponding correction mistake was simulated in each patient ($\varepsilon = 1$). Furthermore, if COVER detected a real mistake in a given patient, the simulation continued on the assumption that the cause would be identified and the mistake eradicated for that patient. Mistake M1 was applied with a value $\Omega = 5$ mm, which would constitute a severe mistake, particularly if it would occur in each patient. In Fig. 7.4, we show the resulting values of $p_{>5}$ as a function of acceptance threshold T for each mistake type. Obviously, $p_{>5}$ is quite independent of the mistake types investigated here. Based on Fig. 7.4, criterion (2) formulated in section 7.2.5 for $p_{>5}$, combined with $p_{>5} = 24\%$ without setup corrections (section 7.3.1), yields $T \approx 6.5$ mm as an appropriate acceptance threshold. This value meets the requirements for the maximum

Type I error frequency, $T \geq 3 \times \sigma_r$ (previous section). Due to reasons of robustness pointed out in the next section, we choose $T = 3 \times \sigma_r = 5.8$ mm in the results presented hereafter. Fig. 7.5 displays the simulated cumulative distributions of d_{Nf} obtained for 100% correction mistake incidence for the three mistake types as compared to treatment without setup verification and correction. For NAL without COVER, the M1 mistakes enhance the shoulder of the cumulative distribution while M2 and M3 mistakes strongly enhance the tail (long-dash curves in Fig. 7.5a-5c). As expected, application of NAL+COVER with $T = 3$ $\times \sigma_r$ and 100% mistake eradication after an alert by Test 1 of COVER, forces the cumulative distributions of d_{Nf} to be similar or even better than the distribution obtained without setup corrections (short-dash curve with closed squares in Fig. 7.5a-5c).

The short-dash curves with open squares in Fig. 7.5a-5c are the cumulative distributions of d_{Nf} obtained if only 50% of the correction mistakes that yield an alert in Test 1 are eradicated (the mistake incidence was kept at $\varepsilon = 1$). In this situation, both the shoulder and the tail of the cumulative distribution are still strongly reduced relative to no COVER stage. Even for this extreme case (only 1 out of 2 detected correction mistakes are solved while mistakes occur for each patient), the criterion $p_{>5} < 24\%$ is still met for all three mistake types, demonstrating the robustness of the procedure against variations in the fraction of solved mistakes. This robustness stems from the fact that large mistakes that are not solved after Test 1 will often be confirmed by Test 2 and subsequently eradicated because of the
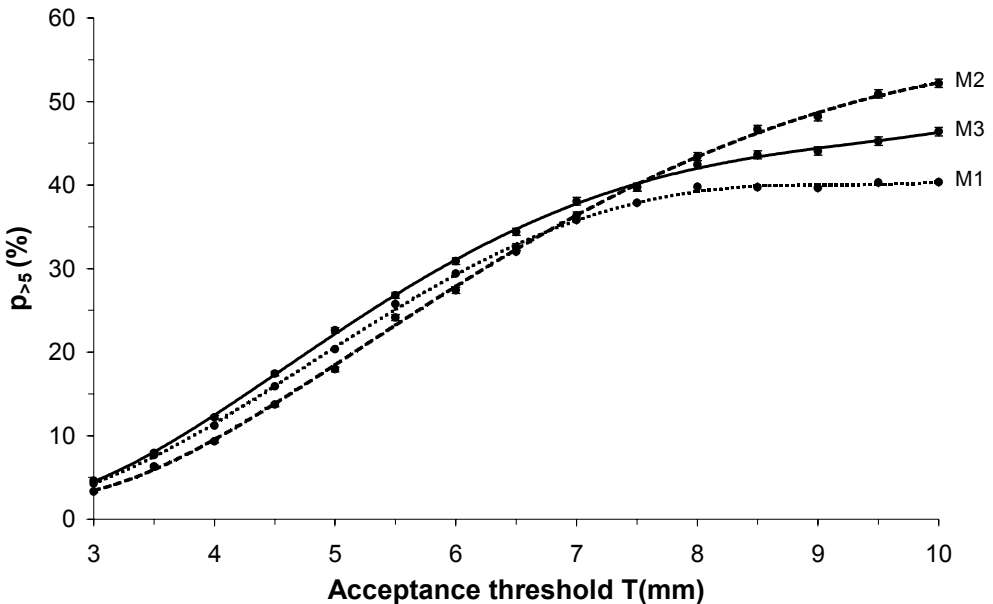

Figure 7.4 Probability that the 3D systematic patient displacement $d_{Nf} > 5$ mm, in case of 100% correction mistake incidence (for each of the three types of correction mistakes in Table 7.2). Obtained from MC simulations for the case that alerts raised by Test 1 always resulted in correction mistake identification and eradication.

complete restart of the protocol (Fig. 7.1).

Fig. 7.5d shows the effect of the COVER stage in case of the more realistic, but still large mistake incidence $\varepsilon = 0.05$ (based on simulations in which each detected mistake is eradicated). As can be seen by comparing the curves for NAL with $\varepsilon = 0$ and NAL with $\varepsilon = 0.05$ in Fig. 7.2, only the tail of the cumulative distribution is significantly influenced by correction mistakes at the latter incidence level. Hence, the cumulative distributions for $\varepsilon = 0.05$ are shown for $d_{Nf} > 4$ mm. For each of the mistake types, COVER reduces $p_{>5}$ from 2-3% to 1%. Moreover, COVER ensures that the fraction of patients with $d_{Nf} > 1$ cm becomes totally negligible. The workload for $\varepsilon = 0.05$ due to COVER is less than 1.08 measurement per patient, independent on the mistake type, which is close to what was obtained for $\varepsilon = 0$ (section 7.3.2).

7.3.4 Robustness to uncertainties in the setup distribution parameters of a patient population

Based on the above, we conclude that for $T = 3 \times \sigma_r$, the requirements of section 7.2.5 are met, at a workload which is very close to the minimum of 1 extra measurement per patient. In this section we investigate the effect on the workload, the number of protocol restarts and the detection efficiency of correction mistakes, if the measured parameters describing the setup distribution of the patient population is not entirely accurate, e.g. when derived from a limited patient data set.

As demonstrated in the Appendix, for $\varepsilon << 1$ the Type I error frequencies and workload of COVER depend solely on the distribution of the random setup errors. For an iso-random population, the accuracy with which the measured value of σ_r from Eq. (7.2) approximates the actual value can be calculated from the chi-square distribution [123]. In our institute, it has become standard to base the setup error distribution parameters on groups of at least 30 patients with at least 5 imaged fractions per patient. For setup studies with 5 fractions imaged in 30 patients, the *measured* value of σ_r will be between $0.87\sigma_r$ and $1.13\sigma_r$ in 95% of the studies. Therefore, if we would perform such a setup study in the clinical population of Table 7.1, and subsequently apply the rule $T = 3 \times \sigma_r$, the derived T value would typically range between 5.1 and 6.6 mm instead of its optimal value of 5.8 mm. To study the workload and number of protocol restarts for this range of T-values we performed MC simulations for the clinical population and $\varepsilon = 0$. The value of T was varied over the above range, and kept independent of setup direction. These simulations were similar to those presented in section 7.3.2, but now the number of patients per MC run was increased to 30000 to precisely determine the restart rate due to complete Type I errors over the T-range of interest. Note that, because in practice the inaccuracy in the measured value of σ_r will differ per direction, the net effect of the inaccuracies in the estimate of T will in general be smaller than presented below; instead of the isotropic variations studied here (i.e., T was under- or over estimated by equal amounts in all 3 directions), the effects of inaccuracies for different directions will partly compensate each other.

Fig. 7.6 shows that, for the range of threshold values of importance, N_{meas} does not exceed 4.14; hence the COVER workload is no more than 1.14 measurement/patient. The fraction of patients with an unnecessary protocol restart is below 1% over this same range,
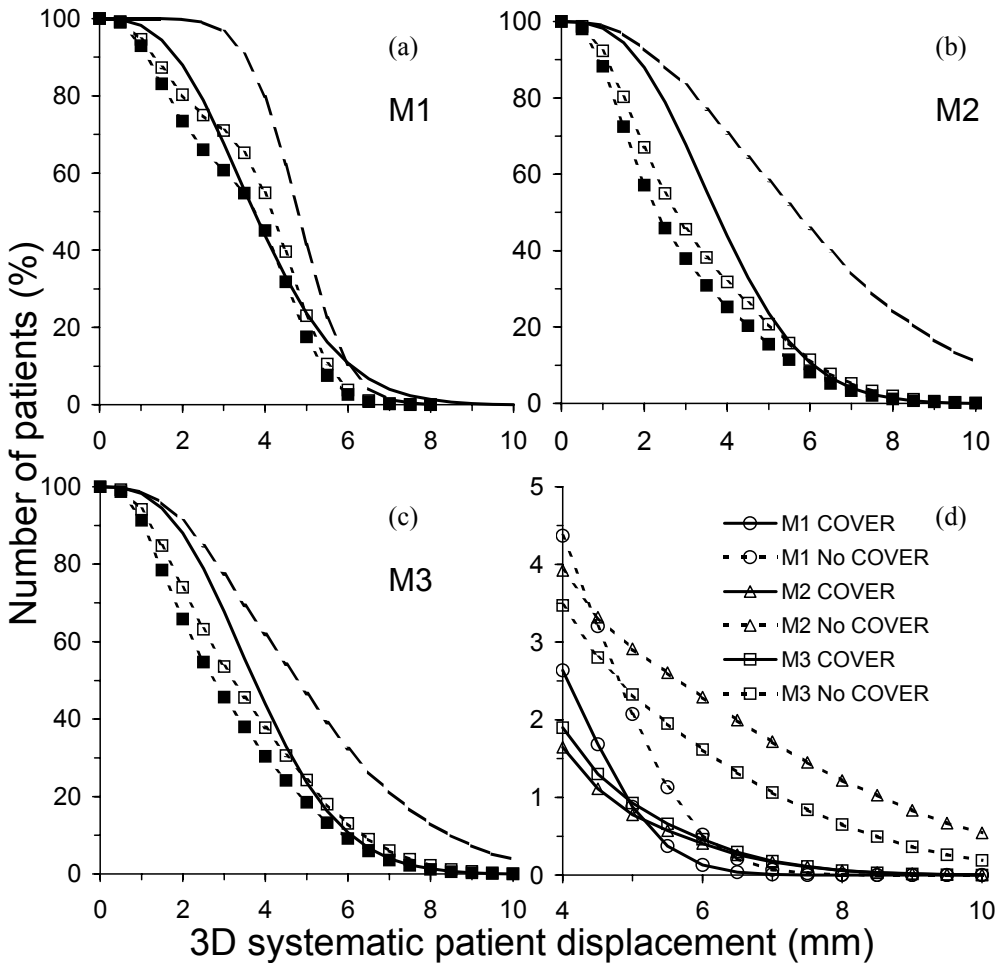
Figure 7.5 Cumulative distributions of the 3D systematic patient displacements d_{Nf}, from MC simulations; (a) The solid thick curve is the distribution obtained without setup corrections. The long dash curve is obtained if setup corrections are established with the NAL protocol, but wrongly executed with Type M1 mistakes (Table 7.2) for *all* patients ($\varepsilon = 1$) while no COVER stage was applied. The short dash curve with closed squares is obtained if COVER with $T = 3\,\sigma_r$ is applied and each correction mistake detected in Test 1 of COVER is eradicated. The short dash curve with open squares is the result of the same COVER procedure, but now for the case that only 50 % of such detected mistakes are immediately identified and eradicated. (b) and (c) are as (a), but obtained for M2 respectively M3 mistake types. (d) Tails of the cumulative distributions for COVER and no COVER stage if the correction mistake incidence is 5% ($\varepsilon = 0.05$) for each of the three mistake types. These distributions were obtained under the assumption that each mistake indicated by COVER in Test 1 is eradicated.

which we consider acceptably small. The robustness of the detection efficiency of correction mistakes against variation in T can be directly derived from Fig. 7.4. For a 100%

incidence of the three mistake types studied, the COVER stage establishes that the criterion $p_{>5} < 24\%$ is satisfied for the range of threshold values of interest, although just so for the largest values ($T = 6.6$ mm).

We conclude that the threshold T of the COVER stage can be determined on the basis of a setup study of the limited size described above. One could further reduce the number of patients if the number of setup measurements per patients is increased, but such an approach is only useful if the inter-patient variation $\delta\sigma_p$ is small.

7.4 Discussion

We have described a correction verification procedure (designated COVER) to detect large mistakes in the execution of setup corrections at a small additional workload per patient. The criteria for the selection of the parameter T in the COVER procedure (formulated in section 7.2.5) were chosen conservatively. In particular, the requirement that the cumulative distribution of systematic displacements should not worsen due to correction mistakes for the very unlikely situation that each patient suffers from such a mistake ($\varepsilon = 1$), constrains the acceptance threshold T. An argument for choosing conservative criteria is that we have incomplete knowledge of the mistakes that may occur, and only a limited sample of possibilities can be studied. Nevertheless, we could reconcile these criteria with a workload due to the COVER stage that is very close to the minimal workload of 1 additional measurement per patient.

The final result of the combined NAL+COVER procedure is partly determined by the type of correction mistakes that occur, but most importantly by their incidence. From Fig. 7.5d, we see that the benefit of COVER for the relatively large incidence of $\varepsilon = 0.05$ lies mainly in reducing the tail of the distribution of systematic displacements. Although the height of this tail is already small without COVER (few %), COVER precludes large systematic displacements (> 1 cm). This result can be quite relevant to individual patients if treatment planning margins are based on the expected small systematic displacements after setup correction (see the tightness of the $\varepsilon = 0$ distribution in Fig. 7.2). An equally important property of the COVER stage is that it will alert operators frequently in case of an error in the SCP procedure itself. For instance, a software configuration error may generate mistakes in the generated setup correction in each patient. As a result, the mistake incidence would make a jump transition from $\varepsilon \approx 0$ to $\varepsilon = 1$, and the workload and alert frequency of Test 1 in COVER would strongly increase. If, for example, a software error would lead to M2 type mistakes (Table 7.2), then for the clinical population, the frequency of operator alerts would increase from 4% to 46%. Such an increase would undoubtedly be noticed and hence an $\varepsilon = 1$ situation will be detected rapidly.

The acceptance threshold of the COVER protocol is based on *a priori* knowledge of the setup error distributions of the patient population. Alternatively, T could be made patient specific if the random setup errors measured during the N_m pre-correction fractions are taken into account. However, for the 2 to 4 pre-correction measurements available in a typical NAL protocol [24], the accuracy of the measured value of σ_p is small. In fact, for $N_m = 3$, the measured value would be related to the true value through the chi-square distribution with 2 degrees of freedom [123]. As a result, the 95% confidence interval for the actual value of σ_p would range from 0.5 to 6.3 times the measured value in any given patient. Hence it is not possible to develop an appropriate test (small Type I and Type II

error frequency, low workload) without use of *a priori* knowledge of the random setup errors, a result we confirmed by replacing the global geometrical test of Eq. (7.10) by a Student's *t* test [123] in our MC simulations. We found that, with the latter test, sufficient mistake detection efficiency could only be obtained at large Type I error frequencies [25].

The NAL+COVER protocol aims at providing a low workload solution to off-line setup corrections. For our prostate population, the NAL+COVER procedure would require < 4.1 measurements/patient to achieve the final systematic displacement distribution of the NAL $\varepsilon = 0$ curve in Fig. 7.2. Because the NAL SCP was proposed as a low EPI analysis workload alternative to the SAL protocol [24], we should compare the workload of NAL+COVER to that of the SAL protocol.

First, the workload of the COVER stage is not only determined by the number of required measurements, but also by the time required to falsify partial Type I errors (section 7.2). We take the following actions to determine the nature of a possible error detected by COVER. In case of a setup correction, our EPID software automatically prints a form containing all relevant data so copying by hand cannot be the source of error. However, the corrections are performed manually by a technologist. Although a second technologist always inspects this correction, we first consider if the detected error is likely due to a manual execution error, e.g. in reading the numbers from the form. If not, we check if the data on the correction form used are identical to those presented by the EPID software. When the form is correct, the patient data input, i.e. the quality of the reference setup
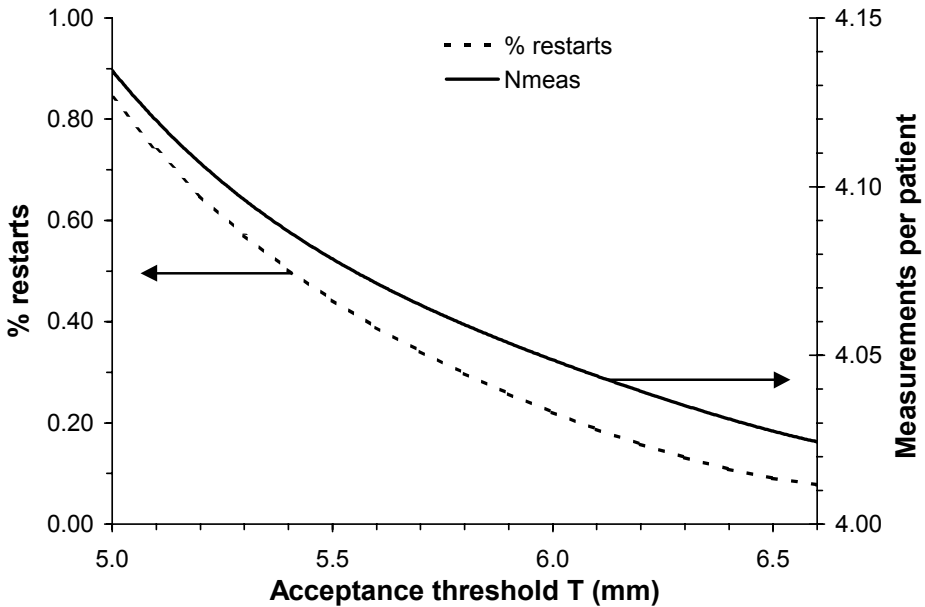

Figure 7.6 EPI workload (right vertical axis) and % protocol restarts (left vertical axis) in the clinical population for a range of acceptance thresholds *T*. This range reflects the expected statistical uncertainty in σ_r (hence in $T = 3\sigma_r$) when obtained from a limited setup study.

images, the treatment plan data, the contours used in matching and the selected SCP, are verified. If no peculiarities are found, the search ends and the second verification measurement will be performed in the next fraction. The workload introduced by the above actions is small compared to the entire EPI workload, not only because each action requires little operator time but more importantly because they are necessary in less than 5% of the patients and generate further actions in an even much smaller group (section 7.3.2).

Second, we compared the EPI workload of the NAL and SAL SCPs. We used methods described elsewhere [24] to determine SAL protocol parameters which yield a cumulative distribution comparable to that achieved with the NAL+COVER procedure (i.e. $\Sigma_{eff} = 1$ mm). The optimal parameters are those that achieve such a distribution with a minimum value of N_{meas}. The optimal SAL parameters thus found were (α, N_{max}) = (7.5 mm, 6), requiring N_{meas} = 9.5. Other SAL protocol parameters that achieved the specified cumulative distribution with less frequent multiple corrections (hence clinically more acceptable [24]), required $N_{meas} \geq 11$. We also determined the minimum value of Σ_{eff} that can be achieved with a SAL protocol of the same EPI workload as the described NAL+COVER protocol, i.e. N_{meas} = 4. We found that for this workload, the SAL SCP could not reduce Σ_{eff} below 1.6 mm. In summary, the NAL+COVER protocol remains markedly more efficient than the SAL protocol, with the additional benefit that the reduction of systematic errors is realized early in treatment with a single setup correction.

Third, we must evaluate how the SAL SCP treats correction mistakes because this is the reason behind COVER. If a setup correction is generated by the SAL protocol, the SCP restarts, and additional setup corrections may follow [13]. In practice, such multiple setup corrections are added together to provide the net correction [17]. For the clinical population described in this paper, patients with a setup correction had multiple corrections in over 20% of the cases. This rate of multiple corrections is expected for the SAL protocol [17] and therefore operators are not prompted to search for a possible correction mistake. Multiple setup corrections in themselves increase the risk of correction mistakes, because the numerical values of the setup correction change over time. More importantly, a large correction mistake will also lead to multiple setup corrections indistinguishable from 'standard' setup corrections. As a result, whether the SAL protocol corrects such mistakes is dependent on the origin of the mistake. To illustrate this, we distinguish between software and operator mistakes.

Operator mistakes may be dealt with effectively by the SAL SCP in some cases. For instance, if a prescribed setup correction is copied wrongly onto a form, which is subsequently used to execute the correction manually, the SAL SCP may detect this mistake through a large residue setup error and will generate a new correction. Because the SCP can not distinguish between actual setup errors and correction mistakes, the net correction will be calculated as if the previous correction was executed correctly. It can be shown that two additional setup corrections are required to eliminate the effect of the mistake, if the two subsequent corrections are both executed correctly. A disadvantage of this mechanism is that the origin of the mistake is never identified, and alertness to prevent mistakes is not raised.

Software mistakes, on the other hand, may not be corrected but instead amplified by the SAL SCP. For example, we mentioned previously the clinical occurrence of an M2 type mistake. This was the result of wrong input of the patient orientation to the software at the start of treatment. As a result, all corrections generated by the SAL SCP contained the M2 mistake for that patient. Although the SAL SCP detected the correction mistakes by

generating multiple corrections, each of these corrections suffered from the M2 mistake. The magnitude of the net setup correction grew to such a large value, that operators investigated the cause and solved the problem. However, nothing in the output of the SAL SCP indicated correction mistakes. The moment of mistake detection therefore depended on human observation. This is particularly worrisome as increasing automation tends to make software mistakes more important than human mistakes. Hence, it may be a better approach in general to extend SCPs by a COVER stage instead of using the SCP itself as an iterative verification tool.

Although the final setup accuracy obtainable with NAL+COVER is high, actual results obtained with any SCP are only as good as the measurement accuracy. For instance, if the anatomical structures selected in the portal or reference image are not mutually consistent or a systematic patient rotation is present in the reference image, the applied setup correction can be inaccurate while post-correction measurements, based on the same structures, still indicate a seemingly satisfactory setup. This problem of *systematic* measurement inaccuracy occurs in all SCPs and solutions must be found in careful image analysis procedures. For instance, for all prostate patient setups, we analyze opposed lateral fields to reduce as well as visualize the effect of e.g. rotations or patient movement in between beams [17]. Systematic measurement inaccuracies may also be detected if two observers perform the anatomy match independently. As indicated in section 7.2.3, COVER helps to detect large *random* measurement inaccuracies. If, for a specific patient, setup error measurements are significantly less accurate than in most patients, this will increase both the apparent random setup errors as well as the actual post-correction systematic setup error. Consequently, it is probable that the apparent setup error in the verification measurement exceeds the threshold *T*. The actions listed above to be performed after operator alerts may then help identify the source of the problem.

We conclude that the NAL+COVER protocol is a safe and highly efficient means to reduce systematic setup errors. At present, we clinically evaluate the NAL+COVER protocol for prostate patients with the parameters described in this paper.

Acknowledgements

Appendix

A. Derivation of the population distribution of random errors

As described in section 7.2.1, the SDs of the random errors σ_p were assumed to be normally distributed among patients with distribution function $N(\sigma_{mp}, \delta\sigma_p)$. The squares in Fig. 7.7 show the distribution of σ_p values measured in 600 prostate patients [24]. In fact, because the σ_p-distributions were similar for each setup direction, they were added so that the displayed distribution is based on 3×600 values. Because a finite number of setup measurements per patient is available ($N_{\text{meas}} = 10$ [24]), each measured value of σ_p is related to its actual value by the chi-square probability distribution [123]. Hence, the measured distribution of σ_p values is not normally distributed if the actual distribution is, but rather a convolution of the normal distribution and the chi-square distribution. We used MC simulations to derive the actual distribution from the measurements as follows.

3D setup errors were simulated in 10 fractions per patient for 10000 patients, where the actual value of σ_p was drawn from $N(\sigma_{mp}, \delta\sigma_p)$ for each patient. This process was repeated for a range of parameter values $(\sigma_{mp}, \delta\sigma_p)$. Per pair of parameter values, σ_p values were calculated from the simulated setups and compared to the measured set using the Kolmogorov-Smirnov (KS) statistic [104]. The KS statistic expresses the significance of a difference between two distribution functions of two sets of measurements (in our case, one set is measured σ_p, the other set is simulated σ_p). Optimal parameters $(\sigma_{mp}, \delta\sigma_p)$ were sought by minimizing the KS statistic. By considering the significance of this statistic for the optimal parameters, the assumption of a normal distribution function for the actual σ_p can be tested. Note that, for unconstrained parameters $(\sigma_{mp}, \delta\sigma_p)$, the probability distribution $N(\sigma_{mp}, \delta\sigma_p)$ may assign finite probabilities to values of $\sigma_p < 0$. This non-physical behavior was taken into account by truncating the distribution below $\sigma_p = 0$ in the MC simulations. However, the KS statistic reached a minimum in a region of $(\sigma_{mp}, \delta\sigma_p)$ values where the probability of $\sigma_p < 0$ was completely negligible (see below).

In Fig. 7.7 we show the frequency distribution of measured σ_p values in 0.25 mm wide bins. The 2 SD wide error bars were derived from the available number of measurements per bin [123]. The optimal parameters were $(\sigma_{mp}, \delta\sigma_p) = (1.9 \text{ mm}, 0.4 \text{ mm})$ and these values are given in Table 7.1. The corresponding simulated frequency distribution is depicted by the dashed line in Fig. 7.7. The visual correspondence of the two distributions is confirmed by a large significance level of the KS statistic: $p = 0.63$. Therefore, a normal distribution of actual σ_p values is an acceptable assumption; the solid line in Fig. 7.7 shows the actual normal distribution for the optimal parameters given above.
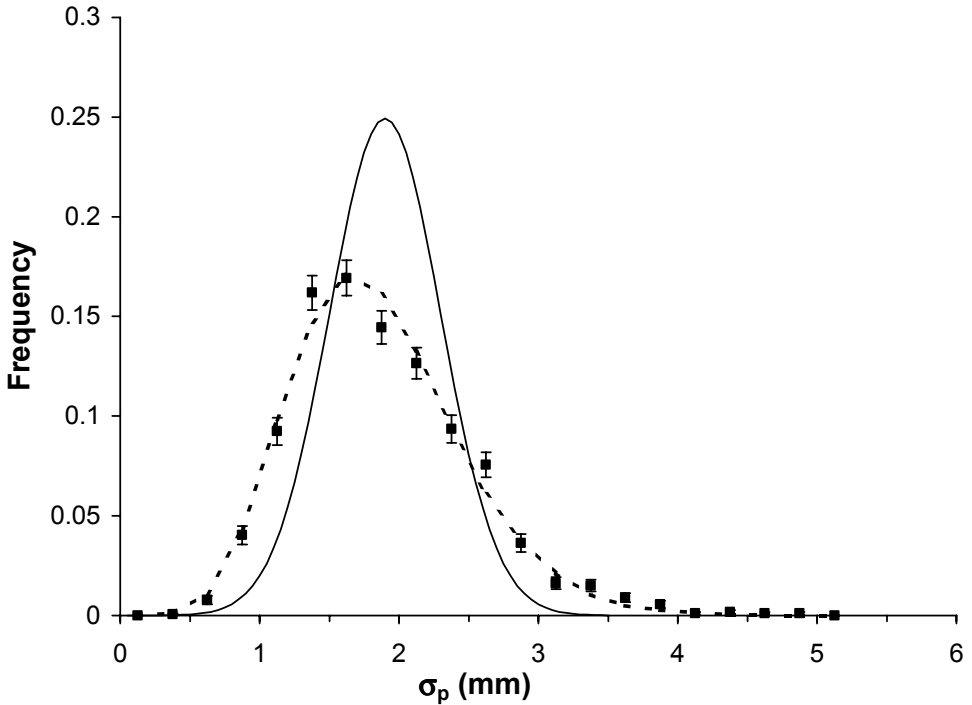
Figure 7.7 Distribution of random setup errors σ_p in 0.25 mm bins. Squares with error bars (2 SD wide) are measured data from the clinical population (600 patients, 10 measurements per patient), the dashed line is the corresponding MC result for distribution parameters $(\sigma_{mp}, \delta\sigma_p) = $ (1.9 mm, 0.4 mm). The solid line is the corresponding actual (normal) distribution of σ_p.

B. Expressions for the Type I error frequency and N_{meas} implied by COVER

In this section, we will use the symbols introduced in section 7.2.3. For readability, we will drop the patient subscript p, as the analysis below will hold for each patient. We will assume that the random setup errors have a normal probability density function equal for each patient (iso-random) and that no correction mistakes occur ($\varepsilon = 0$). Then, the first post-correction verification measurement $x_{v1}^{cor} = \Delta_{v1} + \delta + \tau_{v1}$ (Eq.(7.8)) will be normally distributed with SD $= \sigma_{v1}$ as given by Eq. (7.9). For a single setup direction, the frequency of partial Type I errors, $f_{pI,1D}$ will equal the probability $P(|x_{v1}^{cor}| > T)$, which can be evaluated in terms of the standard error function ("erf"):

$$f_{pI,1D} = P(|x_{v1}^{cor}| > T) = 1 - \mathrm{erf}\left(\frac{T}{\sqrt{2}\sigma_{v1}}\right). \tag{A7.1}$$

For k setup directions, and isotropic distributions of setup errors, the partial Type I error frequency $f_{pI,kD}$ becomes

$$f_{pI,kD} = 1 - (1 - f_{pI,1D})^k = 1 - \text{erf}^k\left(\frac{T}{\sqrt{2}\sigma_{v1}}\right), \tag{A7.2}$$

which was used to obtain the solid curves in Fig. 7.3a for $k = 3$.

The frequency of complete Type I errors is slightly harder to obtain. For such an error to occur, both verification measurements must exceed the acceptance threshold: $|x_{v1}^{cor}| > T$ and $|x_{v2}^{cor}| > T$. However, the probability $P(|x_{v2}^{cor}| > T)$ is not independent of $P(|x_{v1}^{cor}| > T)$, because $x_{v2}^{cor} = \Delta_{v2} + \delta + \tau_{v2}$ contains the same δ-term as x_{v1}^{cor} (in fact, the residue systematic error after correction). We write $x_{v1}^{cor} = R_{v1} + \delta$ and $x_{v2}^{cor} = R_{v2} + \delta$, where the stochastic quantity R_{v1} is independent of R_{v2}, both being normally distributed like $N(0, \sigma_R)$ with $\sigma_R = \sqrt{\sigma_{mp}^2 + \sigma_{couch}^2}$ (section 7.2.3). Because the random setup errors are normally distributed, so is δ with distribution $N(0, \Sigma_{res})$ where the SD Σ_{res} is given by Eq. (7.6) for the NAL protocol. We can now calculate the probability of a complete Type I error in one dimension:

$$f_{cI,1D} = P(|R_{v1} + \delta| > T \wedge |R_{v2} + \delta| > T) = \int_{-\infty}^{\infty} P'(\delta)\left[\int_{|R_{v1}+\delta|>T} P'(R_{v1})\,dR_{v1} \int_{|R_{v2}+\delta|>T} P'(R_{v2})\,dR_{v2}\right]d\delta \tag{A7.3}$$

where $P'(x)$ is the probability density function of x. Substitution of the normal distributions described above for $P'(x)$ yields

$$f_{cI,1D} = \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}\left(\frac{\delta}{\Sigma_{res}}\right)^2}}{\sqrt{2\pi}\Sigma_{res}}\left(1 - \frac{1}{2}\left\{\text{erf}\left(\frac{T-\delta}{\sqrt{2}\sigma_R}\right) + \text{erf}\left(\frac{T+\delta}{\sqrt{2}\sigma_R}\right)\right\}\right)^2 d\delta. \tag{A7.4}$$

The frequency of complete Type I errors in more than one dimension can be obtained from Eq. (A7.2) by substituting f_{cI} for f_{pI}. The expression in Eq. (A7.4) was evaluated using a modified Romberg method [104] to yield the solid curves in Fig. 7.3b (again for k = 3).

With the above expressions, we can also evaluate the number of measurements required by the NAL+COVER procedure. The minimum number of measurements required to complete the procedure equals $N_m + 1 \equiv N_{min}$. For a fraction of patients equal to $f_{pI,kD}$, at least one additional measurement must be performed and for a fraction $f_{cI,kD}$ the protocol must be restarted. For the latter group, at least another N_{min} measurements must be performed, etc. As a result, the average number of measurements required per patient becomes

$$N_{\text{meas}} = N_{\min} + f_{p\text{I,kD}} + f_{c\text{I,kD}}(N_{\min} + f_{p\text{I,kD}} + f_{c\text{I,kD}}(N_{\min} + f_{p\text{I,kD}} + f_{c\text{I,kD}}(.....))) =$$

$$(N_{\min} + f_{p\text{I,kD}})(1 + f_{c\text{I,kD}} + f_{c\text{I,kD}}^2 + f_{c\text{I,kD}}^3 + ...) = (N_{\min} + f_{p\text{I,kD}})/(1 - f_{c\text{I,kD}})$$

(A7.5).

By putting k = 3, the solid curves in Fig. 7.3c were obtained.

8 Electronic portal image assisted reduction of systematic set-up errors in head and neck irradiation

J.C.J. de Boer, J.R. van Sörnsen de Koste, C.L. Creutzberg, A.G. Visser, P.C. Levendag and B.J.M. Heijmen

Abstract

Purpose: To quantify systematic and random patient set-up errors in head and neck irradiation and to investigate the impact of an off-line correction protocol on the systematic errors.

Methods and Materials: Electronic portal images were obtained for 31 patients treated for primary supraglottic larynx carcinoma who were immobilised using a PVC cast. The observed patient set-up errors were input to the shrinking action level (SAL) off-line decision protocol and appropriate set-up corrections were applied. To assess the impact of the protocol, the positioning accuracy without application of set-up corrections was reconstructed.

Results: The set-up errors obtained without set-up corrections (1 standard deviation (SD) = 1.5-2 mm for random and systematic errors) were comparable to those reported in other studies on similar fixation devices. On average, 6 fractions per patient were imaged and the set-up of half the patients was changed due to the decision protocol. Most changes were detected during weekly check measurements, not during the first days of treatment. The application of the SAL protocol reduced the width of the distribution of systematic errors to 1 mm (1 SD), as expected from simulations. A retrospective analysis showed that this accuracy should be attainable with only 2 measurements per patient using a different off-line correction protocol, which does not apply action levels.

Conclusions: Off-line verification protocols can be particularly effective in head and neck patients due to the smallness of the random set-up errors. The excellent set-up reproducibility that can be achieved with such protocols enables accurate dose delivery in conformal treatments.

8.1 Introduction

Radiotherapy treatment delivery in head and neck tumors requires a highly accurate and reproducible treatment set-up. The steep relation between tumor dose and tumor control probability (TCP) frequently observed in this region [86, 92, 119], together with the large value of the 50% tumor control dose [92] implies that the TCP will be very sensitive to 'cold spots' in the dose distribution and necessitates a good coverage of the entire target volume with a prescribed minimum dose [32]. Therefore, the local recurrence rate may be quite sensitive to set-up errors, as was pointed out early by Goitein and Busse for supraglottic lesions [47]. Similarly, due to the close proximity of sensitive structures (e.g. spinal cord, salivary glands) such set-up errors may lead to severe complications [86]. Specifically *systematic* set-up errors (repetitive errors that occur in each treatment fraction) may lead to an inhomogeneous dose distribution in the target volume and above tolerance doses in the critical normal tissues [55, 125].

Numerous investigations on the patient set-up accuracy during head and neck irradiation have been reported, both using films (e.g. [19, 59, 140]) and electronic portal imaging devices (EPIDs) [15, 145, 146]. The latter devices allow for easy repeated imaging and image processing, rapid quantitative assessment of treatment set-up errors, and automated verification of the field shape [116]. Hence, apart from their application in studies on the set-up accuracy for a certain immobilisation technique, EPIDs facilitate accurate set-up verification and, if necessary, the application of set-up corrections.

Although EPID guided set-up corrections based on a quantitative protocol have been applied successfully for some treatment sites [17, 28, 91], to date the application of such a protocol in the head and neck region has been either speculative [15, 39] or realised only in very small patient groups of two to six patients [102, 135, 146]. In an off-line decision protocol, the set-up correction for a given treatment fraction is entirely based on set-up measurements in previous fractions. Hence, these protocols can only reduce the systematic set-up errors [13], not the random day-to-day variations in set-up. As will be discussed later, the above mentioned small patient number studies are insufficient to determine the distribution of systematic set-up errors with an appreciable accuracy. Therefore, the effect of an off-line protocol should be studied in a larger patient group. In this paper, we describe the impact of a shrinking action level (SAL) protocol [13] on the set-up accuracy of 31 patients treated for supraglottic larynx carcinoma. These patients were positioned using a commercially available head rest and mask system. Data are presented on both the set-up accuracy without application of the SAL protocol (i.e., the accuracy achieved with the fixation device alone) and with set-up corrections (i.e., the improvement achieved with the SAL protocol). Based on the obtained results, we also evaluated a new correction protocol aimed at the reduction of systematic errors with less portal imaging workload.

8.2 Methods and Materials

8.2.1 Patient group and set-up method

The study group consisted of 31 patients with a (T1-4) supra-glottic larynx tumor without detectable positive lymphnodes in the neck (N0). In the first 23 fractions (nearly opposed beams at 70° and 290°), both the elective neck and primary tumor target volumes were irradiated to a dose of 46 Gy. Next, the primary target volume was boosted in 12 fractions (with beams at 90° and 270°) to a dose of 70 Gy. Details on the CTV definition and planning optimisation techniques are found in [37]. The patients were positioned in supine position using a commercially available polyurethane head support (Sinmed BV, The Netherlands; type CUS/14) with accompanying small base plate (Posifix^{TM}). A transparent PVC mask fabricated from a plaster of Paris cast was attached to the base using 2-point fixation at both sides of the head (Fig. 8.1). Next, a planning CT scan was performed using 5-mm slice thickness and a 3D-dose distribution was calculated using the CadPlan (Varian-Dosetek, Finland) treatment planning system. Subsequently, set-up reference lines and the isocentre as defined by the treatment plan was marked on the patient mask on a simulator, and films of all portals were obtained for documentation. The films obtained at 270° were used as reference images for set-up verification in most patients; the last few patients who entered our study were scanned on a Picker CT-simulator and thereby high quality DRRs were available as reference images (based on 3 mm thick CT slices).
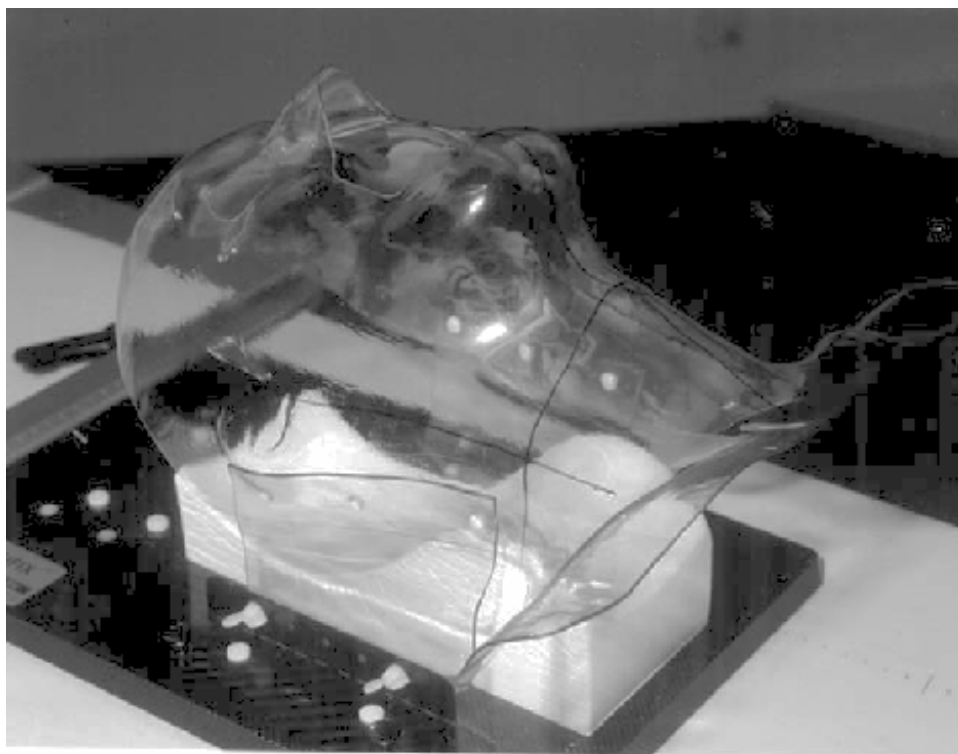

Figure 8.1 Illustration of the PVC mask and base plate used in this study.

At the treatment unit, patients were positioned using the head support and PVC mask, after which the lateral and midline laser beams were aligned on the definitive isocentre demarcation on the mask.

8.2.2 Image acquisition and analysis

Since treatment beams were at slightly oblique angles during the first 23 fractions, and treatment fields were usually too small to discern anatomical structures useful in portal image analysis, an enlarged specific portal imaging field (PIF) was added [28] at a gantry angle of 270°. Patient set-up verification and correction was based entirely on the PIF. Given the orientation of the treatment beams, this was considered sufficient for this patient group.

The exposure per PIF was 6 accelerator monitor units (MU), which, at a clinical doserate (200 MU/min), was the minimum exposure feasible at the accelerator at which this study was performed. No more than 8 fractions were imaged per patient (see section 8.2.3), so the cumulative PIF exposure was always less than 0.5 Gy. The contribution of the maximum PIF dose to the treatment plan was reviewed by the clinicians to evaluate the impact on the critical structures (e.g. parotid- and submandibular glands, spinal cord).

The images were obtained with the new Theraview-NT EPID (Cablon Medical, the Netherlands) [26]. Although mechanically this EPID resembles the previous generation of Theraview EPIDs, there are notable changes. For instance, the tube camera was replaced by a CCD camera in order to facilitate accurate portal dosimetry. In addition, completely new software was developed which features the integration of set-up correction protocols in the image analysis process.

After image acquisition, the field edge of the image is extracted and an automatic template match (which allows for variation of scaling in both directions, rotation and translation) with a prescribed field definition from the planning system is performed [90]. Once this match is completed, the co-ordinate system of the portal image has been established. At the same time, this match constitutes an automated independent verification of the delivered treatment field shape. From an analysis of long term measurements on the reproducibility of these field matches, we have found that the position of the projected isocentre within our the EPID images is known with an accuracy of 0.5 mm (1 SD) in any direction [90] and so we can measure patient set-up errors with respect to the isocentre with a high precision.

The EPID software allows for delineation of anatomical structures in the reference images, which can be projected onto the corresponding portal images and subsequently shifted and rotated until a satisfactory match is achieved [28]. The anatomical structures we used in this template match are depicted in Fig. 8.2. The white contours were considered the most appropriate structures for matching since they are expected to exhibit negligible internal movement with respect to the target volume. These represent the spinous process of C2, the lower back of the skull, and the posterior aspect of the vertebral bodies (drawn as a single continuous line). Sometimes, the outlines of the separate vertebral bodies and the lower jaw were used to aid in the matching process. The latter structure was used only to detect possible jaw movement, and was not allowed to influence match results.

Once a template match was accepted by the user, it was automatically input to the selected off-line correction protocol (i.e., the SAL protocol described in the next section) and the necessary set-up corrections were presented by the EPID software together with a review of all relevant data.

The reproducibility of the match results was determined from a repetitive analysis of the same set of portal images. For 5 patients, 2 fractions were selected and the corresponding 10 images were matched 5 times, in 5 subsequent runs of 10 images by one of the authors (JRvSdK). Because the images of a single patient were always matched by a single observer, in most cases the aforementioned author, we consider the thus derived intra-observer variation as representative of the measurement uncertainty in our study. The images were selected to represent the various image qualities and visibility of anatomical structures encountered in clinical practice. The interval between match runs was at least two hours and the runs were distributed over 2 days.

8.2.3 *Set-up correction protocol and corrections*

 To improve the set-up accuracy, in particular to reduce the systematic errors, we applied the off-line SAL protocol [13]. From measurement $N = 1$ onwards, the set-up error is averaged over the last N measurements for each measured direction, yielding a mean error vector V_N of length d_N. In our study, this vector has two components, the translation along the cranio-caudal direction (Y) and the ventro-dorsal direction (Z). The vector-length d_N is tested against the action level for this measurement, $\alpha_N = \alpha / \sqrt{N}$ where α is the initial action level. If the vector length exceeds the action level, a set-up correction is performed in all subsequent fractions and the procedure is repeated with N reset to 1. This process continues until $N = N_{max}$ subsequent measurements have been obtained without the
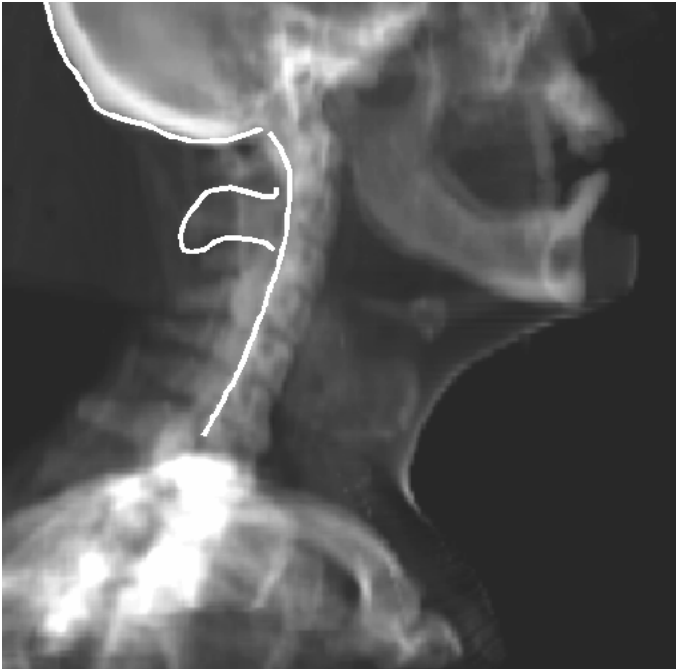

Figure 8.2 The anatomical structures used in our portal image analysis, indicated by white contours.

need for a correction, after which the 'first stage' of the protocol is finished. The first stage of the SAL protocol intends to correct initial large systematic set-up errors, and we aimed at finishing this stage no later than the second week of treatment. Applicable values of the two SAL protocol parameters α and N_{max} were determined from Monte Carlo simulations (see section 8.2.5) based on typical patient set-up accuracy values found in literature [15, 59, 145]. The selected parameters (α, N_{max}) = (5 mm, 2) ensured that the maximum remaining systematic patient set-up error is approximately 3.5 mm (i.e., $\alpha_{final} = \alpha / \sqrt{N_{max}} =$ 5 mm/$\sqrt{2}$) while at the same time the PIF exposure remains below 24 MU (few patients would require multiple corrections with these parameters, and so the number of imaged fractions rarely exceeds 4 in the first stage).

The second stage of the protocol is concerned with follow up measurements to assure that proper set-up accuracy is maintained throughout the treatment. In this study, second stage images were obtained once every two weeks during the first 46 Gy (irradiation includes the elective neck) and once a week during the last 24 Gy (irradiation of the primary target volume only). The set-up errors derived from these images were included in a running average of the set-up error vector over the last N_{max} measurements. The length of the latter vector was compared with the fixed action level α_{final}. If a correction was necessary during the second phase, the protocol was completely restarted. This procedure is identical to the procedure we applied previously to other treatment sites [17, 28].

Set-up corrections were performed by first setting up the patient using the original isocentre demarcation, after which the treatment couch was shifted according to the set-up correction vector, i.e. $-V_N$. We verified that these corrections could be performed accurately (1 SD = 0.3 mm, [28]) and with negligible added treatment time. Note that due to this procedure, we can assume that the execution of set-up corrections will have no influence on the rotation component of the set-up errors, i.e. rotations of patient anatomy relative to the intended set-up do not depend on whether a correction is executed or not.

8.2.4   Set-up accuracy analysis

We adopt the definitions introduced by Bijhold *et al.* [22] for systematic and random errors. The systematic set-up error for a specific patient is the mean set-up error of that patient. Formally, if the set-up error in a certain direction during fraction f is x_f, and the total number of treatment fractions is N_f, then the patient's systematic error is $S = N_f^{-1} \sum_{f=1}^{N_f} x_f$. Each patient p will have a systematic error, S_p. The mean of S_p over all patients in a given population is denoted by μ whereas its standard deviation (SD) is given by Σ .

In addition to systematic errors, the patient set-up will be different for each treatment fraction, due to random set-up errors. The random errors can be characterised by the SD of the variations in x_f. This SD, which we refer to by σ_p for patient p, may be different for each patient. To characterise the random errors in a population an appropriate average is calculated over the patient group, which is denoted by σ [28].

Our measured patient set-up data include the effect of a correction protocol. However, since we know the magnitude of the applied corrections and we can perform these corrections with high accuracy (see previous section) we can obtain the patient set-up data

that would have been measured without a set-up protocol. This method has also been applied in set-up studies on prostate [17] and lung cancer patients [28]. The data thus derived will be referred to as 'initial' and labelled with 'init'. Quantities directly based on the measured set-up data (which were obtained with the SAL protocol) will be labelled with SAL, e.g. Σ_{SAL}. The SDs of the random error will be calculated exclusively from the initial data, so that $\sigma_{\text{SAL}} \equiv \sigma_{\text{init}} \equiv \sigma$. The SD calculated from the initial data represents only the true random set-up variations (and of course the accuracy with which they can be measured; details are in [24, 28]). The SD of the random errors obtained from data with set-up corrections, however, represents the true random set-up variations as well as the spread introduced by the set-up corrections.

Note that with the off-line correction protocol we applied, only a subset of the N_f fractions was imaged and analysed. This has some impact on the calculated value of Σ obtained from the measurements, particularly if the number of fractions imaged per patient (N_{meas}) is small. If no set-up corrections are applied, S_p is best estimated by $N_{\text{meas}}^{-1} \sum_{i=1}^{N_{\text{meas}}} x_i$ [28], where the summation is over the subset of fractions which have been imaged and analysed. The value of Σ derived from these estimates is always slightly overestimated due to the random set-up errors, and the amount of overestimation can be easily calculated [28]. We corrected for this bias in our calculation of Σ_{init}. In case of set-up corrections, the measured value of Σ can also be biased with respect to its true value, and this bias depends on the set-up correction specifics, the number of imaged fractions, and the relative weights assigned to the available measurements. For simple limiting cases, such a bias can be expressed analytically, but in case of the SAL protocol we cannot derive such closed form expressions [13]. We therefore estimated the bias in Σ_{SAL} from Monte Carlo simulations as described in the next section.

8.2.5 *Monte Carlo simulations*

We developed a 3D Monte Carlo simulation program [24] analogous to Bel et al. [13] to study the impact of correction protocols on patient set-up error distributions. The program simulates the systematic and random set-up errors occurring in a population of patients for a specific number of fractions per patient by drawing numbers from appropriate statistical distributions (in the current implementation, normal distributions) describing such set-up errors. Furthermore, per patient, the decisions that would be generated by an off-line verification protocol based on these simulated set-up errors are derived, and the effect of the application of the prescribed set-up corrections on subsequent fractions are simulated. Details of this process and validation of the results can be found in [13, 17, 24]. Here, we describe the program input relevant to the present study. The program requires, for each relevant direction, the parameters that describe the patient set-up distribution if no corrections are applied (μ_{init}, σ, Σ_{init}) and the SD of the accuracy with which the set-up corrections can be executed by treatment couch shifts. The type of off-line verification protocol can be selected (e.g. the SAL protocol) together with the protocol parameters (e.g., (α, N_{max})). A correction damping scheme (i.e., the magnitude of the set-up corrections is reduced according to a priori information on systematic and random errors [45]) can be optionally selected. If applicable (as in the case of the SAL protocol), a 'second stage' follow-up protocol can be defined (see section 8.2.3). Finally, the number of treatment

fractions per patient (N_f) and the number of patients per simulation must be given. All treatment fractions are always simulated, so that the accuracy of the results is only limited by the number of patients of the simulation run (10^4 patients in all simulations, unless mentioned otherwise). However, statistics are also calculated for the sample of fractions that would have been imaged in clinical practice, i.e. those fractions for which imaging is mandatory according to the selected correction protocol. Hence, biases in results based on actually measured data can be quantified.

8.2.6   Evaluation of a new off-line protocol

In previous work [23, 24], we have shown that the reduction of Σ achieved by the first stage of the SAL protocol is much lower than expected from the required number of measurements per patient. We therefore introduced the   "No Action Level" or NAL protocol, which has only one parameter, N_m. Within this protocol, for each patient the first N_m fractions are imaged and the systematic error vector is estimated to be the mean of these measurements V_{N_m} . In subsequent fractions a correction equal to $-V_{N_m}$ is always applied, irrespective of the magnitude of V_{N_m} (apart from vector components with a magnitude below the accuracy with which table corrections can be executed) and the protocol ends. Hence, the number of measurements per patient is always N_m. The protocol is based on our clinical experience that the application of set-up corrections by couch translations requires nearly negligible treatment time [17], whereas the workload introduced by image analysis is considerable.

To verify if clinical application of the NAL protocol would yield satisfactory results, we started with a retrospective application to a large database of measured set-up errors in prostate patients [23]. The database consisted of 3D set-up data for 600 patients where, on average, set-up errors for 10 fractions per patient were available. Next, for each patient, we applied both SAL and NAL protocols to the available set-up data and calculated the effect of the corrections that would have been prescribed by these protocols. We found that the reduction of systematic set-up errors obtained with any SAL protocol could be achieved with significantly less measurements per patient using the NAL protocol. We then performed both a theoretical analysis of both types of correction protocols as well as Monte Carlo simulations analogous to those described above [24]. We found that, quite independent from the exact magnitudes of the initial systematic and random errors, a target magnitude of systematic errors (i.e. Σ) can be achieved with a NAL protocol using 2-3 times less portal images than a SAL protocol. Apart from the reduced imaging workload, this result is of particular importance in patient groups requiring dedicated PIF fields for set-up verification (section 8.2.2), as it enables a reduction of the dose given to healthy tissue without compromising on set-up accuracy.

At the time the patient set-up study presented in this paper was performed, we had not yet devised the NAL protocol. However, because the NAL protocol allows for a completely analytical calculation of the attainable set-up accuracy if the SD of the initial systematic errors (Σ_{init}) and random set-up errors (σ) are known, we can reliably establish its outcome in retrospect [24]. In fact, the SD of the systematic errors obtained with NAL is approximately $\Sigma_{\text{NAL}} = \sigma / \sqrt{N_m}$, and hence depends primarily on the magnitude of the random set-up errors [24]. In addition, as mentioned above, we have incorporated the NAL protocol in our Monte Carlo simulation software (section 8.2.5) so we can verify both

analytical predictions and simulated results. We have used these tools to evaluate the efficacy of a NAL protocol in head and neck patients using the distribution parameters of the initial (no correction) set-ups obtained in our SAL protocol study.

8.3 Results

8.3.1 Match accuracy

For each of the ten images, the SD of the reproducibility over the 5 matches was determined. For the set-up error measured in the Y direction, these SDs ranged from 0.2 mm to 0.9 mm, yielding a mean reproducibility over these images of 1 SD = 0.5 mm. For the Z direction, the corresponding range of SDs was 0.2 to 0.8 mm, and the mean was again 0.5 mm. For rotations, the range over the ten images was 0.0° to 0.6°, with a mean SD of 0.4°. These small random match inaccuracies (i.e., intra-observer variations) have a negligible impact on the derived systematic and random patient set-up errors, as will be discussed below.

8.3.2 Measured patient set-up accuracy

Of the 31 patients we studied, 15 patients had at least one set-up correction (48%). The protocol resulted in an average number of imaged and analysed fractions of 6.1 per patient. Most corrections were detected during the second stage of the protocol (see below), during which the action level is actually constant. Nevertheless, the protocol did reduce the systematic errors significantly (Table 8.1). The mean systematic errors (μ) were reduced to below 0.5 mm, whereas the SD of the systematic errors (Σ) was reduced from approximately 2 mm to 1 mm in both directions. The SDs of the random errors (σ) were approximately 1.5 mm in both directions, indicating that the fixation system ensures adequate patient set-up reproducibility. This partly explains why the systematic errors were already quite small initially [22], and why only half of the patients had their set-up corrected.

    The value of σ given in Table 8.1 is an average of the SDs of random errors measured in individual patients (σ_p). Since σ_p is slightly different in each patient, we can calculate the SD of σ_p in order to quantify the inter-patient variation in the random errors. However, the latter SD is partly determined by the sample size: if each patient had a true σ_p identical to the σ value given in Table 8.1, the measured σ_p's would not be identical due to the finite sample of 6 measurements per patient. The sample size thus presents a lower limit on the SD of σ_p measured in a real population, which can be quantified [28, 123]. In our study, these limits equal 0.5 respectively 0.4 mm for the Y respectively Z directions. The measured SD of σ_p is 0.6 mm in both directions. This is close to the lower limit, indicating that inter-patient variation in random set-up errors is small in the patient group we studied.

    For one patient, a particularly large set-up error of 5 cm was detected during the first fraction. The source of this error was traced to the simulation, during which the isocentre demarcation had been placed at a wrong position on the patient's mask. Once this mistake

had been corrected, the patient required no more set-up corrections. The loss of tumor dose was calculated and corrected for during the subsequent fractions. This case illustrates that portal imaging can very effectively minimise the possible damage that could ensue from incidental huge set-up errors.

Table 8.1 Measured distribution parameters of patient set-up errors in 31 patients. Values are given for patient set-ups with and without the application of the corrections generated by the SAL protocol described in the text, labelled "SAL" and "init", respectively.

	μ_{init}	Σ_{init}	σ	μ_{SAL}	Σ_{SAL}
Y (mm)	0.6	1.6	1.6	0.3	1.1
Z (mm)	-0.5	2.1	1.4	-0.5	1.2
θ (°)	0.0	0.5	1.1		

The SAL protocol allows for multiple set-up correction detections per patient [13], since after an applied correction, the protocol restarts. We found that 10% of the patients had multiple corrections detected and applied, comparable to the multiple correction frequency observed in applications of the SAL protocol to other treatment sites [17, 28]. In Fig. 8.3, the cumulative distributions of the 2D systematic errors (length of the average set-up error vector, where the average is calculated over all treatment fractions) with and without set-up corrections according to the SAL protocol are shown. With application of the protocol, all
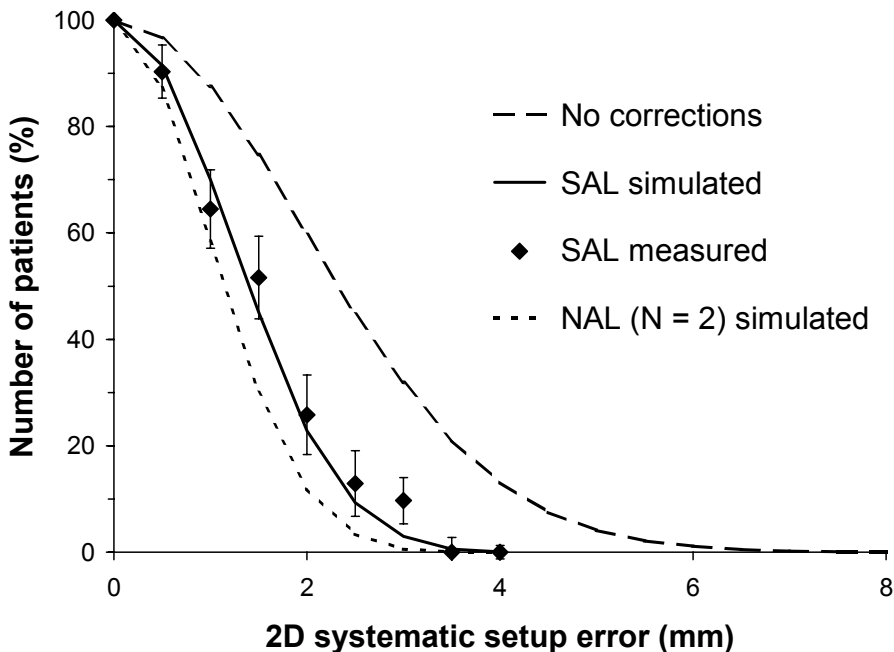

Figure 8.3 The cumulative distributions of the 2D systematic errors (i.e., length of the 2D vector) obtained without and with the SAL or the NAL correction protocols.

mean 2D errors are < 3.5 mm, consistent with expectations (section 8.2.3).

In Table 8.1, the measured rotation errors in the sagittal plane (θ) are listed. As can be seen, the systematic components are small ($0.0 \pm 0.5°$ (1 SD)), and will have had only a small impact on match results. This result, together with the small value of the (less important) random rotations, confirms that rotation errors posed no problem in the studied patient group. Note that we did not separate rotations into values measured either with or without a correction protocol; they are identical due to the reasons pointed out at the end of section 8.2.3.

If we correct the numbers in Table 8.1 for the finite match accuracy expressed by the SDs given in section 8.3.1, the impact is very small. Since the systematic errors are based on averages over approximately 6 fractions per patient, the correction of Σ is minute. The effect on σ is approximately 5% for both translations and rotations, which we consider negligible.

We analysed the rate of corrections during the 2nd stage of the protocol (see section 8.2.3) in more detail, since it is the source of a significant workload: without this stage, the number of required images per patient is approximately halved. In Fig. 8.4, we show the correction rates in the second stage as the number of measurements in this stage progresses. The correction rate for a certain measurement number N_{2nd} in the second stage is defined as the probability of detecting the necessity for a set-up correction in that measurement. We chose to let N_{2nd} count only the second stage measurements, i.e. not the 'in between' first stage measurements due to possible corrections detected in the second stage. We illustrate this procedure by considering a patient for which $N_{2nd} = 1$ at fraction 11. A correction is detected and applied after which the first stage SAL protocol is restarted. Given an N_{max} value of 2, and the low probability of multiple corrections, this stage will be finished near fraction 13. Next, the second stage protocol starts again and a follow up measurement is performed at, say, fraction 19. This now becomes measurement 2 in the second stage (N_{2nd} =2). If no correction is detected, the next follow up measurement increases N_{2nd} to 3, etc. By not counting the measurements in the short periods of the first stage SAL protocol initiated during the second stage, the correction rates thus derived are representative for the spread in time of the corrections detected in the second stage. For instance, if the interval between 2nd stage measurements is 5 fractions, the x-axis of Fig. 8.4 should be multiplied by 5, to obtain the correction rate in time. From Fig. 8.4, we see that the measured correction rate decays slowly with time, certainly considering the statistical uncertainty expressed by the $\pm$ 1 SD error bars in that figure. These uncertainties were estimated as follows. If the correction rate at a certain N_{2nd} is r, the expected number of correction detections is $r \times N$ where N is 31 patients in our study. We assume that the correction rate r is equal for each patient; this is a reasonable approximation once the first stage of the protocol is completed and if the inter-patient variation in the magnitude of the random set-up errors is small (see above). The actual number of corrections observed will then be binomially distributed with success probability $p = r$ and N = 31 [123]. For such a distribution, the SD on the number of actually observed corrections can be estimated as $\sqrt{N \times r(1-r)}$, where r is the observed correction rate.

8.3.3   *Comparison with simulation results*

Monte Carlo simulations including both the first and second stage SAL protocol were performed assuming a correction accuracy of 0.3 mm (1 SD; see section 8.2.3). The values of (μ_{init}, σ_{init}, Σ_{init}) were taken from Table 8.1. The mean imaging frequency during the

second stage was measured to be once every 8 fractions (in between the weekly and biweekly frequencies described in section 8.2.3) and we used this frequency in our simulations.

The simulations yielded set-up corrections in 49% of the patients, while on average 6.2 fractions were imaged per patient. These values are very close to the actual values obtained in our clinical study (previous section). The fraction of patients with multiple correction detections was 16%, consistent with the measured value of 10% to within the noise introduced by the group size of 31 patients (this was confirmed by repeating simulations with only $N = 31$ patients instead of $N = 10^4$). Final systematic errors (Σ_{SAL}) were calculated both from the simulated fractions that would have been imaged using the set-up protocol (superscripted 'imaged' in Table 8.2), as well as from all simulated fractions (superscripted 'all', which by definition equals the true value of Σ_{SAL}). As can be seen from Table 8.2, differences between the two values are very small, and both are almost identical to the measured values. Hence, we do not need to correct our measured values for the bias introduced by the finite sample of imaged fractions. In Fig. 8.3, the simulated cumulative distribution of the 2D systematic errors in case of application of the SAL correction protocol is shown. Again, the simulated and measured distributions correspond to within the statistical uncertainties posed by the measured data.
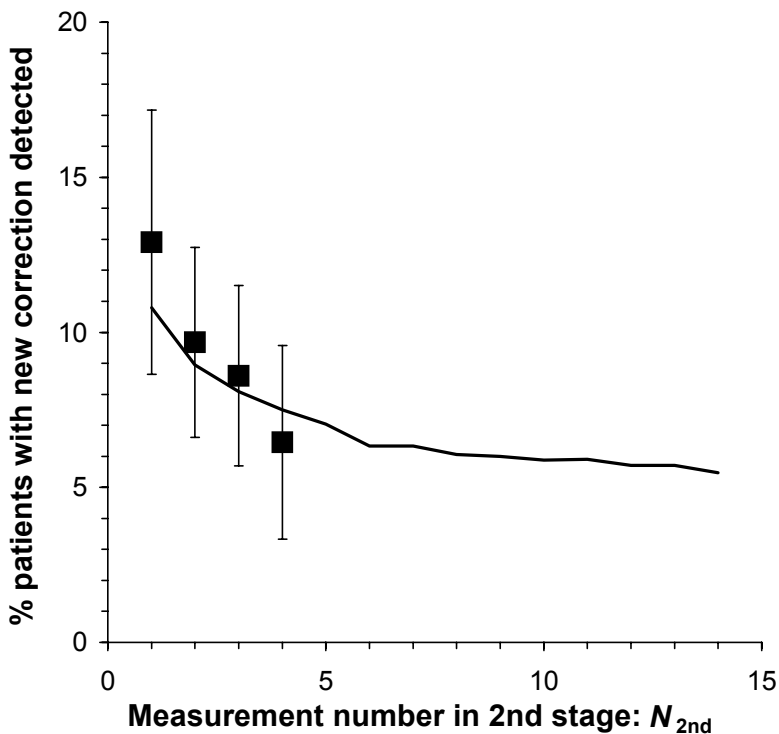

Figure 8.4 Frequency of corrections during the second stage of the off-line protocol as function of the number of check measurements in that stage. The solid line was obtained from simulations, whereas the closed squares were measured. The error bars represent $\pm$ 1 SD estimates of the statistical uncertainty on the measured correction rates.

It is of interest that, although no time trends nor transient events in the patient set-ups were simulated, the simulations confirm the tendency observed in the clinical data that slightly more than 50% of the set-up corrections is detected during the second stage of the protocol. This second stage was initially intended to correct for time dependent changes in the systematic set-up error [17], but obviously its practical effect is quite different. In fact, the simulations show that the rate with which corrections are detected during the second stage decays slowly with measurement number (Fig. 8.4), and the simulated rates are fully consistent with the measured rates. Hence, we conclude that the measured second stage corrections do not require time dependent changes in the systematic set-up errors. Instead, they can be ascribed to the statistical uncertainties in the decision process of the SAL protocol. The second stage corrections are inefficient if no time trends are present due to their late occurrence in treatment. Nevertheless, they do contribute to a reduction of systematic errors: our simulations showed that without the second stage, Σ_{SAL} would have been approximately 1.5 mm, instead of the 1 mm presently achieved (Table 8.1).

Table 8.2 Patient set-up distribution parameters derived from simulations. The systematic variations obtained in Monte Carlo simulations of the SAL protocol were split into values that would have been obtained from the "imaged" fractions only, as well as from "all" fractions. The systematic variations obtained by application of a NAL protocol with 2 set-up measurements per patient are given in the last column. The resulting systematic errors (from which Σ_{NAL} was obtained) were calculated from all simulated fractions per patient.

	$\Sigma_{\text{SAL}}^{\text{imaged}}$	$\Sigma_{\text{SAL}}^{\text{all}}$	Σ_{NAL}
Y (mm)	1.0	1.1	1.0
Z (mm)	1.1	1.2	0.9

8.3.4   Evaluation of the NAL protocol

Based on the approximate expression for the SD of systematic errors achievable with NAL (Σ_{NAL}) given in section 8.2.6, we estimated that a NAL protocol with $N_m = 2$ measurements per patient should yield similar results to the SAL protocol we actually applied. In the last column of Table 8.2, we show that for such a protocol, the remaining systematic variations Σ_{NAL} are indeed approximately 1 mm in each set-up direction. In Fig. 8.3, we give the cumulative distribution we would have obtained for a NAL protocol with $N_m = 2$ (the curve is taken from the Monte Carlo simulation, which is indistinguishable from the analytical result). As expected from the Σ_{NAL} values in Table 8.2, this cumulative distribution is slightly better than the result obtained with the first and second stage SAL protocol combined, while the number of necessary measurements per patient is reduced from 6 to 2. This reduction is consistent with the expected reduction in workload discussed in section 8.2.6, and is made possible because the SAL protocol is rather inefficient. The SAL inefficiency is mainly due to the restart of the first stage of the protocol after each set-up correction. Set-up information obtained before the restart is lost and so information on maximally N_{max} measurements is used for any set-up correction. As a result, the value of

Σ can not be reduced appreciably below $\sigma / \sqrt{N_{\max}}$ by the SAL protocol, even though the number of imaged fractions per patient can be much higher than $N_{\max}$. For instance, in the present study the number of imaged fractions per patient was 6, three times the $N_{\max}$ value of 2. On the other hand, the NAL protocol uses information of each imaged fraction, and due to its lack of action levels does not require subsequent iterations.

8.4 Discussion and conclusions

Our data show that with careful daily positioning using a simple head and neck fixation device, random patient set-up errors with an SD of 1.5 mm were obtained, whereas the SD of the systematic errors becomes approximately 2 mm for both directions in the sagittal plane (Table 8.1). The rotations we measured in this plane were small. Furthermore, because such small random set-up variations can be achieved, the systematic errors could be approximately halved using an off-line set-up correction protocol. In this section, we will first evaluate our 'baseline' set-up accuracy, obtained without any set-up corrections, by comparison to similar studies (in fact, using portal imaging at the treatment unit and comparable fixation devices) in the literature. Next, we will discuss the increase in accuracy obtained with set-up corrections and further improvements for the future.

The published studies on head and neck patient set-up accuracy can be roughly divided into two groups, the small patient number studies (N ≈ 10) and the medium to large patient number studies (N > 20). Unfortunately, the small patient number studies do not provide constraining information on the distribution of the systematic set-up errors. From statistical estimation theory, the confidence intervals for the actual value of Σ can be readily obtained from the value measured in a group of N patients [123]. For instance, if a value of Σ = 1.5 mm was measured in a group of 10 patients, the 95% confidence interval for the actual Σ would be 1.0-2.7 mm, a range covering both quite narrow and relatively broad distributions of systematic errors. In contrast, the numbers on the random set-up accuracy are usually quite reliable even if N is small, provided there are sufficient measurements per patient and the inter-patient variability is limited. We considered the small number studies in which the magnitude of both systematic and random errors are clearly reported. These were (1) Hunt et al. [59], on the positioning of 6 nasopharynx patients immobilised with a customised aquaplast head restraint, (2) Bel et al. [15], investigating the set-up of 10 patients with parotid gland or tonsillar tumors positioned in individual PVC head casts and (3) Yan et al. [145], who obtained set-up data for 12 head and neck patients immobilised in thermal plastic masks. For each study, we calculated the confidence interval of Σ and found it overlapped with the values of Σ_{init} listed in Table 8.1. Values of σ were in the range 1-2 mm in each study (the average values were σ_Y = 1.4 mm, σ_Z = 2.0 mm). Again, these values were consistent with our results (Table 8.1).

We also considered four studies with larger patient numbers. Weltens et al. [140] performed a set-up study on 43 head and neck patients positioned in either PVC or thermoplastic masks. Irrespective of the fixation device, they found random errors of magnitude σ = 2.1 mm in both the Y and Z directions. In their presentation of systematic errors, the three set-up directions can not be uniquely distinguished, but their values of Σ are approximately 3.5 mm. These values are significantly larger than the systematic errors we obtained and outside the confidence intervals of the small patient number studies discussed above. Even larger systematic set-up errors were found by Mitine et al [87], for 27 head and neck patients immobilised in plastic masks. They measured Σ = 4.3 and 4.6

mm in the *Y* respectively *Z* directions, and much smaller values of the random errors (σ = 2.5 respectively 2 mm). They concluded that set-up correction decisions based on a first-day image is an effective procedure for head and neck patients, because in their case the large systematic errors could be found with a relatively high accuracy from such a single image. This is obviously not true for patient groups where the systematic errors are comparable to the random errors, as is the case in our study was well as some of the previously mentioned ones. Finally, Hess et al. [57] performed a study on 95 head and neck patients immobilised with individual synthetic casts, and derived deviations between simulation and treatment set-up from a verification film of a single treatment fraction. From their data we can derive the SD of the overall set-up error, which equals $\sqrt{\Sigma^2 + \sigma^2}$. By fitting an error function to the closed curve in Fig. 1 in [57] we obtained a value of 3.4 mm as a mean for both set-up directions. Again, this number is significantly larger than the value of 2.3 mm we calculate from Table 8.1.

The above studies pertain to various head and neck tumor sites and treatment techniques, and hence to various patient conditions, possible weight loss, etc. In addition, several manufacturing materials and methods were applied to construct the fixation devices. Nevertheless, the magnitude of systematic errors that were reported in these studies are comparable to, or sometimes larger than, the values listed in Table 8.1. The lower limit on the spread of systematic errors that can be achieved with conventional mask fixation devices therefore seems to be approximately 2 mm (1 SD).

By applying set-up corrections, however, we could reduce the systematic errors further to Σ = 1 mm. Such small systematic errors will be of importance in highly conformal treatments. Moreover, the correction protocol ensures that the "tail" of the distribution of systematic errors can be controlled (Fig. 8.3), establishing a "guaranteed" accuracy in all patients. One extremely large set-up error, resulting from a misjudgement during simulation, was quickly detected and corrected for. The timely detection of such events requires regular portal imaging in all patients.

Because portal imaged guided set-up corrections can yield small systematic set-up errors, the reference images against which these errors are measured become increasingly important. For most patients in the present study digitized simulator films were used. However, in our current daily routine the reference is defined by DRRs whenever appropriate CT data are available. Various studies have shown that set-up errors at the simulator (which become systematic if the simulator defines the reference position) are comparable to those at the treatment unit [28, 76]. Hence, optimal set-up accuracy for patients with CT planning requires the use of DRRs as reference images.

In this study, we applied the SAL protocol for set-up verification and correction, and required on average 6 imaged fractions per patient. We found that the reduction of systematic errors was partly established during the second stage of the protocol. Originally, the second stage was introduced to detect possible transient effects in the patient set-up [17], i.e. to verify if the concept of a constant systematic set-up error breaks down in certain patients. For instance, weight loss could introduce a time dependent factor in the systematic set-up error, and perhaps also in the magnitude of the random errors. In our study, we observed a slowly decaying rate of correction during the second stage (Fig. 8.4). Our simulations showed that this result can be entirely explained by the limited statistics on which correction decisions are based and does not require the presence of transient effects. In fact, the corrections for systematic set-up errors during the second stage are partly triggered by the noise introduced by the random set-up errors, and as such the effectiveness of this second stage is not optimal. Firstly, the relatively late point in time at which the set-

up corrections are applied (on average, half way the total irradiation period) decrease the possible benefit of these corrections. Secondly, whether or not a patient has an actual set-up correction becomes somewhat arbitrary.

It can be shown that, in general, systematic set-up errors can be reduced faster and with less imaging workload using the NAL instead of the SAL protocol [23, 24]. The calculations presented in this paper confirm this result for head and neck patients: with information from only 2 fractions per patient, we predict a set-up accuracy slightly better than actually obtained with the SAL protocol using 6 imaged fractions per patient. The main reason for this increased efficiency was given in section 8.3.4. When dedicated PIF fields are used for set-up verification, as in our study, an additional benefit of the NAL protocol is the significant reduction of dose to healthy tissue without compromising on set-up accuracy. At present, we are conducting a randomised prospective study to evaluate the NAL and SAL protocols in prostate patients, and are studying the necessity of verification measurements or other follow up measurements. If the expected reduction in workload is confirmed, we will rapidly broaden the application of this protocol to other treatment sites, including head and neck cancer patients.

In conclusion we found that, although not optimally efficient, the SAL off-line protocol enabled the reduction and control of systematic set-up errors for head and neck patients immobilised with conventional PVC masks. The remaining systematic errors after application of this protocol were small (1 SD = 1 mm) and would have been very difficult to achieve using only masks.

9 Set-up verification of cervix cancer patients treated with long treatment fields; implications of a non-rigid bony anatomy

S. Quint, J.C.J. de Boer, J.R. van Sörnsen de Koste, B.J.M. Heijmen and M.J.J. Olofsen-van Acht

Abstract

Purpose: For cervix cancer patients, treatment fields may extend up to vertebra L1. In clinical practice, set-up verification is based on measured displacements of the pelvic rim as visible in the caudal part of the treatment fields. The implications of this procedure for the positions of bony structures in the cranial part of the fields were investigated.

Methods and materials: Twelve patients had 4 repeat simulator sessions. Both during treatment simulation (the reference) and the repeat sessions anterior radiographs were acquired covering the whole treatment field. The films were used to investigate differences between the cranial and the caudal part of the treatment field in day-to-day bony anatomy displacements.

Results: Both in the transversal and the longitudinal direction these differences were significant (3.5 mm, 1 standard deviation). Indications were found that large differences in cranio-caudal direction may be correlated with (non-rigid) internal pelvic rim rotations around a lateral axis. In the longitudinal direction, the position of L1 correlated much better with the position of vertebra S1 than with the position of the pelvic rim, which is usually used for set-up verification.

Conclusions: Due to the non-rigid bony anatomy of the studied patients, the usual set-up verification and correction procedure can result in set-up errors of 10 mm and more for structures in the cranial part of the treatment field, even in case of a perfect set-up of the pelvic rim. Possibly, other patient set-up and immobilization procedures may result in a better day-to-day reproducibility of the 3D bony anatomy shape. (Remaining) differences in anatomy position changes between the caudal and cranial field ends may be accounted for by using non-uniform CTV-to-PTV planning margins, or by an adapted patient set-up verification and correction protocol.

9.1 Introduction

Generally, treatment fields applied in radiotherapy account for uncertainties in the daily set-up of the patient in the treatment beams as well as for internal organ movement e.g. due to breathing or variations in rectum or bladder filling [62, 63, 79, 80, 125]. The larger these uncertainties are, the bigger the selected treatment fields, and the higher the probability of treatment related morbidity.

Patient set-up verification with Electronic Portal Imaging Devices (EPID) has been applied to reduce the uncertainty of the set-up of the bony anatomy of the patient in the treatment beams. EPID-images acquired during the first few treatment sessions have been used to minimize systematic deviations between the patient set-up during treatment and the reference set-up as defined by a simulator film or a digitally reconstructed radiograph (DRR) [17, 28]. Both systematic *and* random set-up errors have been minimized using daily-acquired images for on-line set-up verification and correction [91, 130]. Especially systematic set-up errors have a major impact on the magnitude of the required planning margins [125]. Several studies have been performed to investigate the use of EPIDs for detection and correction of internal organ movement, i.e. position variations of internal organs relative to the bony anatomy of the patient [3, 10, 36, 69, 129, 136].

For subgroups of cervix cancer patients [21, 94] we include the para-aortic nodes in the target volume, resulting in field lengths of up to 38 cm at isocentre distance. At the time that this study was performed, a Philips SRI-100 EPID was used for set-up verification of these patients. The (fixed) focus-to-detector distance of 160 cm resulted in a maximum detectable field length in cranial-caudal direction of 19 cm, defined at 100 cm from the focus. Therefore, bony structures in the cranial part of the treatment fields were not visible in the acquired EPID-images.

In the present study we have used radiographic films to investigate implications of the clinically applied set-up verification, based on the pelvic rim as visible in an EPID-image, for the position of the (non-visible) anatomy in the cranial part of the field.

9.2 Methods and materials

9.2.1 Patients

Twelve patients with a gynaecological malignancy were included in this study. All patients were treated with postoperative external beam radiotherapy. The Clinical Target Volume (CTV) [62] included the primary tumor region, the parametria, the proximal 2/3 of the vagina, and the para-aortic, iliacal and obturator lymphnodes. The dose prescription was 48.6 Gy, delivered in daily fractions of 1.8 Gy. All patients were treated in supine position. Long lateral, sagittal and transversal laser lines were marked on the patient. The intersection points of these lines were tattooed, together with the caudal field border. In case of discrepancy between the skin marks in longitudinal direction, the caudal field border tattoo was the decisive parameter for set-up [35].

9.2.2 *Discrepancies between caudal and cranial anatomy displacements*

To study correlations between day-to-day variations in the bony anatomy position in the caudal part of the anterior treatment field with position variations in the cranial part of the field, all twelve patients had 4 repeat sessions on the treatment simulator with 1 week intervals. In order to always fully capture the complete treatment field, both during treatment simulation and during the repeat sessions, two anterior simulator films were taken, one including the caudal field border, the other with the cranial border; the time interval in between acquisition of the films was typically 1-2 minutes. The isocentre and the corresponding transversal and longitudinal cross hairs, and a common part of the bony anatomy of the patient of at least 20 cm in cranio-caudal direction were visible on both films. In the analyses, the imaged isocentres and cross hairs were used to position the two films relative to each other. The common anatomical structures visible in the two films were then used to inspect for patient movement on the treatment couch in between acquisition of the two simulator films. In that case the films were excluded from the analyses. Patient positioning on the simulator table was performed as on the treatment unit, using lasers, and lines and tattoos on the patients' skin (see section 9.2.1).

All films were digitised using a CCD camera and a frame grabber coupled to a PC. Delineation of bony anatomy and registration of delineated structures in the digitised reference simulator film and a corresponding repeat film were performed with in-house developed software written in AVS (Advanced Visual Systems, Waltham MA). Corresponding cranial and caudal films were analysed independently. In-plane rotations
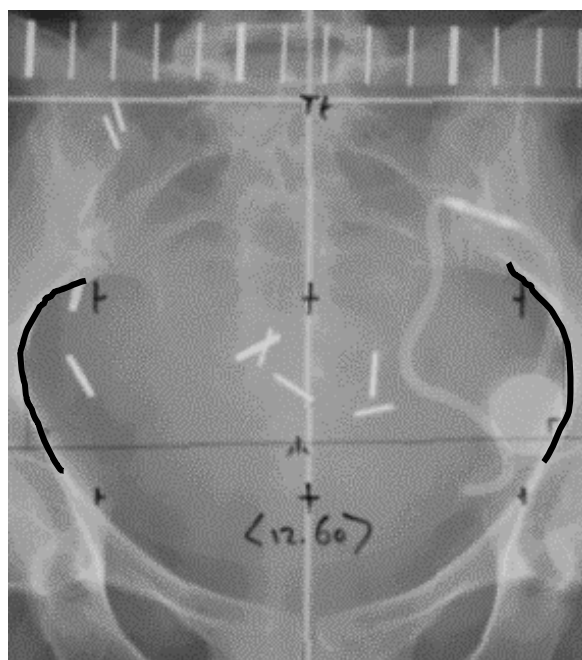

Figure 9.1 Partial delineation of the pelvic rim in a repeat film for measurement of the caudal anatomy translation (δXcaud, δYcaud) in relation to the reference set-up.

were small (<1°) and neglected in the analyses. A description of the applied software is given by Stroom et al. [130].

For the caudal repeat films, the position of the delineated lateral parts of the pelvic rim (see figure 9.1) - which were judged to be appropriate indicators for the (primary) target volume position - was compared to the position in the corresponding reference film as acquired during treatment simulation. For the cranial repeat films, delineations of the vertebrae L1 and L2 were used to assess position deviations with respect to the reference situation during treatment simulation. In the remainder of the paper, the symbols ΔX and ΔY refer to differences in deviations from the reference set-up between the cranial and the caudal field ends. E.g., $\Delta X = \delta X_{caud} - \delta X_{cran}$, with δX_{caud} the transversal displacement of the pelvic rim in the repeat film relative to its position in the reference film, and δX_{cran} the shift in transversal direction of L1/L2 in the repeat film relative to the reference position (see figure 9.2 for a graphical definition of δX_{cran}, δY_{cran}, δX_{caud}, δY_{caud}, ΔX, and ΔY).

To assess the inter-observer variation, 80% of all analyses were performed by two independent observers.

9.2.3 Pelvic rim out-of-plane rotations around a lateral axis

It is well known that pelvic patients may have significant internal (i.e. non-rigid) rotations of the pelvic rim around a lateral axis. In an anterior film such a rotation may appear as a translation in longitudinal direction. In those cases, the rotations may contribute to the measured ΔY.

The influence of internal pelvic rim rotations around lateral axes on measured ΔY values was (indirectly) studied by also assessing δY_{caud} displacements using fully delineated pelvic rims, instead of partial delineation (figure 9.1). In the absence of the above described out-of-plane rotations, δY_{caud}, derived from partially or fully delineated pelvic rims, should be equal. Observed differences in δY_{caud} between partial and full delineation for the subgroup of repeat set-ups with $\Delta Y < 4$ mm (group I) were compared to δY_{caud} differences in group
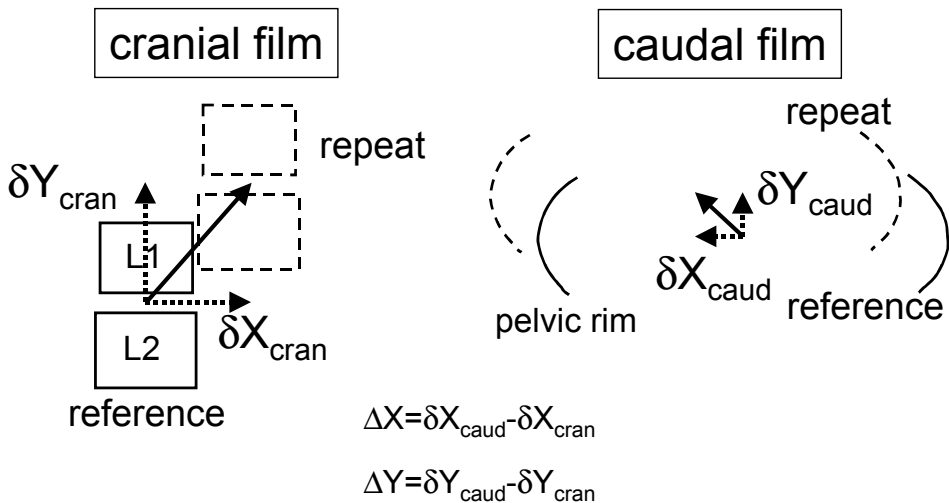
Figure 9.2 Schematic definition of the quantitites δX_{cran}, δY_{cran}, δX_{caud} δY_{caud}, ΔX, and ΔY.

II, defined as $\Delta Y \geq 4$ mm.

For the repeat set-ups in group II, the changes with respect to the reference situation in the distance between the center of mass of L5 (point III in figure 9.3) and the center of mass of the pelvic rim (point IV) were determined. The latter investigations required simultaneous analyses of cranial and caudal films. This could not be done with the software mentioned above. Therefore, corresponding caudal and cranial films were positioned on a light box with the proper overlap (section 9.2.2) and measurements were performed with a ruler.

Hanley et al. [49] have described techniques to investigate the influence of out-of-plane rotations also using lateral fields.

9.2.4   *Vertebra S1 as a predictor for the position of L1*

For the studied patient group, the vertebra S1 is always visible in (the upper part of) an EPID-image. For the repeat set-ups with translation differences ΔY of 4 mm or more (group II, see above), we investigated whether the center of mass of S1 (point II in figure 9.3) was a better predictor for the Y-coordinate of L1 (point I in figure 9.3) than the commonly used, partially delineated pelvic rim (see above). These analyses were also based on measurements performed with a ruler and films positioned on a light box (previous section).

Figure 9.3 The anatomical structures used for the ruler measurements on film. Points I, II, and III are the centre of mass of L1, sacrum and L5, respectively. Point IV is situated on the horizontal tangential defined by the two acetabuli, centred between the two vertical tangentials to the pelvic rim.

9.3 Results

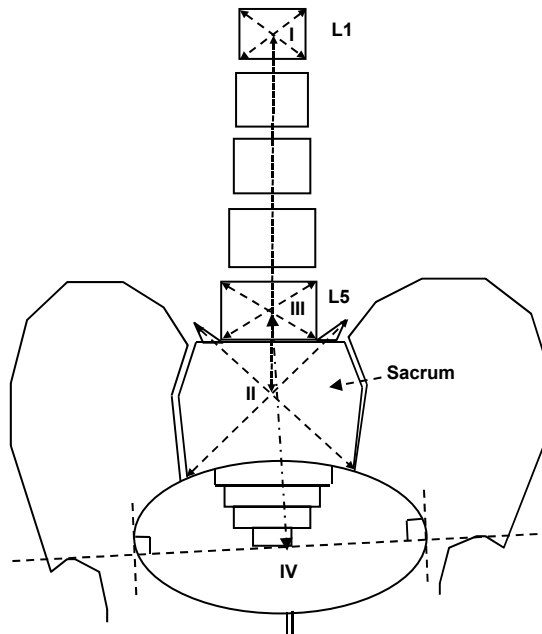### 9.3.1   Discrepancies between caudal and cranial anatomy displacements

Of the maximum number of 48 differences ΔX and ΔY in anatomy translation between the caudal and the cranial part of the treatment field that could be determined with the 12x4 repeat simulator sessions, 2 observations ΔX and ΔY had to be excluded from the analyses because of noticeable patient movement in between acquisition of the cranial and the caudal film. For the remaining 46 observations, the obtained translation differences, given in the format mean ± 1 standard deviation (SD), were $\Delta X = -0.4 \pm 3.7$ mm (range –10, +9 mm) and $\Delta Y = 1.0 \pm 3.6$ mm (range –5, +15 mm), see Table 9.1.

   A second observer repeated an arbitrary selection of 36 of the 46 analyses. The observed mean differences between the first and second observer in δX_{cran} and δY_{cran}, as derived from measured displacements of L1/L2, and the standard deviations were 0.3 ± 1.2 mm and 0.1 ± 1.5 mm, respectively. Differences in observed pelvic rim displacements yielded inter-observer variations in δX_{caud} and δY_{caud} of 0.0 ± 0.8 mm and 0.5 ± 1.2 mm, respectively. The inter-observer variations in ΔX and ΔY were 0.3 ± 1.4 mm and 0.0 ± 1.8 mm, respectively, which is consistent with the inter-observer variations given above as found for the separate cranial and caudal analyses.

   To correct the translation differences ΔX and ΔY found by the first observer for inter-observer variations, the standard deviations describing the variations between observers 1 and 2 in ΔX (1.4 mm) and ΔY (1.8 mm) were divided by $\sqrt{2}$ and then quadratically subtracted from the squared standard deviation of observer 1 for ΔX (3.7 mm) and ΔY (3.6 mm), respectively. The corrected values for the standard deviations of ΔX (3.6 mm) and ΔY (3.5 mm) show that observation errors are negligible compared to the found differences between the cranial and caudal parts of the anterior treatment field in anatomy position variations (Table 9.1). The observed large standard deviations for ΔX and ΔY point at a non-rigid bony anatomy in our patient group.

Table 9.1 Translation differences. ΔX and ΔY: mean displacement differences between the anatomy in the caudal field end and the anatomy in the cranial field end. The errors are standard deviations (1SD). X is the lateral direction, Y the longitudinal direction.

	based on 1 observer	corrected for inter-observer variation
ΔX (mm)	-0.4 ± 3.7	-0.4 ± 3.6
ΔY (mm)	1.0 ± 3.6	1.0 ± 3.5

9.3.2   Pelvic rim out-of-plane rotations around a lateral axis

For the groups I (N=36), and II (N=10), with small and large ΔY, respectively, (section 9.2.3) the observed differences in δY_{caud} between partially or fully delineated pelvic rims are 0.4 ± 0.7 mm (group I) and 0.7 ± 1.7 mm (group II); the latter standard deviations are corrected for inter-observer variations (1 SD = 1.1 mm) in the measured differences in

δY_{caud} related to partial or full delineation of the pelvic rim. Using a two-sided F-test to test statistical significance of the difference in the observed standard deviations (0.7 mm versus 1.7 mm) yielded a p-value of 0.0001. The difference in the observed standard deviations for the two groups suggests larger pelvic rim rotations around a left-right axis for group II. This is confirmed by the standard deviation of 3.0 mm that was found for the measured distance variations between L5 (point III in figure 9.3) and the center of mass of the pelvic rim (point IV in figure 9.3) for film pairs belonging to group II.

9.3.3   *Vertebra S1 as a predictor for the position of L1*

For repeat set-ups with ΔY larger than 4 mm (group II), the standard deviation describing the variation in the measured distance along the Y-axis between the center of mass of L1 or L2 and the center of mass of S1 was only 1.0 mm. This is much less than the observed standard deviation of 3.6 mm (for the groups I and II together, section 9.3.1), describing ΔY variations determined using the pelvic rim instead of S1. This data suggests that S1 is a much more accurate predictor for the Y-position of L1 than the pelvic rim.

9.4 Discussion and Conclusions

For subgroups of cervix cancer patients the para-aortic nodes may be included in the target, resulting in treatment fields as long as 38 cm (defined at isocentre). Using repeat simulator films for twelve of these patients, large discrepancies (standard deviations of 3.6 mm and 3.5 mm in X- and Y-direction, respectively, and maximum deviations of 10 and 15 mm) were observed between displacements of the pelvic rim with respect to the reference situation (treatment simulation) and the corresponding displacements of the vertebrae L1 and L2. These discrepancies are due to 3D bony anatomy shape changes (non-rigid bony anatomy), that are projected in the applied long treatment fields. Indications were found that the occurrence of the larger discrepancies in the longitudinal direction correlated with non-rigid, internal rotation of the pelvic rim around a lateral axis. In clinical practice, the usual set-up verification and correction procedure may result in errors of 10 mm or more in the position of structures in the cranial end of the treatment field, even in case of a perfect (corrected) set-up of the pelvic rim.

   At the time that we performed this study, a Philips SRI-100 EPID was used for set-up verification of these patients. Due to the fixed focus-to-detector distance (FDD) of 160 cm, the maximum detectable field length in cranial-caudal direction was only 19 cm (defined at isocentre distance). Therefore, bony structures in the cranial part of the treatment fields were not visible in the acquired EPID-images. In this study we have shown that the vertebra S1, which is visible in an anterior image acquired with the SRI-100, is a much better predictor for shifts in the Y-direction of bony structures in the non-visible cranial part of the treatment field than the pelvic rim.

   Our currently used Theraview[NT] EPIDs (manufactured by Theraview Technology by Cablon Medical), have an adjustable FDD, resulting in a maximum detectable field length of 29 cm for an FDD of 130 cm. However, even if also displacements of structures in the cranial part of the fields may be detected, their use in set-up verification and correction may be not trivial in case non-rigid bony anatomy changes result in different effective required couch translations for cranial and caudal structures.

Other patient set-up and immobilization techniques may possibly result in a more reproducible 3D bony anatomy than observed in this study. (Remaining) anatomy displacement differences between the caudal and cranial field ends, as detected with the EPID, may be corrected by couch translations to cancel pelvic rim displacements, combined with field shape changes to properly include cranial structures in the treatment fields. Alternatively, increased treatment planning margins may be applied around the cranial part of the delineated CTV. Further investigations are needed to test or implement these, and possibly other strategies, in clinical practice.

In this study, we investigated implications of the non-rigid bony anatomy of cervix cancer patients treated with long fields for set-up correction and verification. Also other patients are sometimes treated with very long fields, e.g. patients with Hodgkins disease. Detailed investigations for these other patient categories seem warranted.

Acknowledgements

10 Discussion

Although EPIDs were introduced clinically decades ago [5, 31, 72, 116], how to use them effectively is still a topic of research and debate [56, 72]. Apart from the relatively widespread use of the SAL set-up correction protocol [13], the general tendency is to develop and implement in-house methods for EPID-based set-up and dosimetric treatment verification. Furthermore, the formats in which results have been reported and the methods used to derive these results from measured data are still far from uniform. As a result, comparison of study outcomes is often difficult and sometimes impossible. Perhaps this non-uniformity is to be expected for a relatively young field of research. However, the current lack of standardisation of EPID-based verification methods is awkward as EPIDs serve to assure the quality of radiotherapy treatments, while radiotherapy relies in many other aspects on broadly accepted protocols (e.g. as issued by the ICRU).

One of the main tasks of radiotherapy physics is to ensure that the prescribed dose is actually delivered. Although it is now well accepted that mobility of patient anatomy is an important issue in dose delivery, emphasis is still on QA in treatment preparation and treatment unit performance. The high degree of QA thus achieved in planning systems and treatment units is quite misleading if the volume in which dose is delivered is not equally well controlled.

The lack of generally accepted strategies to apply EPIDs routinely is all the more surprising if one considers the suitability of EPIDs for QA purposes. In early studies, it was already shown that positioning errors could be accurately measured and corrected for with EPIDs [116], as confirmed by the generally small values of set-up errors reported in this thesis. Similarly, EPIDs based on CCD cameras have been shown capable of accurate dosimetry [chapter 2 of this thesis and references therein]. Nevertheless, a recurrent theme in meetings on EPI is the low degree of utilisation of EPIDs outside research purposes [56]. Below, we discuss some of the reasons for the infrequent use of EPIDs, and how these problems relate to the subjects treated in this thesis.

10.1 Are EPIDs hard to use?

AAPM Radiation Therapy Committee Task Group 58 (TG58) was initiated "to provide materials to help the medical physicist and colleagues in the clinical implementation of EPIDs in radiation oncology". In their recent publication [56], TG58 quantifies the disappointing use of EPIDs in the US and describes the main properties of commercially available EPIDs. Furthermore, procedures are described for routine maintenance of EPIDs and some clinical applications are discussed. This report is valuable as an introduction to new users, and provides guidelines that will assist in clinical EPID introduction. Unfortunately, TG58 does not provide much more information than previous EPID reviews [31, 116] as to why EPIDs are infrequently used, apart from the often-reported insufficient image quality. However, some of the essential reasons behind the lack of clinical EPID use are perhaps quite straightforward and will be discussed below.

There exists a major gap between the software to acquire and analyse EPID images used in many published studies and the software commercially available. For example, our initial studies on patient set-up and dosimetry using the Philips SRI-100 EPID [2, 38, 52, 99], were only possible because of our knowledge of the accompanying software. As this EPID

was developed in collaboration with our institute, we could modify the source code when necessary. This software-tuning phase, as well as the transfer of knowledge accumulated during that phase, is often not sufficienty appreciated by EPID manufacturers. For instance, many of the procedures required for routine calibration and maintenance of EPIDs described by TG58 are not integrated in EPID software. In addition, the possibility to extract relevant patient information from the treatment planning system or treatment unit software is not universally available, requiring considerable manual labor to enter new patients in the EPID database. Similar limitations exist for importing reference images.

Another hurdle is the lack of integrated and easily configurable patient set-up verification and correction protocols. Here, the user must often fall back on homemade software. Such protocol software must combine information measured in EPIs of various treatment beams and handle the involved co-ordinate systems properly. Furthermore, the possibility to query and review statistics obtained for a group of patients defined by user-specified criteria is usually lacking. Given these examples of important limitations in commercially available software, routine use of EPIDs often implies a high workload and it is not surprising that EPIDs are used infrequently or incompletely in many radiotherapy centers.

Finally, the relative success of EPID application in Europe, and certainly in the Netherlands, may be partly ascribed to the important role of technologists in the analysis and correction of patient set-ups. If the analyses of EPI and subsequent actions have been fully protocolized, and radiotherapists have agreed upon these protocols, the actual execution can be transferred to the technologists. The technologists in our hospital have, in general, found this a stimulating approach, as it helps them to visualize the accuracy of their work and to quantify and correct the set-up errors for patients that seem to be hard to position. At the same time, radiotherapists are relieved from the burden of tracking the EPI results of many patients on various treatment units, although they are informed in case of irregularities.

10.2   Will flat panel imagers boost the clinical use of EPIDs?

Lack of proper image quality is often mentioned as a reason for limited EPID use. However, the TG58 report [56] mentions that this observation is in contrast with the excellent quality reported in literature. The gap between what is achieved in specialised centers and the image quality available to the general user confirms our suggestion that both software and hardware of commercially available EPIDs insufficiently transfer the knowledge gained by research groups. We find that well-calibrated EPIDs have long been capable of providing proper image quality for most practical purposes (see below), but perhaps well-calibrated EPIDs are rare. The procedures described in the TG58 report, which are largely aimed at establising a working QA program for the EPIDs themselves, can help alleviate this problem. Because of promising image quality and recent commercial availibility, the attention has currently shifted to flat panel EPIDs (either based on direct or indirect radiation detection [5]) and the hope is that they will solve many EPI-related problems. However, below we argue that the benefit of flat panel detectors for regular portal imaging may be quite limited.

We have achieved small set-up errors ($\Sigma \sim 1$ mm, $\sigma \sim 1$-2 mm) using camera-based EPIDs for various treatment sites (this thesis). Furthermore, other groups have obtained similar results when using similar set-up protocols [17, 73, 75, 135, 138, 139]. Hence, the image quality of EPIDs over the past years has certainly been sufficient for accurate

determination of the position of relevant (bony) anatomical landmarks. This is confirmed by the small inter- and intra-observer inaccuracies reported in this thesis. As described in chapter 2, CCD camera-based EPIDs (CEPIDs) have a track record for precise transit dosimetry as CCDs are extremely stable and highly linear detectors [38, 52, 69, 97, 99-101]. Although the optical cross-talk is much smaller in flat panels than in CEPIDs [5], cross-talk in CEPIDs can be accurately corrected for [chapter 2 and references therein]. Moreover, the glare generated in the fluorescent screen is also present in indirect flat panel imagers [82]. Furthermore, frame-transfer CCDs allow for instantaneous complete sensor read-out and therefore are highly suited for verification of dynamically delivered IMRT.

It has been shown in chapters 4, 5, and 8 that EPIs of 10 MU or less, obtained with a CEPID with an uncooled CCD camera (chapter 2), allow for accurate set-up verification based on anatomical landmarks. Even if dedicated "imaging only" fields (PIFs) are required, and the total dose delivered in these fields to uninvolved tissue may not exceed ~ 1 Gy, this exposure is sufficiently low for off-line correction protocols, particularly in case of the NAL protocol (chapters 6 and 7). Using a cooled CCD camera, we have shown that exposures can be lowered to such levels that on-line protocols based on PIFs become possible [11] (see below).

EPIDs can also be used for the visualisation of organ position and movement through small implanted markers. Marker positions have already been measured using various camera-based EPIDs [9, 65, 126, 136, 137]. We have obtained images using a CEPID equipped with a sensitive, cooled CCD [11]. Various clinically applied markers, including small platinum ones (diameter 0.9 mm, length 3 mm) could be observed at exposures of 2-3 MU in lateral images. The straightforward image processing applied to these images and the less than optimal use of the CCD camera dynamic range implies that we can push exposures even lower. The spatial resolution, as expressed by the half-value frequency of the normalised MTF [106] was f_{50} = 0.36 lp/mm, markedly larger than in conventional CEPIDs [118]. Equally important, the occurrence of 'hot pixels' due to radiation damage by photo-neutrons produced at photon beam energies > 10 MV, was strongly reduced by the Peltier cooling of the CCD. After 8 months of operation on an intensively used Linac (radiating most of the time at 23 MV) no hot pixel has yet been observed and the dark current pixel-to-pixel variations are still extremely small. This result promises a dramatic increase in lifetime compared to non-cooled radiation damaged CCDs, which usually do not last much longer than 12 months at beam energies > 15MV. Hence the largest problem of CEPIDs (gradual loss of image quality) has been tackled.

Another benefit of flat panel devices that is often mentioned is their compact design. Although this design allows for a more compact state when retracted, the large cross-section when deployed offers less clearance than camera-based EPIDs when in actual use. Furthermore, large fields cannot be imaged with flat panel imagers and asymmetrical fields in general require extra caution, because radiation damage to the electronics surrounding the actual detector area must be prevented. Given the large financial costs still involved with these panels and the uncertainty about their lifetime in clinical environments, the case for flat panel EPIDs may not be entirely obvious.

We conclude that the limited clinical use of conventional EPIDs is mainly due to poor calibration and provision of relevant tools by their manufacturers, and therefore not inherent to the technical capabilities of these devices. Hence, even if flat panel EPIDs become the default choice, their clinical use will be largely limited by the same factors as in previous EPID types. In particular, the lack of user-friendly software tools means that local efforts are required and quantitative regular use, especially if set-up corrections are involved, may still be unattainable in many clinics.

10.3   Off-line set-up corrections: is early verification sufficient?

In this thesis, we described the development of treatment verification and correction at an acceptable increase of portal imaging related workload. The focus was on reduction of systematic errors, as it was shown in a population statistical approach that such errors require larger geometrical planning margins than random set-up errors (chapter 3).

In contrast to the huge number of set-up studies in which the magnitude of observed set-up errors is reported for various treatment sites, the number of studies in which patient set-up is actually *corrected* on the basis of EPI is still quite limited. In particular, there are only a handful of studies in which set-up corrections are based on a more than very crude protocol [60, chapter 6].

For many years, the shrinking action level (SAL) protocol has been the standard for off-line corrections [13]. In this protocol, the magnitude of the systematic error is estimated as the mean of the measured set-up errors and compared to an action level which shrinks as the number of measurements increases (in fact as $\alpha / \sqrt{n}$ where α is the initial action level (mm) and n is the number of measurements). The shrinking level reflects the increased accuracy of the estimate of the systematic error with increasing number of measurements. If no correction was necessary during a specific number of measurements (N_{max}), the protocol has finished. If a set-up correction is required, the protocol starts again and hence requires at least N_{max} more measurements to finish (see Chapters 4-8 for details). The No Action Level (NAL) protocol introduced in chapters 6 and 7 has a different approach. It is based on the observation that, since off-line protocols correct for a systematic set-up error, one should not correct before this error is known with sufficient accuracy. It is straightforward to derive that if set-up corrections are determined from the average set-up error in N_m measurements, one achieves $\Sigma_{\text{residue}} = \sigma / \sqrt{N_m}$ *after correction* (chapter 6). As the distributions of random set-up errors and systematic errors *before correction* are often comparable ($\Sigma \approx \sigma$, chapters 3-9, and references therein), off-line set-up corrections based on a single measurement ($N_m = 1$, e.g. a check during the first fraction) are ineffective. In the NAL protocol, early in the fractionated treatment, each patient is imaged for N_m subsequent fractions without any correction and the systematic error is estimated as the mean set-up error in those fractions. In the remaining fractions, this error is always corrected for by a table shift; there is no action level involved. By the formula given above, N_m fully determines the residue systematic errors for the patient population. Due to various reasons, discussed in chapters 6 and 7, the NAL approach reduces systematic errors with much less measurements than the SAL protocol and a reduction in workload of a factor 3 can be achieved.

The NAL protocol may serve as a model for a protocol in which the systematic errors are corrected for early in the fractionated treatment. For such protocols, however, systematic shifts of patients over the course of treatment (time trends) would not be detected. We will therefore address the issue of these time trends at some length. Since the first quantitative report on their magnitude [41], it is widely accepted that repeat measurements during the entire treatment period are mandatory. We will first discuss available literature on time trends, both on bony anatomy as well as on prostate movement. We will also present initial results of a prospective study on the efficacy of the NAL protocol in prostate patients, which included an analysis of time trends. Finally, we investigate whether a practical

protocol could be devised to detect significant trends at an early stage based on a statistically valid criterion.

10.3.1 Time trends in published studies

Time trends in set-up errors have been mainly analysed in terms of 'linear trends': when the patient is positioned on external marks, the actual position of the (bony) anatomy shifts by a fixed amount relative to the previous treatment fraction. A clear mechanism for such linear trends is lacking, and it is important to stress that large trends can never be linear. After all, there are physical limits to the amount of mobility of patient anatomy relative to the fixation device or skin, particularly when one considers a systematic time dependent shift. Linear trends may therefore only serve as a first order description pertaining to small effects.

The first systematic study on time trends was performed by El-Gayed et al. [41], who studied two small patient groups (10 prostate and 10 rectal carcinoma patients) and no detailed studies have become available since, with the exception of Hanley et al. [49]. As discussed in chapter 5, there are reasons to doubt the statistical value of the evidence on trends given in these publications. But there are other, non-statistical doubts about the importance of the trends reported in these studies. El-Gayed et al. [41] mention that progressive displacements with time due to gradual changes in muscle tension would predominantly cause non-linear trends, i.e. the fraction-to-fraction displacements would decrease in magnitude as treatment progresses. However, these authors subsequently state that such non-linear trends were not observed in their study. They propose that the large linear trends they did observe (leading to displacements over time of 1 cm) were due to the non-randomness of re-inking skin marks. Therefore, their paper, which inspired most subsequent trend investigations because of the large magnitude of the reported trends, suggests that these trends may in fact not be related to patient posture or anatomical variation, but are a consequence of their procedure of re-marking. Contradictory to this observation, they maintain that repeated imaging over the course of treatment is necessary for proper set-up correction, a thought that has found foothold since.

In the study of Hanley et al. [49] on set-up accuracy in 50 prostate carcinoma patients, it is concluded that the statistically significant trends are clinically insignificant, i.e. the net displacements due to significant trends are no larger than 2-3 mm. In a retrospective analysis, Yan et al. [145] reported time dependent drifts in 4 out of 27 lung patients and 8 out of 25 pelvic patients. Unfortunately, the magnitudes of the trends are not mentioned, and it is unclear if the Kalman-filter method they applied to determine optimal set-up corrections actually tests for statistical significance. As such, this study does not present quantitative information on the importance of trends. In a recent study by Alasti et al. [1], in which 32-41 fractions per patient were imaged in 23 prostate patients, no evidence for time trends was found, but it is not evident how the set-up correction protocol applied by these authors influenced that result.

The set-up studies described in this thesis strongly suggest that time trends are unimportant. In the N = 600 prostate patients study (chapter 6), we have shown retrospectively that the NAL protocol reduces systematic errors efficiently over the entire treatment period. The distribution of systematic errors measured prospectively with the SAL protocol followed by weekly verification measurements were consistent with the results from Monte Carlo simulations (MC) of set-ups without time trends. This consistency with MC includes important observed variables such as detection frequency of corrections,

distribution of number of corrections per patient and the fraction of patients with corrections. In particular, the relatively large frequency of corrections detected during the weekly measurements in the so-called second stage of the protocol [17] is reproduced in the MC simulations. If we consider the sub-group of 454 prostate patients with SAL parameters α = 9 mm, N_{max} = 4 [13], both measurements and simulations without time trends show that the SAL protocol yields corrections in approximately 30% of the patients, whereas the second stage increases this number to 50%. Therefore, frequent and large trends are no pre-requisite for explaining frequent corrections in the second stage. Instead, MC simulations show that such corrections are rather the result of repeated application of the action-level test. Nevertheless, the second stage is necessary in combination with SAL, not in particular to reduce the effect of trends, but to reduce systematic errors to the commonly accepted levels of Σ = 1-1.5 mm which are usually not achieved by SAL alone [17, chapter 6]. Similar results regarding the second stage and trends were found for head and neck patients (chapter 8). In fact, because possible weight loss in the latter group may incur increased mobility within the patient mask, we investigated in detail whether the correction frequencies generated in the second stage required time dependent changes in the systematic set-up errors. Within the error of measurement, this was not the case (section 8.3.3).

Based on the available publications, we conclude that there are no published population-based set-up studies for bony anatomy to support the claim that time trends need to be taken into account. Rather, there seems to be substantial support for the statement that trends are unimportant. Nevertheless, we thought that definitive acceptance of a NAL-like protocol, based solely on set-up measurements in the initial part of the fractionated treatment, demanded a prospective study to exclude potential adverse effects due to time trends. The initial results of such a study are presented below.

10.3.2 *A prospective study on NAL and the importance of time trends; preliminary results*

We studied 30 prostate cancer patients in the NAL protocol with N_m = 3, according to the methods described in chapter 6. A set-up correction was executed if it was larger than 1.5 mm in the left-right or superior-inferior directions and larger than 1.0 mm in the anterior-posterior (AP) direction. After correction, the accuracy of the applied correction was validated with the COVER procedure described in chapter 7. Once this procedure was completed, image analysis was finished for the technologists at the treatment unit, who applied the same set-up correction for all subsequent fractions. For the purpose of our study, as many fractions as possible were imaged. These images were intended solely for post-correction monitoring of the systematic error and were analysed in 'background' with no feedback to the treatment unit. In the same period, 55 patients were treated with our standard SAL protocol for prostate patients, which includes a second stage of weekly verification measurements ([17], chapters 4 and 8).

 The number of measurements per patient required for the NAL + COVER procedure was 4.0, for SAL it was 9.6. However, for the NAL patients, set-up data was available for, on average, 23 fractions per patient with a good coverage over the entire treatment period. This allowed for an accurate determination of residue systematic errors in this group. The resultant cumulative distributions of the 3D systematic displacements are given in Fig. 10.1. There is no statistically significant difference between the distributions achieved with NAL

and SAL, despite the fact that, contrary to NAL, SAL did include weekly verification measurements.

Since no 3D systematic displacement exceeded 4 mm for NAL, this result demonstrates that time trends played no role of importance. In fact, NAL reduced the values of Σ from 1.6-2.2 mm to 1.1 mm in each direction, whereas MC simulations without time trends yielded a Σ-value of 1.0 mm in each direction. Given the sample size of 30 patients, MC and measured values obviously correspond within statistical uncertainty ($p = 0.47$). As a result, there is no need for time trends to describe the set-up data measured in this study.

Nevertheless, given the well-sampled set-up data in the NAL patient group, we could analyse linear trends independently. The trend magnitude and statistical significance were determined with linear regression analysis [41, 49]. Apart from $p < 0.05$, we required that the trend magnitude should lead to a net displacement over 30 post-correction fractions of at least 3 mm in order to be clinically significant. In addition, post-correction set-up data had to be available over a time period spanning at least 15 daily fractions to identify lasting trends. The largest trend thus observed was 0.14 mm/fraction (AP direction), yielding a maximum displacement from fraction 5 to 35 of 4 mm. Hence, the increase in the systematic error for this particular patient and direction was 2 mm. Note that part of this trend may well be due to statistical noise, as the SD of the trend magnitude due to finite sampling, even in this well sampled study, is approximately 0.06 mm/fraction. A more general statistical analysis was performed, the details of which are beyond the scope of this
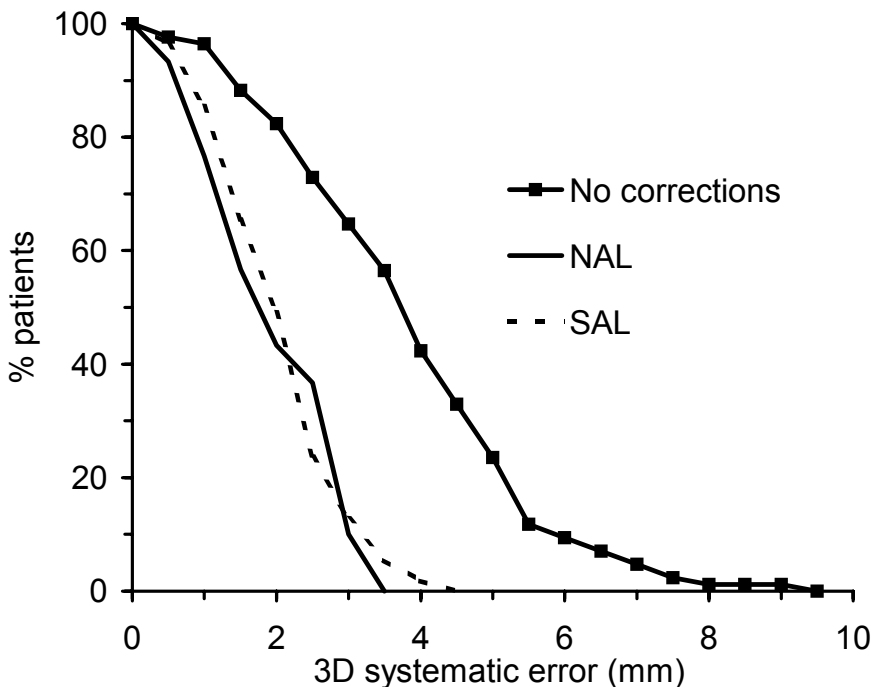

Figure 10.1 The cumulative distribution of the 3D systematic setup errors obtained for 30 patients treated with NAL (solid line) and 55 patients with SAL (dashed line). The distribution obtained if no corrections would have been applied was identical in both groups and is given by the solid line with squares.

discussion. We only mention here that the distribution of the trend magnitudes over the patients is approximately given by 0.0 ± 0.03 (1 SD) mm/fraction (corrected for the statistical noise described above), all under the assumption that linear trends did indeed occur. This narrow distribution of trends confirms that trends played no role of importance in the studied patient group.

10.3.3 Internal organ motion and time trends

The studies discussed above pertain to bony anatomy, whereas internal organ motion might introduce different trends. The largest and most detailed study presently available investigated the position of the prostate and seminal vesicles in 50 patients in prone position through multiple CT scans over an 8-9 week period [85]. Trends were not observed in the population as a whole, nor where any strong individual trends seen over the entire treatment period. These findings are consistent with a similar earlier study on prone positioning in prostate patients [127]. In the latter study, a time trend for patients in supine position was observed in the sagital plane but this trend was non-linear and its effect vanished towards the end of treatment. The trend was associated with a difference in laxation procedure between planning and repeat scans. The net displacement ascribed to the trends was 1-1.5 mm in the AP and cranial-caudal directions during part of the treatment period, quite small compared to the systematic displacements due to prostate movement alone (Σ = 2.5-2.7 mm) in the same group. Hence, at present no indication for important trends in prostate position exists. Furthermore, the small observed trends can be counteracted with appropriate procedures (e.g. prone positioning). We are not aware of similar studies on time trends for other target organs.

10.3.4 Can time trends be detected in time using a statistical criterion?

Suppose that a treatment site could be found in which trends are an important limiting factor to the set-up accuracy achievable with off-line correction protocols. In that case, trends should be identified at a sufficiently early stage. Since systematic errors of Σ = 1-1.5 mm can be achieved on a routine basis with off-line corrections, a systematic error larger than 3 mm in a specific direction has become rare. Hence, if systematic set-up errors are reduced effectively during the first treatment week, a gradual patient shift of 5 mm over the entire treatment period would introduce an unexpectedly large error. As pointed out above, such a shift is close to what one would expect to be the maximum trend that can be described linearly. If 30 fractions are delivered after the initial set-up correction, the shift per fraction is 0.17 mm. This magnitude is much larger than trends actually observed in our NAL study (previous section) and its probability of occurrence, based on the SDs given there, is entirely negligible. Nevertheless, even such a large shift would be hard to find amidst random displacements with a typical SD of σ = 2 mm. We assume that trends will be detected by the linear regression method described above [41, 49], i.e. the slope of the linear fit to the measured set-up data is tested for statistical significance. If the significance level p = 0.05 and set-up errors are measured *every day*, we derive that approximately 20 fractions need to be imaged and analysed to detect the 0.17 mm/fraction trend with 50% probability. Obviously, this would be a highly inefficient procedure, requiring a vast workload to detect trends at a late stage. Furthermore, if so much imaging per patient is feasible, on-line corrections, which render time trends unimportant, become the evident

choice. Instead, if time trends are sought with weekly imaging (say once in 5 fractions), the spread in time helps to detect trends with fewer measurements. Nevertheless, we then find that at least 8 measurements would be required and hence important trends are not likely to be detected before fraction 35.

The required number of measurements derived above holds if first all measurements are obtained, and next the significance of a possible trend is established. In practice however, one would reassess the significance of a trend with each new measurement. This implies a rapid increase of false trend detections for conventional significance levels of about 0.05. Suppose that 3 set-up measurements are performed during the first week of treatment and that the significance of a time trend is determined on the basis of subsequent weekly measurements. The probability of false trend detection increases with each measurement, particularly since 3 set-up directions are tested independently. We performed MC set-up simulations for patients in which no time trends occurred and who had random set-up errors $\sigma = 2$ mm in all 3 directions. We assumed that 4 weekly check measurements were performed (i.e. approximately up to fraction 25) and a trend was deemed significant as soon as a measurement yielded $p < 0.05$ and a predicted shift over 30 fractions of at least 3 mm. Because no trends were simulated, any significant trend was a false detection due to statistical uncertainty. Such false detections occurred in 40% of the patients. To reduce this unacceptably large false detection rate to 5%, the statistical significance level must be lowered to $p = 0.005$. However, large actual trends of the magnitude discussed above would then be detected in only 6% of the cases and even huge trends, causing displacements of 1 cm, would have a probability of detection of less than 30%. We conclude that no satisfactory procedure exists to detect large trends at a suitably early stage without incurring a tremendous rate of false positives.

10.3.5 NAL-based set-up corrections: conclusions

As stressed above, most studies have found no evidence for relevant time trends, even in large patient groups. Hence, there are obviously positioning methods that render the occurrence of trends rare to non-existent in all studied sites. This implies that the NAL protocol can yield a significant workload reduction compared to conventional protocols without loss of positioning accuracy, as demonstrated clinically in section 10.3.2. It should be appreciated that off-line protocols are aimed at creating an acceptable distribution of systematic errors for a patient *population*. As indicated above, weekly verification measurements are not suited for discrimination between time trends and random set-up fluctuations. Therefore, only if a huge number of false positive detections are allowed, the effect of hypothetical time trends may be reduced. It is questionable if the very infrequent or even purely speculative occurrence of an actual trend warrants the large workload involved with weekly imaging for all patients. This is particularly true if that workload limits the application of set-up corrections to a small group of patients, while the increase of accuracy in that small group due to weekly imaging is negligible. One could even argue that, in case of dedicated PIFs (chapters 4, 8) where healthy tissue is irradiated only for imaging purposes, the lack of a convincing demonstration of trends further precludes weekly check measurements and a protocol like NAL is the optimal solution.

Recently, the effect of NAL set-up corrections was analysed from a radiobiological point of view for prostate cancer [4]. It was demonstrated that, for $\Sigma \geq \sigma$, a NAL protocol with $N_m = 3$ was sufficient to gain most of the benefit in tumor control probability incurred by

set-up corrections. The optimal efficiency of such a protocol under similar conditions was also derived in chapter 6 on purely geometrical grounds.

In chapter 7, we developed the COVER procedure as an extension of NAL, in order to verify the proper execution of set-up corrections with minimal workload. We think that set-up corrections will ultimately become fully automated. EPID and accelerator control software will be interconnected and all corrections will be performed automatically through treatment couch shifts. Consequently, the validation of the correction execution will occur primarily in the acceptance phase of the EPID-accelerator-treatment couch interface. In principle, this would allow for a NAL protocol with only N_m and not N_m+1 measurements. Nevertheless, it is likely that a post-correction image will still be required to satisfy the physician, as well as for administrative or legal purposes and so the COVER procedure would remain a useful method to detect systematic errors in the procedures used for set-up correction.

The NAL protocol, combined with COVER, is now applied in our clinic to position prostate, lung, head-and-neck and gynaecology patients. Furthermore, we are investigating the use of NAL for correction of systematic positioning errors rooted in internal organ movement (as detected in EPI with markers), since the prevailing importance of decreasing systematic errors early in treatment also holds true for internal motion.

174

References

1. Alasti H, Petric MP, Catton CN and Warde PR. Portal imaging for evaluation of daily on-line setup errors and off-line organ motion during conformal irradiation of carcinoma of the prostate. *Int J Radiat Oncol Biol Phys* 2001; 49: 869-884
2. Althof VGM, de Boer JCJ, Huizenga H, Stroom JC, Visser AG and Swanenburg BN. Physical characteristics of a commercial electronic portal imaging device *Med Phys* 1996; 23: 1845-1854
3. Althof VGM, Hoekstra CJM, Te Loo HJ. Variation in prostate position relative to adjacent bony anatomy. *Int.J Radiat Oncol Biol Phys*. 1996; 34: 709-715
4. Amer AM, Mackay RI, Roberts SA, Hendry JH and Williams P.C. The required number of treatment imaging days for an effective off-line correction of systematic errors in conformal radiotherapy of prostate cancer-a radiobiological analysis. *Radiother Oncol* 2001; 61: 143-150
5. Antonuk LE. Electronic portal imaging devices: a review and historical perspective of contemporary technologies and research. *Phys Med Biol* 2002; 47: R31-R65
6. Antonuk LE, El-Mohri Y, Huang W, Jee KW, Siewerdsen JH, Maolinbay M, Scarpine VE, Sandler H and Yorkston J. Initial performance evaluation of an indirect-detection, active matrix flat-panel imager (AMFPI) prototype for megavoltage imaging. *Int J Radiat Oncol Biol Phys* 1998; 42: 437-454
7. Austin-Seymour M, Kalet I, McDonald J, Kromhout-Schiro S, Jacky J, Hummel S and Unger J. Three dimensional planning target volumes: a model and a software tool. *Int J Radiat Oncol Biol Phys* 1995; 33: 1073-1080
8. Balter JM, Ten Haken RK, Lawrence TS, Lam KL and Robertson JM. Uncertainties in CT-based radiation therapy treatment planning associated with patient breathing. *Int J Radiat Oncol Biol Phys* 1996; 36: 167-174
9. Balter JM, Lam KL, Sandler HM, Littles JF, Bree RL, Ten Haken RK. Automated localization of the prostate at the time of treatment using implanted radiopaque markers: technical feasibility. *Int J Radiat Oncol Biol Phys* 1995; 33: 1281-1286
10. Balter JM, Sandler HM, Lam KL, Bree RL, Lichter AS, and Ten Haken RK. Measurement of prostate movement over the course of routine radiotherapy using implanted markers. *Int J Radiat Oncol Biol Phys* 1995; 31: 113-118
11. Barnhorn JC, de Boer JCJ and Heijmen BJM. High quality images with a cooled CCD-camera based EPID for setup verification based on implanted radiopaque markers. Proceedings of the 7th international workshop on electronic portal imaging, Vancoucer; 2002; p. 102-103
12. Bel A, Bartelink H, Vijlbrief RE and Lebesque JV. Transfer errors of planning CT to simulator: a possible source of setup inaccuracies? *Radiother Oncol* 1994; 31: 176-180
13. Bel A, van Herk M, Bartelink H and Lebesque JV. A verification procedure to improve patient set-up accuracy using portal images. *Radiother Oncol* 1993; 29: 253-260
14. Bel A, van Herk M, and Lebesque JV. Target margins for random geometrical treatment uncertainties in conformal radiotherapy. *Med Phys* 1996; 23: 1537-1545
15. Bel A, Keus R, Vijlbrief RE and Lebesque JV. Setup deviations in wedged pair irradiation of parotid gland and tonsillar tumors with an electronic portal imaging device. *Radiother Oncol* 1995; 37: 153-159
16. Bel A, Petrascu O, Van de Vondel I, Coppens L, Linthout N, Verellen D and Storme G. A computerized remote table control for fast on-line patient repositioning: Implementation and clinical feasibility. *Med Phys* 2000; 27: 354-358
17. Bel A, Vos PH, Rodrigus PT, Creutzberg CL, Visser AG, Stroom JC and Lebesque JV. High-precision prostate cancer irradiation by clinical application of an offline patient setup correction procedure, using portal imaging. *Int J Radiat Oncol Biol Phys* 1996; 35: 321-332
18. Belshi R, Pontvert D, Rosenwald JC, and Gaboriaud G. Automatic three-dimensional expansion of structures applied to determination of the clinical target volume in conformal radiotherapy. *Int J Radiat Oncol Biol Phys* 1997; 37: 689-696
19. Bentel GC, Marks LB, Hendren K and Brizel DM. Comparison of two head and neck immobilization systems. *Int J Radiat Oncol Biol Phys* 1997; 38: 867-873

20. Bentel GC, Marks LB and Krishnamurthy R. Impact of cradle immobilization on setup reproducibility during external beam radiation therapy for lung cancer. *Int J Radiat Oncol Biol Phys* 1997; 38: 527-531

21. Van den Berg HA, Olofsen-van Acht MJJ, Van Santvoort JPC, Seven M, Quint S and Levendag PC. Definition and validation of a reference target volume in early stage node-positive cervical carcinoma, based on lymphangiograms and CT-scans. *Radiother Oncol* 2000; 54: 163-170

22. Bijhold J, Lebesque JV, Hart AAM and Vijlbrief RE. Maximizing setup accuracy using portal images as applied to a conformal boost technique for prostatic cancer. *Radiother Oncol* 1992; 24: 261-271

23. de Boer JCJ and Heijmen BJM. An optimal off-line set-up correction protocol based on numerical simulations with measured set-up data of 600 prostate cancer patients. In: Abstract book of the XIIIth ICCR. Schlegel W, Bortfeld T, editors. Heidelberg: Springer-Verlag. 2000; 553-555

24. de Boer JCJ and Heijmen BJM. A protocol for the reduction of systematic patient set-up errors with minimal portal imaging workload. *Int J Radiat Oncol Biol Phys* 2001; 50: 1350-1365

25. de Boer JCJ and Heijmen BJM. Reducing patient set-up errors: to minimize workload and maximize safety. 6th biennial ESTRO meeting on physics in clinical radiotherapy, Sevilla, 17-20 Sep 2001. *Radiother Oncol* 2001; 61 (Suppl. 1): S65

26. de Boer JCJ, Heijmen BJM, Pasma KL and Visser AG. Characterisation of a high elbow, fluoroscopic electronic portal imaging device for portal dosimetry. *Phys Med Biol* 2000; 45: 197-216

27. de Boer JCJ, van Sörnsen de Koste JR, Creutzberg CL, Visser AG, Levendag PC and Heijmen BJM. Electronic portal image assisted reduction of systematic setup errors in head and neck irradiation. *Radiother Oncol* 2001; 61: 299-308

28. de Boer JCJ, van Sörnsen de Koste JR, Senan S, Visser AG and Heijmen BJM. Analysis and reduction of 3D systematic and random set-up errors during the simulation and treatment of lung cancer patients with CT-based external beam radiotherapy dose planning. *Int J Radiat Oncol Biol Phys* 2001; 49: 857-868

29. Byhardt RW, Scott C, Sause WT, Emami B, Komaki R, Fisher B, Lee JS and Lawton C. Response, toxicity, failure patterns, and survival in five Radiation Therapy Oncology Group (RTOG) trials of sequential and/or concurrent chemotherapy and radiotherapy for locally advanced non-small-cell carcinoma of the lung. *Int J Radiat Oncol Biol Phys* 1998; 42: 469-478

30. Boellaard R, van Herk M, Uiterwaal H and Mijnheer BJ. First clinical tests using a liquid-filled electronic portal imaging device and a convolution model for the verification of the midplane dose. *Radiother Oncol* 1998; 47: 303-312

31. Boyer AL, Antonuk L, Fenster A, van Herk M, Meertens H, Munro P, Reinstein .E, and Wong J. A review of electronic portal imaging devices (EPIDs). *Med Phys* 1992; 19: 1-16

32. Brahme A, Chavaudra J, Landberg T, *et al*. Accuracy requirements and quality assurance of external beam therapy with photons and electrons. Brahme A, editor. *Acta Oncol* 1988; Suppl. 1

33. Le Chevalier T, Arriagada R, Quoix E, Ruffie P, Martin M, Douillard JY, Tarayre M, Lacombe-Terrier MJ and Laplanche A. Radiotherapy alone versus combined chemotherapy and radiotherapy in unresectable non-small cell lung carcinoma. *Lung Cancer* 1994; 10 (Suppl .1): S239-244

34. Le Chevalier T, Arriagada R, Tarayre, Lacombe-Terrier MJ, Laplanche A, Quoix E, Ruffie P, Martin M and Douillard JY. Significant effect of adjuvant chemotherapy on survival in locally advanced non-small cell lung carcinoma. *J Natl Cancer Inst* 1992; 84: 58

35. Creutzberg CL, Althof VGM, de Hoog MD, Visser AG, Huizenga H, Wijnmaalen A, and Levendag PC. A quality control study of the accuracy of patient positioning in irradiation of pelvic fields. *Int J Radiat Oncol Biol Phys* 1996; 34: 697-708

36. Crook JM, Raymond Y, Salhani D, Yang H, and Esche B. Prostate motion during standard radiotherapy as assessed by fiducial markers. *Radiother Oncol* 1995; 37: 35-42

37. van Dieren EB, Nowak PJ, Wijers OB van Sörnsen de Koste JR, van der Est H, Binnekamp DP, Heijmen BJM and Levendag PC. Beam intensity modulation using tissue compensators or dynamic multileaf collimation in three-dimensional conformal radiotherapy of primary cancers

of the oropharynx and larynx, including the elective neck. *Int J Radiat Oncol Biol Phys* 2000; 47: 1299-1309

38. Dirkx MLP, Kroonwijk M, de Boer JCJ and Heijmen BJM. Daily dosimetric quality control of the MM50 Racetrack Microtron using an electronic portal imaging device *Radiat Oncol* 1995; 37: 55-60

39. Dunscombe PB, Fox K, Loose S and Leszczynski K. The investigation and rectification of field placement errors in the delivery of complex head and neck fields. *Int J Radiat Oncol Biol Phys* 1993; 26: 155-61

40. Ekberg L, Holmberg O, Wittgren L, Bjelkengren G and Landberg T. What margins should be added to the clinical target volume in radiotherapy treatment planning for lung cancer? *Radiother Oncol* 1998; 48:71-77

41. El-Gayed AA, Bel A, Vijlbrief R, Bartelink H and Lebesque JV. Time trend of patient setup deviations during pelvic irradiation using electronic portal imaging. *Radiother Oncol* 1993; 26: 162-171

42. Essers M, Hoogervorst BR, van Herk M, Lanson H and Mijnheer BJ. Dosimetric characteristics of a liquid-filled electronic portal imaging device. *Int J Radiat Oncol Biol Phys* 1995; 33: 1265-1272

43. Essers M and Mijnheer BJ. *In vivo* dosimetry during external photon beam radiotherapy. *Int J Radiat Oncol Biol Phys* 1999; 43: 245-59

44. Evans PM, Hansen VN, Mayles WP, Swindell W, Torr M and Yarnold JR. Design of compensators for breast radiotherapy using electronic portal imaging. *Radiother Oncol* 1995; 37 43-54

45. Gluhchev G. The magnitude of treatment field set-up parameter correction in radiation therapy. *Radiother Oncol* 1998; 48: 79-82

46. Goitein M. Calculation of the uncertainty in the dose delivered during radiation therapy. *Med Phys* 1985; 12: 608-612

47. Goitein M, Busse J. Immobilization error: Some theoretical considerations. *Radiology* 1975; 117: 407-412

48. Halperin R, Roa W, Field M, Hanson J and Murray B. Setup reproducibility in radiation therapy for lung cancer: a comparison between T-bar and expanded foam immobilization devices. *Int J Radiat Oncol Biol Phys* 1999; 43: 211-216

49. Hanley J, Lumley MA, Mageras GS, Sun J, Zelefsky MJ, Leibel SA, Fuks Z and Kutcher GJ. Measurement of patient positioning errors in 3D conformal radiotherapy of the prostate. *Int J Rad Onc Biol Phys* 1997; 37: 435-444

50. Hansen VN, Evans PM and Swindell W. The application of transit dosimetry to precision radiotherapy. *Med Phys* 1996*;* 23: 713-721

51. Hayman JA, Martel MK, Ten Haken RK, Normolle DP, Todd RF 3rd, Littles JF, Sullivan MA, Possert PW, Turrisi AT and Lichter AS. Dose escalation in non-small cell lung cancer using conformal 3-dimensional radiation therapy: Update of a phase I trial. (Abstr.) *J Clin Onc* 1999; 18: 459a (Abstract)

52. Heijmen BJM, Pasma KL, Kroonwijk M, Althof VGM, de Boer JCJ, Visser AG and Huizenga H. Portal dose measurement in radiotherapy using an electronic portal imaging device (EPID). *Phys Med Biol* 1995; 40: 1943-1955

53. van Herk M, Boellaard R, Brugmans M and Van Dalen A. Feasibility of the PortalVision system for verification of dynamic therapy. Proceedings of the 4th International Workshop on Electronic Portal Imaging. 1996; p. 53

54. van Herk M, Bruce A, Kroes AP, Shouman T, Touw A, and Lebesque JV. Quantification of organ motion during conformal radiotherapy of the prostate by three dimensional image registration. *Int J Radiat Oncol Biol Phys* 1995; 33: 1311-1320

55. van Herk M, Remeijer P, Rasch C and Lebesque JV. The probability of correct target dosage-dose population histograms for deriving treatment margins in radiotherapy. *Int J Radiat Oncol Biol Phys* 1999;47:1121-1135

56. Herman MG, Balter JM, D. Jaffray A, McGee KP, Munro P, Shalev S, van Herk M and Wong JW. Clincal use of electronic portal imaging: Report of AAPM Radiation Therapy Committe Task Group 58. *Med Phys* 2001; 28: 712-737

57. Hess CF, Kortmann RD, Jany R, Hamberger A and Bamberg M. Accuracy of field alignment in radiotherapy of head and neck cancer utilizing individualized face mask immobilization: a retrospective analysis of clinical practice. *Radiother Oncol* 1995; 34: 69-72

58. van den Heuvel F, de Neve W, Verellen D, Coghe M, Coen V abd Storme G. Clinical implementation of an objective computer-aided protocol for intervention in intra-treatment correction using electronic portal imaging. *Radiother Oncol* 1995; 35: 232-239

59. Hunt MA, Kutcher GJ, Burman C, Fass D, Harrison L, Leibel S and Fuks Z. The effect of set-up uncertainties on the treatment of nasopharynx cancer. *Int J Radiat Oncol Biol Phys* 1993; 27: 437-447

60. Hurkmans CW, Remeijer P, Lebesque JV and Mijnheer BJ. Set-up verification using portal imaging; review of current clinical practice. *Radiother Oncol* 2001; 58: 105-120

61. Hunt MA, Shultheiss TE, Desobry G, and Hanks GE. Convolving set-up uncertainties with dose distributions. *Med Phys* 1993; 20: 929

62. International Commission on Radiation Units and Measurements. *ICRU report 50: Prescribing, recording, and reporting photon beam therapy*, ICRU Publications, Bethesda MD 1993

63. International Commission on Radiation Units and Measurements. *ICRU Report 62: Prescribing, recording and reporting photon beam therapy*, Bethesda, MD 1999

64. Jaffray DA and Siewerdsen JH. Cone-beam computed tomography with a flat-panel imager: initial performance characterization. *Med Phys* 2000; 27: 1311-1323

65. Kaatee RSJP, Olofsen MJJ, Verstraate M, Quint S, de Boer JCJ, van Santvoort JPC and Heijmen BJM. Detection of movement of the cervix during radiotherapy, using a fluoroscopic EPID and radiopaque markers. *Radiother Oncol* 2001; 61 (Suppl. 1): 204

66. Killoran JH, Kooy HM, Gladstone DJ, Welte FJ, and Beard CJ. A numerical simulation of organ motion and daily setup uncertainties: implications for radiation therapy. *Int J Radiat Oncol Biol Phys* 1997; 37: 213-221

67. Kirby MC The consequences of fixed-pattern noise and image movement on electronic portal images. *Phys Med Biol* 1996; 41: 2345-2356

68. Kirby MC and Williams PC. The use of an electronic portal imaging device for exit dosimetry and quality control measurements. *Int J Radiat Oncol Biol Phys* 1995; 31: 593-603

69. Kroonwijk M, Pasma KL, Quint S, Koper PCM, Visser AG, and Heijmen BJM. *In vivo* dosimetry for prostate cancer patients using an electronic portal imaging device (EPID); detection of internal organ motion. *Radiother Oncol* 1998; 49: 125-32

70. Kubo HD and Hill BC. Respiration gated radiotherapy treatment: a technical study. *Phys Med Biol* 1996; 41: 83-91

71. Lagerwaard F, van Sörnsen de Koste JR, Nijssen-Visser MRJ, Schuchhard-Schipper RH, Oei SS, Munne A and Senan S. Multiple "Slow" CT scans for incorporating Lung tumor mobility in radiotherapy planning. *Int J Radiat Oncol Biol Phys* 2001; 51: 932-937

72. Langmack KA. Portal imaging. *Br J Radiology* 2001; 74: 789-804

73. Lebesque JV, Remeijer P, van Riel V, Vos P, de Boer JCJ, Visser AG, van Lin E, Feenstra S. Clinical Evaluation of Setup Verification and Correction Protocols: Results of Multicentre Studies of the Dutch Cooperative EPID Group. Proceedings of the 5th international workshop on electronic portal imaging, Phoenix AZ 1998; p. 20-23

74. Leong J. Use of digital fluoroscopy as an on-line verification device in radiation therapy. *Phys Med Biol* 1986; 31: 593-603

75. van Lin ENJT, Nijenhuis E, Huizenga H,. van der Vight L and Visser AG. Effectiveness of couch-height based patient set-up and an off-line correction protocol in prostate cancer radiotherapy. *Int J Radiat Oncol Biol Phys* 2001; 50: 569-577

76. Lohr F, Schramm O, Schraube P, Sroka-Perez G, Seeber S, Schlepple G, Schlegel W and Wannenmacher M. Simulation of 3D-treatment plans in head and neck tumors aided by matching of digitally reconstructed radiographs (DRR) and on-line distortion corrected simulator images. *Radiother Oncol* 1997; 45: 199-207

77. Luchka K, Chen D, Shalev S, Gluhchev G and Rajapakshe R. Assessing radiation and light field congruence with a video based electronic portal imaging device. *Med Phys* 1996; 7: 1245-1252

78. Lujan AE, Larsen EW, Balter JM and Ten Haken RK. A method for incorporating organ motion due to breathing into 3D dose calculations. *Med Phys* 1999; 26: 715-720

79. Mageras GS, Fuks Z, Leibel SA Ling CC, Zelefsky MJ, Kooy HM, van Herk M and Kutcher GJ. Computerized design of target margins for treatment uncertainties in conformal radiotherapy. *Int J Radiat Oncol Biol Phys* 1999; 43: 437-445

80. Mageras GS, Kutcher GJ, Leibel SA, Zelefsky MJ, Melian E, Mohan R, and Fuks Z. A method of incorporating organ motion uncertainties into three- dimensional conformal treatment plans. *Int J Radiat Oncol Biol Phys* 1996; 35: 333-342

81. Martel MK, Ten Haken RK, Hazuka MB, Kessler ML, Strawderman M, Turrisi AT, Lawrence TS, Fraass BA and Lichter AS. Estimation of tumor control probability model parameters from 3-D dose distributions of non-small cell lung cancer patients. *Lung Cancer* 1999; 24: 31-37

82. McCurdy BM, Luchka K and Pistorius S. Dosimetric investigation and portal dose image prediction using an amorphous silicon electronic portal imaging device. *Med Phys* 2001; 28: 911-924

83. McNutt TR, Mackie TR, Reckwerdt P and Paliwal BR. Modeling dose distributions from portal dose images using the convolution/superposition method. *Med Phys* 1996; 23: 1381-1392

84. McNutt TR, Mackie TR, Reckwerdt P, Papanikolao N and Paliwal BR. Calculation of portal dose using the convolution/superposition method. *Med Phys* 1996; 23: 527-535

85. Mechelakos JG, Mageras GS, Zelefsky MJ, Lyass O, van Herk M, Kooy HM, Leibel S and Ling CC. Time trends in organ position and volume in patients receiving prostate three-dimensional conformal radiotherapy. *Radiother Oncol* 2002; 62: 261-265

86. Mijnheer BJ, Batterman JJ and Wambersie A. What degree of accuracy is required and can be achieved in photon and neutron therapy ? *Radiother Oncol* 1987; 8: 237-252

87. Mitine C, Dutreix E, van der Schueren E. Black and white in accuracy assessment of megavoltage images: the medical decision is often grey. *Radiother Oncol* 1993; 28: 31-36

88. Morrill SM, Langer M, and Lane RG. Real-time couch compensation for intra-treatment organ motion: theoretical advantages. *Med Phys* 1996; 23: 1083

89. Munro P. Portal Imaging Technology: Past, Present and Future. *Seminars in Radiation Oncology* 1995; 5:115-133

90. Murrer LHP, Kaatee RSJP, de Boer JCJ and Heijmen BJM. Sagging correction for Theraview electronic portal imaging systems: stability analysis and correction procedure. Proceedings of the 6th international workshop on electronic portal imaging, Brussels 2000; Abstract 32

91. De Neve W, Van den Heuvel F, De Beukeleer M, Coghe M, Thon L, De Roover P, Van Lancker M, and Storme G. Routine clinical on-line portal imaging followed by immediate field adjustment using a tele-controlled patient couch. *Radiother Oncol* 1992; 24: 45-54

92. Okunieff, P, Morgan D, Niemierko A, and Suit HD. Radiation dose-response of human tumors. *Int J Radiat Oncol Biol Phys* 1995; 32:1227-1237

93. Olofsen-van Acht MJ, van den Berg H, Quint S, de Boer JCJ, Seven M, van Sörnsen de Koste JR, Creutzberg CL and Visser AG. Reduction of irradiated small bowel volume and accurate patient positioning by use of a bellyboard device in pelvic radiotherapy of gynecological cancer patients. *Radiother Oncol* 2001; 59: 87-93

94. Olofsen-van Acht MJ, Quint S, Seven M, van Santvoort JP, van den Berg HA, , Logmans A, and Levendag PC. Three-dimensional treatment planning for postoperative radiotherapy in patients with node-positive cervical cancer. Comparison between a conventional and a conformal technique. *Strahlenther Onkol* 1999; 175: 462-469

95. Partridge M, Evans PM, van Herk M, Ploeger LS, Budgell GJ and James HV. Leaf position verification during dynamic beam delivery: a comparison of three applications using electronic portal imaging. *Med Phys* 2000; 27: 1601-1609

96. Partridge M, Evans PM, Mosleh-Shirazi A and Convery D. Independent verification using portal imaging of intensity-modulated beam delivery by the dynamic MLC technique. *Med Phys* 1998; 25**:** 1872-1879

97. Pasma KL, Dirkx ML, Kroonwijk M, Visser AG and Heijmen BJM. Dosimetric verification of intensity modulated beams produced with dynamic multileaf collimation using an electronic portal imaging device. *Med Phys* 1999; 26: 2373-2378

98. Pasma KL, Heijmen BJM, Kroonwijk M and Visser AG. Portal dose image (PDI) prediction for dosimetric treatment verification in radiotherapy. I. An algorithm for open beams. *Med Phys* 1998; 25: 830-840

99. Pasma KL, Kroonwijk M, de Boer JCJ, Visser AG and Heijmen BJM. Accurate portal dose measurement with a fluoroscopic electronic portal imaging device (EPID) for open and wedged beams and dynamic multileaf collimation *Phys Med Biol* 1998; 43: 2047-2060

100. Pasma KL, Kroonwijk M, Quint S, Visser AG, and Heijmen BJM. Clinical application of in-vivo dosimetry with a fluoroscopic EPID for prostate cancer patients (N=115) Proceedings of the 5th International Workshop on Electronic Portal Imaging, Phoenix AZ 1998; p. 118-119

101. Pasma KL, Kroonwijk M, van Dieren EB, Visser AG and Heijmen BJM. Verification of compensator thicknesses using a fluoroscopic electronic portal imaging device. *Med Phys* 1999; 26: 1524-1529

102. Pisani L, Lockman D, Jaffray D, Yan D, Martinez A and Wong J. Set-up error in radiotherapy: on-line correction using electronic kilovoltage and megavoltage radiographs. *Int J Radiat Oncol Biol Phys* 2000; 47: 825-839

103. Pouillot J and Lirette A. Verification and correction of setup deviations in tangential breast irradiation using EPID: Gain versus workload. *Med Phys* 1996; 23:1393-1398

104. Press WH, Flannery BP, Teukolsky SA and Vetterling WT. *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1990

105. Quint S, de Boer JCJ, van Sörnsen de Koste JR, Heijmen BJM and Olofsen-van Acht MJ. Set-up verification of cervix cancer patients treated with long treatment fields; implications of a non-rigid bony anatomy. *Radiother Oncol* 2001; 60: 25-29

106. Rajapakshe R, Luchka K and Shalev S. A quality control test for electronic portal imaging devices *Med Phys* 1996; 23: 1237-1244

107. Rimkus D and Bailey NA. Patient exposure requirements for high contrast resolution digital radiographic systems *Am J Roentgenol* 1984; 142: 603-608

108. Robertson JM, Ten Haken RK, Hazuka MB, Turrisi AT, Martel MK, Pu AT, Littles JF, Martinez FJ, Francis IR, Quint LE and Lichter AS. Dose escalation for non-small cell lung cancer using conformal radiation therapy. *Int J Radiat Oncol Biol Phys* 1997; 37: 1079-1085

109. Roeske JC, Forman JD, Mesina CF, He T, Pelizzari CA, Fontenla E, Vijayakumar S and Chen GT. Evaluation of changes in the size and location of the prostate, seminal vesicles, bladder, and rectum during a course of external beam radiation therapy. *Int J Radiat Oncol Biol Phys* 1995; 33: 1321-1329

110. Ross CS, Hussey DH, Pennington EC, Stanford W and Doornbos JF. Analysis of movement of intrathoracic neoplasms using ultrafast computerized tomography. *Int J Radiat Oncol Biol Phys* 1990; 18: 671-677

111. Rudat V, Flentje M, Oetzel D, Menke M, Schlegel W, and Wannenmacher M. Influence of the positioning error on 3D conformal dose distributions during fractionated radiotherapy. *Radiother Oncol* 1994; 33: 56-63

112. Samson MJ, van Sörnsen de Koste JR, de Boer JCJ, Tankink H, Verstraate M, Essers M, Visser AG and Senan S. An analysis of anatomic landmark mobility and setup deviations in radiotherapy for lung cancer. *Int J Radiat Oncol Biol Phys* 1999; 43: 827-832

113. Schewe JE, Balter JM, Lam KL, ten Haken RK. Measurement of patient setup errors using port films and a computer-aided graphical alignment tool. *Med Dosim* 1996; 21: 97-104

114. Senan S, van Sörnsen de Koste J, Samson M, Tankink H, Jansen P, Nowak PJ, Krol AD, Schmitz P and Lagerwaard FJ. Evaluation of a target contouring protocol for 3D conformal radiotherapy in non-small cell lung cancer. *Radiother Oncol* 1999; 53: 247-255

115. Seppenwoolde Y, Shirato H, Kitamura K, Shimizu S, van Herk M, Lebesque JV and Miyasaka K. Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. *Int J Radiat Oncol Biol Phys* 2002; 53: 822-834

116. Shalev S. Treatment Verification Using Digital Imaging in *Radiation Therapy Physics* ed. AR Smith. Springer, Berlin 1995; 155-173.

117. Shalev S and Gluhchev G. Decision rules for the correction of field set-up parameters. *Med Phys* 1994; 21: 952

118. Shalev S, Luchka K and Rajapakshe R Techniques for commissioning electronic portal imaging devices *Proceedings of the XIIth International Conference on the Use of Computers in Radiation Therapy* ed DD Leavitt and G Starkschall. Medical Physics Publishing, Madison WI 1997; 272-275

119. Skladowski K, Tarnawski R, Maciejewski B, Wygoda A, Slosarek K. Clinical radiobiology of glottic T1 squamous cell carcinoma. *Int J Radiat Oncol Biol Phys* 1999; 43: 101-106

120. Song PY, Washington M, Vaida F. Hamilton R, Spelbring D, Wyman B, Harrison J and Chen GT. A comparison of four patient immobilization devices in the treatment of prostate cancer patients with three dimensional conformal radiotherapy. *Int J Radiat Oncol Biol Phys* 1996; 34: 213-219

121. van Sörnsen de Koste JR, Lagerwaard F, Schuchhard-Schipper RH, Nijssen-Visser MR, Voet PW, Oei SS and Senan S. Dosimetric consequences of tumor mobility in radiotherapy of stage I non-small cell lung cancer - an analysis of data generated using 'slow' CT scans. *Radiother Oncol* 2001; 61: 93-99

122. van Sörnsen de Koste JR, de Boer JCJ, Senan S and Heijmen BJM. Reduction of PTVs in lung cancer patients by CT-simulation and use of DRRs in setup verification. Proceedings of the 6th International Workshop on Electronic Portal Imaging, Brussels 2000; Abstract 7

123. Spiegel, MR. Schaum's outline of theory and problems of statistics (Schaum's outline series). 2nd ed., McGraw-Hill, New York 1997

124. Van de Steene J, Van den Heuvel F, Bel A, Verellen D, De Mey J, Noppen M, De Beukeleer M and Storme G. Electronic portal imaging with on-line correction of setup error in thoracic irradiation: clinical evaluation. *Int J Radiat Oncol Biol Phys* 1998; 40: 967-976

125. Stroom JC, de Boer JCJ, Huizenga H, Visser AG. Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability. *Int J Radiat Oncol Biol Phys* 1999; 43: 905-919

126. Stroom JC, de Boer JCJ, Koper PC, Murrer LHP, van Os M, van Kampen P, Jansen PP and Heijmen BJM. Prostate motion measurement and reduction during external beam radiotherapy based on implanted brachytherapy markers. *Radiother Oncol* 2002; 64 (Suppl. 1): S115

127. Stroom JC, Koper PC, Korevaar GA, van Os M, Janssen M, de Boer JCJ, Levendag PC and Heijmen BJM. Internal organ motion in prostate cancer patients treated in prone and supine treatment position. *Radiother Oncol* 1999; 51: 237-248

128. Stroom JC, Korevaar GA, Koper PC, Visser AG, and Heijmen BJM. Multiple two-dimensional versus three-dimensional PTV definition in treatment planning for conformal radiotherapy. *Radiother Oncol* 1998; 47: 297-302

129. Stroom JC, Kroonwijk M, Pasma KL, Koper PC, van Dieren E, Levendag PC and Heijmen BJM. Detection of internal prostate movement using portal images. *Radiother Oncol* 1998; 48: S108

130. Stroom JC, Olofsen-van Acht MJJ, Quint S, Seven M, de Hoog M, Creutzberg CL, de Boer JCJ, and Visser AG. On-line set-up corrections during radiotherapy of patients with gynecologic tumors. *Int J Radiat Oncol Biol Phys* 2000; 46: 499-506

131. Stroom JC and Storchi PR. Automatic calculation of three-dimensional margins around treatment volumes in radiotherapy planning. *Phys Med Biol* 1997; 42: 745-755

132. Stroom JC, Visser AG, de Boer JCJ and Huizenga H. Proceedings of the XIth International Conference on the Use of Computers in Radiation Therapy, Hounsell AR, Wilkinson JM and Williams PC (eds), North Western Medical Physics Department, Christie Hospital, Manchester 1994; 264-265

133. Valicenti RK, Michalski JM, Bosch WR, Gerber R, Graham MV, Cheng A, Purdy JA and Perez CA. Is weekly port filming adequate for verifying patient position in modern radiotherapy? *Int J Radiat Oncol Biol Phys* 1994; 30: 431-438

134. Valicenti RK, Waterman FM, Corn BW and Curran WJ. A prospective, randomized study addressing the need for physical simulation following virtual simulation *Int J Radiat Oncol Biol Phys* 1997; 39: 1131-1135

135. van der Vight L, van Lin E, Nijenhuis E, Huizenga H and Visser AG. Implementation of an off-line portal imaging patient set-up and verification procedure in head and neck irradiation. Proceedings of the 6th international workshop on electronic portal imaging, Brussels 2000; Abstract 6

136. Vigneault E, Pouliot J, Laverdière J, Roy J and Dorion M. Electronical portal imaging device detection of radio-opaque markers for the evaluation of prostate position during megavoltage irradiation: a clinical study. *Int J Radiat Oncol Biol Phys* 1997; 37: 205-212

137. Visser AG, van Lin E and van der Vight L. Fiducial markers in the prostate: prostate motion and how to correct for it. Proceedings of the 7th international workshop on electronic portal imaging, Vancoucer 2002; 82-83

138. Vos PH, Bel A, Poortmans PMP, Vlaun V, Rodrigus PTR, Lebesque JV. Set-up accuracy for pelvic fields after application of a 3-D correction procedure using digital portal imaging. *Int J Radiat Oncol Biol Phys* 1994; 30 (Suppl. 1): 229

139. Vos P, van Riel V, de Winter K. 3D Measurement and correction of systematic patient setup errors in external irradiation of patients with a brain tumor, using digital portal imaging. Abstract book of the XIIth ICCR, Leavitt DD and Starkschall G (eds), Medical Physics publishing 1997; 223-226

140. Weltens C, Kesteloot K, Vandevelde G and Van den Bogaert W. Comparison of plastic and Orfit masks for patient head fixation during radiotherapy: precision and costs. *Int J Radiat Oncol Biol Phys* 1995; 33: 499-507

141. Wong JW, Sharpe MB, Jaffray DA, Kini VR, Robertson JM, Stromberg JS, and Martinez AA. The use of active breathing control (ABC) to reduce margin for breathing motion. *Int J Radiat Oncol Biol Phys* 1999; 44: 911-919

142. Wong WJ, Slessinger ED, Hermes RE, Offutt CJ, Roy T and Vannier MW. Portal dose images I: Quantitative treatment plan verification *Int J Radiat Oncol Biol Phys* 1990; 18: 43-50

143. Wowk B and Shalev S. Grooved phosphor screens for on-line portal imaging *Med Phys* 1993; 20: 1641-1651

144. Yan D, Vicini F, Wong J, and Martinez A. Adaptive radiation therapy. *Phys Med Biol* 1997; 42: 123-132

145. Yan D, Wong J, Vicini F, Michalski J, Pan C, Frazier A, Horwitz E and Martinez A. Adaptive modification of treatment planning to minimize the deleterious effects of treatment setup errors. *Int J Radiat Oncol Biol Phys* 1997; 38: 197-206

146. Yan D, Ziaja E, Jaffray D, Wong J, Brabbins D, Vicini F and Martinez A. The use of adaptive radiation therapy to reduce setup error: a prospective clinical study. *Int J Radiat Oncol Biol Phys* 1998; 41: 715-720

147. Yin FF, Schell MC and Rubin P. Input/output characteristics of a matrix ion-chamber electronic portal imaging device. *Med Phys* 1994; 21: 1447-1454.

148. Ying X, Geer LY and Wong JW. Portal dose images II: patient dose estimation. *Int J Radiat Oncol Biol Phys* 1990; 18: 1465-14

List of publications

Althof VGM, de Boer JCJ, Huizenga H, Stroom JC, Visser AG and Swanenburg BN. Physical characteristics of a commercial electronic portal imaging device. *Med Phys* 1996; 23: 1845-1854

de Boer JCJ and Heijmen BJM. A protocol for the reduction of systematic patient set-up errors with minimal portal imaging workload. *Int J Radiat Oncol Biol Phys* 2001; 50: 1350-1365

de Boer JCJ and Heijmen BJM. A new approach to off-line setup corrections: combining safety with minimum workload. *Med Phys* 2002*; 29: 1998-2012

de Boer JCJ, Heijmen BJM, Pasma KL and Visser AG. Characterisation of a high elbow, fluoroscopic electronic portal imaging device for portal dosimetry. *Phys Med Biol* 2000; 45: 197-216

de Boer JCJ, van Sörnsen de Koste JR, Creutzberg CL, Visser AG, Levendag PC and Heijmen BJM. Electronic portal image assisted reduction of systematic setup errors in head and neck irradiation. *Radiother Oncol* 2001; 61: 299-308

de Boer JCJ, van Sörnsen de Koste JR, Senan S, Visser AG and Heijmen BJM. Analysis and reduction of 3D systematic and random set-up errors during the simulation and treatment of lung cancer patients with CT-based external beam radiotherapy dose planning. *Int J Radiat Oncol Biol Phys* 2001; 49: 857-868

Dirkx MLP, Kroonwijk M, de Boer JCJ and Heijmen BJM. Daily dosimetric quality control of the MM50 Racetrack Microtron using an electronic portal imaging device *Radiat Oncol* 1995; 37: 55-60

Heijmen BJM, Pasma KL, Kroonwijk M, Althof VGM, de Boer JCJ, Visser AG and Huizenga H. Portal dose measurement in radiotherapy using an electronic portal imaging device (EPID). *Phys Med Biol* 1995; 40: 1943-1955

Olofsen-van Acht MJ, van den Berg H, Quint S, de Boer JCJ, Seven M, van Sörnsen de Koste JR, Creutzberg CL and Visser AG. Reduction of irradiated small bowel volume and accurate patient positioning by use of a bellyboard device in pelvic radiotherapy of gynecological cancer patients. *Radiother Oncol* 2001; 59: 87-93

Pasma KL, Kroonwijk M, de Boer JCJ, Visser AG and Heijmen BJM. Accurate portal dose measurement with a fluoroscopic electronic portal imaging device (EPID) for open and wedged beams and dynamic multileaf collimation *Phys Med Biol* 1998; 43: 2047-2060

Quint S, de Boer JCJ, van Sörnsen de Koste JR, Heijmen BJM and Olofsen-van Acht MJ. Set-up verification of cervix cancer patients treated with long treatment fields; implications of a non-rigid bony anatomy. *Radiother Oncol* 2001; 60: 25-29

Samson MJ, van Sörnsen de Koste JR, de Boer JCJ, Tankink H, Verstraate M, Essers M, Visser AG and Senan S. An analysis of anatomic landmark mobility and setup deviations in radiotherapy for lung cancer. *Int J Radiat Oncol Biol Phys* 1999; 43: 827-832

van Sörnsen de Koste JR, de Boer JCJ, Schuchhard-Schipper RH, Senan S and Heijmen BJM. Procedures for high precision setup verification and correction of lung cancer patients using CT-simulation and digitally reconstructed radiographs (DRR). *Int J Radiat Oncol Biol Phys* 2002 (in press)

van Sörnsen de Koste JR , Lagerwaard FJ, de Boer JCJ, Nijssen-Vissera MRJ and Senan S. Are multiple CT scans required for planning curative radiotherapy in lung tumors of the lower lobe? *Int J Radiat Oncol Biol Phys* (accepted for publication)

Stroom JC, de Boer JCJ, Huizenga H and Visser AG. Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability. *Int J Radiat Oncol Biol Phys* 1999; 43: 905-919

Stroom JC, Koper PC, Korevaar GA, van Os M, Janssen M, de Boer JCJ, Levendag PC and Heijmen BJM. Internal organ motion in prostate cancer patients treated in prone and supine treatment position. *Radiother Oncol* 1999; 51: 237-248

Stroom JC, Olofsen-van Acht MJJ, Quint S, Seven M, de Hoog M, Creutzberg CL, de Boer JCJ and Visser AG. On-line set-up corrections during radiotherapy of patients with gynecologic tumors. *Int J Radiat Oncol Biol Phys* 2000; 46: 499-506

Samenvatting

For a summary in English, see section 1.5

Dit proefschrift behandelt de theoretische onderbouwing en klinische toepassing van efficiënte methoden ter bewaking van de kwaliteit van radiotherapeutische externe bundel bestralingen met behulp van 'Electronic Portal Imaging Devices' (EPIDs). Deze methoden zijn ontwikkeld vanuit de gedachte dat ze, naast effectief, ook relatief simpel te implementeren moeten zijn en dat de toegevoegde werklast op de bestralingstoestellen klein is. Een belangrijke reden achter deze criteria is dat de acceptatie van het gebruik van EPIDs in de dagelijkse praktijk op een systematische, geprotocolleerde wijze, sterk achter is gebleven bij initiële verwachtingen.

Met EPIDs worden afbeeldingen gemaakt van de bestralingsbundel waarin de doorstraalde patiëntanatomie, zoals geprojecteerd door deze bundel, zichtbaar wordt. Op deze manier kunnen positioneringsfouten (afwijkingen ten opzichte van de geplande situatie) van de patiënt ten opzichte van de bundel worden gemeten. Daarbij blijkt het essentieel te zijn deze fouten te scheiden in systematische fouten (die gedurende iedere bestraling van een bepaalde patiënt optreden) en willekeurige of 'random' fouten (die voor iedere bestralingssessie ('fractie') van een bepaalde patiënt anders kunnen zijn). Met name de reductie van systematische fouten blijkt een efficiënte methode om onderdosering van de tumor te voorkomen.

Door de intensiteiten in de EPID afbeeldingen te analyseren kan ook de bestralingsdosis die in de patiënt is afgegeven worden gecontroleerd. Met een toenemende complexiteit van de technieken die gebruikt worden om de juiste driedimensionale dosisverdeling in de patiënt te verkrijgen wordt een dergelijke controle steeds belangrijker.

Hieronder volgt een beknopt overzicht van hoe bovenstaande onderwerpen in de diverse hoofdstukken aan bod komen.

Hoofdstuk 1. Introductie
De scheiding van positioneringsfouten in systematische en random fouten wordt in detail uitgelegd. De beschrijving van de distributie van deze fouten in een populatie van patiënten, essentieel voor latere hoofdstukken, wordt toegelicht.

Hoofdstuk 2. Beschrijving van de TNT EPID en dosimetrische toepassingen
In dit hoofdstuk geven we een technische omschrijving van de EPID die in al het beschreven werk is toegepast. Deze EPID is in onze kliniek ontwikkeld, in eerste instantie als prototype in het kader van een NKB project, en momenteel in een samenwerking met Cablon Medical (Leusden, Nederland). Alhoewel de techniek zich sterk heeft ontwikkeld sinds het prototype dat in dit hoofdstuk wordt beschreven, zijn belangrijke elementen (zoals het gebruik van een CCD camera als lichtdetector) behouden gebleven. Tevens is de beschreven methode om EPID beelden om te rekenen naar 2D transmissiedosisverdelingen onveranderd. Deze methode, die kan worden toegepast op basis van een vrij beperkte hoeveelheid calibratiemetingen, is specifiek voor EPID systemen waarin de afstand tussen de fluorescerende laag en spiegel (Fig. 1.1) groter is dan in conventionele EPIDs.

Hoofdstuk 3. Relatie tussen positioneringsfouten en planningmarges

In de radiotherapie wordt het tumorvolume dat bestraald moet worden met een bepaalde minimum dosis (clinical target volume of CTV) bepaald door de radiotherapeut in bijvoorbeeld een CT scan. Echter, door positioneringsfouten tijdens de bestraling kan dit volume verplaatst zijn ten opzichte van de bestralingsbundel waardoor onderdoseringen kunnen optreden. Om dit te voorkomen, wordt het CTV meestal uitgebreid met een bepaalde marge naar het planning target volume (PTV) en tijdens de computerplanning van de bestraling wordt zorg gedragen dat dit uitgebreide volume met de vereiste minimumdosis wordt bestraald.

Hoe groot de uitbreidingsmarge moet zijn als de verdeling van systematische en random fouten in een populatie van patiënten bekend is, is het onderwerp van hoofdstuk 3. Hiertoe wordt het zogenaamde 'coverage probability' (CP) model gebruikt, waarin de waarschijnlijkheden dat een bundel de tumor juist bedekt worden berekend aan de hand van eerdergenoemde verdelingen van de positioneringsfouten. Op deze manier kan worden geëvalueerd wat, uitgaande van een bepaalde marge rond het CTV en een gegeven dosisverdeling in het zo gecreëerde PTV, de dosisverdeling in het CTV, gewogen over alle mogelijke positioneringsvariaties, zal worden.

Door bepaalde voorschriften op te leggen aan deze gemiddelde dosisverdeling kunnen toepasselijke marges worden uitgedrukt in te grootte van de systematische en random fouten in de populatie. Daarbij blijkt dat systematische fouten leiden tot zo'n 3 maal grotere marges dan random fouten van dezelfde grootte. Het is daarom vooral van belang om systematische fouten zoveel mogelijk te beperken, opdat de bestraalde volumina zo klein mogelijk kunnen blijven en anderzijds de kans op onderdoseringen kan worden geminimaliseerd. Naast een dergelijk marge 'recept' voor een populatie, kan het CP model ook worden gebruikt om op individuele basis veilige uitbreidingsmarges af te leiden.

Hoofdstukken 4 en 5: onnauwkeurigheden optredend tijdens het vastleggen van de referentiepositionering voor bestraling van longkankerpatiënten en effectiviteit van positioneringscorrecties tijdens de bestraling

In hoofdstuk 4 wordt ingegaan op de nauwkeurigheid waarmee het tijdens de computerplanning vastgestelde bestralingsplan wordt vastgelegd tijdens de zogenaamde bestralingssimulatie, toegespitst op de situatie voor longkanker patiënten. Tijdens de simulatie wordt de bestraling nagebootst en worden markeringen geplaatst op de patiënt of op patiënt-fixatiemiddelen om deze situatie vast te leggen en later te reproduceren op het bestralingstoestel. Daarbij worden doorlichtingbeelden gemaakt (simulatiefilms) die tijdens de bestraling vaak worden gebruikt als referentiebeelden ter vergelijking met EPID beelden. Dit betekent dat een positioneringsfout tijdens de simulatie een systematische fout is, die tijdens iedere fractie van een gegeven patiënt optreedt, en bovendien met simulatiefilms als referentiebeelden niet zal worden opgemerkt. Om deze simulatiefouten te kwantificeren werden simulatiefilms vergeleken met digitally reconstructed radiographs (DRRs). DRRs zijn doorlichtingbeelden die worden berekend uit de CT data in combinatie met het daarop gebaseerde bestralingsplan en geven daarmee per definitie de geplande positionering perfect weer. Door de positie van anatomische structuren in de DRRs te vergelijken met die in de corresponderende simulatiefilms konden de simulatiefouten worden bepaald.

Omdat deze fouten groot bleken te zijn en, zoals genoemd, systematisch van aard, wordt in hoofdstuk 5 een procedure uitgewerkt waarbij de simulatie wordt overgeslagen en de patiëntmarkering benodigd voor de positionering tijdens de bestraling direct op de CT-scanner plaats vindt. In hoofdstukken 4 en 5 wordt bovendien het effect beschreven van

positioneringscorrecties aan de hand van EPID beelden in een zogenaamd 'off-line' correctie protocol. Off-line betekent dat positioneringsfouten gemeten in EPID beelden tijdens een aantal bestralingssessies worden gebruikt om een correctie voor volgende sessies te bepalen. Dientengevolge kunnen hiermee alleen systematische positioneringsfouten worden gereduceerd. Uit deze studies blijkt dat de benodigde correcties nauwkeurig kunnen worden bepaald op basis van zowel simulatiefilms als DRRs als referentiebeelden. Omdat om bovengenoemde redenen de simulatiefilms aanzienlijke systematische fouten kunnen bevatten kan de procedure beschreven in hoofdstuk 5 gecombineerd met correcties gebaseerd op DRRs en EPID beelden worden toegepast om systematische fouten effectief te verkleinen.

Hoofdstukken 6 en 7. Een methode om systematische positioneringsfouten nauwkeurig te corrigeren met weinig werklast.
In hoofdstuk 6 beschrijven we een nieuw off-line protocol om systematische fouten te reduceren. In dit 'No Action Level' (NAL) protocol wordt eerst de gemiddelde positioneringsfout gedurende een klein aantal fracties (bijv. 3) vroeg in de behandeling bepaald. Gedurende deze fracties vinden geen correcties plaats, maar in de volgende fracties wordt een correctie uitgevoerd ter grootte van de fout bepaald in de eerste fracties. Het NAL protocol werd retrospectief toegepast op met EPIDs gemeten positioneringsfouten in 600 prostaatkankerpatiënten. Daarnaast werd gebruik gemaakt van Monte Carlo simulaties en analytische afleidingen om het gedrag van het NAL protocol nauwkeurig te bestuderen. De conclusie van dit werk is dat met het NAL protocol dezelfde reductie in systematische fouten bereikt kan worden als met het meest toegepaste protocol van dit moment, het 'shrinking action level' (SAL) protocol, maar met een factor 3 maal minder EPID gerelateerde werklast.

   In hoofdstuk 7 wordt dit protocol uitgebreid met een extra meting gericht op het verifiëren van de juiste uitvoering van de positioneringscorrectie berekend in het NAL protocol (correction verification ofwel COVER). Deze uitbreiding is nuttig omdat in het NAL protocol slechts één maal een correctie wordt berekend, waarna niet meer wordt afgebeeld. Om te voorkomen dat er een systematische fout wordt gemaakt in de uitvoering van deze correctie gedurende alle volgende fracties, werd een methode bedacht om significante uitvoeringsfouten met weinig metingen te detecteren. Uit een risicoanalyse gebaseerd op Monte Carlo simulaties werd een testcriterium afgeleid voor de positioneringsfout gemeten in de eerste fractie waarin de positioneringscorrectie wordt uitgevoerd. Met dit criterium hoeft, indien geen systematische fouten worden gemaakt in de uitvoering, voor 95% van de patiënten niet verder te worden gemeten, terwijl grote systematische afwijkingen toch worden gedetecteerd. Daarmee werd het mogelijk om een verificatiestap toe te voegen die de lage werklast van het NAL protocol niet noemenswaardig verhoogt.

Hoofdstuk 8: Positioneringscorrecties voor hoofd-hals patiënten.
In hoofdstuk 8 wordt een studie beschreven naar de positioneringsnauwkeurigheid van hoofd-hals patiënten, gefixeerd met een conventioneel masker en gecorrigeerd met het SAL correctie protocol. Daarnaast wordt de effectiviteit van het NAL protocol voor deze patiëntgroep geanalyseerd, en wordt bevestigd dat een factor 3 werklast reductie mogelijk is ten opzichte van SAL. Daarbij wordt vastgesteld dat de wekelijkse verificatiemetingen, zoals in het algemeen toegepast in combinatie met SAL, in deze patiëntgroep geen ander

effect lijken te sorteren dan compensatie van het gebrek aan correcties in het eerste deel van het SAL protocol.

Hoofdstuk 9: Positioneringsproblemen als gevolg van een flexibele anatomie
Het meeste werk op het gebied van positioneringscorrecties is gericht op het uitvoeren van correcties door verschuivingen van de behandeltafel (Fig. 1.1). Deze aanpak werkt zolang relatieve veranderingen binnen de patiëntanatomie geen rol spelen. Dit is echter niet vanzelfsprekend zodra vervormingen of rotaties van belang worden en dit belang neemt toe naarmate de behandelde volumina groter worden. De meeste modellen trachten het effect van rotaties in de 2D projecties van EPID beelden te beschrijven aan de hand van een rigide patiënt anatomie. De aanname van dergelijke rigiditeit lijkt echter inconsistent met het optreden van diezelfde rotaties. Om dit probleem te illustreren wordt in hoofdstuk 9 onderzocht of positioneringsfouten gemeten aan de hand van benige structuren in het caudale deel van grote cervix bestralingsvelden overeenkomen met de simultaan gemeten fouten in het craniale deel van het veld. Dit blijkt sterk afhankelijk van de keuze van de structuren waarmee deze fouten worden gemeten, en in het algemeen zijn de afwijkingen tussen caudaal en craniaal gelegen structuren van dezelfde orde als de positioneringsfouten van de caudale structuren.

Hoofdstuk 10: Discussie
De discussie spitst zich toe op twee onderwerpen. Ten eerste worden enkele redenen achter de tegenvallende verspreiding van klinisch gebruik van EPIDs besproken, waarbij tot de conclusie wordt gekomen dat niet zozeer de beeldkwaliteit maar eerder de gebrekkige software een belangrijke rol speelt. Ten tweede wordt aandacht geschonken aan zogenaamde time trends: geleidelijke, systematische verplaatsingen van de patiënt gedurende de totale bestralingsperiode. Uit deze discussie wordt geconcludeerd dat er weinig concreet gerapporteerd is over time trends en dat de gerapporteerde trends ofwel voorkomen kunnen worden ofwel klein zijn. Deze discussie wordt met name toegevoegd omdat time trends van belang zijn voor de effectiviteit van het NAL protocol. In dit kader worden de resultaten behaald met het NAL protocol bij prostaatpatiënten in een prospectieve studie gepresenteerd.

Dankwoord

De artikelen zijn geworden tot een proefschrift en dit wordt, zoals gebruikelijk, uitgebracht onder één naam. Maar, zoals de lijsten van (co)auteurs bij de diverse hoofdstukken al aangeven zijn er vele anderen bij betrokken geweest en dit is het moment minstens een aantal daarvan nog eens bij naam te noemen.

Allereerst copromotor en onderzoeksmotor, Ben Heijmen. Zijn vermogen om dwars door alle ruis de zin en onzin van onderzoeksresultaten te doorzien, de tijd die hij altijd nam voor langdurige en intensieve discussies (naast de vele prettige gesprekken over de meest uiteenlopende onderwerpen buiten het werk), de bereidheid mijn artikelen zeer grondig te lezen en van commentaar en verbeteringen te voorzien en zijn voortdurende motivatie tot het doen van onderzoek zijn niet alleen van groot belang geweest voor het tot stand komen van dit proefschrift, maar vormen ook elementen die het dagelijkse werk tot een plezier maken. Ik hoop dan ook dat we nog lange tijd zullen kunnen samenwerken.

Maar uiteraard ben ik vele anderen dank verschuldigd. Ten eerste de EPID-fysici van het eerste uur, Henk Huizenga en Andries Visser, die in mij het vertrouwen stelden en de middelen aanreikten om 'van scratch' een nieuwe EPID te ontwikkelen, een project waar 'RAPID' (zie hoofdstuk 2) het gevolg van was. Daarbij moet ook Boudewijn Swanenburg worden genoemd, indertijd directeur van de Stichting Ruimteonderzoek (SRON)-Leiden. Dankzij hem kwam ik in contact met de afdeling klinische fysica van de Daniel den Hoed Kliniek, een groep waar hij mee had samengewerkt bij de ontwikkeling van de Philips SRI-100 EPID. Ook zijn latere inbreng in de werkgroep ter ontwikkeling van RAPID was bijzonder waardevol.

 Ten tweede is daar de groep 'MM50 laboranten' die er al snel voor zorgden dat ik me thuis voelde in 'de Daniel' dankzij hun collegialiteit en enthousiasme. Hun inzet was altijd heel stimulerend. Dankzij Gert, Sandra Quint, Merik, Henri, de Marjoleinen, Petra en nog vele anderen werd de toepassing van EPIDs een succes in onze kliniek. De enorme database met positioneringsfouten van honderden patiënten, waar ik voor dit proefschrift gebruik van kon maken, zou er zonder hen nooit zijn gekomen en al zeker niet van dezelfde kwaliteit zijn geweest.

John van Sörnsen de Koste was de laborant die samen met mij het NKB project uitvoerde waar het meeste werk uit dit proefschrift op gebaseerd is. Uit diverse hoofdstukken wordt zijn belangrijke bijdrage duidelijk, en zijn uitputtende analyses van duizenden EPID beelden zijn onder insiders legendarisch. Ik heb altijd de neiging gehad resultaten, aangeleverd door anderen, niet zomaar over te nemen zonder er zelf goed naar gekeken te hebben. Bij John durfde ik dat principe vaak los te laten omdat hij altijd goed voorbereid, precies en systematisch werkte, ook als er tijdsdruk was.

Een andere essentiële groep vormen de programmeurs en automatiseerders. Het prototype van RAPID had nooit tijdig in klinisch gebruik genomen kunnen worden zonder de inspanningen van Bert van der Leije, die bij het hele ontwikkeltraject betrokken was. Hij bouwde eigenhandig en binnen enkele maanden een compleet nieuwe EPID user-interface en wist daarmee ook van Visual C++ turbotaal te maken. Gelukkig werkt hij nog steeds mee aan de software voor de opvolger van RAPID, het TNT systeem, dat we in samenwerking met Cablon Medical ontwikkelen. Op die basis zijn Paul en Jaco erbij gekomen, die er niet alleen voor hebben gezorgd dat we het aandurfden de software op een volledig nieuwe basis op te zetten maar die ook nieuwe technieken en ideeën inbrachten.

Het blijft inspirerend om met dit groepje te werken. Bart, Hafid, Norman (en weer Bert) wil ik bedanken voor de snelle ondersteuning bij computerproblemen, en de vrijheid die ze lieten om eigen experimenten op de PC te ondernemen.

Uiteraard zijn er ook de klinisch fysici (in opleiding) en projectmedewerkers met wie het niet alleen prettig omgaan was maar die ook een belangrijke inbreng hebben gehad in dit werk. Ten eerste Joep Stroom, met wie ik vooral tijdens de eerste jaren vaak uitgebreid ging lunchen wat leidde tot de meest uiteenlopende gesprekken. Soms leidden die tot slappe lachaanvallen, maar ook tot stevige discussies over wat te doen met al die onnauwkeurigheden in de radiotherapie, met name over de concepten die hij in hoofdstuk 3 van dit proefschrift beschrijft. Ook Kasper en Marco (lange tijd kamergenoten), Maarten, Sandra Vieira, Lars en Robert hebben altijd voor een aangename sfeer met veel discussie over EPIDs en de rest van de wereld gezorgd. Van Marco heb ik een groot aantal nuttige IDL routines kunnen overnemen, waar ik op kon bouwen voor het werk in hoofdstuk 2. Lars en Robert wil ik ook bedanken voor het grotendeels overnemen van de klinische EPID-ondersteuning waardoor ik de ruimte kreeg dit boekje eindelijk af te schrijven.

Uiteraard was veel van het werk uit dit proefschrift alleen mogelijk dankzij de betrokkenheid van radiotherapeuten die het nut van positioneringsverbetering met EPIDs nadrukkelijk erkenden. Manouk Olofsen, Carien Creutzberg en Suresh Senan wil ik hartelijk danken voor hun actieve betrokkenheid bij diverse studies en bijdrage aan het stimuleren van systematisch EPID gebruik. Daarnaast heeft mijn promotor, Peter Levendag, regelmatig zijn enthousiasme over het EPID-werk in onze kliniek getoond, evenals over het onderzoek in dit boekje. Ik wil hem en de overige leden van de promotiecommissie bedanken voor het doornemen van dit proefschrift. Dat geldt ook voor de leden die nog niet zijn vermeld: Ben Mijnheer, Pieter Vos, Vincent Grégoire and Phil Evans, thank you very much for coming over to Rotterdam to take part in the opposition.

Ik doe tekort aan veel mensen als ik niet de overige klinisch fysici, klinisch fysisch medewerkers, laboranten, artsen, instrumentmakers, electronici, secretaressen, etc., zou bedanken, niet alleen voor eventuele bijdragen aan dit boekje (want die zijn er zeker), maar ook voor het samen maken van de 'Rotterdamse sfeer' waarin ik het prettig werken vind.

Maar, er is meer dan collega's. Mijn vader was door zijn nieuwsgierigheid, technische inslag en creativiteit een inspiratiebron en ik zal hem op de dag van mijn promotie bijzonder missen. Daar staat tegenover dat mijn moeder en twee zussen onvermoeibaar belangstelling en vertrouwen tonen in wat ik doe, en ook veel gezelligheid brengen. Christophe en Karin bleven trouwe vrienden, ook toen ik steeds minder tijd kreeg om uitgebreid dingen te ondernemen. Ook Homme bleef dat tot het laatste moment. Zijn afwezigheid zal ik zeker blijven voelen.

Stephen, Tanya and Julius, thank you for being great friends in the land down under.

Om Mariëtte hier te bedanken is eigenlijk te simpel. In de periode waarin dit boekje tot stand kwam, en waarin ik tevens de opleiding tot klinisch fysicus afrondde, werden ook Niels en Veerle geboren. Dankzij Mariëtte draaide alles zonder probleem door, ook op haar werk, en daarom weet ik niet zeker wie nu eigenlijk het meeste gedaan heeft voor dit proefschrift. Wat ik wel weet is dat Niels en Veerle voor ons beiden een enorme bron van energie en plezier zijn, die maken dat het ondanks de drukte toch nooit echt zwaar wordt.

Curriculum Vitae

Hans de Boer werd geboren op 8 april 1965, te Leiden. Aldaar behaalde hij in 1983 zijn VWO diploma aan het St. Agnes College. Vervolgens studeerde hij sterrenkunde aan de Rijksuniversiteit Leiden, waar hij in maart 1989 zijn doctoraalexamen behaalde op onderzoek aan het magneto-hydrodynamisch evenwicht in onze melkweg (de evenwichtstoestand tussen het interstellaire gas en de daaraan gekoppelde magneetvelden en kosmische straling). Aansluitend werkte hij tot februari 1993 bij de Stichting Ruimteonderzoek Nederland (SRON) als NWO projectmedewerker aan de ontwikkeling van software voor calibratie, simulatie en data-analyse voor de ruimtetelescoop COMPTEL, een instrument dat hoog energetische fotonenstraling (0.5-30 MeV) kon afbeelden. Om bewust eens een ander werkveld te proberen trad hij in april 1993 in dienst van de Daniel den Hoed Kliniek, waar hij zich in eerste instantie toelegde op het klinische toepasbaar maken van prototype EPID apparatuur die eerder binnen de Daniel den Hoed in samenwerking met SRON en Philips Medical Systems was ontwikkeld. Daarnaast analyseerde hij de fysische eigenschappen van de toenmalige EPIDs om daarmee in de periode 1995-1997 een nieuw, verbeterd EPID systeem te ontwikkelen ("RAPID") en begeleidde hij de technische en klinische introductie van RAPID. Gelijktijdig begon hij in 1996 aan een NKB project waarbinnen de toepassingen die met RAPID binnen bereik kwamen werden onderzocht, wat leidde tot dit proefschrift. Tevens werd hij in 1999 projectleider binnen de samenwerking die met Cablon Medical werd aangegaan om RAPID tot een commercieel verkrijgbaar product te maken en technologisch verder te ontwikkelen. Gedurende het NKB project startte hij ook de opleiding tot klinisch fysicus, die hij in november 2002 heeft afgerond.

Hij is getrouwd met Mariëtte Sanderse en de vader van Niels (1999) en Veerle (2001).