UNIVERSITY OF HELSINKI REPORT SERIES IN PHYSICS

HU-P-D163

ACCURACY OF ABSORBED DOSE IN EXTERNAL PHOTON BEAM RADIOTHERAPY: what level is sufficient and how to approach it?

Mika Kapanen

Department of Oncology
Helsinki University Central Hospital
Helsinki, Finland

Radiation in Health Care
Radiation and Nuclear Safety Authority
Helsinki, Finland

Department of Physics
Faculty of Science
University of Helsinki
Helsinki, Finland

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public criticism in Auditorium D101, the Department of Physics (Physicum), Gustaf Hällströmin katu 2, on June 18th, 2009, at 12 o'clock noon.

Helsinki 2009

Supervised by:

Docent Mikko Tenhunen
Department of Oncology
Helsinki University Central Hospital
Finland

Professor Juhani Keinonen
Department of Physics
University of Helsinki
Finland

Reviewed by:

Docent Tapani Lahtinen
Department of Oncology
Kuopio University Hospital
Finland

Docent Maunu Pitkänen
Department of Oncology
Tampere University Hospital
Finland

Opponent:

Docent Simo Hyödynmaa
Department of Oncology
Tampere University Hospital
Finland

Report Series in Physics HU-P-D163
ISBN 978-952-10-4237-9 (printed version)
ISSN 0356-0961
Helsinki University Print
Helsinki 2009

ISBN 978-952-10-4238-6 (PDF version)
available at http://ethesis.helsinki.fi/
Helsinki University E-thesis
Helsinki 2009

M. Kapanen: Accuracy of absorbed dose in external photon beam radiotherapy: what level is sufficient and how to approach it?, University of Helsinki, 2009, 76 pp. + appendices.
University of Helsinki, Report Series in Physics, HU-P-D163.
Classification (INSPEC): A8770H, B7520C, A8760J, A8760M, A8750G, C1330.
Key words: radiotherapy, photon, linear accelerator, quality assurance, quality control, dosimetry, dose calculation, dose response, salivary gland, patient set-up.

ABSTRACT

Radiation therapy (RT) plays currently significant role in curative treatments of several cancers. External beam RT is carried out mostly by using megavoltage beams of linear accelerators. Tumor eradication and normal tissue complications correlate to dose absorbed in tissues. Normally this dependence is steep and it is crucial that actual dose within patient accurately correspond to the planned dose. All factors in a RT procedure contain uncertainties requiring strict quality assurance.

From hospital physicist´s point of a view, technical quality control (QC), dose calculations and methods for verification of correct treatment location are the most important subjects. Most important factor in technical QC is the verification that radiation production of an accelerator, called output, is within narrow acceptable limits. The output measurements are carried out according to a locally chosen dosimetric QC program defining measurement time interval and action levels. Dose calculation algorithms need to be configured for the accelerators by using measured beam data. The uncertainty of such data sets limits for best achievable calculation accuracy. All these dosimetric measurements require good experience, are workful, take up resources needed for treatments and are prone to several random and systematic sources of errors. Appropriate verification of treatment location is more important in intensity modulated radiation therapy (IMRT) than in conventional RT. This is due to steep dose gradients produced within or close to healthy tissues locating only a few millimetres from the targeted volume.

The thesis was concentrated in investigation of the quality of dosimetric measurements, the efficacy of dosimetric QC programs, the verification of measured beam data and the effect of positional errors on the dose received by the major salivary glands in head and neck IMRT. A method was developed for the estimation of the effect of the use of different dosimetric QC programs on the overall uncertainty of dose. Data were provided to facilitate the choice of a sufficient QC program. The method takes into account local output stability and reproducibility of the dosimetric QC measurements. A method based on the model fitting of the results of the QC measurements was proposed for the estimation of both of these factors. The reduction of random measurement errors and optimization of QC procedure were also investigated. A method and suggestions were presented for these purposes. The accuracy of beam data was evaluated in Finnish RT centres. Sufficient accuracy level was estimated for the beam data. A method based on the use of reference beam data was developed for the QC of beam data. Dosimetric and geometric accuracy requirements were evaluated for head and neck IMRT when function of the major salivary glands is intended to be spared. These criteria are based on the dose response obtained for the glands.

Random measurement errors could be reduced enabling lowering of action levels and prolongation of measurement time interval from 1 month to even 6 months simultaneously maintaining dose accuracy. The combined effect of the proposed methods, suggestions and criteria was found to facilitate the avoidance of maximal dose errors of up to even about 8 %. In addition, their use may make the strictest recommended overall dose accuracy level of 3 % (1SD) achievable.

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis consists of a summary and the following publications referred to by the Roman numerals in the text.

I **M. Kapanen**, M. Tenhunen, T. Hämäläinen, P. Sipilä, R. Parkkinen, H. Järvinen. Analysis of quality control data of eight modern radiotherapy linear accelerators: the short- and long-term behaviours of the outputs and the reproducibility of quality control measurements. Phys. Med. Biol. 2006; 51:3581-3592.

II **M. Kapanen**, M. Tenhunen, R. Parkkinen, P. Sipilä, H. Järvinen. The influence of output measurement time interval and tolerance on treatment dose deviation in photon external beam radiotherapy. Phys. Med. Biol. 2006; 51:4857-4867.

III **M. Kapanen**, P. Sipilä, R. Bly, H. Järvinen, M. Tenhunen. Accuracy of central axis dose calculations for photon external radiotherapy beams in Finland: the quality of local beam data and the use of averaged data. Radiother. Oncol. 2008; 86:264-271.

IV **M. Kapanen**, J. Collan, K. Saarilahti, J. Heikkonen, K. Kairemo, M. Tenhunen. Accuracy requirements for head and neck intensity-modulated radiation therapy based on observed dose response of the major salivary glands. Radiother. Oncol. 2009; in press.

AUTHOR'S CONTRIBUTIONS

All publications included in this thesis are a result of a group effort. In all the studies (I-IV), the author of this thesis analyzed the data, developed analyzing methodology and wrote the manuscript. In particular, he developed the method for estimation of effect of the use of different dosimetric quality control programs on treatment doses in the study II. The author designed the study II and participated actively in the design of the studies I, III and IV. He interpreted the results in the studies I-III and substantially participated in the interpretation of the results in the study IV. The author carried out the most substantial contribution to literature reseach (studies I-IV) and participated in patient treatment planning in study IV. None of the included publications has been previously used in a thesis by other author.

LIST OF ESSENTIAL ABBREVIATIONS AND SYMBOLS

1D	one dimensional
2D	two dimensional
3D	three dimensional
A	parameter in equivalent collimator relation
AAPM	American Association of Physicists in Medicine
A-P	antero posterior direction
C-C	craniocaudal direction
CBCT	cone beam computed tomography
CC	constancy check
CI	confidence interval
CT	computed tomography
D	absorbed dose
d	depth
D_{50}	a parameter determining location of sigmoidal dose response curve
d_{max}	depth of dose maximum
D_{mean}	mean value of absorbed dose
dose gradient	dose derivative with respect to distance, defined here as dD_{mean}/dr
ESTRO	European Society for Therapeutic Radiology and Oncology
FS	field size
Γ	a parameter determining steepness of sigmoidal dose response curve
γ	normalized steepness of dose response curve (= ΔNTCP(%)/ΔD(%))
HUCH	Helsinki University Central Hospital
IAEA	International Atomic Energy Agency
ICRU	International Commission on Radiation Units and Measurements
IMRT	intensity-modulated radiation therapy
IPEM	Institute of Physics and Engineering in Medicine
J_{20}^{10}	depth ionisation ratio ($\equiv J_{10}/J_{20}$)
LAT	lateral direction
MC	Monte Carlo
MLC	multileaf collimator
MRI	magnetic resonance imaging
MU	monitor unit
MV	mega volt
NACP	Nordic Association of Clinical Physics
NCS	Netherlands Commission on Radiation Dosimetry
NTCP	normal tissue complication probability
OF	output factor
OM	absorbed dose measurement for accelerator dose production
PBC	pencil beam convolution algorithm
PDD	percentage depth dose
PMMA	polymethyl methacrylate
PRV	planning organ at risk volume
PTV	planning target volume
PUC	probability of uncomplicated cure
QA	quality assurance
QC	quality control
QI	quality index

rEF	relative ejection fraction of a salivary gland
RF	radiofrequency
ρ	correlation coefficient
ROI	region of interest
RPC	Radiological Physics Center
RT	radiation therapy
SAD	source to axis distance
SD	standard deviation
SPD	source to phantom distance
SSD	source to skin distance
SSDL	secondary standard dosimetry laboratory
STUK	Finnish Radiation and Nuclear Safety Authority
TCP	tumor control probability
TLD	thermoluminescence dosimeter
TMR	tissue maximum ratio
TPR_{10}^{20}	tissue phantom ratio; ratio of doses at depths of 20 and 10 cm
TPS	treatment planning system (computer based)
X	width of a field
Y	height of a field

1. INTRODUCTION

The purpose of radiation therapy (RT) is to maximize killing of cancer cells in a tissue by using ionizing radiation while maximizing or keeping the sparing of healhy cells at acceptable level. Currently, external RT is most usually carried out by using linear accelerators proving high energy x-rays. This so called "megavoltage (MV)" radiation penetrates deeply into tissues enabling treatment of central parts of body with reasonable amount of dose delivered to superficial structures near the skin. RT is applied for several types of cancer but as well for some benign diseases. RT can be given as an exclusive treatment technique or combined with surgery and/or chemotherapy. RT is currently well available treatment technique in western world.

The role of RT as a curative or palliative treatment technique is significant and is currently increasing (Delaney *et al.* 2005, Elshaikh *et al.* 2006). Considering Finland alone, the number of RT centres was 9 at the time of data collection for this thesis (years 2003-2005) and has increased to 12 until time of the writing of this thesis (the beginning of year 2009), and, moreover, one private RT centre is starting treatment activity during a few months. One of the main reasons for the increase of the use of RT has been the development of technology of linear accelerators providing faster, more accurate and more reproducible treatment delivery in the sense of both dose and geometry. As a consequence, irradiation of normal tissues near a targeted volume has been reduced since smaller marginals are needed around the target. On the other hand, knowledge of radiation response of tissues has increased.

Considering the physical radiation quantities, biological effects correlate to absorbed dose received by a tissue. In some cases, the dependence is observed to be steep suggesting strict accuracy requirements for both absolute value of dose and the location where it is delived. Considering huge amount of patients treated annually by RT, it is expected that increased accuracy for both of the former factors has significant effect. Several international and national recommendations have been made for procedures to reach and to maintain a sufficient accuracy level. Most of the recommendations give accuracy limits for physical parameters based on practically achievable limits. Since the success of tumor destruction and avoidance of adverse biological effects are the only clinical measures for the success of the curative RT, estimation of accuracy requirements and benefits of improvements on biological basis is desirable. Unfortunately, the estimation of biological effects is difficult due to their complex nature including many sources of uncertainty and factors specific for a certain cancer and a treatment technique. Therefore, estimations based on biological effects are at their best only for special cases. General situation can be approached by doing estimations for 'average' dose response characteristics but application of such estimations may be limited.

The accuracy of absorbed dose delivered to a patient depends on several factors. These include quality of dosimetry, dosimetric and geometric accuracy of dose delivery, accuracy of dose calculation within a patient, quality of radiological images, treatment planning process, reproducibility of patient set-up during a treatment and correction strategy of patient positional errors during a treatment. From hospital physicist's point of view, quality control (QC) procedures concentrate on all these factors from the choice of good quality equipments and softwares to appropriate measurement procedures. Also development of methods for image quidance and patient set-up verification and immobilization is essential.

The most important aim for good dose accuracy is the maintenance of a narrow deviation in the radiation production of an accelerator. The radiation production is called 'output' and it is defined as the dose absorbed at a reference point in a beam per a monitor unit (MU) measured by accelerator monitoring chamber (Gy/MU). Shift of the output level results in changes of the doses received by all patients. The maintenance of good dosimetric accuracy in treatments depends on the stability of an accelerator, reproducibility of QC measurements and the ability of the used QC procedure in the detection of true drifts in the beam parameters. The output is measured at regular time intervals accurately in a water tank by using a very stabile reference instrument. The measurements are carried out with relatively long time intervals since they are workful and time consuming. Approximative constancy checks (CC) of output are carried out more often by using by fast and easily movable equipment. A CC device may not be very stabile requiring regular calibrations against the local reference instrument.

The use of an appropriate QC procedure for output should maximize the detection of both 'normal' and unexpected changes in output levels. QC programs given by different international or national authorities have recommended considerable variable time intervals for the output measurements ranging from one week to even one year. Intuitively it is safer to have short output measurement intervals but too redundant measurements waste resources and time that could be used for treatments. The RT centres have to choose their QC program intuitively without quantitative data of the efficacy of the use of the chosen program on the overall accuracy of dose. The elaboration and optimization of a local QC program would require the knowledge of normal time pattern in output level to estimate output stability and suitable time intervals for the measurements. Such knowledge might also facilitate the detection of potential malfunctions and measurement errors. The knowledge of measurement reproducibility is crucial in the choice of appropriate action levels for the measurements and in the evaluation of appropriate remedying actions for measured output changes. Due to relatively long output measurement time intervals, the reproducibility of the approximate CCs should be sufficient in the detection of output changes of only a few per cent with very few check repetitions. Moreover, long-term stability of a CC device should be sufficient (or known) with respect to the chosen output measurement time interval. All the factors mentioned above may depend on accelerator and dosimetric equipment types.

In addition to output, accuracy of dose calculations is one of the most crucial factors in the achievement of good overall dosimetric accuracy. A hospital physicist should choose appropriate calculation algorithm and verify that the measured basic beam data used to configure the calculation algorithm are accurate. The quality of measured beam data sets baseline for the accuracy level achievable for all dose calculations. The recognition of systematic measurement errors is difficult and methods are desired to identify such errors preferably already during a measurement session. The identification of sources for potentially poor calculation accuracy is desirable to avoid unnecessary and workful remeasurements of correct beam data sets. There is no consencus between the centres what accuracy level should be considered sufficient for the basic beam data.

Patient set-up errors have an important effect on the dose received by the target and normal tissues. Their magnitude, effect on the dose and related biological effects can be accurately estimated only in special cases. In intensity modulated radiation therapy (IMRT), the role of patient set-up errors is more prominent than in conventional RT due to steep dose gradients within or close to normal tissues located near the target produced by moving leafs of multileaf collimator during a treatment. An interesting special case of IMRT is head and neck IMRT, where substantial part of function of one or more of the major salivary glands is intended to

be spared. Salivary glands are quite radiation sensitive, they might locate at a distance of only few millimetres from the target and they are relatively small when compared to the target. The knowledge of geometric accuracy required for the treatments would be of a great clinical value. Evaluations of such requirement based on the dose response characteristics of the glands have not been published.

This thesis was concentrated in the investigation of the quality of dosimetric QC measurements, the efficacy and optimization of QC programs for the dosimetric measurements, and the quality of basic beam data used to configure dose calculation algorithm. The uncertainty related to these factors has a direct effect on the accuracy of dose received by all patients. Methods, suggestions and criteria were proposed to improve dose accuracy and to verify that the uncertainty related to the investigated factors is not too high when compared with the uncertainty of other factors currently achievable. Overall uncertainty of dose and the combined effect of the proposed factors to improve it were estimated. Biological effects related to the improvements were estimated for the general situation by using literature data. In addition to these factors, patient set-up accuracy and the effect of positional errors on the dose received by the major salivary glands were investigated in head and neck IMRT. Both dosimetric and geometric accuracy requirements were estimated for the case when substantial part of salivary gland function is intended to be spared. The requirements were based on the dose response characteristics obtained for the major salivary glands and dose gradients within the glands.

2. THEORETICAL BACKGROUND

This section contains a literature review and is aimed to give an overall impression of RT describing the basic concepts, therapy procedures and quality assurance. It clarifies the importance and central role of the investigated topics for the quality of treatments. On the other hand, the theory and international recommendations related to the investigated topics are presented in this section.

2.1. DOSE RESPONSE IN RADIOTHERAPY

Absorbed dose D is defined as the mean energy E imparted by ionizing radiation to matter of mass m:

$$D = \frac{dE}{dm} \tag{1}$$

The unit of D is Gray (Gy) [J/kg]. This physical quantity of radiation is found to correlate with biological effects releated to RT and the word "dose" means absorbed dose in this thesis.

Biological effects such as normal tissue complications and tumor eradication depend on the dose distributions within these tissues. Dose volume dependence has been parametrized by proposing several mathematical models (Schultheiss *et al.* 1983, Lyman 1985, Kutcher *et al.* 1991, Källman *et al.* 1992, Niemierko 1997, 1999). Two extreme types observed for dose volume relationship suggest that tissue inner structure can be modelled by either a serial or parallel architecture of functional units (Burman *et al.* 1991). Destruction of one functional unit may lead to a complication in a serially structured tissue rendering the NTCP dependent on maximum dose within the tissue. In a parallelly structured tissue, e.g. kidney, lung and salivary gland, the NTCP correlates best with mean dose within the tissue. TCP correlates usually best to minimum dose or minimum dose delivered to 95 % volume of a tumor (Withers 2000).

The general relationship between the dose and the probability of a biological effect has been observed to be sigmoidal for over 70 years ago (Holthusen 1936). This relationship, called dose response, is commonly parametrized by fitting a logistic model function of form

$$\mathrm{NTCP} = \frac{1}{1 + \left(\frac{D_{50}}{D} \right)^{\Gamma}} \tag{2}$$

to clinical data, where NTCP is the probability of a complication, and D_{50} and Γ are fitting parameters specific for the tissue and the complication. The parameter D_{50} gives the dose resulting in 50% probability of complication. Also the probability of eradication of a tumor, known as tumor control probability (TCP), can be parametrized by using the sigmoidal model.

The sigmoidal dose response can be approximated by a linear model within a narrow range of dose values.

The most important feature of the sigmoidal dose response is weak dependence of biological effects on dose with low ($D << D_{50}$) and high ($D >> D_{50}$) dose values. On the contrary, with intermediate dose values ($D \approx D_{50}$) a small change of dose results in significant change of NTCP or TCP (Figure 2-1). The steepness of the dose response is usually expressed by normalizing the derivative of NTCP as (Brahme 1984)

$$\gamma = \frac{d\text{NTCP}}{dD} D \approx \frac{\Delta\text{NTCP}(\%)}{\Delta D(\%)} \tag{3}$$

giving a dimensionless quantity for the steepness. The reported values of γ have varied typically from about 1 (Cheung *et al.* 2005a) to about 6 (Levegrün *et al.* 2001) depending on selected biological endpoint, cancer type and subgroup of patients. The value of about 3 has been considered as a general mean (Okunieff *et al.* 1995).

Irradiated volume in a patient contains both targeted and normal tissues. An overall risk related to a treatment can be estimated by combining the parametrized models of NTCP and TCP resulting in an estimate for the probability of uncomplicated cure PUC = TCP(1-NTCP) (Figure 2-1). The NTCP and TCP curves are usually steep and close to each other in the dose axis resulting in a narrow PUC curve. This sets biological basis for strict dose accuracy requirements in RT. Besides the biological aspects, high dose accuracy is essential for comparisons of treatment techniques and results between centres.

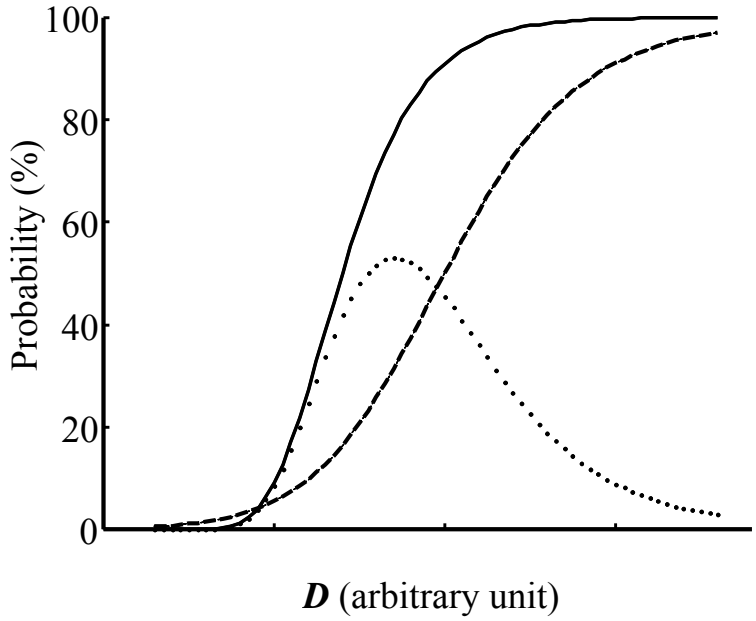

Figure 2-1. *The sigmoidal dose response models for tumor control probability (TCP, solid line) and normal tissue complication probability (NTCP, dashed line). The probability of uncomplicated cure (PUC, dotted line) is close to its maximal value within a relatively narrow range of dose values.*

The current dose response models are not capable for describing individual variations in radiation sensitivity. The models provide only percentages of complications and tumor control for a group of patients receiving equal and uniform dose. Combined effect of individual radiation sensitivity and uncertainty in quantification of biological effects is remarkable resulting in large scatter of data points and relatively large confidence limits for the values of the fitting parameters.

2.2. DETERMINATION OF DOSE WITHIN A PATIENT IN EXTERNAL PHOTON BEAM RADIOTHERAPY

2.2.1. Dose measurements

The determination of absorbed dose within a patient is based on the measurement of absorbed dose in water, since mean electron density of soft tissues is close to that of water. The measurements are based on Bragg-Gray cavity theory (Bragg 1912, Gray 1936) and its further improvements (Burlin 1966) assuming that i) the air cavity of a measuring instrument does not change the energy spectrum of secondary electrons, ii) photons produce only negligible amount of secondary electrons in the air cavity and iii) the flux of photons is constant in the air cavity and in its surroundings. A small (< 1 cm^3) air filled thimble type ionization chamber with thin walls fullfills best these assumptions and is recommended for the reference tool of clinical dosimetry. Chambers with graphite walls have been found to have better long-term stability and more constant response for different x-ray beam energies than chambers with plastic walls. A commonly used ionization chamber for photon beam dosimetry is open 0.6 cc Farmer-type NE 2571 (NE Technology Ltd, Reading, UK).

The dosimetry of the RT equipment is carried out by following of one of the internationally recommended dosimetric code of practice. At the moment, the TRS-398 given by the International Atomic Energy Agency (IAEA 2000) is used in Finland. Electric charge (ionization) formed in a chamber is collected by high voltage and measured by an electrometer and needs to be calibrated to absorbed dose value. The radiotherapy centres calibrate their ionization chamber in a secondary standard dosimetry laboratory (SSDL) where the calibration is carried out in the ^{60}Co beam against a secondary standard of absorbed dose. The calibration is valid in reference conditions with temperature 20 ^{o}C and athmospheric pressure 101.3 kPa, and correction is required for actual measurement conditions. The calibration is traceable to a primary standard. According to IAEA, the accuracy of the calibration of a clinical dosimetry instrument is 0.6 % (1SD) when calibrated in a SSDL.

The calibration coefficient of a chamber should be corrected for the nominal energy of user's x-ray beam. The correction factor is given in the dosimetric code of practice and is based on the value of a beam quality index (QI). Finally, the dose absorbed in water for radiation of quality Q is given as

$$D_{\mathrm{w},Q} = M \cdot k_{\mathrm{Tp}} \cdot k_{\mathrm{s}} \cdot k_{\mathrm{pol}} \cdot k_{\mathrm{elec}} \cdot N_{\mathrm{w},Q} \qquad (4)$$

where M is reading of an electrometer and correction factors k_{Tp}, k_{s}, k_{pol} and k_{elec} are for conversion of actual measurement conditions to reference conditions, for recombination of ion pairs before they are collected, for polarity of collecting voltage and for sensitity of

electrometer, respectively. $N_{w,Q}$ is the calibration coefficient of the chamber for dose absorbed in water for radiation of quality Q.

For the QI, the TRS-398 recommends measurement of tissue phantom ratio (TPR_{10}^{20}, ratio of doses at depths of 20 and 10 cm) or ratio of depth ionizations at depths 10 and 20 cm ($J_{20}^{10} \equiv J_{10}/J_{20}$) for a field size of 10x10 cm^2. According to the TRS-398 these quantities are empirically related as

$$\mathrm{TPR}_{10}^{20} = \frac{1.2661}{J_{20}^{10}} - 0.0595 \qquad (5)$$

According to IAEA (2000), overall uncertainty of clinical absorbed dose measurements is 1.5 % (1SD) and a measurement procedure carried out in user's beam gives the most significant contribution to this being even 1.4 % (1SD). The latter is devided into 5 factors: 1) long-term stability of user dosimeter 0.3 %, 2) establishment of reference conditions 0.4 %, 3) dosimeter reading relative to beam monitor 0.6 %, 4) correction for influence quantities (k_i) 0.4 % and 5) beam quality correction 1.0 %. The accuracy of absorbed dose measurements could be significantly improved if the uncertainty of the measurements carried out in the user's beam could be reduced.

2.2.2. Dose calculation

A treatment is administered in monitor units (MUs) measured by two independent monitoring chambers within an accelerator's head. The purpose of dose calculation is to estimate dose within a patient and to calculate the number of MUs needed for planned dose delivery for each field. Reference methods for dose calculation are based on Monte Carlo (MC) (Demarco *et al. 1998*). For clinical purposes, however, dose calculation should be done in times of few minutes and approximative analytical calculation models are used instead of the more physical MC models. In the Nordic countries, the most commonly used calculation methods for photon beams are the pencil beam convolution (PBC) (Storchi *et al.* 1999) and the collapsed cone convolution (CCC) (Ahnesjö 1989). In this thesis, the dose calculations are carried out by using the PBC and quality of beam data used to configure this model are evaluated. The principles of the PBC model and the equations relating the configuration beam data to measured dose values are presented.

The PBC method

The PBC method devides therapy beam into small cylindrical beams of radius 0.5 cm. Absorbed dose is calculated in a water phantom for a field F using the equation (Storchi *et al.* 1999)

$$D(x,y,d;F) = \frac{(\mathrm{SPD}+d_{\mathrm{ref}})^2}{(\mathrm{SPD}+d)^2} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(x',y') P_{\mathrm{int}}(x',y',d) K(x-x',y-y',d)\,dx'\,dy' \qquad (6)$$

where $F(x,y)$ is field intensity matrix, $P_{int}(x,y,d)$ intensity profile and $K(x,y,d)$ is pencil beam kernel. The $F(x,y)$ include field geometry and its value is 1 in an open part of the field, 0 outside the field and equal to the fraction of transmission of MLC under the part of the field shielded by the MLC. $P_{int}(x,y,d)$ is normalized fluence of primary photons accounting for the non-flatness of the open field beam and variations of photon fluence due to variation of energy spectrum as a function of off-axis distance. The x and y are coordinate axis in the directions parallel to the width and height of the field, respectively, d is the depth to which dose is calculated and d_{ref} is the reference depth. SPD is the distance from the x-ray source to the surface of the phantom. The convolution is done at 5 standard depths and the dose is interpolated along fanlines for other depths. Finally, the dose distribution is corrected to take into account correct source to skin distance (SSD), tissue inhomogeneities and oblique incidence of the beam to a patient.

Basic beam data and their relation to calculated number of MUs

Three types of basic beam data should be measured to configure the PBC model. The $P_{int}(x,y,d)$ and $K(x,y,d)$ are determined iteratively from the measured data (Storchi *et al.* 1999). The measurements are carried out in a large water phantom with recommended miminum size of 30x30x30 cm^3 conventionally by using the SPD of 100 cm. The measured data are converted into relative form and, therefore, absorbed dose calibration of a beam has to be done only at one reference point. The types of the required beam data are:

1) Measurement of depth dose data for square fields XxX is recommeded at least for field side X = smallest, 4, 6, 8, 10, 12, 15, 20, 25, 30, 35 and 40 cm. The data is normalized to dose (D) at the depth of dose maximum (d_{max}) giving percentage depth dose data (PDD) defined as

$$\mathrm{PDD}_{X\mathrm{x}X}(d) = \frac{D_{X\mathrm{x}X}(d)}{D_{X\mathrm{x}X}(d_{\max})} \cdot 100\% \qquad (7)$$

PDD data for a rectangular field XxY are taken from the PDD of an equivalent square field calculated to give the same dose at 10 cm depth as the field XxY.

2) The number of monitor units (MU) providing a certain dose (e.g. 1 Gy) at d_{max} should be measured for all rectangular field sizes. Field size resolution of 1 cm is recommended. The number of MUs obtained for a field size XxY is normalized to the number of MUs obtained for a reference field size 10x10 cm^2 resulting in an output factor (OF) defined as

$$\mathrm{OF}_{X\mathrm{x}Y} = \frac{\mathrm{MU}_{10x10}}{\mathrm{MU}_{X\mathrm{x}Y}} \qquad (8)$$

The OFs of rectangular fields can be used to calculate OFs for asymmetric and dynamic wedge fields (Tenhunen 1994).

3) The PBC model requires measurement of beam profiles for the square fields at least at three depths, but five measurement depths of d_{max}, 5, 10, 20 and 30 cm are recommended.

Measurements should be done for the same field sizes for which PDDs were measured. Peaks caused by noise in the profiles capable of causing error in the normalization of the profile on the central axis are handled by curve fitting through the center of the profile. A weighted average between the fit and measured curve is used. In addition, the diagonal profile is needed for the largest open field size.

At the geometry of beam data measurements, the dose D delivered to a water phantom is calculated from a simple equation showing explicit dependence of D on the measured beam data. The number of MUs required for a delivery of D to the depth d at the central axis of a square non-MLC field XxX is calculated as

$$\mathrm{MU}_{X\mathrm{x}X} = \frac{\mathrm{MU}_{10\mathrm{x}10} \cdot 100\%}{\mathrm{OF}_{X\mathrm{x}X} \cdot \mathrm{PDD}_{X\mathrm{x}X}(d)} \tag{9}$$

where $\mathrm{MU}_{10\mathrm{x}10}$ is the number of MUs measured to deliver the same dose to the $d_{\max}$ of the reference 10x10 cm^2 field. The Equation (9) demonstrates the impact of the accuracy of OF and PDD data on the accuracy of the dose calculated at the central axis of a field. Dose calculated at an off-axis point is normalized to the dose at the central axis and, therefore, the accuracy of the central axis dose forms a baseline for the accuracy of dose calculation at the other points in a field. For a rectangular field XxY, the Equation (9) becomes

$$\mathrm{MU}_{X\mathrm{x}Y} = \frac{\mathrm{MU}_{10\mathrm{x}10} \cdot 100\%}{\mathrm{OF}_{X\mathrm{x}Y} \cdot \mathrm{PDD}_{c\mathrm{x}c}(d)} \cdot \frac{\mathrm{PSF}_{X\mathrm{x}Y}}{\mathrm{PSF}_{c\mathrm{x}c}} \tag{10}$$

where PSF is the phantom scatter factor and c is the side of an equivalent square field resulting in the same central axis D at a depth of 10 cm as the field XxY (calculated according to the Eq. 6). The PSF values are taken from a table obtained by the NCS (Van Gasteren *et al.* 1991) by taking into account the QI of the beam (Storchi and Van Gasteren 1996). If SSD is different from that used in the beam data measurements (SPD), the PDD data are corrected by using the formula (Khan 2003, p. 172)

$$\mathrm{CF}_{\mathrm{MF}}(d) = \frac{\mathrm{T}(d, c_2)}{\mathrm{T}(d, c_1)} \cdot \left[\frac{\mathrm{SSD} + d_{\max}}{\mathrm{SSD} + d}\right]^2 \cdot \left[\frac{\mathrm{SPD} + d}{\mathrm{SPD} + d_{\max}}\right]^2 \tag{11}$$

where

$$c_1 = \frac{\mathrm{SPD} + d}{\mathrm{SPD}} \cdot c$$

$$c_2 = \frac{\mathrm{SSD} + d}{\mathrm{SSD}} \cdot c$$

being based on the Mayneord F factor (Mayneord and Lamerton 1944) corrected by T(d,c) ratio, where T(d,c) is tissue maximum ratio (TMR) or tissue air ratio (TAR) value at depth d for a field having an equivalent field size c at the field entry level. The TMR is determined from the measured beam data as (Khan 2003, p. 196)

$$\mathrm{TMR}(d,\mathrm{FS}_d) = \frac{\mathrm{PDD}_{\mathrm{FS}_d}(d)}{\mathrm{PDD}_{\mathrm{FS}_{d_{\mathrm{ref}}}}(d_{\mathrm{ref}})} \cdot \left[\frac{\mathrm{SSD}+d}{\mathrm{SSD}+d_{\mathrm{ref}}}\right]^2 \cdot \frac{\mathrm{PSF}(\mathrm{FS}_{d_{\mathrm{ref}}})}{\mathrm{PSF}(\mathrm{FS}_d)} \qquad (12)$$

where FS_d is field size at depth d and $\mathrm{FS}_{d_{\mathrm{ref}}}$ is field size at depth d_{ref}.

2.2.3. Radiation therapy procedures

First phase of the RT procedures is aquisition of adequate anatomical and functional information for the planning of a target volume in a patient. Geometric reference of patient anatomy is a 3D tomography image (CT) series or sometimes a magnetic resonance image (MRI) series, in which the target is drawn by a radiation oncologist. The planning of the target may require registration of images from different imaging modalities. Patient position in the reference image series should be accurately reproduced in each treatment fraction requiring sufficient patient immobilization technique and patient set-up verification during the treatment. Isocenter of therapy beams is currently chosen when the reference image series is acquired. When necessary, the position of the isocenter can be verified by simulating the patient set-up in a treatment simulator where x-ray images can be aquired from the directions of the therapy beams. For the set-up verification, at least two orthogonal reference images containing suitable land marks (usually bones and/or air cavities) are acquired or reconstructed from the reference series by using a treatment planning system (TPS).

Treatment planning is carried out by choosing the number, orientation, shape, radiation type and energy of the therapy beams and also dose fractionation scheme. The location of the target and critical organs is accurately visualized by the TPS and the dose distribution within a patient can be calculated to the reference anatomical image series. Uniform dose distribution of prescribed dose is aimed for the target while irradiation of normal tissues is minimized by proper choice of field orientations and shapes. Treatments are often delivered by using static fields shaped conformal to target by using multileaf collimator (MLC). MLC consists usually from 80 to 120 leafs of a width 0.5-1.0 cm independently adjustable to bound the target while blocking normal tissues.

Image guidance is used in RT meaning that anatomic images are acquired during the treatment. The locations of the land marks between those images and the reference image are determined by on-line image registration and observed differences are corrected before treatment delivery. The use of cone beam CT acquistions (CBCT) during the treatment has currently increased providing information of the deformation of tissue structures. The quality of CBCT images sometimes makes it possible to replan a treatment to correspond potentially changed patient anatomy.

With recent years, the use of intensity modulated treatment technique (IMRT) (Brahme *et al.* 1982, AAPM 2003, Bortfeld *et al.* 2006, Meyer 2007) has increased significantly. In that technique, highly conformal dose distribution is achieved usually by moving the leafs of the

MLC during the irradiation. In IMRT, flatness of dose distribution within the target is compromised to enhanced normal tissue sparing near the target. Inverse treatment planning method is utilized by setting constraints for dose and volume of the target and critical normal tissues (Brahme 1988, Brahme 1995, Webb 1997). A function containing suitable penalties for unfulfilled constraints is minimized resulting in an optimized photon fluence matrix for each treatment field providing optimal dose distribution within a patient. The optimal fluence is converted into actual fluence by calculating tables for MLC movement and by taking tissue inhomogeneities within a patient into account (Svensson *et al.* 1994, Webb 1997). An alternative approach to inverse IMRT treatment planning is forward planning (fIMRT) achieved recently increasing interest (Kestin *et al.* 2000, Xiao *et al.* 2000, Chen *et al.* 2001, Lee *et al.* 2004). In that technique, basic field set-up is planned first and dose distribution within patient calculated. After this, low dose regions within the target and hot spots within normal tissues are corrected by adding constraints or fields targeted to the low dose areas while shielding the areas of hot spots.

Each RT procedure is prone to several sources of uncertainty requiring appropriate quality assurance. With a physical point of view, the important factors are proper contrast and geometric scaling of the images used for the treatment planning, accuracy of image registration, appropriate patient fixation providing well reproducible positioning, proper patient set-up verification, knowledge of the nature and effects of positional errors, and correction methods for such errors during the treatment.

2.3. CLINICAL QUALITY ASSURANCE

2.3.1. General aspects

International Organization for Standardization (ISO) has defined that the quality is ”*the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs*” (ISO 1994). Practically this means that a patient should have relevantly planned and realized therapy. The quality assurance (QA) should comprehend whole treatment process from the prescription of the therapy to the follow up of the patient. At a clinical level QA should concern quality policy, organization, methods, equipment, and education as described in the comprehensive international QA recommendations (Kutcher *et al.* 1994, Thwaites *et al.* 1995). From a physicist’s point of view, the QA can be captured to mean the verification that the planned dose value is delivered to the planned location in a patient. Any deviation from the treatment plan may lead to unexpected complications or loss of treatment efficacy, and hampers comparison of methods in a centre and between centres. Methodological QA is required to ensure the choise of relevant treatment technique and planning method, while technical QA is needed for accurate realization of the treatment plan. Quality control (QC) describes practical actions related to QA.

The ISO has recommended that estimated accuracy should be expressed in uncertainties (ISO 1995). Fundamentally, the uncertainties can be classified into type-A which can be estimated by statistical means and type-B estimated by other means. Both of these types are expressed by standard deviation (SD) and overall uncertainty is the combination of the uncertainties of the both types estimated for all factors involved in a process. Gaussian uncertainties of independent sources are summed in squares. QA aims to reduce the uncertainty of the both type A and B.

Several recommendations exist for maximal uncertainty of dose within a RT patient: ± 3 % (1SD) given by Brahme *et al.* (1988), ± 3.5 % (1SD) given by Mijnheer *et al.* (1987) and ± 5 % (interpreted as 95% CI) given by ICRU (1976). The estimations made by Brahme *et al.* and ICRU are based on conclusions made for dose response of some types of tumors while Mijnheer *et al.* investigated NTCP. Moreover, IAEA has concluded in the TRS-398 that ± 5 % (1SD) might be a sufficient accuracy level. The differences in these recommendations can be understood by considering the complex nature of biological effects rendering recommendations at general level difficult. Since the studies included in this thesis concentrate to both dosimetric and geometric QA, these topics and also most important recommendations relating to the investigations are introduced.

2.3.2. Dosimetric quality assurance

Dosimetric quality assurance should comprehend all radiation therapy procedures affecting the value of the dose absorbed to a patient. From a physical point of view, the main topics of dosimetric QA are the accuracy of dosimetric measurements, the maintainance of dose output of an accelerator in the predefined tolerance limits and the accuracy of dose calculation.

Dose measurements

QC for a reference ionizing chamber comprehends regular calibrations at a standards laboratory and verification for long-term stability of the calibration before each absorbed dose measurement by using a radioactive check source. Before an absorbed dose measurement, flatness of beam profiles indicating correct beam orientation and depth dose data indicating correct beam energy should be measured for reference field size. Absorbed dose measurements used for the calibration of a beam are often called ‘output measurements’ since they measure dose production (Gy/MU) of a linear accelerator. In this thesis, the dose production is called ‘output’ and the absorbed dose measurements are called ‘output measurements’.

The output measurements are carried out in a large water phantom (minimum size 30x30x30 cm^3) at reference geometry at a reference point for a reference field size, usually 10x10 cm^2. The output measurements carried out as described above are workfull and are often called “definitive calibrations” and they are performed rarely. “Routine calibrations” are performed more often by using the reference ionizing chamber but they can be carried out in a smaller water or solid (e.g. PMMA) phantom. These calibrations can be converted into definitive calibrations (Seuntjens *et al.* 2005). Approximative constancy checks (CC) are performed frequently, preferably daily. Very commonly used istrument is a sealed plane parallel ionizing chamber which is larger and has lower spatial resolution than the reference chamber but is not so fragile. Unfortunately, CC devices are usually not stable requiring regular calibrations against the reference chamber.

QC programs in RT centres have remarkable differencies. Several international QC programs have been published recommending considerable different time intervals for the output measurements ranging from one week (IPEM 1999, SSRPM 2003) or two weeks (NCS 1996) to even one year (WHO 1988, Dunscombe *et al.* 2007). Also time intervals laying somewhere between these have been proposed being one month (NACP 1980, Kutcher *et al.* 1994, IPEM 1999) and 3 months (NACP 1980). The most recommended action levels for the output measurements are ± 2 % (NACP 1980, IEC 1989, Kutcher *et al.* 1994, IPEM 1999, SSRPM 2003). Also wider action levels being ± 3 % (Aird *et al.* 2007) have been suggested, while

some centres have been reported to use very narrow tolerance levels from ± 0.75 % to ± 1.5 % (Meijer *et al.* 1998, Luketina and Greig 2004). Moreover, an output calibration program without tolerances, interpreted as an output calibration exactly to a measured value, has been proposed (WHO 1988). Some centres define tolerance as action level while other consider excess of tolerance level as an indicator for increased control for the treatment unit performance. In unclear situations, the term 'tolerance' is used here as it was in the cited literature. The centres have to define and document procedures for remedies relating to the excess of the predefined action levels. For the daily CCs, action levels of ± 3 % (Kutcher *et al.* 1994, SSRPM 2003) or ± 5 % (IPEM 1999) have been recommended.

Approaches to optimize a local QC program

The investigation of time trends in dosimetric parameters such as output (Watanabe 2000, Kristensen *et al.* 2003, Krutman *et al.* 2004, Luketina and Greig 2004, Tuncel *et al.* 2004) or beam energy (Biggs 2003) has recently been increasing. The results have been utilized for fine tuning a local QC program in the sense of measurement time interval and action levels but the latter has received more attention. The action levels are conventionally estimated by using the mean and standard deviation of a measured parameter (Cozzi and Fogliata-Cozzi 1998, Van Esch *et al.* 2000), but also a more sophisticated method to improve the detection of systematic shifts in outputs based on statistical process control has been published (Pawlicki *et al.* 2005). Only a few publications exist on seeking evidence for the effectiveness of a use of a certain output measurement time interval and action levels (McKenzie 2003, McKenzie *et al.* 2005, 2006). These studies have attempted to balance the costs of resources spent for the QC procedures and the costs related to significant underdosage of patients not detected by the QC procedures resulting in tumor recurrence and loss of lives. In these studies, it was clearly stated that such cost-benefit analysis includes several sources of uncertainty and they are based on theoretic assumptions or worst case scenarios of output changes and their detection instead of real observations. Therefore, the benefits and costs related to a QC program remain highly uncertain.

Beam data and dose calculation

QC of dose calculation concerns the quality of basic beam data used for the configuration of TPS and the accuracy of dose calculations for realistic situations emphasizing uncertainty inherent to a calculation algorithm. These categories are usually mixed together mistakenly but considerable different accuracy criteria exist for these categories.

Dose calculation accuracy at the central axis of open (non-MLC) rectangular fields is most explicitely related to the accuracy of basic beam data used for the TPS configuration (as was seen in 2.2.2.) and reflects best accuracy achievable for all dose calculations for that beam. Several recommendations exist for accuracy criteria for open square fields ranging from 1 % (Fraass *et al.* 1998, IPEM 1999) to 2 % (Ahnesjö and Aspradakis 1999, Venselaar *et al.* 2001). For open rectangular fields the corresponding criteria are from 1.5 % (Fraass *et al.* 1998) to 2 % (Ahnesjö and Aspradakis 1999, Venselaar *et al.* 2001). For a normalization point at central axis of a regtangular open field, defined as OF, stricter accuracy criterion of 0.5 % (Fraass *et al.* 1998) has been recommended.

For dose calculations in realistic treatment situations, lower accuracy criteria are suggested. The most cited ICRU Report 42 (1987) has recommended the tightest accuracy criterion of 2% for low dose gradient areas or 2 mm for high dose gradient areas. Consistently, by

combining uncertainty related to the RT procedures it has been concluded that dose calculation does not need to be more accurate than 2-3 % while 1 % is as an ultimate goal (Ahnesjö and Aspradakis 1999). On practical basis it has been estimated that accuracy of 3-4 % or 3-4 mm for high dose gradient regions is achievable (Van Dyk *et al.* 1993, IPEM 1999, Venselaar *et al.* 2001). In the presence of large 3D inhomogeneities, remarkably lower accuracy criteria of 5-7 % or 7 mm in penumbra region have been considered acceptable (Fraass *et al.* 1998). The criteria are interpreted to concern all calculation points or concerned as 95 % CI for these points. Several QC programs have been published for the TPS utilizing conventional static therapy fields (Fraass *et al.* 1998, AAPM 2001, Venselaar and Welleweerd 2001, Venselaar *et al.* 2001, ESTRO 2004, IAEA 2004) and also for systems utilizing modern delivery methods such as IMRT fields (Van Esch *et al.* 2002, ESTRO 2008, ICRU 2009).

Some elaborate QC methods have been proposed to evaluate the accuracy of central axis beam data based on parametric presentation of the TPR (Bjärngard *et al.* 1997, Xiao *et al.* 1998). The implementation of these methods requires measurements of beam attenuation and hardening coefficients in a narrow-beam geometry and/or several measurements of the TPRs. Due to the generality of these methods, they provide an accuracy level of about 1-2 % for FSs larger than 10x10 cm^2 and in the region of electron equilibrium. Another approach to QC of beam data is the comparison of measured beam data to standardized reference beam data. Based on observations of great consistency in beam data measured for the accelerators of the same make, model and energy, the Radiological Physics Center (RPC) in the USA has started construction of comprehensive reference data sets for several accelerator models of different vendors (Followill *et al.* 2004). The utilization of reference data for QC is much simpler than the use of the elaborated methods. It is not clear, however, whether the proposed methods are accurate enough in the detection of relatively small errors exceeding the strict accuracy criteria of ± 1 % recommended for the basic beam data.

External audits for radiotherapy dosimetry

The purpose of external audits is the review of local RT procedures by an independent external body. The dosimetry audits can be carried out by on-site review visits and/or off-site reviews by using mailed dosimeters. The on-site audits comprehend checks of local dosimetry systems, tests of dosimetry parameters of RT equipment, tests of TPSs, review of QA programmes and review of dosimetry records. On on-site audits, the dosimetric checks are usually carried out by ionizing chamber measurements in a water phantom while thermoluminescence dosimeters (TLDs) are most commonly used for the mailed reviews. Most on-site audits are carried out at national level by a national authority such as the Radiation and Nuclear Safety Authority (STUK) in Finland (Järvinen *et al.* 2001). The three major TLD networks are the IAEA/WHO (World Health Organization) TLD postal dose audit service operating worldwide (Svensson *et al.* 1990, Izewska and Andreo 2000), the European Society for Therapeutic Radiology and Oncology (ESTRO) EQUAL program operating in the European Union (Ferreira *et al.* 2000) and the RPC network operating in North America (Hanson *et al.* 1991, Aguirre *et al.* 2002). The EQUAL network has been linked to the European Commission network (Dutreix *et al.* 1994).

2.3.3. Geometric quality assurance

Geometric QA should comprehend all radiation therapy procedures affecting the location to which the dose is delivered in a patient. Geometric accuracy is more crucial in IMRT than in conventional RT due to steep dose gradients.

Geometric accuracy requirements for a treatment unit concerns size of radiation field, size of light field, position of MLC leafs, position of radiation isocenter, position of geometric isocenter, collimator rotation, gantry rotation, position of treatment couch and quality of an imaging device. High geometrical accuracy is required also for anatomical data used for the treatment planning concerning the phases from the data acquisition and correct data transfer to the capability of a TPS to handle the data accurately. Several QC recommendations and programs have been published for testing all these technical factors including action levels for remedies (Kutcher *et al.* 1994, IPEM 1999, SSRPM 2003).

Accuracy of patient set-up depends on a treatment location in a patient, patient fixation technique, method used for isocenter positioning, image guidance method used for position verification and, to some extent, co-operation of the patient. In addition, the target volume may move within a patient e.g. due to breathing movement. Uncertainty related to these factors is taken into account by planning a treatment to a volume including margins added around clinical target volume (CTV), encompassing clinically obvious gross tumor volume (GTV) and its potential subclinical spreading, resulting in the use of planning target volume (PTV) (ICRU 1993, ICRU 1999). QA is aimed to ensure that acceptable methodology is used to confirm sufficient reproducibility for patient set-up and coverage of PTV irradiation while minizing margins needed around CTV to minimize irradiation of normal tissues. Sometimes margins are added around critical normal tissues in a treatment plan to maximize their sparing resulting in the use of planning organ at risk volume (PRV). Comprehensive and detailed summary of sources of positional uncertainty and their estimated magnitudes for different treatment locations is given by the British Institute of Radiology (BIR 2003).

3. AIMS OF THE THESIS

The aim of the thesis was to estimate uncertainty of absorbed dose of radiotherapy patients related to treatment and dosimetry procedures, such as quality control measurements, commissioning of a treatment planning system, and patient set-up. Some practical methods, suggestions and criteria were proposed to improve the accuracy of the absorbed dose. The specific aims of the study were:

1. to estimate reproducibility of output measurements of linear accelerators, to assess reproducibility and long-term stability of constancy checks of output, to investigate typical time behaviour of output to elaborate and optimize quality control procedures, to propose methods for the estimation of all these subjects. (Study I).

2. to evaluate the effectiveness of different internationally recommended quality control programs for output measurements by taking into account reproducibility of output measurements and constancy checks of output, and stability of output. (Study II).

3. to evaluate the accuracy of basic beam data used for the configuration of treatment planning systems in Finnish radiotherapy centres, to determine dose calculation accuracy at reference geometry, to evaluate practically achievable accuracy criterion for central axis dose calculations, to find out potential reasons for reduced dose calculation accuracy, to estimate consistency of beam data for the most commonly used accelerator models, to develop a method for quality control of basic beam data. (Study III).

4. to determine the magnitude of set-up errors in head and neck IMRT, to estimate the effect of positional errors on the dose received by the major salivary glands, to determine dose response model parameters for planned dose of the salivary glands, to estimate dosimetric and geometric accuracy requirements for the head and neck IMRT in the case when substantial fraction of salivary gland function is intended to be spared. (Study IV).

5. In the thesis, the effectiveness of the methods, suggestions and criteria proposed in Studies I-IV was summed. Combined uncertainty of dose and its improvements were estimated. Related biological effects were also assessed.

4. MATERIALS AND METHODS

4.1. STUDY MATERIAL

4.1.1. Quality control data

The QC data collected by the Department of Oncology, Helsinki University Central Hospital (HUCH), between January 2000 and December 2004 were analyzed to determine short- and long-term time trends in accelerator outputs and the reproducibility of the output measurements and the CCs of outputs. The data had been collected for photon external beams of eigth linear accelerators, including one Varian Clinac 600 C (6 MV), two 600 CDs (6 MV), one 2100 C (6 and 15 MV), two 2100 CDs (6 and 15 or 18 MV) (Varian Medical Systems, Palo Alto, CA), one BrainLab Novalis (6 MV) (BrainLab, Heimstetten, Germany) and one Elekta SL18 (6 and 15 MV, Elekta Oncology Systems, Crawley, UK). The output measurements had been carried out by experienced hospital physicists at the depth of 10 cm in a 10×10 cm^2 beam at maximal time intervals of 6 months by using a 0.6 cc open ionization chamber type NE 2571 and NE Farmer 2570 electrometer (NE Technology Ltd, Reading, UK). The equipment had been calibrated at years 2000 and 2003 in the national secondary standard laboratory (STUK) and the constancy of the calibration had been checked before each measurement by using a check source and also by comparative measurements in connection with on-site visits by the STUK several times a year. The CCs of the outputs had been carried out two or three times a week by the RT technologists by using four sealed plane parallel ionization chambers of type PTW-Linacheck T42010 (PTW, Freiburg, Germany). One device was used per two accelerators. The CC devices had been calibrated in connection with the output measurements. Relative dosimetry comprehending the checks of beam profiles and depth dose (QI) was carried out before each output measurement by using a 0.13 cc open ionization chamber (Wellhöfer IC15, Scanditronix-Wellhöfer, Uppsala, Sweden) and a Wellhöfer WP700 scanning water tank with Dosimetrie WP700 software (Wellhöfer Dosimetrie, Schwarzenbruck, Germany).

4.1.2. Measurements for dose calculation accuracy and beam data

Dose measurements

Measurements used for the evaluation of dose calculation accuracy were carried out by the STUK during site visits in the Finnish RT centres between the years 2003 and 2005. The measurements were performed for 48 photon beams from 28 accelerators including all 9 RT centres in Finland. The absolute dosimetry was carried out by using a 0.6 cc Farmer type ionization chamber (NE 2571) and an electrometer (NE Farmer 2570). The chamber and electrometer calibrations came from the Finnish SSDL in STUK and were traceable to the primary normal of the International Bureau of Weights and Measures (BIPM). On the site visits, parameters related to beam profiles (symmetry, flatness and penumbra) and energy (J_{20}^{10} and/or TPR_{10}^{20}) were verified to be consistent with the values measured in connection with institution's beam data measurements.

The accelerators included in the analysis were 7 Varian Clinac 600 Cs or CDs (4-6 MV), 18 Varian Clinac 2100 Cs or CDs (6, 15 or 18 MV), 1 Varian Clinac 2100 SC (6, 10 MV), 1 Elekta SL 75-5 (6 MV) and 1 Elekta SL 18 (6, 15 MV). One out of 29 accelerators in Finland was excluded because it was a stereotactic machine with limited field sizes.

Dose calculations

Dose calculations for the dose delivery measured by the STUK were carried out by a hospital physicist. The number of monitor units (MUs) for an isocentric central axis dose delivery of 2 Gy at the depth of 10 cm in water (SAD=100 cm, SSD=90 cm) were calculated for the following FSs (*X*x*Y*): 5x5, 7x7, 10x10, 15x15, 20x20, 30x30, 40x40, 5x30 and 30x5 cm^2. Calculations were performed by using the single PBC algorithm implemented in CadPlan® (Varian Medical Systems, Helsinki, Finland) or Eclipse® (Varian Medical Systems, Palo Alto, CA) TPS software.

Basic beam data

OF and PDD data used for the configuration of TPS were collected from all 9 Finnish RT centres for the analysis. The geometry set-ups used for the basic beam data and the absorbed dose measurements are shown in Figure 4-1.

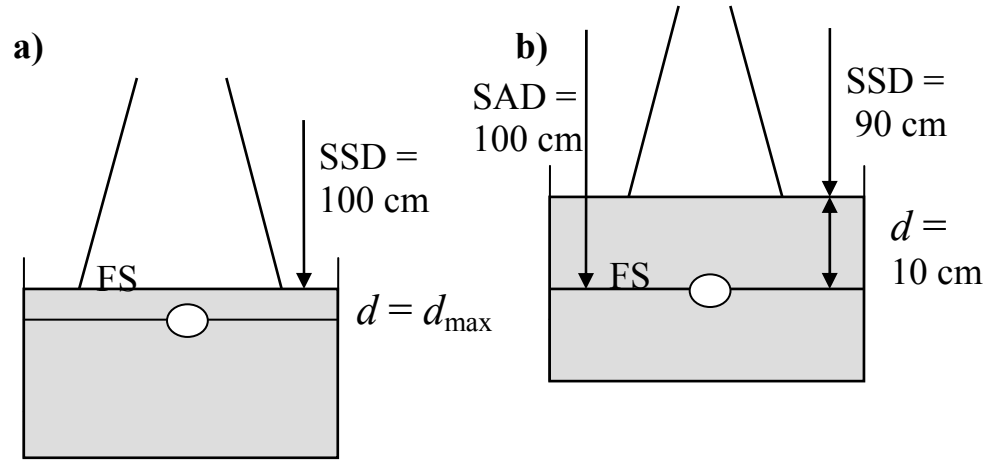

Figure 4-1. *a) The geometry set-up used for basic beam data measurements in connection with local TPS configuration and for local beam calibrations. b) The geometry set-up used for the dose measurements to determine calculation accuracy. The depth of field size definition is indicated as FS.*

4.1.3. Patient material

Patient and tumor characteristics

Data of 25 patients scheduled to be treated by bilateral radiation therapy for head and neck squamous cell cancer between the years 2005-2007 were analyzed to estimate patient set-up errors, the effect of set-up errors on the dose received by the major salivary glands and dose response characteristics of the major salivary glands. Pretreatment staging of the tumors had been done by clinical examination, MRI and endoscopy, and biopsies had been taken from the primary tumor and, when necessary, from clinically suspicious lymph nodes. The tumors had been staged according to the International Union Against Cancer (UICC) Tumor-Node-Metastasis (TNM) classification version 2002. More detailed main characteristics of the

patients and tumors are presented in Study IV (Table 1). Thirteen patients received definitive radiotherapy with curative intent and 12 patients received postoperative radiotherapy. Total doses varied from 50 to 70 Gy received at 2 Gy fractions. Concurrent weekly cisplatin 40mg/m^2 was given to all 13 patients treated by definitive radiotherapy and to 6 of the 12 patients treated by postoperative radiotherapy. The patients belong to a group participating in a prospective study of head and neck radiotherapy approved by an Ethics Committee of the Helsinki University Central Hospital. The current investigations did not change the treatment practice of these patients.

Measurement of salivary gland function

Individual salivary gland function had been carried out by scintigraphy. The method has been described earlier in detail by Tenhunen *et al.* (2008). In brief, the scintigraphy had been performed by using an intravenous injection of Tc-99m-pertechnetate (185 MBq) and a dynamic series of anterior images had been collected with Toshiba GCA-7200A/UI gamma camera (Toshiba Corporation, Tokyo, Japan) equipped with a low-energy high-resolution collimator. Regions of interest (ROI) had been placed on each major salivary gland and on the background. The individual salivary gland ejection fraction (EF) describing the magnitude of salivary secretion due to stimulus was calculated from the ROIs. Stimulated saliva production had been achieved by using 20-40 ml of 25 % lemon juice administered 15 min after the tracer injection. The assessment of the EFs had been performed before radiotherapy and 6 months after the therapy for each patient.

Major salivary glands without observable pretreatment function were excluded from the analysis. The volumes (mean ± SD) of the included parotids were 17.0 ± 7.3 cm^3 (range: 7.1-32.4 cm^3, *N*=41) and those of the submandibular glands were 4.9 ± 1.9 cm^3 (range: 2.8-9.1 cm^3, *N*=17).

4.1.4. Determination of patient set-up errors and their effect on major salivary gland mean dose

Determination of set-up errors and their effect on dose

A stereotactic head and neck immobilization device was used (BrainLab, Heimstetten, Germany). The isocenter localization was based on stereotactic coordinates computed by using BrainLab stereotactic TPS. Positional shift of a salivary gland with respect to the therapy beams was approximated from the positional shift of the bony structures near the gland. The shift was defined as the difference between their locations in two orthogonal portal images and two reference X-ray images taken in connection with the treatment simulation. The shifts were obtained in three orthogonal dimensions (1D) and they were combined to obtain also a three dimensional shift (3D). Portal imaging was carried out in two first treatment fractions and, thereafter, at least weekly for each patient. Mean of set-up variations during treatment indicating systematic set-up error that would have occurred without correction of patient positioning and standard deviation (SD) of the set-up variations indicating random set-up error were calculated for each patient. Rotational shifts were converted into translational shifts and were not explicitly presented. At the time of the collection of the patient data, a systematic positional shift in one of the three orthogonal directions (1D) ≥ 3 mm detected by the portal images was verified by resimulating the patient positioning and when necessary by correcting the positioning with respect to the stereotactic coordinates.

The effect of patient set-up errors on the mean doses of the major salivary glands that would have occurred without corrections of patient positioning was estimated by using the TPS. The planned isocenter was moved according to the obtained systematic shift and the dose distributions within the glands were recalculated without changing the field fluence distributions and the number of monitor units.

Treatment planning

The IMRT technique has been previously described in detail (Saarilahti *et al.* 2005, 2006). Dose distributions were calculated by using the single PBC algorithm implemented in Eclipse® TPS software (Varian Medical Systems, Palo Alto, CA). Tissue inhomogeneity correction was carried out by using the modified Batho method (Batho 1964, Young and Gaylord 1970). After conventional TPS configuration, a proper width of the PBC kernel, crucial for accurate dose calculations in IMRT, was readjusted and confirmed through film measurements of treatment fields to produce acceptable dose calculation accuracy for IMRT test fields (Van Esch *et al.* 2002). All 25 patients were treated with the same Varian Clinac 600 CD (6 MV) linear accelerator (Varian Medical Systems, Palo Alto, CA) equipped with Millenium 120 MLC (leaf width 0.5 cm).

4.2 ANALYZING METHODOLOGY

4.2.1. Analysis of quality control measurements

Short-term behaviour of an output level

A systematic short-term (time period between two successive output measurements) drift in an output level of a linear accelerator was determined by fitting a linear model

$$g(t) = a' + b't \qquad (13)$$

to the results of the CCs (Figure 4-2), where a' and b' are the fitting parameters and t is the time. The results of the CCs were corrected for a systematic drift in the calibration of the CC device. Each obtained linear trend was compared to a trend formed by the successive output measurements.

For statistical analysis of output behaviour, empirical probability distributions were constructed for the obtained linear output trends for both low and high photon energy regions. The probability of a trend was determined by dividing its duration by the sum of the durations of all trends. Total duration of the trends was about 8400 and 4300 days for 6 and 15-18 MV, respectively.

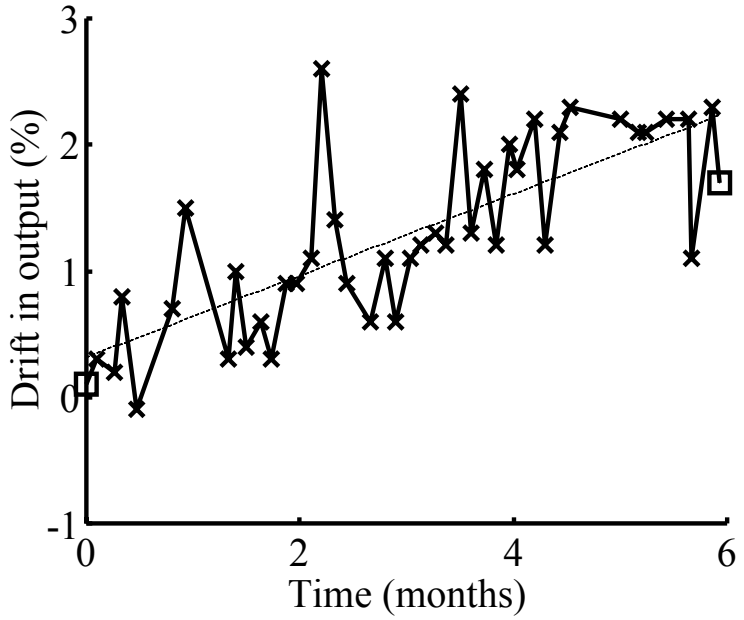

Figure 4-2. *According to the constancy checks ('x's with a solid line), a systematic short-term drift in an output level was assumed linear (dashed line) between two successive output measurements (squares).*

Long-term behaviour of an output level

A 'free' cumulated systematic long-term (time period of more than two successive output measurements) drift in an output level, which most likely would have occurred without any dose adjustments, was estimated for each accelerator and photon energy. This was done by arranging consecutively the linear short-term trends formed by two successive output measurements. The linear model (Eq. (13)) or a non-linear model with an empirically determined form was fitted to each constructed data set.

Since the model with most fitting parameters is not necessarily the best one due to random errors in the data, the model chosen for a data set gave minimal value for the Akaike information criterion AIC (Akaike 1974) defined as

$$\mathrm{AIC} = N \ln[\sum (\Delta err)^2] + 2\beta \qquad (14)$$

where β is the number of parameters ($\beta = 2$ for $g(t)$), N is the number of data points and $\Sigma(\Delta err)^2$ is the sum of the squared fitting errors of a model.

If the monitor chamber of an accelerator was replaced, a construction of new set of data was started from that day on. If a dosimetrically significant repair or an adjustment was carried out between successive output measurements (such as adjustment of beam profile, energy, dose rate, RF-power or RF-frequency), a trend in output for that time interval was determined by fitting the linear model to the results of the CCs measured before that repair or adjustment.

As complementary results to those presented in the publications, the efficacy of the empirical model fitting of the QC measurement results was investigated in the improvement of dose accuracy. The reduction of random errors was evaluated when a current measurement result is estimated and future value predicted from the fitted model. These were carried out by simulating random measurement errors by using two different measurement reproducibility levels and by model fitting the simulated data. Linear time trend was assumed in the fitted parameter. Finally, the effect of potentially reduced random measurement errors on the deviation of output was estimated.

Reproducibility of output measurements

The reproducibility of the output measurements was estimated by using two different approaches: i) from the SD of the fitting errors of the long-term models of the output and ii) from the SD of differences between comparative measurements of the hospital and the STUK carried out on the site visits by using independent measurement devices and measurement set-ups. The former method i) combines the uncertainty related purely to the measurements (equipment, conditions and procedure) and random drifts in accelerator output. The latter method ii) assumes comparison of two independent output measurements. The method can be simplified by assuming either the same reproducibility level for the two measurements when the reproducibility (1SD) is obtained by deviding the obtained SD by $\sqrt{2}$, or more cautiously that comparative measurements include no errors when the reproducibility is obtained directly from the obtained SD.

Due to limited number of data points used for the model fitting, random measurement errors have an effect on the fitted model and the SD of the fitting errors does not necessarily correspond exactly to the SD of random errors (defined as measurement reproducibility). The correspondence between these SDs and considerations for sufficient number of data points needed to fit the linear model were investigated by simulating measurement reproducibility levels of 0.5 and 1.0 % (1SD) and by model fitting the simulated data. The estimation is for linear parameters changes and is provided for the output measurements, CCs and calibration coefficient of the CCs. These results have not been published earlier.

Feasibility of constancy checks

The reproducibility of the CCs was estimated from the SD of the fitting errors of the linear model fitted to the CC results. The obtained reproducibility includes short-term variations in output and in calibration of a CC device. An empirical distribution was constructed from the fitting errors to quantify the scatter of the CCs around the model. The reproducibility of the CCs was estimated for two different positioning methods of the CC devices: by using an optical scale of the accelerator or by using positioning lasers.

A long-term stability of the CCs was estimated by quantifying the time behaviour of the calibration factor for each CC device. This was done by the empirical model fitting.

Measurement of energy parameters

Time behaviour of the PDD at the depth of 10 cm and J_{20}^{10} were determined for the reference FS of 10x10 cm^2 at 100 cm SSD by using the empirical model fitting as described above. In addition, the reproducibility of the measurements was assessed for these parameters from the

SD of the fitting errors as described above. Default settings of the software were used for the data smoothing.

4.2.2. Estimation of efficacy of quality control programs

A method for the estimation of treatment dose variations due to output shifts

A method based on MC principle was developed to assess the accuracy of total treatment dose (D) in external photon beam radiotherapy. The method was used to estimate the variations in treatment dose (ΔDs) due to shifts in output level with respect to the predefined reference dose ($\Delta D = 0$) achievable with the use of a QC program for output, i.e. a certain output measurement time interval and action levels for the output and CC measurements (Figure 4-3). The probability that an output is outside of the predefined action levels can be estimated for different QC programs. The method takes into account the stability of the output and random errors (reproducibility) of the output measurements and the relative CCs. These are given in the form of empirically determined or simulated probability distributions. The method was accomplished with MATLAB^{TM} (The MathWorks Inc., Natick, MA).

During a treatment course, an output level was assumed to change linearly and the output stability is generated from a distribution of linear output changes. Two different practices of output level adjustments were simulated: i) the *'normal case'*, when an output exceeding the action level is adjusted to a measured reference value; ii) the *'anticipatory case'*, when a rising output level is adjusted to a value below a measured reference level and a decreasing output level is adjusted to a value above a measured reference level.

The simulation method assumes one type of output measurements making no difference between the definitive and routine calibrations. Both of these are carried out by using the same reference ionization chamber and the relation between the definitive calibrations carried out in a water phantom and the routine calibrations carried out potentially in a PMMA or solid water phantom is assumed accurate. The relation should be determined locally and the corresponding uncertainty should be added to the estimated overall uncertainty.

The performance of relative CCs several times a week was included in the simulation. If the average level of the CC results exceeded the predefined check action level, an output measurement was carried out. The check device reading was calibrated against every output measurement. Therefore the error of the check calibration factor includes random errors of both output and CC measurements.

Reproducibility of output measurements

The simulations were carried out for two different levels of output measurement reproducibility. For the best accuracy level, random errors were generated from an empirical distribution with a SD of 0.7 % and a 95 % CI of - 1.4 … + 1.4 % being consistent (within 95 % CI) for the reproducibility obtained in the Study I. Worse measurement accuracy level was estimated theoretically by widening the empirical distribution obtained in the Study I resulting in a SD of 1.4 % and a 95 % confidence interval of - 2.8 … + 2.8 % corresponding to estimated uncertainty of output measurements in user's beam made by IAEA in the TRS-398 (IAEA 2000). The empirical distribution was considered more appropriate than e.g. the Gaussian distribution because the empirical distribution lacks very large measurement errors

(> 3.6SD). Such errors occur very rarely, they are usually indentified and the estimation of probability for undetected large errors is uncertain.

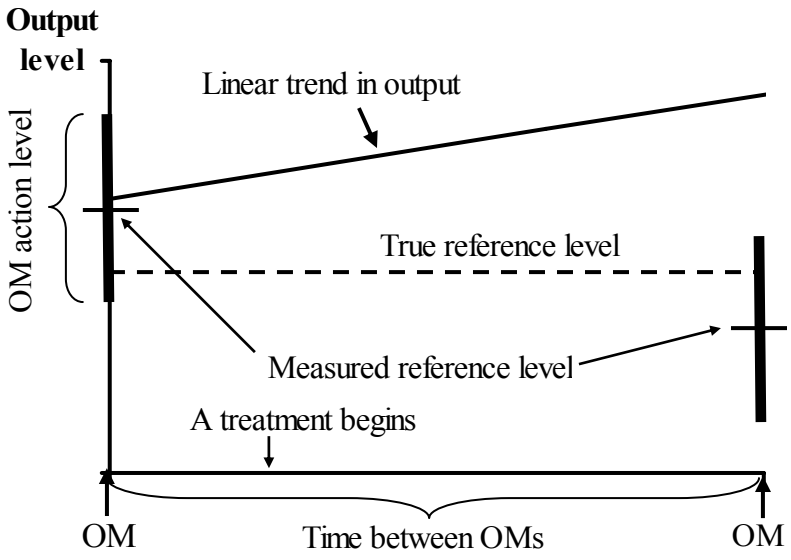

Figure 4-3. *The simulation of total dose (D) received in a treatment. A treatment begins at a random time point between two successive output measurements (denoted as OMs). A measured reference output level differs from the true reference level (ΔD = 0) due to a random OM error. When 'anticipatory' output adjustments are used, the output level at the OM, carried out prior to the beginning of the treatment, is generated uniformly within the predefined OM action levels. In the 'normal case', this output level is generated uniformly within the upper half (e.g. [0 ... +2 %]) in the case of a rising output level and within the lower half (e.g. [-2 ... 0 %]) in the case of a decreasing output level. A linear trend assumed in the output during a treatment time is generated randomly from the distribution given. If the action level is exceeded in the subsequent OM, as is the case in this figure, the output level is adjusted to a measured reference level. An average treatment time of 1.5 months was assumed for the calculations (30 fractions).*

Reproducibility of constancy checks

In the Study I, the distribution of random errors was estimated for the CCs having a SD of 0.5 %, when the device was positioned conventionally by using an optical scale. An independent effect of long-term instability of ± 1 % in a year (reported recently by several manufacturers of the CC devices) was approximated by widening the empirical distribution resulting in a SD of 1.1 % and a 95 % confidence interval of - 2.1 … + 2.1 %. The investigation was limited within the range of output measurement intervals from 0.25 to 6 months to minimize risk of too optimistic estimations due to larger long-term instability than assumed here.

Output stability

The simulations were carried out for different levels of output stability to achieve estimations for different practical situations. The estimations made for linear output changes of 0, 0.3, 0.6 and 1.2 %/month offer approaches for individual accelerators. The fully stable case (trend

equals to zero) was included to estimate dose errors merely due to the reproducibility of the QC measurements. In addition, the simulations were carried out for the distributions of linear output changes determined in the Study I to approach the spread of outputs more realistic during a long-term period for a group of accelerators.

The estimation of the spread of outputs achievable by using a QC program was carried out cautiously by assuming that the maximal trend in the empirical data (being 1.2 %/month) is just within 95 % confidence interval of the range of output trends observed in practice. For the simulation, a Gaussian distribution of output trends with a mean of zero and a SD of 0.6 %/month was used.

Assessment of biological influence

The variations of treatment doses (ΔDs) were converted into variations of TCP (ΔTCPs) to assess biological effects related to the use of a QC program and also the other factors investigated in this thesis. The linear conversion was assumed sufficient due to relatively small ΔDs and the ΔTCPs were obtained by using the Equation 3. A normalized steepness value $\gamma = 2.7$, reported by Okunieff *et al.* (1995), was chosen to describe average tumor dose response.

The dose-response characteristics of normal tissue complications are very different depending crucially on the realization of treatment (e.g. irradiated tissue type, organ, irradiated volume and dose distribution). Therefore, numerical estimations related to the NTCP included in this thesis concerns only the dose response of the major salivary glands in the head and neck IMRT (investigated in Study IV). For known dose-response steepness, however, one can estimate an increase in NTCP from the results presented for the ΔDs.

4.2.3. Evaluation of beam data

Determination of dose calculation accuracy

Dose calculation accuracy for open rectangular fields in Finnish RT centres was estimated from the dose measurements carried out by STUK. Difference between the measured and calculated dose was corrected for a shift in accelerator output level by subtracting the difference between the measured and reference values of output calibrations (Gy/MU). Therefore, the difference between the measured and calculated dose was considered to reflect the calculation accuracy. Possible misinterpretations due to random measurement errors were minimised by comparing the results obtained on at least two successive site visits within two years. Calculation accuracy of open rectangular fields was considered to reveal the quality of basic beam data used for the TPS configuration according to the Equations (9)-(10). The factors for the correction of different geometries used for the basic beam data and absorbed dose measurements (the Equations (11)-(12)) depend on the ratios of PDDs were assumed to give negligible contribution to calculation errors.

Averaged data sets for OF and PDD

For the identification of reasons for poor calculation accuracy exceeding the limit of ± 2 %, averaged reference data sets were constructed for both OFs and PDDs by using the data collected from the 9 RT centres. The local beam data were compared to the obtained averaged data to investigate whether potential differences explain the poor calculation accuracy. The

construction of averaged data sets was motivated by several investigations suggesting consistent values of the basic beam data for accelerators of the same manufacturer, model and energy (Fontenla *et al.* 1992, Watts 1999, Followill *et al.* 2004, Cho *et al.* 2005).

Reference data sets were constructed for the OFs, defined at d_{max} (for each field size) measured at 100 cm SSD. Averaged PDD data were constructed at a 10 cm depth for data measured at 100 cm SSD. Only the OFs and PDDs giving high dose calculation accuracy within ± 1 %, fulfilling the calculation accuracy recommended for open square fields, were averaged to construct accurate reference data sets. PDD data were available only from five out of the nine centres. The presentation of averaged data requires the use of beam QI. Since J_{20}^{10} is usually measured in clinical practice instead of TPR_{10}^{20}, the accuracy of the empirical relation between these parameters given in the Equation (5), was estimated through measurements.

Averaged data sets for reference MUs

The number of MUs calculated by local physicists were corrected to result in accurate dose delivery. First, the calculated MUs were converted into normalised MUs by multiplying them by the output calibration factor (Gy/100 MU). These normalised MUs simply corresponded to the numbering of MUs required when the output is calibrated to 1 Gy/100 MU. Then, the normalised MUs were corrected to give an exact dose delivery of 2 Gy. This was done by scaling the result of the dose measurement (corrected for output level shifts) and correspondingly the normalised MU count. The averages of the corrected MU counts from 24 accelerators were used for the reconstruction of averaged MU data sets. Only two Clinac 600 Cs were excluded from the analysis due to non-standard shadow trays used in the beam. Basically the MUs can be considered as OFs defined at 10 cm depth in an isocenter.

Parametrization of beam data

The averaged OF and MU data sets were parametrised for the open square fields in order to present them in a compact way. A non-linear model function with an empirically determined form was used.

In order to parametrise OFs for rectangular open fields, the feasibility of an empirical equivalent square field formula of (Vadash and Bjärngard 1993)

$$C = \frac{(A+1)XY}{AX+Y} \tag{15}$$

obtained for head scatter factors, was evaluated, where A is an empirical parameter and C x C is the equivalent collimator opening for a field XxY. The combinations with X,Y = 5, 10, 15, 20 and 35 cm were used to determine A.

4.2.4. Estimation of accuracy requirements based on dose response

Modelling of dose response

Volume based (Lyman) normal tissue complication probability model (Lyman 1985) was chosen for the describtion of salivary gland function after RT. Mean dose (D_{mean}) within a gland was chosen as a dosimetric measure for the biological effect since parotid glands have been shown to behave like parallel organs (Eisbruch *et al.* 1999). The ratio (rEF) of EF values obtained for a gland before (t = 0) and 6 months after RT was assumed to follow sigmoidal dose response relationship as

$$\mathrm{rEF} = \frac{\mathrm{EF}(6\,\mathrm{months})}{\mathrm{EF}(0)} = \frac{1}{1+\left(\frac{D_{mean}}{D_{50}}\right)^{\Gamma}} \qquad (16)$$

where D_{50} and Γ are fitting parameters. The data obtained for the parotid and submandibular glands were combined since no significant differences have previously been observed in the dose response characteristics of these glands by using the rEF as an endpoint (Tenhunen *et al.* 2008). In addition to the sigmoidal dose response model, a linear model was fitted to the data points forming the steepest slope region.

Estimation of dosimetric and geometric accuracy requirements

To predict rEF at an accuracy level considered clinically significant, ΔrEF = ± 0.1, analogous to ΔNTCP = ± 10 %, was chosen. Accuracy criterion for D_{mean} was estimated by using the maximal slope of the obtained sigmoidal dose response curve. The criterion of D_{mean} was converted into geometric accuracy requirement by using the average 3D dose gradient value within the glands separately for the parotid and submandibular glands.

5. RESULTS

5.1. ANALYSIS OF QUALITY CONTROL MEASUREMENTS

5.1.1. Observed time trends

Short-term output

The obtained probability distributions for linear short-term output trends are presented in Figure 5-1. The output was observed to increase with time for the Clinac 2100 C/CDs and to decrease for the Clinac 600 C/CDs. Only very shallow negative output trends were observed for the Clinac 2100 C/CDs. The steepest trends were about 1 %/month for both high and low photon energy regions, and were observed for new accelerators just after commissioning. During the observation period 3 new accelerators were commissioned. High probability for very shallow trends demonstrates the tendency of the outputs to stabilize with time.

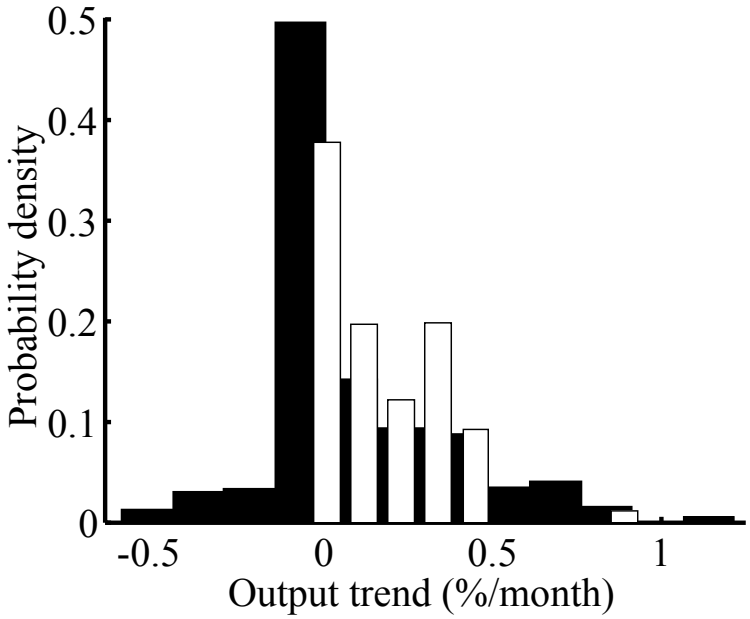

Figure 5-1. *The empirical probability distributions of linear short-time output trends constructed from the quality control data collected for 6 MV (black bars) and for 15-18 MV (white bars). For 6 MV, the mean ± SD of the output trends, weighted by their relative durations, were 0.05 ± 0.27 %/month. For 15-18 MV, the corresponding values were 0.15 ± 0.19 %/month.*

Long-term output

The time behaviour of output was highly individual and depended on the age of the accelerator (Figure 5-2). An exponential model of form

$$f(t) = a'' + b'' \exp(c''t) \qquad (17)$$

was empirically found suitable for non-linear outputs, where *a´´, b´´* and *c´´* are the fitting parameters. The exponential model was preferred only for 2 new accelerators with long observation time (≥ 40 months). The linear model was sufficient for new accelerators or monitor chambers with short observation time (≤ 21 months) and for older accelerators. More detailed results of the long-term output modelling are summarized in Study I (Table 1). Maximal fitting errors of the preferred long-term output models ranged normally from 0.1 to 1.0 % with a mean of 0.7 %. The results of three output measurements (2.4 % of the total) were clearly outlying (1.2, 1.4 and 2.1 %) from the time trends formed by other output measurements and CCs due to unknown reasons.

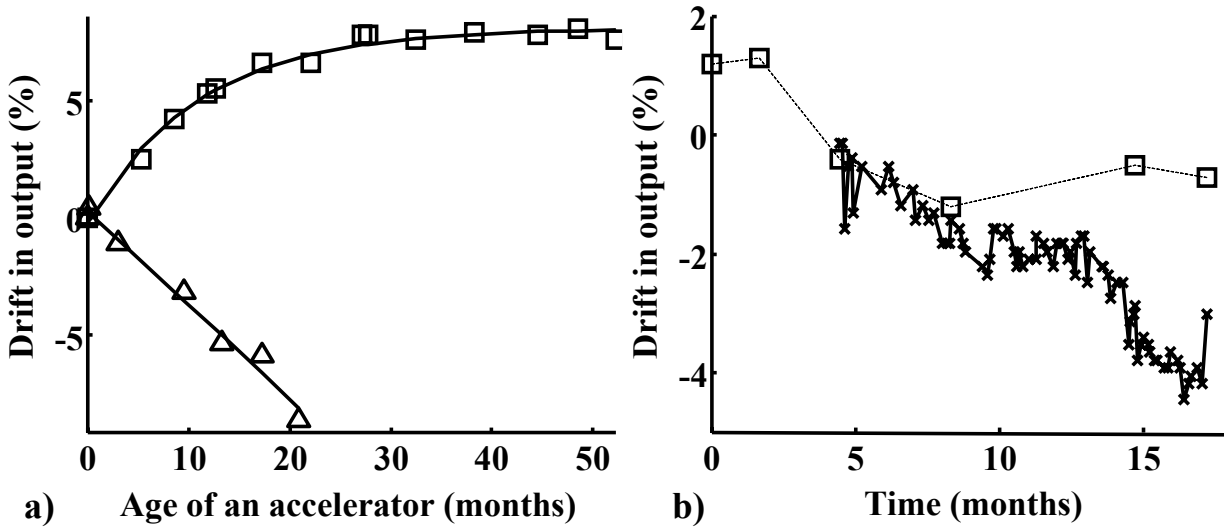

Figure 5-2. *Systematic cumulated long-term drifts in output levels. Accelerators with sealed monitoring chambers: a) Example of both an exponentially increasing level (output measurements illustrated with squares) and a linearly decreasing level (output measurements illustrated with triangles). The proposed models (solid lines) describe well the constructed drifts. An accelerator with an unsealed monitoring chamber: b) According to CCs (stars with a solid line) and output measurements (squares with a dashed line), a periodic behaviour (period about a year) due to environmental conditions was observed. Differences between the output and CC measurements increase significantly with time demonstrating great systematic long-term drift in the stability of the CC device (regular recalibrations removed). It should be reminded that the constructed output drifts do not describe dose errors in treatments because of dose adjustments.*

The average time for an output drift of 2 %, corresponding to the most commonly used action level for the output measurements, was 18 ± 12 months (averaged for the beams listed in Study I, Table 1). The non-linearly changing outputs of 2 new accelerators showed steep upward trends just after commissioning reaching the drift of 2 % even within 2 - 3 months. The nonlinear outputs stabilized essentially during the first two or three years.

With one Clinac 600 CD, one time period was excluded from the long-term modelling because the quantitative determination of output trends was greatly uncertain due to frequent dosimetrically significant repairs and adjustments (magnetron replacements, adjustment of

beam energy and dose rate). During this period, the CCs indicated exceptional output behaviour being both greatly increasing and decreasing. Before and after this period, however, the output was again decreasing and could be modelled, consistently with the other 600 CDs.

Unfortunately, only short data sets were collected for an accelerator with unsealed monitor chamber (SL 18) because of multiple monitor chamber and magnetron replacements during the observation period. These data sets indicated periodic output behaviour due to environmental conditions (humidity) with amplitude of about 1 %. The effect seems to limit the accuracy of output modelling and prediction of output level for accelerators with unsealed monitor chambers.

Stability of constancy checks

The calibration factors of two out of the four CC devices increased linearly during the whole observation period (1-3 years) at the rate of 0.30 ± 0.08 %/month and 0.29 ± 0.04 %/month for 6 MV and 15-18 MV, respectively, (correlation coefficient $|\rho| > 0.95$). The effect of the drift in the calibration is demonstrated in Figure 5-2 b). The calibration factors of the rest two devices remained constant with time ($|\rho| < 0.7$). The SD of the relative fitting errors of the linear models was 0.6 %.

Beam energy parameters

The energy parameter J_{20}^{10} was constant ($|\rho| \leq 0.5$) for all the beams during the observation period. Consistently, the PDD at 10 cm depth for the reference field was constant ($|\rho| \leq 0.7$).

5.1.2. Reproducibility of measurements

Output measurements

The SD of the differences between the comparative measurements of the hospital and STUK was 0.37 % resulting in a rough estimation for the reproducibility of the local output measurements carried out at the depth of 10 cm. By incorporating the uncertainty of the PDDs measured at the 10 cm depth being 0.30 %, the estimated reproducibility of the local output measurements becomes about 0.5 % at the d_{max}.

The SDs of the fitting errors of the long-term output models were 0.4 and 0.5 % for 6 and 15-18 MV, respectively. According to the simulation of random measurement errors, the estimation of output measurement reproducibility from the SD of the fitting errors results in systematic underestimation depending on the number of data points used to fit the models. For 10 data points, as was the case on the average in the current analysis, the underestimation is -0.05 ± 0.12 % and -0.09 ± 0.23 % for measurement reproducibility levels of 0.5 % and 1.0 % (1SD), respectively. By considering these, the estimated uncertainty of the local output measurements becomes about 0.5 ± 0.2 % when presented with 95 % CI. Importantly, the both included methods resulted in consistent values for the reproducibility of the local output measurements.

If accuracy level of 0.1 % is desired for the determination of measurement reproducibility, many fitted data sets with minimal number of included data points of 10 are needed. The importance of the number of data points used for a modelled data set is emphasized by realizing that for 5 data points the underestimation of the reproducilibity is significantly

greater, being -0.10 ± 0.17 % and -0.20 ± 0.33 %, for measurement reproducibility levels of 0.5 and 1.0 %, respectively.

Constancy checks

The reproducibility of the CCs was 0.5 % when the device was positioned conventionally by using optical scale while it was improved to 0.2 % when the device was positioned by using lasers. The obtained distributions of random CC measurement errors are presented in Study I (Figure 4). Comparison of these values suggests that the reproducibility of the CCs is predominated by random distance errors of the positioning of a device.

The uncertainty of a CC calibration factor includes the uncertainties of both output and CC measurements and two successively determined calibration factors differed often more than 1 % from each other (Figure 5-3). Therefore an output trend can be determined best by using the same calibration factor or by using the readings of a device instead of calibrated values.

Beam energy parameters

The SD of the measured PDDs at a 10 cm depth for the reference 10x10 cm^2 field was 0.3 ± 0.1 %, ranging from 0.2 to 0.4 % for individual accelerators. The SD of the J_{20}^{10} s was 0.5 ± 0.2 %, ranging from 0.2 to 0.7 %. The SD of d_{max} was 0.5 ± 0.2 mm for 6 MV and 1.2 ± 0.3 mm for 15-18 MV. It should be realized that these values depend on smoothing method used for the measured data.

5.1.3. Methods proposed for quality control

The proposed single-exponential and linear models described normally (without dosimetrically significant malfunctions, adjustments or repairs) well the constructed systematic long-term drifts in output while linear model was sufficient for short-time drifts. The proposed methods proved well suited for the estimation of parameter changes and measurement reproducibility. The model fitting of the QC results seems to facitate the regognition of outlying measurements reducing the number of unnecessary adjustements of output level.

The determination of systematic drift in CC device reading is crucial in the confirmation of a good clinical dose control and in the determination of output trends according to the CCs. The robustness of the CC calibration seems to be improved by comparing a newly obtained calibration factor to a value determined by model fitting of the previous values (Figure 5-3). For this ‘predicted’ value, the effect of random errors is minimized (provided that enough data is fitted) and thus the comparison may reveal the cases when the level of the CC results would ‘jump’ away from a true output level due to an erroneous calibration factor. The accuracy of the calibration factor may not necessarily be improved by simply averaging the previously determined values because a device reading may drift systematically with time.

The result of an output measurement could usually be estimated within 0.5 % from the previous measurement by using the output trend determined by fitting a linear model to the CC results (with systematic long-term drift in device reading taken into account). This offers approach for estimation of individual time intervals for output measurements. The approach is independent of the uncertainty of the calibration factor of a CC device.

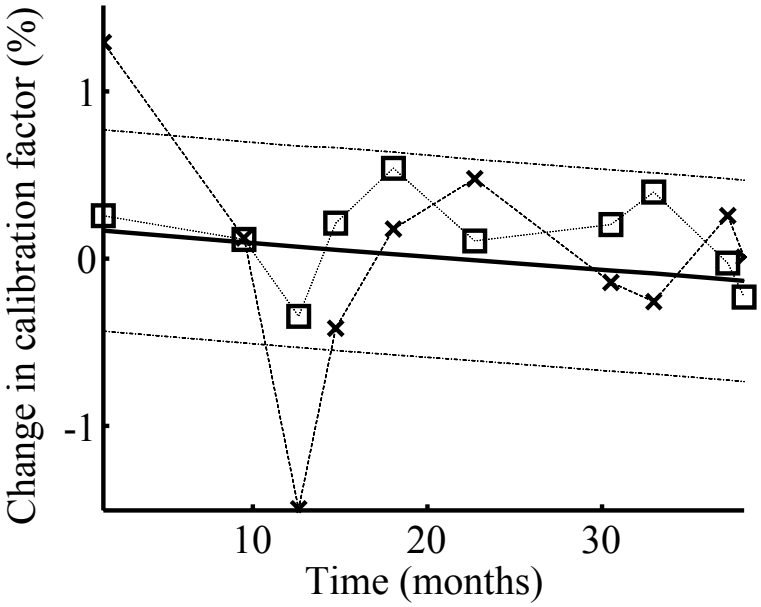

Figure 5-3. *The calibration factors obtained for a CC device in connection with output measurements. The values are given as differences from the average value. The original values are indicated with 'x's, while the squares indicate the obtained values when random errors of both output and CC measurements are estimated and corrected. The errors were estimated by using the proposed output modelling and by averaging the first 5 check readings after the calibration, respectively. Large random errors exceeding ± 0.6 % (1SD, horizontal dashdotted lines) in some calibration factors (the 1st and 3rd 'x' from the left) were also recognized by fitting a linear model (solid line) to the obtained (uncorrected) calibration factors demonstrating the usefulness of empirical model fitting.*

5.2. EFFICACY OF QUALITY CONTROL PROGRAMS

Achievable deviation of outputs

For a given QC program, the 95 % and 99 % CI limits of Δ*D*s were similar for all the included distributions of output variations. In contrast, for single output trends, great dependence on the steepness of a trend was observed showing that the average of output variations has more prominent effect than their deviation. The limits for the measurement interval of 3 months were usually about two times as large as those estimated for the interval of 1 week. The effect of output stability on the Δ*D*s practically vanished when the measurements were carried out weekly and outputs could be kept within a spread defined by the action levels.

In the case of the steepest observed output trends of up to 1.2 %/month, the use of the measurement interval of 3 months and action level of ± 2 % kept an output practically within about ± 4.0 % or ± 5.5 % (99 % CI) for the measurement reproducibility level of 0.7 % or 1.4 %, respectively. By using the interval of 1 month, these limits were about ± 3.0 % or ± 4.0 %, respectively. The probability of steep output changes is normally very small and they are detected by the CCs and, therefore, the output is kept within narrower limits as estimated in

this ‘worst case scenario’. All the presented 95 and 99 % limits of the Δ*D*s are one-directional (in the direction of trend).

The effect of use of different action levels is demonstrated in Figure 5-4. Because the included distributions of output changes gave quite similar results, it is most likely that these 95 % limits provide generally feasible long-term estimations for the Δ*D*s. Results for linear output trends are presented in the Study II (Table 1).

The effect of anticipatory adjustments for output levels was relatively small on the 95 and 99 % limits of the Δ*D*s. By using such adjustments, the limits could be reduced only by about 0.3 % when a measurement interval from 1 to 3 months was used and the measurement reproducibility was 0.7 %. If measurement reproducibility was 1.4 %, these limits could be reduced by about 0.4 %.

For the empirical distributions of output variations, the SD of Δ*D*s was about 1.5 %, 1.1 % and 0.8 % for the measurement intervals of 3, 1 and 0.25 months, respectively, when the output measurement reproducibility was 0.7 %. For the theoretical Gaussian distribution of output variations together with measurement reproducibility of 1.4 %, the use of these measurement intervals resulted in the SD of about 2.0 %, 1.4 % and 0.9 %, respectively.

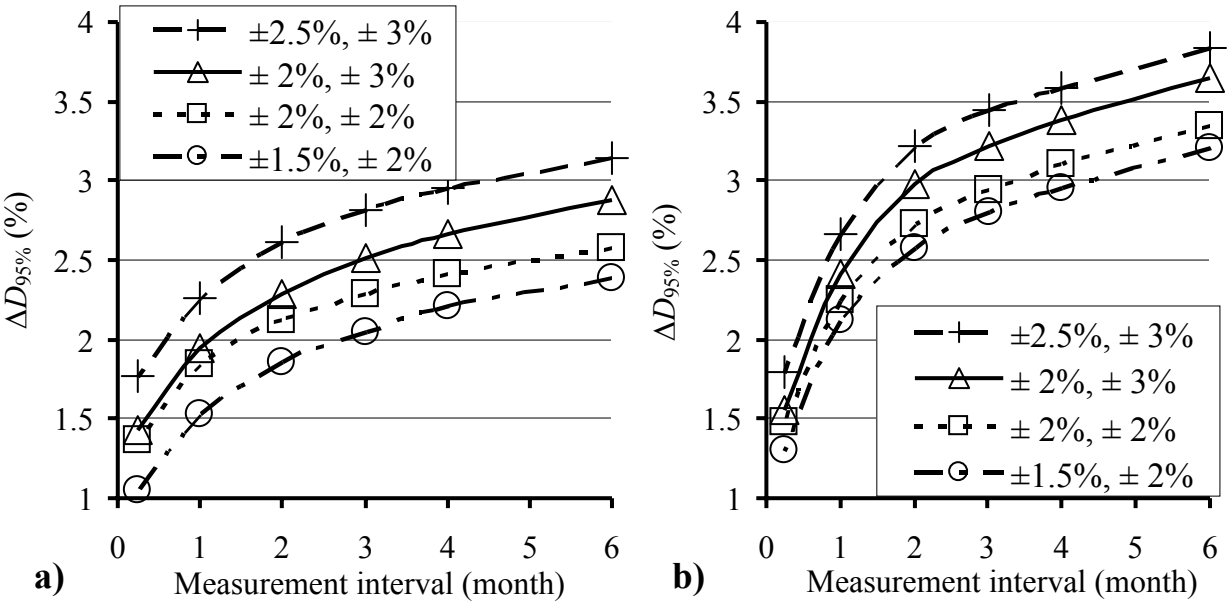

Figure 5-4. *The upper 95 % limit of ΔDs estimated for empirical distribution of linear output changes determined for the 15-18 MV beams. The estimations were carried out for the output measurement reproducibility levels of a) 0.7 % and b) 1.4 %. The action levels of the QC measurements are expressed in the figures as: ± OM %, ± CC %, where OM is the action level for the output measurements and CC is that for the constancy checks. ‘Normal’ practice of output adjustments was assumed, where output is adjusted to a measured reference value.*

Probability of unacceptable treatments

The probability of unacceptable treatments (defined as $|\Delta D| > 2$ % due to output shifts) achievable with the use of a QC program is summarized in Study II (Table 2) for different levels of output stability and measurement reproducibility. The probabilities depended crucially on both the output stability and the reproducibility of the output measurements. The fraction of unacceptable treatments could be significantly decreased by using anticipatory adjustments for output.

The influence of different combinations of output measurement time intervals and action levels are demonstrated in Figure 5-5. The presented curves have the greatest gradients when measurement interval is between 0.25 and 2 months, since the random errors of the QC measurements become more prominent when the number of output measurements during a treatment time decreases.

It was observed that the probability of unacceptable treatments remained equal when the time interval for the output measurements was doubled and the action level for the output measurements was lowered from ± 2 % to ± 1.5 % and the action level for the CCs from ± 3 % to ± 2 %. Moreover, the measurement interval of 3 months could be doubled to 6 months if merely the action level of the CCs was lowered from ± 3 % to ± 2 %. In these cases, the spread of output (99 % limit of ΔDs) widened about 0.4 % or 0.7 % when the output measurement reproducibility was 0.7 % or 1.4 %, respectively. The use of slightly relaxed action level of ± 2.5 % for the output measurements increased the probability of unacceptable treatments to the level achievable by using a doubled output measurement time interval.

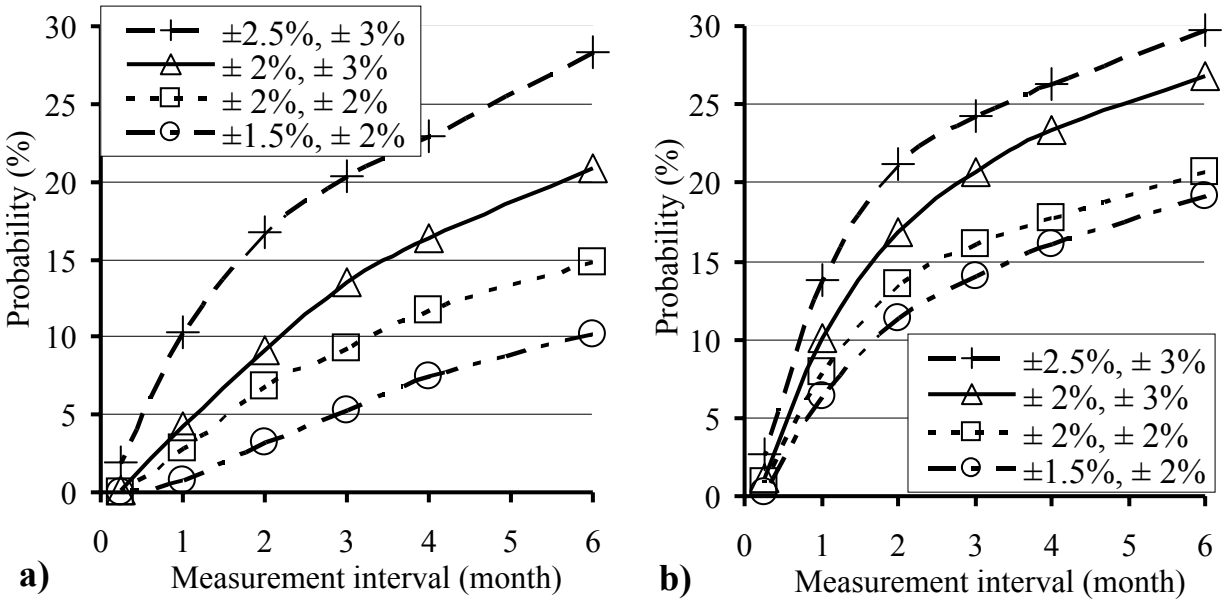

Figure 5-5. *The probability of unacceptable treatments with $\Delta D > +2$ % estimated for the empirical distribution of linear output trends determined for the 15-18 MV beams. The estimations were carried out for the output measurement reproducibility levels of a) 0.7 % and b) 1.4 %. The action levels of the QC measurements were expressed in the figures as: ± OM %, ± CC %. 'Normal' practice of output adjustments was assumed.*

The 95 and 99 % limits of the Δ*D*s and the probabilities of unacceptable treatments are more practical measures for achievable dose accuracy than the SDs of the Δ*D*s. This is because the SDs may be small for single output trends but the mean of Δ*D*s may be considerable different from zero. Biological risks related to tumour recurrence or normal tissue complications, related to the use of a QC program for output, can be assessed from the results presented in this chapter.

5.3. ACCURACY OF BEAM DATA

5.3.1. Finnish survey

The differences between measured and calculated central axis dose are summarized for the Finnish RT centres in Figure 5-6. For all the test fields, the fraction of fields with a delivered dose within ±1, ±2 and ±3 % was 79, 97 and 99 %, respectively. For the elongated rectangular field of 5x30 cm^2, a systematic underdosage of about -0.8 % was observed and the fraction of fields with a delivered dose within ±1, ±2 and ±3 % was 63, 96 and 100 %, respectively. For the worst observed cases, calculation accuracy could be improved by 2-3 % when compared to the most accurate dose calculations observed. The results of the centres are summarized in detail in Study III (Table 1).

The measured and calculated dose agreed within $\pm$1 % in all test fields for 7 (25%) and 6 (30%) of the 4-6 MV and 10-18 MV beams, respectively. The limit $\pm$1 % was exceeded only for the 40 x 40 cm^2 field or the extremely elongated fields (5x30 cm^2 or 30x5cm^2) for 8 (29%) and 6 (30%) of the 4-6 MV and 10-18 MV beams, respectively. A tolerance of ± 2 % was exceeded at least in one test field for 4 out of the 28 investigated beams of 4-6 MV (14 %) and for 2 out of 20 beams of 10-18 MV (10 %).

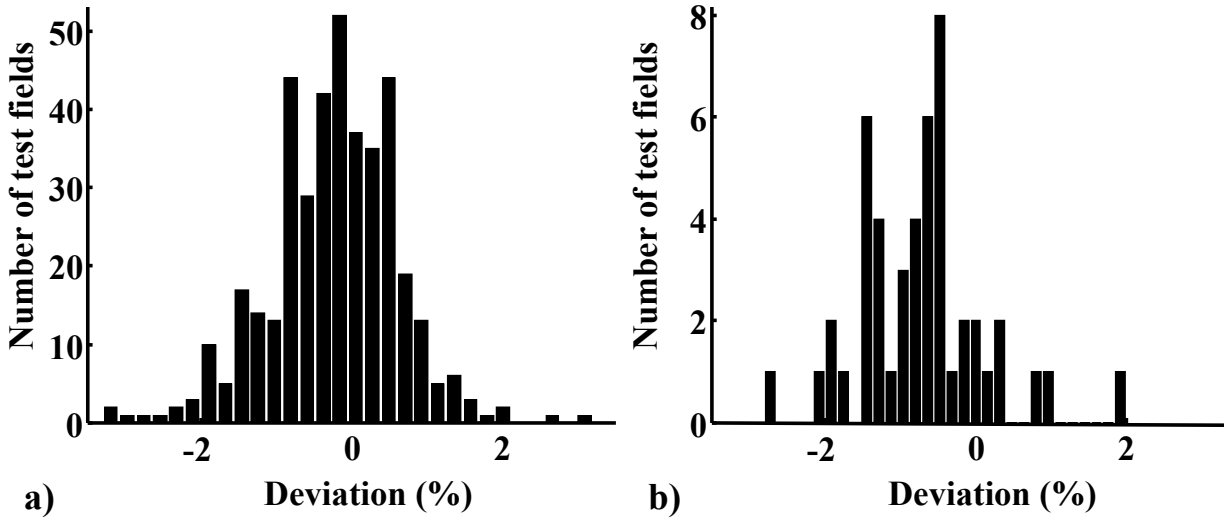

Figure 5-6*. The difference between measured and calculated central axis dose at a 10 cm depth for the open test fields. a) The distribution of differences for all fields (N=403). The mean ± SD were -0.2 ± 0.9 %. b) The distribution of differences for elongated rectangular fields of 5x30 cm^2 (N=48). The mean ± SD were -0.8 ± 0.9 %.*

5.3.2. Method proposed for quality control

Consistency and parametrization of beam data

The numbers of MUs corrected to result in correct isocentric dose delivery of 2 Gy at a depth of 10 cm (100 cm SAD) for the open square fields were equal within ± 1.0 % for all the 24 accelerators included. A non-linear function of form

$$h(X) = a + \frac{b}{(X-c)^d} \tag{18}$$

was empirically found suitable (accuracy within ± 0.3 %) for the parametrisation of the averaged MUs, where a,b,c and d are the fitting parameters, and X is the side of a square field given in cm. The fitting parameters for FS ranging from 5x5 to 40x40 cm^2 are presented in Study III (Table 2). The SD of the MUs was ≤ 0.6 % for the most FSs.

Where the agreement between measured and calculated dose was within ± 1 %, the OFs of the open test fields measured by the centres were in good agreement (within ±0.7%) with the averaged values. The parametrisation of the averaged OFs defined at d_{max} for the open square fields ranging from 4x4 to 40x40 cm^2 is presented in Study III (Table 3). The Eq. (18) was used (accuracy within ± 0.3 %).

Where the agreement between measured and calculated dose was within ± 1 %, the PDDs of the test fields measured by the centres were in good agreement (relative difference within ±1.2%) with their averaged values. The averaged PDDs are presented in Study III (Table 4). The PDDs showed about two times larger maximal deviation than the OFs. Due to relatively large deviation and scarcity of the PDD data provided, the averaged PDDs were published only for nominal energies with several investigated beams. Since the empirical relationship between J_{20}^{10} and TPR_{10}^{20} (Equation 5) was experimentally found to be accurate within ± 0.1 % for the investigated beams (except ± 0.3 % for J_{20}^{10} = 1.712), the TPR_{10}^{20} values for the beams were calculated by using this equation.

In the cases of good calculation accuracy within ± 1 %, the OFs for rectangular fields XxY at d_{max} could be determined for nominal energies 4, 6, 10 and 15 MV by using the equivalent square field relation (Eq. 15) with a maximal error of 0.7 % (SD ≤ 0.3 %). This accuracy was achieved for the Clinac 600 CDs and 2100 CDs by using A values of 1.31 and 1.53, respectively. The values of A depended on the accelerator model and not on the nominal energy. For 18 MV beams from Clinac 2100 CDs, the best value for A was also 1.53 but the maximal error increased to about 1.1 % (SD ≈ 0.4 %). The SD of the A values was ≤ 0.05 for all nominal energies. The number of the accelerators investigated is shown for each nominal energy in Study III (Table 3).

The use of reference data in identification of reasons for calculation disagreements

For two 6 MV and two 15 MV beams having maximal dose calculation error of about 3 %, the OFs used for the TPS deviated maximally by about 2.5 and 1.5 %, respectively, from the obtained averaged OFs (Figure 5-7). The correlation between the deviation and the error was

0.31 and 0.90 for the 6 and 15 MV beams, respectively. When the OFs were replaced by the averaged data, the maximal calculation errors decreased to about 2 % for all the four beams. For these beams, also the PDDs used for the TPS deviated maximally by about 2 % from the obtained averaged PDDs. The correlation between the deviation and the calculation error was 0.95 and 0.99, respectively. When also the PDD data were replaced by the averaged data, maximal calculation errors decreased to about 1 % for all the four beams, revealing the existence of significant errors in beam data.

For one 6 MV beam of Clinac 600 CD, dose calculation errors of up to about 3 % were observed. The TPS was configured by using OFs measured for a 6 MV beam of Clinac 2100 CD. The average OFs of 2100 CDs differed from those of 600 CDs maximally by about 2.2 % (even when the beam energy parameters were equal) explaining the observed calculation error ($|\rho|$=0.97).

For one 6 MV beam of Clinac 2100 CD, calculation error of 2.3 % was observed for the 40x40 cm^2 field. The OFs used for the TPS configuration were consistent with the averaged data within 0.3 % while the PDD data deviated from the averaged data for that field size by about 2.4 % explaining the observed calculation error.

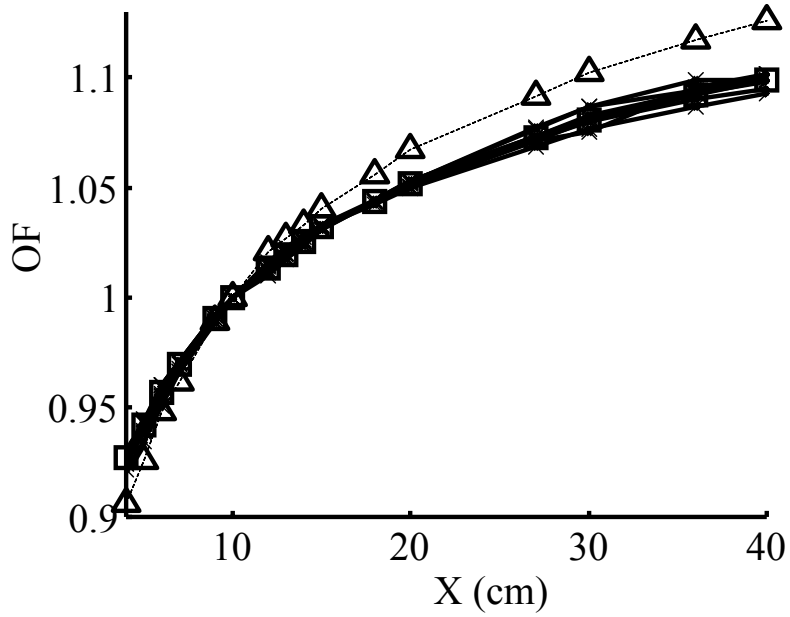

Figure 5-7. *The OFs at d_{max} ('x's with solid lines) measured for the 6 MV beams of Clinac 2100 CDs. One set of OFs (triangles with dashed line) deviated maximally by about 2.5 % from the obtained averaged OFs (squares with solid line). The large deviations were related to large calculation errors for central axis dose.*

5.4. SET-UP ACCURACY

Set-up errors and their effect on glandular mean dose

Summary of patient set-up variations is presented in Table 1. SD of random errors ranged from 0.4 to 3.3 mm among the patients. The shifts of D_{mean} values due to systematic set-up

errors are summarized in Table 2. Mean ± SD of 3D dose gradients within the parotid glands (defined as dD_{mean}/dr) was 0.8 ± 0.3 Gy/mm (range from 0.2 to 1.7 Gy/mm) in the steepest gradient direction toward a gland (usually anterior, cranial and lateral). The corresponding gradient within the submandibular glands was 0.9 ± 0.6 Gy/mm (range from 0.1 to 2.3 Gy/mm). No significant correlation was found between the gland volume and ΔD_{mean} or dose gradient ($|\rho| \leq 0.42$).

Table 1. Summary of patient set-up variations. The results are in millimetres and are expressed as mean ± SD calculated for all 25 included patients.

	Direction						
Shift type	A-P	C-C	LAT	|A-P|	|C-C|	|LAT|	3D
Systematic	0.5 ± 2.2	1.1 ± 2.1	0.1 ± 1.9	1.7 ± 1.3	1.9 ± 1.4	1.5 ± 1.2	3.4 ± 1.6
Random	0.0 ± 1.5	0.0 ± 1.4	0.0 ± 1.4	1.1 ± 0.9	1.1 ± 0.9	1.1 ± 0.9	2.2 ± 1.1

Directions: A-P = anteroposterior (+ = posteriorly), C-C = craniocaudal (+ = caudally), LAT = lateral (+ = to right).
| | denotes absolute values or magnitudes of shifts.

Table 2. Shifts of the glandular D_{mean}s due to the observed systematic patient set-up variations. The results are given as mean ± SD for the glands and the range is in parenthesis.

Gland	Planned D_{mean} (Gy)	Shifted D_{mean}(Gy)	Shift (Gy)	|Shift| (Gy)	Shift (%)	|Shift| (%)
Parotids (N=41)	23.2 ± 6.5 (13.8...55.2)	22.3 ± 7.1 (11.5...56.2)	-0.8 ± 1.6 (-4.5...2.9)	1.4 ± 1.2 (0...4.5)	-4.3 ± 8.3 (-25...15)	6.7 ± 6.4 (0...25)
Subm. (N=17)	33.8 ± 11.3 (22.6...55.1)	33.5 ± 11.1 (22.1...52.5)	-0.3 ± 2.7 (-8.0...3.9)	1.9 ± 1.9 (0...8.0)	-0.3 ± 8.2 (-20...17)	6.0 ± 5.5 (0...20)
Both (N=58)	26.3 ± 9.4 (13.8...55.2)	25.6 ± 9.8 (11.5...56.2)	-0.7 ± 2.0 (-8.0...3.9)	1.5 ± 1.4 (0...8.0)	-3.1 ± 8.4 (-25...17)	6.5 ± 6.1 (0...25)

The shift is defined as shifted – planned dose value obtained for a gland.
| | denotes absolute values or magnitudes of the shifts.

Dose response parameters

The relative ejection fraction of a major salivary gland as a function of the planned D_{mean} is presented in Figure 5-8. The 95 % confidence intervals of the parameter fits are presented in the figure. The maximal slope of the sigmoidal model was -0.041 1/Gy. In the form of normalized slope this is γ = - 0.041 1/Gy * 30.4 Gy = 1.2. Dose region for the fitting of the linear model was 10…40 Gy but equal slope value was obtained by using a narrower region 20…40 Gy.

Estimation of dosimetric and geometrical accuracy requirements

The accuracy criterion of ±10 % for the rEF results in an accuracy criterion of ± 2.4 Gy for D_{mean} of both the parotid and submandibular glands. Cautious criteria for 3D shifts can be achieved by considering positional shifts only in the most critical dose gradient direction (shift in all three directions toward a gland). On the average, the above accuracy requirement of D_{mean} was achieved with maximal systematic 3D shift of 3.0 mm for the parotid glands and of 2.7 mm for the submandibular glands.

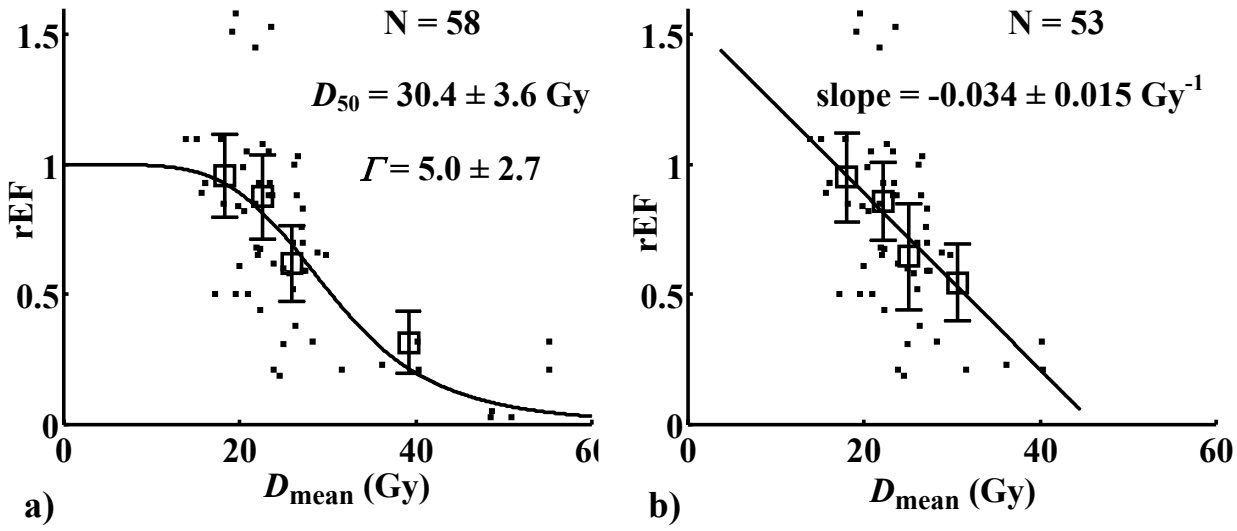

Figure 5-8. *The dose response models (solid lines) for the relative ejection fraction (rEF) of the major salivary glands 6 months after radiotherapy. Squares are the means of the rEFs for four subgroups of equal sizes with 95 % confidence limits of the means presented. a) The sigmoidal model and b) the linear model (fitted to a narrower range of data points between 10...40 Gy). The sizes of the subgroups were n = 58/4 = 14…15 and n' = 53/4 = 13…14, respectively.*

5.5. ACHIEVABLE IMPROVEMENT IN DOSE ACCURACY

5.5.1. Effectiveness of the investigated factors

The efficacy of the proposed empirical model fitting of the QC measurement results in the reduction of the uncertainty related to random measurement errors is summarized in Table 3. The presented results are applicable for the output and CC measurements and for the calibration coefficient of a CC device. The benefit of the modelling depends on the number of data points used for the fit.

The uncertainties estimated for the factors investigated in this thesis and the efficacy of the proposed methods, suggestions and criteria in the reduction of these uncertainities are summarized in Table 4. Two aspects of improvements are presented: i) those based on the reduction of SD indicate reduction of uncertainty observable in national and international

multicentre comparisons and increase in the consistency of treatment quality, ii) those based on the reduction of maximal errors indicate the benefit obtainable for individual patients and improvement in consistency of treatment quality.

Table 3. Uncertainty (1SD) (%) of a current measurement result when its value is obtained from the linear model fitted to the results of the current and previous measurements. *The uncertainty (1SD) (%) estimated for a current measurement result when its value is predicted by fitting the linear model to the results of the previous measurements (in italics).*

Measurement reproducibility (1SD) (%)	Number of data points used to fit the model			
	5	10	15	30
0.50	0.39	0.29	0.24	0.18
1.00	0.77	0.58	0.49	0.36
0.50	*0.53*	*0.34*	*0.27*	*0.19*
1.00	*1.05*	*0.68*	*0.54*	*0.38*

The results are indepent of the steepness of a linear time trend and the length of equal time intervals between the measurements.

The improvements for the reproducibility of the output (OM) and constancy check (CC) measurements and the calibration coefficient of a CC device are presented for typical size of data sets achievable for the model fitting of QC measurements consisting of ≥ 10 data points. The maximal effects are based on special cases observed in the analysis of QC data such as outlying output measurements and CC calibration factor. Maximal effect of systematic long-term drift in stability of a CC device is calculated for output measurement time interval of 3 months used widely in Finland.

Table 4. Summary of the uncertainties estimated for the investigated factors, the proposed improvements and their effects.

Investigated factor	Improvement	SD (%)		Maximal effect (%)
		Estimated	Improved	
Reproducib. of OMs	Model fitting	0.5	0.3	≈ 2[1]
Reproducib. of CCs	Model fitting	0.5	0.3	≈ 2[1]
Calibr. coeff. of CCs	Model fitting	0.6	0.4	≈ 2[1]
Drift in CC stability	Model fitting	-	-	≈ 1[1]
Beam data	Accuracy criterion of ±1 %	0.9	0.6	≈ 2[1]
QC program	Interval from 3 to 1 month[2]	1.5	1.1	≈ 1
QC program	Interval from 1 to 0.25 month[2]	1.1	0.8	≈ 1
QC program	Lowering of action levels[3]	ΔSD = -0.3		≈ 0.5
Output adjustments	Normal -> anticipatory	ΔSD = -0.2 or -0.1[4]		≈ 0.3 or 0.2[4]
Glandular mean dose	Correction of positional shifts[5]	ΔSD ≈ -8.4		≈ 25

[1]Improvements observed for special cases (see text), the effect is not necessarily general.
[2]Shortening of OM time interval from 3 months to 1 month or from 1 month to 1 week.
[3]Lowering of OM action level from ±2 % to ±1.5 % and CC action level from ±3 % to ±2 %.
[4]For OM time interval of 3 months or 1 month, respectively.
[5]Correction of all systematic positional shifts based on bony landmarks near the glands.

The improvement of SD for central axis beam data is taken from the SD of uniform distribution between the limits of the proposed accuracy criterion (from -1.0 to 1.0 %). The maximal effect is taken from the difference between the largest observed calculation errors for the test fields in Finland and the limits of the proposed accuracy criterion.

The presented maximal effect related to the shortening of output measurement time interval is within ± 0.4 % with the maximal effects (99 % limits) obtainable for the steepest output trends and the empirical distributions of output trends, and it is for the commonly used action levels of ± 2% and ±3% for the output measurements and the CCs, respectively, and for the output measurement reproducibility of 0.7 %. For the lower output measurement reproducibility level of 1.4 %, the corresponding maximal effect is 1.5 %. The improvements of the SDs are presented for output measurement reproducibility of 0.7 % and for the empirical distributions of output trends. For output measurement reproducibility of 1.4 %, the corresponding estimated values are 2.0 and 1.4 %, while the improved values are 1.4 and 0.9 %, respectively.

The presented maximal effect related to the lowering of the action levels is for the both output measurement reproducibility levels and the investigated range of output measurement time intervals, and it is within ±0.2 % with the maximal effects (99 % limits) obtained for the steepest output trends and the empirical distributions of output trends. The presented improvement of the SD is for the output measurement reproducibility of 0.7 % and measurement time intervals of 3 months and 1 month while that for the interval of 1 week is -0.2 %. The corresponding improvements of the SDs are 0.1 % lower for the output measurement reproducibility level of 1.4 %.

The improvements of SD and the maximal effects (99 % limits), when the type of the output adjustments is changed from normal to anticipatory, were estimated for the commonly used action levels of ±2 % and ±3 % for the output measurements and the CCs, respectively, and for the output measurement reproducibility of 0.7 %. The corresponding improvements of SD and the maximal effects were 0.1 % lower and 0.2 % greater, respectively, for the measurement reproducibility level of 1.4 %. For the output measurement interval of 0.25 month, the improvements were negligible.

For the major salivary glands, the maximal effect achievable for the ΔD_{mean} and its SD by correcting all systematic positional shifts based on bony landmarks is presented in the Table 4. For the conventional strategy correcting systematic 1D shifts > 3 mm (to 3 mm), the remaining shifts of the D_{mean}s were -2.8 ± 7.6 %.

5.5.2. Effectiveness of the proposed model fitting

The proposed model fitting of the QC measurement results facilitates regognition of random measurement errors and reduction of their effect improving measurement reproducibility and offering possibility to lower action levels. The deviation of total doses achievable is presented in Table 5. The results were obtained for output measurement reproducibility of 0.4 % (1SD) by improving the previously used value 0.7 % by about 40 % according to the results presented in Table 3. The uncertainty of the calibration coefficient of the CCs was 0.6 % (1SD) obtained by assuming that its value is predicted cautiously through model fitting to short data point sets to correct its value for long-term drifts in device stability (but resulting in no improvement in the reproducibility according to the results presented in Table 3). The SDs

were calculated by using the empirical distribution of output changes obtained for 6 MV and the limits by using the steepest observed output trends. Further improvement achievable by using anticipatory output adjustments is only about 0.1 % for the presented 95 and 99 % limits, regardless of output measurement time interval. The results are accurate for linear changes in output and calibration coefficient of the CCs providing theoretical maximal effect that can be approached in practice.

If the model fitting of the QC measurements is not used, the presented results are applicable when the measurement reproducibility level is 0.4 % (1SD) for the output measurements and 0.5 % (1SD) for the CCs, provided that potential systematic long-term drifts of a CC device are known and their effect corrected. If the model fitting is used only for the regognition of outlying measurements, conformed by remeasurements, the action levels may be lowered and the presented result can be approached when reproducibility level is close to 0.5 % (1SD) for the both output and CC measurements, provided that long-term drifts of a CC device are known and corrected.

Table 5. Deviation of total doses achievable by using the proposed model fitting of the QC measurement results in the case of normal output adjustments. The SD, 95 % limit and 99 % limit are presented for different action levels. The results for the case of anticipatory output adjustments are given in brackets.

Output measurement time interval (month)	Action levels: ±output measurements, ±CCs								
	±2 %, ±3 %			±1.5 %, ±2 %			±1 %, ±2 %		
	SD	95 %	99 %	SD	95 %	99 %	SD	95 %	99 %
6	1.6 (1.3)	2.8	3.3	1.2 (1.0)	2.0	2.5	1.0 (0.9)	2.0	2.4
3	1.4 (1.2)	2.7	3.2	1.1 (0.9)	2.0	2.4	0.9 (0.8)	2.0	2.4
1	1.1 (1.0)	2.2	2.6	0.9 (0.8)	1.7	2.0	0.6 (0.6)	1.4	1.8
0.25	0.9 (0.9)	1.5	1.7	0.7 (0.6)	1.2	1.4	0.4 (0.4)	0.9	1.0

The results suggest that the 95 and 99 % limits of the shifts in total doses obtainable by using a short output measurement time interval of 1 or 0.25 month, are achievable by using significantly longer measurement intervals of 6 months or 1 month, respectively. Consistently, the SD of dose spread obtainable by using a short measurement interval of 1 or 0.25 month is achievable by using significantly longer interval of 6 or 3 months, respectively, being mainly due to the lowering of action levels. On the other hand, the uncertainty of total doses achieved by using a certain measurement time interval can be reduced up to about half. Reduction of 99 % limit is about 2.5, 2.0, 1.0 and 1.0 % for the measurement interval of 6, 3, 1 and 0.25 months, respectively.

5.5.3. Combined improvement in dose accuracy

Maximal improvement

Maximal (99 % CI) error of dose related to the use of a dosimetric QC program and accuracy of basic beam data is presented in Table 6. The errors due to beam data and long-term drift in CC stability are systematic and they are summed linearly to the 99 % limit of the ΔD related

to the use of a QC program. Considerable maximal dose errors can be seen. Maximal improvement of dose accuracy achievable by choosing the strictest accuracy level is 9.7%-2.0% = 7.7% when compared with the worst case presented. Considerable improvements of about 3…5 % can be achieved with reasonable effort by using output measurement time interval of 3 months if the proposed model fitting is used and effort is made for accurate beam data measurements. Poor accuracy of beam data renders the effort made on frequent workfull output measurements meaningless.

Relative work load was estimated by assuming similar types of output measurements (water tank or solid phantom). No work load was defined for beam data measurements since they are not carried out regularly. Work/quality-ratio was calculated by multiplying work load and total ΔD. Both of these were normalized to those obtainable by using measurement interval of 0.25 month.

Table 6. Maximal combined error of dose (99 % limit of ΔD) allowed by different dosimetric QC programs and quality of central axis beam data.

Action levels: ± OMs, ± CCs	± 2 %, ± 3 %				± 1.5 %, ± 2 %	± 1 %, ± 2 %
Meas. interval (month)	3	3	1	0.25	6	0.25
Reproducibility (%)	0.7	1.4	0.7	0.7	0.4[1]	0.4[1]
ΔD due to output (%)	4.2	5.4	2.9	1.9	2.5	1.0
ΔD due to beam data (%)	1.0…3.3	1.0…3.3	1.0…3.3	1.0…3.3	1.0…3.3	1.0…3.3
Drift in CC stability (%)	1	1	0.3	0	corrected	corrected
Total ΔD (%)	**6.2...8.5**	**7.4…9.7**	**4.2…6.5**	**2.9…5.2**	**3.5…5.8**	**2.0…4.3**
Relative work load (%)	8.3	8.3	25	100	4.2	100
Relative work/quality (%)	18…24	21…28	36…56	100…179	5.2…8.3	69…148

[1]improved from the value 0.7 % by using the proposed model fitting

An error of 1 % in total dose results in average ΔTCP of 2.7 % and ΔNTCP of γ_{NTCP} %, where γ_{NTCP} is normalized slope of a NTCP curve. The maximal effect of the proposed actions in Table 6 is about 8 % facilitating the avoidance of significant decrease of 21 % in TCP or increase of $8\gamma_{NTCP}$ % in NTCP. The maximal effect is significant also for the NTCP of the major salivary glands being 8 %*1.2 %/% ≈ 10 %.

Improvement of the SD

In this chapter, the uncertainty of total dose (1SD) achievable by using the methods, suggestions and criteria proposed in this thesis are underlined and the SDs achievable otherwise are expressed *in italics*. The uncertainties are combined like independent Gaussians. The estimated uncertainty of dose related to the choice of a QC program for output is between 0.4 and *2.0* % (1SD). The uncertainty of dose measurements estimated by the IAEA is 1.2 % (1SD) when components interpreted to be related to measurement reproducibility are removed (factors 1-3 in chapter 2.2.1.). The uncertainty of central axis beam data was determined to be between 0.6 and *0.9* % (1SD). The combination of these factors results in "baseline" uncertainty of dose being between 1.4 and *2.5* % (1SD). The uncertainty of dose calculation

and physical treatment planning of 3-4 %, estimated by several groups, is interpreted here as 2SD with the SD being about 2 %. By combining this to the results obtained above, the uncertainty of dose is between 2.4 and *3.2* % (1SD).

Acceptable uncertainty of dose remaining for other factors (such as positional errors) to reach the strictest recommended overall accuracy requirement for dose of 3 % (1SD) is between 1.7 and *0* % (1SD). To reach the more relaxed recommended overall accuracy level of 5 % (1SD), the remaining acceptable uncertainty is between 4.4 and *3.8* % (1SD). The achievement of these overall accuracy levels for dose results in that the SD of TCP and NTCP is on the average about 8.1…14 % and $3\gamma_{NTCP}$ …$5\gamma_{NTCP}$, respectively, where γ_{NTCP} is normalized dose response slope for a complication.

Salivary glands

For the major salivary glands, the SD of ΔDs estimated above has quite small effect on the SD of the ΔNTCPs being about *3.2* %*1.2 %/% ≈ *3.8* %. As was presented in chapter 5.5.1., patient position correction in head and neck IMRT based on bony landmarks results in a SD of 7.6 % for the glandular D_{mean}s. By roughly assuming from literature data (Robar *et al.* 2007, Vakilha *et al.* 2007, Lee *et al.* 2008a, Lee *et al.* 2008b) that the deformation and movement of parotid glands toward patient midline has a variating component of about 6 % (1SD), the combined effect of positional shifts remained on the D_{mean}s of the parotid glands is ≈ 10 % (1SD) in the current data. This results in the SD of NTCPs of about 10 %*1.2 %/% ≈ 12 %. Geometric accuracy clearly remains the most crucial factor for the sparing of the major salivary glands as far as physical criteria are concerned. The dose gradient within the glands can be converted into the gradient of the ΔNTCP resulting in the value of 3.3 and 3.7 %/mm for the parotid and submandibular glands, respectively.

6. DISCUSSION

6.1. ANALYSIS OF QUALITY CONTROL DATA

Time trends

Systematic drifts were observed in accelerator output and calibration of CC device. Empirical models were found suitable for quantification of normal (no malfunctions) drifts for the investigated equipment.

Interestingly, the output of the Clinac 2100 CDs tended to increase with time while that for the 600 CDs tended to decrease with time. Similarly, increasing output trend has been reported for the 2100 CD (Luketina and Greig 2004) and the 2300 CD (Ravichandran *et al.* 2007), and decreasing trend for the 600 CD (Ravichandran *et al.* 2007). Since possible shifts in beam energy and profiles were checked and corrected before each output measurement, the obtained output drifts were attributed to the dose monitoring capability of accelerator. Changes in this capability may be due to chemical gas emission from chamber electric insulators, minor or major leakage of a sealed chamber and drifts in measurement electronics (Greene 1986). The decreasing output trends observed for the 600 CDs suggest increasing sensitivity for dose monitoring while the increasing output trends observed for the 2100 CDs suggest a decreasing sensitivity for dose monitoring (both accelerators have sealed monitoring chambers with metal plates). The investigation of the contribution of the above factors to the observed time trends was beyond the scope of this study.

The average time of 18 ± 12 months observed for an output drift of 2 % is remarkably longer than the most recommended output measurement time intervals, ranging from 1 week to 1 month. Because of the large variety of output trends, including also steep trends related to some accelerators just after commissioning, it is obvious that the obtained distributions represent quite realistic samples of normally observable long-term output variations for the investigated accelerator types. The long average drift time together with great variance in drift times indicate that the use of individual output measurement time intervals would be a more time saving procedure than the use of a common measurement time interval (Puurunen *et al.* 1985). By using the presented output modelling, current trends in outputs can be quantified and thus accelerator-specific time intervals for the output measurements can be approached. The results suggested that it is reasoned to apply a short output measurement time interval for a new Varian accelerator just after the commissioning.

The use of the presented long-term output modelling has some practical limitations. The rate of output adjustments may have an effect on the output stabilization. Despite drifts in beam energy and profile are maintained within narrow limits, they may have a slight effect on the determined output trends. Reproducibility of the measurements sets a limit for the number of data points needed for accurate determination of a time trend. Due to all these uncertainties, the values of the model parameters should be interpreted cautiously and be recalculated after each output measurement to estimate current output trend.

Dosimetrically significant malfunctions, repairs, adjustments and also wide output adjustments may cause unexpected shifts or complex fluctuations on the output level and may limit the use of the output modelling. In connection with this kind of a repair or an adjustment, the collection of new set of data for the model fitting may be required. The ageing

of an accelerator, without any malfunctions, may increase the risk of unexpected output behaviour (Roth *et al.* 1998), especially with accelerators having unsealed monitor chambers with aluminium plates (Blad *et al.* 1996). Despite all these limitations and disadvantages, however, the proposed output modelling was feasible for 7 of 8 accelerators during the whole observation period. Thus it seems to serve as a useful tool in QC and also in the optimization of QC program. For curiosity it should be mentioned that modelling of time trends for QC purposes has also been suggested for other technical parameters such as gun current considered to facilitate the evaluation component "health" and the detection of abnormal behaviour related to malfunctions (Crichton *et al.* 1999, Haas *et al.* 2000).

Periodic output changes from ± 2 % to ± 3 % have been reported by using unsealed CC instruments for output not correcting for temperature and pressure (Saw *et al.* 1997, Watanabe 2000). This should be taken into account in the estimation of short-term output trends from the CC measurements and the stability of the calibration of a CC device. The adjustment of dose rate, RF power or RF frequency may cause a shift in the calibration factor of a CC device. This should be corrected or collection of a new data set may be required for the estimation of device stability.

Reproducibility of measurements

The fitting errors of the long-term output models included random errors of the output measurements, incompleteness of the output models and random short-term shifts in outputs. Since the fitting errors were comparable with the differences of the comparative measurements between the hospital and STUK, the fitting errors seemed to reflect the reproducibility of the local output measurements. The comparative measurements were independent but they were, however, carried out in the same environmental conditions (humidity) and by using the same QI value resulting in potential underestimation of the local measurement reproducibility. By accounting the uncertainty due to the former estimated by the IAEA (0.4 %) (IAEA 2000) and the effect of the currently estimated uncertainty of the QI (0.5 %) on the calibration factor of a reference chamber being ≤ 0.1 % (1SD), the estimated uncertainty of the local output measurements becomes about 0.6 % (1SD).

The estimated reproducibility of the local output measurements was 0.5 % (1SD) and is remarkably better than the uncertainty of 1.4 % (1SD) estimated by the IAEA (2000). The estimation done by the IAEA is for a group of beams and ionization chambers including errors that are both systematic and random for an individual beam and chamber. The estimation and correction of such systematic errors, e.g. an error due to the use of the beam QI for the determination of the calibration coefficient for a reference ionization chamber, can not usually be approached by a user. Therefore, the measurement reproducibility level that can be estimated by a user is due to random errors. Similar reproducibility values of 0.4 % (1SD) for the output measurements and 0.8 % (1SD) for the CCs have been reported (Luketina and Greig 2004).

In principle, the high reproducibility of output measurements enables the lowering of the action level from the currently recommended value of ± 2 %. It was observed that 2.4 % of the output measurements were clearly outlying from the trends of other measurements and from the CCs. It would thus be anticipated that the lowering of the action level under ± 1.5 % (≈ 3SD) might lead to small fraction of unnecessary output adjustments. To approach even lower action levels, the indentification of outlying or erroneous output measurements should be improved. This can be approached by using the proposed model fitting for the output

measurements, the CCs and the stability of a CC device and by careful comparison and interpretation of these results.

6.2. ESTIMATION OF EFFICACY OF QUALITY CONTROL PROGRAMS

The developed method offers a tool to estimate the efficacy of QC programs for output levels on a statistical basis. The results obtained for single linear output changes are suitable for individual accelerators, while the results obtained for the distributions of output changes provide long-term estimations for a set of accelerators. However, if the average value of the output changes for a given set of accelerators differs remarkably from zero, the results obtained for a single output change corresponding to the average value may provide a more realistic estimation.

The fraction of unacceptable treatments (with $|\Delta D| > 2$ %) could be maintained when the output measurement time interval is doubled, if the action levels of the QC measurements are lowered. In principle, this enables the optimization of the measurement interval while maintaining the treatment quality. Lowering of the action levels requires high measurement reproducibility to avoid unnecessary or even erroneous output adjustments. The recognition of erroneous measurements can be improved by using the proposed model fitting of the results of the QC measurements. The most important limitation to the use of a prolonged output measurement time interval is the quality of the CCs related to both the equipment technology and the calibration procedure. A method to improve the calibration procedure was proposed in the section 5.1.3. Naturally, CCs should also be carried out for beam energy and symmetry parameters having an effect on the results of the CCs of output. When time interval for output measurements is aimed to be prolonged, it is useful to verify the stability of a CC device with shorter intervals e.g. through output measurements carried out in a solid phantom for only one nominal beam energy.

The presented estimations of output spreads were obtained by assuming that the long-term average of the drifts in the CC calibration is zero. The estimations are valid only when these drifts are recoverable or when systematic calibration drifts are known and their effect is corrected. This assumption was made to maximize the generality of the results since long-term drifts may depend drastically on a device used. The validity of the CC reproducibility level used in the estimations can be verified for a device by using the proposed model fitting.

The spread of total doses achievable by using a QC program was presented cautiously by using the 99 % limits of the ΔDs obtained for the steepest empirical output shifts of up to 1.2 %/month. The overestimation of dose spread achievable by using the longest measurement interval of 3 months (due to more frequent control of such unstable accelerators regardless of the choice of QC program) is only about 0.5 % when estimated by using the commonly detected output trend of 0.3%/month.

The empirical distributions used for the simulation of the random errors for both the output and CC measurements are practically identical with the Gaussian distribution between −2SD and +2SD. Therefore, it is expected that the estimated 95% limits of the ΔDs and also the probabilities of unacceptable treatments are generally applicable. The empirical distributions do not include very large errors (> 3.6SD) since they are usually identified in practice and the tails of the distributions are short. Therefore, these distributions provide slightly greater

probabilities for large errors close to 3SD when compared to the Gaussian distribution. Due to the principal uncertainty related to the estimation of probabilities for large measurement errors, the presented 99% limits of the ΔDs should be interpreted cautiously.

In practice, output is not necessarily adjusted exactly to a measured reference value as assumed in the presented estimations suggesting that the obtained dose spreads may be slightly underestimated. On the other hand, it was assumed that the results of the output measurements are accepted uniformly within the predefined action limits. Output may often be adjusted even though these limits are not exceeded and the presented dose spreads may slightly overestimate the dose spread achieved in this case.

The SD, 95 % and 99 % limits of actual ΔDs were estimated for 6 MV beams from the results of the absorbed dose measurements resulting in the values of 1.1, 1.9 and 3.2 %, respectively. These values are quite close to the corresponding values of 1.3, 2.2 and 3.0 %, respectively, estimated by using the developed method for the QC practice used in HUCH during the observation period (output measurement time interval up to 6 months, average action levels: ± 1.7 % and ± 2.3 % for the output and CC measurements, respectively, 'normal' output adjustments). Great consistency in these results suggests the feasibility of the developed method. Slightly smaller actual values may be due to the use of anticipatory output adjustments for some steep systematic output trends.

6.3. DOSE CALCULATION ACCURACY

Photon calculation accuracy in Finland compared with the international level

The current analysis revealed that the OF and PDD data in the Finnish RT centres deviated more than ± 2 % from the current averaged data for about 6 and 10 % of the investigated beams, respectively. Quite similar percentages of about 4 and 13 %, respectively, have been reported by the RPC in the USA (Bencomo *et al.* 1999), obtained by comparing mailed copies of measured beam data for 235 photon beams (from 75 centres) with the reference data constructed by the RPC. In the UK, an on-site dosimetric review for 159 photon beams (through ionization chamber measurements at a 5 cm depth) revealed that the mean ± SD of differences between measured and calculated central axis dose were 0.3 ± 1.5 % (range from -4.0 to 6.6 %) for FSs of 5x5, 10x10 and 15x15 cm^2 (Thwaites *et al.* 1992). The corresponding differences in the current study were 0.0 ± 1.2 % (range from -2.5 to +3.9 %), respectively (obtained for the same FSs by omitting the correction of the measured differences for output shifts since this was not done in the referred study). Dose calculation errors exceeded the limit of ± 3 % in about 6 % of the investigated beams in Finland. The percentage is significantly lower than 23 % obtained for such large deviations in OFs alone observed in the ESTRO EQUAL-program (Ferreira *et al.* 2000) (obtained for 227 open 7x7, 20x20 and 7x20 cm^2 fields at a 10 cm depth). The comparison of the current results with those obtained in the EQUAL, however, is limited mainly due to different measurement techniques and contributions of set-up errors (averaged ion chamber measurements in this study versus single TLDs in the EQUAL).

Achievable calculation accuracy level for central axis dose with the PBC algorithm

The current results suggest that the uncertainty of central axis dose calculation due to uncertainty of measured beam data is about 0.9 % (1 SD) with range from -3.3 to 3.2 % for

the investigated beams. By considering together the SD of the averaged OFs (0.5 %), the reproducibility of the local PDD measurements (0.3 %) and the percentage of test fields having calculation accuracy within ± 1 % (79 %), it can be concluded that the calculation accuracy level within ± 1 % is achievable for open square fields in water with reasonable effort and should, therefore, be recommended. As a matter of fact, the SD of the OFs includes inter-accelerator differences and gives, therefore, too pessimistic estimation for reproducibility of OFs measured for a single accelerator. An accuracy level of ± 1.5 % is achievable for elongated rectangular fields (accounting for the observed systematic underestimation). These conclusions are consistent with the recommendations made by the AAPM (Fraass *et al.* 1998). The large observed calculation errors suggest that the quality of basic beam data should always be verified to obtain relevant inter-centre comparisons of dose calculations, such as comparison of different TPSs (Kosunen *et al.* 1993, Venselaar and Welleweerd 2001).

The use of reference beam data

The current results indicated that when dose calculation accuracy was within ± 1%, the consistency of beam data was within about ± 1 % (at least up to a 10 cm depth) for all the investigated accelerators. Greater SD with FSs ≥ 30x30 cm^2 may be due to limited size of water phantom used for the beam data measurements. The use of the current averaged data sets proved useful in the detection of reasons for large dose calculation errors exceeding ± 2%. These together suggest that reference data sets can be used for accurate QC of beam data. The data sets obtained for the 6, 15 and 18 MV beams of Clinac 2100 CDs have high confidence due to several individual beams investigated in these nominal energy groups. The generality of the data obtained for the other investigated beams may be limited due to fewer individual beams analyzed. These data sets, however, were also published because they provided accurate dose calculations for the investigated beams. By investigating beam data from 2350 photon beams, the RPC (Followill *et al.* 2004) demonstrated that dosimetric beam data are in an agreement within ± 2 % for individual accelerators of same manufacturer, model and energy. Since the data were not corrected for measured calculation accuracy, it is not known to what extent the observed variation is caused by measurement errors and real variations in beam characteristics. The accuracy of the reference data, however, may be limited due to slight differences in the geometry of treatment head, flattening filter and monitor chamber. Therefore, accurate beam data measurements should not be replaced with the use of reference data.

Some current accelerator models, such as tomotherapy units, are supplied with their own TPS ready configurated and the user only verifies the accuracy of the dose calculation algorithm through point measurements. This rises a question how to confirm that the algorithm is tuned to result in best possible calculation accuracy for maximal number of clinical situations.

The current results support the previous findings of high consistency in beam data for the Clinac 600 CDs and 2100 CDs. The averaged OFs (at d_{max}) obtained for the 6 MV beams of the 2100 CDs are consistent within 0.4 % with those reported by Cho *et al.* (2005) and by Watts (1999). For the 15 MV beams of the 2100 CDs, the current averaged OF data were consistent within 0.4 % with those reported by Watts (1999). For the 18 MV beams of the 2100 CDs, the current OF data were consistent within 1.0 % with the values presented by Followill *et al.* (2004). For the 6, 15 and 18 MV beams of the 2100 CDs, the averaged PDDs at a 10 cm depth were within 0.7 % with those reported by Coffey *et al.* (1980), within 0.7 % with those reported by the BIR (1996) and within 1.0 % with those reported by the BIR

(1996), respectively. The averaged PDDs at a 10 cm depth obtained for the 4 MV beam of the Clinac 600 CD and for the 6 (old filter) and 10 MV beams of the Clinac 2100 CDs were within 0.5 % with the values published by the BIR (1996). The data published by Fontenla *et al.* (1992) were the best match found within 0.8 % for the averaged PDDs at a 10 cm depth for the 6 MV beams of the 600 CDs. A more detailed analysis concerning the use of reference data for different accelerator models has been performed by the RPC (Followill *et al.* 2004).

The current value of the parameter $A \approx 1.5$ in the equivalent square field relation (Eq. 15) obtained for the Clinac 2100 CDs is consistent with the value $A = 1.5$ reported for head-scatter photons of the 2100 C (Kim *et al.* 1997) and the 2100 CD (Zhu *et al.* 2001). On the contrary, the current value of $A \approx 1.3$ obtained for the 600 CDs is slightly smaller than the value $A = 1.5$ reported for head-scatter photons from the 6 MV beam of the Varian 6/100 (Zhu *et al.* 2001). Smaller collimator exchange effect observed for the 600 CDs when compared to that of the 2100 CDs is due to thicker shielding around a monitoring chamber partly eliminating radiation backscatter from the upper collimator jaws to the monitoring chamber (Yu *et al.* 1996). The current results suggest that the exchange effect of the collimator jaws on the OFs is smaller at a 10 cm depth than at d_{max} being understood by a lower amount of low energy extrafocal radiation from treatment head at 10 cm when compared to that at d_{max} (Liu *et al.* 1997). The effect of extrafocal radiation is more prominent in the 5x30 cm^2 field than in the 30x5 cm^2 field explaining the systematic underestimation of dose calculations observed in the 5x30 cm^2 field at a depth of 10 cm when the PBC dose calculation algorithm is used.

6.4. EFFECT OF SET-UP ERRORS

In this thesis, positional shifts of the bony structures near the major salivary glands were determined to approximate positional shifts of the glands by using the conventional method of patient set-up verification in head and neck IMRT. The method is based on two orthogonal portal images taken from one to three times a week and by correcting for set-up errors exceeding the predefined 1D tolerance. Better positional accuracy is expected with current cone beam CT implementations such as Varian On-Board Imager (OBI). However, the current results of the observed positional errors of the bony structures in different orthogonal directions (means and SDs) were consistent within 1 mm with those obtained by using OBI (Mechalakos *et al.* 2007). The current SDs obtained for the systematic positional shifts are close to the average of the values reported in other current studies (Hong *et al.* 2005, Guckenberger *et al.* 2006, Linthout *et al.* 2006) and are clearly smaller than the median of values reviewed in 2001 (Hurkmans *et al.* 2001).

It is expectable that positional shifts of the bony structures near the major salivary glands result in positional shifts of the glands due to the basic anatomy but a crucial question is the magnitude of movement of the glands with respect to these bony structures resulting in positional uncertainty of the glands. According to the recent studies of several groups the parotid glands tend to shrink and move toward patient midline during the radiotherapy of head and neck cancer resulting in an increase in their D_{mean} (Robar *et al.* 2007, Han *et al.* 2008, Lee *et al.* 2008a). By roughly assuming constant parotid volume and shape, and by using the most consistent literature value for the shift toward midline of about 3 mm (Lee *et al.* 2007, Robar *et al.* 2007, Lee *et al.* 2008b), a net effect of the shift is 1.2 ± 0.5 Gy for the current data. The estimated magnitude is consistent with an accuracy level of 1.0 Gy reported achievable for D_{mean} by using bony landmarks (O`Daniel *et al.* 2007). The relative effect is about 5 % (range

from 0.5 to 12 %) being consistent with the results of several other studies (Robar *et al.* 2007, Vakilha *et al.* 2007, Lee *et al.* 2008a). Existence of potentially similar effect of volume change and movement toward patient midline has not yet been reported for the submandibular glands and remains unknown.

The dosimetric and geometric accuracy requirements proposed for the major salivary glands concern overall accuracy and they are valid whether positional changes have not significant effect on the steepness of the dose response. The accuracy criteria were given cautiously to account for steep dose response and dose gradients within the glands. Steep dose response slopes from 4 to 5%/Gy consistent with our current results have been reported (Eisbruch *et al.* 1999, Münter *et al.* 2004, Dijkema *et al.* 2008) supporting the current proposal of strict accuracy criteria. The effect of gland deformation may limit the accuracy level of the rEF achievable by using the proposed geometric accuracy criterion but the desired accuracy level can be approached by applying the criterion for the shift of the mass center of a gland.

The slope of the linear model fitted to the steepest dose regions was expectedly too shallow when compared to that obtained by using the sigmoidal model. This is because of relatively large dose region used for the fitting of the linear model in order to reach enough data points. Despite this limitation, however, the use of the linear model independently confirmed the dependence of the rEF on the D_{mean} providing an estimate of the minimal steepness of the dose response curve.

In the current work, biological effects observable only 6 months after radiotherapy was used as an endpoint. This was chosen cautiously since the recovery of the salivary function with time is individual and depends on glandular dose (Eisbruch *et al.* 1999, Murdoch-Kinch *et al.* 2008). Conveniently, the slopes obtained for later effects do not seem to be different (Dijkema *et al.* 2008, Tenhunen *et al.* 2008) from those obtained for 6 months by using the same methodology but D_{50} value is greater for later effects (Dijkema *et al.* 2008). The current D_{50} value is consistent within 2 Gy with those obtained by other groups (Eisbruch *et al.* 1999, Dijkema *et al.* 2008) by using Lyman dose response model and binary endpoint (complication or not) with complication defined as reduction of salivary gland flow below 25 %. Since some authors have observed (Murdoch-Kinch *et al.* 2008) that radiation sensitivity of submandibular glands may be lower than that of parotid glands, the sigmoidal model was also fitted separately for these glands. The obtained D_{50} and γ values were within 0.3 Gy and 0.4, respectively, for both parotid and submandibular glands showing no significant differences. The observation is consistent with that of other group (Münter *et al.* 2004) and with that of our previous study carried out by using the same endpoint based on scintigraphy (Tenhunen *et al.* 2008).

Salivary gland scintigraphy is estimated to detect changes of 5-10 % in parenchymal function due to its small observer dependency and high reproducibility (Bohuslavizki *et al.* 1997). The SD of the fitting errors of the sigmoidal dose response model was about 0.3 (= 30 %) containing variation in radiation sensitivity, limitations of the scintigraphic method and changes in gland function due to non-systematic dose deviations from the planned dose. Since the estimated effect of dose shifts is about 12 %, the effect of the other sources is about 27 %. This suggests that individual radiobiological sensitivity of a gland and its scintigraphic assessment remain the most prominent factors of uncertainty in the determination of the dose response parameters when a logistic dose response model is applied together with the D_{mean}. The use of a more sophisticated NTCP model accounting for inhomogeneous dose distributions within the glands might reduce the scatter of data points. The SD of the EF(0)

and EF(6) was about 0.2 for the patients being consistent to the estimated uncertainty of a similar scintigraphic method observed for the parotid glands by an other group (Firat *et al.* 2006). Interestingly, by combining these uncertainties, the overall uncertainty becomes about 0.3 (1SD) corresponding to the large variation seen in the values of the rEF.

As a curiosity, the sigmoidal and linear dose response models were also fitted to the shifted D_{mean} values (shift due to the effect of uncorrected realized systematic positional error). Interestingly, the fitting errors of the models reduced about 10 % and the steepness of the fitted dose response curves increased about 10 %. Despite that actual positional shifts of the major salivary glands were detected potentially only partly by using the bony landmarks near the glands, the correction of D_{mean} values for the obtained positional shifts had a slight effect on the obtained dose response parameters. The observation is consistent to some expectations that any source of error may reduce the steepness of the slope of a fitted dose response model. The results suggest that corrections for errors made individually for the glands seem to reduce the scatter of the data and increase the steepness of a fitted dose response model. The correction of D_{mean} by taking into account also the deformation of a gland and its movement toward patient midline would be of a great interest.

Several methods have been proposed to estimate the effect of random positional errors on dose (Ploquin *et al.* 2006). As suggested by the results by other groups (Samuelsson *et al.* 2003, Siebers *et al.* 2005), an average effect of random set-up errors is much smaller than that of systematic errors. Since intrafractional positional shifts have also been found to be far less than interfractional shifts (Mechalakos *et al.* 2007), it can be considered that the systematic shifts obtained from the portal images sufficiently describe the prominent effect of positional shifts detectable by the portal imaging.

The effect of positional shifts on D_{mean} values of the salivary glands depends on margins around the glands (PRV) and target volume used in the treatment planning (Manning *et al.* 2001, Ballivy *et al.* 2006). This essentially determines the steepness of the dose gradients faced by the parts of the glands nearest to the target volume. In the current study, a margin of about 3 mm was used. Maximal 3D dose gradient in the current study was 2.3 Gy/mm but 1D gradients being about two times steeper have been reported (Prabhakar *et al.* 2007).

6.5. LIMITATIONS OF BIOLOGICAL ESTIMATIONS

The aim of the accuracy criteria is that absorbed dose is kept within acceptable limits around the clinical reference value which should be traceable to an absolute dosimetry standard. This is essential for sufficient and consistent treatment quality, multicentre trials and dose escalation studies. In addition to physical accuracy criteria, the extent of biological effects related to QC procedures should be considered. E.g. proper comparison of different treatment protocols requires sufficient limits for the variations of TCP and NTCP. The biological estimations, however, are limited due to large variation in steepness of dose response depending drastically on cancer type, irradiated normal tissues, biological endpoint chosen (Levegrün *et al.* 2001, Cheung *et al.* 2005a, b), follow-up time, treatment practice (including other treatment modalities), dose distributions within tissues (Levegrün *et al.* 2001) and possibly even slightly different dosimetry standards for the absorbed dose. Moreover, physiological factors such as differences in tumor size (Suit 1966), acute and chronic hypoxia (Elkind *et al.* 1965, Kallman 1972, Dasu *et al.* 2005) affect tumour dose response. As a

consequence, considerably steeper dose response slopes have been reported e.g. for some prostate cancer subgroups than for unclassified prostate cancer patients (Levegrün *et al.* 2001, Cheung *et al.* 2005a) complicating and limitating quantitative biological estimations even further.

The recently reported slopes for biochemical control of prostate cancers, being from 1.4 to 2.2, are within 95% CI consistent with the average slope of 2.7 used in the current biological estimations. This supports the feasibility of the presented estimations of the ΔTCPs for the prostate cancer (Fowler *et al.* 2001, Cheung *et al.* 2003, Cheung *et al.* 2005a). The existence of much steeper dose-response slopes should not, however, be omitted since any errors and variations in treatment techniques and in radiobiological tumor characteristics tend to decrease the slopes of the fitted TCP models (Fischer and Moulder 1975, Zagars *et al.* 1987, Thames *et al.* 1992). This suggests that special criteria for dose accuracy may have to be considered when relevant and possible.

6.6. ARE CURRENT LEVEL AND REQUIREMENTS OF QUALITY CONTROL SUFFICIENT?

The overall accuracy requirement of absorbed dose is 3-5 % (1SD) which could be interpreted so that the 95 % CI of the dose should be about 6-10 % (estimated as 2SD). Unfortunately, the uncertainty of dose related to several factors, e.g. positional errors and dose calculation, depends on treatment, planning and immobilization techniques rendering its accurate estimation limited. Therefore, the proposals for sufficient QC procedures and requirements for the investigated topics are given so that their contributions to the overall uncertainty are small when compared with those of other factors that have been reported currently achievable. In this light, the use of procedures stricter than the proposed ones would increase workload without significant improvements in overall dose accuracy. The systematic errors are summed linearly and random errors in squares.

QC program of output

The choice of an output measurement time interval has the most significant effect on accuracy when the interval is between 0.25 and about 2 months, while it becomes less important when the interval is longer than 2 months. The overall uncertainty of 1.5 % (1SD) estimated for the absorbed dose measurements by the IAEA (2000) does not contain the uncertainty related to a QC program of output. One criterion for a choice of an appropriate QC program might be that the uncertainty related to a QC program has not significant contribution when it is combined to the uncertainty of the dose measurements. By interpreting that some factors in the IAEA estimation are related to random measurement errors (such as long-term stability of user dosimeter 0.3 %, establishment of reference conditions 0.4 % and dosimeter reading relative to beam monitor 0.6 %), the combined uncertainty is rounded to 1.5 % (with 0.5 % precision) when the SD of dose spread (ΔDs) due to a QC program is up to 1.2 %. This criterion would require the use of a very short output measurement time interval of 1 or 0.5 month for output measurement reproducibility levels of 0.7 or 1.4 %, respectively. The use of the proposed model fitting of the QC measurement results in order to reduce the effect of random errors and to lower the measurement action levels from the conventional values (as seen in Table 5) results in that the criterion can be met even by using a long measurement interval of up to 6 months. As far as the presented criterion is concerned, the choise of an appropriate QC

program depends on the reproducibility of the QC measurements, and on the handling of random measurement errors and long-term drifts in the stability of a CC device.

Potential development of dosimetry protocols may reduce the uncertainty of the dose measurements. Then, acceptable uncertainty related to the use of a QC program of output should be reconsidered to maintain it in a sufficiently low level with respect to potentially reduced overall uncertainty. The presented results can be applied for such reconsideration provided that reproducibility of the measurements will not be significantly improved from the best level used for the presented estimations with potentially improving equipment technology.

The accuracy requirement based merely on the SD of the ΔDs is not necessarily sufficient, e.g. for unstable accelerators showing significant systematic time patterns in output. The SD of ΔDs may be small but the average dose level may be considerable different from zero and/or significant fraction of total doses may exceed unacceptable limits. Therefore, maximal limits achievable for the ΔDs should also be considered in the estimation of an appropriate QC program. The estimation of these limits is difficult due to uncertainty related to the recognition of potential large output deviations and large measurement errors in practice. As a general rule, the results suggest the use of a short output measurement time interval when the measurement reproducibility level is poor. For the poorer measurement reproducibility level of 1.4 %, the use of a short measurement interval of 1 month results in similar 99 % limits of the ΔDs as the use of a longer interval of 3 months for better measurement reproducibility of 0.7 %. For poor measurement reproducibility level close to 1.4 %, the use of a long interval of 3 months seems to be insufficient since maximal ΔDs are about ± 5 % being about half of the acceptable limits for the overall dose accuracy. The use of a measurement interval of 1 week keeps the ΔDs practically within the redefined action levels independently from the measurement reproducibility. The use of the proposed output modelling of the QC measurements results in that the use of a long output measurement interval of even up to 6 months seems to be sufficient (as seen in Table 5) consistently to the criterion based on the SD of the ΔDs.

The use of the proposed model fitting of the QC measurement results enables the optimization of resources spent on workful output measurements while maintaining treatment quality with the investigated accelerator types (as was seen in Tables 5 and 6). It is interesting to notice that after the publication of the study II, doubling of output measurement time interval by simultaneously reducing measurement tolerance by about 0.2-0.8 % based on accelerator stability has also been proposed by other authors (Bouchard and Carrier 2007). As an alternative goal, the spread of output can be significantly reduced improving treatment quality when output measurement time interval is not changed. Anticipatory adjustment technique is recommended for output. In such technique, systematically increasing output level is adjusted to a value below the measured reference value and decreasing output to a value above the measured reference value. If the measurement reproducibility is close to 1.4 % or even worse, improvements in measurement practice or equipment should be considered. The choice of appropriate general output measurement time interval has been discussed so far but the use of individual measurement intervals seems to be the most optimal procedure. Especially new accelerators may have to be controlled frequently just after commissioning. Output stability and appropriate time intervals can be estimated and updated by using the proposed model fitting. The ultimate ‘truth’ for the choice of a sufficient QC program remains unknown due to its dependence on the uncertainty of other factors being case-specific. The results and

suggestions presented in this thesis, however, provide some general advice and should facilitate in the choice of an adequate QC program for output.

Workfull output measurements carried out in a water tank can not easily be replaced by less workfull and equally accurate measurements. Some QC programs accept the use of PMMA or solid water phantom measurements from weekly to monthly basis to minimize the effect of potentially poor long-term stability of a device used for the daily CCs (Kutcher *et al.* 1994, IPEM 1999). Then, water tank measurements are carried out rarely, e.g. at least annually. Problems may arise, however, if the solid phantom measurements are used only to check output constancy and they are not converted into water tank measurements. Due to finite reproducibility of water tank measurements, output level may be shifted. Therefore, accurate conversion should be carried out and rare water tank measurements should be used to regular verification of the conversion due to potential changes in phantom material. The currently proposed model fitting method should be usefull also in the evaluation of the conversion.

Beam data and dose calculation

The importance of QC for basic beam data was demonstrated by showing a few dose calculation errors between 2-3 % in the open rectangular test fields exceeding international recommendations of calculation accuracy for such fields. These systematic errors accumulate with uncertainty of other sources jeopardizing the fulfilment of the accuracy recommended for clinical dose calculation and absorbed dose within a patient. Therefore, the best accuracy level practically achievable should be aimed at for the open rectangular fields to ensure sufficient quality of basic beam data used for TPS.

As reviewed in chapter 2.3.2., several groups have estimated that the uncertainty related to the dose calculation and physical treatment planning is typically about 3-4 % most likely including small level of uncertainty (≈ ± 1 %) related to basic beam data. This is about half of the overall acceptable uncertainty. In this light, it seems that systematic errors in beam data should not exceed the value of about ± 1 % justifying the strict accuracy criteria proposed for central axis beam data for field sizes from 5x5 to 40x40 cm^2.

The choice of appropriate dose calculation algorithm should also be considered in QA procedures. E.g. for still commonly used pencil beam convolution algorithm large calculation errors exceeding 10 % have been reported in inhomogeneous media (Knöös *et al.* 1995). For static beams, the QC of central axis beam data is the most important procedure but more elaborate tests are required to evaluate the performance of calculation algorithms in practical situations.

For the head and neck IMRT, the clinically commonly used PBC and superposition convolution algorithms tend to underestimate the mean dose of the parotid glands by about 4 % when compared with MC techniques (Schwarz *et al.* 2003, Boudreau *et al.* 2005, Sakthi *et al.* 2006, Mihaylov *et al.* 2007). The underestimation seems to be rather systematic than random for an individual gland and, therefore, its effect is included in the dose response parameters obtained for the glands by using dose values calculated by using these algorithms.

Positional verification in head and neck IMRT

Positional errors and their effects were investigated only in a special case of salivary gland sparing head and neck IMRT. In that technique, patient position verification is carried out

conventionally based on bony landmarks near the glands detected by the portal imaging. The literature data reviewed in section 6.4. suggest, however, that the parotid glands tend to shrink and move toward patient midline during the treatment rendering the positioning based on bony landmarks inaccurate. In that section it was estimated (based on the current and literature data) that positional corrections based merely on the bony landmarks near the major salivary glands seem to reduce the positional errors of the glands roughly to about half of their actual values (O`Daniel *et al.* 2007). Corrective actions based on positional shifts of rigid anatomy may not be adequate, especially if both of the parotid and/or submandibular glands are intended to be spared. According to literature, the average shift toward midline may be about 3 mm being as large as the currently proposed geometric accuracy criterion for the glands. In this light, the currently proposed accuracy requirements might be approached on the average by using the conventional patient position verification based on portal imaging provided that all detected systematic positional errors could be completely corrected. Pursuing this, however, raises questions whether the conventional verification method based merely on daily images and tolerance for 1D shifts is accurate enough in the detection of very small systematic 3D shifts. One approach may be a strategy based on sequential estimations of systematic shift from the portal, kV or CT images accumulating during treatment. Patient set-up is corrected for the obtained systematic error, not only for random error exceeding the tolerance of 1D shifts.

Considerable variation with range from -2 to 12 mm has been reported for the shifts of individual parotids toward patient midline (Robar *et al.* 2007, Vakilha *et al.* 2007, Lee *et al.* 2008b). This suggests that sparing of parotid function may be compromised for some patients even when all systematic positional shifts are corrected based on the bony landmarks. The detection and correction for tissue deformations and movements with respect to bony structures can be currently approached with the use of adaptive radiation therapy technique and deformable structure fitting (Han *et al.* 2008, Lee *et al.* 2008a). In such technique, the dose distribution can be replanned according to current patient anatomy. The technique seems promising and may prove useful in head and neck IMRT. Slight improvement of 5 % * 1.2 %/% ≈ 6 % may be expected on the average in spared parotid gland function (estimated by using current dose response slope and literature data of the relative dose effect). Due to large variation in biological data, however, it is not clear whether all possible positional corrections and adaptive radiotherapy crucially reduce the variation of NTCP of the glands rendering the estimation of outcome more accurate for an individual gland and patient. The question will hopefully be answered in future by collecting dose response data with dose errors due to all positional shifts corrected.

The presented accuracy requirements should be adopted when a glandular mean dose is close to about 23-37 Gy. Due to differences in dose gradients between individual treatment plans, geometric accuracy criterion of the major salivary glands should be evaluated individually for each treatment plan e.g. by simulating set-up errors. For this purpose, the accuracy requirement obtained for the D_{mean} can be used.

The accuracy of glandular mean dose may be affected by several system related sources of errors when IMRT technique is being used. Systematic errors of MLC leaf movements of 1 mm may cause about 10 % differences in D_{mean} values of the parotids (Mu *et al.* 2008) requiring strict quality assurance.

7. SUMMARY AND CONCLUSIONS

The aspects investigated in this thesis have an impact on the accuracy of absorbed dose in radiation therapy. These are reproducibility of dosimetric quality control (QC) measurements, stability of accelerator radiation output, effectiveness of a chosen dosimetric QC program, accuracy of beam data used to configure dose calculation algorithm and patient positioning. Methods, suggestions and criteria were proposed to improve dose accuracy and to optimize workload related to dosimetric QC. These were demonstrated to be applicable in QC of the investigated Varian Clinac 600 and 2100 CDs having sealed monitoring chambers in normal clinical situations.

The developed method revealed that appropriate choice of a dosimetric QC program should take into account the measurement reproducibility and output stability. A method based on empirical model fitting of QC measurement results was found suitable for the quantification of these factors. The change of measurement action levels was shown to have more prominent relative effect than the change of measurement time interval. The proposed model fitting facilitated identification and reduction of random measurement errors enabling the lowering of measurement action levels. As a consequence, workload of dosimetric measurements can be significantly reduced by prolonging output measurement interval from 1 month to even 6 months while maintaining treatment quality. Alternatively, by maintaining the workload, dose accuracy can be improved by even about 3 %. The method can be easily incorporated in the electronic archives of QC results.

The resources spend on QC measurements can be further optimized if individual measurement time intervals are used for the accelerators instead of a common measurement interval. Frequent checks were reasoned for some accelerators just after the commissioning but these accelerators seemed to stabilize with time. The proposed empirical model fitting was found suitable for the evaluation of individual measurement time intervals.

The importance of QC for beam data used for the dose calculations was demonstrated by showing errors of up to about 3 % in such data. The magnitude of these errors was comparable to the benefit obtainable by using a short output measurement time interval. Robust reference beam data sets were constructed for the Varian Clinac 2100 CDs. They were shown useful in the identification of error sources and accurate tools for the QC of beam data. The use of strict accuracy criterion of ± 1 % was found justified for central axis beam data.

Required dosimetric QC actions were presented for pursuing even the strictest overall dose accuracy recommendation of 3 % (1SD). Maximal combined effect of the investigated dosimetric QC actions was up to about 8 %. It is expected that improvement of dose accuracy makes the modelling and estimation of biological effects more reliable, i.e., tumor control and normal tissue complications.

In the head and neck IMRT, narrow tolerances of ± 2.4 Gy and ± 3 mm were proposed for dosimetric and positional accuracy of the major salivary glands, respectively, based on the dose response obtained for the glands. These facilitate the evaluation of performance of position verification methods and the improvement of the prediction of spared salivary gland function. The current results combined with literature data suggest, that the proposed tolerances can be met on the average if all systematic positional errors detected from the bony landmarks near the glands can be corrected.

ACKNOWLEDGEMENTS

This thesis is based on work done in a research project at the Radiation and Nuclear Safety Authority (STUK, Finland) and on clinical work at the Department of Oncology, Helsinki University Central Hospital (HUCH, Finland), between the years 2004 and 2008. Both of these institutions are greatly acknowledged. Especially I wish to express my gratitude to Eero Kettunen, M.Sc., Director of Radiation Practices Regulation, for the opportunity to work in the research project in STUK.

I warmly acknowledge Professor Heikki Joensuu, Head of the Department of Oncology (HUCH), for the opportunity to work in the clinic and analyze clinical data.

I am most grateful to my supervisor Docent Mikko Tenhunen, Ph.D., Chief Physicist of the Department of Oncology (HUCH) for proposing me to the research project in STUK, guiding throughout this research, valuable advice, brilliant ideas, always positive and encouraging attitude. I admire his ability to analyze problems from every perspective and to utilize every observation to constantly develope radiation therapy procedures with uncompromising effort.

I express deep gratitude to my other supervisor Professor Juhani Keinonen, Head of the Department of Physics, University of Helsinki, for guidance to accomplish this work.

I am greatly indebted to the official reviewers of this thesis, Docent Tapani Lahtinen, Ph.D. and Docent Maunu Pitkänen, Ph.D. for their valuable comments and constructive criticism improving the quality of the thesis.

I owe my most profound thanks to all the co-authors of the publications included in this thesis, Hannu Järvinen, M.Sc.Tech., Docent Jorma Heikkonen, Ph.D. and Docent Kalevi Kairemo, M.D., Ph.D.. Especially I acknowledge Petri Sipilä, M.Sc., Tuomo Hämäläinen, M.Sc. and Ritva Bly, Lic.Phil. for providing dose measurement data for the analysis and for several interesting discussions, and Juhani Collan, M.D. and Kauko Saarilahti, M.D., Ph.D. for the opportunity of analyzing a part of patient data of a large prospective IMRT investigation at the Department of Oncology (HUCH).

The physicists of the Department of Oncology (HUCH), are greatly acknowledged for the giving of their quality control data for the analysis and for pleasant and co-operative atmosphere at the clinic. I sincerely thank the physicists in the radiotherapy centres of Jyväskylä, Kotka, Kuopio, Lappeenranta, Oulu, Tampere, Turku and Vaasa for providing their beam data for the national analysis.

I owe my warmest gratitude to my dear wife, Satu, for her love, unconditional patience, encouragement and sacrifices. Finally, but not least, I feel deep gratitude to my mother, Terttu, for her support, loving patience and sacrifices. Without these this thesis would never have been accomplished.

Helsinki, May 2009
Mika Kapanen

REFERENCES

Aguirre J, Tailor R, Ibbott G, *et al*.. Thermoluminescence dosimetry as a tool for the remote verification of output for radiotherapy beams: 25 years of experience. In: Proceeding of the International Symposium on Standards and Codes of Practice in Medical Radiation Dosimetry. IAEA-CN-96/82. Vienna, Austria, IAEA; 2002, p. 191-199.

Ahnesjö A. Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media. Med. Phys. 1989; 16:577-592.

Ahnesjö A and Aspradakis MM. Dose calculations for external photon beams in radiotherapy. Phys. Med. Biol. 1999; 44:R99-R155.

Aird E, Mayles P and Mubata C. Quality control of megavoltage equipment. In: Handbook of radiotherapy physics, Theory and practice. Mayles P, Nahum A and Rosenwald JC (eds.). New York, NY, USA, Taylor&Francis; 2007, p. 831.

Akaike H. A new look at the statistical model identification. IEEE Trans. Automat. Control 1974; 19:716-723.

American Association of Physicists in Medicine (AAPM). Basic applications of multileaf collimators. Radiation Therapy Committee Task Group 50. AAPM report no. 72. Madison, WI, USA, AAPM; 2001.

American Association of Physicists in Medicine (AAPM). Intensity-Modulated Radiation Therapy. The State of the Art. Medical Physics Monograph no. 29. Madison, WI, USA, AAPM; 2003.

Ballivy O, Parker W, Vuong T, *et al*.. Impact of geometric uncertainties on dose distribution during intensity modulated radiotherapy of head-and-neck cancer: the need for a planning target volume and a planning organ-at-risk volume. Current Oncol. 2006; 13:108-115.

Batho H. Lung corrections in Cobalt 60 beam therapy. J. Can. Ass. Radiol. 1964; 15:39-83.

Bencomo J, Lowenstein J, Davis D, *et al*.. Analysis of Radiological Physics Center remote tools program data. Med. Phys. 1999; 26:1161.

Biggs PJ. Review of the energy check of an electron-only linear accelerator over a 6 year period: sensitivity of the technique to energy shift. Med. Phys. 2003; 30:635-639.

Bjärngard BE, Vadash P and Ceberg CP. Quality control of measured x-ray beam data. Med. Phys. 1997; 24:1441-1444.

Blad B, Nilsson P and Knöös T. The influence of air humidity on an unsealed ionization chamber in a linear accelerator. Phys. Med. Biol. 1996; 41:2541-2548.

Bohuslavizki K, Brenner W, Lassmann S, *et al*.. Quantitative salivary gland scintigraphy – A recommended examination prior to and after radioiodine therapy. Nuklearmedizin 1997; 36:103-109.

Bortfeld T, Schmidt-Ullrich R, De Neve W and Wazer DE. Image-Guided IMRT. Berlin, Heidelberg, New York, Springer-Verlag; 2006.

Bouchard H and Carrier J. Reducing the frequency of linac output check: a statistical model of linac output fluctuation based on a 3 year history to evaluate new tolerances as a function of test frequency for 12 linear accelerators. Med. Phys. 2007; 34:2612.

Boudreau C, Health E, Seuntjens J, *et al*.. IMRT head and neck treatment planning with a commercially available Monte Carlo based planning system. Phys. Med. Biol. 2005; 50:879-890.

Bragg WH. Studies in radioactivity. New York, NY, USA, Macmillan; 1912, p. 94.

Brahme A. Dosimetric precision requirements in radiation therapy. Acta Radiol. Oncol. 1984; 23:379-391.

Brahme A. Optimization of stationary and moving beam radiation therapy techniques. Radiother. Oncol. 1988; 12:129-140.

Brahme A. Treatment optimization using physical and radiobiological objective functions. In: Radiation Therapy Physics. Smith A (ed). Berlin, Germany, Springer; 1995.

Brahme A, Chavaudra J, Landberg T, *et al*.. Accuracy requirements and quality assurance of external beam therapy with photons and electrons. Acta Oncol. (Stoch) 1988; suppl. 1:1-76.

Brahme A, Roos J and Lax I. Solution of an integral equation in rotation therapy. Phys. Med. Biol. 1982; 27:1221-1229.

British Institute of Radiology (BIR). Central axis depth dose data for use in radiotherapy. Br. J. Radiol. 1996; 25:62-109.

British Institute of Radiology (BIR). Geometric uncertainties in radiotherapy. Defining the planning target volume. London, BIR; 2003.

Burlin TE. Further examination of theories relating the absorption of gamma-ray energy in a medium to the ionization produced in a cavity. Phys. Med. Biol. 1966; 11:255-266.

Burman C, Kutcher G, Emami B and Goitein M. Fitting of normal tissue tolerance data to an analytical function. Int. J. Radiat. Oncol. Biol. Phys. 1991; 21:123-135.

Chen Y, Michalski D, XiaoY and Galvin J. Automatic aperture selection and IMRT plan optimisation by beamweight renormalisation. Int. J. Radiat. Oncol. Biol. Phys. 2001; 51:74-75.

Cheung R, Tucker S, Dong L and Kuban D. Dose-response for biochemical control among high-risk prostate cancer patients after external beam radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. 2003; 56:1234–1240.

Cheung R, Tucker S, Lee A, *et al*.. Dose-response characteristics of low- and intermediate-risk prostate cancer treated with external beam radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. 2005a; 61:993-1002.

Cheung R, Tucker S, Lee A, *et al*.. Assessing the impact of an alternative biochemical failure definition on radiation dose response for high-risk prostate cancer treated with external beam radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. 2005b; 61:14–19.

Cho S, Vassiliev O, Lee S, *et al*.. Reference photon dosimetry data and reference phase space data for the 6 MV photon beam from Varian Clinac 2100 series linear accelerators. Med. Phys. 2005; 32:137-148.

Coffey C, Beach J, Thompson D and Mendiondo M. X-ray beam characteristics of the Varian Clinac 6-100 linear accelerator. Med. Phys. 1980; 7:716-722.

Cozzi L and Fogliata-Cozzi A. Quality assurance in radiation oncology. A study of feasibility and impact on action levels of an in vivo dosimetry program during breast cancer irradiation. Radiother. Oncol. 1998; 47:29-36.

Crichton R, Sharpe P, Haas O, *et al*.. Modelling parameters variation for linear accelerator: a quality control issue. 2nd Workshop on European Scientific and Industrial collaboration WESIC 99, Newport, UK, Sept. 1st-3rd, 1999; pp. 257-263.

Dasu A, Toma-Dasu I and Karlsson M. The effects of hypoxia on the theoretical modelling of tumour control probability. Acta Oncol. 2005; 44:563–571.

Delaney G, Jacob S, Featherstone C and Barton M. The role of radiotherapy in cancer treatment. Cancer 2005; 104:1129-1137.

Demarco J, Solberg T and Smathers. A CT-based Monte Carlo simulation tool for dosimetry planning and analysis. Med. Phys. 1998; 25:1-11.

Dijkema T, Terhaard C, Roesink J, *et al*.. Large cohort dose-volume response analysis of parotid gland function after radiotherapy: intensity-modulated versus conventional radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. 2008; 72:1101-1109.

Dunscombe P, Johnson H, Arsenault C, *et al*.. Development of quality control standards for radiation therapy equipment in Canada. J. Appl. Clin. Med. Phys. 2007; 8:108-118.

Dutreix A, Derremaux S, Chavaudra J and Van der Schueren E. Quality control of radiotherapy centres in Europe: beam calibration. Radiother. Oncol. 1994; 32:256-264.

Eisbruch A, Ten-Haken R, Kim H, *et al*.. Dose, volume, and function relationships in parotid salivary glands following conformal and intensity-modulated irradiation of head and neck cancer. Int. J. Radiat. Oncol. Biol. Phys. 1999; 45:577-587.

Elkind M, Swain R, Alescio T, *et al*.. Oxygen, nitrogen, recovery and radiation therapy. In: Cellular radiation biology. Shalek R (ed.). Baltimore, MD, Williams&Wilkins; 1965, p. 442-466.

Elshaikh M, Ljungman M, Ten Haken R and Lichter A. Advances in radiation oncology. Annu. Rev. Med. 2006; 57:19-31.

European Society for Therapeutic Radiology and Oncology (ESTRO). Quality assurance of treatment planning systems. Practical examples for non-IMRT photon beams. Booklet no. 7. Brussels, Belgium, ESTRO; 2004.

European Society for Therapeutic Radiology and Oncology (ESTRO). Guidelines for the verification of IMRT. Booklet no. 9. Brussels, Belgium, ESTRO; 2008.

Ferreira IH, Dutreix A, Bridier A, *et al*.. The ESTRO-QUALity assurance network (EQUAL). Radiother. Oncol. 2000; 55:273-284.

Firat F, Cermik T, Sarikaya A and Berkarda S. Effects of gender and age on the quantitative parameters of ^{99m}Tc pertechnetate salivary gland scintigraphy in normal subjects. Nucl. Med. Commun. 2006; 27:447-453.

Fischer J and Moulder J. The steepness of the dose-response curve in radiation therapy. Theoretical considerations and experimental results. Radiology 1975; 117:179–184.

Followill D, Lowenstein J and Ibbott G. High-energy photon standard dosimetry data: a quality assurance tool. Med. Phys. 2004; 31:1782.

Fontenla DP, Napoli JJ and Chui CS. Beam characteristics of a new model of 6-MV linear accelerator. Med. Phys. 1992; 19:343-349.

Fowler J, Chappell R and Ritter M. Is α/β for prostate tumors really low? Int. J. Radiat. Oncol. Biol. Phys. 2001; 50:1021–1031.

Fraass B, Doppke K, Hunt M, *et al*.. American Association of Physicists in Medicine (AAPM). Radiation Therapy Committee Task Group 53: Quality assurance for clinical radiotherapy treatment planning. Med. Phys. 1998; 25:1773-1829.

Gray LH. An ionization method for the absolute measurement of gamma-ray energy. Proc. Roy. Soc. 1936; A156:578-596.

Greene D. Linear accelerators for radiation therapy. In: Medical Physics Handbooks 17. Lenihan JMA (ed.). Bristol, UK, Adam Hilger Ltd.; 1986, p. 151.

Guckenberger M, Meyer J, Vordermark D, *et al*.. Magnitude and clinical relevance of translational and rotational patient setup errors: a cone-beam CT study. Int. J. Radiat. Oncol. Biol. Phys. 2006; 65:934-942.

Haas O, Mills J, Crichton R, *et al*.. Systems modelling approach to quality control for linear accelerators. J. Comput. Sim. Mod. Med. 2000; 1:46-53.

Han C, Chen YJ, Liu A, *et al*.. Actual dose variation of parotid glands and spinal cord for nasopharyngeal cancer patients during radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. 2008; 70:1256-1262.

Hanson W, Shalek R and Kennedy P. Dosimetry quality assurance in the United States from the experience of the Radiological Physics Center. In: Quality Assurance in Radiotherapy Physics. Starkschall G and Horton J (eds.). Madison, WI, USA, Medical Physics Publishing; 1991.

Holthusen H. Erfahrungen über die Verträglichkeitsgrenze für Röntgenstrahlen und deren Nutzanwendung zur Verhütung von Schäden. Strahlenther. 1936; 57:254-269.

Hong T, Tome W, Chappell R, *et al*.. The impact of daily setup variations on head-and-neck intensity-modulated radiation therapy. Int. J. Radiat. Oncol. Biol. Phys. 2005; 61:779-788.

Hurkmans C, Remeijer P, Lebesque J and Mijnheer B. Set-up verification using portal imaging; review of current clinical practice. Radiother. Oncol. 2001; 58:105-120.

Institute of Physics and Engineering in Medicine (IPEM). Physics aspects of quality control in radiotherapy. Report 81. York, UK, IPEM; 1999.

International Atomic Energy Agency (IAEA). Absorbed dose determination in external beam radiotherapy. Code of practice for high energy photon beams. TRS-398. Vienna, Austria, IAEA; 2000.

International Atomic Energy Agency (IAEA). Commissioning and quality assurance of computerized planning systems for radiation treatment of cancer. TRS-430. Vienna, Austria, IAEA; 2004.

International Commission on Radiation Units and Measurements (ICRU). Determination of absorbed dose in a patient irradiated by beams of x or gamma rays in radiotherapy procedures. Publication 24. Bethesda, MD, ICRU; 1976.

International Commission on Radiation Units and Measurements (ICRU). Use of computers in external beam radiotherapy procedures with high-energy photons and electrons. Report 42. Bethesda, MD, ICRU; 1987.

International Commission on Radiation Units and Measurements (ICRU). Prescribing, recording and reporting photon beam therapy. Report 50. Bethesda, MD, ICRU; 1993.

International Commission on Radiation Units and Measurements (ICRU). Prescribing, recording and reporting photon beam therapy. Supplement to Report 50. Report 62. Bethesda, MD, ICRU; 1999.

International Commission on Radiation Units and Measurements (ICRU). Prescribing, recording and reporting IMRT. ICRU Report (no. unavailable). Washington DC, USA, ICRU; in press, 2009.

International Electrotechnical Commission (IEC). Medical electrical equipment. Medical electron accelerators in the range 1 MeV to 50 MeV. Guidelines for performance characteristics. IEC 977. Geneva, Switzerland, IEC; 1989.

International Organization for Standardization (ISO). Quality management and quality assurance: Vocabulary. ISO 8402. Geneva, Switzerland, ISO; 1994.

International Organization for Standardization (ISO). Guide to the expression of uncertainty in measurement, 2nd edn. Geneva, Switzerland, ISO; 1995.

Izewska J and Andreo P. The IAEA/WHO TLD postal programme for radiotherapy hospitals. Radiother. Oncol. 2000; 54:65-72.

Järvinen H, Sipilä P, Parkkinen R, *et al*.. Quality systems for radiotherapy: impact by a central authority for improved accuracy, safety and accident prevention. Presented at the IAEA Symp. on Radiological Protection of Patients in Diagnostic and Interventional Radiology, Nuclear Medicine and Radiotherapy, Malaga 2001.

Kallman RF. The phenomenon of reoxygenation and its implication for fractionated radiotherapy. Radiol. 1972; 105:135-142.

Kestin L, Sharpe M, Frazier R, *et al*.. Intensity modulation to improve dose uniformity with tangential breast radiotherapy: initial clinical experience. Int. J. Radiat. Oncol. Biol. Phys. 2000; 48:1559-1568.

Khan F. The physics of radiation therapy, 3rd ed. Philadelphia, PA, Lippincott Williams & Wilkins; 2003.

Kim S, Zhu T and Palta J. An equivalent square field formula for determining head scatter factors of rectangular fields. Med. Phys. 1997; 24:1770-1774.

Knöös T, Ahnesjö A, Nilsson P and Weber L. Limitations of a pencil beam approach to photon dose calculations in lung tissue. Phys. Med. Biol. 1995; 40:1411-1420.

Kosunen A, Järvinen H, Vatnitskij S, *et al*.. Intercomparison of radiotherapy treatment planning systems for external photon and electron beam dose calculations. Radiother. Oncol. 1993; 29:327-335.

Kristensen I, Björk P, Englund E, *et al*.. Quality control of linear accelerators. Radiother. Oncol. 2003; 68:S65.

Krutman Y, El-Dan O, Rachamim B, *et al*.. The quality assurance equipment program for a 200-patients/day RT department with 3 linacs. Radiother. Oncol. 2004; 73:S415.

Kutcher G, Burman C, Brewster M, *et al*.. Histogram reduction method for calculating complication probabilities for three-dimensional treatment planning evaluations. Int. J. Radiat. Oncol. Biol. Phys. 1991; 21:137-146.

Kutcher G, Coia L, Gillin M, *et al*.. Comprehensive QA for radiation oncology: Report of AAPM Radiation Therapy Committee Task Group 40. Med. Phys. 1994; 21:581-618.

Källman P, Ågren A and Brahme A. Tumour and normal tissue responses to fractionated non-uniform dose delivery. Int. J. Radiat. Biol. 1992; 62:249-262.

Lee N, Akazawa C, Akazawa P, *et al*.. A forward-planned treatment technique using multisegments in the treatment of head-and-neck cancer. Int. J. Radiat. Biol. Phys. 2004; 59:584-594.

Lee C, Langen K, Kupelian P, *et al*.. Positional and volumetric changes in parotid glands during head and neck radiation therapy assessed using deformable image registration. Int. J. Radiat. Oncol. Biol. Phys. 2007; 69:S203.

Lee C, Langen K, Weiguo L, *et al*.. Assessment of parotid gland dose changes during head and neck cancer radiotherapy using daily megavoltage computed tomography and deformable image registration. Int. J. Radiat. Oncol. Biol. Phys. 2008a; 71:1563-1571.

Lee C, Langen K, Weiguo L, *et al*.. Evaluation of geometric changes of parotid glands during head and neck cancer radiotherapy using daily MVCT and automatic deformable registration. Radiother. Oncol. 2008b; 89:81-88.

Levegrün S, Jackson A, Zelefsky M, *et al*.. Fitting tumor control probability models to biopsy outcome after three-dimensional conformal radiation therapy of prostate cancer: pitfalls in deducing radiobiological parameters for tumors from clinical data. Int. J. Radiat. Oncol. Biol. Phys. 2001; 51:1064-1080.

Linthout N, Verellen D, Tournel K and Storme G. Six dimensional analysis with daily stereoscopic X-ray imaging of intrafractional patient motion in head and neck treatments using five points fixation masks. Med. Phys. 2006; 33:504-513.

Liu H, Mackie T and McCullough E. Calculating output factors for photon beam radiotherapy using a convolution/superposition method based on a dual source photon beam model. Med. Phys. 1997; 24:1975-1985.

Luketina IA and Greig L. Linear accelerator output variability. Austral. Phys. Engin. Scien. Med. 2004; 27:155-159.

Lyman JT. Complication probability as assessed from dose-volume histograms. Radiat. Res. 1985; 104:S13-S19.

Manning M, Quiwen W, Cardinale R, *et al*.. The effect of setup uncertainty on normal tissue sparing with IMRT for head-and-neck cancer. Int. J. Radiat. Oncol. Biol. Phys. 2001; 51:1400-1409.

Mayneord W and Lamerton L. A survey of depth dose data. Br. J. Radiol. 1944; 14:255.

McKenzie A. What is the optimum frequency of quality control tests in radiotherapy? Radiother. Oncol. 2003; 68:S74.

McKenzie A, Briggs G, Buchanan R, *et al*.. Cost-benefit analysis of quality control in UK. Radiother. Oncol. 2005; 76 (Suppl. 2):S33.

McKenzie A, Briggs G, Buchanan R, *et al*.. Balancing costs and benefits of checking in radiotherapy. Report 92. York, UK, IPEM; 2006.

Mechalakos J, Hunt M, Lee N, *et al*.. Using an onboard kilovoltage imager to measure setup deviation in intensity-modulated radiation therapy for head-and-neck patients. J Appl. Clin. Med. Phys. 2007; 8:28-44.

Meijer G, van Kleffens H and Mijnheer B. Consistency in quality control programmes for electron accelerators in radiotherapy centres. Radiother. Oncol. 1998; 48:103-110.

Meyer JL, ed.. IMRT.IGRT.SBRT. Advances in the Treatment Planning and Delivery of Radiotherapy. Basel, Switzerland, S. Karcher AG; 2007.

Mihaylov IB, Lerma FA, Fatyga M and Siebers JV. Quantification of the impact of MLC modeling and tissue heterogeneities on dynamic IMRT dose calculations. Med. Phys. 2007; 34:1244-1252.

Mijnheer B, Battermann J and Wambersie A. What degree of accuracy is required and can be achieved in photon and neutron therapy? Radiother. Oncol. 1987; 8:237-252.

Mu G, Ludlum E and Xia P. Impact of MLC leaf position errors on simple and complex IMRT plans for head and neck cancer. Phys. Med. Biol. 2008; 53:77-88.

Murdoch-Kinch C, Kim H, Vineberg K, *et al*.. Dose-effect relationships for the submandibular salivary glands and implications for their sparing by intensity modulated radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. 2008; 72:373-382.

Münter M, Karger C, Hoffner S, *et al*.. Evaluation of salivary gland function after treatment of head-and-neck tumors with intensity-modulated radiotherapy by quantitative pertechnetate scintigraphy. Int. J. Radiat. Oncol. Biol. Phys. 2004; 58:175-184.

Netherlands Commission on Radiation Dosimetry (NCS). Quality control of medical linear accelerators. Current practice and minimum requirements. NCS report 9. Delft, Netherlands Measurements Institute; 1996.

Niemierko A. Reporting and analyzing dose distributions: a concept of equivalent uniform dose. Med. Phys. 1997; 24:103-110.

Niemierko A. A generalised concept of equivalent uniform dose (EUD). Med. Phys. 1999; 26:1100.

Nordic Association of Clinical Physics (NACP). Procedures in external radiation therapy dosimetry with electron and photon beams with maximum energies between 1 and 50 MeV. Acta Radiol. Oncol. 1980; 19:55–79.

O`Daniel J, Garden A, Schwartz D, *et al*.. Parotid gland dose in intensity-modulated radiotherapy for head and neck cancer: is what you plan what you get? Int. J. Radiat. Oncol. Biol. Phys. 2007; 69:1290-1296.

Okunieff P, Morgan D, Niemierko A and Suit H. Radiation dose-response of human tumors. Int. J. Radiat. Oncol. Biol. Phys. 1995; 32:1227-1237.

Pawlicki T, Whitaker M and Boyer A. Statistical process control for radiotherapy quality assurance. Med. Phys. 2005; 32:2777–2786.

Ploquin N, Kay I, Ranger-Baltazar A, *et al*.. A comparison of techniques for simulating set-up error and uncertainty in head and neck IMRT. Med. Phys. 2006; 33:3213-3219.

Prabhakar R, Ganesh T, Haresh K, *et al*.. Impact of patient setup error in the treatment of head and neck cancer with intensity modulated radiation therapy. Int. J. Radiat. Oncol. Biol. Phys. 2007; 69S:S462.

Puurunen H, Lahtinen T, Vainio P, *et al*.. Experiences about the quality control of the new linear accelerators. Med. Biol. Engin. Comput. 1985; 23(S1):637-638.

Ravichandran R, Binukumar J, Davis C, *et al*.. Evaluation methods for detecting changes in beam output and energy in radiation beams from high-energy linear accelerators. J. Med. Phys. 2007; 32:92-96.

Robar J, Day A, Clancey J, *et al*.. Spatial and dosimetric variability of organs at risk in head-and-neck intensity-modulated radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. 2007; 68:1121-1130.

Roth J, Roser HW, Nemec HW, *et al*.. Quality assurance of physical parameters in radiation oncology at the University Hospital of Basel - a retrospect. Radiother. Oncol. 1998; 48:83-88.

Saarilahti K, Kouri M, Collan J, *et al*.. Intensity modulated radiotherapy for head and neck cancer: evidence for preserved salivary gland function. Radiother. Oncol. 2005; 74:251-258.

Saarilahti K, Kouri M, Collan J, *et al*.. Sparing of the submandibular glands by intensity modulated radiotherapy in the treatment of head and neck cancer. Radiother. Oncol. 2006; 78:270-275.

Sakthi N, Keall P, Mihaylov I, *et al*.. Monte Carlo-based dosimetry of head-and-neck patients treated with SIB-IMRT. Int. J. Radiat. Oncol. Biol. Phys. 2006; 64:968-977.

Samuelsson A, Mercke C and Johansson KA. Systematic set-up errors for IMRT in the head and neck region: effect on dose distribution. Radiother. Oncol. 2003; 66:303-311.

Saw C, Korb L and Krishna K. Local temperature and pressure variations and its effects on daily ionometric measurements of radiation beam outputs from linear accelerators. Med. Dosim. 1997; 22:1-4.

Schultheiss T, Orton C and Peck R. Models in radiotherapy: volume effects. Med. Phys. 1983; 10:410-415.

Schwarz M, Bos LJ, Mijnheer BJ, *et al*.. Importance of accurate dose calculations outside segment edges in intensity modulated radiotherapy treatment planning. Radiother. Oncol. 2003; 69:305-314.

Seuntjens J, Olivares M, Evans M and Podgorsak E. Absorbed dose to water reference dosimetry using solid phantoms in the context of absorbed dose protocols. Med. Phys. 2005; 32:2945-2953.

Siebers J, Keall P, Qiuwen W, *et al*.. Effect of patient setup errors on simultaneously integrated boost head and neck IMRT treatment plans. Int. J. Radiat. Oncol. Biol. Phys. 2005; 63:422-433.

Storchi P, van Battum L and Woudstra E. Calculation of a pencil beam kernel from measured photon beam data. Phys. Med. Biol. 1999; 44:2917-2928.

Storchi P and van Gasteren J. A table of phantom scatter factors of photon beams as a fuction of the quality index and field size. Phys. Med. Biol. 1996; 41:563-571.

Suit HD. Radiation biology: a basis for radiotherapy. In: Textbook of radiotherapy. Fletcher GH (ed.). Philadelphia, PA, Lea&Febiger; 1966.

Svensson H, Hanson G and Zsdansky K. The IAEA/WHO TL dosimetry programme for radiotherapy centres 1969-1987. Acta Oncol. 1990; 29:461-467.

Svensson R, Källman P and Brahme A. Analytical solution for the dynamic control of multileaf collimators. Phys. Med. Biol. 1994; 39:37-61.

Swiss Society of Radiobiology and Medical Physics (SSRPM/SGSMP). Quality control of medical electron accelerators. Recommendations nr 11. Bern, Switzerland, SSRPM; 2003.

Tenhunen M. Photon beam output factors of medical linear accelerators with special reference to asymmetric collimation and dynamic wedge fields. Doctoral dissertation. Kuopio, Finland, Kuopio University Printing Office; 1994.

Tenhunen M, Collan J, Kouri M, *et al*.. Scintigraphy in prediction of the salivary gland function after gland-sparing intensity modulated radiation therapy for head and neck cancer. Radiother. Oncol. 2008; 87:260-267.

Thames H, Schultheiss T, Tucker S, *et al*.. Can modest escalations of dose be detected as increased tumor control ? Int. J. Radiat. Oncol. Biol. Phys. 1992; 22:241-246.

Thwaites D, Scalliet P, Leer J and Overgaard J. Quality assurance in radiotherapy. Radiother. Oncol. 1995; 35:61-75.

Thwaites D, Williams J, Aird E, *et al*.. A dosimetric intercomparison of megavoltage photon beams in UK radiotherapy centres. Phys. Med. Biol. 1992; 37:445-461.

Tuncel N, Kizildag A, Ibili A, *et al*.. The in-house management of data: analyses of output stability and performance of elekta SLI25 linac. Radiother. Oncol. 2004; 73:S412.

Vadash P and Bjärngard B. An equivalent square formula for head scatter factors. Med. Phys. 1993; 20:733-734.

Vakilha M, Hwang D, Breen S, *et al*.. Changes in position and size of parotid glands assessed with daily cone-beam CT during image-guided IMRT for head and neck cancer: implications for dose received. Int. J. Radiat. Oncol. Biol. Phys. 2007; 69S:S438.

Van Dyk J, Barnett R, Cygler J and Shragge P. Commissioning and quality assurance of treatment planning computers. Int. J. Radiat. Oncol. Biol. Phys. 1993; 26:261-273.

Van Esch A, Bogaerts R, Kutcher G and Huyskens D. Quality assurance in radiotherapy by identifying standards and monitoring treatment preparation. Radiother. Oncol. 2000; 56:109-115.

Van Esch A, Bohsung J, Sorvari P, *et al*.. Acceptance tests and quality control (QC) procedures for the clinical implementation of intensity modulated radiotherapy (IMRT) using inverse planning and the sliding window technique: experience from five radiotherapy departments. Radiother. Oncol. 2002; 65:53-70.

Van Gasteren J, Heukelom S, van Kleffens H, *et al*.. The determination of phantom and collimator scatter components of the output of megavoltage photon beams: measurement of the scatter part with a beam-coaxial narrow cylindrical phantom. Radiother. Oncol. 1991; 20:250-257.

Venselaar J and Welleweerd H. Application of a test package in an intercomparison of the performance of treatment planning systems used in a clinical setting. Radiother. Oncol. 2001; 60:203-213.

Venselaar J, Welleweerd H and Mijnheer B. Tolerances for the accuracy of photon beam dose calculations of treatment planning systems. Radiother. Oncol. 2001; 60:191-201.

Watanabe Y. Analysis of daily beam output data for teletherapy machines. Engin. Med. Biol. Soc. 2000; 2:1243-1245.

Watts RJ. Comparative measurements on a series of accelerators by the same vendor. Med. Phys. 1999; 26:2581-2585.

Webb S. The Physics of Conformal Radiotherapy. Bristol, UK, Institute of Physics Publishing; 1997.

Withers HR. Biological aspects of conformal therapy. Acta Oncol. 2000; 39:569-577.

World Health Organisation (WHO). Quality assurance in radiotherapy. Geneva, Switzerland, WHO; 1988.

Xiao Y, Altschuler MD and Bjärngard BE. Quality assurance of central axis dose data for photon beams by means of a functional representation of the tissue phantom ratio. Phys. Med. Biol. 1998; 43:2195-2206.

Xiao Y, Galvin J, Hossain M and Valicenti R. An optimized forward-planning technique for intensity modulated radiation therapy. Med. Phys. 2000; 27:2093-2099.

Young M and Gaylord J. Experimental tests of corrections for tissue inhomogeneities in radiotherapy. Br. J. Radiol. 1970; 43:349-355.

Yu M, Sloboda R and Mansour F. Measurement of photon beam backscatter from collimators to the beam current monitor chamber using target current pulse counting and telescope techniques. Phys. Med. Biol. 1996; 41:1107-1117.

Zagars G, Schultheiss T and Peters L. Inter-tumor heterogeneity and radiation dose control curves. Radiother. Oncol. 1987; 8:353-362.

Zhu T, Bjärngard B, Xiao Y and Yang CJ. Modeling the output ratio in air for megavoltage photon beams. Med. Phys. 2001; 28:925-937.