Results of the

IMRT dosimetry intercomparison 2008

I Introduction

An issue often discussed in radiotherapy is the ability of the planning systems to take into account inhomogeneities, especially in the thorax region. For IMRT, dose calculation problems can be enhanced due to partly very small field segments. So a national intercomparison considering this problem is appropriate to check the ability of the calculation algorithms used in the own clinic and to raise a discussion which helps to sensitize the participants to this topic. Until now, no intercomparisons were available to test this issue. Consequently, it has been decided to perform a national intercomparison in Switzerland dealing with IMRT in the thorax region. The intercomparison has been organized by the team of the Cantonal Hospital of St.Gallen.

Ahead of the IMRT dosimetry intercomparison, a pilot study with six participants has been conducted in order to test the reliability of the film and the TLD dosimetry in the phantom environment. The results are presented in the Bulletin 2/2008 [1] of our society.

Some institutions tested more than one algorithm or tested an algorithm in a situation which they knew as essentially inadequate.

So, we want to stress that larger deviations <u>give no information</u> on the quality of the irradiation process achieved by an institution!

II Materials and methods

A. General

For the IMRT intercomparison of the SGSMP, the thorax phantom 002LFC (CIRS Inc.) has been used. A standard slice has been modified with drillings to accommodate the TLDs (Figure 1).

The CT scan has been carried out by the institutions themselves. The applied CT dose has been measured with additional TLDs, attached to the phantom surface.

TLD-100 discs (4.5 mm Ø x 0.9 mm; Harshaw Inc.) and a TLD reader model “5500” (Harshaw Inc.) have been used. The tempering procedure has been done in a PTW-TLDO oven (PTW Freiburg). Reference irradiations were performed using a “Theratron 60” cobalt unit (AECL of Canada).

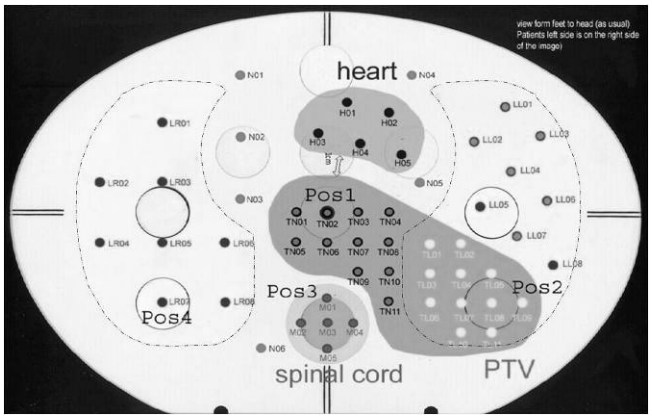

Fig. 1 Standard slices with drillings accommodating TLDs. Left: Used for the pilot study. Right: Used for the IMRT intercomparison

B. IMRT dosimetry intercomparison

The institutions carried out the ionisation chamber measurement with their own equipment. EDR2 films (Eastman Kodak Co.) have been developed (“Optimax 2010”, PROTEC GmbH) and scanned (“Diagnostic Pro”, Vidar Systems Corporation) by the physics team in St.Gallen. Additionally, a calibration film from the same batch was generated in St.Gallen.
All 23 institutions irradiating patients in Switzerland participated between July 2008 and February 2009. 24 machines have been tested. It has been suggested applying an IMRT technique, but other techniques were also accepted. Some institutions carried out the calculation with two different calculation algorithms (five) or participated twice in the intercomarison by applying different machines (one) or irradiation techniques (one). Altogether 30 plan-measurement combinations have been evaluated, which will be treated as independent in this study. Due to differing technical situations, some parameters were evaluated in less than 30 combinations. The applied calculation algorithms and irradiation techniques are shown in Table 2.

<u>Absolute dosimetry with TLDs</u>

A special slice (length: 6.3 cm) contains a cubic cavity in the "sternum" (see figure 2, right side). A “mini phantom” containing 8 TLDs can be placed in the cavity so that the depth of the TLDs is 10 cm. By applying a 10 cm square field (gantry angle: 0°, source to surface distance: 90 cm), the irradiation condition is comparable to a basic single field irradiation under standard conditions in water.

Fig. 2 Left: CIRS Thorax phantom with Perspex structure slices. Right: „mini phantom“, dedicated to check the absolute calibration

The institutions have calculated the dose $D_{s,P}$ to the TLDs in the phantom with the same algorithm as used for the treatment plan. Additionally, they have stated the dose under standard conditions in water, $D_{s,W}$, when applying the same number of monitor units. Ahead of the intercomparison, conversion factors $k_{P\rightarrow W}$ have been determined by measurements in St.Gallen: The same number of monitor units has been applied to TLDs in the phantom with the basic single field, as described above, and to TLDs in the water phantom under standard conditions. The conversion factors $k_{P\rightarrow W}$ is the ratio of these measurements. Similar irradiations have been repeated with an ionisation chamber. The conversion factors allow to calculate the dose under standard conditions in water, $D_{m,W}$, given the dose in the phantom under approximated standard conditions, $D_{m,P}$: $D_{m,W} = k_{P\rightarrow W} \times D_{m,P}$. For 6X, the conversion factor is 1.01. Thus, $D_{m,W}/D_{s,W}$ is a measure for the dose calibration of the machine and should be unity. $D_{m,P}/D_{s,P}$ does include the systematic errors which already arise in a homogeneous part of the phantom. Consequently, $(D_{m,P}/D_{s,P})^{-1}$ allows correcting the TLD measurements of the plan irradiation for systematic errors, originating from the planning process and the TLD measurement or the machine calibration. Due to technical reasons, for the TomoTherapy machine, measurements in the mini phantom have not been carried out.

Absolute dosimetry with an ionisation chamber

Slice 01 accommodates adapters for ionisation chamber measurements at different positions (see figure 1, Pos1 to Pos4). For Pos1, and analogous to the TLD measurements, conversion factors are available to calculate the dose expected in water under standard conditions. For 6X, the factor is 1.00. The same quantities can be checked as stated for the TLD measurements. This allows cross checking the TLD to the ionisation chamber measurements.

Contouring and calculation of the IMRT plan

Two identical Perspex slices form the longitudinal phantom ends (see figure 2, left side). They contain shapes needed for the contouring. All structures required for planning are placed symmetrically around the measurement plane. The PTV and the heart are 8 cm long, the other structures cover the entire phantom length. Hence, it can be expected that dose gradients in the longitudinal direction do not seriously affect the measurement accuracy. The transversal PTV area is about 70 cm^2. It covers parts of the left lung and parts of normal tissue (each containing 11 TLD measurement positions). Thus, the calculation algorithm can be reliably tested in both kinds of tissue. Other positions for TLDs outside the PTV are grouped in the right lung, the left lung (eight each), the spinal cord, and the heart (five each). Additional six TLD positions are distributed outside these structures. All together 54 different TLD positions have been evaluated. Each TLD measurement point consists of two TLD discs. The large number of absolute dosimeters allows a statistical analysis of the calculation accuracy in different parts of the phantom.
To avoid effects of air gaps between slices, the measurement planes are placed 5.0 cm off axis to the field isocentre.
The plan had to fulfill the following constraints: a) PTV: prescribed median Dose = 2.00 Gy. b) Spinal cord: < 75 % of the prescribed dose. c) Both lungs outside PTV: < 20 % of the lungs receive > 35 % of the prescribed dose. d) Heart: < 55 % of the prescribed dose.

Application of the IMRT plan and evaluation

Calculations have shown that the absorption properties of the phantom are invariant within 1 % in the longitudinal direction. So, measurements have been conducted in the same plane relative to the isocentre with TLDs, film and ionisation chambers (Pos1 to Pos4), but in different slices of the phantom (figure 3). This allows cross checking the measurements done in the same points relative to the isocentre.

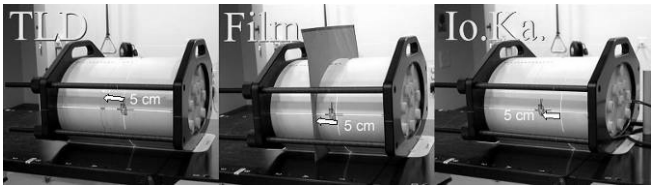

Fig. 3 Measurements in the CIRS thorax phantom with TLD, film and ionisation chamber.

Evaluation

The film measurements have been evaluated using the “Verisoft” software (PTW Freiburg). Due to file format inconsistencies, some DICOM dose distributions provided by the institutions could not be evaluated. We hope that the “Verisoft” version intended to be used in the next intercomparison is able to handle more formats then the actual one. The evaluation has been performed relatively by applying a scaling factor to the measured dose distribution. Since “Verisoft” does not support nu-

merical parameters which characterize the integral result of the gamma index evaluation and information about the outline of the phantom in the gamma index image is missing, it is difficult to do a meaningful evaluation. So the institutions are asked to interpret the results themselves.
For the evaluation, the algorithms used by the institutions are classified as "type a" and "type b" algorithms [2, 3]: "Type b" models are able to treat the electron transport in an approximate way as well as the secondary photon transport in the medium, accounting for density changes, sampled along the full three dimensions. "Type a" algorithms are 1D and primarily based on equivalent path length for inhomogeneity correction.

III Results

Structure volume measurements

Figure 4 shows the PTV volume measurements for 23 structures. 8 structures show a deviation larger than 2 % in respect to the median value (dotted line), one structure deviates for more than 5 %.

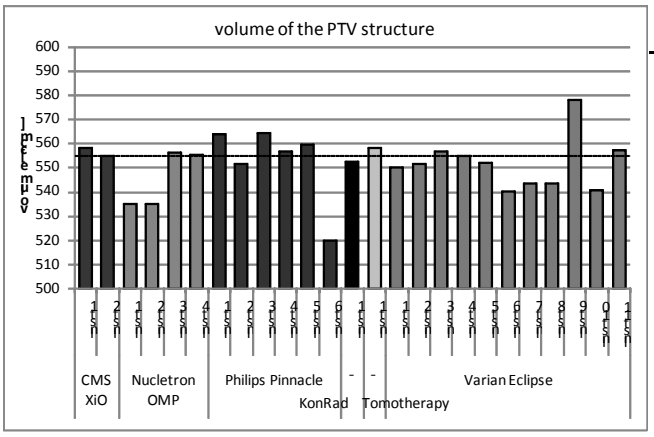

Fig. 4 volume measurements of the PTV structure (23 structures)

Some measurement values, especially for non PTV structures, support the assumption that the contoured structure length did not coincide with the length specified in the instructions which resulted in larger deviations. So a final interpretation has to be done by the institution itself.
It is generally accepted that the IMRT technique is more suitable to comply enhanced demands with the dose distribution than the 3D-CRT technique. Unexpectedly, the dose and percentage information stated by the institutions to describe the fulfillment of the constraints give no hint that static or dynamic IMRT techniques have advantages compared to the 3D-CRT technique. Also, in all groups of treatment techniques, large single outliers can be observed. Again, a final interpretation has to be done by the institution.

Check of the absolute dosimetry

Table 1 shows the mean ratios of the measured to the stated dose in the normal phantom tissue, $D_{m,P}/D_{s,P}$. $D_{m,W}/D_{s,W}$ describes the machine calibration. For this situation, no parameter shows a significant difference between "type a" (11 evaluations) and "type b" (17 evaluations) algorithms. The $D_{m,P}/D_{s,P}$ values for the TLD measurements are slightly higher than the ionisation chamber measurements. Different reasons can be considered: The angular orientation of the TLD disc and the phantom environment differs from the calibration conditions in water which can slightly influence the TLD sensitivity. Further measurements are scheduled on this topic. 16 from 28 TLD correction factors for systematic errors (57 %) are within 1.00 ± 0.01. The mean value is 1.005 ± 0.015, the mean absolute deviation from unity is 0.012 ± 0.010. The check of the machine calibration shows good results for both ion chamber and TLD measurements.

	$D_{m,P}/D_{s,P}$		$D_{m,W}/D_{s,W}$	
	ion. chamber	TLD	ion. chamber	TLD
"type a"	0.996 ± 0.006	1.012 ± 0.016	1.005 ± 0.006	1.000 ± 0.014
"type b"	0.987 ± 0.010	1.001 ± 0.013	1.008 ± 0.012	1.004 ± 0.014
all	0.990 ± 0.010	1.005 ± 0.015	1.007 ± 0.010	1.002 ± 0.014

Table 1 Results of the "mini phantom" measurements (28 evaluations)

TLD and ionisation chamber measurements at other positions

A measure for the accuracy of the applied inhomogeneity correction algorithms, independent from systematic deviations, is the difference between the values of $(D_m-D_s)/D_{prescribed}$ in the lung tissue and normal tissue within the PTV. Figure 5 shows an overview of the differences, separated for the ionisation chamber (Pos1 and Pos2) and TLD measurements. Both TLD and ionisation chamber measurements demonstrate the well known tendency of "type a" algorithms to overestimate the dose in the lung region.

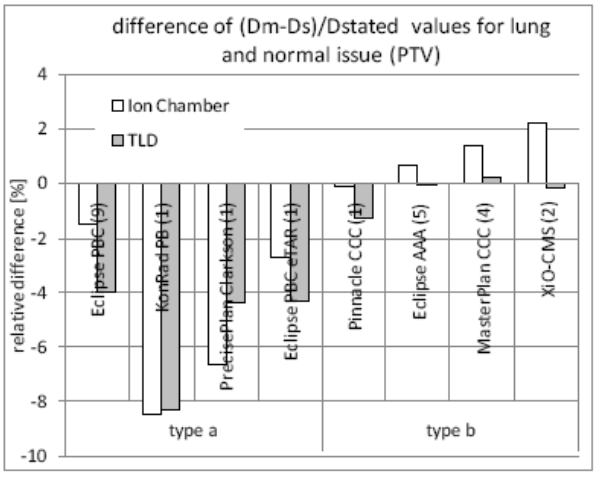

Fig. 5 Difference of the $(D_m-D_s)/D_{stated}$ values between the lung and normal tissue within the PTV

Mann-Whitney-tests for both measurement equipments prove unambiguously that "type b" algorithms are superior to "type a" algorithms in calculating the dose in the lung region ($p<0.001$). There is insufficient statistics to argue that the XiO CMS algorithm overcorrects for the lung tissue.
Table 2 shows statistics to the applied calculation algorithms and irradiation techniques.

		# of comp.			mean TLD msmt: $(D_m-D_s)/D_{prescribed}$ [%]						
	algorithm	3D-CRT	dyn. IMRT	static IMRT	PTV: Norm. t.	PTV: Lung t.	left lung	right lung	normal tissue	heart	spinal cord
„type a"	Eclipse PBC		7	2	-	-	-	-	-	-	-
	KonRad PB			1	-	-	-	-	-	-	-
	Prec.Plan int.			1	-	-	-	-	-	-	1.7
	Ecl. PBC	1			0.5	-	1.2	-	1.0	1.6	0.4
	mean "type a"	-	-	-	-	-	-	-	-	-	-
"type b"	Pinnacle CCC	2		4	0.8	-	0.8	0.1	0.7	0.5	-
	Eclipse AAA		4	1	-	-	-	-	-	-	-
	MasterPlan	3		1	1.8	2.1	0.0	1.2	0.7	0.4	1.5
	XiO-CMS			2	0.0	-	1.1	0.4	-	-	2.9
	Tomotherapy		1		-	-	0.8	-	-	0.4	-
	mean „type b"	-	-	-	*0.1*	-	*0.2*	*0.2*	*0.2*	*0.0*	*0.1*

Table 2. Left columns: Calculation type and irradiation technique statistics. Right columns: Mean results of the plan measurements for 30 evaluations (m=TLD measured; s=stated). 3D-CRT values are corrected for measurements in the high gradient area.

Regarding all TLD measurement groups, the mean absolute difference, related to the prescribed dose, is 3.0 ± 2.7 % for the "type a" and 1.9 ± 1.9 % for the "type b" algorithms. For regions outside the lungs, the figures are 2.2 ± 2.0 % and 1.9 ± 1.8 %. So, "type b" algorithms show no advantages in homogeneous regions compared with "type a" algorithms. The mean stated doses for the right lung, left lung, normal tissue, heart, spinal cord and structures are: 0.79, 1.06, 0.93, 0.77 and 1.18 Gy. This information helps to estimate the relative local difference between the TLD measured and stated doses.

IV Discussion and conclusion

Due to the limited number of participants, it is not possible to issue reliable statements on the properties of the single calculation algorithms. Nevertheless, there are some trends to observe: Generally, „Type b" algorithms take inhomogeneities better into account than "type a" algorithms. Some "type a" algorithms show deviations over 5 % in the PTV lung region, but there are still differences within the "type a" and the "type b" groups. Outside inhomogeneities, "type a" algorithms show in general good calculation results. This finding coincides with other statements [2, 3].
The intercomparison procedure has turned out to be feasible and yields convincing results. Although the effort for the participants is comparatively large, the feedback was mainly positive. In the future, the IMRT intercomparison will be repeated regularly with modified objectives.
The results of the intercomparison exceed the expectations. They suggest that cancer patients in Switzerland get a suitable radiation therapy in any of the centers offering this treatment modality.
Due to the reliability of all participants the intercomparison could be completed within the scheduled time frame.

We thank all institutions for their pleasing co-operation.

Hans Schiefer *Wolf Seelentag*

REFERENCES

Schiefer H, Seelentag WW, Results of the Pilot study to the IMRT dose intercomparison 2008. STSMP Bulletin 2/2008 66:14-16

Knöös T, Wieslander E, Cozzi L et al. (2006) Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situations. Phys Med Biol 51:5785-5807

Fogliata A, Vanetti E, Albers D et al. (2007) On the dosimetric behavior of photon dose calculation algorithms in the presence of simple geometric heterogeneities: comparison with Monte Carlo calculations. Phys Med Biol 52:1363-1385

With this article the authors have fulfilled their obligation arising from the SSRMP Research Grant.

Thank you, Hans Schiefer und Wolf Seelentag for your work for our society!

Luca Cozzi, President of

SGSMP
SSRPM
SSRFM