Journal Pre-proof

Planning benchmark study for SBRT of liver metastases: Results of the DEGRO/DGMP working group stereotactic radiotherapy and radiosurgery

C Moustakis , O Blanck , MKH Chan , J Boda-Heggemann , N Andratschke , M-N Duma , D Albers , C Bäumer , R Fehr , SA Körber , D Schmidhalter , M Alraun , WW Baus , E Beckers , M Dierl , S Droege , F Ebrahimi Tazehmahalleh , J Fleckenstein , M Guckenberger , C Heinz , C Henkenberens , A Hennig , J Köhn , C Kornhuber , T Krieger , B Loutfi-Krauss , M Mayr , M Oechsner , T Pfeiler , G Pollul , J Schöffler , H Tümmler , C Ullm , M Walke , R Weigel , M Wertman , R Wiehle , T Wiezorek , L Wilke , U Wolf , HT Eich , D Schmitt

PII: S0360-3016(22)00018-9
DOI: https://doi.org/10.1016/j.ijrobp.2022.01.008
Reference: ROB 27420

To appear in: *International Journal of Radiation Oncology, Biology, Physics*

Received date: 26 June 2021
Revised date: 19 December 2021
Accepted date: 7 January 2022

Please cite this article as: C Moustakis , O Blanck , MKH Chan , J Boda-Heggemann , N Andratschke , M-N Duma , D Albers , C Bäumer , R Fehr , SA Körber , D Schmidhalter , M Alraun , WW Baus , E Beckers , M Dierl , S Droege , F Ebrahimi Tazehmahalleh , J Fleckenstein , M Guckenberger , C Heinz , C Henkenberens , A Hennig , J Köhn , C Kornhuber , T Krieger , B Loutfi-Krauss , M Mayr , M Oechsner , T Pfeiler , G Pollul , J Schöffler , H Tümmler , C Ullm , M Walke , R Weigel , M Wertman , R Wiehle , T Wiezorek , L Wilke , U Wolf , HT Eich , D Schmitt , Planning benchmark study for SBRT of liver metastases: Results of the DEGRO/DGMP working group stereotactic radiotherapy and radiosurgery, *International Journal of Radiation Oncology, Biology, Physics* (2022), doi: https://doi.org/10.1016/j.ijrobp.2022.01.008

Planning benchmark study for SBRT of liver metastases: Results of the DEGRO/DGMP working group stereotactic radiotherapy and radiosurgery

Moustakis C[1§], Blanck O[2,21], Chan MKH[2], Boda-Heggemann J[3], Andratschke N[4], Duma M-N[5,6], Albers D[7], Bäumer C[8], Fehr R[9], Körber SA[10], Schmidhalter D[11], Alraun M[12], Baus WW[13], Beckers E[14], Dierl M[15], Droege S[16], Ebrahimi Tazehmahalleh F[1,13], Fleckenstein J[3], Guckenberger M[4], Heinz C[17], Henkenberens C[18], Hennig A[19], Köhn J[20,21], Kornhuber C[22], Krieger T[23], Loutfi-Krauss B[20,21], Mayr M[24], Oechsner M[5], Pfeiler T[8], Pollul G[25], Schöffler J[26], Tümmler H[27], Ullm C[28], Walke M[29], Weigel R[30], Wertman M[9,31], Wiehle R[32], Wiezorek T[6], Wilke L[4], Wolf U[33], Eich HT[1], Schmitt D[10,34]

[1]University Hospital Münster, Department of Radiation Oncology, Münster, Germany
[2]University Medical Center Schleswig Holstein, Department of Radiation Oncology, Kiel, Germany
[3]University Medicine Mannheim, Department of Radiation Oncology, Medical Faculty Mannheim, University Heidelberg, Mannheim, Germany
[4]University Hospital Zürich, Department of Radiation Oncology, Zürich, Switzerland
[5]University Hospital Rechts der Isar, Department of Radiation Oncology, Technical University, Munich, Germany
[6]University Hospital Jena, Department of Radiation Oncology, Jena, Germany
[7]University Medical Center Hamburg-Eppendorf, Department of Radiotherapy and Radiation Oncology, Hamburg, Germany
[8]West German Proton Therapy Center, Essen, Germany
[9]University Medicine Rostock, Department of Radiation Oncology, Rostock, Germany
[10]Heidelberg University Hospital, Department of Radiation Oncology, Heidelberg, Germany
[11]Division of Medical Radiation Physics and Department of Radiation Oncology, Inselspital, Bern University Hospital, and University of Bern, Switzerland
[12]Schwarzwald-Baar Hospital, Department of Radiation Oncology, Villingen-Schwenningen, Germany
[13]University Hospital Cologne, Department of Radiation Oncology, Cologne, Germany
[14]Gamma Knife Center Krefeld, Krefeld, Germany
[15]Hospital Bayreuth, Department of Radiation Oncology, Bayreuth, Germany
[16]Lung Clinic Hemer, Department of Radiation Oncology, Hemer, Germany
[17]University Hospital Munich, Department of Radiation Oncology, LMU Munich, Munich, Germany
[18]Medical University Hannover, Department of Radiation Oncology, Hannover, Germany
[19]Private Practice for Radiation Oncology, Distler Bautzen, Germany
[20]University Hospital Frankfurt, Department of Radiation Oncology, Frankfurt, Germany
[21]Saphir Radiosurgery Center, Frankfurt and Northern Germany, Kiel, Germany
[22]University Hospital Halle, Department of Radiation Oncology, Halle (Saale), Germany
[23]University Hospital Würzburg, Department of Radiation Oncology, Würzburg, Germany
[24]Private Practice for Radiation Oncology South, Kaufbeuren, Germany
[25]University Hospital Mainz, Department of Radiation Oncology, Mainz, Germany
[26]Private Practice for Radiation Oncology, Böblingen, Germany
[27]Private Practice for Radiation Oncology, Dresden, Germany
[28]Radiologie Muenchen, Private Practice for Radiation Oncology, Munich, Germany
[29]University Hospital Magdeburg, Department of Radiation Oncology, Magdeburg, Germany
[30]University Hospital Innsbruck, Department of Radiation Oncology, Innsbruck, Austria
[31]Ernst von Bergmann Hospital, Department of Radiation Oncology, Potsdam, Germany
[32]University Hospital Freiburg, Department of Radiation Oncology, Freiburg, Germany
[33]University Hospital Leipzig, Department of Radiation Oncology, Leipzig, Germany
[34]University Medical Center Göttingen, Department of Radiation Oncology, Göttingen, Germany

[§]Corresponding Author:

Dr. Christos Moustakis, christos.moustakis@ukmuenster.de
Universitätsklinikum Münster, Klinik für Strahlentherapie
Albert-Schweitzer-Campus 1, Gebäude A1, D-48149 Münster

Disclosures: None

Funding: None

Acknowledgments

We would like to thank the independent experts Dr. Marco Esposito[a] and Dr. Victor Hernandez[b] for critically reading the work, discussing and validating the RATING score.

[a] Medical Physics Unit, AUSL Toscana Centro, Florence, Italy
[b] Department of Medical Physics, Hospital Universitari Sant Joan de Reus, Tarragona, Spain

ABSTRACT

Purpose

To investigate, if liver SBRT treatment planning can be harmonized across different treatment planning systems, delivery techniques and institutions by using a specific prescription method and to minimize the knowledge gap concerning inter-system and inter-user differences. To provide best practice guidelines for all used techniques.

Methods

A multiparametric specification of target dose ($GTV_{D50\%}$, $GTV_{D0.1cc}$, $GTV_{V90\%}$, $PTV_{V70\%}$) with a prescription dose of $GTV_{D50\%}$ = 3 x 20 Gy and OAR limits were distributed with CTs and structure sets from three liver metastases patients. Thirty-five institutions provided 132 treatment plans using different irradiation techniques. These plans were first analyzed for target and OAR doses. Four different renormalization methods were performed (PTV_{Dmin}, $PTV_{D98\%}$, $PTV_{D2\%}$, PTV_{Dmax}). The resulting 660 treatments plans were evaluated regarding target doses in order to study the effect of dose renormalization to different prescription methods. A relative scoring system was used for comparisons.

Results

$GTV_{D50\%}$ prescription can be performed in all systems. Treatment plan harmonization was overall successful with standard deviations for D_{max}, $PTV_{D98\%}$, $GTV_{D98\%}$ and PTV_{Dmean} of 1.6 Gy, 3.3 Gy, 1.9 Gy and 1.5 Gy, respectively. Primary analysis showed 55 major deviations from clinical goals in 132 plans, while in only <20% of deviations GTV/PTV dose was traded for meeting OAR limits. $GTV_{D50\%}$ prescription produced the smallest deviation from target planning objectives and between techniques, followed by the PTV_{Dmax}, $PTV_{D98\%}$, $PTV_{D2\%}$ and PTV_{Dmin} prescription. Deviations were significant for all combinations but for the PTV_{Dmax} prescription compared with $GTV_{D50\%}$ and $PTV_{D98\%}$. Based on the various dose prescription methods, all systems significantly differed from each other, while $GTV_{D50\%}$ and $PTV_{D98\%}$ prescription showed the least differences between the systems.

Conclusions

This study showed the feasibility of harmonizing liver SBRT treatment plans across different treatment planning systems and delivery techniques when a sufficient set of clinical goals is given.

INTRODUCTION

Stereotactic body radiotherapy (SBRT) is defined as highly precise hypo-fractionated radiotherapy [1] and is well-established for many extracranial tumor manifestations like primary and secondary lung and liver malignancies [2-7] and spinal metastases [8], among others. Due to knowledge and technology gains in the recent years [10], the treatment of liver metastases with SBRT found its way into clinical routine, especially for those originating from colorectal, breast and lung cancer [7, 10, 11]. However, currently, SBRT for liver metastases is not often performed and treatment planning as well as techniques may vary widely. [7, 12]. To harmonize liver SBRT practice, also in the context of multi-center multi-platform clinical trial preparations, several investigations on minimally needed biological effective dose (BED) [7, 15-17] and technology comparison on phantom and in-vivo dose delivery accuracy [16] have been published. The aim of the study was to fill the knowledge gap concerning the inter-system and inter-user differences for treatment techniques and treatment planning systems.

For SBRT in general, machine/technology and planner experience variability have been investigated in the past [17-20]. An overview of planning benchmark studies has also recently been published [21]. The results from those studies show that treatment plan quality only slightly depends on machine/technology whilst the highest variability seems to originate from planner experience [18, 22]. One way to overcome those differences can be addressed through benchmark studies and crowd knowledge-based learning [19, 22-24]. Another way is the strict specification of planning goals and dose prescription like a recent study on lung SBRT demonstrated [25]. In the lung, tissue heterogeneities are mostly responsible for discrepancies in tumor dose distribution between different patients using common circumferential planning target volume (PTV) prescription methods [26, 27]. However, there is a lack of benchmark planning studies for larger tumors in more homogeneous tissue closer to critical organs at risk (OAR), like liver metastases.

In this joint planning benchmark study from the XXX working group radiosurgery and stereotactic radiotherapy and the XXX working group for physics and technology in stereotactic radiotherapy we assessed the harmonization of treatment plan quality for liver metastases SBRT in a multi-institutional multi-platform context on an international level. Besides previous knowledge gains from the harmonization of SBRT treatment planning for

lung tumors through gross/clinical/internal tumor/target volume (GTV/CTV/ITV) mean dose prescription [25] we implemented stricter planning goals for GTV and PTV dose coverage requirements based on recent findings on various dose-response relationships for liver metastases [11, 13, 28]. The aims of the present study were: (i) harmonization of liver SBRT treatment plan quality for large lesions, (ii) investigation on the possibility to use GTV median dose prescription in the liver and (iii) validation that the As Low As Reasonably Achievable (ALARA) principle is better implemented after our initial planning benchmarks [17, 18].

MATERIALS AND METHODS

Case Selection and Patient Characteristics

After approval from the leading ethics committee of the medical faculty of the University of XXXX (reference number D 457/18), the SBRT liver databases of the lead institutions were screened for three suitable cases for this treatment planning benchmark study. The number of cases originated from previous studies and was found to be a balance between statistical power in analysis and workload of the participants [17, 18]. For the search we defined the primary characteristics of the cases in accordance with prior pattern-of-care studies [7, 11] to be (1) each a different primary histology commonly treated with SBRT with follow-up ≥ 1 year, to be (2) of varying tumor volume, but larger than a GTV of 25 cc and to include (3) at least one case with (a) central location and/or (b) smaller whole liver volume < 1200 cc and/or (c) close location to gastrointestinal organs (approx. 1.5 cm based on [12]). From the databases twelve cases were preselected based on the aforementioned characteristics from which finally three cases were selected for this benchmark based on study committee consensus decision.

Case 1 (criterion a) was a 52-year-old patient with liver metastases from rectal carcinoma. After several systemic and local treatments one oligo-persistent liver metastasis remained. It was located in segment IV close to the portal vein and hence unresectable and was treated with SBRT (GTV = 52 cc, whole liver = 2350 cc).

Case 2 (criterion b) was an 83-year-old patient with breast carcinoma. Liver metastases were first discovered 3 years after initial diagnosis and several asynchronous oligo-metastases were resected or ablated over the course of twelve years. After recurrence of yet another two adjacent oligo-metastases in segment IV/VIII the treatment was continued with SBRT due to the small remaining liver volume (GTV for both metastases = 69 cc, one merged CTV, whole liver = 1134 cc).

Case 3 (criterion c) was a 63-year-old patient with early stage non-small-cell lung cancer which was initially treated with SBRT due to multiple co-morbidities. Subsequently, an

inoperable oligo-metastasis in liver segment V with close proximity to the intestine (1.5 cm) was treated with SBRT (GTV = 25 cc, whole liver = 1600 cc).

Treatment planning image data sets and contours were initially obtained from the treating institution after full case data anonymization. The planning CT was acquired head first supine with vacuum bag at end expiration breath hold with 1.0 mm in-plane resolution and 1.5 mm slice thickness and the planning MRI was performed with the same resolution using a standard liver SBRT protocol as described previously [28]. For the purpose of this planning benchmark study the OAR contours were harmonized for all three cases and adapted/added where necessary according to XXX and international guidelines [12, 19], based on study committee consensus decision. The original GTV and CTV contours, for these three cases CTV = GTV + 5 mm isotropic in-liver expansion [30], were not modified and an isotropic margin of 3 mm was added to the CTV to generate the PTV assuming an active motion compensation technique during treatment [9, 16]. The final PTVs for this study were 126 cc (case 1), 152 cc (case 2) and 73 cc (case 3) as illustrated in Figure 1.

Treatment techniques and treatment planning systems

The anonymized treatment planning CT images and the respective radiotherapy structure sets (RTSS) in Digital Imaging and Communications in Medicine (DICOM) standard format were distributed to all institutions participating in this trial. Treatment techniques and treatment planning systems [9] was performed in each institution´s treatment planning system (TPS) using institutional-specific methods/techniques and society guidelines [12, 18]. The use of a type B [26] dose calculation algorithm was recommended. All submitted treatment plans had to be clinically acceptable concerning OAR limits, judged by the residing radiation oncologist responsible for stereotactic radiotherapy. Plans were supposed to meet pre-defined dose prescription and clinical goals. If a goal was not met, a deviation was documented. To characterize the extent of the deviation, we pre-defined a set of thresholds to differentiate between “minor” and “major” deviations. In the following, the goals for “no deviation” and the thresholds for “minor deviation” (in brackets) are presented. “Major deviations” meant that thresholds for “minor deviations” were not met.

1) Based on previous studies [7, 13, 14, 18, 20, 25], the prescription dose for all three cases was defined as median GTV dose ($GTV_{D50\%}$) = 3 x 20 Gy at the 100% isodose line ($BED_{\alpha/\beta=10Gy}$ = 180 Gy_{10}). Further target planning objectives were (a) GTV near maximum (i.e. $GTV_{D0.1cc}$) ≤ 107% = 64.2 Gy (≤ 110% = 66 Gy) [26, 27], (b) GTV coverage at 54 Gy (i.e. $GTV_{V90\%}$) [28] ≥ 98% (≥ 95%) and (c) PTV coverage at 42 Gy (i.e. $PTV_{V70\%}$) ≥ 98% (≥ 95%) [7, 13].

2) Based on commonly available OAR limits for liver SBRT in three fractions [12, 29, 30] the

major limitations for this study were (a) healthy liver minus $Liver_{V15Gy}$ ($Liver_{V17Gy}$) ≥ 700 cc, (b) gastrointestinal organs D_{max} (D_{1cc}) ≤ 24 Gy, (c) heart D_{max} (D_{1cc}) ≤ 30 Gy and (d) esophagus/stomach D_{max} (D_{1cc}) ≤ 21 Gy. Further limitations are presented in the Appendix 1 Table A.1 and all OARs were to be handled based on the As Low As Reasonable Achievable (ALARA) concept.

Dosimetric Evaluation

The resulting dose distributions were collected from all participating institutions in DICOM format and imported into a common TPS (Eclipse, version 15.6, Varian Medical Systems, Palo Alto, USA) for combined evaluation. A detailed questionnaire for each case (see Appendix 2 Section II) was also completed by the participants. The dosimetric evaluation was based on the International Commission on Radiation Units and Measurements (ICRU) report 91 on prescribing, recording, and reporting of stereotactic treatments with small photon beams [26].

Primary Dosimetric Evaluation

Since each TPS will have minor differences in contour interpretation, not all submitted dose distributions may perfectly fulfill the dose prescription requirement of 60 Gy to the 100% isodose line after the import in Eclipse. Hence, we first performed a minimal dose correction in Eclipse by re-normalization of the $GTV_{D50\%}$ to 60 Gy for all treatment plans.

Renormalization to different prescription methods

In order to further assess different dose prescription methods, we also re-normalized the dose to (a) $PTV_{D98\%}$ (near minimum) = 42 Gy, (b) PTV_{Dmin} (absolute minimum) = 42 Gy, (c) $PTV_{D2\%}$ (near maximum) = 64.2 Gy and (d) PTV_{Dmax} (absolute maximum) = 64.2 Gy resulting in overall five dose distributions for every submitted treatment plan. These are all common methods, while prescribing to a single voxel (minimum and maximum prescription) is in principle less robust than a volume based prescription.

Target Volume Doses and Indices Evaluation

For each prescription method we evaluated GTV and PTV D_{min}, $D_{98\%}$, $D_{50\%}$, $D_{2\%}$ and D_{max} and GTV_{V54Gy} and PTV_{V42Gy}. Additionally, commonly used plan quality indices were evaluated based on previous studies and ICRU report 91 recommendations [17, 18, 26, 31, 33]:

1a) Homogeneity Index PTV (HI_{PTV}) = ($PTV_{D2\%}$ - $PTV_{D98\%}$) / $PTV_{D50\%}$
1b) Homogeneity Index GTV (HI_{GTV}) = ($GTV_{D2\%}$ - $GTV_{D98\%}$) / $GTV_{D50\%}$

2a) ICRU 91 conformity index (CI) = (V_{PTV} * V_{42Gy}) / PTV_{V42Gy}^2

2b) External conformity index (CΔ) = (V_{42Gy} - PTV_{V42Gy}) / V_{PTV}

3a) Gradient Index PTV (GI_{PTV}) = V_{21Gy} / V_{42Gy}
3b) Gradient Index GTV (GI_{GTV}) = V_{30Gy} / V_{60Gy}

where PTV_{VxGy} represents the planning target volume part and V_{XGy} denotes the total tissue volume which received at least X Gy, respectively. For delivery efficiency comparison, the monitor units (MU) and the estimated irradiation times were collected via the aforementioned questionnaire (Appendix 2 Section II).

Organs at Risk Evaluation

Since the OAR limits will likely be violated when re-normalizing based on different dose prescription methods than originally planned on, we only evaluated the OAR dosimetry based on the re-normalized treatment plans of 60 Gy to the $GTV_{D50\%}$ in Eclipse. Based again on the suggestions of the ICRU report 91 we extracted D_{mean}, $D_{2\%}$ and D_{max} and a set of appropriate volume doses (V_{XGy}) and (D_{Ycc}) derived from guidelines and international protocols, as described above, and partly provided in Table 1.

Plan Quality Ranking

In order to compare the treatment plans qualitatively, independent of specific case properties and combined for various technology, we applied the previously published and well-established relative plan ranking method [17, 18, 21, 33]. In brief, with this ranking method the analyzed data is split into four categories (1 = excellent, 2 = above average, 3 = below average, 4 = poor) using the normal distribution (bell curve) of the best and the worst results for that data (first order ranking). For larger amounts of parameters in the data, each parameter was ranked separately. These separate ranks (1-4) were then summed for each plan and this sum was ranked again using the same normal distribution ranking method (second order ranking). The following data was ranked using these methods:

1) For the analysis of overall plan quality for the primary dose prescription ($GTV_{D50\%}$ = 60 Gy) we included all relevant and non-redundant plan parameters into the second order ranking method (i.e. PTV_{V42Gy}, GTV_{V54Gy}, GTV_{D1cc}, CI, CΔ, GI_{PTV}, Intestine/Duodenum D_{1cc}/V_{15Gy}, Esophagus/Stomach/Spinal-Cord/Aorta/Inferior-Vena-Cava/Skin D_{1cc} and Heart D_{1cc}/V_{24Gy}). We also created sub-scores for only the OAR (calculated like the overall plan quality rank, but without the target parameters) for all three cases combined and for only the OAR close to the PTV (the same as before without the non-close OAR) for each case separately then combined into final sub-score over the three cases. Here we also analyzed (a) individual planner and (b) technology.

2) For the analysis of deviations from GTV/PTV objectives and OAR limits for the primary dose prescription ($GTV_{D50\%}$ = 60 Gy) we gave penalty points for each minor (1 penalty point) or major (2 penalty points) deviation. The penalty points were summed for each plan and the first order ranking method was used to rank the plans based on the penalty sum (deviation score). Again, we analyzed (a) individual planner and (b) technology.

Since each of the four presented plan scores (i.e. overall plan score, all OAR score, PTV close OAR score and deviation score) may not fully reflect a specific quality of a treatment plan we decided to also evaluate the average over all four scores and defined this average as final benchmark score for each participant.

Dose Prescription Evaluation

For the analysis of the renormalization to different prescription methods, we calculated the absolute deviation from the ideal GTV/PTV dose objectives as described in the previous section "Treatment techniques and treatment planning systems". For each plan, the sum of these absolute deviations in Gy was calculated and used to score each plan with each of the five prescription methods and to compare (a) prescription methods and (b) technology over all three cases.

Statistical Analysis

Statistical differences and variances of the dose metrics and ranking scores between different planning techniques were analyzed by Kruskall-Wallis test and Levene's test, respectively. When statistical differences were found with p value < 0.05, further post-hoc pair-wise comparisons between planning techniques were performed applying Bonferroni's correction. All statistical analyses were conducted using R software version 3.5.1 (The R Foundation, Vienna, Austria) and were considered significant if p value was < 0.05.

RESULTS

Thirty-five institutions with experience in liver-SBRT participated in this study. Examples for dose distributions for case 1 can be found in Figure 2.

Treatment techniques and treatment planning systems

For all three cases combined, 132 treatment plans (44 for case 1, 45 for case 2 and 43 for case 3) were generated using multiple techniques within various treatment planning systems. Notably, some institutions provide more than one plan per case. For case 1, 1 institution provided 3 plans, 7 institutions provided 2 plans. For case 2, 1 institution provided 5 plans, 6

institutions provided 2 plans. For case 3, 1 institution provided 3 plans, 6 institutions provided 2 plans. Sixty-eight plans (52%) were generated with Intensity Modulated Arc Therapy (IMAT), 11 plans (8%) with Static Field Intensity Modulated Radiation Therapy (SF-IMRT), 16 (12%) with Three Dimensional Conformal Radiation Therapy (3D-CRT), 16 plans (12%) with Robotic Radiosurgery (RRS), 12 plans (9%) with Helical Radiotherapy (HRT) and 9 plans (7%) with Proton Therapy (PT) techniques. Of all plans, 53%, 23% and 14% used a photon spectrum with a maximum energy of 6, 10 and 15 MeV, of those 27, 6 and 18 plans used a flattening filter (FF) and 48, 24 and 0 plans were flattening filter free (FFF), respectively.

Dose calculation algorithms were mostly Collapsed Cone (CC, 29%), Monte Carlo (MC, 25%) and Anisotropic Analytical Algorithms (AAA, Varian Medical Systems, Palo Alto, USA, 25%), while Superposition/Convolution was used in 9% of the plans and Acuros (Varian Medical Systems, Palo Alto, USA) in 5%, respectively. Pencil Beam (PB) was used for the proton plans (7%). The dose calculation grid was smaller than or equal to 2 mm for all plans and most used the planning CT resolution (1 x 1 x 1.5 mm^3). All plans were clinically accepted in each institution in terms of dose to critical organs before submission. Import of the submitted dose files was not straightforward in all cases and some files had to be modified to be imported into the common planning system used for this study (Eclipse).

Primary Dosimetric Evaluation

Initial Dose Re-Normalization

The differences between the submitted plan doses from the questionnaire and the Eclipse doses for the mandatory prescription of 3 x 20 Gy to the GTV $D_{50\%}$ were small. For Case 1, 2 and 3 we found differences of 0.1 ± 0.2 Gy (range -0.6-1.3%), 0.0 ± 0.2 Gy (range -1.4-1.0%) and 0.0 ± 0.1 Gy (range -0.5-1.1%), respectively. This may point to very similar volume and dose interpretation of the varying treatment planning systems. Hence, the dose re-normalization in Eclipse of 3 x 20 Gy to the GTV $D_{50\%}$ for all cases was regarded as dosimetrically negligible.

Target Volume Doses and Indices

Despite strict requirements in prescription and several other parameters, the dose and homogeneity inside the GTV and PTV varied between individual planners, even between similar technologies. However, the inter-planner inter-system treatment plan harmonization in this study was successful overall (see DVHs for all cases in Figure 3), with mean ± standard deviation (SD) plan D_{max}, $PTV_{D98\%}$, $GTV_{D98\%}$ and PTV_{mean} of 63.9 ± 1.6 Gy (goal ≤ 64.2 Gy), 42.9 ± 3.3 Gy (goal ≥ 42.0 Gy), 55.8 ± 1.9 Gy (goal ≥ 54.0 Gy) and 54.4 ± 1.5 Gy (no goal given), respectively. The resulting HI_{PTV} and HI_{GTV} were 0.35 ± 0.07 and 0.11 ± 0.04, and the CI and CΔ were 1.19 ± 0.18 and 0.15 ± 0.19, respectively. On the other hand, the dose

gradients outside the PTV varied to a much greater degree with GI_{PTV} of 3.54 ± 0.72 (see Table 1).

Organs at Risk Doses

The closest organs at risk were the esophagus (case 1), heart (case 2) and duodenum (case 3). Their corresponding absolute maximum doses were 18.8 ± 3.5 Gy (goal ≤ 21 Gy), 35.3 ± 5.3 Gy (goal ≤ 30.0 Gy) and 24.0 ± 4.3 Gy (goal ≤ 24 Gy) for the esophagus, heart and duodenum, respectively (see Table 1). The V_{15Gy} of the liver was 694.5 ± 149.5 cc (case 1, goal ≤ 1650 cc), 489.5 ± 74.7 cc (case 2, goal ≤ 434 cc) and 244.3 ± 28.0 cc (case 3, goal ≤ 900 cc). Further details are presented in Figure 4.

Deviations from clinical goals

Overall, only 20 treatment plans (18 for case 1 using various techniques and 1 PT plan each for case 1 and 3) had no deviations from the clinical goals. Out of the total 132 we found 55 treatment plans with at least one major deviation from clinical goals. Concerning specific deviations from GTV and PTV objectives we found 2/0 and 6/0 minor/major (case 1), 4/3 and 9/8 minor/major (case 2) and 6/2 and 12/3 minor/major (case 3) deviations. For the OAR dose, we found 33/3 minor/major (case 1, major = stomach and skin D_{1cc}), 73/39 minor/major (case 2, major mainly = liver V_{17Gy} and heart D_{1cc}) and 80/49 minor/major (case 3, major mainly = duodenum $V_{15/18Gy}$ and skin D_{1cc}) deviations. Based on a case-by-case evaluation, we found that only 23.1%, 18.2% and 7.1% of the plans with these deviations traded GTV/PTV dose coverage for meeting the OAR limits , which was the individual decision of the residing radiation oncologist.

Best Practice Guidelines

Based on the individual relative plan scoring and ranking system used for this study (averaged sub-scores for combined and selective plan metrics and deviations from clinical goals as described in materials and methods), we selected five individual planners with the best scores for IMAT, SF-IMRT, RRS, HT and PT to present their best practice approach for liver SBRT (Appendix 3). The ALARA concept was in general followed in this study; however, outliers from this concept were also noted for all three cases (Figure 4).

Treatment techniques

Proton Therapy significantly outperformed all other planning techniques showing the best averaged and selective sub-scores for overall plan quality, dosimetry of close OAR and all OAR and summed deviations from clinical goals (Table 2, $p < 0.001$). Excluding proton therapy, robotic radiosurgery significantly outperformed the other techniques in terms of sub-score for close OARs ($p < 0.01$ for paired tests, Table 2). However, regarding all OARs

combined, proton therapy outperformed only helical radiotherapy ($p < 0.01$, Table 2). Robotic radiosurgery showed the best dose conformity over all systems and the highest GTV and PTV homogeneity index (meaning the least homogeneous dose). Further details for the primary evaluation are presented in Tables 2 and 3.

Treatment delivery times per fraction (without patient setup), were lowest for FFF IMAT (mean 2.0 min, range 1.7-2.8 min), followed by 3D-CRT (mean 4.4 min, range 3.0-7.0 min), IMAT (all energies, mean 6.3 min, range 1.7 - 15.0 min), SF-IMRT (mean 6.8 min, range 5.0-7.0 min) and PT (mean 9.6 min, range 7.7-10.0 min), tailed by HRT (mean 36.0 min, range 26.6-60.9) and RRS (mean 59.0 min, range 26.0-91.0 min).

Renormalization to different prescription methods

Five different renormalization methods were performed as described in materials and methods (PTV D_{min}, $D_{98\%}$, $D_{2\%}$, D_{max}, GTV $D_{50\%}$) resulting in a total of 660 treatments plans which were evaluated based on GTV and PTV metrics in order to study the effect of dose renormalization to different prescription methods. In terms of absolute deviation from the planning objectives for GTV and PTV we found that the GTV $D_{50\%}$ prescription had the smallest mean differences across all parameters (5.5 ± 3.9 Gy) followed by the PTV D_{max}, $D_{98\%}$ and $D_{2\%}$ prescription (6.5 ± 4.2 Gy, 8.9 ± 9.6 Gy, and 10.7 ± 6.9 Gy) and, finally, the PTV D_{min} prescription (71.9 ± 101.3 Gy). DVH graphs for all cases are presented in Figure 3. Using the adjusted Kruskal Wallis test, the pairwise comparisons for the deviations were significant ($p < 0.001$) for all combinations but for the PTV D_{max} prescription compared with GTV $D_{50\%}$ ($p = 0.36$) and PTV $D_{98\%}$ ($p = 0.74$) prescriptions while the GTV $D_{50\%}$ and PTV $D_{98\%}$ had again significant differences in deviations ($p = 0.02$).

In terms of comparing different systems with each other for the deviations to the GTV and PTV planning objectives and based on the various dose prescription methods we found that all systems significantly differed from each other ($p < 0.04$ for all systems and prescription, Table 3). That did not change when the group ranking method was applied to the absolute deviations. Aside from the PTV D_{min} prescription, which had the largest deviations for all systems, we found that the GTV $D_{50\%}$ and the PTV $D_{98\%}$ prescription showed the least significant differences between the systems in the pair-wise comparisons (see Table 3). Here, only PT showed significant differences in comparison with IMAT ($p = 0.010$-0.012) and RRS ($p = 0.003$-0.024). For the group ranking evaluation of the absolute deviation PT kept significant differences with IMAT ($p = 0.014$) for the PTV $D_{98\%}$ prescription while for the GTV $D_{50\%}$ prescription PT was significantly different to IMAT ($p = 0.002$), SF-IMRT ($p = 0.029$) and RRS ($p = 0.003$)

RATING Score

Recently, radiotherapy treatment planning study guidelines (RATING) were published along with a scoring metric for study quality assessment [34]. Based on self-assessment of our study and the evaluation of two independent experts we achieved a mean RATING score of 98% (RATING fraction 201 out of 205 points, Appendix 4 Table A.2), which was validated during the authors' review process of the manuscript.

DISCUSSION

To our knowledge, with 132 submitted plans from 35 institutions, this is one of the largest treatment planning studies to date, demonstrating the large interest in inter-institutional collaboration and exchange of information. The present study demonstrates the possibility to generate very similar SBRT treatment plans with various treatment planning systems for a variety of treatment delivery systems including all common techniques. The multiparametric specification of target dose (GTV $D_{50\%}$, GTV $D_{0.1cc}$, $GTV_{V90\%}$, $PTV_{V70\%}$) leads to more harmonized plans than in our previous NSCLC planning study [18] with the objective of a PTV encompassing dose and the prescription isodose line. More details are presented in Appendix 5.

Prospective and retrospective clinical studies can suffer from a large variability of target doses, even if the dose prescription is nominally the same for all patients, especially if different planning and delivery techniques are used in a multi-center setting [24, 35]. This is due to unspecific or a too limited set of planning objectives and may lead to inconsistent dose-to-outcome correlations, if not at least a sufficent set of dose parameters is reported for analysis. Because of that, we recommend multiparametric target dose objectives for prospective clinical trials, which can lead to a harmonized patient plan collective, as demonstrated in this study. Another important aspect for clinical studies is the recommendation or at least documentation of the dose calculation algorithm type. Here type B algorithms were proposed, which was fulfilled from all participants (only AAA is an intermediate algorithm, according to [26]), except the proton facilities, where the Pencil Beam algorithm is the clinical standard. Because of that and the small involvement of tissue inhomogeneities in the liver cases, we do not expect an influence of the dose calculation algorithm on our results.

The OAR limits were not met in all plans and for the three cases only 23.1%, 18.2% and 7.1% of the plans with deviations traded GTV/PTV dose coverage for meeting the OAR limits. Because of the recommendation to submit only clinically acceptable plans this may reflect different clinical practice between the institutions. The “best practice” guideline in the

supplement might be helpful in balancing the target and OAR goals given the differences in approach we saw.

The ranking of treatment techniques must be seen under the limitation of the assumption of an active motion management during the beam delivery using no ITV and a CTV-to-PTV margin of only 3 mm. The accuracy of different treatment delivery and motion monitoring techniques was not considered. This would only be possible through individual margin definition, which counteracts our method of plan evaluation. In particular, the fact that proton therapy plans outperformed all photon techniques in all cases regarding target and OAR goals does not necessarily mean a better treatment in liver SBRT. Active motion management techniques in proton therapy are still under investigation [36, 37].

In the current investigation the prescription to the GTV $D_{50\%}$ showed the smallest amount of deviations from the planning objectives and, even more importantly, between different delivery techniques. It is the main limitation of the study that only for this single dose prescription the optimization was done and thus the result is likely biased, especially for the PTV D_{min} prescription. To minimize the bias, all other dose prescriptions used for re-normalization of all plans were part of the set of dose objectives, hence a plan was not necessarily changed through re-normalization, if all objectives are met. The alternative would have included obtaining different optimized and prescribed plans for all cases from the planners. However, this would have resulted in 3x5 plans per planner and we considered the workload involved too excessive for study participants in this scenario.

CONCLUSIONS

This study shows the feasibility of harmonizing liver SBRT treatment plans across different treatment planning systems and delivery techniques, when a sufficient set of clinical goals is given. The method of GTV $D_{50\%}$ prescription can be performed in all systems, improving overall consistency. The ALARA principle was followed for most OARs, but in many plans dose limits in OARs close to the target were exceeded to meet the target dose. Besides the comparison between different treatment techniques and platforms, advice for planning strategies is provided in the appendix.

REFERENCES

[1] XXX

[2] Klement RJ, Sonke JJ, Allgäuer M, et al. Correlating Dose Variables with Local Tumor Control in Stereotactic Body Radiation Therapy for Early-Stage Non-Small Cell Lung Cancer: A Modeling Study on 1500 Individual Treatments. Int J Radiat Oncol Biol Phys. 2020;107(3):579-586. https://doi.org/10.1016/j.ijrobp.2020.03.005

[3] Mazzola R, Ruggieri R, Figlia V, et al. Stereotactic body radiotherapy of central lung malignancies using a simultaneous integrated protection approach: A prospective observational study. Strahlenther Onkol. 2019;195(8):719-724. https://doi.org/10.1007/s00066-018-01419-0

[4] Hörner-Rieber J, Bernhardt D, Blanck O, et al. Long-term Follow-up and Patterns of Recurrence of Patients With Oligometastatic NSCLC Treated With Pulmonary SBRT. Clin Lung Cancer. 2019;20(6):e667-e677. https://doi.org/10.1016/j.cllc.2019.06.024

[5] Chiang CL, Chan MKH, Yeung CSY, et al. Combined stereotactic body radiotherapy and trans-arterial chemoembolization as initial treatment in BCLC stage B-C hepatocellular carcinoma. Strahlenther Onkol. 2019;195(3):254-264. https://doi.org/10.1007/s00066-018-1391-2

[6] Gkika E, Strouthos I, Kirste S, et al. Repeated SBRT for in- and out-of-field recurrences in the liver. Strahlenther Onkol. 2019;195(3):246-253. https://doi.org/10.1007/s00066-018-1385-0

[7] XXXX

[8] Guckenberger M, Sweeney RA, Hawkins M, et al. Dose-intensified hypofractionated stereotactic body radiation therapy for painful spinal metastases: Results of a phase 2 study. Cancer. 2018;124(9):2001-2009. https://doi.org/10.1002/cncr.31294

[9] XXXX

[10] Méndez Romero A, Schillemans W, van Os R et al. The Dutch-Belgian Registry of Stereotactic Body Radiation Therapy for Liver Metastases: Clinical Outcomes of 515 Patients and 668 Metastases. Int J Radiat Oncol Biol Phys. 2020 Dec 29:S0360-3016(20)34567-3. https://doi.org/10.1016/j.ijrobp.2020.11.045.

[11] Mahadevan A, Blanck O, Lanciano R, et al. Stereotactic Body Radiotherapy (SBRT) for liver metastasis - clinical outcomes from the international multi-institutional RSSearch® Patient Registry. Radiat Oncol. 2018;13(1):26. https://doi.org/10.1186/s13014-018-0969-2

[12] XXX

[13] Klement RJ, Guckenberger M, Alheid H, et al. Stereotactic body radiotherapy for oligo-metastatic liver disease - Influence of pre-treatment chemotherapy and histology on local tumor control. Radiother Oncol. 2017;123(2):227-233. https://doi.org/10.1016/j.radonc.2017.01.013

[14] Klement RJ, Abbasi-Senger N, Adebahr S, et al. The impact of local control on overall survival after stereotactic body radiotherapy for liver and lung metastases from colorectal cancer: a combined analysis of 388 patients with 500 metastases. BMC Cancer. 2019;19(1):173. https://doi.org/10.1186/s12885-019-5362-5

[15] Brunner TB, Blanck O, Lewitzki V, et al. Stereotactic body radiotherapy dose and its impact on local control and overall survival of patients for locally advanced intrahepatic and extrahepatic cholangiocarcinoma. Radiother Oncol. 2019;132:42-47. https://doi.org/10.1016/j.radonc.2018.11.015

[16] Boda-Heggemann J, Jahnke A, Chan MKH, et al. In-vivo treatment accuracy analysis of active motion-compensated liver SBRT through registration of plan dose to post-therapeutic MRI-morphologic alterations. Radiother Oncol. 2019;134:158-165. https://doi.org/10.1016/j.radonc.2019.01.023

[17] XXXX

[18] XXXX

[19] Habraken SJM, Sharfo AWM, Buijsen J, et al. The TRENDY multi-center randomized trial on hepatocellular carcinoma - Trial QA including automated treatment planning and benchmark-case results. Radiother Oncol. 2017;125(3):507-513. https://doi.org/10.1016/j.radonc.2017.09.007

[20] Esposito M, Maggi G, Marino C, et al. Multicentre treatment planning inter-comparison in a national context: The liver stereotactic ablative radiotherapy case. Phys Med. 2016;32(1):277-283. https://doi.org/10.1016/j.ejmp.2015.09.009

[21] Giglioli FR, Garibaldi C, Blanck O, et al. Dosimetric Multicenter Planning Comparison Studies for Stereotactic Body Radiation Therapy: Methodology and Future Perspectives. Int J Radiat Oncol Biol Phys. 2020;106(2):403-412. https://doi.org/10.1016/j.ijrobp.2019.10.041

[22] Villaggi E, Hernandez V, Fusella M, et al. Plan quality improvement by DVH sharing and planner's experience: Results of a SBRT multicentric planning study on prostate. Phys Med. 2019;62:73-82. https://doi.org/10.1016/j.ejmp.2019.05.003

[23] Esposito M, Masi L, Zani M, et al. SBRT planning for spinal metastasis: indications from

a large multicentric study. Strahlenther Onkol. 2019;195(3):226-235. https://doi.org/10.1007/s00066-018-1383-2

[24] Giglioli FR, Strigari L, Ragona R, et al. Lung stereotactic ablative body radiotherapy: A large scale multi-institutional planning comparison for interpreting results of multi-institutional studies. Phys Med. 2016;32(4):600-606. https://doi.org/10.1016/j.ejmp.2016.03.015

[25] XXXX

[26] Seuntjens J, Lartigau EF, Cora S, et al. ICRU report 91. Prescribing, recording, and reporting of stereotactic treatments with small photon beams. J ICRU. 2014. 14(2):1–160. https://doi.org/10.1093/jicru/ndx017

[27] XXXX

[28] Stera S, Balermpas P, Chan MKH, et al. Breathing-motion-compensated robotic guided stereotactic body radiation therapy: Patterns of failure analysis. Strahlenther Onkol. 2018;194(2):143-155. https://doi.org/10.1007/s00066-017-1204-z

[29] XXXX

[30] Grimm J, LaCouture T, Croce R, et al. Dose tolerance limits and dose volume histogram evaluation for stereotactic body radiotherapy. J Appl Clin Med Phys. 2011;12(2):3368. https://doi.org/10.1120/jacmp.v12i2.3368

[31] Paddick I, Lippitz B. A simple dose gradient measurement tool to complement the conformity index. J Neurosurg. 2006;105(Supplement):194-201. https://doi.org/10.3171/sup.2006.105.7.194

[32] Wagner TH, Bova FJ, Friedman WA, et al. A simple and reliable index for scoring rival stereotactic radiosurgery plans. Int J Radiat Oncol Biol Phys. 2003;57(4):1141-1149. https://doi.org/10.1016/s0360-3016(03)01563-3

[33] XXXX

[34] Rønn Hansen C, Crijns W, Hussein M, et al. RAdiotherapy Treatment plannINg study Guidelines (RATING): A framework for setting up and reporting on scientific treatment planning studies. Radiother Oncol. 2020 Dec;153:67-78. https://doi.org/10.1016/j.radonc.2020.09.033

[35] Guckenberger M, Andratschke N, Dieckmann K, et al. ESTRO ACROP consensus guideline on implementation and practice of stereotactic body radiotherapy for peripherally located early stage non-small cell lung cancer. Radiother Oncol. 2017 Jul;124(1):11-17. https://doi.org/10.1016/j.radonc.2017.05.012.

[36] Paganetti H, Beltran CJ, Both S, et al. Roadmap: proton therapy physics and biology. Phys Med Biol. 2020 Nov 23. https://doi.org/10.1088/1361-6560/abcd16.

[37] Ribeiro CO, Visser S, Korevaar EW et al. Towards the clinical implementation of intensity-modulated proton therapy for thoracic indications with moderate motion: robust optimised plan evaluation by means of patient and machine specific information. Radiother Oncol. 2021 Feb 2:S0167-8140(21)00014-1. https://doi.org/10.1016/j.radonc.2021.01.014.

Figure 1: Axial (left) and topogram (right) view of case 1 with PTV = 126 cc (a), case 2 with PTV = 152 cc (b) and case 3 with PTV = 73 cc (c). The red line illustrates the planning target volume (PTV) and the orange line shows the gross tumour volume (GTV).

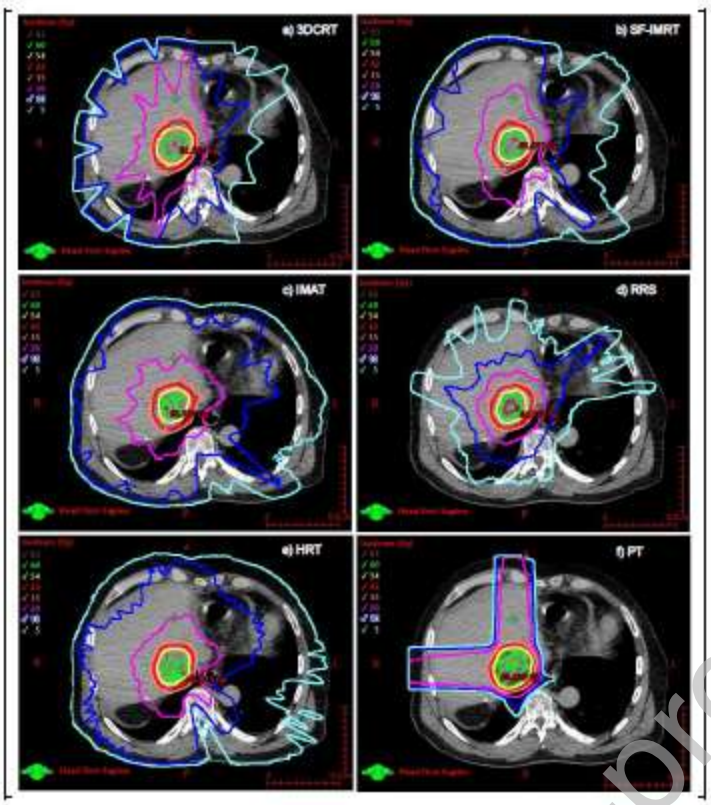

Figure 2: Dose distributions for different treatment techniques for case 1 with a) three-dimensional conformal radiation therapy (3D-CRT), b) static field intensity modulated radiation therapy (SF-IMRT), c) intensity modulated arc therapy (IMAT), d) robotic radiosurgery (RRS), e) helical radiotherapy (HT) and f) proton therapy (PT).

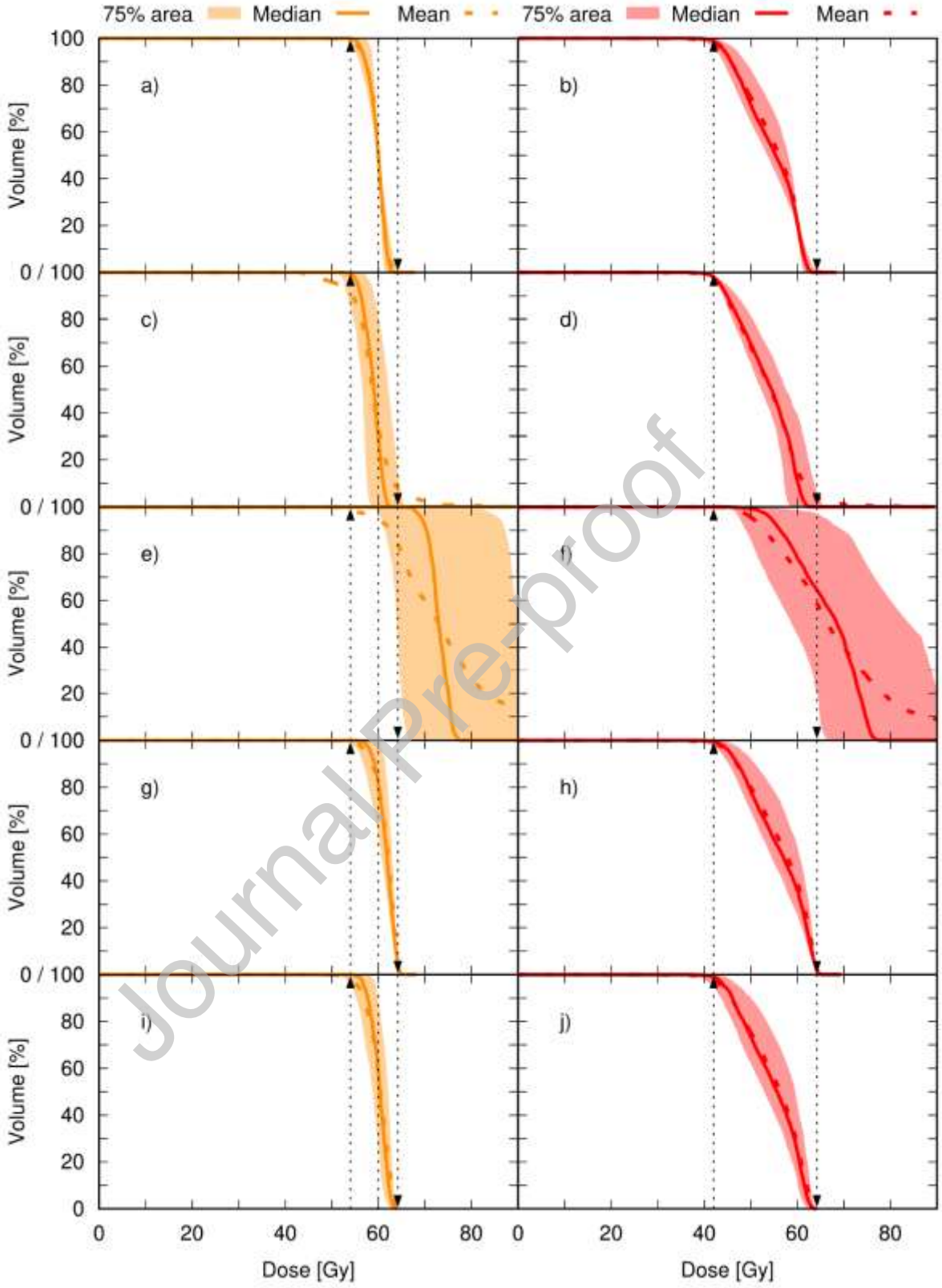

Figure 3: DVHs for different prescription methods for all plans of all cases. Clinical goals are marked with arrows. For GTV (orange) and PTV (red) the median and mean curve and the area of the central 75% of data is shown. The subplots refer to the follwing prescriptions: Reference prescription of 60 Gy to GTV $D_{50\%}$ (a, b), prescription of 42 Gy to PTV $D_{98\%}$ (c, d), 42 Gy to PTV D_{min} (e, f), 64.2 Gy to PTV $D_{2\%}$ (g, h) and 64.2 Gy to PTV D_{max} (i, j).

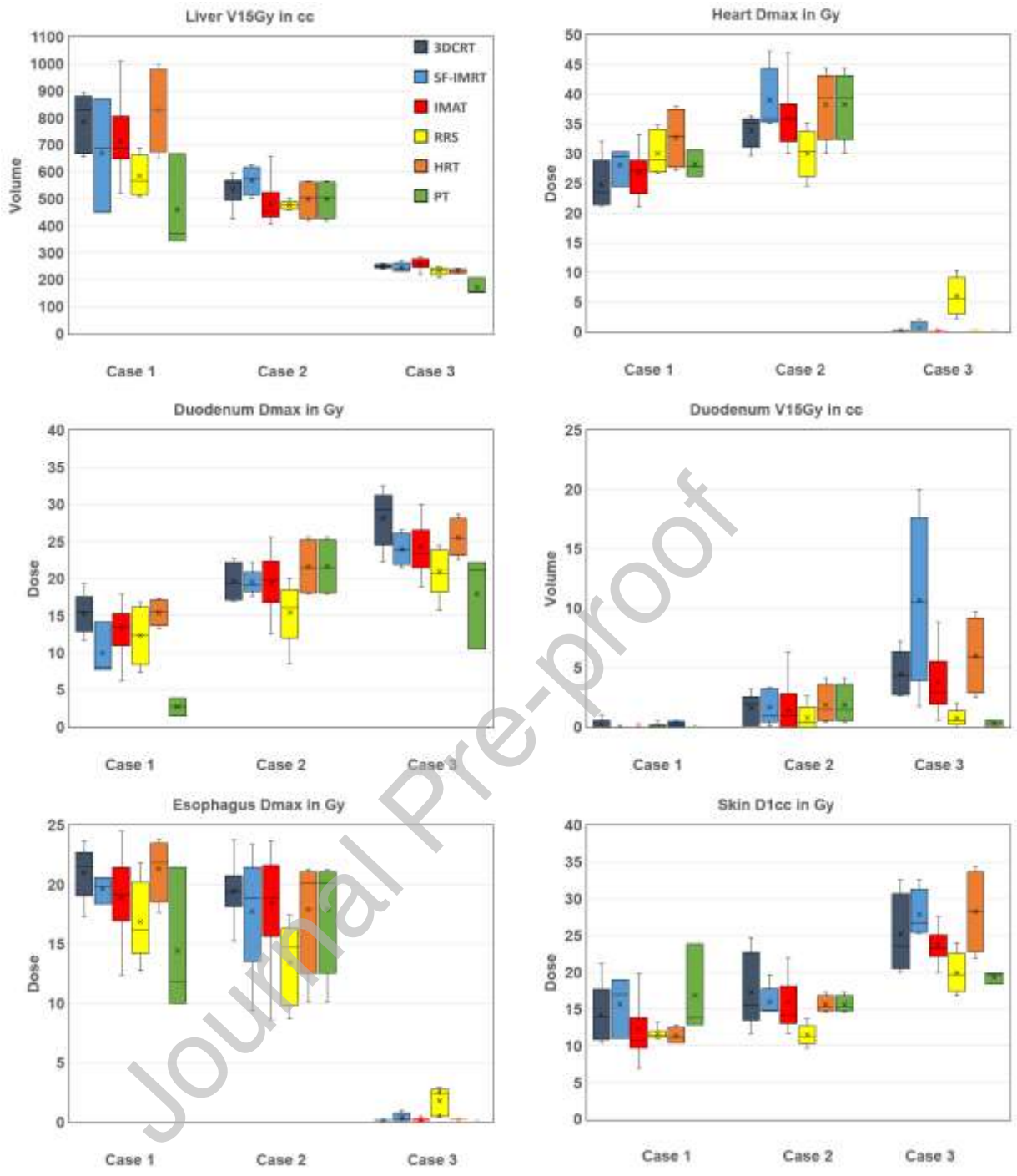

Figure 4: Organs at risk dosimetry as boxplots for various organs at risk for the 3 benchmark cases and for different techniques: three-dimensional conformal radiation therapy (3D-CRT), static field intensity modulated radiation therapy (SF-IMRT), intensity modulated arc therapy (IMAT), robotic radiosurgery (RRS), helical radiotherapy (HT) and proton therapy (PT).

Table 1: Key components from the dosimetric evaluation for $GTV_{D50\%}$ dose re-normalization at 3 x 20 Gy. PTV = Planning Target Volume. GTV = Gross Target Volume. IMAT = Intensity Modulated Arc Therapy (IMAT). SF-IMRT = Static Field Intensity Modulated Radiation Therapy. 3D-CRT = Three Dimensional Conformal Radiation Therapy. RRS = Robotic Radiosurgery. HRT = Helical Radiotherapy. PT = Proton Therapy. HI = Homogeneity Index. CI = ICRU 91 Conformity Index. CΔ = External Conformity Index. GI = Gradient Index

Mean ± standard deviation	**All Plans**	**3D-CRT**	**SF-IMRT**	**IMAT**	**RRS**	**HRT**	**PT**
Plan D_{max} (Gy)	63.9 ±1.6	63.2 ±1.4	64.5 ±1.2	64.0 ±1.6	64.7 ±1.3	62.5 ±1.3	63.5 ±2.0
GTV $D_{2\%}$ (Gy)	62.5 ±1.1	62.5 ±1.0	63.0 ±0.9	62.5 ±1.1	63.3 ±0.7	61.6 ±1.0	61.9 ±1.0
GTV $D_{98\%}$ (Gy)	55.8 ±1.9	55.8 ±1.3	55.6 ±1.4	55.9 ±1.7	53.7 ±1.9	56.9 ±1.9	57.1 ±1.8
PTV $D_{2\%}$ (Gy)	62.2±1.0	62.2 ±0.9	62.5 ±0.8	62.3 ±0.9	62.9 ±0.7	61.4 ±0.8	61.7 ±1.1
PTV $D_{98\%}$ (Gy)	42.9 ±3.3	42.0 ±1.6	42.5 ±1.7	43.4 ±3.4	41.2 ±3.6	42.6 ±4.2	44.7 ±4.8
PTV D_{mean}(Gy)	54.4 ±1.5	55.5 ±0.8	53.9 ±1.9	54.4 ±1.6	53.7 ±0.9	55.2 ±1.9	56.3 ±1.0
HI_{PTV}	0.35 ±0.07	0.36 ±0.04	0.37 ±0.05	0.34 ±0.07	0.40 ±0.07	0.33 ±0.08	0.29 ±0.1
HI_{GTV}	0.11±0.04	0.10 ±0.04	0.12 ±0.03	0.11 ±0.04	0.16 ±0.04	0.08 ±0.04	0.08 ±0.03
CI	1.19 ±0.18	1.22 ±0.08	1.19 ±0.18	1.19 ±0.21	1.13 ±0.06	1.30 ±0.24	1.21 ±0.08
CΔ	0.15 ±0.19	0.17 ±0.05	0.11 ±0.08	0.15 ±0.23	0.08 ±0.05	0.24 ±0.25	0.18 ±0.10
GI_{PTV} Case 1	3.86±0.90	3.91 ±0.20	4.39 ±0.94	3.94 ±0.39	3.19 ±0.30	3.75 ±0.20	2.68 ±1.25
GI_{PTV} Case 2	3.54±0.59	3.78 ±0.54	4.01 ±0.58	3.71 ±0.28	2.79 ±0.19	3.54 ±0.21	2.30 ±0.82
GI_{PTV} Case 3	3.20±0.45	3.14 ±0.33	3.43 ±0.69	3.40 ±0.25	2.80 ±0.11	3.25 ±0.32	2.19 ±0.21
GI_{PTV} all Cases	3.54 ±0.72	3.88 ±1.23	3.91 ±0.75	3.68 ±0.38	2.95 ±0.28	3.51 ±0.32	2.39 ±0.79
GI_{GTV} Case 1	4.31±0.78	3.75 ±0.14	4.83 ±0.43	4.45 ±0.61	4.48 ±0.47	4.07 ±0.84	2.73 ±0.16
GI_{GTV} Case 2	8.89±4.63	7.27 ±3.55	13.0 ±3.686	8.96 ±4.59	11.5 ±5.83 7	5.99 ±1.76	4.18 ±0.50

GI_{GTV} Case 1	11.4±1.940	11.4 ±0.575	12.3 ±1.296	11.8 ±2.139	10.5 ±0.981	11.3 ±0.850	7.98 ±1.44
GI_{GTV} all Cases	8.18 ±4.15	7.64 ±3.66	±4.28 10.77	9.21 ±7.80	8.59 ±4.52	7.12 ±3.38	4.96 ±2.47
Case 1 esophagus D_{max} (Gy)	18.8 ±3.5	21.8 ±1.1	19.6 ±1.1	18.9 ±3.1	16.9 ±3.3	20.3 ±3.3	14.4 ±6.2
Case 2 heart D_{max} (Gy)	35.3 ±5.3	33.9 ±2.7	39.5 ±5.0	36.2 ±4.5	30.0 ±4.1	38.3 ±6.0	31.6 ±8.2
Case 3 duodenum D_{max} (Gy)	24.0 ±4.3	28.1 ±3.9	24.0 ±2.1	24.3 ±3.8	20.9 ±3.3	25.5 ±2.6	18.0 ±6.4
Case 1 liver V_{15Gy} (cc)	694.±149.55	±92. 800.23	±211. 669.86	±115. 714.43	±72. 583.27	±180. 809.80	±178. 460.18
Case 2 liver V_{15Gy} (cc)	±74.7 489.5	±59. 535.24	±52.6 566.8	±66.1 482.5	±16. 476.40	±71.5 498.4	±43.5 346.1
Case 3 liver V_{15Gy} (cc)	±28.0 244.3	±7.6 249.9	±17.9 244.8	±18.6 258.0	±15. 232.04	±8.3 231.5	±30.8 171.5
Number of plans	132	16	11	68	16	12	9

Table 2: Ranking evaluation for the GTV $D_{50\%}$ prescription grouped by system for main and sub-scores (left side, score 1-4 as described in materials and methods) and Kruskal Wallis test results (right side, omnibus for all systems and post-hoc Tukey's Honest adjusted for pairwise comparison) of technique differences. PTV = Planning Target Volume, OAR = Organ at Risk, SF-IMRT = Static Field Intensity Modulated Radiation Therapy, 3D-CRT = Three Dimensional Conformal Radiation Therapy, RRS = Robotic Radiosurgery, HRT = Helical Radiotherapy, PT = Proton Therapy.

Ranking evaluation grouped by system	**3D-CRT**	**SF-IMRT**	**IMAT**	**RRS**	**HRT**	**PT**	**Kruskal Wallis and post-hoc Tukey's Honest test**	
All plan metrics, all systems							*All systems*	p < 0.001
Minimum	2.0	2.0	1.0	1.0	2.0	1.0	*PT vs. 3D-CRT*	p < 0.001
Maximum	4.0	4.0	4.0	3.0	4.0	2.0	*PT vs. SF-IMRT*	p = 0.002
Mean	3.1	2.7	2.6	2.3	2.9	1.1	*PT vs. IMAT*	p < 0.001
Standard Deviation	1.0	0.8	1.5	0.8	0.8	0.3	*PT vs. RSS*	p = 0.047
Median	3.0	3.0	2.0	2.0	3.0	1.0	*PT vs. HRT*	p < 0.001

All OARs, all systems							*All systems*	p < 0.001
Minimum	2.0	2.0	1.0	1.0	2.0	1.0	*PT vs. 3D-CRT*	p < 0.001
Maximum	3.0	4.0	4.0	4.0	4.0	1.0	*PT vs. SF-IMRT*	p < 0.001
Mean	2.7	3.0	2.6	2.4	2.9	1.0	*PT vs. IMAT*	p < 0.001
Standard Deviation	0.9	0.9	1.4	0.8	0.8	0.3	*PT vs. RSS*	p = 0.008
Median	3.0	3.0	3.0	2.0	3.0	1.0	*PT vs. HRT*	p < 0.001
All OARs, PT excluded							*PT excluded*	p = 0.01
Minimum	2.0	1.0	1.0	1.0	1.0	n/a	*RSS vs. 3D-CRT*	p = 0.199
Maximum	3.0	4.0	4.0	3.0	4.0	n/a	*RSS vs. SF-IMRT*	p = 0.077
Mean	2.5	2.7	2.4	1.8	2.9	n/a	*RSS vs. IMAT*	p = 0.185
Standard Deviation	0.8	0.8	1.4	0.6	0.8	n/a	*RSS vs. HRT*	p = 0.007
Median	2.0	3.0	2.0	2.0	3.0	n/a	*Other systems*	p > 0.19
OAR close to PTV, PT excluded							*PT excluded*	p < 0.001
Minimum	1.0	1.0	1.0	1.0	1.0	n/a	*RSS vs. 3D-CRT*	p < 0.001
Maximum	4.0	4.0	4.0	3.0	4.0	n/a	*RSS vs. SF-IMRT*	p = 0.002
Mean	2.9	2.9	2.3	1.4	2.8	n/a	*RSS vs. IMAT*	p = 0.01
Standard Deviation	1.0	0.8	1.4	0.5	0.8	n/a	*RSS vs. HRT*	p = 0.001
Median	3.0	3.0	2.0	1.0	3.0	n/a	*Other systems*	p > 0.25
Protocol deviations, all systems							*All systems*	p = 0.027
Minimum	2.0	1.0	1.0	1.0	2.0	1.0	*PT vs. HRT*	p = 0.035
Maximum	4.0	4.0	4.0	3.0	4.0	3.0	*Other systems*	p > 0.12
Mean	2.3	2.5	2.1	1.9	2.6	1.4		
Standard Deviation	0.8	0.8	1.3	0.7	0.8	0.4		
Median	2.0	2.0	2.0	2.0	2.0	1.0		

Table 3: Absolute deviations from planning objectives (left) and for the ranking evaluation of the absolute differences (right) and Kruskal Wallis test results (Omnibus for all systems and post-hoc Tukey's Honest adjusted for pairwise comparison) of technique differences. Significant differences are marked with an asterisk (*). GTV = Gross Target Volume, PTV = Planning Target Volume, IMAT = Intensity Modulated Arc Therapy (IMAT), SF-IMRT = Static Field Intensity Modulated Radiation Therapy, 3D-CRT = Three Dimensional Conformal Radiation Therapy, RRS = Robotic Radiosurgery, HRT = Helical Radiotherapy, PT = Proton Therapy.

	Absolute deviations from planning objectives (Gy)					**Ranking evaluation of the absolute deviations**				
Prescription method	GTV $D_{50\%}$	PTV $D_{98\%}$	PTV $D_{2\%}$	PTV D_{min}	PTV D_{max}	GTV $D_{50\%}$	PTV $D_{98\%}$	PTV $D_{2\%}$	PTV D_{min}	PTV D_{max}
IMAT	5.26	7.71	10.0	66.12	6.04	2.03	1.78	2.12	1.81	2.10
SF-IMRT	4.27	6.83	9.7	35.71	4.58	2.00	1.81	2.27	1.27	1.78
HRT	7.46	12.54	12.87	107.52	9.34	2.50	2.25	2.67	2.33	2.75
RRS	4.43	9.45	7.24	60.84	5.21	1.87	1.81	1.50	2.06	1.87
PT	8.97	17.99	15.68	116.63	10.44	3.22	2.89	3.00	2.56	3.11
3D-CRT	4.57	7.19	9.07	80.30	6.41	2.25	1.87	2.06	2.50	2.31
	p-values									
Comparing all systems	0.003*	0.011*	<0.001*	0.003*	0.001*	0.002*	0.037*	<0.001*	<0.001*	0.002*
IMAT vs. SF-IMRT	1.000	0.992	1.000	0.746	0.922	1.000	0.998	0.990	0.382	0.807
IMAT vs. HRT	0.317	0.362	0.239	0.227	0.248	0.493	0.757	0.338	0.444	0.285
IMAT vs. RRS	0.838	0.999	0.223	0.995	0.911	0.984	1.000	0.127	0.878	0.936
IMAT vs. PT	0.010*	0.012*	0.007*	0.375	0.012*	0.002*	0.014*	0.032*	0.168	0.020*
IMAT vs. 3D-CRT	0.998	0.935	0.994	0.045	0.969	0.899	0.993	1.000	0.037*	0.937
SF-IMRT vs. HRT	0.577	0.840	0.674	0.069	0.155	0.747	0.985	0.909	0.046*	0.108
SF-IMRT vs. RRS	0.976	0.992	0.485	0.653	1.000	0.998	1.000	0.204	0.160	0.999
SF-IMRT vs.	0.067	0.170	0.106	0.117	0.011	0.029	0.209	0.386	0.014	0.009

PT					*	*			*	*
SF-IMRT vs. 3D-CRT	0.999	0.999	0.992	0.018 *	0.738	0.968	1.000	0.991	0.003 *	0.525
HRT vs. RRS	0.110	0.402	0.008 *	0.707	0.116	0.380	0.902	0.007 *	0.982	0.153
HRT vs. PT	0.811	0.798	0.832	1.000	0.872	0.464	0.554	0.921	0.991	0.903
HRT vs. 3D-CRT	0.744	0.940	0.228	1.000	0.832	0.985	0.982	0.501	0.987	0.904
RRS vs. PT	0.003 *	0.024 *	<0.001*	0.787	0.006 *	0.003 *	0.071	<0.001*	0.799	0.012 *
RRS vs. 3D-CRT	0.801	0.899	0.777	0.437	0.713	0.750	0.999	0.436	0.695	0.679
3D-CRT vs. PT	0.102	0.238	0.012 *	1.000	0.195	0.117	0.155	0.086	1.000	0.309