Consequences of anorectal cancer atlas implementation in the cooperative group setting: Radiobiologic analysis of a prospective randomized in silico target delineation study

Panayiotis Mavroidis[a,b,*], **Drosoula Giantsoudis**[c], **Musaddiq J. Awan**[d], **Jasper Nijkamp**[e], **Coen R. N. Rasch**[f], **Joop C. Duppen**[e], **Charles R. Thomas Jr.**[g], **Paul Okunieff**[h], **William E. Jones III**[a,i], **Lisa A. Kachnicc**[j], **Niko Papanikolaou**[a], and **Clifton D. Fuller**[d] **on Behalf of the Southwest Oncology Group Radiation Oncology Committee**

[a]Department of Radiation Oncology, The University of Texas Health Science Center at San Antonio, San Antonio, TX, USA [b]Department of Oncology-Pathology, Karolinska Institutet and Stockholm University, Stockholm, Sweden [c]Department of Radiation Oncology, Massachusetts General Hospital Cancer Center, Boston, MA, USA [d]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA [e]Department of Radiotherapy, The Netherlands Cancer Institute, Antoni van Leeuwenhoek Huis, Amsterdam, The Netherlands [f]Department of Radiotherapy, Academic Medical Center, Amsterdam, The Netherlands [g]Department of Radiation Medicine, Oregon Health & Science University, Portland, OR, USA [h]Department of Radiation Oncology, University of Florida, Gainesville, FL, USA [i]Department of Radiation Oncology, Audie L. Murphy Veterans Administration Hospital, San Antonio, TX, USA [j]Department of Radiation Oncology, Boston University Medical Center, Boston, MA, USA

Abstract

Purpose—The aim of this study is to ascertain the subsequent radiobiological impact of using a consensus guideline target volume delineation atlas.

Materials and methods—Using a representative case and target volume delineation instructions derived from a proposed IMRT rectal cancer clinical trial, gross tumor volume (GTV) and clinical/planning target volumes (CTV/PTV) were contoured by 13 physician observers (Phase 1). The observers were then randomly assigned to follow (atlas) or not-follow (control) a consensus guideline/atlas for anorectal cancers, and instructed to re-contour the same case (Phase 2).

Results—The atlas group was found to have increased tumor control probability (TCP) after the atlas intervention for both the CTV (p < 0.0001) and PTV1 (p = 0.0011) with decreasing normal

---

[*]Corresponding author: Panayiotis Mavroidis, Ph.D., Division of Medical Physics, Department of Radiological Sciences, Cancer Therapy and Research Center, University of Texas Health Sciences Center San Antonio, 7979 Wurzbach Rd, MC 7889, San Antonio TX 78229-4427, USA, mavroidis@uthscsa.edu (P. Mavroidis), Tel: +1 210-450-1027, +1 210-478-9703.

tissue complication probability (NTCP) for small intestine, while the control group did not. Additionally, the atlas group had reduced variance in TCP for all target volumes and reduced variance in NTCP for the bowel. In Phase 2, the atlas group had increased TCP relative to the control for CTV (p = 0.03).

Conclusions—Visual atlas and consensus treatment guidelines usage in the development of rectal cancer IMRT treatment plans reduced the inter-observer radiobiological variation, with clinically relevant TCP alteration for CTV and PTV volumes.

Keywords

Anorectal cancer; Target delineation; Radiobiological analysis; Atlas implementation; Inter-observer variation

Introduction

In the pre-conformal radiotherapy era, standardized fields based on bony anatomy were utilized to ensure uniformity of treated regions. However, in the era of volume-based delineation, considerable operator dependent variation exists in target volume delineation. This factor affects the planned dose distributions complicating the clinical trial quality assurance and preventing the compatible comparison of treatment protocols.

The location of organs-at-risk (OAR) and their tolerance doses constitute a major factor that determines the prescribed dose in radiation treatment planning. OARs are usually located in the immediate vicinity of the CTV limiting dose deliverable to target volumes [1]. Intensity Modulated Radiotherapy (IMRT) generates more conformal distributions as compared to older techniques resulting in reduction of radiation dose and toxicity to OARs and thus potentially improving clinical outcomes.

Comparatively low tolerance doses, which characterize involved OARs relative to tumoricidal dose thresholds, are usually the major constraints in pelvic radiotherapy, especially when gross tumor volume (GTV) and clinical target volume (CTV) arise from potentially dose limiting normal tissue (as in cancer of the rectal mucosa). Isodose charts, dose volume histograms (DVH), dose-volume parameters and conformity-based indices are currently used for treatment plan evaluation. However, these evaluation measures do not account for radiobiological characteristics of tumors nor normal tissues [2], and thus are, at best, indirect correlates of clinically relevant parameters. Consequently, radiobiological measures should ideally be considered in order to estimate the expected treatment outcome. The applied radiobiological measures provide the expected treatment outcome within a clinical range of uncertainty, whereas the DVHs and other dosimetric quantities do not provide any association to the treatment outcome. This analysis uses tumor control probability (TCP), normal tissue complication probability (NTCP) and complication-free tumor control probability (P_+) as direct treatment plan evaluation parameters [3–5] to assess the utility of atlas-based educational intervention on plan quality.

In a previous prospective randomized effort [6], implementation of a consensus guideline-based atlas [7] demonstrably improved CTV but not GTV volumetric concordance with an

expert reference for a standardized rectal cancer case. Additionally, consensus atlas use reduced inter-observer CTV delineation variance to a statistically significant degree.

The primary aim of this secondary analysis was determining whether the aforementioned alteration in volumetric coverage resulted in clinically meaningful differences in tumor control probabilities. Secondly, this analysis sought to estimate radiobiological parameter (e.g. TCP, NTCP, P_+) variability demonstrable in a standardized contouring protocol to serve as a benchmark for future cooperative group trials. Thus, an evaluation of radiobiological differentials attributable to consensus guideline atlas implementation could be achieved.

Methods and materials

This prospective *in silico* study was deemed exempt and was conducted under the auspices of the University of Texas Health Science Center at San Antonio institutional review board. Pilot data from the study have been presented previously [6]. Briefly, thirteen radiation oncologist observers from eight SWOG-affiliated institutions were recruited and were asked to contour a standardized case (an anonymized patient with Stage T3N0M0 adenocarcinoma of the rectum) with instructions from an (at that time) in-development SWOG protocol (S0713: "*A Phase II Study of Oxaliplatin, Capecitabine, Cetuximab and Radiation in Pre-operative Therapy of Rectal Cancer*", ClinicalTrials.gov Identifier NCT00686166) (Supplementary Fig. 1). The observers were experienced in the treatment of carcinoma of the rectum and in the delineation of rectal carcinomas. Subsequently, the observers were randomly assigned to receive an electronic copy of an unpublished (at that time) rectal cancer atlas [7]. The observers re-contoured the same case with (atlas group – six observers) or without the atlas (control group – six observers). The use of the atlas for the re-contouring of the tissues will be notated as intervention. Data collection was performed using "Big Brother", a custom target volume delineation evaluation software platform developed at The Netherlands Cancer Institute [6]. The observers were asked to contour the GTV, CTV_A, and CTV_B targets (Supplementary Table A) [6, 7]. The CTV encompassed the GTV as well as the peri-rectal, pre-sacral, internal and external iliac nodal regions. The PTV1 is defined as a GTV expansion of 2.0–3.0 cm, including the CTV, whereas the PTV2 (the boost volume) is defined as an expansion of the GTV by 2.0 cm including the whole of the sacral hollow. Contours from a "reference expert" involved in the development of the RTOG consensus atlas and guidelines [LAK] served as a comparator for the observer-derived contours. During the study period, none of the observers other than the reference expert had a previous knowledge of this atlas. A statistical comparison of the volume differentials and *post hoc* exploratory contour surface variability analysis [8, 9] was previously reported [6]. In this analysis, the statistical significance of the presented results is investigated.

Treatment Planning

Treatment planning was performed using a commercial treatment planning software (Pinnacle, Philips Medical Systems, Inc.). A volumetric modulated arc technique (VMAT), which employs 2 arcs of 6 MV photons, was applied. The organs-at-risk were delineated as ROIs by a single observer [CDF]. The individual treatment plans were produced by a single

physicist [DG] using the dosimetric constraints for the target volumes and organs at risk that were specified in the SWOG S0713 protocol (Supplementary Table B). The individual treatment plans were produced using the first set of delineations of each observer. The same treatment plans were subsequently applied on the second sets of delineations of each observer (no re-planning took place, only renormalization), in order to determine the impact of delineation/segmentation alone upon plan quality.

Radiobiological measures for treatment plan evaluation

Secondary radiobiological evaluation was performed using previously defined literature-derived metrics [10]. Tumor response was calculated using the Poisson model, with parallel tumor structural organization assumed (i.e. 100% clonogenic kill required for tumor control). Thus, tumor control probability (TCP) for a tumor volume is given by the expression:

$$TCP=\prod_{i=1}^{M}P(D_i)^{\Delta v_i} \quad (1)$$

where M is the total number of voxels or sub-volumes in the target. Response of a normal tissue to a non-uniform dose distribution was obtained using the relative seriality model, with normal tissue complication probability expressed as [3]:

$$NTCP=1-\prod_{j=1}^{N_{organs}}\left(1-P_{\mathrm{I}}^{j}\right)=1-\prod_{j=1}^{N_{organs}}\left(1-\left[1-\prod_{i=1}^{M_j}\left(1-P^j(D_i)^{s_j}\right)^{\Delta v_i}\right]^{1/s_j}\right) \quad (2)$$

where P_{I}^{j} is the probability of injuring organ j and N_{organs} is the total number of OARs. $P^j(Di)$ is the probability of response of the organ j having the reference volume and being irradiated to dose D_i. $v_i = V_i/V_{\mathrm{ref}}$ is the fractional subvolume of the organ (V_i) that is irradiated at the dose level D_i compared to the reference volume (V_{ref}) for which the values of the model parameters have been calculated. M_j is the total number of voxels or subvolumes in the organ j, and s_j is the relative seriality parameter that characterizes the internal organization of that organ.

Complication-free tumor control probability (P_+) was used to estimate the overall effectiveness of a treatment plan, expressed in terms of PTV tumor control probability (TCP) and normal tissue complication probability (NTCP) [3]:

$$P_+=TCP(1-NTCP) \quad (3)$$

Here, the TCP that was used for calculating the P_+ values was based on PTV2. Biologically effective uniform dose, $\overline{\overline{D}}$, is defined as the dose that causes the same TCP or NTCP as the actual dose distribution delivered to the patient [4]. Generalized equivalent uniform dose (gEUD) was used as a mean dose to a given tissue accounting for radiobiological characteristics of that tissue [5]. Dose-response parameters for organs involved in this study, were derived from previously published data [8, 9, 11–13] and are shown in Table 1.

Statistical analysis

Statistical analysis was performed using the JMP software package (SAS Insititute, Cary, NC, USA). The one-sided Wilcoxon Signed-Rank test was used as a non-parametric measure for matched pair analysis (e.g. Phase 1 vs. Phase 2). The Wilcoxon Rank Sums Test was used to assess distributional equivalence/nonequivalence between post-intervention cohorts for both groups. The Brown-Forsythe test was used as a non-parametric measure to determine whether variance in TCP/NTCP changed across an ROI for both interventions.

Results

Table 2 presents an outline of dosimetric and radiobiological measures that evaluate treatment plan efficacy. In this table, for every observer's treatment plan and organ delineation set, the values the different measures were derived. Fig. 1 shows normalized cumulative DVHs of the targets, GTV, CTV and both PTVs for the expert, the atlas-assisted group and the control group.

Atlas intervention increases TCP for CTV and PTV1 and decreases NTCP for small intestine

TCP or NTCP was calculated for each user across each structure and each phase of the trial. This means that on a given observer's treatment plan the organ delineation sets of all the observers in the group (atlas or control) were applied and the corresponding TCPs were calculated. Based on those TCPs, the average TCP was calculated. This was repeated for every plan and the overall average TCP of each target was derived (the same dataset was used for Tables 4 and 5). Mean TCP (for tumor ROIs) and NTCP (for small intestine) for both cohorts before and after the intervention are shown in Table 3. Based on Tables 2 and 3, for CTV the atlas intervention improved TCP while the control group showed much smaller differences between the two phases. This suggests that the intervention itself led to improved TCP and was not resulting from re-contouring the same case. Additionally, NTCP for small intestine was significantly improved showing that atlas intervention led to decreased tissue complication probability. Table 3 shows that there were small changes in the TCP for the GTV for both the control group and the atlas intervention group leaving a non-definitive answer as to whether the use of an atlas improved dose to this target.

Atlas intervention reduces variance in TCP for all tumor volumes and variance in NTCP for bowel

The TCP or NTCP standard deviations between users stratified by ROI and each phase of the trial and are shown in Table 4. Across all ROIs for the control group there was no significant alteration in TCP or NTCP variance. However, for the atlas-assisted group, across GTV, CTV and small intestine there was significant reduction in TCP and NTCP variance, implying atlas use improved standardization of plan parameters. These findings are illustrated in Fig. 1, where the reduction in inter-observer DVH variation in phase 2 in the atlas-assisted group is much larger than that in the control group.

Atlas exposure yielded increased TCP relative to control for CTV but not GTVs or PTVs

Regarding atlas use in the delineation of PTV1 (Wilcoxon p=0.3, n.s.) and PTV2 (Wilcoxon p=0.3, n.s.), the corresponding TCP values between both cohorts were not significantly different. For CTV (Wilcoxon p=0.03), a statistically significant difference in the corresponding TCP values between the atlas and control cohorts after atlas exposure was observed with the atlas cohort having a larger mean TCP. This was opposite for GTV (Wilcoxon p=0.02), for which a statistically significant difference in the corresponding TCP values between the atlas and control cohorts was observed, but the control group had greater TCP, suggestive that even with the atlas, the control cohort was better at contouring the GTV. Conjunctively, these results suggest that atlas use improved coverage to the CTV for atlas users relative to the control group after exposure but the addition of margins for PTV coverage were sufficient to overcome inter-observer variance. Fig. 1 provides an illustration of the DVHs for the different groups, phases and structures, which confirm the aforementioned findings.

Atlas intervention increases P_+

P_+ was calculated for each user for each phase of the trial. Mean P_+ for the control group and atlas group for the PTV2 before and after the intervention are shown in Table 5. The atlas cohort showed an improvement in P_+ while the control group failed to show change, suggesting that atlas intervention increased complication-free tumor control probability (Supplementary Fig. 2). Importantly, the difference between atlas and control P_+ was negligible (p>0.05).

Discussion

Inter-observer delineation variance adds an oft-unaccounted for level of "noise" to clinical and dosimetric data in cooperative group trials, an area of great concern as protocol non-compliance has been demonstrably associated with reduced outcomes [14]. The present study indicates that the use of an atlas or a protocol with guidelines for performing target delineation may partially ameliorate this source of uncertainty [6, 15].

Based on the results shown in Fig. 1 and Table 2, it seems that the control group delineated the targets similarly in the first and second phases because they used the same approach and they did not have additional information that would alter their way of dealing with the case. In the atlas-assisted group, the differences are more pronounced because they got access to information that changed considerably their initial approach by which they delineated those targets.

Our data, as per Tables 3–5 demonstrate that an atlas-aided target delineation intervention may both improve and standardize radiobiologic parameters for treatment planning in anorectal cancer. Atlas intervention results in increased TCP for the atlas cohort for both the GTV and CTV, and the atlas group also showed significantly better TCP for the CTV relative to the control group after the intervention. Additionally, atlas use significantly reduced variance in TCPs for all tumor volumes and variance in NTCP for small intestine. Finally, the complication-free tumor control probability was improved after atlas

intervention. Conjunctively, these findings suggest that atlas utilization may increase and improve conformality of radiobiologic probabilities of tumor control and tissue complication amongst multiple observers, as in a cooperative group setting.

The TCP values depend on the radiobiological parameters of the targets and the dose distribution that each target receives. In the study, since the same parameters were used for all the plans and delineation phases the varying factor was the dose distributions produced by the different plans. These dose distributions are the final outcome of treatment planning and the degree of satisfaction of the constraints, which were used during the optimization process, may be affected by the patient geometry at hand (as this is expressed through the organ delineations). This happens because the organ delineation and the treatment plan optimization processes have a perplexed relation.

Due to the fact that the qualities of the produced treatment plans are related to the targets that were used, another comparison between the delineations of the different observers was performed using the plan that was produced based on the target delineations of the reference expert (Table 3, lower panel). This comparison verified the results that were obtained when the treatment plans of the observers were used, which showed that the TCP differences of the atlas group were statistically significant for GTV and CTV. This comparison also verified that the differences in inter-observer variance between the atlas-assisted and control groups were not statistically significant (neither before not after the intervention of atlas).

Inter-observer delineation variability can be estimated by several criteria [16]. Our previous effort utilized simple volumetric analyses [6]. In the present study, dose-response models based on radiobiological parameters derived from clinical data, were also used to directly assess the expected clinical impact of said volumetric alterations. In our estimation, these dose-response models prove a more direct estimator of clinically relevant endpoints. The accuracy of the radiobiological models is necessarily dependent on the accuracy of input parameters that describe the dose-response relations of tumors and normal tissues. In this study, most of the tissue response parameters have been taken from recently published clinical studies, where the determined confidence intervals are at clinically acceptable levels (e.g. uncertainty of approximately 5% in D_{50}). The uncertainty in the absolute knowledge of the expected responses does not affect the principal conclusions, given the relative magnitude of radiobiologic differentials seen at equivalent dose prescription levels. Additionally, the radiosensitization effect of concurrent chemotherapy was not explicitly modeled, though such effect should theoretically be independent of target definition differentials. Nonetheless, failure to inculcate chemotherapy effects leads to underestimation of TCP and subsequent overestimation of P_+.

Since the importance of using delineation instructions and an atlas in clinical trials is known to the radiation oncology community, the outcome of the study is predictable from a qualitative point of view, i.e. that the use of an atlas is superior compared to the situation where no atlas is used. However, the extent of improvement in terms of treatment outcome had not been adequately investigated in the past using volumetric-based radiobiological indices, which would be able to quantify the clinical impact of the differences observed in the isodose distributions and DVHs of the different targets.

It is recognized that only one rectal cancer case was used in this study and a statistically significant sample of cases would give a more accurate picture of the comparisons. However, for each observer, the treatment plans were produced using optimization algorithms, which calculated the final dose distributions based on predefined physical constrains such as prescribed doses to the targets and dose limits to the OARS (inverse physical optimization). For purposes of this study, the use of a single case enhances the individualization of the different delineations and demonstrates clearer the inter-observer variability. On the other hand, incorporating multiple patients would provide a more robust statistical validity for the findings increasing however the number of the factors to be correlated and the complexity of the analysis.

The current study provides an outline of how cooperative groups including SWOG might address concerns regarding target delineation practice standardization. Our data suggest that atlas utilization, in an *in silico* rectal cancer study model, provides detectable alteration of plan quality. Use of a "dummy run" approach with subsequent radiobiologic modeling, as performed herein, offers an efficient, reproducible mechanism to evaluate quality assurance steps (e.g. atlas use [6, 17], educational interventions [18, 19], or auto-segmentation methods [20]).

Conclusion

The analysis of a single case indicates that the use of a visual atlas and consensus treatment guidelines in the development of rectal cancer IMRT treatment plans may result in a reduction in the inter-observer variability of the dosimetric and radiobiological parameters for the GTV/CTV ROIs, and an *in silico* improvement in group tumor control expectations which may be potentially clinically meaningful. Implementation of supplementary visual atlas-based target volume delineation procedures should be employed as a low-cost, effective methodology of quality-assurance in clinical trials involving highly conformal radiotherapy modalities for anorectal cancers. Further studies are needed to quantify the utility of atlas-based interventions for cooperative group studies in other anatomic sites.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Wambersie A, Landberg T, Gahbauer R. Prescribing, recording and reporting photon beam therapy: the problem of margins (the recent ICRU recommendations, Report #62, 1999). Patras Medical Physics. 1999; 99:25–31.

2. Brahme, A. Which parameters of the dose distribution are best related to the radiation response of tumors and normal tissues?. Proceedings fo the Interregional Seminars for Europe, the Middle East and Africa Organized by the IAEA; Leuven. 1994;

3. Källman P, Lind BK, Brahme A. An algorithm for maximizing the probability of complication-free tumour control in radiation therapy. Phys Med Biol. 1992; 37:871–90. [PubMed: 1589457]

4. Mavroidis P, Lind BK, Brahme A. Biologically effective uniform dose (D⁻) for specification, report and comparison of dose response relations and treatment plans. Phys Med Biol. 2001; 46:2607–30. [PubMed: 11686278]

5. Niemierko A. A generealied concept of equivalent uniform dose (abstr.) Med Phys. 1999; 26:1100.

6. Fuller CD, Nijkamp J, Duppen JC, et al. Prospective randomized double-blind pilot study of site-specific consensus atlas implementation for rectal cancer target volume delineation in the cooperative group setting. Int J Radiat Oncol Biol Phys. 2011; 79:481–9. [PubMed: 20400244]

7. Myerson RJ, Garofalo MC, El Naqa I, et al. Elective clinical target volumes for conformal therapy in anorectal cancer: a radiation therapy oncology group consensus panel contouring atlas. Int J Radiat Oncol Biol Phys. 2009; 74:824–30. [PubMed: 19117696]

8. Mavroidis P, Lind BK, Van Dijk J, et al. Comparison of conformal radiation therapy techniques within the dynamic radiotherapy project 'Dynarad'. Phys Med Biol. 2000; 45:2459–81. [PubMed: 11008949]

9. Mavroidis P, al-Abany M, Helgason AR, et al. Dose-response relations for anal sphincter regarding fecal leakage and blood or phlegm in stools after radiotherapy for prostate cancer. Radiobiological study of 65 consecutive patients. Strahlenther Onkol. 2005; 181:293–306. [PubMed: 15900425]

10. Kutcher GJ. Quantitative plan evaluation: TCP/NTCP models. Front Radiat Ther Oncol. 1996; 29:67–80. [PubMed: 8742889]

11. Emami B, Lyman J, Brown A, et al. Tolerance of normal tissue to therapeutic irradiation. Int J Radiat Oncol Biol Phys. 1991; 21:109–22. [PubMed: 2032882]

12. Ågren Cronqvist, AK. Quantification of the response of heterogeneous tumors and organized normal tissues to fractionated radiotherapy. University of Stockholm; Stockholm: University Press; 1995.

13. Ferreira BC, Svensson R, Löf J, et al. The clinical value of non-coplanar photon beams in biologically optimized intensity modulated dose delivery on deep-seated tumours. Acta Oncol. 2003; 42:852–64. [PubMed: 14968947]

14. Peters LJ, O'Sullivan B, Giralt J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02. 02. J Clin Oncol. 2010; 28:2996–3001. [PubMed: 20479390]

15. Rosenthal DI, Asper JA, Barker JL Jr, et al. Importance of patient examination to clinical quality assurance in head and neck radiation oncology. Head Neck. 2006; 28:967–73. [PubMed: 16823872]

16. Fotina I, Lütgendorf-Caucig C, Stock M, et al. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlenther Onkol. 2012; 188:160–7. [PubMed: 22281878]

17. Nijkamp J, de Haas-Kock DF, Beukema JC, et al. Target volume delineation variation in radiotherapy for early stage rectal cancer in the Netherlands. Radiother Oncol. 2012; 102:14–21. [PubMed: 21903287]

18. Kalpathy-Cramer J, Bedrick SD, Boccia K, et al. A pilot prospective feasibility study of organ-at-risk definition using Target Contour Testing/Instructional Computer Software (TaCTICS), a training and evaluation platform for radiotherapy target delineation. AMIA Annu Symp Proc. 2011:654–63. [PubMed: 22195121]

19. Gwynne S, Spezi E, Wills L, et al. Toward semi-automated assessment of target volume delineation in radiotherapy trials: The SCOPE 1 Pretrial Test Case. Int J Radiat Oncol Biol Phys. 2012; 84:1037–42. [PubMed: 22878126]

20. Anders LC, Stieler F, Siebenlist K, et al. Performance of an atlas-based autosegmentation software for delineation of target volumes for radiotherapy of breast and anorectal cancer. Radiother Oncol. 2012; 102:68–73. [PubMed: 21962822]
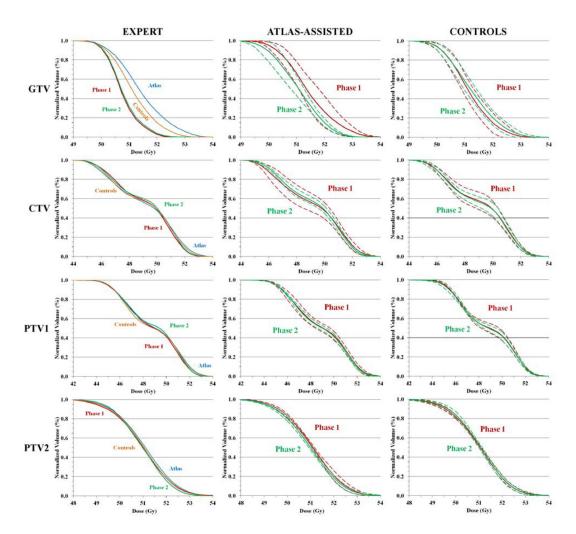
Fig. 1.
The normalized cumulative dose volume histograms (DVHs) of the GTV, CTV, PTV1 and PTV2. The DVHs are based on the first (Phase 1) and second (Phase 2) delineations of the expert (left panel), the atlas-assisted group (middle panel) and control group (right panel). In the left panel, the Phase 1 and Phase 2 delineations of the expert are plotted against the first (Phase 1) delineations of the atlas-assisted and control groups. In the middle and right panels, the solid lines indicate the average DVHs of the Phase 1 and Phase 2 delineations and the dashed lines indicate the one standard deviation inter-observer DVH variation within the corresponding groups.

Table 1

Summary of the model parameter values for the targets and organs-at-risk involved in the investigated cancer type. D_{50} is the 50% response dose, γ is the maximum normalized value of the dose-response gradient and s is the relative seriality, which characterizes the volume dependence of the organ.

Structure	D_{50} (Gy)	γ	s	a/β	Endpoint
Target Volume	41.7	2.2	—	10.0	control
Rectum	80.0	2.2	0.7	3.0	necrosis, stenosis
Bladder	80.0	3.0	0.18	3.0	symptomatic contracture
Small intestine	53.6	2.3	1.5	3.0	obstruction perforation
Femur	65.0	2.7	1.0	3.0	necrosis

Table 2

Summary of the dosimetric and radiobiological results for the treatment plans that were developed based on the target delineations of the expert (upper panel), control group (middle panel) and atlas-assisted group (lower panel). The organs at risk were contoured by a single physician. The response probability, generalized equivalent uniform dose, mean dose, maximum dose and minimum dose of each target and organ at risk are presented. The odd columns show the results of the first organ delineation, whereas the even columns express the differences between the first and second organ delineations. Here, for every observer's treatment plan and organ delineation set, the presented values were derived.

Tissue	TCP or NTCP %	TCP or NTCP %	gEUD/Gy	gEUD/Gy	$\bar{D}$/Gy	$\bar{D}$/Gy	D_{max}/Gy	D_{max}/Gy	D_{min}/Gy	D_{min}/Gy
					EXPERT					
GTV	84.1	0.8	50.6	0.3	50.6	0.3	53.0	0.5	49.0	0.3
CTV	78.5	0.1	48.6	0.0	49.1	0.1	53.5	0.3	43.5	0.3
PTV1	76.5	0.2	48.0	0.1	48.6	0.1	54.0	−0.3	41.0	1.0
PTV2	84.4	0.1	50.7	0.1	50.8	0.0	55.0	−0.5	46.0	0.8
Small intestine	4.0	0.0	43.2	−0.2	21.8	0.2	49.5	0.3	1.0	0.3
					CONTROL GROUP					
GTV	85.1 ± 0.6	0.2 ± 0.3	51.0 ± 0.2	0.1 ± 0.1	51.0 ± 0.2	0.1 ± 0.1	53.3 ± 0.2	0.2 ± 0.3	49.0 ± 0.3	0.1 ± 0.3
CTV	78.1 ± 1.3	0.0 ± 0.7	48.5 ± 0.4	0.0 ± 0.3	49.1 ± 0.4	0.0 ± 0.2	53.9 ± 0.4	−0.3 ± 0.5	43.3 ± 0.7	0.1 ± 0.3
PTV1	76.4 ± 0.6	−0.1 ± 0.9	47.9 ± 0.2	0.0 ± 0.3	48.6 ± 0.2	0.0 ± 0.2	54.1 ± 0.5	0.8 ± 0.8	40.9 ± 1.0	−0.2 ± 2.0
PTV2	84.6 ± 0.2	0.1 ± 0.1	50.7 ± 0.1	0.0 ± 0.1	50.9 ± 0.1	−0.1 ± 0.4	54.3 ± 0.8	0.0 ± 1.2	46.5 ± 0.7	0.3 ± 0.8
Small intestine	3.4 ± 0.4	−0.4 ± 0.9	34.8 ± 1.5	0.8 ± 2.3	20.8 ± 2.7	−0.2 ± 4.5	49.5 ± 0.3	0.2 ± 0.5	1.3 ± 0.9	−0.5 ± 0.9
					ATLAS-ASSISTED GROUP					
GTV	85.7 ± 0.8	−0.9 ± 1.1	51.2 ± 0.3	−0.4 ± 0.5	51.3 ± 0.4	−0.4 ± 0.5	53.5 ± 0.6	−0.2 ± 0.5	49.1 ± 0.5	−0.3 ± 0.8
CTV	78.1 ± 2.0	0.4 ± 1.7	48.5 ± 0.6	0.1 ± 0.5	49.1 ± 0.6	0.1 ± 0.5	53.9 ± 0.8	−0.2 ± 0.6	43.3 ± 0.3	0.3 ± 0.4
PTV1	76.8 ± 1.1	−0.2 ± 1.0	48.0 ± 0.3	−0.1 ± 0.3	48.8 ± 0.3	−0.1 ± 0.3	54.8 ± 1.4	−0.4 ± 1.1	40.0 ± 2.3	0.4 ± 1.9
PTV2	84.7 ± 0.1	−0.3 ± 0.2	50.8 ± 0.0	−0.1 ± 0.1	50.9 ± 0.1	−0.1 ± 0.1	54.5 ± 0.8	0.2 ± 0.8	46.5 ± 0.2	0.0 ± 0.3
Small intestine	4.7 ± 0.8	−0.5 ± 0.9	36.6 ± 2.1	0.1 ± 0.6	23.1 ± 4.3	−0.3 ± 2.3	50.2 ± 0.6	−0.5 ± 0.9	1.1 ± 0.8	−0.5 ± 0.7

Table 3

The results of the statistical analysis of the tumor control probabilities (TCP) of the gross and clinical target volumes (GTV and CTV) and the normal tissue complication probability (NTCP) of small intestine are shown for the Atlas-assisted and Control groups (p value for one sided Wilcoxon Signed-Rank Test). The small intestine was the only OAR that showed some risk for complications, which consequently also represents the overall NTCP. Here, the TCPs were calculated for every observer's treatment plan and the organ delineation sets of all the observers of the group (atlas or control).

ROI	Group	Mean Phase 1	Mean Phase 2	Mean Difference	SEM	p-value (* = p < 0.05)
Comparisons using the individual plans of the observers						
GTV	Control	84.9	85.2	0.3	0.04	<0.001*
	Atlas	85.4	84.8	−0.6	0.17	0.016*
CTV	Control	72.5	74.1	1.6	2.67	0.35
	Atlas	69.2	78.0	8.8	2.70	<0.001*
Small Intestine	Control	3.5	3.8	0.3	0.12	0.97
	Atlas	4.6	4.2	−0.4	0.12	0.004*
Comparisons using the plan of the expert						
TVG	Control	84.1	84.6	0.6	0.04	0.016*
	Atlas	84.1	84.7	0.7	0.03	0.016*
CTV	Control	78.5	75.3	−3.2	2.26	0.66
	Atlas	69.1	77.5	8.4	6.82	0.047*

Table 4

The results of the statistical analysis of the variances in tumor control probability (TCP) for the gross and clinical target volumes (GTV and CTV) as well as in normal tissue complication probability (NTCP) for small intestine are shown for the Atlas-assisted and Control groups (p-value for Brown-Forsythe test for difference in variance). The small intestine was the only OAR that showed some risk for complications, which consequently also represents the overall NTCP. Here, the TCPs were calculated for every observer's treatment plan and the organ delineation sets of all the observers of the group (atlas or control).

	ROI	Group	Standard Deviation Phase 1	Standard Deviation Phase 2	p-value (* = p < 0.05)
TCP	GTV	Control	0.57	0.50	0.62
		Atlas	0.94	0.40	<0.001*
	CTV	Control	15.50	7.28	0.54
		Atlas	18.11	1.23	0.001*
NTCP	Small Intestine	Control	0.42	0.70	0.06
		Atlas	0.76	0.68	0.049*

Table 5

The results of the statistical analysis of the complication-free tumor control probabilities (P_+) are shown for the Atlas-assisted and Control groups (p-value for one-sided Wilcoxon Signed-Rank Test). Here, the TCPs were calculated for every observer's treatment plan and the organ delineation sets of all the observers of the group (atlas or control).

Cohort	Mean P_+ Phase 1	Mean P_+ Phase 2	Mean Difference	SEM	p-value (* = p < 0.05)
Control	80.8	80.5	−0.3	0.14	0.75
Atlas	77.6	80.0	2.4	0.95	<0.001*