

Published in final edited form as:

Int J Radiat Oncol Biol Phys. 2011 February 1; 79(2): 481–489. doi:10.1016/j.ijrobp.2009.11.012.

Prospective randomized double-blind pilot study of site-specific consensus atlas implementation for rectal cancer target volume delineation in the cooperative group setting

Clifton D. Fuller, MD^{1,2,3,*}, Jasper Nijkamp, MSc^{4,*}, Joop Duppen⁴, Coen R.N. Rasch, MD, PhD⁴, Charles R. Thomas Jr, MD³, Samuel J. Wang, MD, PhD³, Paul Okunieff, MD⁵, William E. Jones III, MD¹, Daniel Baseman, MD¹, Shilpen Patel, MD⁶, Carlo G. N. Demandante, MD^{1,7}, Anna M. Harris, MD¹, Benjamin D. Smith, MD^{7,8}, Alan W. Katz, MD, MPH⁵, Camille McGann, MD⁸, Jennifer L. Harper, MD¹⁰, Daniel T. Chang, MD¹¹, Stephen Smalley, MD¹², David T. Marshall, MD¹⁰, Karyn A. Goodman, MD¹³, Niko Papanikolaou, PhD^{1,2}, and Lisa A. Kachnic, MD¹⁴ on behalf of the Radiation Oncology Committee of the Southwest Oncology Group (SWOG)

¹Department of Radiation Oncology The University of Texas Health Science Center at San Antonio, San Antonio, TX, USA. ²Graduate Division of Radiological Sciences/Department of Radiology, The University of Texas Health Science Center at San Antonio, San Antonio, TX, USA. ³Department of Radiation Medicine, Oregon Health & Science University, Portland, OR, USA. ⁴Department of Radiotherapy, The Netherlands Cancer Institute, Antoni van Leeuwenhoek Huis, Amsterdam, The Netherlands. ⁵Department of Radiation Oncology, University of Rochester, Rochester, NY, USA. ⁶Department of Radiation Oncology, University of Washington Medical Center, Seattle, WA, USA. ⁷Radiation Oncology Flight, Wilford Hall Medical Center, Lackland Air Force Base, San Antonio, TX, USA. ⁸Department of Radiation Oncology, MD Anderson Cancer Center, Houston, TX, USA. ⁹Department of Radiation Oncology, Brooke Army Medical Center, San Antonio, TX, USA. ¹⁰Department of Radiation Oncology, Medical University of South Carolina, Charleston, SC, USA. ¹¹Department of Radiation Oncology, Stanford University, Stanford, CA, USA. ¹²Department of Radiation Oncology, University of Kansas, Kansas City, KS, USA; Radiation Oncology Center of Olathe, Olathe, KS, USA. ¹³Department of Radiation Oncology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA. ¹⁴Department of Radiation Oncology, Boston University Medical Center, and Massachusetts General Hospital Cancer Center, Boston, MA, USA.

Abstract

Purpose—Variation in target volume delineation represents a significant hurdle in clinical trials involving conformal radiotherapy. We sought to determine the impact of a consensus guideline-based visual atlas on contouring of target volumes.

© 2009 Elsevier Inc. All rights reserved.

Correspondence to: Clifton D. Fuller, M.D., Department of Radiation Oncology, The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Dr., MSC7889, San Antonio, TX, USA, fullercd@uthscsa.edu, Ph: 210-450-1127, Fax: 210-450-5543. *CDF and JN contributed equally to the execution of this study.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflicts of Interest Notification: The author(s) assert no conflict(s) of interest to declare germane to this effort.

Methods—A representative case and target volume delineation instructions derived from a proposed rectal cancer clinical trial involving conformal radiotherapy were contoured (Scan1) by 14 physician observers and a reference expert. Gross tumor volume (GTV), and 2 clinical target volumes (CTVA, comprising internal iliac, pre-sacral, and peri-rectal nodes, and CTVB, external iliac nodes) were contoured. Observers were randomly assigned to receipt (Group_A) /non-receipt (Group_B) of a consensus guideline and atlas for anorectal cancers, then instructed to re-contour the same case/images (Scan2). Observer variation was analyzed volumetrically using conformation number (CN, where CN=1 equals a total agreement).

Results—In 14 evaluable contour sets (1 expert, 7 Group_A, 6 Group_B), there was greater agreement for GTV (mean CN 0.75) than CTVs (mean CN 0.46–0.65). Atlas exposure for Group_A led to a significant increased inter-observer agreement for CTVA (mean initial CN 0.68, post-atlas 0.76; $p=0.03$), as well as increased agreement with the expert reference (initial mean CN 0.58, 0.69 post-atlas; $p=0.02$). For GTV and CTVB, neither inter-observer nor expert agreement was altered after atlas exposure.

Conclusion—Consensus guideline atlas implementation resulted in a detectable difference in inter-observer agreement and greater approximation of expert volumes for CTVA, but not GTV or CTVB, in the specified case. Visual atlas inclusion should be considered as a feature in future clinical trials incorporating conformal radiotherapy.

Keywords

Target volume delineation; cooperative group; conformal radiotherapy; atlas; consensus guideline

Introduction

Inter-observer differences in target volume delineation are a demonstrated source of potential treatment variability in the context of clinical trials that incorporate conformal radiotherapy approaches^{1, 2}. Recent publications suggest that target delineation consensus documentation is highly desirable for clinical trials³ and that specific instructional or educational interventions may afford a measurable effect in terms of physician contouring^{4, 5}.

As part of efforts to improve radiotherapy implementation for Southwest Oncology Group trials, and consistent with its focus on quality improvement in cooperative studies, the SWOG Radiation Oncology Committee authorized this study as a pilot project to achieve the following primary specific aims:

1. Feasibility of centralized target volume delineation evaluation as a pre-trial adjunct to a SWOG-sponsored study (SWOG S0713), and
2. Determination of the impact of implementation of a consensus anatomical atlas on target volume variability.

Methods

This prospective IRB-exempt study was conducted under the auspices of the University of Texas Health Science Center San Antonio Institutional Review Board. This study was designed as a double-blind, randomized hypothesis generating pilot study (Figure 1). Statistical power for agreement analysis was estimated for a non-Bonferroni-corrected paired-measures Wilcoxon test (assuming a minimum asymptotic relative efficiency of ≥ 0.863 compared with a paired t-test), with a specified $1-\beta$ of 0.7, and α of ≤ 0.05 , resulting in a minimum requisite sample size of 6 observers (radiation oncologists) per group,

calculated using G*Power 3 statistical software⁶. Goal enrollment was 10–12 observers per cohort.

Participant radiation oncologists (observers) were recruited from SWOG-participating institutions. Those who indicated interest were sent study documentation which included a standardized case report, description of target volumes to be contoured, and a compact disc (CD) containing 3 mm axial CT images derived from the DICOM file of the standardized case study's simulation CT, to be contoured twice using the Big Brother target delineation software program. "Big Brother" is a custom target volume delineation evaluation software platform developed at the Netherlands Cancer Institute^{7, 8}. Big Brother consists of a user interface with target delineation features common to most commercial treatment planning systems⁹ and collects a wide array of volumetric and target delineation data unobtrusively during the contouring session. The included case study depicted the history and clinical findings from an anonymized patient with T3N0M0 adenocarcinoma of the rectum with instructions modeled on a SWOG protocol in development at the time which included detailed instructions regarding 3-DCRT and IMRT treatment plan design (SWOG S0713: *A Phase II Study of Oxaliplatin, Capecitabine, Cetuximab and Radiation in Pre-operative Therapy of Rectal Cancer*, ClinicalTrials.gov Identifier NCT00686166), with terminology modified to fit nomenclature established in the then-unpublished RTOG consensus guideline for target delineation in anorectal cancers¹⁰. Observers were asked to contour the structures as per Table 1. Axial CT images were extracted using a single DICOM dataset; identical copies of reconstructed (axial, sagittal, and coronal views) were then designated Scan1 and Scan2.

Half of distributed CDs contained an automated HTML link, which, after submission of the first contouring session (Scan1) and the subsequent electronic survey, directed users to a pre-publication version of the RTOG consensus guidelines for target volumes in anorectal cancer¹⁰ as well as instructions to re-contour the exact same axial CT images a second time (Scan2) with the same case presentation, instructions and target definitions, using the RTOG consensus guideline visual atlas as a guide (Group_A). All other CDs contained HTML pop-up directions to re-contour the same volumes on the identical CT-simulation-derived dataset (Scan2), using the same aforementioned case data/instructions as previous (Group_B). Hence, Group_B did *not* receive consensus atlas guidance for re-contouring the case. CDs with and without the HTML link to the consensus atlas were randomly shuffled before labeling and delivery to participants; both study personnel and physician observers were unaware of which CD had been distributed to each participant until electronic survey completion.

After completion of the GTV and CTV delineation on Scan1, observers submitted the case by email and were directed to an electronic survey (Table 2). Subsequently, participants were provided with instructions to re-contour the case either with or without the assistance of an anatomically specific consensus atlas. The recently published RTOG consensus atlas¹⁰ was utilized in pre-publication form (http://www.rtog.org/pdf_document/AnorectalContouringGuidelines.pdf).

In addition, one of the members involved in the development of the RTOG consensus guideline was asked to delineate Scan1 and Scan2. This observer [LK] was designated as a "reference expert", with her contours serving as a *de facto* gold standard against which to compare observer-derived contours. During the study period, only the reference expert user had a prior knowledge of this atlas, and thus study participants represented a *tabula rasa* with regard to consensus guidelines.

Delineation agreement analysis

All delineations were first visually analyzed (Figure 2) and protocol deviations from the delivered instructions were identified by review of all axial contours. Total volume encompassed in cubic centimeters for all structures were calculated and tabulated. Statistical comparison of volume differentials between Scan1 and Scan2 were performed for each structure for Group_A and B respectively.

Baseline inter-observer variation for the SWOG protocol was derived from the delineations on Scan1 from all observers, excepting reference expert. Baseline intra-observer variation was derived from a comparison of the volume of Scan1 and Scan2 in Group_B. The effect of the atlas on inter-observer variation was quantified by comparing the inter-observer variation for Scan1 and Scan2 in Group_A. For comparison within cohorts, a composite median delineation was calculated for each group. The median delineation represents a 50% coverage isosurface of the observers, such that each voxel inside is designated by at least 50% of the observers, and was calculated for GTV, CTVA and CTVB. The CTVC structure was not designated in the instructions as a necessary volume be contoured for this clinical case, and was therefore not analyzed. For volumetric agreement analysis for Group_A, first, the common volume (CV) was calculated between either the median or expert contour (V1) and the observer contour (V2). Subsequently, as a modification of the concept introduced by van't Riet *et al.*^{11, 12}, a conformation number (CN) was derived as $CN = CV^2/(V1*V2)$. Differences in CN values for target structures (e.g. GTV, CTVA, CTVB) for Scan1 and Scan2 for Group_A, using both reference expert and group median delineation isosurface as a comparator, were calculated and formally assessed for statistical significance by paired-measures Wilcoxon test.

For Group_B, intra-observer CN values were calculated using the aforementioned van't Riet formula¹²; the common volume (CV) was calculated between either the Scan1 (V1) and Scan2 (V2).

Post-hoc exploratory contour surface variability analysis

After completion of planned volumetric analysis, surface distance analysis was performed to identify regional delineation variation within the CTVA volume, using virtual volume unfolding as previously presented elsewhere^{13, 14}. Briefly, for the surface distance variation calculation, the reference structure (median or expert) was first re-sampled to 100 equidistant points per delineated slice. Secondly, for each point on the reference structure, the distance, perpendicular to the surface, to the observer derived contour was calculated. For the observer variation analysis, the standard deviation (SD) of all observers was calculated for each point on the reference structure. For comparison with the expert, the group median was calculated by taking the median of the distances for each point on the expert derived reference structure to the perpendicular surface of every observer-derived contour. Regional differentials in surface variation were then explored graphically and numerically.

Results

Eight SWOG institutions had at least one user submitting contours, as well as a single non-SWOG affiliated participant. Of the twenty-six observers directly asked to participate, 15 submitted contour set pairs, of which 14 were technically evaluable (1 expert, 7 Group_A, 6 Group_B). The non-evaluable contour set consisted of non-connected, non-overlapping contours, which precluded ready analysis with the cohort at large. Survey results were pooled and are tabulated in Table 2.

All 14 remaining observers delineated the GTV and CTVA on Scan1 and Scan2. While the CTVB was mandatory in the specific delineation instructions, it was only delineated by 10 of the 13 observers. The CTVC, which should not have been delineated, was contoured by two observers both on Scan1 and Scan2, by one observer on Scan1 only, and by one other observer on Scan2 only. For one observer of Group_A and four observers of Group_B, major deviations from the delineation protocol (e.g. GTV was not encompassed by the CTVA) were visible on axial slice review. For one other observer of Group_A, the CTVB covered the internal iliac vessels instead of the externals on both ScanA and ScanB. For the 5 observers where the CTVA did not fully cover the GTV, the CTVA was manually edited so that the observer-contoured GTV was encompassed for the volume analysis; preliminary statistical evaluation evidenced minimal alteration of volumetric statistics by this modification.

Between Scan1 and Scan2, only the increase in volume of the CTVB in Group_A approached statistical significance ($p=0.06$) (Table 3). In Group_B, the number of CTVB slices covered by all observers dropped from 14 to 3 axial CT slices, while the average delineated number of axial CT slices contoured only dropped from 20 to 16 slices. The median GTV delineated on Scan1 for all observers had a volume of 74 cc. The average CN for baseline inter-observer variation of the GTV was 0.75 (range 0.60–0.81). The median CTVA had a volume of 709 cc and a CN of 0.65 (range 0.47–0.75), which indicates comparatively greater inter-observer disagreement for CTVA compared to the GTV. For CTVB, with a median volume of 70 cc, even less inter-observer agreement could be found, with an average CN of 0.46 (range 0.24–0.70).

Atlas exposure led to a statistically significant increase in volumetric agreement on CTVA between observers (Figure 3a) and with the expert (Figure 3b), as measured by CN. The average inter-observer CN (i.e. agreement with the median surface) increased from 0.68 (range 0.41–0.78) on Scan1 to 0.76 (range 0.57–0.87) on Scan2 ($p=0.031$, paired Wilcoxon signed rank test; Figure 3a). The average CN, compared with the expert, increased from 0.58 pre-atlas (range 0.42–0.70) to 0.69 post-atlas (0.58–0.78, $p=0.016$; Figure 3b). For the CTVB, however, neither inter-observer variation (mean CN 0.39 (range 0.26–0.67) vs. 0.45 (range 0.13–0.68, $p=0.4$), nor agreement with the expert (mean CN 0.31 (range 0.16–0.49) vs. 0.30 (range 0.11–0.44)) was altered to a statistically significant degree after atlas exposure ($p=0.8$).

Since the atlas only affected observer variation for CTVA, exploratory *post-hoc* surface distance variation analysis was limited to CTVA only (Figure 4 and 5). To translate surface maps into numbers, first the reference structure (median/expert CTVA) was divided into anterior, lateral and posterior regions, and subdivided into upper- and lower sub-regions at the level of the tip of the coccyx. For each of the 6 regions, the SD value covering 5%–95% of regional surface distance difference was taken to characterize the minimum and maximum regional variation (Table 4), though no formal statistical comparison of regional sub-volumes was performed. Visual inspection (Figure 4 and Figure 5) showed introduction of the atlas resulted in modification of surface distance between observers and the expert primarily in limited regions of the CTVA, rather than the CTVA volume globally. Modification of target volumes was most notably localized to the upper-anterior region adjacent to the bladder, lower-posterior and lateral CTVA (not shown); however, statistical significance was not formally assessed. For all defined regions, except upper-posterior and upper-lateral, the upper 95% CI of the inter-observer surface standard deviation was reduced by 0.2–0.8 cm after the introduction of the atlas. As Table 4 demonstrates, >1 cm of surface variation was observed for multiple regions before atlas implementation for all users, and though reduced after atlas administration, >1cm was still needed to cover 95% of the surface variation in CTV sub-regions.

Regarding intra-observer variation, the absolute volume of all respective structures contoured was essentially equivalent (Table 3). Comparison between delineations on Scan1 and Scan2 in Group_B yielded an average CN of 0.80 (range 0.75–0.82), 0.68 (range 0.47–0.89) and 0.54 (range 0.16–0.72) for the GTV, CTVA and CTVB respectively. Regional intra-observer variability is illustrated graphically in Figure 5.

Discussion

Despite the well-known consequences of geometric inaccuracy in target volume delineation^{15–17}, inter-observer variability in target definition has been demonstrated in a host of studies, in various anatomical sites¹⁸. Simply put, “inter-observer variability in the definition of GTV and CTV is a major – for some tumor locations probably the largest – factor contributing to the global uncertainty in radiation treatment planning”¹⁸. Consequently, there are continuing efforts to implement solutions to possible sources of variability/error in the target volume delineation process. These solutions have included optimization of imaging inputs^{19–22}, instructional protocol modification^{5, 23, 24}, integration of specific training programs^{25, 26}, development of software tools^{27–31}, and implementation of standardized guidelines^{32–37} for distinct anatomic sub-sites. For clinical trials, the situation is potentially more vexing, as insurance of adequate treatment uniformity between comparison cohorts necessitates comparatively increased attention to both protocol construction and enrollee plan review, costing significant time in terms of resources for the primary investigator(s).

In terms of feasibility, the study was readily completed (total study duration of 5 months). A total of 12/26 (46%) invited SWOG institutions confirmed intent to participate; however, only 8/26 (31%) had resultant submissions. Nonetheless, our findings suggest that a reasonably powered target delineation trial might be implemented with a modicum of cooperative group resource allocation in timely manner, and that such a study is both technically and logistically feasible. Analysis of resultant data alludes to the difficulty of executing clinical trials in the conformal radiotherapy era. The high proportion of major protocol deviations was consistent with previous reports. The substantial variation from expert reference and median contour surfaces observed for all users pre-intervention (Figure 4 and 5, Table 4) suggests that efforts to further minimize inter-observer variability are imperative. As Table 4 demonstrates, substantial inter-observer surface deviation was observed for multiple CTVA sub-regions before atlas implementation. After atlas administration, a reduction of 0.3, 0.6 and 0.8 cm was achieved for the upper-anterior, lower-lateral and lower-posterior CTVA sub-region upper limit of standard deviation from the median isosurface. Although >1 cm would *still* be needed to cover 95% of all contouring variability, the achieved reduction(s) would result in a decrement in required PTV expansion margins. However, further reduction of variation is desired, because the PTV margins required to encompass the residual variation in target delineation would limit the practical advantages of IMRT over conventional RT.

Several limitations to this pilot study are evident. The sample size of this study is limited, and only a single case was contoured. Utilization of a reference expert’s contours as a *de facto* “gold standard” points to the fact that the “ground truth” in contouring clinical target volumes remains ambiguous (Table 3, noting variation within the reference expert user’s sequential contours). Some variance in the study might be attributable to instructions which were distinct from standard clinical practice (e.g. the external iliacs are not typically contoured for T3 rectal cases). Our invitation was limited to SWOG institutes, creating potential sampling bias, and the fact that only interested observers participated creates an avenue for selection bias. Nonetheless, our data suggest that inclusion of a visual atlas in addition to written instructions can improve conformance to a reference expert’s contours

(Figure 3a), as well as reducing inter-observer variability, to a statistically detectable degree (Figure 3b). However, our data also suggest substantial residual variability in rectal target volume delineation, even after atlas utilization (Table 3, 4).

This study is consistent with previous investigations of educational interventions and consensus guideline application in contouring studies. Recently, Bekelman et al.²⁵ demonstrated improvement in contour quality after a directed teaching intervention, echoing previous work by Tai et al. showing increased protocol compliance after a site-specific educational experience²⁶. With regard to consensus guideline application as an avenue towards target variability reduction, Dimopolous et al.³² detail a study in which 19 cervical cancer cases were contoured using GEC-ESTRO guidelines by two observers, with a resultant between-user conformity index (CI) 11 ± 21 in the range 0.6–07 for target volumes, roughly consistent with CIs in the current series. Likewise, Wong et al. recently demonstrated, using a test-retest sequence, that improved consistency in seroma contouring could be observed after exposure to consensus guidelines³⁸. In the clinical trial setting, it is likely that “trial specific” atlases should be used, based on patterns of failure data (as per Roels et al.³⁹) or, possibly, after a pilot contouring trial similar to the present study. For instance, RTOG anorectal consensus guidelines stipulate coverage “extending CTVA ~1 cm into the posterior bladder, to account for day-to-day variation in bladder position¹⁰.” This incorporation of motion into CTV generation, rather than PTV expansion, represents a conceptual break with ICRU 6240 and other guidelines³⁹, wherein the posterior bladder wall would not be contoured. No users in Group_A included significant portions of posterior bladder pre-atlas, whereas a majority did so after atlas exposure (in compliance with the presented atlas¹⁰ and consistent with the reference expert).

Future studies will be required to ascertain if observed effects of atlas administration are transferable to other anatomic sites with potentially more complicated anatomic relationships^{5, 24}. The SWOG Radiation Therapy Committee intends to suggest building target delineation studies into clinical trial protocol development/quality assurance processes. Aspects of this dataset may also be integrated into design of educational materials for a proposed Dutch cooperative group rectal study workshop. We plan to use portions of this dataset to construct composite models accounting for rectal motion and set-up variability^{14, 41}, as well as development of novel software strategies for evaluation⁴² and minimization of target delineation variance.

Conclusion

The addition of a visual atlas and consensus treatment guidelines to a written protocol increased clinical target volume delineation conformance with expert derived contours and increased contour agreement between participants. The addition of a visual atlas and consensus treatment guidelines to a written protocol increased clinical target volume delineation conformance with expert derived contours and increased contour agreement between participants for CTVA, but not GTV or CTVB, for the included rectal cancer case. Detected inter-observer (both with and without atlas) and intra-observer variation in contouring target structures was substantial. Visual atlas-based supplementary target volume specification materials should be considered for clinical trials involving conformal radiotherapy approaches.

Acknowledgments

Funding acknowledgement(s)

C.D.F. is supported by a training grant from the National Institutes of Health/National Institute of Biomedical Imaging and Bioengineering, “Multidisciplinary Training Program in Human Imaging” (5T32EB000817-04), a

Technology Transfer Grant from the European Society for Therapeutic Radiology Oncology, and the Product Support Development Grant from the Society for Imaging Informatics in Medicine. The funder(s) played no role in study design, in the collection, analysis and interpretation of data, in the writing of the manuscript, nor in the decision to submit the manuscript for publication. Portions of this data have been selected for a Poster Recognition Award at the 2009 American Society for Radiation Oncology Annual Meeting, November 1–5, 2009, Chicago, IL.

References

1. Foppiano F, Fiorino C, Frezza G, et al. The impact of contouring uncertainty on rectal 3D dose-volume data: results of a dummy run in a multicenter trial (AIROPROS01-02). *Int J Radiat Oncol Biol Phys* 2003;57:573–579. [PubMed: 12957271]
2. Yamazaki H, Nishiyama K, Tanaka E, et al. Dummy run for a phase II multi-institute trial of chemoradiotherapy for unresectable pancreatic cancer: inter-observer variance in contour delineation. *Anticancer Res* 2007;27:2965–2971. [PubMed: 17695480]
3. Li XA, Tai A, Arthur DW, et al. Variability of target and normal structure delineation for breast cancer radiotherapy: an RTOG Multi-Institutional and Multiobserver Study. *Int J Radiat Oncol Biol Phys* 2009;73:944–951. [PubMed: 19215827]
4. Mitchell DM, Perry L, Smith S, et al. Assessing the effect of a contouring protocol on Postprostatectomy Radiotherapy Clinical Target Volumes and Interphysician Variation. *Int J Radiat Oncol Biol Phys*. 2009 Article in Press.
5. Senan S, van Sörnsen de Koste J, Samson M, et al. Evaluation of a target contouring protocol for 3D conformal radiotherapy in non-small cell lung cancer. *Radiother Oncol* 1999;53:247–255. [PubMed: 10660205]
6. Faul F, Erdfelder E, Lang AG, et al. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007;39:175–191. [PubMed: 17695343]
7. Steenbakkens R, Duppen J, Fitton I, et al. Observer variation in target volume delineation of lung cancer related to radiation oncologist–computer interaction: A ‘Big Brother’ evaluation. *Radiother Oncol* 2005;77:182–190. [PubMed: 16256231]
8. Steenbakkens RJ, Duppen JC, Fitton I, et al. Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis. *Int J Radiat Oncol Biol Phys* 2006;64:435–448. [PubMed: 16198064]
9. Duppen JC, Steenbakkens RJHM, Fitton I, et al. Delineation software for teaching in target volume definition. *Radiother Oncol* 2004;73:S437–S438.
10. Myerson RJ, Garofalo MC, Naqa IE, et al. Elective Clinical Target Volumes for Conformal Therapy in Anorectal Cancer: An Radiation Therapy Oncology Group Consensus Panel Contouring Atlas. *Int J Radiat Oncol Biol Phys* 2008;74:824–830. [PubMed: 19117696]
11. Feuvret L, Noel G, Mazon JJ, et al. Conformity index: a review. *Int J Radiat Oncol Biol Phys* 2006;64:333–342. [PubMed: 16414369]
12. van't Riet A, Mak AC, Moerland MA, et al. A conformation number to quantify the degree of conformality in brachytherapy and external beam irradiation: application to the prostate. *Int J Radiat Oncol Biol Phys* 1997;37:731–736. [PubMed: 9112473]
13. Hoogeman M, van Herk M, de Bois J, et al. Quantification of local rectal wall displacements by virtual rectum unfolding. *Radiother Oncol* 2004;70:21–30. [PubMed: 15036848]
14. Nijkamp J, de Jong R, Sonke JJ, et al. Target volume shape variation during hypo-fractionated preoperative irradiation of rectal cancer patients. *Radiother Oncol* 2009;92:202–209. [PubMed: 19454365]
15. Jeanneret-Sozzi W, Moeckli R, Valley J, et al. The Reasons for Discrepancies in Target Volume Delineation. *Strahlenther Onkol* 2006;182:450–457. [PubMed: 16896591]
16. Logue JP, Sharrock CL, Cowan RA, et al. Clinical variability of target volume description in conformal radiotherapy planning. *Int J Radiat Oncol Biol Phys* 1998;41:929–931. [PubMed: 9652859]
17. Njeh CF. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. *J Med Phys* 2008;33:136–140. [PubMed: 19893706]

18. Weiss E, Hess C. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences. *Strahlenther Onkol* 2003;179:21–30. [PubMed: 12540981]
19. Bassi MC, Turri L, Sacchetti G, et al. FDG-PET/CT imaging for staging and target volume delineation in preoperative conformal radiotherapy of rectal cancer. *Int J Radiat Oncol Biol Phys* 2008;70:1423–1426. [PubMed: 17931795]
20. Cattaneo G, Reni M, Rizzo G, et al. Target delineation in post-operative radiotherapy of brain gliomas: Interobserver variability and impact of image registration of MR(pre-operative) images on treatment planning CT scans. *Radiother Oncol* 2005;75:217–223. [PubMed: 15885831]
21. Rasch C, Barillot I, Remeijer P, et al. Definition of the prostate in CT and MRI: a multi-observer study. *Int J Radiat Oncol Biol Phys* 1999;43:57–66. [PubMed: 9989514]
22. Rasch C, Keus R, Pameijer FA, et al. The potential impact of CT-MRI matching on tumor volume delineation in advanced head and neck cancer. *Int J Radiat Oncol Biol Phys* 1997;39:841–848. [PubMed: 9369132]
23. Coles C. Quantitative assessment of inter-clinician variability of target volume delineation for medulloblastoma: quality assurance for the SIOP PNET 4 trial protocol. *Radiother Oncol* 2003;69:189–194.
24. Vorwerk H, Beckmann G, Bremer M, et al. The delineation of target volumes for radiotherapy of lung cancer patients. *Radiother Oncol* 2009;91:455–460. [PubMed: 19339069]
25. Bekelman JE, Wolden S, Lee N. Head-and-neck target delineation among radiation oncology residents after a teaching intervention: a prospective, blinded pilot study. *Int J Radiat Oncol Biol Phys* 2009;73:416–423. [PubMed: 18538494]
26. Tai P, Van Dyk J, Battista J, et al. Improving the consistency in cervical esophageal target volume definition by special training. *Int J Radiat Oncol Biol Phys* 2002;53:766–774. [PubMed: 12062623]
27. Chao KS, Bhide S, Chen H, et al. Reduce in variation and improve efficiency of target volume delineation by a computer-assisted system using a deformable image registration approach. *Int J Radiat Oncol Biol Phys* 2007;68:1512–1521. [PubMed: 17674982]
28. Ciernik IF, Huser M, Burger C, et al. Automated functional image-guided radiation treatment planning for rectal cancer. *Int J Radiat Oncol Biol Phys* 2005;62:893–900. [PubMed: 15936575]
29. Pasquier D, Lacornerie T, Vermandel M, et al. Automatic segmentation of pelvic structures from magnetic resonance images for prostate cancer radiotherapy. *Int J Radiat Oncol Biol Phys* 2007;68:592–600. [PubMed: 17498571]
30. van Sörnsen de Koste JR, Senan S, Underberg RW, et al. Use of CD-ROM-based tool for analyzing contouring variations in involved-field radiotherapy for Stage III NSCLC. *Int J Radiat Oncol Biol Phys* 2005;63:334–339. [PubMed: 16168828]
31. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 2006;31:1116–1128. [PubMed: 16545965]
32. Dimopoulos JC, Vos VD, Berger D, et al. Inter-observer comparison of target delineation for MRI-assisted cervical cancer brachytherapy: Application of the GYN GEC-ESTRO recommendations. *Radiother Oncol*. 2008
33. Gregoire V. CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC, RTOG consensus guidelines. *Radiother Oncol* 2003;69:227–236. [PubMed: 14644481]
34. Lawton C, Michalski J, Elnaqa I, et al. RTOG GU Radiation Oncology Specialists Reach Consensus on Pelvic Lymph Node Volumes for High-Risk Prostate Cancer. *Int J Radiat Oncol Biol Phys* 2008;74:383–387. [PubMed: 18947938]
35. Poortmans P, Bossi A, Vandeputte K, et al. Guidelines for target volume definition in post-operative radiotherapy for prostate cancer, on behalf of the EORTC Radiation Oncology Group. *Radiother Oncol* 2007;84:121–127. [PubMed: 17706307]
36. Sidhom MA, Kneebone AB, Lehman M, et al. Post-prostatectomy radiation therapy: consensus guidelines of the Australian and New Zealand Radiation Oncology Genito-Urinary Group. *Radiother Oncol* 2008;88:10–19. [PubMed: 18514340]

37. Small W, Mell LK, Anderson P, et al. Consensus guidelines for delineation of clinical target volume for intensity-modulated pelvic radiotherapy in postoperative treatment of endometrial and cervical cancer. *Int J Radiat Oncol Biol Phys* 2008;71:428–434. [PubMed: 18037584]
38. Wong EK, Truong PT, Kader HA, et al. Consistency in seroma contouring for partial breast radiotherapy: impact of guidelines. *Int J Radiat Oncol Biol Phys* 2006;66:372–376. [PubMed: 16965989]
39. Roels S, Duthoy W, Haustermans K, et al. Definition and delineation of the clinical target volume for rectal cancer. *Int J Radiat Oncol Biol Phys* 2006;65:1129–1142. [PubMed: 16750329]
40. International Commission on Radiation Units and Measurements. ICRU Report 62: Prescribing, recording, and reporting photon beam therapy. Bethesda, Md: International Commission on Radiation Units and Measurements; 1999.
41. Ippolito E, Mertens I, Haustermans K, et al. IGRT in rectal cancer. *Acta Oncol* 2008;47:1317–1324. [PubMed: 18661433]
42. Jena R, Kirkby NF, Burton KE, et al. A novel algorithm for the morphometric assessment of radiotherapy treatment planning volumes. *Br J Radiol*. 2009

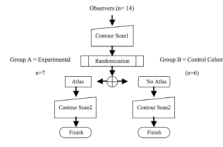


Figure 1.
Study design.

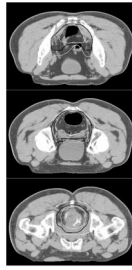


Figure 2. Overview of the largest and smallest CTV_A delineations (white) and the median CTV_A delineation (black) from all observers on Scan_1.

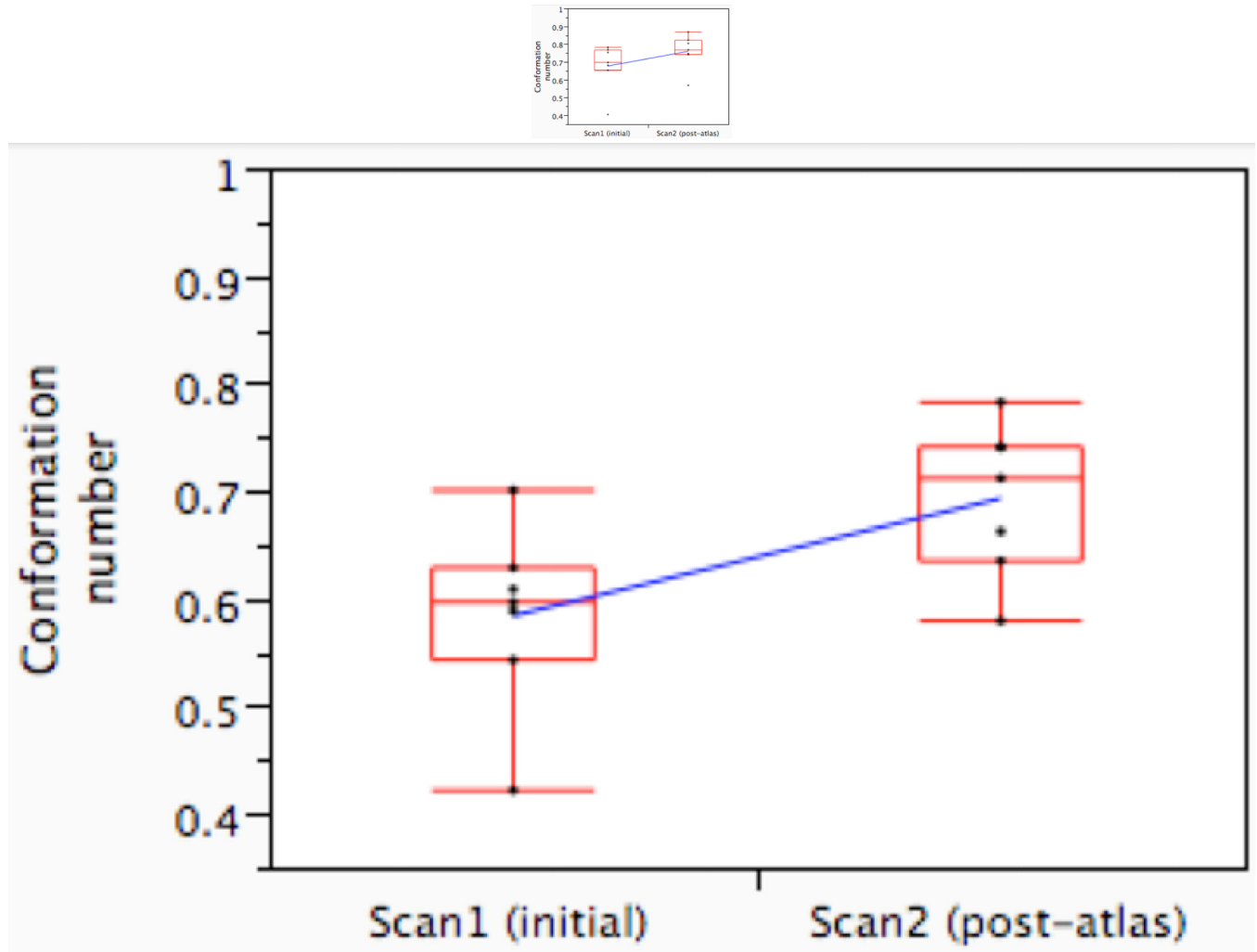


Figure 3. Conformation number (CN) for the observers in Group_A before (Scan1) and after the introduction of the atlas (Scan2), compared to the median surface (Figure 3a) and compared to the expert delineation (Figure 3b); CN=1 indicates complete volumetric agreement. The boxplot displays the interquartile range (i.e. the 25–75th percentiles), with a horizontal line demonstrating the group median; whiskers include all points within ± 1.5 times the interquartile range from the 25th and 75th percentiles. Pre- and post-atlas mean CNs are connected by a diagonal blue line.

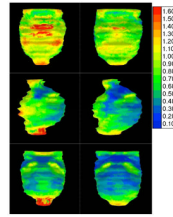


Figure 4.

Inter-observer variation changes with the introduction of the atlas. On the left the inter-observer variation (SD) in Group_A before introduction of the atlas is shown in an anterior, sagittal and posterior view. On the right the inter-observer variation in the same group is shown after introduction of the atlas. Group standard deviation of distance to the expert surface is shown as a color scale to the right side of figure (values in cm from expert surface).

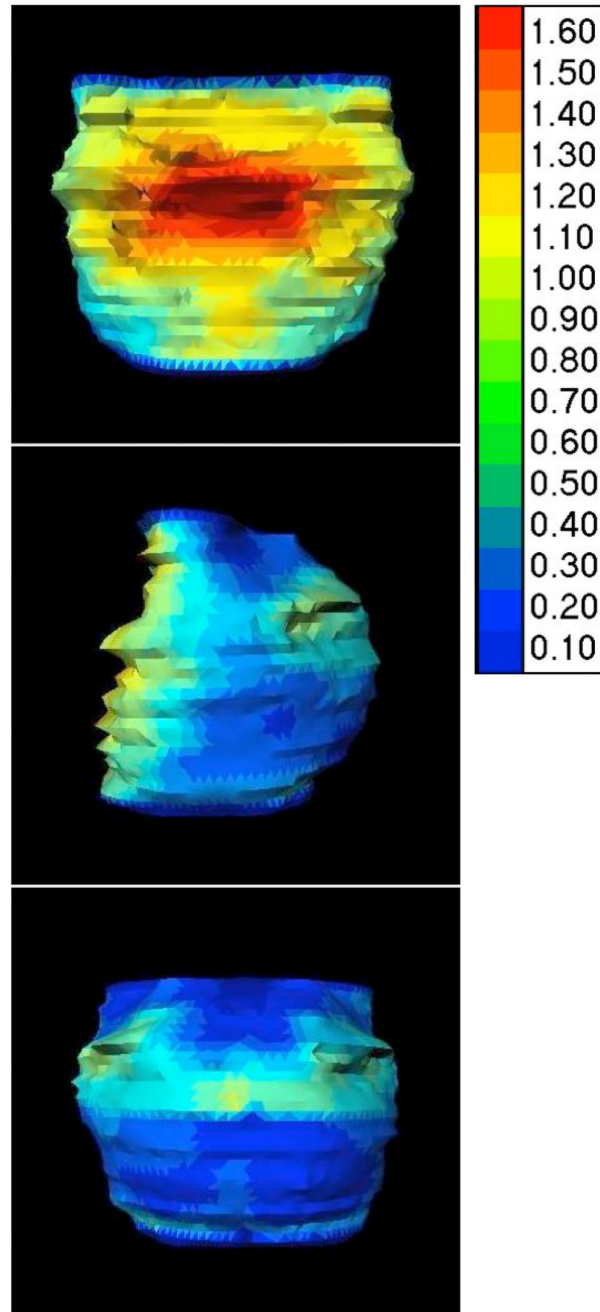


Figure 5. Intra-observer variation. The intra-observer variation (SD) is shown in an anterior, sagittal and posterior view for the observers who delineated the case twice without any changes in between (Group_B). Group standard deviation of distance between equivalent points for Scan1 and Scan2 for each user are shown as a color scale to the right side of figure (in cm).

Table 1

Target volume definitions and instructions.

Structure	Definition	Specific instructions
GTV	<ul style="list-style-type: none"> Includes the primary tumor and any pelvic node felt involved grossly with metastatic disease. Assessment of grossly involved nodal disease may be made on the basis of CT for this study. 	
CTV	<ul style="list-style-type: none"> Consists of CTVA, CTVB, and CTVC. Includes the GTV and the following nodal groups: perirectal nodes; presacral nodes; internal iliac; and common iliac nodes below the L5-sacral junction. 	<ul style="list-style-type: none"> “Contour the structures denoted GTV, CTVA, CTVB, and (optionally) CTVC, based on the preceding target volume definitions.” “Do NOT add ANY margin to account for positional variability. It will be assumed that a standardized planning target volume (PTV) expansion will be applied of 0.7–1.0 cm omnidirectionally using the GTV/CTV structures that you generate.”
CTVA	<ul style="list-style-type: none"> For this study, defined to be the nodal regions that would regularly be treated for rectal cancers: internal iliac, presacral, and peri-rectal. 	
CTVB	<ul style="list-style-type: none"> For this study, defined to be the external iliac nodal region. 	
CTVC	<ul style="list-style-type: none"> For this study, defined to be inguinal nodal region. 	

Table 2

Selected survey query results.

Query	Response/Parameter	All observers* n=13	Group_A (atlas exposure) n=7	Group_B (controls) n=6
“Do you have specific training or expertise in this anatomic site?”	“Yes”	5 (38)	3 (43)	2 (33)
	“No”	8 (62)	4 (57)	4 (67)
“Please estimate the amount of time you spent on the contouring task alone for this session (minutes)”	Mean±SD (Range)	34±12 (18–65)	30±11 (20–65)	34±13 (18–60)
“How confident are you in contouring in this anatomical region?”	“Very confident/Expert”	1 (8)	1 (14)	0 (0)
	“Somewhat confident/Intermediate”	10 (77)	5 (71)	5 (83)
	“Not confident/Novice”	2 (15)	1 (14)	1 (17)
“Please indicate the relative difficulty of the contouring session”	1-Easy	0 (0)	0 (0)	0 (0)
	2-Somewhat easy	2 (15)	1 (14)	1 (17)
	3-Average	8 (62)	5 (71)	3 (50)
	4-Somewhat difficult	3 (23)	1 (14)	2
	5-Difficult	0 (0)	0 (0)	0 (0)
“How clear were the directions for contouring this case?”	“Unclear”	1 (8)	1 (14)	0 (0)
	“Somewhat clear”	2 (15)	0 (0)	2 (33)
	“Clear”	5 (39)	4 (57)	1 (17)
	“Very clear and understood”	5 (39)	2 (28)	3 (50)
“Please indicate how confident you are that the contours you created for the submitted case are correct.”	“Not confident”	0 (0)	0 (0)	0 (0)
	“Somewhat confident”	7 (54)	4 (57)	3 (50)
	“Confident”	6 (46)	3 (43)	3 (50)
“Did you find the atlas was a useful aid to your contouring of the case?”	“Very confident”	0 (0)	0 (0)	0 (0)
	“Not helpful”	-	0 (0)	-
	“Somewhat helpful”	-	4 (57)	-
	“Very helpful”	-	3 (43)	-

* A single user’s contours were not technically amenable for analysis; responses from this user were discarded from survey results.

Table 3

Selected volumetric and axial slice measures.

Parameter	Structure	Reference Expert			Group_A (Atlas)			Group_B (Control)		
		Scan 1	Scan 2	Scan 1	Scan 2	p-value [†]	Scan 1	Scan 2	p-value [†]	
Mean volume (cm ³)*	GTV	68	78	78 [15.2]	78 [15.3]	1.0	68 [9.7]	68 [6.9]	0.9	
	CTVA	784	820	800 [276]	809 [172]	0.7	590 [208]	642 [251]	0.4	
	CTVB	67	91	77 [62]	100 [78]	0.06	71 [26]	51 [32]	0.3	
Mean delineated length* (number of axial CT slices)	GTV	12	12	19 [8.3]	19 [8.3]	-	16 [7.4]	16 [7.4]	-	
	CTVA	45	43	44 [6.8]	41 [5.9]	-	40 [6.3]	39 [6.2]	-	
	CTVB	19	21	13 [7.5]	15 [8.5]	-	20 [4.7]	16 [7.7]	-	
No. axial CT slices covered by all observers	GTV	12	12	1	12	-	12	12	-	
	CTVA	45	43	32	31	-	31	33	-	
	CTVB	19	21	9	11	-	14	3	-	

* standard deviation [SD] in brackets;

[†] using paired Wilcoxon rank test to compare absolute volume of Scan 1 and Scan 2.

Table 4

CTVA inter-observer surface distance, expressed as the 95% confidence interval of the group standard deviation from the median group isosurface, by regional subdivision, before and after atlas exposure.

		Anterior	Posterior	Lateral
<i>All observers Scan 1 (e.g. initial total inter-observer variation)</i>	Upper	0.6–1.4 cm	0.2–0.7 cm	0.3–1.1 cm
	Lower	0.6–1.2 cm	0.3–1.2 cm	0.4–1.1 cm
<i>Group A Scan 1 (e.g. initial inter-observer variation)</i>	Upper	0.5–1.4 cm	0.2–0.7 cm	0.2–1.0 cm
	Lower	0.7–1.3 cm	0.4–1.4 cm	0.5–1.4 cm
<i>Group A Scan 2 (e.g. post-atlas inter-observer variation)</i>	Upper	0.5–1.1 cm	0.1–0.7 cm	0.2–0.9 cm
	Lower	0.6–1.1 cm	0.2–0.6 cm	0.3–0.8 cm