Int. J. Radiation Oncology Biol. Phys., Vol. ■, No. ■, pp. 1–9, 2010

0360-3016/10/$–see front matter

ELSEVIER

doi:10.1016/j.ijrobp.2009.11.012

**CLINICAL INVESTIGATION**

# 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, M.D.,[1]*† Jasper Nijkamp, M.Sc.,[1§] Joop C. Duppen, Ing.,[§] Coen R. N. Rasch, M.D., Ph.D.,[§] Charles R. Thomas, Jr., M.D.,[‡] Samuel J. Wang, M.D., Ph.D.,[‡] Paul Okunieff, M.D.,[¶] William E. Jones, III, M.D.,* Daniel Baseman, M.D.,* Shilpen Patel, M.D.,[||] Carlo G. N. Demandante, M.D.,*[#] Anna M. Harris, M.D.,* Benjamin D. Smith, M.D.,[†]** Alan W. Katz, M.D., M.P.H.,[¶] Camille McGann, M.D.,[††] Jennifer L. Harper, MD,[‡‡] Daniel T. Chang, M.D.,[§§] Stephen Smalley, M.D.,[¶¶] David T. Marshall, M.D.,[‡‡] Karyn A. Goodman, M.D.,[||||] Niko Papanikolaou, Ph.D.,*† and Lisa A. Kachnic, M.D. [##] on behalf of the Radiation Oncology Committee of the Southwest Oncology Group

*Department of Radiation Oncology and †Graduate Division of Radiological Sciences, Department of Radiology, University of Texas Health Science Center at San Antonio, San Antonio, TX; ‡Department of Radiation Medicine, Oregon Health & Science University, Portland, OR; §Department of Radiotherapy, The Netherlands Cancer Institute, Antoni van Leeuwenhoek Huis, Amsterdam, The Netherlands; ¶Department of Radiation Oncology, University of Rochester, Rochester, NY; ||Department of Radiation Oncology, University of Washington Medical Center, Seattle, WA; #Radiation Oncology Flight, Wilford Hall Medical Center, Lackland Air Force Base. San Antonio, TX; **Department of Radiation Oncology, M. D. Anderson Cancer Center, Houston, TX; ††Department of Radiation Oncology, Brooke Army Medical Center, San Antonio, TX; ‡‡Department of Radiation Oncology, Medical University of South Carolina, Charleston, SC; §§Department of Radiation Oncology, Stanford University, Stanford, CA; ¶¶Department of Radiation Oncology, University of Kansas, Kansas City, and Radiation Oncology Center of Olathe, Olathe, KS; ||||Department of Radiation Oncology, Memorial Sloan-Kettering Cancer Center, New York, NY; ##Department of Radiation Oncology, Boston University Medical Center, and Massachusetts General Hospital Cancer Center, Boston, MA

**Purpose: Variations in target volume delineation represent a significant hurdle in clinical trials involving conformal radiotherapy. We sought to determine the effect of a consensus guideline-based visual atlas on contouring the target volumes.**

**Methods and Materials: A representative case was contoured (Scan 1) by 14 physician observers and a reference expert with and without target volume delineation instructions derived from a proposed rectal cancer clinical trial involving conformal radiotherapy. The gross tumor volume (GTV), and two clinical target volumes (CTVA, including the internal iliac, presacral, and perirectal nodes, and CTVB, which included the external iliac nodes) were contoured. The observers were randomly assigned to receipt (Group A) or nonreceipt (Group B) of a consensus guideline and atlas for anorectal cancers and then instructed to recontour the same case/images (Scan 2). Observer variation was analyzed volumetrically using the conformation number (CN, where CN = 1 equals total agreement).**

**Results: Of 14 evaluable contour sets (1 expert and 7 Group A and 6 Group B observers), greater agreement was found for the GTV (mean CN, 0.75) than for the CTVs (mean CN, 0.46–0.65). Atlas exposure for Group A led to significantly increased interobserver agreement for CTVA (mean initial CN, 0.68, after atlas use, 0.76; $p = .03$) and increased agreement with the expert reference (initial mean CN, 0.58; after atlas use, 0.69; $p = .02$). For the GTV and CTVB, neither the interobserver nor the expert agreement was altered after atlas exposure.**

Reprint requests to: Clifton D. Fuller, M.D., Department of Radiation Oncology, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Dr., MSC7889, San Antonio, TX. Tel: (210) 450-1127; Fax: (210) 450-5543; E-mail: fullercd@uthscsa.edu

C. D. Fuller was supported by a training grant from the National Institutes of Health/National Institute of Biomedical Imaging and Bioengineering, ''Multidisciplinary Training Program in Human Imaging'' (Grant 5T32EB000817-04), a Technology Transfer Grant from the European Society for Therapeutic Radiology and Oncology, and the Product Support Development Grant from the Society for Imaging Informatics in Medicine.

The funders played no role in the study design, collection, analysis, interpretation of data, manuscript writing, or decision to submit the report for publication.

Portions of the data were selected for a Poster Recognition Award at the 2009 American Society for Radiation Oncology Annual Meeting, November 1–5, 2009, Chicago, IL.

[1] C.D.F. and J.N. contributed equally to this work.

Conflict of interest: none.

Received Aug 1, 2009, and in revised form Oct 8, 2009. Accepted for publication Nov 5, 2009.

**Conclusion: Consensus guideline atlas implementation resulted in a detectable difference in interobserver agreement and a greater approximation of expert volumes for the CTVA but not for the GTV or CTVB in the specified case. Visual atlas inclusion should be considered as a feature in future clinical trials incorporating conformal RT.** © 2010 Elsevier Inc.

**Target volume delineation, cooperative group, conformal radiotherapy, atlas, consensus guideline.**

## INTRODUCTION

Interobserver differences in target volume delineation are a demonstrated source of potential treatment variability in the context of clinical trials that incorporate conformal radiotherapy (RT) approaches (1, 2). Recent publications have suggested that target delineation consensus documentation is highly desirable for clinical trials (3) and that specific instructional or educational interventions might afford a measurable effect in terms of physician contouring (4, 5).

As a part of efforts to improve RT implementation for the Southwest Oncology Group (SWOG) trials and consistent with its focus on quality improvement in cooperative studies, the SWOG Radiation Oncology Committee authorized the present study as a pilot project to achieve the following primary specific aims: the feasibility of centralized target volume delineation evaluation as a pretrial adjunct to a SWOG-sponsored study (SWOG S0713), and the determination of the effect of implementation of a consensus anatomic atlas on target volume variability

## METHODS AND MATERIALS

This prospective institutional review board-exempt study was conducted under the auspices of the University of Texas Health Science Center at San Antonio institutional review board. The present study was designed as a double-blind, randomized hypothesis-generating pilot study (Fig. 1). Statistical power for agreement analysis was estimated for a non–Bonferroni-corrected paired-measures Wilcoxon test (assuming a minimum asymptotic relative efficiency of $\geq$0.863 compared with a paired $t$ test), with a specified $1 - \beta$ of 0.7 and $\alpha$ of $\leq$0.05, resulting in a minimal requisite sample size of 6 observers (radiation oncologists) per group, calculated using the G*Power 3 statistical software (6). Goal enrollment was 10–12 observers per cohort.

The participating radiation oncologists (observers) were recruited from SWOG-participating institutions. Those who indicated interest were sent the study documentation, which included a standardized case report, description of the target volumes to be contoured, and a compact disc (CD) containing 3-mm axial computed tomography (CT) images derived from the Digital Imaging and Communications in Medicine (DICOM) file of the standardized case study's simulation CT scan. The volumes were 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). It consists of a user interface with target delineation features common to most commercial treatment planning systems (9) 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 Stage T3N0M0 adenocarcinoma of the rectum with instructions modeled on a SWOG protocol in development at that time that included detailed directives regarding 3-dimensional conformal RT and intensity-modulated RT 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 the terminology modified to fit the nomenclature established in the then-unpublished Radiation Therapy Oncology Group (RTOG) consensus guidelines for target delineation in anorectal cancers (10). The observers were asked to contour the structures as listed in Table 1. The axial CT images were extracted using a single Digital Imaging and Communications in Medicine data set; identical copies of the reconstructed (axial, sagittal, and coronal views) were then designated as Scan 1 and Scan 2.

One-half of the distributed CDs contained an automated HTML link, which, after submission of the first contouring session (Scan 1) and the subsequent electronic survey, directed users to a prepublication version of the RTOG consensus guidelines for target volumes in anorectal cancer (10) and instructions to recontour the exact same axial CT images a second time (Scan 2), 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 recontour the same volumes on the identical CT simulation-derived data set (Scan 2), using the same aforementioned case data/instructions as previously (Group B). Thus, Group B did not receive consensus atlas guidance for recontouring the case. The CDs with and without the HTML link to the consensus atlas were randomly shuffled before labeling and delivery to the participants; the study personnel and physician observers were both unaware of which CD had been distributed to each participant until electronic survey completion.

After completion of the gross tumor volume (GTV) and clinical target volume (CTV) delineation on Scan 1, the observers submitted the case by e-mail and were directed to an electronic survey (Table 2). Subsequently, the participants were provided with instructions to recontour the case with or without the assistance of an anatomically specific consensus atlas. The recently published RTOG consensus atlas (10) was used in prepublication form (available from: www.rtog.org/pdf_document/AnorectalContouringGuidelines.pdf).

In addition, one of the members involved in the development of the RTOG consensus guidelines was asked to delineate Scans

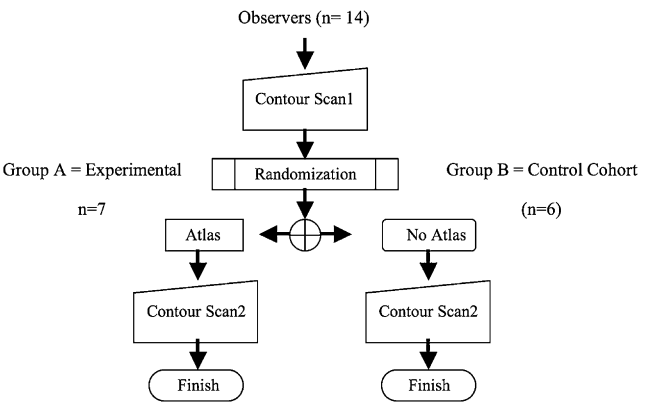


Fig. 1. Study design.

Table 1. Target volume definitions and instructions

| Structure | Definition/Instructions |
|---|---|
| GTV | Includes primary tumor and any pelvic node thought to be involved grossly with metastatic disease<br>Assessment of grossly involved nodal disease can be made according to computed tomography scan for this study |
| CTV | Consists of CTVA, CTVB, and CTVC; should include GTV and following nodal groups: perirectal nodes, presacral nodes, and internal iliac and common iliac nodes below L5–S junction |
| CTVA | For this study, defined as nodal regions that would regularly be treated for rectal cancer (*i.e.,* internal iliac, presacral, and perirectal nodes) |
| CTVB | For this study, defined as external iliac nodal region |
| CTVC | For this study, defined as inguinal nodal region |
| Specific instructions | Contour structures denoted GTV, CTVA, CTVB, and (optionally) CTVC, according to preceding target volume definitions<br>Do NOT add ANY margin to account for positional variability. It will be assumed that a standardized planning target volume expansion of 0.7–1.0 cm omnidirectionally will be applied using the GTV/CTV structures generated |

*Abbreviations:* GTV = gross tumor volume; CTV = clinical target volume.

1 and 2. This observer (L.K.) was designated as the ''reference expert,'' with her contours serving as the *de facto* reference standard against which to compare the observer-derived contours. During the study period, only the reference expert user had any previous knowledge of this atlas; thus, the study participants represented a *tabula rasa* with regard to the consensus guidelines.

### *Delineation agreement analysis*

All delineations were first analyzed visually (Fig. 2), and any protocol deviations from the delivered instructions were identified by a review of all axial contours. The total volume encompassed in cubic centimeters for all structures was calculated and tabulated. A statistical comparison of the volume differentials between Scans

Table 2. Selected survey query results

| Query | All observers* ($n$ = 13) | Group A (atlas exposure; $n$ = 7) | Group B (control; $n$ = 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 amount of time spent on contouring task alone for this session (minutes)'' | | | |
| Mean ± SD | 34 ± 12 | 30 ± 11 | 34 ± 13 |
| Range | 18–65 | 20–65 | 18–60 |
| ''How confident were you in contouring in this anatomic 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 relative difficulty of 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) |
| Very confident | 0 (0) | 0 (0) | 0 (0) |
| Did you find the atlas was a useful aid to your contouring of the case | | | |
| Not helpful | — | 0 (0) | — |
| Somewhat helpful | — | 4 (57) | — |
| Very helpful | — | 3 (43) | — |

* One user's contours were not technically amenable for analysis; responses from this user were discarded from survey results.

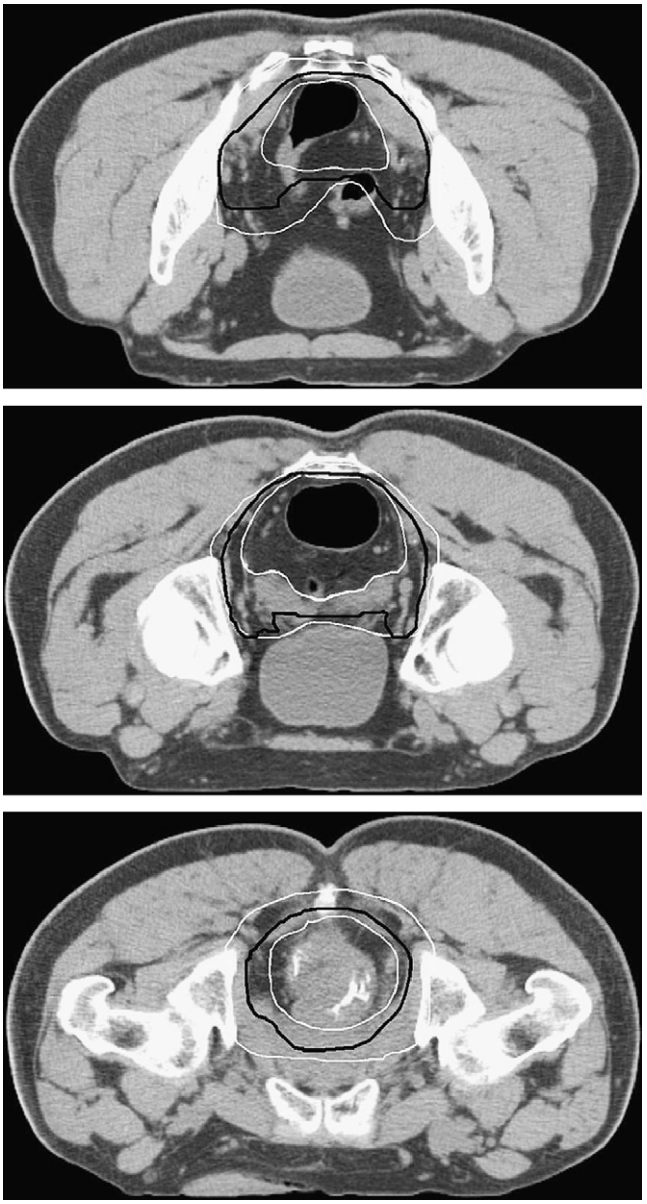

Fig. 2. Overview of largest and smallest clinical target volume A delineations (white) and median clinical target volume A delineation (black) from all observers on Scan 1.

1 and 2 was performed for each structure for Group A and B, respectively.

The baseline interobserver variation for the SWOG protocol was derived from the delineations on Scan 1 from all observers, except for the reference expert. The baseline intraobserver variation was derived from a comparison of the volume of Scans 1 and 2 from Group B. The effect of the atlas on interobserver variation was quantified by comparing the interobserver variation for Scans 1 and 2 from Group A. For a comparison within cohorts, a composite median delineation was calculated for each group. The median delineation represented 50% coverage of the isosurface by the observers, such that each voxel inside was designated by $\geq$50% of the observers and was calculated for GTV, CTVA (internal iliac, presacral, and perirectal nodes), and CTVB (external iliac nodes). 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, the common volume was first calculated between either the median or expert contour (V1) and the observer contour (V2). Subsequently, as a modification of the concept introduced by Feuvret *et al.* (11) and van't Riet *et al.* (12), a conformation number (CN) was derived as $CN = CV^2/(V1 \times V2)$. Differences in the CN values for the target structures (*e.g.,* GTV, CTVA, CTVB) for Scans 1 and 2 for Group A, using both the reference expert and the group median delineation isosurface as a comparator, were calculated and formally assessed for statistical significance using the paired-measures Wilcoxon test.

For Group B, the intraobserver CN values were calculated using the aforementioned van't Riet formula (12); the common volume was calculated between either Scan 1 (V1) or Scan 2 (V2).

### *Post hoc exploratory contour surface variability analysis*

After completion of the planned volumetric analysis, surface distance analysis was performed to identify the regional delineation variation within the CTVA volume, using virtual volume unfolding, as previously published (13, 14). In brief, for the surface distance variation calculation, the reference structure (median or expert) was first resampled to 100 equidistant points per delineated slice. Second, 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 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 26 observers directly asked to participate, 15 submitted contour set pairs, of which 14 were technically evaluable (1 expert, 7 in Group A, and 6 in Group B). The nonevaluable contour set consisted of nonconnected, nonoverlapping contours that precluded ready analysis with the cohort at large. The survey results were pooled and tabulated (Table 2).

All 14 remaining observers delineated the GTV and CTVA on Scans 1 and 2. Although the CTVB was mandatory in the specific delineation instructions, it was only delineated by 11 of the 14 observers. The CTVC, which should not have been delineated, was contoured by 2 observers on both Scan 1 and Scan 2, by 1 observer on Scan 1 only, and by 1 observer on Scan 2 only. For 1 observer in Group A and 4 observers in Group B, major deviations from the delineation protocol (*e.g.,* the GTV was not encompassed by the CTVA) were visible on axial slice review. For an additional observer in Group A, the CTVB covered the internal iliac vessels instead of the external iliac on both Scan 1 and Scan 2. For the 5 observers for whom the CTVA did not fully cover the GTV, the CTVA was manually edited such that the observer-contoured GTV was encompassed for the volume analysis; a preliminary statistical evaluation evidenced minimal alteration of the volumetric statistics by this modification.

Table 3. Selected volumetric and axial slice measures

| | Reference expert | | Group A (atlas) | | | Group B (control) | | |
|---|---|---|---|---|---|---|---|---|
| Parameter | Scan 1 | Scan 2 | Scan 1 | Scan 2 | *p** | Scan 1 | Scan 2 | *p** |
| Mean volume ($cm^3$) | | | | | | | | |
| GTV | 68 | 78 | 78 ± 15.2 | 78 ± 15.3 | 1.0 | 68 ± 9.7 | 68 ± 6.9 | .9 |
| CTVA | 784 | 820 | 800 ± 276 | 809 ± 172 | .7 | 590 ± 208 | 642 ± 251 | .4 |
| CTVB | 67 | 91 | 77 ± 62 | 100 ± 78 | .06 | 71 ± 26 | 51 ± 32 | .3 |
| Mean delineated length (no. of axial CT slices) | | | | | | | | |
| GTV | 12 | 12 | 19 ± 8.3 | 19 ± 8.3 | — | 16 ± 7.4 | 16 ± 7.4 | — |
| CTVA | 19 | 21 | 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 | — |
| Axial CT slices covered by all observers *(n)* | | | | | | | | |
| GTV | 12 | 12 | 1 | 12 | — | 12 | 12 | — |
| CTVA | 45 | 43 | 32 | 31 | — | 31 | 33 | — |
| CTVB | 19 | 21 | 9 | 11 | — | 14 | 3 | — |

Abbreviations as in Table 1.
Data presented as mean ± standard deviation, unless otherwise noted.
* Using paired Wilcoxon rank sum test to compare absolute volume of Scans 1 and 2.

Between Scans 1 and 2, only the increase in the volume of the CTVB in Group A approached statistical significance ($p = .06$; Table 3). In Group B, the number of CTVB slices covered by all observers decreased from 14 to 3 axial CT slices, and the average delineated number of axial CT slices contoured only decreased from 20 to 16 slices. The median GTV delineated on Scan 1 for all observers had a volume of 74 $cm^3$. The average CN for the baseline interobserver variation of the GTV was 0.75 (range, 0.60–0.81). The median CTVA had a volume of 709 $cm^3$ and a CN of 0.65 (range, 0.47–0.75), indicating comparatively greater interobserver disagreement for CTVA compared with the GTV. For CTVB, with a median volume of 70 $cm^3$, even less interobserver 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 (Fig. 3a) and with the expert (Fig. 3b), as measured by CN. The average interobserver CN (*i.e.,* agreement with the median surface) increased from 0.68 (range, 0.41–0.78) on Scan 1 to 0.76 (range, 0.57–0.87) on Scan 2 ($p = .031$, paired Wilcoxon signed rank test; Fig. 3a). The average CN, compared with the expert, increased from 0.58 before the atlas (range, 0.42–0.70) to 0.69 after the atlas (range, 0.58–0.78, $p = .016$; Fig. 3b). For the CTVB, however, neither the interobserver variation (mean CN, 0.39 [range, 0.26–0.67] vs. mean CN, 0.45 [range, 0.13–0.68]; $p = .4$) nor the agreement with the expert (mean CN, 0.31 [range, 0.16–0.49] vs. mean CN, 0.30 [range, 0.11–0.44]) was altered to a statistically significant degree after atlas exposure ($p = .8$).

Because the atlas only affected the observer variation for the CTVA, the exploratory *post hoc* surface distance variation analysis was limited to CTVA (Figs. 4 and 5). To translate the surface maps into numbers, first the reference structure (median/expert CTVA) was divided into the anterior, lateral, and posterior regions and subdivided into the upper and lower subregions at the level of the coccyx tip. For each of the six regions, the standard deviation value covering 5–95% of the regional surface distance difference was taken to characterize the minimal and maximal regional variation (Table 4), although no formal statistical comparison of the

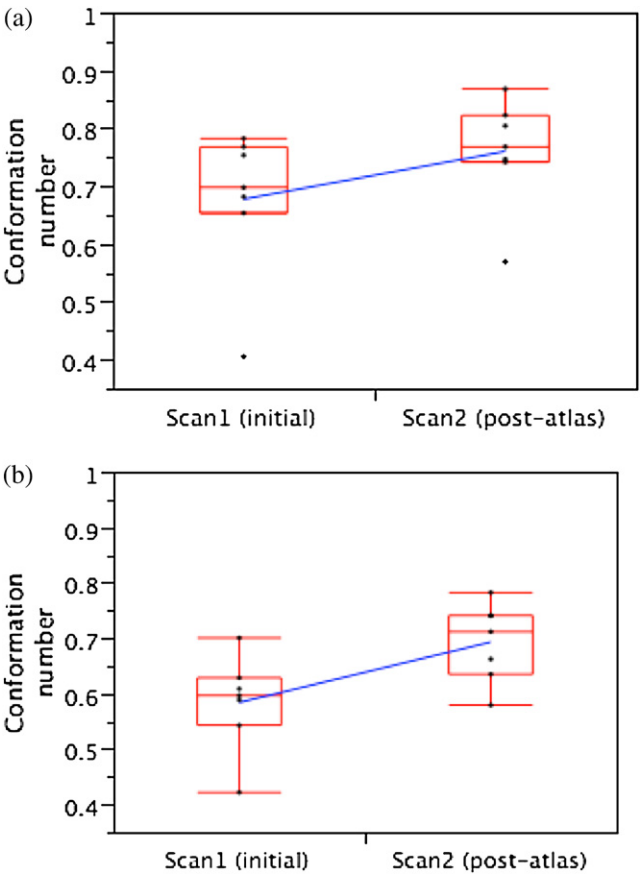


Fig. 3. Conformation number (CN) for observers in Group A before (Scan 1) and after introduction of atlas (Scan 2) compared with median surface (a) and expert delineation (b). CN = 1 indicates complete volumetric agreement. Box plot displays interquartile range (*i.e.,* 25th to 75th percentiles), with horizontal line demonstrating group median; whiskers include all points within ±1.5 times interquartile range from 25th and 75th percentiles. Pre- and postatlas mean CNs connected by diagonal blue line.

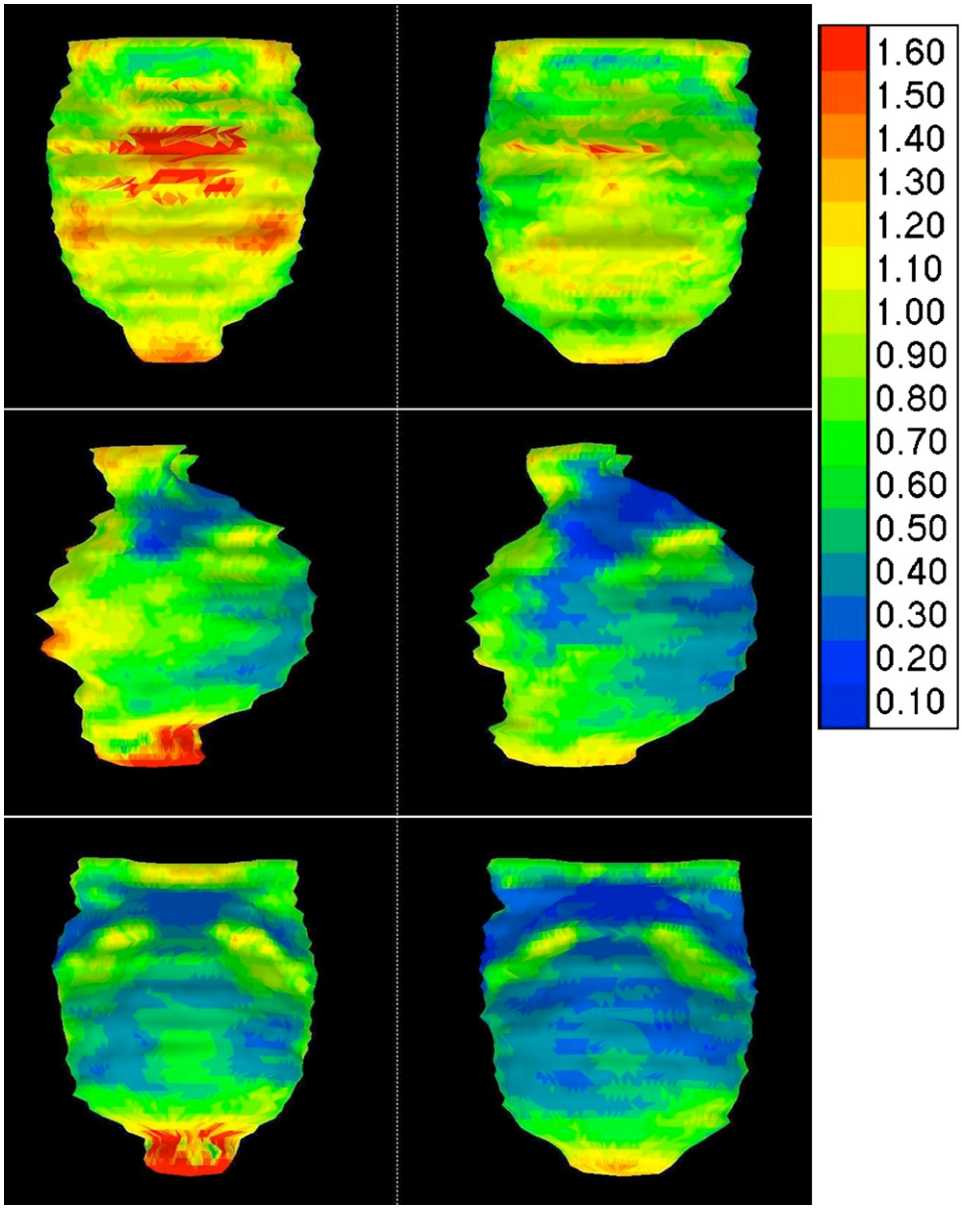


Fig. 4. Interobserver variation changes with introduction of atlas. (Left) Interobserver variation (standard deviation) in Group A before introduction of atlas shown for anterior, sagittal, and posterior view. (Right) Interobserver variation in same group shown after introduction of atlas. Group standard deviation of distance to expert surface shown as color scale on right (values in centimeters from expert surface).

regional subvolumes was performed. Visual inspection (Figs. 4 and 5) showed that the introduction of the atlas resulted in modification of the surface distance between the observers and expert primarily in limited regions of the CTVA, rather than the CTVA volume globally. Modification of the target volumes was most notably localized to the upper-anterior region adjacent to the bladder, lower-posterior, and lateral CTVA (data not shown); however, statistical significance was not formally assessed. For all defined regions, except for the upper-posterior and upper-lateral, the upper 95% confidence interval of the interobserver surface standard deviation was reduced by 0.2–0.8 cm after the introduction of the atlas. As the data in Table 4 demonstrate, >1 cm of surface variation was observed for multiple regions before atlas implementation for all users, and although reduced after atlas administration, >1 cm was still needed to cover 95% of the surface variation in the CTV subregions.

Regarding intraobserver variation, the absolute volume of all respective structures contoured was essentially equivalent (Table 3). A comparison between the delineations on Scans 1 and 2 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. The regional intraobserver variability is illustrated graphically in Fig. 5.

## DISCUSSION

Despite the well-known consequences of geometric inaccuracy in target volume delineation (15–17), interobserver

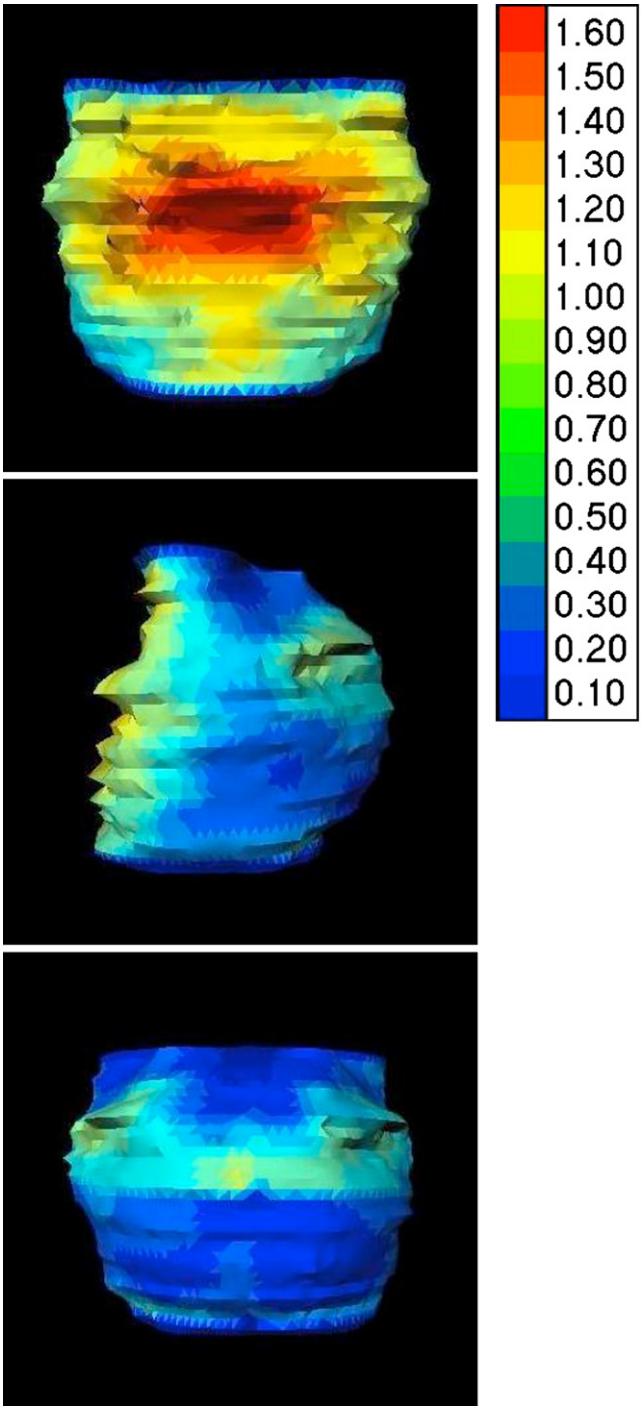


Fig. 5. Intraobserver variation. Intraobserver variation (standard deviation) shown for anterior, sagittal, and posterior views for observers who delineated case twice without any atlas exposure (Group B). Group standard deviation of distance between equivalent points for Scans 1 and 2 for each user shown as color scale on right (in centimeters).

variability in target definition has been demonstrated in a host of studies and at various anatomic sites (18). Simply put, ''interobserver 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, efforts to implement solutions to possible sources of variability/error in the target volume delineation process have continued. 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 subsites. For clinical trials, the situation is potentially more vexing, because to ensure 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 investigators.

In terms of feasibility, the study was readily completed (total study duration, 5 months). Of the 26 invited SWOG institutions, 12 (46%) confirmed an intent to participate; however, only 8 (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 would be both technically and logistically feasible.

The analysis of the resultant data alludes to the difficulty in executing clinical trials in the conformal RT era. The high proportion of major protocol deviations was consistent with that in previous reports. The substantial variation from the expert reference and median contour surfaces observed for all users before the intervention (Figs. 4 and 5 and Table 4) suggests that efforts to further minimize interobserver variability are imperative. As the data in Table 4 demonstrate, substantial interobserver surface deviation was observed for multiple CTVA subregions 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 subregion 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 reductions would result in a decrement in the required planning target volume expansion margins. However, an additional reduction of variation is desired, because the planning target volume margins required to encompass the residual variation in target delineation would limit the practical advantages of intensity-modulated RT compared with conventional RT.

Several limitations to the present pilot study are evident. The sample size was limited, and only a single case was contoured. The use of a reference expert's contours as the *de facto* reference standard points to the fact that the ''ground truth'' in contouring the CTV remains ambiguous (Table 3; note the variation within the reference expert user's sequential contours). Some variance in the study might be attributable to the instructions, which were distinct from standard clinical practice (*e.g.,* the external iliac nodes are not typically contoured for T3 rectal cases). Our invitation was limited to SWOG institutes, creating a potential sampling bias and that only interested observers participated created 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 (Fig. 3a) and reduce interobserver variability to a statistically

Table 4. CTVA interobserver surface distance*

| Variable | Anterior (cm) | Posterior (cm) | Lateral (cm) |
|---|---|---|---|
| All observers, Scan 1 (*i.e.,* initial total interobserver variation) | | | |
| Upper | 0.6–1.4 | 0.2–0.7 | 0.3–1.1 |
| Lower | 0.6–1.2 | 0.3–1.2 | 0.4–1.1 |
| Group A, Scan 1 (*i.e.,* initial interobserver variation) | | | |
| Upper | 0.5–1.4 | 0.2–0.7 | 0.2–1.0 |
| Lower | 0.7–1.3 | 0.4–1.4 | 0.5–1.4 |
| Group A, Scan 2 (*i.e.,* postatlas interobserver variation) | | | |
| Upper | 0.5–1.1 | 0.1–0.7 | 0.2–0.9 |
| Lower | 0.6–1.1 | 0.2–0.6 | 0.3–0.8 |

*Abbreviation:* CTVA = clinical target volume A (see text for details).

* Expressed as 95% confidence interval of group standard deviation from median group isosurface, by regional subdivision, before and after atlas exposure.

detectable degree (Fig. 3b). However, our data also suggest substantial residual variability in rectal target volume delineation, even after atlas use (Tables 3 and 4).

The results of the present study are consistent with those from 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.* (26) showing increased protocol compliance after a site-specific educational experience. With regard to consensus guideline application as an avenue toward target variability reduction, Dimopoulos *et al.* (32) reported a study in which 19 cervical cancer cases were contoured using the Groupe Européen de Curiethérapie and the European Society for Therapeutic Radiology and Oncology (GEC-ESTRO) guidelines by 2 observers (11, 21), with a resultant between-user conformity index in the range 0.6–0.7 for target volumes, roughly consistent with the CNs in the present series. Likewise, Wong *et al.* (38) recently demonstrated, using a test-retest sequence, that improved consistency in seroma contouring could be observed after exposure to consensus guidelines. In the clinical trial setting, it is likely that ''trial-specific'' atlases should be used according to patterns of failure data (as per Roels *et al.* [39]) or, possibly, after a pilot contouring trial similar to the present study. For instance, the Radiation Therapy Oncology Group anorectal consensus guidelines stipulate coverage ''extending CTVA ~1 cm into the posterior bladder, to account for day-to-day variation in bladder position'' (10). This incorporation of motion into CTV generation, rather than the planning target volume expansion, represents a conceptual break with International Commission on Radiation Units and Measurements 6240 and other guidelines (39), in which the posterior bladder wall would not be contoured. No users in Group A included significant portions of the posterior bladder before using the atlas, although most did so after atlas exposure (in compliance with the presented atlas [10] and consistent with the reference expert).

Future studies are required to ascertain whether the 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 data set might also be integrated into the design of educational materials for a proposed Dutch cooperative group rectal study workshop. We plan to use portions of this data set to construct composite models accounting for rectal motion and setup variability (14, 41), as well as the development of novel software strategies for evaluation (42) and minimization of target delineation variance.

## CONCLUSION

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

## 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.
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.
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.

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;75:990–993.
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.
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.
7. Steenbakkers 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.
8. Steenbakkers 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.
9. Duppen JC, Steenbakkers 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: A Radiation Therapy Oncology Group consensus panel contouring atlas. *Int J Radiat Oncol Biol Phys* 2008;74:824–830.
11. Feuvret L, Noel G, Mazeron JJ, *et al*. Conformity index: A review. *Int J Radiat Oncol Biol Phys* 2006;64:333–342.
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.
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.
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.
15. Jeanneret-Sozzi W, Moeckli R, Valley J, *et al*. The reasons for discrepancies in target volume delineation. *Strahlenther Onkol* 2006;182:450–457.
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.
17. Njeh CF. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. *J Med Phys* 2008;33:136–140.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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.
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;89:164–171.
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.
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.
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.
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.
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.
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.
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.
40. International Commission on Radiation Units and Measurements. ICRU report 62: Prescribing, recording, and reporting photon beam therapy. Bethesda: 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.
42. Jena R, Kirkby NF, Burton KE, et al. A novel algorithm for the morphometric assessment of radiotherapy treatment planning volumes. *Br J Radiol*. In press Jul 20. [Epub ahead of print].