(Un-)certainties in radiotherapy of rectal cancer

Nijkamp, J.A.

Citation for published version (APA):
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. Fuller¹,²,₃, Jasper Nijkamp⁴, Joop C. Duiven⁴, Coen R.N. Rasch⁴, Charles R. Thomas³, Samuel J. Wang³, Paul Okunieff⁵, William E. Jones¹, Daniel Baseeman¹, Shilpen Patel⁶, Carlo G.N. Demandante¹,, Anna M. Harris¹, Benjamin D. Smith⁷,⁸, Alan W. Katz⁵, Camille McGann⁸, Jennifer L. Harper¹⁰, Daniel T. Chang¹¹, Stephen Smalley¹², David T. Marshall¹⁰, Karyn A. Goodman¹³, Niko Papanikolaou¹,₂, and Lisa Kachnic¹⁴ on behalf of the Radiation Oncology Committee of the South-West Oncology Group

¹Department of Radiation Oncology and 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 Radiation Oncology, The Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, 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
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.

Material and methods
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 re-contour 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 inter-observer agreement for CTVA (mean initial CN, 0.68, after atlas use, 0.76; p=0.03) and increased agreement with the expert reference (initial mean CN, 0.58; after atlas use, 0.69; p=0.02). For the GTV and CTVB, neither the inter-observer nor the expert agreement was altered after atlas exposure.

Conclusions
Consensus guideline atlas implementation resulted in a detectable difference in inter-observer 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.
**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 (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 pre-trial 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.

**Material and methods**

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. 3.1). 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-β of 0.7 and α of ≤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 enrolment was 10–12 observers per cohort.

![Fig. 3.1: Study design](image)
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 modelled 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 3.1. 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 re-contour 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 re-contour 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 re-contouring 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. Subsequently, the participants were provided with instructions to re-contour 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 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.

### Table 3.1: Target volume definitions and instructions

<table>
<thead>
<tr>
<th>Structures</th>
<th>Definition/Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTV</td>
<td>Includes primary tumor and any pelvic node thought to be involved grossly with metastatic disease. Assessment of grossly involved nodal disease can be made according to computed tomography scan for this study.</td>
</tr>
<tr>
<td>CTV</td>
<td>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.</td>
</tr>
<tr>
<td>CTVA</td>
<td>For this study, defined as nodal regions that would regularly be treated for rectal cancer (i.e., internal iliac, presacral, and perirectal nodes).</td>
</tr>
<tr>
<td>CTVB</td>
<td>For this study, defined as external iliac nodal region.</td>
</tr>
<tr>
<td>CTVC</td>
<td>For this study, defined as inguinal nodal region.</td>
</tr>
</tbody>
</table>

Specific instructions:
- Contour structures denoted GTV, CTVA, CTVB, and (optionally) CTVC, according to preceding target volume definitions.
- 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.

### Delineation agreement analysis

All delineations were first analyzed visually (Fig. 3.2), 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 1 and 2 was performed for each structure for Group A and B, respectively.

The baseline inter-observer variation for the SWOG protocol was derived from the delineations on Scan 1 from all observers, except for the reference expert. The baseline intra-observer variation was derived from a comparison of the volume of Scans 1 and 2 from Group B. The effect of the atlas on inter-observer variation was quantified by comparing the inter-observer 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 ≥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 to 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.
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 intra-observer 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 non-evaluable contour set consisted of non-connected, non-overlapping contours that precluded ready analysis with the cohort at large.

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.
Between Scans 1 and 2, only the increase in the volume of the CTVB in Group A approached statistical significance (p = 0.06; Table 3.2). 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 cm³. The average CN for the baseline inter-observer variation of the GTV was 0.75 (range, 0.60–0.81). The median CTVA had a volume of 709 cm³ and a CN of 0.65 (range, 0.47–0.75), indicating comparatively greater inter-observer disagreement for CTVA compared with the GTV. For CTVB, with a median volume of 70 cm³, even less inter-observer agreement could be found, with an average CN of 0.46 (range, 0.24–0.70).
Chapter 3

Atlas exposure led to a statistically significant increase in volumetric agreement on CTVA between observers (Fig. 3.3a) and with the expert (Fig. 3.3b), 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 Scan 1 to 0.76 (range, 0.57–0.87) on Scan 2 (p = 0.031, 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=0.016; Fig. 3.3b). For the CTVB, however, neither the inter-observer variation (mean CN, 0.39 [range, 0.26–0.67] vs. mean CN, 0.45 [range, 0.13–0.68]; p = .4)

Fig. 3.4: Inter-observer variation changes with introduction of atlas. (Left) Inter-observer variation (standard deviation) in Group A before introduction of atlas shown for anterior, sagittal, and posterior view. (Right) Inter-observer 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).
Consensus atlas implementation for rectal cancer target volume delineation

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 = 0.8).

Because the atlas only affected the observer variation for the CTVA, the exploratory post hoc surface distance variation analysis was limited to CTVA (Figs. 3.4 and 3.5). 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 sub-regions 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 3.3), although no formal statistical comparison of the regional sub-volumes 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 inter-observer surface standard deviation was reduced by 0.2–0.8 cm after the introduction of the atlas. As the data in Table 3.4 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 sub-regions.

Fig. 3.5: Intra-observer variation. Intra-observer 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).
Regarding intra-observer variation, the absolute volume of all respective structures contoured was essentially equivalent (Table 3.2). 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 intra-observer variability is illustrated graphically in Fig. 3.5.

### Table 3.2: Selected volumetric and axial slice measures

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Reference expert</th>
<th>Group A (atlas)</th>
<th>Group B (control)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scan 1</td>
<td>Scan 2</td>
<td>Scan 1</td>
</tr>
<tr>
<td>Mean volume (cm³)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTV</td>
<td>68</td>
<td>78</td>
<td>78 ± 15.2</td>
</tr>
<tr>
<td>CTVA</td>
<td>784</td>
<td>820</td>
<td>800 ± 276</td>
</tr>
<tr>
<td>CTVB</td>
<td>67</td>
<td>91</td>
<td>77 ± 62</td>
</tr>
<tr>
<td>Mean delineated length (no. of axial slices)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTV</td>
<td>12</td>
<td>12</td>
<td>19 ± 8.3</td>
</tr>
<tr>
<td>CTVA</td>
<td>19</td>
<td>21</td>
<td>44 ± 6.8</td>
</tr>
<tr>
<td>CTVB</td>
<td>19</td>
<td>21</td>
<td>13 ± 7.5</td>
</tr>
<tr>
<td>Axial CT slices covered by all observers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTV</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>CTVA</td>
<td>45</td>
<td>43</td>
<td>32</td>
</tr>
<tr>
<td>CTVB</td>
<td>19</td>
<td>21</td>
<td>14</td>
</tr>
</tbody>
</table>

### Table 3.3: CTVA interobserver surface distance, expressed as 95% confidence interval of group standard deviation from median group isosurface, by regional subdivision, before and after atlas exposure

<table>
<thead>
<tr>
<th>Variable</th>
<th>Anterior (cm)</th>
<th>Posterior (cm)</th>
<th>Lateral (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All observers, Scan 1 (i.e., initial total interobserver variation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td>0.6-1.4</td>
<td>0.2-0.7</td>
<td>0.3-1.1</td>
</tr>
<tr>
<td>Lower</td>
<td>0.6-1.2</td>
<td>0.3-1.2</td>
<td>0.4-1.1</td>
</tr>
<tr>
<td>Group A, Scan 1 (i.e., initial interobserver variation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td>0.5-1.3</td>
<td>0.2-0.7</td>
<td>0.2-1.0</td>
</tr>
<tr>
<td>Lower</td>
<td>0.7-1.3</td>
<td>0.4-1.4</td>
<td>0.5-1.4</td>
</tr>
<tr>
<td>Group A, Scan 2 (i.e., postatlas interobserver variation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td>0.5-1.1</td>
<td>0.1-0.7</td>
<td>0.2-0.9</td>
</tr>
<tr>
<td>Lower</td>
<td>0.6-1.1</td>
<td>0.2-0.6</td>
<td>0.3-0.8</td>
</tr>
</tbody>
</table>

## 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 and at various anatomic sites [18]. 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, 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. 3.4 and 3.5 and Table 3.3) suggests that efforts to further minimize inter-observer variability are imperative. As the data in Table 3.4 demonstrate, 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 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 3.2; 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. 3.3a) and reduce inter-observer variability to a statistically detectable degree (Fig. 3.3b). However, our data also suggest substantial residual variability in rectal target volume delineation, even after atlas use (Tables 3.2 and 3.3).
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 1cm 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.

Conclusions
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 inter-observer (both with and without the 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 RT approaches.
References


Chapter 3


Consensus atlas implementation for rectal cancer target volume delineation


