

# Development of a Software for Quantitative Evaluation Radiotherapy Target and Organ-at-Risk Segmentation Comparison

Jayashree Kalpathy-Cramer · Musaddiq Awan · Steven Bedrick · Coen R. N. Rasch · David I. Rosenthal · Clifton D. Fuller

Published online: 17 September 2013  
© Society for Imaging Informatics in Medicine 2013

**Abstract** Modern radiotherapy requires accurate region of interest (ROI) inputs for plan optimization and delivery. Target delineation, however, remains operator-dependent and potentially serves as a major source of treatment delivery error. In order to optimize this critical, yet observer-driven process, a flexible web-based platform for individual and cooperative target delineation analysis and instruction was developed in order to meet the following unmet needs: (1) an open-source/open-access platform for automated/semiautomated quantitative interobserver and intraobserver ROI analysis and comparison, (2) a real-time interface for radiation oncology trainee online self-education in ROI definition, and (3) a source for pilot data to develop and validate quality metrics for institutional and cooperative group quality assurance efforts. The resultant software, Target Contour Testing/Instructional Computer Software (TaCTICS), developed using *Ruby on Rails*, has since been implemented and

proven flexible, feasible, and useful in several distinct analytical and research applications.

**Keywords** Segmentation · Imaging informatics · Image segmentation · Radiation oncology · Algorithms · Teaching

## Introduction

External beam radiotherapy aims to treat tumor and at-risk tissues to doses sufficient to eliminate all viable cancer cells within a specific target volume (TV) [1]. Unfortunately, dose coverage of TVs, as either gross tumor volumes or areas harboring subclinical microscopic tumor clonogens (clinical TV), is substantially complicated technically by the fact that these regions of interest (ROIs) often abut radiosensitive organs-at-risk (OARs). Excessive radiation dose to OARs may lead to deleterious sequelae [1–4]. Thus, a cost–benefit relationship between delivering tumoricidal dose to TVs while maintaining safe doses to OARs is a critical calculation in any radiotherapy plan [4]. Ideally, dose would conform exactly to TVs, while failing to infringe on adjacent OARs; practically, this is almost never possible.

Technical innovation has potentiated a new era in radiotherapy. Beginning in the late 1980s, the rise of widely available 3D imaging (specifically, computed tomography [CT] simulation) birthed conformal radiotherapy approaches with a spatial resolution of radiotherapy delivery technically impossible (outside of intracranial stereotactic applications) even a decade prior [5–7]. The further technical advancement of computing capacity enabled intensity modulated radiotherapy (IMRT) [8], a method by which dose can be further sculpted to conform to TVs using optimization algorithms to select beam weightings consistent with OAR/TV cost–benefit inputs [9]. Finally, the last decade has also seen, through improved

---

J. Kalpathy-Cramer  
Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Harvard Medical School, Department of Radiology and Neuroscience, Massachusetts General Hospital, Charlestown, MA, USA

S. Bedrick  
Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University, Portland, OR, USA

M. Awan · D. I. Rosenthal · C. D. Fuller (✉)  
Head and Neck Section, Division of Radiation Oncology, Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA  
e-mail: cdfuller@mdanderson.org

C. R. N. Rasch  
Department of Radiotherapy, Academic Medical Center, Amsterdam, Netherlands

image-guided radiotherapy (IGRT) technology and practices [10–12], a reduction in the margins required to ensure geometric coverage due to either internal target motion (represented by the internal TV) or setup and immobilization error margin (conceptualized as the planning TV) [13].

Despite this progress, at the level of treatment planning, the process remains heavily user-dependent and variable. The process of radiotherapy treatment planning begins in most cases with the acquisition of a CT dataset, which is then transferred to an FDA-approved treatment planning software (TPS) as a Digital Imaging and Communications in Medicine (DICOM) file. The resultant images are then reconstructed within the TPS for TV delineation. Physician users then manually designated voxels as named ROIs corresponding to TVs or OARs [14]. These ROIs are then used as the primary inputs for all subsequent portions of radiotherapy planning, from beam selection and optimization to plan quality analysis and assurance. After both the decision to treat and the dose prescription, the time-consuming patient-specific process of target delineation is the most critical portion of physician treatment planning.

However, despite the pivotal nature of target definition, data suggest that it may be *the* major source of radiotherapy treatment error [15]. Data from a host of anatomic sites suggest that TV delineation variability results in variability sufficient to potentially alter clinical outcomes for individual patients [16, 17]. Additionally, variation in interobserver target definition may potentially obviate results of cooperative group clinical trials using IMRT or IGRT approaches [16, 18, 19]. In fact, the largest domestic cooperative clinical trialist group has deemed target delineation standardization as “[m]ost critical” in order to “incorporate intensity-modulated radiation therapy and IGRT techniques into the standard of care used in national clinical trials” [20].

Data suggest that specific educational interventions may serve to improve performance in manual target delineation. Examples exist of didactic and workshop-based educational efforts that demonstrate detectable improvement in trainee performance [21–23]. Additionally, the development and validation of atlas-based consensus guidelines have demonstrated some capacity for the standardization of target definition (though this appears to vary by tissue site) [16, 22, 24–31].

Complicating these efforts is the fact that, until recently, few noncommercial tools existed for ROI-based atlas development or quantitative ROI comparison [32–35]. However, several groups have lately developed a series of tools for central collection, atlas development, and ROI analysis. For example, the Netherlands Cancer Institute group has previously developed and implemented “Big Brother” [33–35], a self-contained software GUI that collects task-specific ergonomic data regarding the totality of the target delineation process, affording exceptional informatics insight into the mechanics and process of target delineation. Data from Big

Brother can then be used to create multiuser composite ROIs for consensus development or interobserver ROI comparison [16, 36, 37]. Likewise, Multicentric In Silico Trials In Radiotherapy (MISTIR) is a developed resource allowing pre-execution clinical trial evaluation, with an impressive infrastructure allowing multisite ROI data contribution and central ROI quality assurance [38].

However, we noted that, while TPS systems often had impressive tools for manual segmentation, multisite trials should ideally use software platforms that are “manufacturer agnostic,” since various clinical sites use different commercial TPS platforms; consequently, universal DICOM RT import–export capability was required. In the past, cost and accessibility to the end user has been a significant impediment to groups in software selection, thus an open-access/open-source solution was desired (in contrast to commercial vendor-supplied software or access-restricted academic platforms). Another specific limitation of most extant technologies for multisite ROI delineation analysis and segmentation comparison for educational applications was the lack of real-time or near real-time qualitative and quantitative feedback. In our execution of a previous study of target delineation intervention [16], participants would frequently inquire as to their relative performance vis-à-vis other users and/or reference experts. However, by the time central analysis was performed (often weeks later), recollection of their personal contouring experience by users was often obscured by the passage of time. Additionally, there was a reported desire by trainee participants to have the opportunity to “recontour” over time to track improved correlation with expert contours.

These experiences, as well as our previous research efforts, led to identification of the following unmet needs:

- A web-based platform for automated/semiautomated quantitative interobserver and intraobserver ROI analysis and comparison;
- A real-time interface for radiation oncology trainee online self-education in ROI definition;
- A collection of pilot data to develop and validate quality metrics for institutional and cooperative group quality assurance efforts.

To this end, an application was submitted to the Society of Imaging Informatics in Medicine (SIIM) Product Development Grant mechanism to create open-access, open-source, ROI analysis software capable of:

- providing instructional feedback to end users at multiple institutional sites (e.g., professional skill development),
- implementing multiuser rubrics for quality assurance (e.g., atlas development for clinical trials),
- testing baseline competency of end users in target delineation tasks (e.g., longitudinal learner instruction, clinical trial credentialing),

- and delivering real-time dynamic feedback of user ROIs with other user-derived or expert-derived ROIs (e.g., on-line resident self-assessment and learning).

With generous funding support from SIIM, we have constructed and implemented Target Contour Testing/ Instructional Computer Software (TaCTICS), a novel software platform for TV and OAR analysis. The specific aim of the current manuscript is to report on the structure and initial implementation of the TaCTICS software.

## Materials and Methods

### Approval and Compliance

Institutional Review Board (IRB) approval as an exempt, 45 CFR 46.101(b)(4)(39)-compliant study was obtained, allowing the collection of anonymized DICOM files. Clinical datasets were anonymized and stripped of all 45 CFR 46.101(b)-specified identifiers, and fictionalized case histories were constructed for all resultant efforts detailed herein.

### Software Structure

The system is built using a *Ruby on Rails* framework. We currently use a PostgreSQL database to store user information, information about the studies including location of the CT slices, information extracted from the DICOM header including names, volumes, and slice information for the structures, and metrics derived from all users. However, any relational database could be used, depending on the needs and requirements of the user. We utilized the ruby-dicom gem for parsing

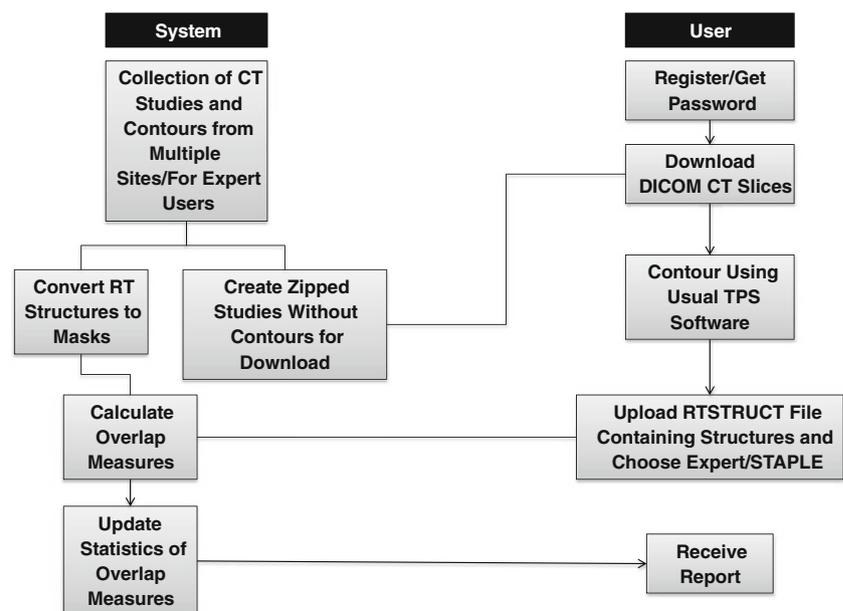
the DICOM files. The main processing of the structures and calculation of the metrics was performed in C++ using the ITK toolkit [39]. These procedures were wrapped in Ruby and called from the website to generate the report. The flow of data and user interaction for the system is given in Fig. 1.

*Configuring the System* Since the system is developed in *Ruby on Rails*, it is a fully functional web application and is implemented in a manner similar to other Rails applications. A system administrator can configure the system for a particular study. The first step is to select and upload the appropriate CT studies (after acquiring the necessary IRB approval). Short case scenarios and instructions are typically uploaded at this stage. Next, the desired metrics (for a description of the available metrics, see the succeeding paragraphs) are selected. Finally, the structure, content, and format of the report are customized. The report can include the metrics, histograms, thumbnails of images, and explanations of the tables and graphs. The generic website contains templates for the website but the look and feel can be easily customized to meet the needs for different study protocols.

*User Experience (Front End)* The “Home” page of the website is configured to describe the present study protocol (Fig. 2). The first step in the case of a new user is registration (Fig. 3). The user registers on the website for a particular study or is given a login and a password by a system administrator. In the case of self-registration, the user enters details requested by the study (name, organization, contact information, etc).

After a user logs in to the system, they can download the studies associated with the protocol as well as any atlases relevant to the study at hand (Fig. 4). Typically, a

**Fig. 1** TaCTICS training program system design



**OREGON HEALTH & SCIENCE UNIVERSITY**

# Target Contour Testing/Instructional Computer Software (TaCTICS)

Home | Registration | Download Data | Submission | Comparisons | Publications

Site

## TaCTICS Overview

TaCTICS is an instructional software for radiation therapy target volume delineation.

We have obtained contours (RT-structures) for different sites including Lung, Liver, Prostate, and Head and Neck from experts. The goal is to match the expert's contour as closely as possible.

### Instructions

1. Register at the registration page. If you have previously registered, log on.
2. Select the anatomical site and download the data. The data is available as zipped DICOM images.
3. Contour the structures using your favorite treatment planning system.
4. Export the RTSTRUCT file.
5. Upload the RTSTRUCT file at the submission page.
6. Select the correct anatomical site (template study) and reference that you want to compare against.
7. Finally, you should get an e-mail with your report within a few minutes.

Home | Registration | Download Data | Submission Site | Comparisons | Publications

2010 Copyright © OHSU Department of Medical Informatics & Clinical Epidemiology

**Fig. 2** Home page of the TaCTICS site

study consists of a zipped file containing the CT slices. Additional information about the study, such as a case scenario and instructions, can be also downloaded at the same time. The users can then upload the CT slices into their preferred TPS and contour the study per the protocol. Finally, the resulting DICOM RTSTRUCT file is exported from the TPS.

The next step in the process is “submission” (Fig. 5). The user provides their login and selects the appropriate study. The DICOM RTSTRUCT file can be uploaded at this stage. Once the file has been processed (typically a few minutes), the users are e-mailed a report containing all the desired metrics, histograms of all the corresponding metrics available in the system with their contour sets highlighted, as well as thumbnails of CT slices with their ROIs and those of the expert/comparison ROIs overlaid. An example for a right parotid contour from a head and neck dataset is shown below in Fig. 6 (top). Users can also identify their relative graphical location on a histogram of all user agreement for a particular contour (Fig. 6 (bottom), red highlighted region) and, subsequently, by their relative histogram position, judge visually as well as

numerically their agreement with a reference ROI set. Alternatively, the user can select another expert user ROI or consensus/composite ROI for comparison.

As illustrated in the system design diagram, the back end of the system consists of a relational database. The users, studies, contours, and metrics are stored in the database. Each time a new DICOM RTSTRUCT file is uploaded, the system makes a call first to the C++ executable to parse the RTSTRUCT file into a set of NifTI [40] files, each containing a binary mask representing the structure. Subsequently, when the user wishes to perform comparisons, other C++ executables are called to perform the comparisons.

The software in its current version is available for download and end user modification at <https://github.com/kalpathy/tacticsRT>.

#### Available Metrics

A number of statistical ROI segmentation analytic metrics have been used to describe the conformity of TVs generated by two

Fig. 3 User registration interface

OREGON HEALTH & SCIENCE UNIVERSITY

## Target Contour Testing/Instructional Computer Software (TaCTICS)

Home | Registration | Download Data | Submission | Comparisons | Publications

Site

**New Users, please register here:**

Login ID:\*

First name:

Last (Family) name:

Your organization:

Your primary area of research:

How did you learn about TaCTICS?

Your website (optional):

Your email address:\*

Your password:\*

Please retype your password:\*

Home | Registration | Download Data | Submission Site | Comparisons | Publications  
 2010 Copyright © OHSU Department of Medical Informatics & Clinical Epidemiology

observers or between an observer and a consensus or gold standard. Some of the most commonly used metrics include those based on volumes, spatial overlap in 2D and 3D, and surface distances. The utility and biases in each specific measure depend on the shape and volumes of the ROIs. The following provides an explanation of these individual metrics that are available in the current iteration of TaCTICS.

#### Volumetric Difference

The volumetric difference (VD) is defined as:

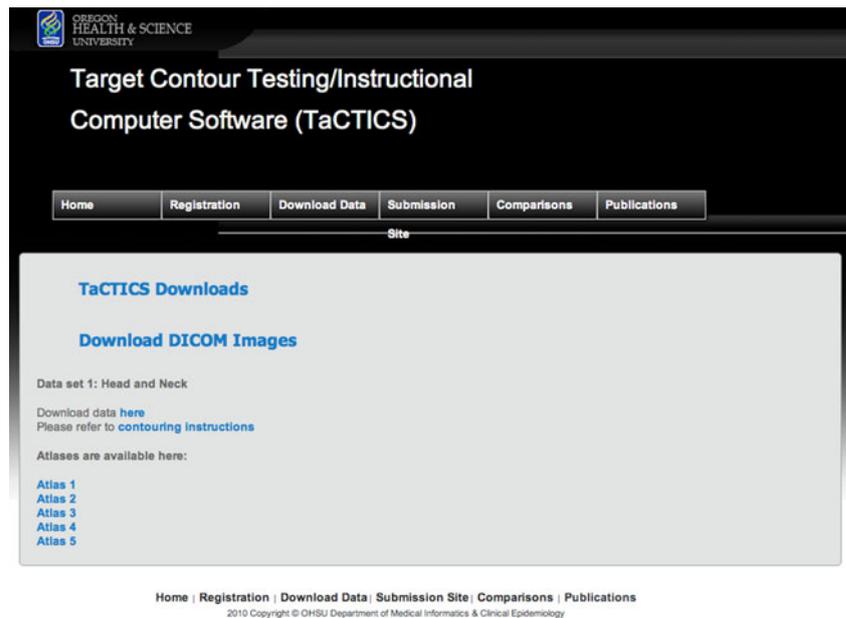
$$VD = \frac{V_a - V_g}{V_g} \times 100$$

where  $V_a$  is the volume of the user-derived contour and  $V_g$  is the volume of the “gold standard” (or the expert-derived contour). As this measure is independent of the spatial locations of the respective volumes, this measure does not address discrepancies resulting from a lack of overlap between the volumes. Additionally, a single pixel/voxel variation between the contours can lead to substantially higher numbers when the volumes are small. Thus, this measure is rarely used in isolation.

#### Dice/Jaccard Coefficients

The Dice [41] and Jaccard [42] coefficients (or Tanimoto [43, 44]) are commonly used measures of spatial overlap, especially for binary labels. Consider Fig. 7 in which  $A$  is the user contour and  $G$  is the gold standard.

**Fig. 4** DICOM file download interface



The Dice and Jaccard coefficients are defined as:

$$D = \frac{2|A \cap G|}{|A| + |G|} \times 100$$

$$J = \frac{|A \cap G|}{|A \cup G|} \times 100$$

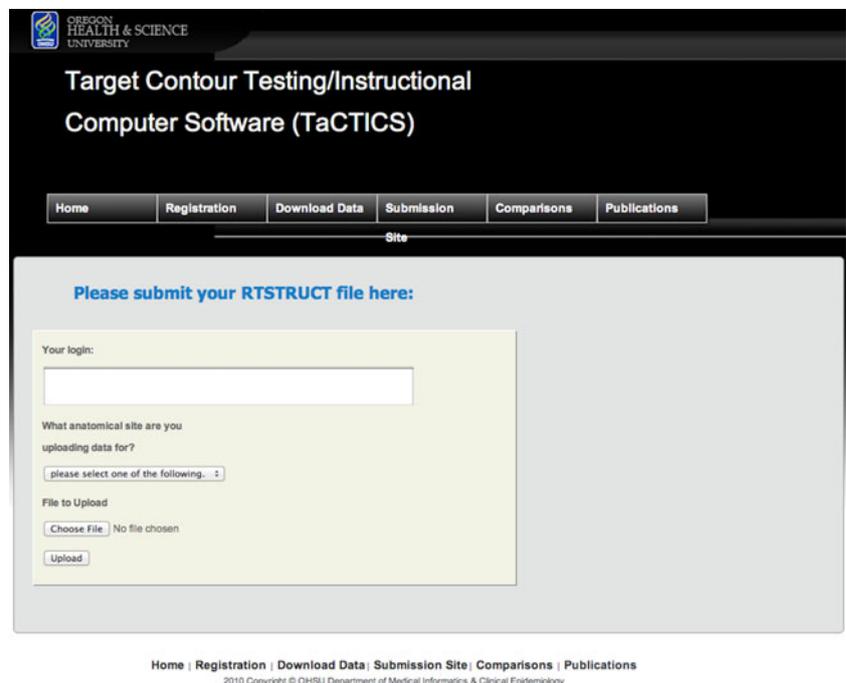
As defined, both of these measures are symmetric. However, in situations where one might desire a higher cost for missing the tumor (such as contouring for radiation oncology), false-positive Dice (FPD) and false-negative Dice (FND) measures can also be used.

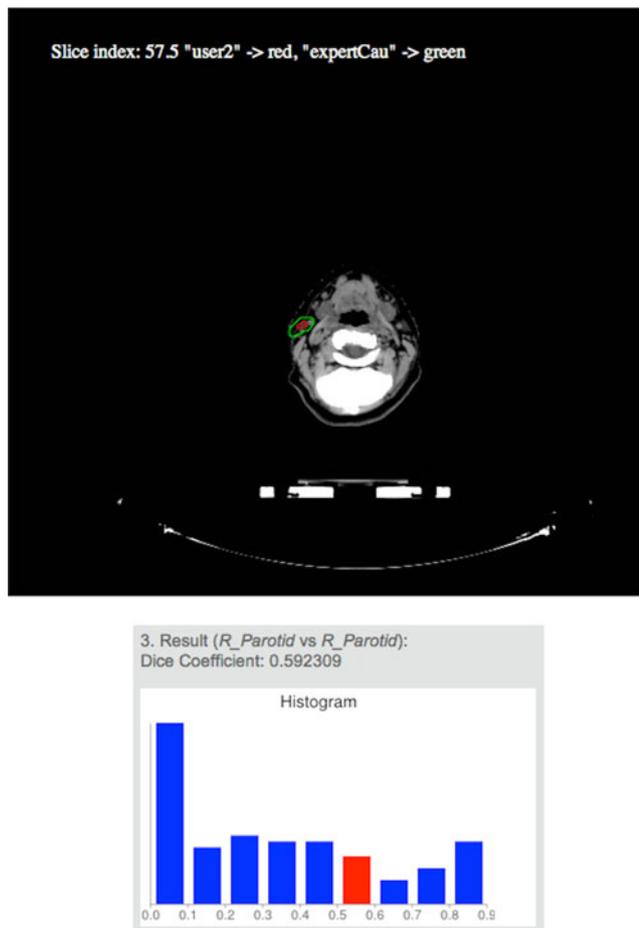
The FPD is measure of voxels that are labeled positive (i.e., 1) by the user but not the expert, while the FND is a measure of the voxels that were considered positive according to the expert but missed by the user being evaluated.

$$FND = \frac{2|\bar{A} \cap G|}{|A| + |G|} \times 100$$

$$FPD = \frac{2|A \cap \bar{G}|}{|A| + |G|} \times 100$$

**Fig. 5** DICOM RT-STRUCT submission interface





**Fig. 6** Top example of the comparison of a user contour (red) with an expert contour set (green) of the right parotid gland on an individual slice. Bottom quantitative histogram showing the relative performance of the user (red) to other users in the dataset

The Dice coefficient is sensitive to the volume of the tumor in that, again, a few pixel/voxel differences between the contours can result in low Dice coefficients in small tumors.

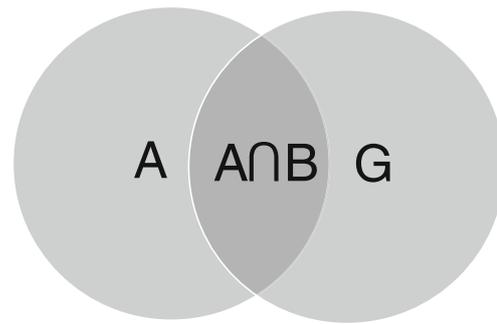
### Hausdorff Distance

Unlike the region-based approaches given previously, surface distance metrics are derived from the contours or the points that define the boundaries of the objects. The Hausdorff distance (HD) is commonly used to measure the distance between point sets defining the objects [43]. The HD between  $A$  and  $G$ ,  $h(A, G)$  is the maximum distance from any point in  $A$  to a point in  $G$  and is defined as:

$$h(A, G) = \max_{a \in A} (d(a, G))$$

where:

$$d(a, G) = \min_{g \in G} \|a - g\|$$



**Fig. 7** Observer ( $A$ ) and gold standard ( $G$ ) regions of interest for illustration of overlap metrics

The symmetric HD is given as:

$$H(A, G) = \max(h(A, G), h(G, A))$$

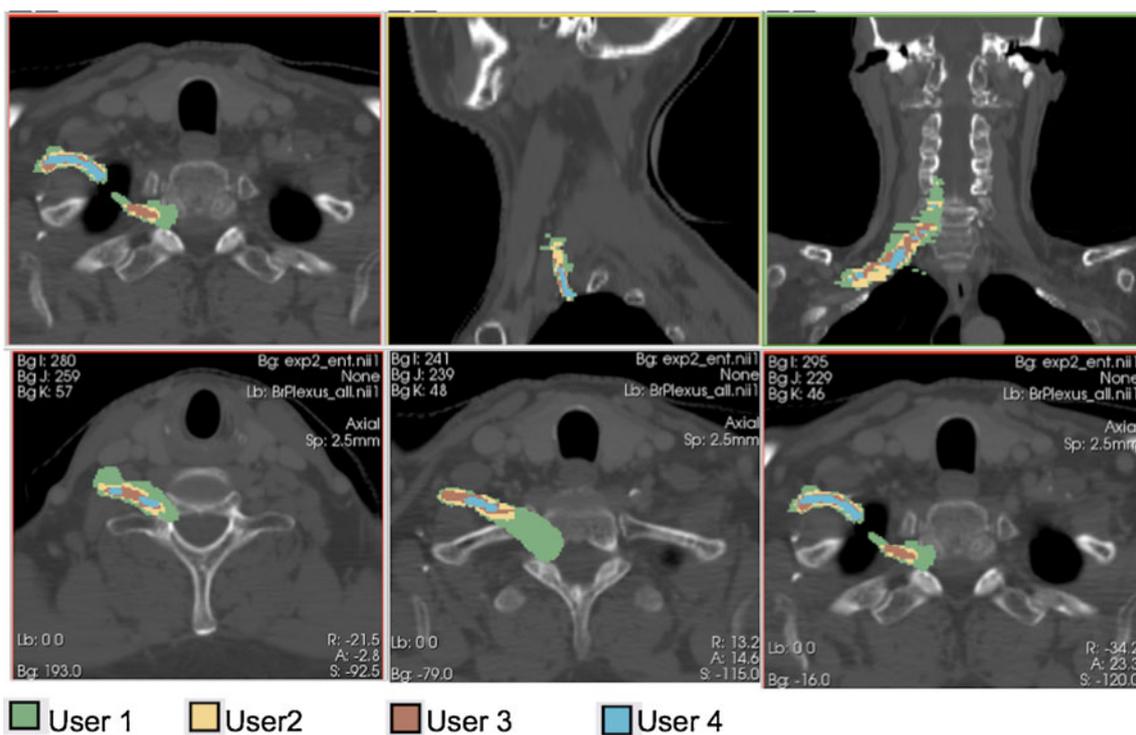
The HD is merely a function of the surface distance between the contours and is independent of the size of the volumes being contoured. The HD, although commonly used, is highly susceptible to outliers resulting from noisy data. Thus, more robust versions of this measure have been developed. A commonly used approach is the 95 % HD where 5 % of the outliers are discarded in identifying the surface distance [45–47].

Our base system has the capacity to calculate all of the previously described measures (volume difference, Dice coefficient, Jaccard coefficient, FPD and FND, and the 95 % HD) from submitted DICOM RTSTRUCT ROIs; however, the system can be configured by the system administrator to utilize one or more of these, depending on the aim of the application.

In addition to providing metrics of conformance between user-delineated ROIs, the system can also be used to develop consensus atlases by incorporating the contours from a number of observers (typically five to seven) using the method proposed by Warfield et al. [48]. The Simultaneous Truth and Performance Level Estimation (STAPLE) is an expectation–maximization algorithm that computes a probabilistic estimate of a “ground truth” segmentation given a set of manual reference contours. The system can be configured to create an additional set of “ground truth” contours using the STAPLE algorithm.

### Results

After website construction, an initial beta version was created for use with existing previously reported datasets [16]. Initial ROI analysis testing capability was performed using evaluative metrics (vide supra) on an existing dataset of rectal cancer TV ROIs. Our data confirm the results of the initial analysis (e.g., TV atlas use resulted in detectable alteration in TV delineation which more closely approximated expert and multi-user composite ROIs) [16].



**Fig. 8** Screen capture of axial, coronal, and sagittal slices, showing brachial plexus ROIs from four expert users

Having demonstrated the ROI analysis capacity of TaCTICS in an existing multi-institutional dataset, we have subsequently used TaCTICS in the initial pilot phase of a project designed to create a usable atlas for thymic malignancies. The International Thymic Malignancy Interest Group has recently created a series of guidelines regarding best practices for the management of thymic neoplasms [49]. As part of this effort, we have sought to provide technical support using TaCTICS. At this time, a pilot case has demonstrated the feasibility of TaCTICS for multi-institutional data collection and STAPLE consensus ROI generation, with data presentation at a scientific congress [50]. Our plans call for three to four additional cases to be included using multimodality imaging inputs, with possible subsequent expert atlas generation to follow.

A prospective interinstitutional implementation for consensus ROI generation has also been completed as part of an evaluation of contouring practices for stereotactic body radiotherapy [51]. Several expert user-derived OAR ROIs were used to create a composite ROI (Fig. 8) for comparing trainee and expert (Fig. 9) manual segmentation of brachial plexus ROIs, as well as similar analyses for tumor and cardiac ROIs (not shown).

Within the context of this study, a majority observer composite ROI, as well as a STAPLE composite ROI, was created using expert-defined ROIs (Fig. 9) and used as a formal comparator for novice users with access to an OAR atlas document. The resultant data were then used to compare OAR agreement. Data from this effort have been selected for conference presentation, with manuscript preparation under way [51]. A sample

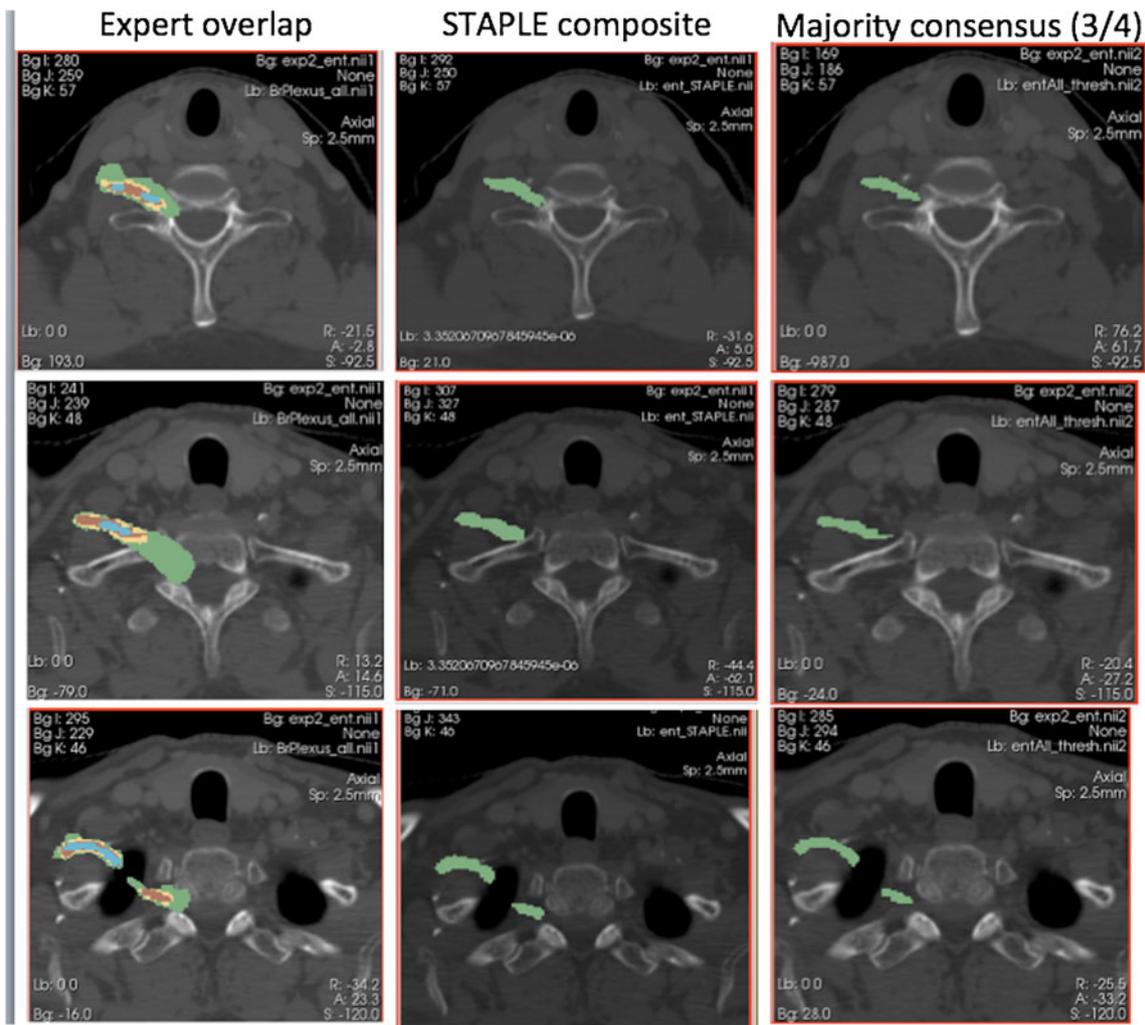
table showing calculated metrics comparing expert user brachial plexus ROI to the STAPLE composite set is seen in Table 1.

A second single-site study has also been performed, examining the effect not only of atlas implementation, but also of the utility of TaCTICS to provide real-time feedback using a credentialing model. In this study, radiation oncology residents were asked to contour OAR ROIs on a test case using their current personal practices. The residents were then given electronic atlas-based resources, as well as access to reference ROIs. Contours outside the reference and a predetermined margin could then be reviewed by the resident who could then assess whether their ROIs conformed to expected ROIs, exceeded a judicious safety margin (“caution” ROI), or fell outside an acceptability margin (“flagged” ROI). Data from this series have been accepted for conference presentation [52], with manuscript submission in progress.

## Discussion

As noted previously, target delineation optimization remains an elusive, though important, goal. In an effort to create an open-access, open-source solution, the TaCTICS software, as detailed previously, serves as an effective mechanism for multiple radiotherapy-related applications for a variety of stakeholders.

For educational institutions, TaCTICS may be readily implemented to serve as a repository for “teaching cases,”



**Fig. 9** Axial slices showing overlaid multi-expert ROIs (*left*), STAPLE composite (*middle*), and majority (*right*) consensus composite ROIs of brachial plexus ROI

with the added value that novice radiation oncologists might then have the capacity to recontour cases in order to self-assess agreement with attending TVs and OARs [21]. While visual inspection of overlapping ROIs is possible with commercial TPS platforms, the lack of a quantitative feedback feature (as in TaCTICS) limits the capacity for “self-scoring.” Though attending instruction one-on-one with a resident carefully assessing each ROI slice-by-slice is ideal, the real-time

feedback afforded by the ability to assess ROIs visually in TaCTICS, as well as by the numeric (Dice) score, gives TaCTICS a natural potential for practice quality improvement efforts in academic radiotherapy departments [21]. Additionally, the price structure (free open-source) is appropriate for cash-strapped academic departmental budgets.

For cooperative groups, TaCTICS has potential added value as well. Cooperative groups have recognized the need for

**Table 1** Metrics calculated by TaCTICS comparing expert brachial plexus agreement metrics as calculated by TaCTICS for ROI contours (see Fig. 9, left) to a reference STAPLE composite ROI (e.g., Fig. 9, middle)

User	Comparison set	Jaccard coefficient	Dice coefficient	FND	FPD	HD	Hausdorff average distance
Expert A	STAPLE	0.30	0.46	0.41	0.62	18.17	1.80
Expert B	STAPLE	0.90	0.95	0.10	0.00	5.00	0.07
Expert C	STAPLE	0.26	0.42	0.59	0.58	22.38	2.45

anatomic or trial-specific atlas generation [16, 24, 25, 28, 29, 53]. As such, many clinical trials will likely include a customized “study atlas” defining TV and OAR ROI parameters in an effort to standardize IMRT/IGRT applications [26, 29, 37, 54, 55]. TaCTICS provides the capacity for composite ROI generation using probabilistic coverage, majority voting, or ground truth estimation by STAPLE. Further, cooperative groups might potentially use features of TaCTICS for pretrial quality assurance, much in the manner of a “dummy run” [56, 57]. In this scenario, a test case and instructions would be sent to potential IMRT/IGRT trial participants. Resultant ROIs could then be compared to an expert composite standard. If a majority of users fail to contour ROIs within a prespecified conformance level or show systematic ROI variation in a meaningful manner, protocol modification (e.g., atlas inclusion, improved contouring instructions) might be employed. Alternatively, once instructions are codified, TaCTICS might be used to “flag” users who fail to contour within a prespecified range of ROIs on the test case [52]. One possible application is a “user-credentialing” step similar to that used to perform IMRT phantom quality assurance for dose delivery for current cooperative group trials [58, 59]. Additionally, contours derived from autosegmentation algorithms could feasibly be uploaded into the system and validated versus an expert set of contours for OARs or tumor volumes.

## Conclusions

In sum, our software has thus far demonstrated the capability for multiuser manually segmented ROI statistical analysis using a variety of metrics, to have the capacity for consensus ROI generation, and to be feasible for both intra-institutional and interinstitutional implementation in an accessible open-source platform. Already, several resultant prospective protocols are in various phases of inception, execution, and presentation. It is our hope that, by making our application and code available, end users will also choose to add their own customized modification to the software, serially improving the ability of TaCTICS to meet the individual practical or research needs of the radiotherapy community.

**Acknowledgments** CDF received support by a training grant from the National Institutes of Health/National Institute of Biomedical Imaging and Bioengineering, “Multidisciplinary Training Program in Human Imaging” (T32EB000817), the National Institutes of Health Clinician Scientist Loan Repayment Program (L30CA136381) the European Society for Therapeutic Radiology and Oncology Technology Transfer Grant, and the Society of Imaging Informatics in Medicine (SIIM) Product Development Grant. JKC is supported in part by a National Institutes of Health/National Library of Medicine Pathway to Independence Award (K99/R00LM009889) and NCI grant 5U01CA154601. These funders played no role in the study design, collection, analysis, and interpretation of data, manuscript writing, or decision to submit the report for

publication. Portions of this data were selected for a presentation at the SIIM 2010 Annual Meeting, 3–6 June, Minneapolis, MN, USA.

## References

1. Report 83: prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy (IMRT). *J ICRU* 10:1–106, 2010
2. Rosenthal DI, Chambers MS, Fuller CD, et al: Beam path toxicities to non-target structures during intensity-modulated radiation therapy for head and neck cancer. *Int J Radiat Oncol Biol Phys* 72:747–755, 2008
3. Garden AS, Lewin JS, Chambers MS: How to reduce radiation-related toxicity in patients with cancer of the head and neck. *Curr Oncol Rep* 8:140–145, 2006
4. Rosenthal DI, Blanco AI: Head and neck squamous cell carcinoma: optimizing the therapeutic index. *Expert Rev Anticancer Ther* 5:501–514, 2005
5. Van Kampen M, Levegrun S, Wannemacher M: Target volume definition in radiation therapy. *Br J Radiol* 70(Spec No):S25–S31, 1997
6. Fraass BA: The development of conformal radiation therapy. *Med Phys* 22:1911–1921, 1995
7. Lichter AS, Ten Haken RK: Three-dimensional treatment planning and conformal radiation dose delivery. *Important Adv Oncol* 1995: 95–109, 1995
8. Bortfeld T: IMRT: a review and preview. *Phys Med Biol* 51:R363–R379, 2006
9. Webb S: The physical basis of IMRT and inverse planning. *Br J Radiol* 76:678–689, 2003
10. Smith RP, Heron DE, Huq MS, et al: Modern radiation treatment planning and delivery—from Rontgen to real time. *Hematol Oncol Clin N Am* 20:45–62, 2006
11. Gregoire V, Jeraj R, Lee JA, et al: Radiotherapy for head and neck tumours in 2012 and beyond: conformal, tailored, and adaptive? *Lancet Oncol* 13:e292–e300, 2012
12. Schwartz DL, Garden AS, Thomas J, et al: Adaptive radiotherapy for head-and-neck cancer: initial clinical outcomes from a prospective trial. *Int J Radiat Oncol Biol Phys* 83:986–993, 2012
13. van Herk M: Errors and margins in radiotherapy. *Semin Radiat Oncol* 14:52–64, 2004
14. Multi-Institutional Target Delineation in Oncology Group: Human–computer interaction in radiotherapy target volume delineation: a prospective, multi-institutional comparison of user input devices. *J Digit Imaging* 24:794–803, 2011
15. Njeh CF: Tumor delineation: the weakest link in the search for accuracy in radiotherapy. *J Med Phys Assoc Med Physicists India* 33:136–140, 2008
16. Fuller CD, Nijkamp J, Duppen JC, et al: Prospective randomized double-blind pilot study of site-specific consensus atlas implementation for rectal cancer target volume delineation in the cooperative group setting. *Int J Radiat Oncol Biol Phys* 79:481–489, 2011
17. Eskander RN, Scanderbeg D, Saenz CC, et al: Comparison of computed tomography and magnetic resonance imaging in cervical cancer brachytherapy target and normal tissue contouring. *Int J Gynecol Cancer: Off J Int Gynecol Cancer Soc* 20:47–53, 2010
18. Peters LJ, O’Sullivan B, Giralt J, et al: Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. *J Clin Oncol: Off J Am Soc Clin Oncol* 28:2996–3001, 2010
19. Weber DC, Poortmans PM, Hurkmans CW, et al: Quality assurance for prospective EORTC radiation oncology trials: the challenges of

- advanced technology in a multicenter international setting. *Radiation Oncol: J Eur Soc Ther Radiol Oncol* 100:150–156, 2011
20. Okunieff P, Kachnic LA, Constine LS, et al: Report from the Radiation Therapy Committee of the Southwest Oncology Group (SWOG): Research Objectives Workshop 2008. *Clin Cancer Res: Off J Am Assoc Cancer Res* 15:5663–5670, 2009
  21. 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* 73: 416–423, 2009
  22. 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* 73:944–951, 2009
  23. Tai P, Van Dyk J, Yu E, et al: Variability of target volume delineation in cervical esophageal cancer. *Int J Radiat Oncol Biol Phys* 42:277–288, 1998
  24. Goodman KA, Regine WF, Dawson LA, et al: Radiation Therapy Oncology Group consensus panel guidelines for the delineation of the clinical target volume in the postoperative treatment of pancreatic head cancer. *Int J Radiat Oncol Biol Phys* 83:901–908, 2012
  25. Ng M, Leong T, Chander S, et al: Australasian Gastrointestinal Trials Group (AGITG) contouring atlas and planning guidelines for intensity-modulated radiotherapy in anal cancer. *Int J Radiat Oncol Biol Phys* 83:1455–1462, 2012
  26. Nijkamp J, de Haas-Kock DF, Beukema JC, et al: Target volume delineation variation in radiotherapy for early stage rectal cancer in the Netherlands. *Radiation Oncol: J Eur Soc Ther Radiol Oncol* 102: 14–21, 2012
  27. Vorwerk H, Hess CF: Guidelines for delineation of lymphatic clinical target volumes for high conformal radiotherapy: head and neck region. *Radiat Oncol* 6:97, 2011
  28. Wang D, Bosch W, Roberge D, et al: RTOG sarcoma radiation oncologists reach consensus on gross tumor volume and clinical target volume on computed tomographic images for preoperative radiotherapy of primary soft tissue sarcoma of extremity in Radiation Therapy Oncology Group studies. *Int J Radiat Oncol Biol Phys* 81: e525–e528, 2011
  29. Kong FM, Ritter T, Quint DJ, et al: Consideration of dose limits for organs at risk of thoracic radiotherapy: atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus. *Int J Radiat Oncol Biol Phys* 81:1442–1457, 2011
  30. Vorwerk H, Beckmann G, Bremer M, et al: The delineation of target volumes for radiotherapy of lung cancer patients. *Radiation Oncol: J Eur Soc Ther Radiol Oncol* 91:455–460, 2009
  31. Weiss E, Richter S, Krauss T, et al: Conformal radiotherapy planning of cervix carcinoma: differences in the delineation of the clinical target volume. A comparison between gynaecologic and radiation oncologists. *Radiation Oncol: J Eur Soc Ther Radiol Oncol* 67:87–95, 2003
  32. Allozi R, Li XA, White J, et al: Tools for consensus analysis of experts' contours for radiotherapy structure definitions. *Radiation Oncol: J Eur Soc Ther Radiol Oncol* 97:572–578, 2010
  33. Brouwer CL, Steenbakkers RJ, van den Heuvel E, et al: 3D variation in delineation of head and neck organs at risk. *Radiat Oncol* 7:32, 2012
  34. Rasch CR, Steenbakkers RJ, Fitton I, et al: Decreased 3D observer variation with matched CT-MRI, for target delineation in nasopharynx cancer. *Radiat Oncol* 5:21, 2010
  35. Steenbakkers RJ, Duppen JC, Fitton I, et al: Observer variation in target volume delineation of lung cancer related to radiation oncologist-computer interaction: a 'Big Brother' evaluation. *Radiation Oncol: J Eur Soc Ther Radiol Oncol* 77:182–190, 2005
  36. 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* 64:435–448, 2006
  37. Jansen EP, Nijkamp J, Gubanski M, et al: Interobserver variation of clinical target volume delineation in gastric cancer. *Int J Radiat Oncol Biol Phys* 77:1166–1170, 2010
  38. Roelofs E, Persoon L, Qamhiyeh S, et al: Design of and technical challenges involved in a framework for multicentric radiotherapy treatment planning studies. *Radiation Oncol: J Eur Soc Ther Radiol Oncol* 97:567–571, 2010
  39. Insight Toolkit: National Library of Medicine. <http://www.itk.org/>. Accessed 29 August 2013
  40. Jenkinson M: Nifti-1 data format. <http://nifti.nimh.nih.gov/nifti-1/>. Accessed 29 August 2013
  41. Dice LR: Measures of the amount of ecologic association between species. *Ecology* 26:297–302, 1945
  42. Jaccard P: The distribution of the flora in the alpine zone. *New Phytol* 11:37–50, 1912
  43. Crum WR, Camara O, Hill DL: Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans Med Imaging* 25:1451–1461, 2006
  44. Rogers DJ, Tanimoto TT: A computer program for classifying plants. *Science* 132:1115–1118, 1960
  45. Mitra J, Kato Z, Marti R, et al: A spline-based non-linear diffeomorphism for multimodal prostate registration. *Med Image Anal* 16:1259–1279, 2012
  46. Mitra J, Marti R, Oliver A, et al: Prostate multimodality image registration based on B-splines and quadrature local energy. *Int J CARS* 7:445–454, 2012
  47. Oguro S, Tuncali K, Elhawary H, et al: Image registration of pre-procedural MRI and intra-procedural CT images to aid CT-guided percutaneous cryoablation of renal tumors. *Int J CARS* 6:111–117, 2011
  48. Warfield SK, Zou KH, Wells WM: Simultaneous Truth and Performance Level Estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* 23:903–921, 2004
  49. Gomez D, Komaki R, Yu J, et al: Radiation therapy definitions and reporting guidelines for thymic malignancies. *J Thorac Oncol: Off Publ Int Assoc Study Lung Cancer* 6:S1743–S1748, 2011
  50. Thomas CR, Kalpathy-Cramer J, Senan S, et al: Development of an expert consensus target delineation atlas for thymic cancers: Initial results of an expert survey. International Thymic Malignancy Interest Group, Amsterdam, 2011
  51. Bongers EM, Dafele MR, Slotman BJ: Proton stereotactic body radiation therapy for clinically challenging cases of centrally and superiorly located stage I non-small-cell lung cancer: in regards to Register et al. *Int J Radiat Oncol Biol Phys* 82:492, author reply 492, 2012
  52. Kalpathy-Cramer J, Bedrick SD, Boccia K, et al: A pilot prospective feasibility study of organ-at-risk definition using Target Contour Testing/Instructional Computer Software (TaCTICS), a training and evaluation platform for radiotherapy target delineation. *AMIA Annu Symp Proc* 2011:654–663, 2011
  53. Feng M, Moran JM, Koelling T, et al: Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer. *Int J Radiat Oncol Biol Phys* 79:10–18, 2011
  54. Malone S, Croke J, Roustan-Delattour N, et al: Postoperative radiotherapy for prostate cancer: a comparison of four consensus guidelines and dosimetric evaluation of 3D-CRT versus tomotherapy IMRT. *Int J Radiat Oncol Biol Phys* 84:725–732, 2012
  55. Lim K, Small Jr, W, Portelance L, et al: Consensus guidelines for delineation of clinical target volume for intensity-modulated pelvic radiotherapy for the definitive treatment of cervix cancer. *Int J Radiat Oncol Biol Phys* 79:348–355, 2011
  56. Matzinger O, Poortmans P, Giraud JY, et al: Quality assurance in the 22991 EORTC ROG trial in localized prostate cancer: dummy run and individual case review. *Radiation Oncol: J Eur Soc Ther Radiol Oncol* 90:285–290, 2009

57. 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* 27:2965–2971, 2007
58. Clark CH, Miles EA, Urbano MT, et al: Pre-trial quality assurance processes for an intensity-modulated radiation therapy (IMRT) trial: PARSPORT, a UK multicentre phase III trial comparing conventional radiotherapy and parotid-sparing IMRT for locally advanced head and neck cancer. *Br J Radiol* 82:585–594, 2009
59. Pawlicki T, Yoo S, Court LE, et al: Process control analysis of IMRT QA: implications for clinical trials. *Phys Med Biol* 53:5193–5205, 2008