Prospective International Pilot Study Evaluating the Efficacy of a Self-Guided Contouring Teaching Module With Integrated Feedback for Transitioning From 2D to 3D Treatment Planning

Mustafa Abugideiri, MD[1]; Eduard Schreibmann, PhD[1]; Jeffrey Switchenko, PhD[1]; Mark W. McDonald, MD[1]; Jonathan J. Beitler, MBA, MD[1]; Walter J. Curran, MD[1]; Deborah Bruner, PhD[1]; Pretesh Patel, MD[1]; Wondemagegnhu Tigeneh, MD[2]; Miressa Mijena, PhD[2]; Sibo Tian, MD[1]; Anees Dhabaan, PhD[1]; Natia Esiashvili, MD[1]; Tian Liu, PhD[1]; and Arif N. Ali, MS, MD[1]

PURPOSE Transitioning from two-dimensional to three-dimensional treatment planning requires developing contouring skills. Contouring atlases are excellent resources, but they do not provide users active feedback. Developing countries may not have many radiation oncologists experienced in three-dimensional planning to provide training. We sought to develop a standardized self-guided educational module with integrated feedback to teach contouring skills.

METHODS AND MATERIALS All 18 oncology residents at Black Lion Hospital/Addis Ababa University in Ethiopia were trained to contour the level II lymph node station. Residents took a baseline pretest quiz, survey, and contouring evaluation. Residents then watched an instructional contouring lecture and performed three additional cases with integrated feedback by comparing their contours to gold-standard contours. Residents then took a post-training quiz, survey, and contouring evaluation. Paired *t* tests and analysis of variance were used for analysis.

RESULTS Before training, the average number of total cases ever contoured was 2.4 and the average number of head and neck cases contoured was 0.5. Comfort with contouring improved from being "not at all comfortable" to "quite comfortable" after the 3-hour training ($P < .001$). The standard deviation between the resident contours and gold standard improved from 72.6 cm^3 (pretest) to 7.4 cm^3 (post-test). The average percentage overlap with the gold-standard contours and Dice similarity coefficient improved with each case performed, from 27.7% and 0.26 (pretest) to 80.1% and 0.77 (post-test), respectively ($P < .001$). After training, 16 of 18 (88.9%) residents produced a Dice similarity coefficient greater than 0.7, the threshold generally accepted for excellent agreement.

CONCLUSION This self-guided teaching module was an effective tool for developing level II lymph node contouring skills by providing active feedback and resulted in improved user confidence and accuracy compared with a gold standard. This module can be expanded to other disease sites and countries to further facilitate transitioning to three-dimensional treatment planning in developing countries.

INTRODUCTION

With the advancement of technology in radiation oncology and more widespread installation of three-dimensional treatment planning equipment, there is increasing need for developing three-dimensional contouring skills. Despite the importance, standardization of volume delineation is still needed, and formal education on contouring is lacking and highly variable,[1-3] especially in the developing world.

Contouring may make up one of the largest sources of uncertainty in treatment planning.[1,4-8] Inadequate target coverage can lead to inferior tumor control,[9] and incorrect normal tissue delineation can lead to increased toxicity.[10] Despite availability of consensus contouring atlases, significant contouring variability is still an issue, as demonstrated by more than 80% of all submitted contours requiring revision and more than 45% requiring multiple revisions for patients enrolled in RTOG 0529 (ClinicalTrials.gov identifier: NCT00423293).[11]

National and international cooperative groups and organizations have developed contouring atlases and seminars to help improve accuracy and standardize contouring.[1,2,12,13] Contouring atlases are useful but do

CONTEXT

Key Objective

We developed and evaluated the feasibility and utility of a self-guided contouring teaching module with integrated feedback for the level II lymph node (L2LN) station in a developing country preparing to transition from two-dimensional to three-dimensional treatment planning.

Knowledge Generated

We showed objective improvements in target delineation, anatomic knowledge, and user confidence. At completion, nearly all participants independently contoured the L2LN station with excellent agreement compared with gold-standard contours.

Relevance

This protocol may offer an alternative for low-income countries to deploying their faculty members or residents internationally to learn contouring skills by providing a financially efficient and time-efficient mechanism for developing contouring skills.

not provide their users active feedback. Some contouring training programs provide contouring feedback to users,[14,15] and others do not.[3] In addition, contouring feedback may be given at the end of the training module without additional examination to assess for improvement after feedback.[13] It is a well-known concept in education that feedback is critical to more effective learning.[16,17] Training sessions that included feedback have shown good success.[1,15]

Because of high clinic volumes and busy physician schedules, it is becoming more difficult for attending physicians to provide contouring feedback to residents, despite its critical role in the educational process. A recent study showed approximately one-third of surveyed residents rarely received feedback on their contours,[2] and another showed less than 25% get formal contouring instruction.[3]

In developing countries with newly purchased modern equipment, there may not be many radiation oncologists experienced in three-dimensional planning to provide contouring feedback. Resources such as EduCase (RadOnc eLearning Center, Fremont, CA) offer excellent platforms where contours can be compared with a standard for feedback; however, these models are internet dependent.[2] Because of insufficient internet bandwidth in nearly all developing counties, these Web-based programs would not be feasible. To help address the challenges transitioning from two-dimensional to three-dimensional treatment planning, this study seeks to evaluate a standardized, internet-independent, self-guided educational contouring module with real-time integrated feedback to facilitate this transition and help improve contouring skills in developing countries.

METHODS AND MATERIALS

All 18 second-, third-, and fourth-year oncology residents at Black Lion Hospital/Addis Ababa University in Ethiopia were trained. The approximately 3-hour contouring training sessions were performed daily with two residents with the goal of learning how to contour the level II lymph node (L2LN) station. Head and neck (HN) lymph node (LN) stations were chosen because of their complexity, because HN contouring has been shown to have the greatest level of contouring variability[2,4,5] and also because of the availability of a standardized contouring atlas.[12] Residents performed a pretest survey that was modified from Gunther et al[13] to assess baseline demographic data (Table 1). Residents were also given a pretest quiz on the anatomic boundaries of the L2LN and a post-test quiz to evaluate improvement (Table 1). Our primary objective was to improve contouring accuracy of the L2LN, and secondary objectives were to improve user confidence and knowledge of anatomic boundaries.

Contouring Platform, Gold-Standard Contours, and Training Module Structure

Before training, a gold-standard L2LN contour was created. This contour was performed and verified by consensus of two US radiation oncology residents and two radiation oncology attending physicians against the international consensus guidelines.[12] All residents contoured using VelocityAI (Varian Medical Systems, Palo Alto, CA). Residents gained familiarity with the software platform by contouring a vertebral body with instruction on how to use the contouring tools. Once residents stated they felt comfortable with the tools, they were instructed to contour every third slice of the right L2LN as a baseline. Residents had 20 minutes to complete the contours and were given 10-, 5-, and 2-minute warnings. Once completed, contours were interpolated. No additional edits could be performed after the time limit. No feedback was given after the first case.

Residents then watched an instructional L2LN contouring lecture from the EduCase Web site performed by Robert Amdur, MD.[18] Residents were provided the HN LN consensus guidelines and atlas by Grégoire et al[12] for reference during and after the video lecture.

TABLE 1. Descriptive Statistics, Pretest and Post-Test Survey, and Quiz Results

Variable	Residents (N = 18)	%
Sex		
Male	15	83.3
Female	3	16.7
Year of residency training		
2	9	50.0
3	5	27.8
4	4	22.2
Previous contouring training		
Yes	5	27.8
No	13	72.2
No. of cases contoured		
0	13	72.2
2	1	5.6
3	1	5.6
7	1	5.6
16	2	11.1
No. of head and neck cases contoured		
0	15	83.3
1	1	5.6
4	2	11.1
No. of two-dimensional head and neck cases performed		
0	9	50.0
3	1	5.6
5	1	5.6
15	1	5.6
20	2	11.1
100	1	5.6
140	1	5.6
250	1	5.6
300	1	5.6
Comfort with English		
Extremely/quite	14	77.8
Moderately/slight/not	4	22.2
Comfort with identifying normal structures in an anatomic atlas for the head and neck		
Extremely/quite	8	44.4
Moderately/slight/not	10	55.6
Comfort with CT anatomy		
Extremely/quite	7	38.9
Moderately/slight/not	11	61.1
Comfort with identifying normal structures on a CT scan for the head and neck		
Extremely/quite	4	22.2
Moderately/slight/not	14	77.8

TABLE 1. Descriptive Statistics, Pretest and Post-Test Survey, and Quiz Results (Continued)

Variable	Residents (N = 18)	%
Comfort with contouring before intervention		
Extremely/quite	1	5.6
Moderately/slight/not	17	94.4
Before completing this module, how comfortable are you in your ability to contour level II lymph nodes of the head and neck?		
Extremely/quite	3	16.7
Moderately/slight/not	15	83.3
Before completing this module, how confident are you in your ability to find and use outside references including atlases and articles to aid in contouring the head and neck?		
Extremely/quite	6	35.3
Moderately/slight/not	11	64.7
Missing	1	—
Before completing this module, how confident are you in your knowledge of CT anatomy of level II lymph nodes of the head and neck?		
Extremely/quite	5	27.8
Moderately/slight/not	13	72.2
Before completing this module, how comfortable are you in your ability to use the contouring tools in the treatment planning software?		
Extremely/quite	1	5.6
Moderately/slight/not	17	94.4
Question 1: Pretest		
The superior boundary of the level II lymph node is at the level of the: (a) Base of skull; (b) Superior edge of the submandibular gland; (c) Inferior edge of the transverse process of C-1; (d) Inferior edge of the body of the hyoid bone		
Correct	1	5.6
Incorrect	17	94.4
Question 2: Pretest		
The anterior boundary of the level II lymph node is at the level of the: (a) Jugulodigastric node; (b) Posterior edge of the submandibular gland; (c) Spinal accessory nerve; (d) Anterior edge of the parotid gland		
Correct	7	38.9
Incorrect	11	61.1
Question 3: Pretest		
The posterior boundary of the level II lymph node is at the level of the: (a) Posterior edge of the sternocleidomastoid muscle; (b) Posterior edge of the spinous process of C-2; (c) Inferior edge of the transverse process of C-1; (d) Posterior surface of the internal jugular vein		
Correct	15	83.3
Incorrect	3	16.7
Question 4: Pretest		
The inferior boundary of the level II lymph node is at the level of the: (a) Inferior edge of the transverse process of C-1; (b) Inferior edge of the body of the hyoid bone; (c) Inferior edge of the submandibular gland; (d) Inferior edge of the cricoid cartilage		
Correct	17	94.4
Incorrect	1	5.6
Question 5: Pretest		
A medial boundary of the level II lymph node is the: (a) Medial surface of the sternocleidomastoid muscle; (b) Parotid gland; (c) Internal jugular vein; (d) Internal carotid artery		
Correct	5	27.8
Incorrect	13	72.2

(Continued on following page)

TABLE 1. Descriptive Statistics, Pretest and Post-Test Survey, and Quiz Results (Continued)

Variable	Residents (N = 18)	%
Total No. of questions correct on pretest quiz		
Median	2.50	—
Minimum	1	—
Maximum	4	—
Comfort with contouring after intervention		
Extremely/quite	15	83.3
Moderately/slightly/not	3	16.7
After completing this module, how comfortable are you in your ability to contour level II lymph nodes of the head and neck?		
Extremely/quite	11	61.1
Moderately/slight/not	7	38.9
After completing this module, how confident are you in your ability to find and use outside references including atlases and articles to aid in contouring the head and neck?		
Extremely/quite	11	61.1
Moderately/slight/not	7	38.9
After completing this module, how confident are you in your knowledge of CT anatomy of level II lymph nodes of the head and neck?		
Extremely/quite	15	88.2
Moderately/slight/not	2	11.8
Missing	1	—
After completing this module, how comfortable are you in your ability to use the contouring tools in the treatment planning software?		
Extremely/quite	13	76.5
Moderately/slight/not	4	23.5
Missing	1	—
Question 1: Post-test		
The superior boundary of the level II lymph node is at the level of the: (a) Base of skull; (b) Superior edge of the submandibular gland; (c) Inferior edge of the transverse process of C-1; (d) Inferior edge of the body of the hyoid bone		
Correct	16	88.9
Incorrect	2	11.1
Question 2: Post-test		
The anterior boundary of the level II lymph node is at the level of the: (a) Jugulodigastric node; (b) Posterior edge of the submandibular gland; (c) Spinal accessory nerve; (d) Anterior edge of the parotid gland		
Correct	18	100.0
Question 3: Post-test		
The posterior boundary of the level II lymph node is at the level of the: (a) Posterior edge of the sternocleidomastoid muscle; (b) Posterior edge of the spinous process of C-2; (c) Inferior edge of the transverse process of C-1; (d) Posterior surface of the internal jugular vein		
Correct	15	83.3
Incorrect	3	16.7
Question 4: Post-test		
The inferior boundary of the level II lymph node is at the level of the: (a) Inferior edge of the transverse process of C-1; (b) Inferior edge of the body of the hyoid bone; (c) Inferior edge of the submandibular gland; (d) Inferior edge of the cricoid cartilage		
Correct	17	94.4
Incorrect	1	5.6

TABLE 1. Descriptive Statistics, Pretest and Post-Test Survey, and Quiz Results (Continued)

Variable	Residents (N = 18)	%
Question 5: Post-test		
A medial boundary of the level II lymph node is the: (a) Medial surface of the sternocleidomastoid muscle; (b) Parotid gland; (c) Internal jugular vein; (d) Internal carotid artery		
Correct	14	77.8
Incorrect	4	22.2
Total No. of questions correct on post-test quiz		
Median	5	—
Minimum	3	—
Maximum	5	—
Age, years		
Median	28	—
Minimum	26	—
Maximum	39	—
Missing	0	—
No. of cases contoured		
Median	0	—
Minimum	0	—
Maximum	16	—
Missing	0	—
No. of head and neck cases contoured		
Median	0	—
Minimum	0	—
Maximum	4	—
Missing	0	—
No. of two-dimensional head and neck cases performed		
Mean	47.39	—
Median	1.50	—
Minimum	0	—
Maximum	300	—
SD	91.63	—
Comfort with English		
Question scale: (1) Extremely comfortable; (2) Quite comfortable; (3) Moderately comfortable; (4) Slightly comfortable; (5) Not at all comfortable		
Missing	0	—
Median	2	—
Minimum	1	—
Maximum	3	—
Missing	0	—
Comfort with identifying normal structures in an anatomic atlas for the head and neck		
Median	3	—
Minimum	1	—
Maximum	4	—
Missing	0	—

(Continued on following page)

TABLE 1. Descriptive Statistics, Pretest and Post-Test Survey, and Quiz Results (Continued)

Variable	Residents (N = 18)	%
Comfort with CT anatomy in general		
Median	3	—
Minimum	2	—
Maximum	4	—
Missing	0	—
Comfort with identifying normal structures on a CT scan for the head and neck		
Median	3	—
Minimum	2	—
Maximum	4	—
Missing	0	—
Comfort with contouring before intervention		
Median	5	—
Minimum	1	—
Maximum	5	—
Missing	0	—
Before completing this module, how comfortable are you in your ability to contour level II lymph nodes of the head and neck?		
Median	4	—
Minimum	2	—
Maximum	5	—
Missing	0	—
Before completing this module, how confident are you in your ability to find and use outside references including atlases and articles to aid in contouring the head and neck?		
Median	3	—
Minimum	1	—
Maximum	5	—
Missing	1	—
Before completing this module, how confident are you in your knowledge of CT anatomy of level II lymph nodes of the head and neck?		
Median	3	—
Minimum	1	—
Maximum	5	—
Missing	0	—
Before completing this module, how comfortable are you in your ability to use the contouring tools in the treatment planning software?		
Median	5	—
Minimum	1	—
Maximum	5	—
Missing	0	—
Comfort with contouring after intervention		
Median	2	—
Minimum	1	—
Maximum	3	—
Missing	0	—

TABLE 1. Descriptive Statistics, Pretest and Post-Test Survey, and Quiz Results (Continued)

Variable	Residents (N = 18)	%
After completing this module, how comfortable are you in your ability to contour level II lymph nodes of the head and neck?		
Median	2	—
Minimum	1	—
Maximum	3	—
Missing	0	—
After completing this module, how confident are you in your ability to find and use outside references including atlases and articles to aid in contouring the head and neck?		
Median	2	—
Minimum	1	—
Maximum	4	—
Missing	0	—
After completing this module, how confident are you in your knowledge of CT anatomy of level II lymph nodes of the head and neck?		
Median	2	—
Minimum	1	—
Maximum	3	—
Missing	1	—
After completing this module, how comfortable are you in your ability to use the contouring tools in the treatment planning software?		
Median	2	—
Minimum	1	—
Maximum	4	—
Missing	1	—
Dice (pretest)		
Mean	0.26	—
SD	0.25	—
Missing	0	—
Dice (case 1)		
Mean	0.64	—
SD	0.18	—
Missing	0	—
Dice (case 2)		
Mean	0.70	—
SD	0.098	—
Missing	0	—
Dice (case 3)		
Mean	0.71	—
SD	0.073	—
Missing	0	—
Dice (post-test)		
Mean	0.77	—
SD	0.046	—
Missing	0	—

(Continued on following page)

Variable	Residents (N = 18)	%
Change in Dice from pretest to case 1		
Mean	0.39	—
SD	0.26	—
Change in Dice from case 1 to post-test		
Mean	0.12	—
SD	0.18	—
Change in Dice from pretest to post-test		
Mean	0.51	—
SD	0.25	—

NOTE. Pretest and post-test questions 1 through 5 were based on Radiotherapy and Oncology 2014.[12]
Abbreviations: CT, computed tomography; SD, standard deviation.

After the lecture, residents contoured the right L2LN, as above, on the second case, with the guidelines and atlas as a reference. Once completed, they reviewed and compared their contours with the gold-standard contours on that specific case to get self-directed feedback, while also referring back to the atlas. Residents then performed two additional cases on the left L2LN, as above, with additional self-directed feedback on their contours after each case. After the fourth case, residents contoured the left L2LN without any resources as a post-test. A post-test quiz and survey were given to further assess the objective and subjective efficacy of the training. The pretest and post-test quiz questions were identical. Residents were given $25 gift cards on completion.

Statistics

Resident contours were evaluated against gold-standard contours for accuracy using the Dice similarity coefficient (DSC).[19] Average Hausdorff distance (HD),[20] Jaccard,[21,22] percentage overlap, false-negative error (FNE), and false-positive error (FPE) were also analyzed.[21,23]

DSC was estimated using the following formula:

$$DSC = \frac{2 \ x \ (A \cap B)}{A + B},$$

where A is the volume of the expert/standard, B is the volume of the given rater, and $A \cap B$ is the intersecting overlap of the two volumes. DSC scores range from 0.0 to 1.0, with 0 representing no overlap and 1 representing perfect overlap. A DSC of 0.7 or greater represents excellent agreement.[24] DSCs were measured at each time point, and change in DSC between pretest and post-test (Table 1), pretest and case 1 (Data Supplement), and case 1 and post-test were also estimated (Data Supplement). Paired t tests were used to assess for DSC differences and other similarity metrics between cases. Paired t tests were used to assess for differences in comfort level variables pre- versus postintervention, and McNemar tests were used to assess

for change in correct quiz question response rates pre- versus postintervention.

Resident characteristics, such as sex, age, year of residency, comfort with English, previous contouring training, number of cases contoured, number of HN cases contoured, self-reported comfort levels, and correct quiz question response rates, were summarized descriptively (Table 1). Resident characteristics were compared with change in DSC (post minus pre; Data Supplement) using analysis of variance or Kruskal-Wallis for categorical variables and Pearson's correlation coefficient for continuous variables. In addition, resident physician characteristics were compared with pretest DSC (Data Supplement) and post-test DSC (Data Supplement). A plot of DSC scores across time was also produced (Fig 1).
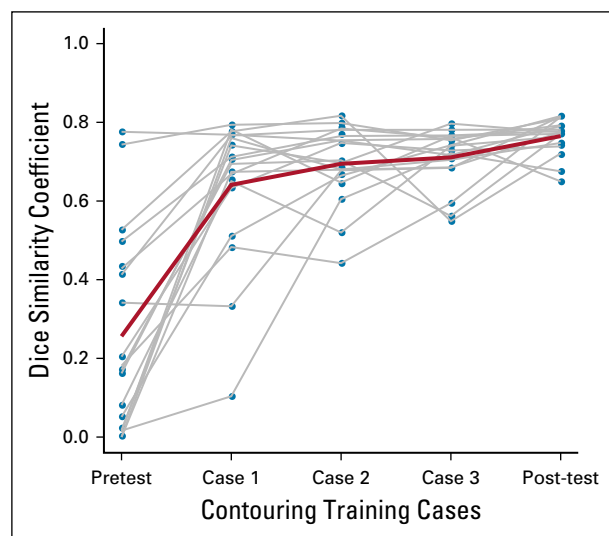

FIG 1. Plot of mean Dice similarity coefficient (DSC; red) and individual DSC of participants (gray) over time. Participants viewed a contouring didactic lecture between the pretest and case 1.

Statistical analysis was performed using SAS 9.4 (SAS Institute, Cary, NC). The significance level was assessed at the 0.05 level.

RESULTS

Study Participants

Eighteen oncology residents from Black Lion Hospital/ Addis Ababa University in Ethiopia completed the training module. The median age was 28 years (range, 26 to 39 years). Fifteen of the 18 participants (83.3%) were male. Nine (50%) were second-year residents, five (27.8%) were third-year residents, and four (22.2%) were in their fourth and final year of training. Five residents (27.8%) had previous contouring experience while performing an away rotation at another facility in Europe that used advanced three-dimensional treatment planning techniques; however, the average number of total cases ever contoured by all residents was 2.4, and the average number of HN cases contoured was 0.5. The median number of cases contoured was 0 for both total cases (range, 0 to 16) and HN cases (range, 0 to 4). Baseline demographic data are listed in Table 1.

Contouring Volume Analysis

The standard deviation (SD) between resident contour volumes and the gold standard improved from 72.6 cm^3 (pretest) to 7.4 cm^3 (post-test). The percentage overlap with the gold standard contours and DSC improved with each case performed from an average percentage overlap of 27. 7% and DSC of 0.26 on the pretest to 80.1% and 0.77, respectively, by the post-test ($P < .001$). After completing the module, 88.9% (16 of 18) of the residents produced a DSC greater than 0.7, and 87.5% of those residents (14 of 16) produced a DSC greater than 0.75. Other similarity measurement indices comparing pretest to post-test produced similar results, with mean HD improving from 43.7 to 17.4 mm, mean Jaccard improving from 0.17 to 0.62, mean FNE improving from 0.72 to 0.2, and mean FPE improving from 0.66 to 0.25 ($P < .001$; Table 2). Improvements in contouring are also illustrated in Figures 2 and 3.

Sex, year of residency training, comfort with English, self-reported comfort with computed tomography (CT) anatomy, comfort with LN anatomy, comfort with contouring in general, comfort with contouring L2LN, and previous HN contouring experience were not associated with differences in mean DSC improvement from pretest to post-test. There were, however, significant differences if residents had even limited previous contouring experience, with a mean DSC improvement of 0.59 for those without contouring experience compared with 0.29 for those with any contouring experience ($P = .019$; Data Supplement). On the contouring pretest, residents who did not have previous contouring training had a mean DSC of 0.17, which was lower than those who did have previous contouring training, with a mean DSC of 0.48 ($P = .018$; Data Supplement);

however, this difference disappeared on the post-test, with a mean DSC of 0.77 for those with previous contouring training and 0.76 for those without it ($P = .89$; Data Supplement).

Post-Training Evaluation

Resident confidence on the pretest with contouring in general and with contouring the L2LN improved from a median of 5 (not at all comfortable; range, 1 to 5) and 4 (slightly comfortable; range, 2 to 5), respectively, to a median of 2 (quite comfortable; range, 1 to 3) for both metrics after the teaching intervention ($P < .001$). On the pretest, only one resident self-reported feeling either extremely or quite comfortable with contouring, but this was likely erroneous, because they also reported no previous contouring experience. Residents were more comfortable using the contouring tools after the training, with an improvement in median scores from 5 (not at all comfortable; range, 1 to 5) to a median of 2 (quite comfortable; range, 1 to 4; $P < .001$). Statistically significant improvements were also noted in the self-assessed ability to find and use contouring atlases and comfort with CT anatomy of the L2LN stations after training (Table 3).

Residents showed improvement in multiple choice quiz scores on the L2LN anatomic borders, with mean and median pretest scores of 2.5 questions answered correctly out of five questions, which improved to a mean of 4.4 and median of five questions answered correctly after the intervention ($P < .001$).

DISCUSSION

We designed and evaluated a self-guided contouring module for the L2LN, which incorporated an instructional video lecture and three contouring cases with integrated feedback allowing users to compare their completed contours to a gold standard. After completion, improved user confidence in contouring and cumulative improvements in accuracy and decreased variability in contouring the L2LN were achieved. The greatest initial improvement occurred after the lecture, which provided a knowledge foundation to build on. The effectiveness of this protocol is underscored, given the fact that the majority of these residents had never contoured before and none had used the VelocityAI software.

Although long-term follow-up was limited to assess for lasting retention of skills and knowledge learned, the initial results of the study are quite impressive and promising. This protocol may offer an alternative for low-income countries to deploying their faculty members or residents internationally to learn contouring skills by providing a financially efficient and time-efficient mechanism for developing contouring skills. Completely automated software could be developed based on this teaching protocol. It could potentially be delivered to a country for on-site training, thus allowing all physicians in a department to be trained, instead of a select few. Because such

TABLE 2. Similarity Metric Improvement Pretest *v* Post-test

Covariate	Statistics	Case		Parametric *P**
		Pretest (n = 18)	Post-test (n = 18)	
Dice	No.	18	18	**< .001**
	Mean	0.26	0.77	
	Median	0.18	0.77	
	Minimum	0	0.65	
	Maximum	0.78	0.82	
	SD	0.25	0.05	
Hausdorff distance, mm	No.	18	18	**< .001**
	Mean	43.69	17.38	
	Median	43.01	15.29	
	Minimum	18.44	8.93	
	Maximum	69.85	58.81	
	SD	15.21	10.8	
Jaccard	No.	18	18	**< .001**
	Mean	0.17	0.62	
	Median	0.1	0.63	
	Minimum	0	0.48	
	Maximum	0.63	0.69	
	SD	0.2	0.06	
Mean overlap	No.	18	18	**< .001**
	Mean	0.26	0.77	
	Median	0.18	0.77	
	Minimum	0	0.65	
	Maximum	0.78	0.82	
	SD	0.25	0.05	
False-negative error	No.	18	18	**< .001**
	Mean	0.72	0.2	
	Median	0.86	0.2	
	Minimum	0.17	0.08	
	Maximum	1	0.3	
	SD	0.3	0.05	
False-positive error	No.	18	18	**< .001**
	Mean	0.66	0.25	
	Median	0.7	0.23	
	Minimum	0.19	0.11	
	Maximum	1	0.47	
	SD	0.25	0.1	
Percentage common	No.	18	18	**< .001**
	Mean	0.28	0.8	
	Median	0.14	0.8	
	Minimum	0	0.7	
	Maximum	0.83	0.92	
	SD	0.3	0.05	

NOTE. Boldface represents statistical significance.

Abbreviation: SD, standard deviation.

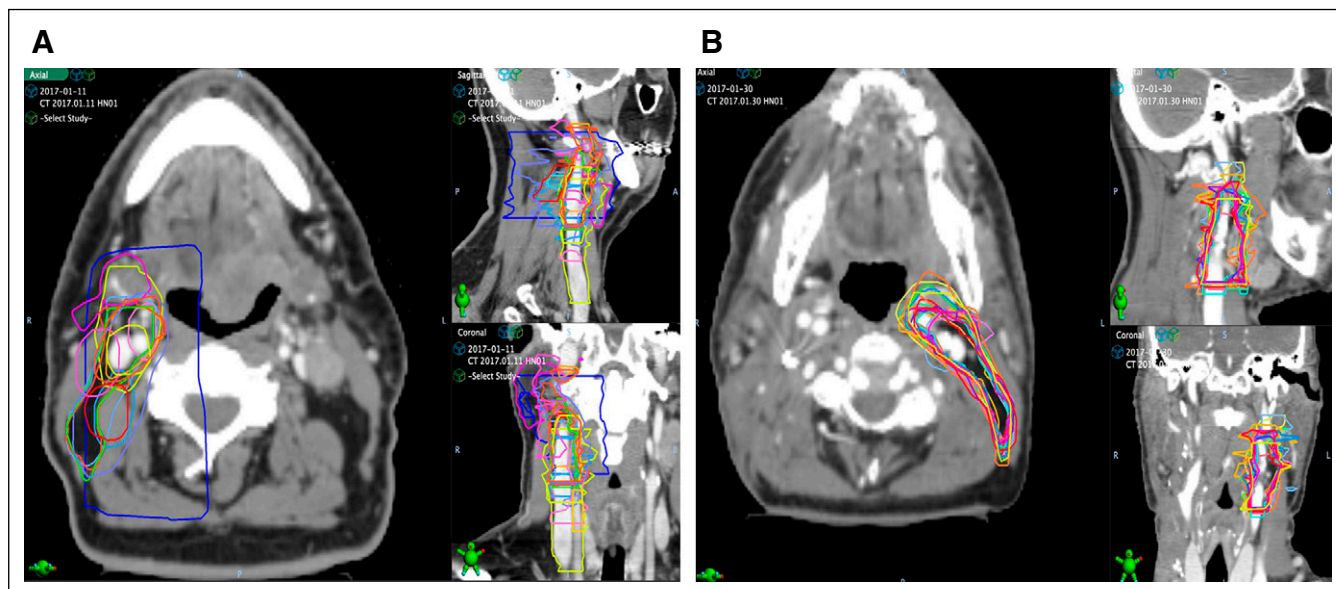*The parametric *P* value is calculated by a paired *t* test.

FIG 2. Selected slices of participant contours relative to gold standard (green). (A) Pretest contours. (B) Post-test contours.

automated software could be run any number of times on-site to achieve proficiency, physicians would not be forced to try to remember skills they learned when training in another country or go back for additional training. Feedback would be obtained on specific cases, as opposed to comparing the case at hand to a different case or contouring atlas and trying to extrapolate the results.

As part of a standardized curriculum, hospitals could acquire the software and training module. Once a certain level of proficiency was achieved, as determined by a contouring skills proficiency examination and objective examination of fund of knowledge, a certificate program could be developed.

With the limited feedback and lack of formalized contouring education domestically, this program could become a critical part of radiation oncology curricula. Gunther et al[14] tried to address the lack of standardization in radiation oncology curricula by developing the introductory radiation oncology curriculum. They found that surveyed residents only felt slightly prepared for residency before their institutional orientation and moderately prepared on completing orientation. After completion of the introductory radiation oncology curriculum, more hands-on training with contouring was requested.[14] A previous nationally delivered survey found only 11.3% of surveyed residents reported that their orientation was essential.[25] This module could supplement orientation providing better preparation for residency and enhancing traditional apprenticeship-based teaching to help provide more consistent and standardized resident educational experiences.

HN LNs were chosen because of their intricate anatomy and increased level of contouring difficulty, with the notion that if residents were able to achieve success contouring such complex structures, then the success of the module would likely be reproducible with less difficult structures. A recent survey showed that up to 30% of surveyed US radiation oncology residents did not feel comfortable identifying normal HN anatomy, and approximately one-third did not feel comfortable identifying at-risk nodal sites.[2] Another found only 27% of residents participating in an HN contouring seminar felt comfortable with target delineation before training.[26] Sura et al[2] showed that after undergoing an HN Webinar-based training, there was a nominal increase in user confidence in identifying the anatomy; however, it was not statistically significant.

In our study, residents showed objective, statistically significant improvements in multiple choice quiz scores on L2LN anatomic borders and contouring volumes assessed by DSC. Subjectively, user confidence in contouring the L2LN and CT anatomy of the L2LN also improved. This potentially supports the notion that incorporating feedback into the teaching protocol improves learning and, thus, confidence. Vega et al[27] also showed that using an interactive contouring session resulted in nominal improvements in pretest to post-test quiz scores; however, contouring proficiency was not examined.

Awan et al[15] used automated feedback as part of their training and showed statistically significant improvement in 16 of the 26 structures contoured. They noted a median DSC greater than 0.7 for four of 11 LNs after just one feedback session[15] and may have produced even higher DSC if more feedback sessions were provided. We provided three opportunities for feedback; by the completion of the module, 88.9% (16 of 18) of the residents produced a DSC greater than 0.7, and 87.5% of those residents (14 of 16) produced a DSC greater than 0.75. A DSC greater than 0.7
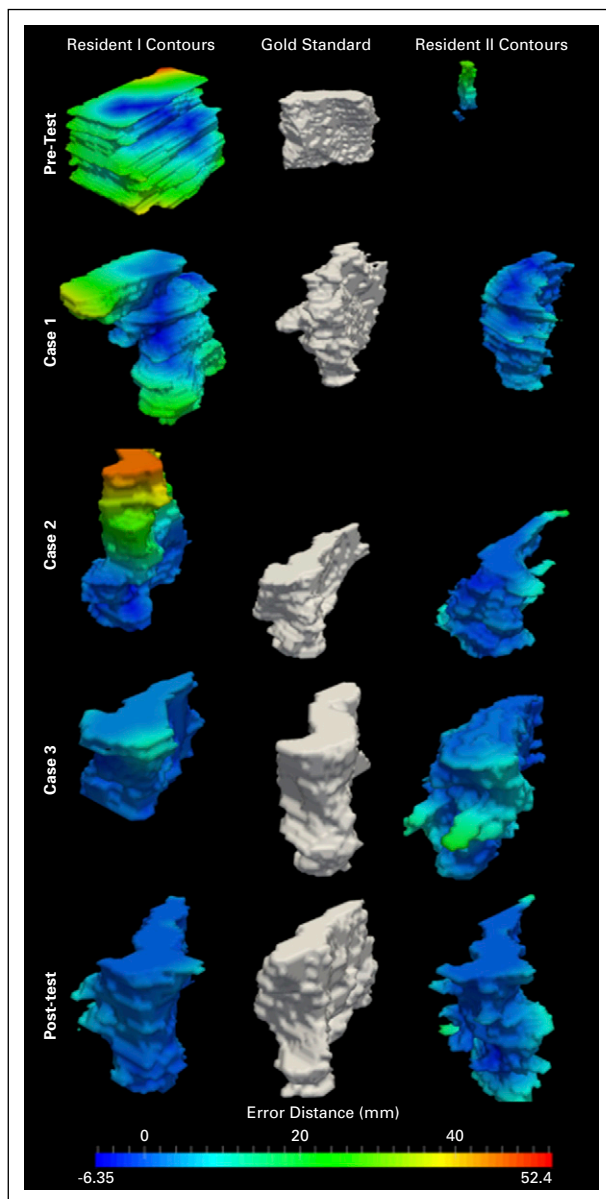
FIG 3. Demonstrates reduction in contouring error (millimeters) of resident sample cases (left and right images) relative to the gold standard (middle).

Differences in contouring may make up one of the largest sources of uncertainty in treatment planning,[1,4,5] although training can reduce variability.[1,6,26,28-30] Because of this, improvements in FPE and FNE are even more significant. FPE represents areas that are labeled positive by the user, but not the expert, whereas FNE measures areas deemed positive by the expert but missed by the user.[23] FPE would likely result in increased toxicity, because an area not deemed to be at risk by the expert was targeted for therapy; FNE would likely result in decreased tumor control, because a critical area that required targeting for therapy was missed. Both of these metrics showed statistically significant improvement from pretest to post-test, with a larger improvement in FNE, arguably the more important of the two.

DSC and Jaccard are the most commonly used measures of spatial overlap.[23] HD is commonly used but is more susceptible to outliers.[23] Because there is no consensus on the optimal measure or a standard method to evaluate inter-rater reliability,[31] we used a combination of statistical measures of agreement and overlap to provide to most clear picture of the accuracy and variability in target volume delineation. This makes our study unique, with the usage of numerous similarity metrics. Because contours were compared with a gold standard, our main focus was the DSC.

Limitations of this study include selection bias to those at Black Lion Hospital/Addis Ababa University and small sample size. Despite survey scores indicating residents were quite comfortable with English, there were likely some imperfections in question interpretation due to language. In addition, there was not a control group to see whether different teaching methods would produce similar results. Although a statistically significant improvement in DSC was also noted between case 1, when feedback was initiated, and the post-test (Table 1; Data Supplement), it was not as drastic as the improvement between the pretest and case 1, when residents watched the instructional lecture, making it difficult to discern whether the feedback or the lecture was more critical to learning or whether the entire module was essential. Our main objective was to evaluate the teaching module holistically, with both the didactic portion and feedback; however, future study could examine whether implementing feedback before the lecture, or without a lecture altogether, would produce similar results. We were also limited in our ability for long-term follow-up with testing for retention of knowledge and skills learned and assessment of the impact on clinical practice.

This study demonstrated the successful implementation of a self-guided contouring teaching protocol with standardized didactic lectures and integrated contouring feedback. Using this teaching intervention in a developing country is feasible and associated with objective improvements in L2LN target delineation, anatomic knowledge, and user confidence. After completion, nearly every participant was

corresponds with excellent agreement on the basis of the literature.[24] Although De Bari et al[1] showed improvement using feedback incorporated into their teaching model, it is difficult to discern the level of success, considering the preintervention DSC for their gross tumor volumes and clinical target volumes were both greater than 0.7. In addition, all contours were performed on the same case, which may have affected their results, whereas our study used five distinct cases. Another study showed improvement in contouring skills without dedicated feedback; however, only two of 18 of the post-test structures produced an average DSC greater than 0.7, and only one was statistically significant.[3]

TABLE 3. Pretest to Post-test Survey Results

Covariate	Statistics	Level	Intervention Before (n = 18)	Intervention After (n = 18)	Parametric *P**
Question 1	No. (%)	Correct	1 (5.56)	16 (88.89)	< .001
	No. (%)	Incorrect	17 (94.44)	2 (11.11)	
Question 2	No. (%)	Correct	7 (38.89)	18 (100)	< .001
	No. (%)	Incorrect	11 (61.11)	0 (0)	
Question 3	No. (%)	Correct	15 (83.33)	15 (83.33)	1.000
	No. (%)	Incorrect	3 (16.67)	3 (16.67)	
Question 4	No. (%)	Correct	17 (94.44)	17 (94.44)	1.000
	No. (%)	Incorrect	1 (5.56)	1 (5.56)	
Question 5	No. (%)	Correct	5 (27.78)	14 (77.78)	.003
	No. (%)	Incorrect	13 (72.22)	4 (22.22)	
Total No. of questions correct	No.		18	18	< .001
	Mean		2.5	4.44	
	Median		2.5	5	
	Minimum		1	3	
	Maximum		4	5	
Comfort with contouring	No.		18	18	< .001
	Mean		4.28	1.94	
	Median		5	2	
	Minimum		1	1	
	Maximum		5	3	
	SD		1.07	0.64	
How comfortable are you in your ability to contour level II lymph nodes of the head and neck?	No.		18	18	< .001
	Mean		3.89	2.33	
	Median		4	2	
	Minimum		2	1	
	Maximum		5	3	
	SD		1.13	0.59	
How confident are you in your ability to find and use outside references including atlases and articles to aid in contouring the head and neck?	No.		17	18	.010
	Mean		3.12	2.17	
	Median		3	2	
	Minimum		1	1	
	Maximum		5	4	
	SD		1.32	0.92	
How confident are you in your knowledge of CT anatomy of level II lymph nodes of the head and neck?	No.		18	17	.001
	Mean		3.11	1.94	
	Median		3	2	
	Minimum		1	1	
	Maximum		5	3	
	SD		1.13	0.56	

(Continued on following page)

TABLE 3. Pretest to Post-test Survey Results (Continued)

Covariate	Statistics	Level	Intervention Before (n = 18)	After (n = 18)	Parametric *P**
How comfortable are you in your ability to use the contouring tools in the treatment planning software?	No.		18	17	**< .001**
	Mean		4.28	2.18	
	Median		5	2	
	Minimum		1	1	
	Maximum		5	4	
	SD		1.07	0.73	

NOTE. Boldface represents statistical significance.

Abbreviations: CT, computed tomography; SD, standard deviation.

*The parametric *P* value is calculated by a paired *t* test for numerical covariates, McNemar test for two-level categorical covariates, and Bowker test of symmetry for categorical covariates with more than two levels.

able to independently contour the L2LN station with excellent agreement with the gold standard. Future work will be to develop fully automated, internet-independent software on the basis of this protocol and include gross tumor volumes with validation across all disease sites and expansion to other training locations with long-term follow-up. We will continue to explore using this protocol as a teaching tool internationally and domestically.

AFFILIATIONS
[1]Winship Cancer Institute of Emory University, Atlanta, GA
[2]Black Lion Hospital/Addis Ababa University, Addis Ababa, Ethiopia

The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CORRESPONDING AUTHOR
Mustafa Abugideiri, MD, 1365 Clifton Rd, AT 225, Atlanta, GA 30322; e-mail: MusAbu5@gmail.com.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST AND DATA AVAILABILITY STATEMENT
Disclosures provided by the authors and data availability statement (if applicable) are available with this article at DOI https://doi.org/10.1200/JGO.18.00224.

AUTHOR CONTRIBUTIONS
Conception and design: Mustafa Abugideiri, Mark W. McDonald, Walter J. Curran, Deborah Bruner, Wondemagegnhu Tigeneh, Natia Esiashvili, Tian Liu, Arif N. Ali
Provision of study material or patients: Wondemagegnhu Tigeneh, Miressa Mijena
Collection and assembly of data: Mustafa Abugideiri, Eduard Schreibmann, Miressa Mijena, Anees Dhabaan, Arif N. Ali
Data analysis and interpretation: Mustafa Abugideiri, Eduard Schreibmann, Jeffrey Switchenko, Mark W. McDonald, Jonathan J. Beitler, Walter J. Curran, Deborah Bruner, Pretesh Patel, Wondemagegnhu Tigeneh, Sibo Tian, Tian Liu, Arif N. Ali
Manuscript writing: All authors

Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST
The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jgo/site/misc/authors.html.

Eduard Schreibmann
Patents, Royalties, Other Intellectual Property: Entitled to royalties derived from Vision RT's sale of products. The terms of this agreement have been reviewed and approved by Emory University in accordance with its conflict of interest policies. (Inst)

Walter J. Curran
Stock and Other Ownership Interests: NantHealth
Honoraria: Bristol-Myers Squibb, AstraZeneca/MedImmune
Research Funding: AbbVie (Inst)

Pretesh Patel
Consulting or Advisory Role: Varian Medical Systems

Natia Esiashvili
Consulting or Advisory Role: Varian Medical Systems
Speakers' Bureau: Varian Medical Systems
Travel, Accommodations, Expenses: Varian Medical Systems

No other potential conflicts of interest were reported.

REFERENCES

1. De Bari B, Dahele M, Palmu M, et al: Short interactive workshops reduce variability in contouring treatment volumes for spine stereotactic body radiation therapy: Experience with the ESTRO FALCON programme and EduCase training tool. Radiother Oncol 127:150-153, 2018

2. Sura K, Lischalk JW, Leckie J, et al: Webinar-based contouring education for residents. J Am Coll Radiol 14:1074-1079.e3, 2017

3. Jaswal J, D'Souza L, Johnson M, et al: Evaluating the impact of a Canadian national anatomy and radiology contouring boot camp for radiation oncology residents. Int J Radiat Oncol Biol Phys 91:701-707, 2015

4. Segedin B, Petric P: Uncertainties in target volume delineation in radiotherapy - are they relevant and what can we do about them? Radiol Oncol 50:254-262, 2016

5. Njeh CF: Tumor delineation: The weakest link in the search for accuracy in radiotherapy. J Med Phys 33:136-140, 2008

6. Boersma LJ, Janssen T, Elkhuizen PHM, et al: Reducing interobserver variation of boost-CTV delineation in breast conserving radiation therapy using a pre-operative CT and delineation guidelines. Radiother Oncol 103:178-182, 2012

7. van Mourik AM, Elkhuizen PHM, Minkema D, et al: Multi-institutional study on target volume delineation variation in breast radiotherapy in the presence of guidelines. Radiother Oncol 94:286-291, 2010

8. Khoo ELH, Schick K, Plank AW, et al: Prostate contouring variation: Can it be fixed? Int J Radiat Oncol Biol Phys 82:1923-1929, 2012

9. Kong F, Ying H, Du C, et al: Patterns of local-regional failure after primary intensity modulated radiotherapy for nasopharyngeal carcinoma. Radiat Oncol 9:60, 2014

10. Furman MJ, Whalen GF, Shah SA, et al: Gastric perforation following stereotactic body radiation therapy of hepatic metastasis from colon cancer. Pract Radiat Oncol 3:40-44, 2013

11. Kachnic LA, Winter K, Myerson RJ, et al: RTOG 0529: A phase 2 evaluation of dose-painted intensity modulated radiation therapy in combination with 5-fluorouracil and mitomycin-c for the reduction of acute morbidity in carcinoma of the anal canal. Int J Radiat Oncol Biol Phys 86:27-33, 2013

12. Grégoire V, Ang K, Budach W, et al: Delineation of the neck node levels for head and neck tumors: A 2013 update. DAHANCA, EORTC, HKNPCSG, NCIC CTG, NCRI, RTOG, TROG consensus guidelines. Radiother Oncol 110:172-181, 2014

13. Gunther JR, Liauw SL, Choi S, et al: A prostate fossa contouring instructional module: Implementation and evaluation. J Am Coll Radiol 13:835-841.e1, 2016

14. Gunther JR, Jimenez RB, Yechieli RL, et al: Introductory radiation oncology curriculum: Report of a national needs assessment and multi-institutional pilot implementation. Int J Radiat Oncol Biol Phys 101:1029-1038, 2018 [Erratum: Int J Radiat Oncol Biol Phys 1012:1602, 2018]

15. Awan M, Kalpathy-Cramer J, Gunn GB, et al: Prospective assessment of an atlas-based intervention combined with real-time software feedback in contouring lymph node levels and organs-at-risk in the head and neck: Quantitative assessment of conformance to expert delineation. Pract Radiat Oncol 3:186-193, 2013

16. Wiggins G: Seven keys to effective feedback. Educational Leadership 70:10-16, 2012

17. Hattie J, Timperley H: The power of feedback. Rev Educ Res 77:81-112, 2007

18. Amdur RJ, Kaylor S: Contouring the Level 2 Nodal Station. Fremont, CA, RadOnc eLearning Center. www.Educase.com

19. Dice LR: Measures of the amount of ecologic association between species. Ecology 26:297-302, 1945

20. Huttenlocher DP, Klanderman GA, Rucklidge WJ: Comparing images using the Hausdorff distance. IEEE Trans Pattern Anal Mach Intell 15:850-863, 1993

21. Hanna GG, Hounsell AR, O'Sullivan JM: Geometrical analysis of radiotherapy target volume delineation: A systematic review of reported comparison methods. Clin Oncol (R Coll Radiol) 22:515-525, 2010

22. Jaccard P: Étude comparative de la distribution florale dans une portion des Alpes et des Jura [in French]. Bull Soc Vaud Sci Nat 37:547-579, 1901

23. Kalpathy-Cramer J, Fuller CD: Target Contour Testing/Instructional Computer Software (TaCTICS): A novel training and evaluation platform for radiotherapy target delineation. AMIA Annu Symp Proc 2010:361-365, 2010

24. Zijdenbos AP, Dawant BM, Margolin RA, et al: Morphometric analysis of white matter lesions in MR images: Method and validation. IEEE Trans Med Imaging 13:716-724, 1994

25. Diavolitsis VM, Shirazi H, Boyle J, et al: Residents' educational needs during transition into radiation oncology residency. J Am Coll Radiol 8:785-788, 2011

26. 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

27. Vega RM, Ishaq O, Rene LM, et al: A novel, pilot curriculum for international education of lymphoma management employing e-contouring. Int J Radiat Oncol Biol Phys 96:E414-E415, 2016

28. Dewas S, Bibault J-E, Blanchard P, et al: Delineation in thoracic oncology: A prospective study of the effect of training on contour variability and dosimetric consequences. Radiat Oncol 6:118, 2011

29. Gillespie EF, Panjwani N, Golden DW, et al: Multi-institutional randomized trial testing the utility of an interactive three-dimensional contouring atlas among radiation oncology residents. Int J Radiat Oncol Biol Phys 98:547-554, 2017

30. Thureau S, Oden S, Mokaouim R, et al: Évaluation de l'apport d'une méthode standardisée dans la définition du lit tumoral à l'aide de clips chirurgicaux dans le cancer du sein [in French]. Cancer Radiother 16:100-106, 2012

31. Fotina I, Lütgendorf-Caucig C, Stock M, et al: Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlenther Onkol 188:160-167, 2012

■ ■ ■