




















# 6 Potential of E-Learning Interventions and Artificial Intelligence–Assisted Contouring Skills in Radiotherapy: The ELAISA Study

Mathis Ersted Rasmussen, MD<sup>1</sup> ; Kamal Akbarov, MD, PhD<sup>2</sup> ; Egor Titovich, MSc<sup>2</sup> ; Jasper Albertus Nijkamp, PhD, MSc<sup>3</sup> ; Wouter Van Elmpt, PhD, MSc<sup>4</sup> ; Hanne Primdahl, MD, PhD<sup>5</sup>; Pernille Lassen, MD, PhD<sup>5</sup>; Jon Cacicedo, MD<sup>6</sup>; Lisbeth Cordero-Mendez, MD<sup>2</sup>; A.F.M. Kamal Uddin, MD, MBBS, DTC, FUICC, FIAEA<sup>7</sup>; Ahmed Mohamed, MD<sup>8</sup> ; Ben Prajogi, MD<sup>9</sup> ; Kartika Erida Brohet, MD<sup>10</sup> ; Catherine Nyongesa, MD<sup>11</sup>; Darejan Lomidze, MD, PhD<sup>12</sup>; Gisupnikha Prasiko, MD, MBBS, DLO<sup>13</sup>; Gustavo Ferraris, MD, PhD<sup>14</sup> ; Humera Mahmood, MD, MBBS<sup>15</sup>; Igor Stojkovski, MD, PhD<sup>16</sup> ; Isa Isayev, MD, PhD<sup>17</sup>; Issa Mohamad, MD<sup>18</sup> ; Leivon Shirley, MD, MBBS<sup>19</sup> ; Lotfi Kochbati, MD, PhD<sup>20</sup>; Ludmila Eftodiev, MD<sup>21</sup>; Maksim Piatkevich, MSc<sup>22</sup> ; Maria Matilde Bonilla Jara, MD<sup>23</sup>; Orges Spahiu, MD<sup>24</sup> ; Rakhat Aralbayev, MD<sup>25</sup>; Raushan Zakirova, MD<sup>26</sup> ; Sandya Subramaniam, MD<sup>27</sup> ; Solomon Kibudde, MD<sup>28</sup> ; Uranchimeg Tsegmed, MD<sup>29</sup>; Stine Sofia Korreman, MSc<sup>3</sup> ; and Jesper Grau Eriksen, MD, PhD<sup>1</sup> 

DOI <https://doi.org/10.1200/GO.24.00173>

## ABSTRACT

**PURPOSE** Most research on artificial intelligence–based auto–contouring as template (AI–assisted contouring) for organs–at–risk (OARs) stem from high–income countries. The effect and safety are, however, likely to depend on local factors. This study aimed to investigate the effects of AI–assisted contouring and teaching on contouring time and contour quality among radiation oncologists (ROs) working in low– and middle–income countries (LMICs).

**MATERIALS AND METHODS** Ninety–seven ROs were randomly assigned to either manual or AI–assisted contouring of eight OARs for two head–and–neck cancer cases with an in–between teaching session on contouring guidelines. Thereby, the effect of teaching (yes/no) and AI–assisted contouring (yes/no) was quantified. Second, ROs completed short–term and long–term follow–up cases all using AI assistance. Contour quality was quantified with Dice Similarity Coefficient (DSC) between ROs’ contours and expert consensus contours. Groups were compared using absolute differences in medians with 95% CIs.

**RESULTS** AI–assisted contouring without previous teaching increased absolute DSC for optic nerve (by 0.05 [0.01; 0.10]), oral cavity (0.10 [0.06; 0.13]), parotid (0.07 [0.05; 0.12]), spinal cord (0.04 [0.01; 0.06]), and mandible (0.02 [0.01; 0.03]). Contouring time decreased for brain stem (–1.41 [–2.44; –0.25]), mandible (–6.60 [–8.09; –3.35]), optic nerve (–0.19 [–0.47; –0.02]), parotid (–1.80 [–2.66; –0.32]), and thyroid (–1.03 [–2.18; –0.05]). Without AI–assisted contouring, teaching increased DSC for oral cavity (0.05 [0.01; 0.09]) and thyroid (0.04 [0.02; 0.07]), and contouring time increased for mandible (2.36 [–0.51; 5.14]), oral cavity (1.42 [–0.08; 4.14]), and thyroid (1.60 [–0.04; 2.22]).

**CONCLUSION** The study suggested that AI–assisted contouring is safe and beneficial to ROs working in LMICs. Prospective clinical trials on AI–assisted contouring should, however, be conducted upon clinical implementation to confirm the effects.

## ACCOMPANYING CONTENT

[Data Supplement](#)

Accepted July 10, 2024

Published September 5, 2024

JCO Global Oncol 10:e2400173

© 2024 by American Society of Clinical Oncology

Licensed under the Creative Commons Attribution 4.0 License

## BACKGROUND

Radiotherapy is a cornerstone in the treatment of cancer, with an estimated 50% of all patients with cancer needing radiotherapy at some point.<sup>1</sup> Radiotherapy has been found to be cost–effective across a variety of cancer types and sites and treatment regimens.<sup>2–6</sup> Despite these benefits,

the utilization of radiotherapy has been found to be sub–optimal across high–, middle–, and low–income countries globally.<sup>1,7,8</sup> Although there are many reasons for this, a major contributing factor is lack of trained clinical staff.<sup>1,9</sup> As the cancer incidence is expected to double in the coming decades,<sup>1</sup> there is an urgent need to develop and implement tools and strategies to reduce the patient–

## CONTEXT

### Key Objective

How does teaching and artificial intelligence (AI)–assisted contouring affect contouring quality and time in a global cohort of radiation oncologists working in low- and middle-income countries (LMICs) contouring organs-at-risk (OARs) for head-and-neck cancer?

### Knowledge Generated

AI-assisted contouring increased contouring quality compared with expert consensus contours regardless of whether teaching was received or not. Teaching increased contouring quality for only two OARs, but increased the time-saving effect of AI-assisted contouring.

### Relevance

AI-assisted contouring in combination with teaching of contouring guidelines is an effective strategy to reduce contouring time and conform contouring practices within and between radiotherapy departments located in LMIC.

specific workload. If the development of radiotherapy does not keep pace with the increasing demands, it may lead to further underutilization globally, which inevitably will harm patients.

Contouring is a time-consuming task for clinical staff<sup>10</sup> and is prone to observer variability,<sup>11,12</sup> but it is essential in modern radiotherapy. Auto-contouring has been studied intensely<sup>11</sup> and is known to reduce contouring time<sup>13,14</sup> and interobserver variation<sup>15–19</sup> across a variety of cancer sites, although manual editing is still required.<sup>20</sup> Today, auto-contouring is often artificial intelligence (AI)–based and is usually integrated into the clinical workflow as templates for contouring (AI-assisted contouring). AI-assisted contouring may serve as a part of the solution to the underutilization of radiotherapy by reducing the manual workload for clinical staff and thereby reducing the diagnosis-to-treatment times, which may improve patient outcome.<sup>21</sup> Most research on AI-assisted contouring, however, stem from high-income countries. This is a challenge in low- and middle-income countries (LMICs) since local factors such as patient abundance, clinical and financial resources, technical expertise, and mindsets of clinicians possibly influence how AI-assisted contouring is used. Hence, AI-assisted contouring should be evaluated in the context of LMICs to secure a safe and beneficial implementation worldwide.

The purpose of this study was to investigate how contouring quality and contouring time were affected among radiation oncologists (ROs) working in LMICs by (1) a single teaching session on contouring guidelines and AI-assisted contouring and (2) having AI-assisted contouring available. Subsequently, the effect of teaching (1) combined with AI-assisted contouring (2) was investigated after 2-week and 6-month follow-up periods. The study was a collaboration between the International Atomic Energy Agency (IAEA) and Aarhus University Hospital, Denmark.

## MATERIALS AND METHODS

### Institutions and Participants

Radiotherapy institutions were selected and enrolled by the IAEA according to the following criteria:

1. Located in an LMIC
2. Treating at least 20 patients with head-and-neck cancer per year
3. Performing computed tomography (CT)–based intensity-modulated radiotherapy
4. Able to enroll at least three ROs
5. Access to stable internet connection

Each institution appointed one chief scientific investigator who was in charge of recruiting ROs from the institution. There were no requirements for the ROs except they should have undergone training in head-and-neck contouring. The ROs completed a questionnaire regarding professional background and knowledge about AI-assisted contouring (Data Supplement, Fig S1). As a result, 23 radiotherapy institutions with 97 ROs were enrolled.

### Case Selection

The aim was to have as many patient cases as possible while having at least seven contour sets made per case. On the basis of the experience from previous studies by IAEA, the expected participant dropout rate was 50%. Therefore, optimally, 14 participants should be assigned to each case. Given the study design (see below) and the number of enrolled institutions and participants, this required 16 head-and-neck cancer cases that were provided by Aarhus University Hospital (Data Supplement, Table S1).

### Study Design and Random Assignment

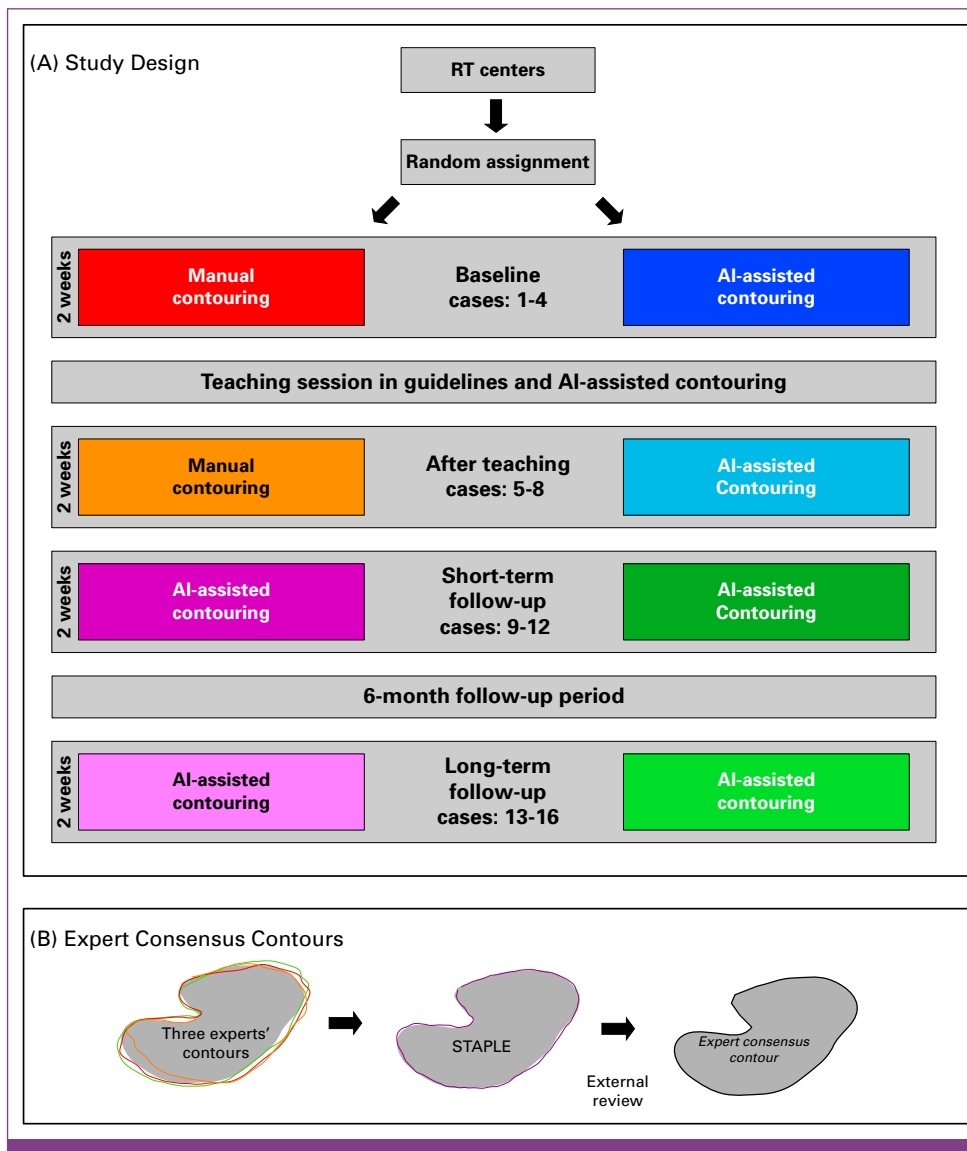
Institutions (including their participants) were randomly assigned to either the control group or the intervention

group. Random assignment was balanced on (1) institutions' annual number of patients with head-and-neck cancer and (2) whether any form of auto-contouring was available at the institution.

With 2 weeks to complete each round, participants were asked to contour eight organs-at-risk (OARs) on one case in each of the four sequential rounds (Fig 1):

1. Before the teaching session (baseline)
2. Immediately after the teaching session (after teaching)
3. Two weeks after the teaching session (short-term follow-up)
4. Six months after short-term follow-up (long-term follow-up)

The control group contoured manually at baseline and after teaching and did AI-assisted contouring in the short-term



**FIG 1.** (A) Overview of the study design. The left column (red, orange, dark purple, and light purple) represents the control group, and the right column (dark blue, light blue, dark green, and light green) represents the intervention group. Inside the colored boxes is the type of contouring used by the given study group in the specific round. Contouring rounds are marked as horizontal gray boxes, and the same four cases are used for both study groups inside these. (B) Overview of the method for obtaining expert consensus contours. (1) Three experts contoured all cases and structures independently, (2) the three contour sets were merged with the STAPLE algorithm and binary thresholded at 0.8, and (3) the final STAPLED structures were reviewed and edited for artifact by one of the initial contourers in consensus with an external head-and-neck expert oncologist. AI, artificial intelligence; RT, radiotherapy; STAPLE, Simultaneous Truth and Performance Evaluation.

and long-term follow-ups. The intervention group used AI-assisted contouring in all four rounds. In each round, four new patient cases were used (same cases in both groups). The cases were assigned randomly institution-wise, which resulted in 2-3 institutions (7-15 ROs) in each study group per case.

The effects of teaching and AI-assisted contouring were quantified in a two-by-two fashion with results at baseline and after teaching. To investigate whether the effects of teaching combined with AI-assisted contouring persisted over time, the short-term and long-term follow-ups were compared with the round after teaching within each group.

## Contouring

Contouring took place in EduCase (RadOnc eLearning Center, Inc, Jackson, WY), and the AI-contours were generated by EduCase professionals using Contour+, Guideline-Based Segmentation Solution (MVision AI Oy, Helsinki, Finland), which is based on well-defined contouring guidelines for brain stem<sup>22</sup> and for the spinal cord, oral cavity, mandible, right submandibular gland, right parotid gland, right optic nerve, and thyroid.<sup>23</sup> To imitate the clinical reality of many LMICs, only CT scans were available to the participants. Participants were instructed to generate clinically acceptable contours in accordance with the contouring guidelines. If deemed necessary, participants were allowed to delete AI contours and start over with manual contouring. The participants did not have access to the contours of others. Contours were handed in individually, but collaboration between participants was, however, not explicitly disallowed.

## Preprocessing of Contours

Digital Imaging and Communications in Medicine Structure Sets were exported from EduCase and converted to Neuroimaging Informatics Technology Initiative files with a voxel spacing of  $x = 0.39$  mm,  $y = 0.39$  mm, and  $z = 2$  mm (a scaling factor in X and Y of 3 times the CT grid).<sup>24,25</sup>

## Contour Quality

Expert consensus contours were generated in the following steps (Fig 1). First, three sets of contours were independently made by three head-and-neck expert oncologists (J.G.E., H.P. and P.L.) without access to the AI contours used in this study. These were merged using Simultaneous Truth and Performance Evaluation (STAPLE).<sup>26</sup> The STAPLE maps were binarized with a threshold of 0.8. The binary STAPLE structures were then reviewed and corrected for artifacts by an external head-and-neck expert oncologist (J.C.) in consensus with J.G.E. Contouring quality was quantified using the Dice Similarity Coefficient (DSC) and Hausdorff Distance 95th percentile (HD95) between participants' contours and expert consensus contours. Increasing DSC and decreasing HD95 indicate increasing

agreement with the expert consensus contours and thus higher contouring quality.

The participants were blinded toward the expert consensus contours throughout the entire study.

## Contouring Time

Contouring time was defined as the time of active contouring (mouse click-and-hold) with any contouring tool. For all contouring interactions, duration in milliseconds, type of interaction, active structure, and participant name were automatically recorded by EduCase. The durations of interactions were summed over structures and are reported in minutes.

## Statistics and Software

Data were assumed to be nonparametric. Therefore, effect sizes were quantified as absolute differences of medians with 95% CIs. This is formatted as: Absolute Difference [CI low; CI high]. CIs were estimated with percentile bootstrapping in 9,999 iterations. Absolute differences are positive when the reference value is the smaller number and vice versa. Data handling was performed in Python 3.9, and DSC and HD95 were calculated using MedPy.<sup>27</sup> Statistics and bootstrapping were performed with SciPy.<sup>28</sup>

## RESULTS

Of the 97 participating ROs, 94 completed the questionnaire on professional background. The random assignment resulted in 11 institutions for the control group and 12 for the intervention group. The characteristics of institutions and participants are found in Table 1. The four rounds of contouring were completed by 89 (92%), 91 (94%), 93 (96%), and 80 (82%) ROs, respectively (Fig 2).

The raw results of contouring quality and contouring time are shown for each group at baseline in Figure 3 and for all contouring rounds in Figure 4. Below is a walk-through of the estimated effect sizes of the four combinations of exposures along with the effect of combining teaching and AI-assisted contouring in the follow-up rounds (Fig 5). Visual comparisons between manual contouring and AI-assisted contouring are provided in the Data Supplement (Fig S2).

## AI-Assisted Contouring Without Teaching

By comparing the two study groups at baseline (Fig 3), the effect of AI-assisted contouring without teaching was quantified (Fig 5, top, circles). This showed a higher (absolute) median DSC with AI-assisted contouring of 0.05 [0.01; 0.10] (CI) for optic nerve, 0.10 [0.06; 0.13] for oral cavity, 0.07 [0.05; 0.12] for parotid, 0.04 [0.01; 0.06] for spinal cord, and 0.02 [0.01; 0.03] for mandible. Correspondingly lower HD95 was only observed for oral cavity and parotid. Contouring times were lower with AI-assisted contouring by  $-1.41$  [ $-2.44$ ;  $-0.25$ ] minutes for brain stem,  $-6.60$  [ $-8.09$ ;  $-3.35$ ]

**TABLE 1.** Baseline Characteristics of Enrolled Institution and Participants

Characteristic	Control	Intervention	Total
Institutions, No.			
No. of centers	11	12	23
Participants invited	45	52	97
Annual cases: 21-50	2	3	5
Annual cases: 51-100	2	2	4
Annual cases: 101-200	4	3	7
Annual cases: 201+	3	4	7
Dedicated head-and-neck unit	7	6	13
Use any contouring guidelines	10	11	21
Use any form of auto-contouring	6	6	12
Participants, No.			
Completed questionnaire	44	50	94
Median working years [95% CI]	8 [7 to 12]	8 [4 to 12]	–
Head-and-neck among specialties	13	14	27
Regular user of auto-contouring	20	20	40
No knowledge of auto-contouring	6	7	13
Basic knowledge of auto-contouring	26	29	55
Intermediate knowledge of auto-contouring	9	9	18
Advanced knowledge of auto-contouring	3	5	8
Active in research and development of auto-contouring	6	6	12

minutes for mandible,  $-0.19$  [ $-0.47$ ;  $-0.02$ ] minutes for optic nerve,  $-1.80$  [ $-2.66$ ;  $-0.32$ ] minutes for parotid, and  $-1.03$  [ $-2.18$ ;  $-0.05$ ] minutes for thyroid.

### AI-Assisted Contouring With Teaching

To address the combined effect of teaching and AI-assisted contouring, the two study groups were compared after teaching (Fig 5, top, triangles). The tendency was similar to that of AI-assisted contouring without teaching, with median DSC being higher with AI-assisted contouring by  $0.03$  [ $0.01$ ;  $0.05$ ] for mandible,  $0.05$  [ $-0.01$ ;  $0.14$ ] for optic nerve,  $0.08$  [ $0.05$ ;  $0.10$ ] for oral cavity,  $0.07$  [ $0.06$ ;  $0.10$ ] for parotid,  $0.01$  [ $0.00$ ;  $0.04$ ] for spinal cord,  $0.03$  [ $0.01$ ;  $0.04$ ] for submandibular, and  $0.02$  [ $0.00$ ;  $0.03$ ] for thyroid. HD95 was only lower for oral cavity and parotid. Contouring time was lower with AI-assisted contouring by  $-1.76$  [ $-3.26$ ;  $-0.93$ ] minutes for brain stem,  $-9.41$  [ $-11.36$ ;  $-6.71$ ] minutes for mandible,  $-0.21$  [ $-0.46$ ;  $-0.04$ ] minutes for optic nerve,  $-3.28$  [ $-5.74$ ;  $-1.99$ ] minutes for oral cavity,  $-2.50$  [ $-3.06$ ;  $-1.56$ ] minutes for parotid,  $-2.65$  [ $-4.30$ ;  $-1.13$ ] minutes for spinal cord,  $-0.78$  [ $-1.08$ ;  $-0.24$ ] minutes for submandibular, and  $-2.35$  [ $-3.39$ ;  $-1.11$ ] minutes for thyroid.

### Teaching Without AI-Assisted Contouring

The effect of teaching without AI-assisted contouring was quantified by comparing the round after teaching with

baseline in the control group (Fig 5, top, crosses). In this comparison, median DSC increased after teaching with  $0.05$  [ $0.01$ ;  $0.09$ ] for oral cavity and  $0.04$  [ $0.02$ ;  $0.07$ ] for thyroid. The differences for the remaining OARs were either small or had inconclusive CIs. Contouring time increased after teaching by  $2.36$  [ $-0.51$ ;  $5.14$ ] minutes for mandible,  $1.42$  [ $-0.08$ ;  $4.14$ ] minutes for oral cavity, and  $1.60$  [ $-0.04$ ;  $2.22$ ] minutes for thyroid. Contouring time was unchanged for the remaining OARs.

### Teaching With AI-Assisted Contouring

For AI-assisted contouring, the effect of teaching was quantified by comparing the round after teaching with baseline in the intervention group (Fig 5, top, horizontal bars). For DSC and HD95, the tendency was similar to teaching without AI-assisted contouring, with increased median DSC after teaching of  $0.03$  [ $0.01$ ;  $0.04$ ] for oral cavity and  $0.07$  [ $0.04$ ;  $0.10$ ] for thyroid. However, contouring time decreased after teaching with  $-1.10$  [ $-2.26$ ;  $-0.10$ ] minutes for oral cavity and with  $-1.03$  [ $-2.18$ ;  $0.05$ ] minutes for thyroid, whereas it was unchanged for the remaining OARs.

### Teaching and AI-Assisted Contouring Over Time

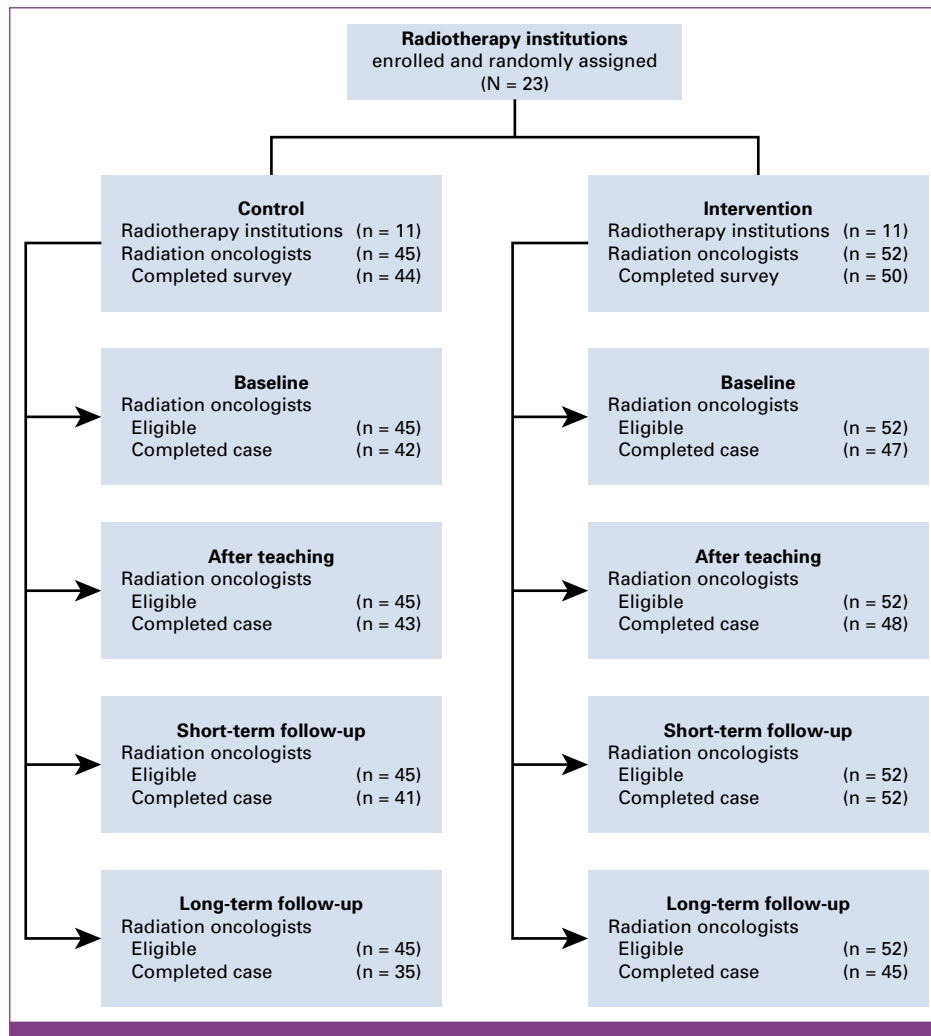
To address whether the effect of teaching combined with AI-assisted contouring persisted over time, the short-term and long-term follow-ups were compared within the study groups with the round after teaching (Fig 5, bottom). For most OARs, the effect sizes were similar in short-term and long-term follow ups, corresponding to a persistent effect over time. However, median DSC was substantially lower for thyroid at long-term follow-up in both study groups.

### Acceptance of AI Contours

Across all submissions with AI-assisted contouring, 335 (24%) AI contours were accepted without editing by a total of 66 (68%) participants. Among these participants, the median number of accepted contours was three per case. Out of the accepted-as-is contours, 100 (30%) contours had a higher DSC than the group average for the specific case and study arm.

## DISCUSSION

This study investigated the effects of teaching and AI-assisted contouring in a large randomized study on ROs from LMICs. The dropout rate was much lower than expected, and therefore, the study was well powered for the research questions. Regardless of contouring method, teaching improved contouring quality for two OARs. Regardless of teaching, AI-assisted contouring increased contouring quality and reduced contouring time for most investigated OARs. The combined effect of teaching and AI-assisted contouring persisted throughout the two follow-up rounds. The study thereby confirms previous research on

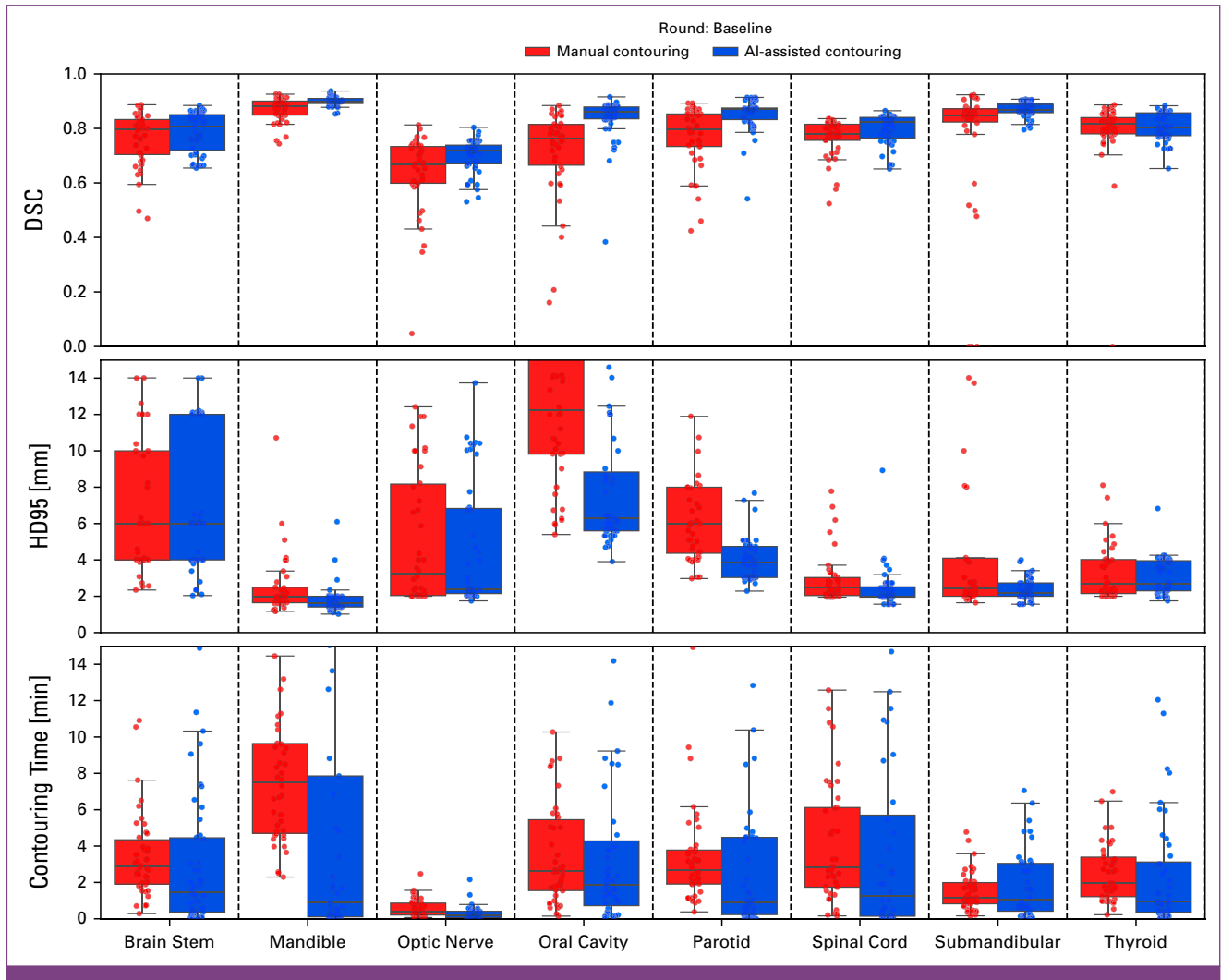


**FIG 2.** A CONSORT diagram for the study participants. Note that missing one submission did not exclude participants from participating later. Therefore, the number of eligible radiation oncologists is the same throughout all study rounds.

time savings<sup>29-32</sup> and reduction in interobserver variability<sup>33</sup> obtained by auto-contouring.

Previous research has shown that teaching improves contour consistency and quality.<sup>34,35</sup> It was therefore somewhat surprising that teaching did not affect contour quality more. This is probably due to participants already contouring well at baseline, and thus there was little room for improvement with the applied metrics. This explanation is supported by the fact that oral cavity was one of the structures that improved with teaching in both study groups. Contouring of oral cavity heavily relies on the guideline definitions. Hence, the increase in contouring quality and contouring time in the control group may indicate that teaching in fact was effective, as the participants spent more time contouring higher quality structures. For the intervention group, this manifested itself with an increase in contouring quality and a drop in contouring time, which could be due to AI contours being in high accordance with the contouring guidelines.

Therefore, fewer adaptations were required with the participants' updated knowledge. Besides this, it can, however, not be ruled out that (1) a single teaching session is not enough to change contouring practice of the participants, (2) the rotation to new cases after the teaching session could make contouring easier/harder, and (3) the metrics were not sensitive to subtle differences in contours. The results are, however, good news as the high baseline quality may be attributed to the recent years' efforts in teaching programs and implementation of contouring guidelines. Although teaching did not increase contouring quality as much as expected, teaching modified the time savings of AI-assisted contouring; for most OARs, AI-assisted contouring alone reduced contouring time (crosses, Fig 5), but even larger time savings were observed when AI-assisted contouring was combined with teaching (lines, Fig 5). In effect, this means that similar levels of contouring quality were obtained faster with AI-assisted contouring when accompanied by teaching.



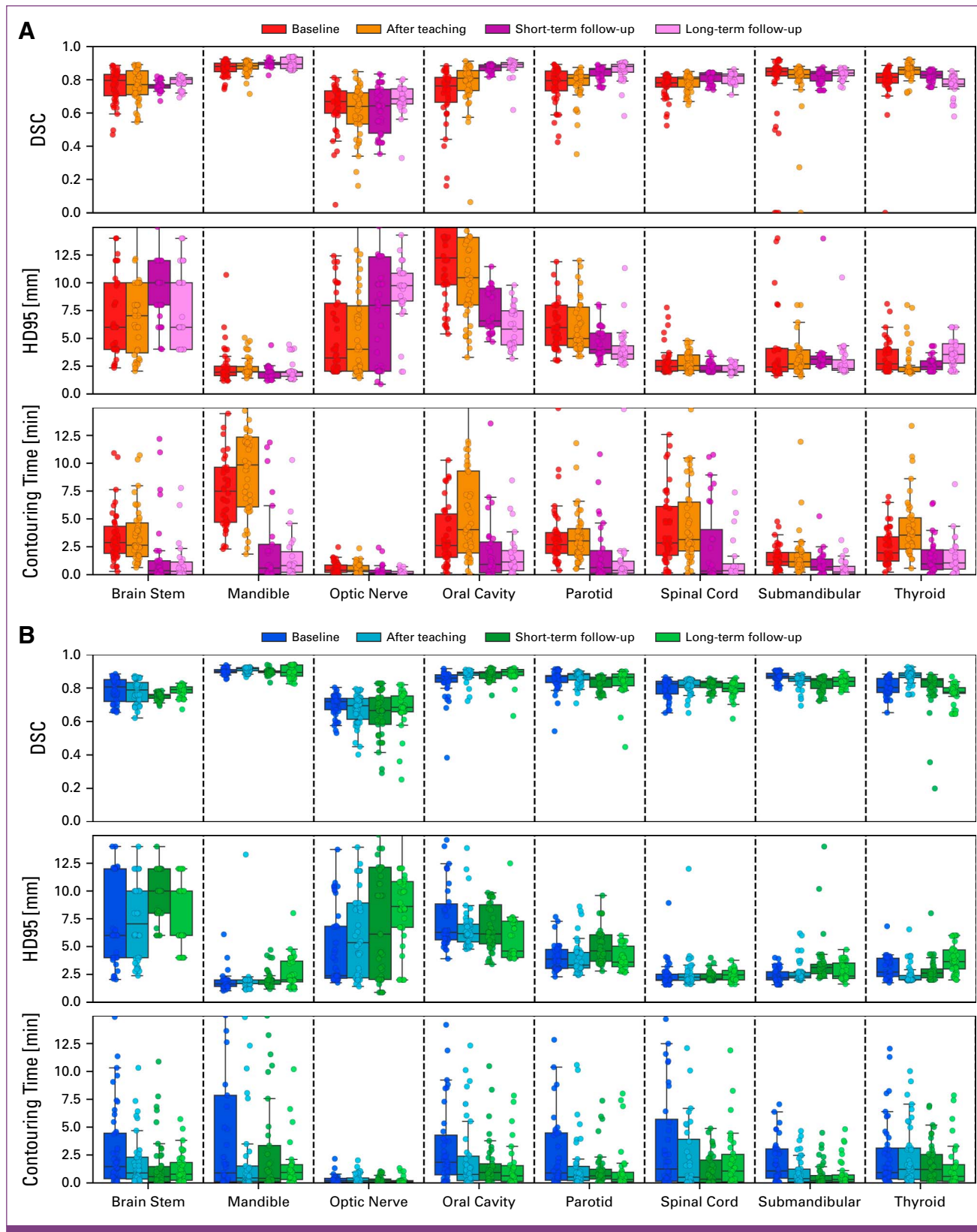
**FIG 3.** The results at baseline of the two groups. In red is the control group contouring manually, and in blue is the intervention group doing AI-assisted contouring. Each circle represents a single contour from a single participant, and boxplots are based on these. The top and middle rows show DSC and HD95 between participants' contours and expert consensus structures, respectively. The bottom row shows the contouring time recorded by the contouring platform. AI, artificial intelligence; DSC, Dice Similarity Coefficient; HD95, Hausdorff Distance 95th percentile.

Given the inevitable interobserver variability that also exists between experts, the consensus structures were considered to be of the highest quality obtainable. It was therefore quite extraordinary that AI-assisted contouring enabled participants to make structures with higher similarity to the expert consensus contours compared with manual contouring. Although the consensus contours were generated independently of the AI contours, they were in high agreement (Data Supplement, Fig S3). This suggested that AI-assisted contouring may be an effective strategy not only to conform contouring practices between individuals but even facilitate the implementation of—and adherence to—contouring guidelines across countries and continents.

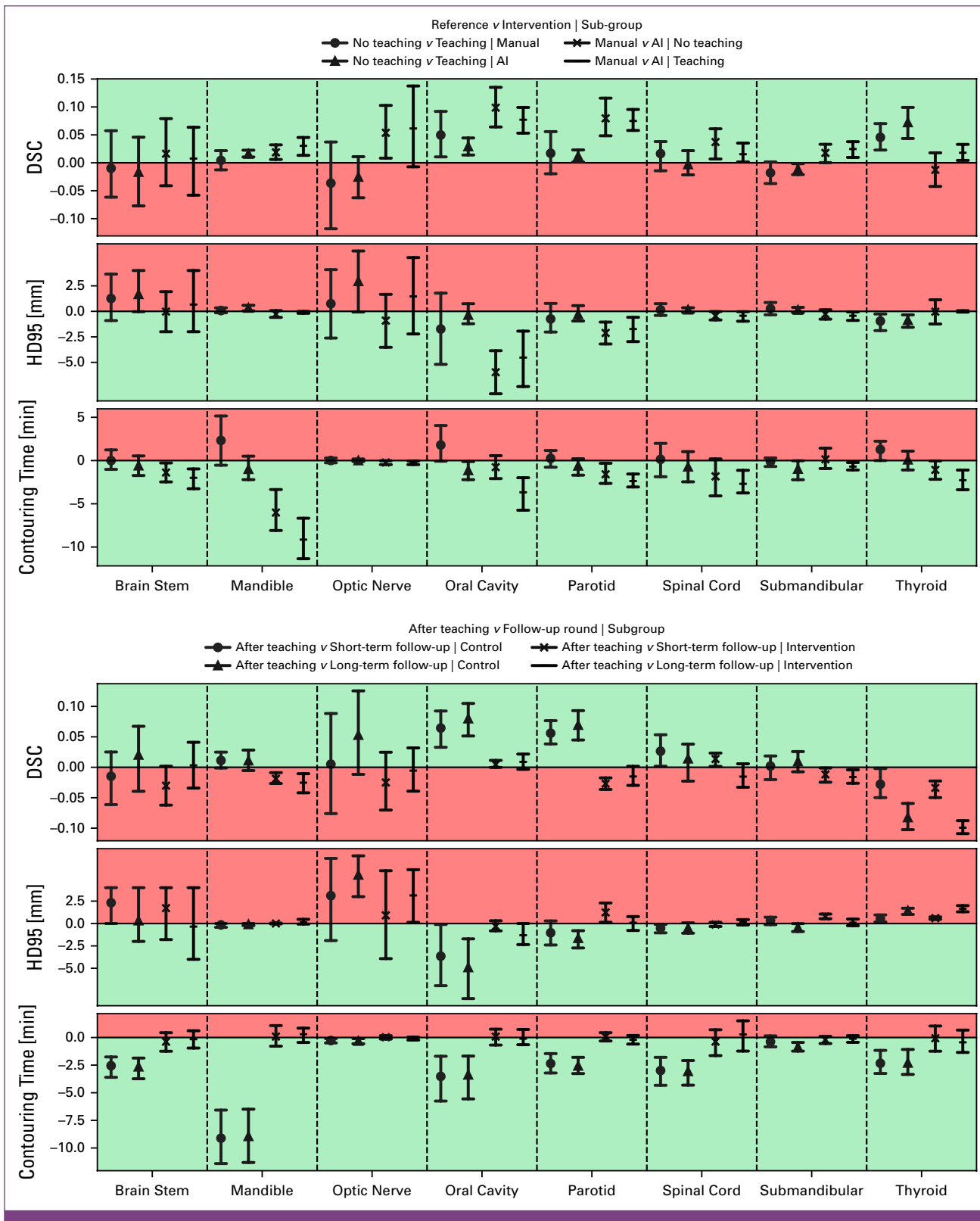
An additional finding was that outliers were effectively eliminated with AI-assisted contouring. With an (arbitrary) threshold in DSC at 0.4, 2.5% of manual contours and only 0.2% of AI-assisted contours were outliers. Although it is

unlikely that all outliers would affect radiotherapy treatment, there were five complete geographical misses observed with manual contouring (Data Supplement, Fig S2). These would likely have affected treatment planning, had they been used clinically. The rate of geographical misses with manual contouring in this study hypothetically translates into one organ missed for every 18 patients contoured. This serves as an imperative reminder that peer review of all contours should be routine clinical practice.

From this study, it is clear that AI-assisted contouring is beneficial when the AI-contouring model locates the right structures and provides decent contours. It is, however, known that these models sometimes fail to do so.<sup>20,36</sup> It remains unknown what effect erroneous AI contours would have on final contours—or, in other words, how wrong AI contours would have to be, before clinicians realize it and fall back on manual contouring. The failure rate is unknown for



**FIG 4.** Raw results for the control group (A) and the intervention group (B) for the entire study. Each contouring round is represented by a color. In the control group (A), red is baseline, orange is the round after teaching, dark purple is short-term follow-up, and light purple is long-term follow-up. In the intervention group (B), dark blue is baseline, light blue is after teaching, dark green is short-term follow-up, and light green is long-term follow-up. Each circle represents a single contour from a single participant, and boxplots are based on these. Within each figure, the top and middle rows show DSC and HD95 between participants' contours and expert consensus structures, respectively. Bottom rows show the contouring time. DSC, Dice Similarity Coefficient; HD95, Hausdorff Distance 95th percentile.



**FIG 5.** The top figure shows the estimated effect sizes of a given combination of exposures, and the bottom figure shows the estimated effect sizes at short-term follow-up. Top, middle, and bottom rows within each subfigure are DSC, HD95, and contouring time. In the top subfigure, no teaching and no AI-assisted contouring are used as reference in all four subanalyses, and in the bottom subfigure, the round after teaching is reference in all four. Absolute differences of medians are shown with 95% CIs. Values above 0 mean that the reference number is smallest and vice versa. AI, artificial intelligence; DSC, Dice Similarity Coefficient; HD95, Hausdorff Distance 95th percentile.

the commercial model used in this study. Therefore, it cannot be determined whether the benefit of avoided outliers supersedes the risk that may come with erroneous AI contours.

A major limitation to the study was that contours were not used clinically. Most participants completed their cases besides their regular clinical duties, which theoretically increased the risk of automation bias due to the risk of time pressure, lack of interest, and lack of accountability for treatment.<sup>37</sup> The fact that two thirds of participants handed in all the accepted-as-is-contours suggests either that automation bias was at play for these participants or that these participants might have cognitively processed the review of AI contours differently. The underlying mechanisms of reviewing and editing AI contours are an important topic for future research but are beyond the scope of this

work. To confirm the findings of this study, prospective clinical studies on AI-assisted contouring in LMICs should be carried out upon clinical implementation.

ROs who worked in low- and middle-income countries contoured most OARs for head-and-neck cancer with higher similarity to expert consensus contours with AI-assisted contouring than with manual contouring. Furthermore, teaching combined with AI-assisted contouring was the most effective strategy to reduce contouring time. The benefits of teaching combined with AI-assisted contouring persisted after a 6-month follow-up period. Therefore, AI-assisted contouring—especially when combined with teaching—is a promising contribution to reaching optimal utilization of radiotherapy in the present and future. Therefore, a global transition toward AI-assisted contouring with appropriate clinical monitoring is encouraged.

## AFFILIATIONS

<sup>1</sup>Experimental Clinical Oncology, Aarhus University Hospital, Aarhus, Denmark

<sup>2</sup>International Atomic Energy Agency, Vienna, Austria

<sup>3</sup>Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

<sup>4</sup>MAASTRO clinic, Maastricht University Medical Centre, Maastricht, the Netherlands

<sup>5</sup>Department of Oncology, Aarhus University Hospital, Aarhus, Denmark

<sup>6</sup>Department of Radiation Oncology, Cruces University Hospital, Bilbao, Spain

<sup>7</sup>Labaid Cancer Hospital and Super Speciality Centre, Dhaka, Bangladesh

<sup>8</sup>National Cancer Institute, University of Gezira, Wad Madani, Sudan

<sup>9</sup>Cipto Mangunkusumo Hospital, Jakarta, Indonesia

<sup>10</sup>Dharmais Cancer Hospital, Jakarta, Indonesia

<sup>11</sup>Kenyatta National Hospital, Nairobi, Kenya

<sup>12</sup>Tbilisi State Medical University and Ingorokva High Medical Technology University Clinic, Tbilisi, Georgia

<sup>13</sup>Nepal Cancer Hospital and Research Center, Lalitpur, Nepal

<sup>14</sup>Centro de Radioterapiya dean Funes, Cordoba, Argentina

<sup>15</sup>Atomic Energy Cancer Hospital NORI, Islamabad, Pakistan

<sup>16</sup>University Clinic of Radiotherapy and Oncology, Skopje, Macedonia

<sup>17</sup>National Center of Oncology, Baku, Azerbaijan

<sup>18</sup>King Hussein Cancer Center, Amman, Jordan

<sup>19</sup>Christian Institute of Health Science and Research, Dimapur, India

<sup>20</sup>Hospital Abderrahmen Mami, Ariana, Tunisia

<sup>21</sup>Moldavian Oncology Institute, Chisinau, Moldova

<sup>22</sup>N. N. Alexandrov National Cancer Centre of Belarus, Minsk, Belarus

<sup>23</sup>Hospital México, San José, Costa Rica

<sup>24</sup>Mother Tereza Hospital, Tirana, Albania

<sup>25</sup>National Centre of Oncology and Hematology, Bishkek, Kyrgyzstan

<sup>26</sup>Center of Nuclear Medicine and Oncology, Semey, Kazakhstan

<sup>27</sup>Hospital Kuala Lumpur, Kuala Lumpur, Malaysia

<sup>28</sup>Uganda Cancer Institute, Kampala, Uganda

<sup>29</sup>National Cancer Center of Mongolia, Ulaanbaatar, Mongolia

## CORRESPONDING AUTHOR

Stine Sofia Korreman, MSc; e-mail: stine.korreman@clin.au.dk.

## PRIOR PRESENTATION

Presented in part at ESTRO2023, Vienna, Austria, May 15, 2023; Danish Cancer Research Days, Odense, Denmark, August 29, 2023; ESTRO2024, Glasgow, Scotland, May 5, 2024; and PhD Day, Aarhus, Denmark, January 19, 2024.

## SUPPORT

Supported in part by Aarhus University with M.E.R.'s salary and in part by IAEA with compensation to participating institutions.

## AUTHOR CONTRIBUTIONS

**Conception and design:** Mathis Ersted Rasmussen, Kamal Akbarov, Egor Titovich, Jasper Albertus Nijkamp, Wouter Van Elmpt, Stine Sofia Korreman, Jesper Grau Eriksen

**Financial support:** Kamal Akbarov, Lisbeth Cordero-Mendez

**Administrative support:** Mathis Ersted Rasmussen, Kamal Akbarov, Egor Titovich, Stine Sofia Korreman, Jesper Grau Eriksen

**Provision of study materials or patients:** Hanne Primdahl, Pernille Lassen, Jon Cacicedo, A.F.M. Kamal Uddin, Ahmed Mohamed, Ben Prajogi, Kartika Erida Brohet, Catherine Nyongesa, Darejan Lomidze, Gisupnikha Prasiko, Gustavo Ferraris, Humera Mahmood, Igor Stojkovski, Isa Isayev, Issa Mohamad, Leivon Shirley, Lotfi Kochbati, Ludmila Eftodiev, Maksim Piatkevich, Maria Matilde Bonilla Jara, Orges Spahiu, Rakhat Aralbayev, Raushan Zakirova, Sandya Subramaniam, Solomon Kibudde, Uranchimeg Tsegmed, Jesper Grau Eriksen

**Collection and assembly of data:** Mathis Ersted Rasmussen, Kamal Akbarov, Egor Titovich, Hanne Primdahl, Pernille Lassen, Jon Cacicedo, A.F.M. Kamal Uddin, Ahmed Mohamed, Ben Prajogi, Kartika Erida Brohet, Catherine Nyongesa, Darejan Lomidze, Gisupnikha Prasiko, Gustavo Ferraris, Humera Mahmood, Igor Stojkovski, Isa Isayev, Issa Mohamad, Leivon Shirley, Lotfi Kochbati, Ludmila Eftodiev, Maksim Piatkevich, Maria Matilde Bonilla Jara, Orges Spahiu, Rakhat Aralbayev, Raushan Zakirova, Sandya Subramaniam, Solomon Kibudde, Uranchimeg Tsegmed, Jesper Grau Eriksen

**Data analysis and interpretation:** Mathis Ersted Rasmussen, Stine Sofia Korreman, Jesper Grau Eriksen, Lisbeth Cordero-Mendez, Kamal Akbarov

**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 unless otherwise noted. 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](http://www.asco.org/rwc) or [ascopubs.org/go/authors/author-center](http://ascopubs.org/go/authors/author-center).

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians ([Open Payments](http://Open Payments)).

### Wouter Van Elmpt

**Speakers' Bureau:** Varian Medical Systems

**Research Funding:** Varian Medical Systems (Inst)

### A.F.M. Kamal Uddin

**Consulting or Advisory Role:** Novartis

### Igor Stojkovski

**Employment:** Siemens Healthineers

**Honoraria:** MSD Oncology, Roche

**Consulting or Advisory Role:** MSD

### Rakhat Aralbayev

**Employment:** National Center of Oncology and Hematology

**Leadership:** National Center of Oncology

**Consulting or Advisory Role:** National Center of Oncology (Inst)

No other potential conflicts of interest were reported.

## ACKNOWLEDGMENT

Our gratitude goes to all oncologists who provided the invaluable data. Thanks to EduCase for their 24/7-support—and for implementing time recording for this study. Thanks to MVision who provided contours through EduCase.

## REFERENCES

- Atun R, Jaffray DA, Barton MB, et al: Expanding global access to radiotherapy. *Lancet Oncol* 16:1153-1186, 2015
- Sun H, Jin C, Wang H, et al: Cost-effectiveness of stereotactic body radiotherapy in the treatment of non-small-cell lung cancer (NSCLC): A systematic review. *Expert Rev Pharmacoecon Outcomes Res* 22:723-734, 2022
- Monten C, Lievens Y: Adjuvant breast radiotherapy: How to trade-off cost and effectiveness? *Radiother Oncol* 126:132-138, 2018
- Verma V, Mishra MV, Mehta MP: A systematic review of the cost and cost-effectiveness studies of proton radiotherapy. *Cancer* 122:1483-1501, 2016
- Abreha SK: Model-based cost-effectiveness analysis of external beam radiation therapy for the treatment of localized prostate cancer: A systematic review. *Cost Eff Resour Alloc* 17:10, 2019
- Verma V, Yegya-Raman N, Sprave T, et al: A systematic review of the cost-effectiveness of stereotactic radiation therapy for cancer oligometastases. *Int J Radiat Oncol Biol Phys* 114:977-988, 2022
- Borras JM, Lievens Y, Dunscombe P, et al: The optimal utilization proportion of external beam radiotherapy in European countries: An ESTRO-HERO analysis. *Radiother Oncol* 116:38-44, 2015
- Lievens Y, Borras JM, Grau C: Provision and use of radiotherapy in Europe. *Mol Oncol* 14:1461-1469, 2020
- Grover S, Xu MJ, Yeager A, et al: A systematic review of radiotherapy capacity in low- and middle-income countries. *Front Oncol* 4:380, 2014
- Hong TS, Tomé WA, Harari PM: Heterogeneity in head and neck IMRT target design and clinical practice. *Radiother Oncol* 103:92-98, 2012
- Vrtovec T, Močnik D, Strojani P, et al: Auto-segmentation of organs at risk for head and neck radiotherapy planning: From atlas-based to deep learning methods. *Med Phys* 47:e929-e950, 2020
- Lin D, Wahid KA, Nelms BE, et al: *E pluribus unum*: Prospective acceptability benchmarking from the contouring collaborative for consensus in radiation oncology crowdsourced initiative for multiobserver segmentation. *J Med Imaging* 10:S11903, 2023
- Cha E, Elguindi S, Onochie I, et al: Clinical implementation of deep learning contour autosegmentation for prostate radiotherapy. *Radiother Oncol* 159:1-7, 2021
- Chen X, Sun S, Bai N, et al: A deep learning-based auto-segmentation system for organs-at-risk on whole-body computed tomography images for radiation therapy. *Radiother Oncol* 160:175-184, 2021
- Tao C-J, Yi J-L, Chen N-Y, et al: Multi-subject atlas-based auto-segmentation reduces interobserver variation and improves dosimetric parameter consistency for organs at risk in nasopharyngeal carcinoma: A multi-institution clinical study. *Radiother Oncol* 115:407-411, 2015
- Chao KSC, Bhide S, Chen H, et al: Reduce in variation and improve efficiency of target volume delineation by a computer-assisted system using a deformable image registration approach. *Int J Radiat Oncol Biol Phys* 68:1512-1521, 2007
- Young AV, Wortham A, Wernick I, et al: Atlas-based segmentation improves consistency and decreases time required for contouring postoperative endometrial cancer nodal volumes. *Int J Radiat Oncol Biol Phys* 79:943-947, 2011
- Reed VK, Woodward WA, Zhang L, et al: Automatic segmentation of whole breast using atlas approach and deformable image registration. *Int J Radiat Oncol Biol Phys* 73:1493-1500, 2009
- Deeley MA, Chen A, Datter RD, et al: Segmentation editing improves efficiency while reducing inter-expert variation and maintaining accuracy for normal brain tissues in the presence of space-occupying lesions. *Phys Med Biol* 58:4071-4097, 2013
- Vaassen F, Boukerroui D, Looney P, et al: Real-world analysis of manual editing of deep learning contouring in the thorax region. *Phys Imaging Radiat Oncol* 22:104-110, 2022
- Chen Z, King W, Pearcey R, et al: The relationship between waiting time for radiotherapy and clinical outcomes: A systematic review of the literature. *Radiother Oncol* 87:3-16, 2008
- Scoccianti S, Detti B, Gadda D, et al: Organs at risk in the brain and their dose-constraints in adults and in children: A radiation oncologist's guide for delineation in everyday practice. *Radiother Oncol* 114:230-238, 2015
- Brouwer CL, Steenbakkens RJHM, Bourhis J, et al: CT-Based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG oncology and TROG consensus guidelines. *Radiother Oncol* 117:83-90, 2015
- Phil T, Albrecht T, Gay S, et al: Sikerdebaard/dcmrstruct2nii: dcmrstruct2nii v2. 2022. <https://doi.org/10.5281/zenodo.6330598>
- Phil T, Albrecht T, Gay S, et al: Aarhus-RadOnc-AI/dcmrstruct2nii: dcmrstruct2nii v2. 2022
- 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
- M. O. medpy—Medical Image Processing in Python. 2023
- SciPy: n.d. <https://scipy.org/>
- Maduro BLA, Sarkar A, Doyle LA, et al: Feasibility evaluation of novel AI-based deep-learning contouring algorithm for radiotherapy. *J Appl Clin Med Phys* 24:e14090, 2023
- Hu Y, Nguyen H, Smith C, et al: Clinical assessment of a novel machine-learning automated contouring tool for radiotherapy planning. *J Appl Clin Med Phys* 24:e13949, 2023
- Radici L, Ferrario S, Borca VC, et al: Implementation of a commercial deep learning-based auto segmentation software in radiotherapy: Evaluation of effectiveness and impact on workflow. *Life* 12: 2088, 2022
- Doolan PJ, Charalambous S, Roussakis Y, et al: A clinical evaluation of the performance of five commercial artificial intelligence contouring systems for radiotherapy. *Front Oncol* 13:1213068, 2023
- Nielsen CP, Lorenzen EL, Jensen K, et al: Consistency in contouring of organs at risk by artificial intelligence vs oncologists in head and neck cancer patients. *Acta Oncol* 62:1418-1425, 2023
- Eriksen JG, Salembier C, Rivera S, et al: Four years with FALCON—An ESTRO educational project: Achievements and perspectives. *Radiother Oncol* 112:145-149, 2014
- Vinod SK, Min M, Jameson MG, et al: A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. *J Med Imaging Radiat Oncol* 60:393-406, 2016
- Temple SWP, Rowbottom CG: Gross failure rates and failure modes for a commercial AI-based auto-segmentation algorithm in head and neck cancer patients. *J Appl Clin Med Phys* 25:e14273, 2024
- Godard K, Roudsari A, Wyatt JC: Automation bias: A systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc* 19:121-127, 2012