PO-1722 AI-driven quality insurance for delineation in radiotherapy breast clinical trials

S. Rivera[1], A. Lombard[2], D. Pasquier[3], S. Wong[4], E. Limkin[5], G. Auzac[5], J. Blanchecotte[6], M.E. Chand-Fouché[7], A. Lamrani-Ghaouti[8], N. Bonnet[8], N. Paragios[2], C. Martineau-Huynh[2], E. Ullmann[2], A. Ruffier[9], E. Deutsch[5]

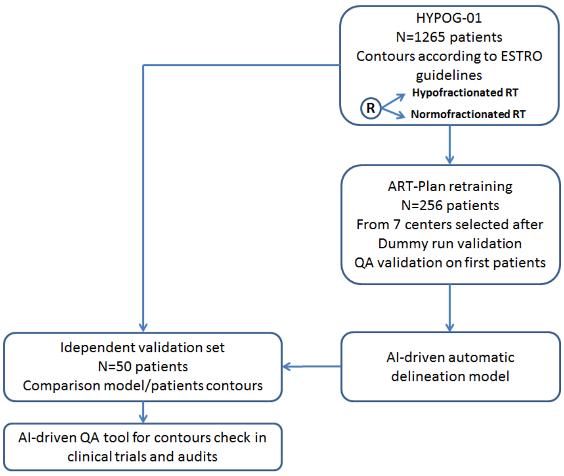
[1]Institut Gustave Roussy, radiotherapy, Villejuif, France ; [2]Therapanacea, artificial intelligence, Paris, France ; [3]Oscar Lambret, radiotherapy, Lille, France ; [4]Assistance Publique des Hôpitaux de Marseille, Radiotherapy, Marseille, France ; [5]Gustave Roussy, radiotherapy, villejuif, France ; [6]Institut de cancérologie de l'ouest site Paul Papin, radiotherapy, Angers, France ; [7]Antoine Lacassagne, radiotherapy, nice, France ; [8]UNICANCER, Unitrad, Kremlin Bicêtre, France ; [9]Institut interrégional de cancérologie- centre Jean-Bernard, Radiotherapy, Le Mans, France

Purpose or Objective

Clinical trials in radiotherapy inherit strong uncertainties on their outcomes due to significant inter/intra-user variability with respect to delineation. The lack of systematic review- in particular for academic trials – and the absence of gold standard for the delineation could have a tremendous impact on trial's outcome. In this work we use artificial intelligence towards the development of a systematic, scalable and bias-free tool for quality control assessment of the delineation step.

Material and Methods

ART-Plan is a CE-marked solution for automatic delineation of 65+ ROI in radiotherapy harnessing anatomically preserving deep learning ensemble networks. In this study, ART-Plan was retrained using 256 patients from 7 investigating centers of the HYPOG-01 phase III randomized trial. HYPOG-01 data inclusion was done using a strict verification protocol. Delineations on random initial samples were assessed and evaluated according to the ESTRO breast contouring guidelines. These delineations were used to amend the ART-Plan pre-trained ensemble network towards the development of the quality insurance delineation tool. The derived solution was compared with human delineations on an independent set of 50 patients from HYPOG-01.



Results

Median Dice Similarity (MDS) and Mean contour distance (MCD) between clinical & deep learning contours were assessed. Organs with a training MDS (MDS_TR) ≥ 0.65 and MCD_TR ≤ 2mm were included to the quality control protocol (coronary artery & brachial plexus were excluded). The spinal cord was included despite low MDS_TR due to variability of practices (z-axis start/end point). Acceptance criteria were set for testing as follows: MDS ≥ 0.8 * MDS_TR & MCD ≤ 1.2*MCD_TR. Quantitative/qualitative results on the testing set are appended:

	MDS	MCD (mm)	Average Volume difference (cc)
Lungs	0.97 ±0.02	0.51 ±0.26	+20
Liver	0.95 ±0.01	0.77 ±0.46	0
Heart	0.91 ±0.03	1.35 ±0.52	+60
Humeral heads	0.90 ±0.05	0.68 ±0.38	+1
CTV Breast/Chest wall	0.89 ±0.06	1.48 ±0.60	+40
Esophagus	0.79 ±0.05	0.72 ±0.42	-1
Spinal cord	0.76 ±0.09	1.98 ±2.00	+19
Thyroid	0.75 ±0.09	0.78 ±0.37	0
CTVn Level 3	0.74 ±0.10	1.00 ±1.16	0
CTVn Level 1	0.72 ±0.10	2.10 ±1.16	0
Larynx	0.72 ±0.20	1.46 ±1.37	+1
CTVn Level 4	0.70 ±0.10	1.39 ±0.57	+1
CTVn Interpectoral	0.66 ±0.10	1.24 ±0.82	-1
CTVn Level 2	0.66 ±0.15	1.49 ±1.03	-2

Conclusion

An anatomically preserving ensemble neural network retrained on high quality contours coming from a multi-center clinical trial could lead to the development of a clinical acceptable control delineation tool. Prospective evaluation in the 30 HYPOG-01 investigating centers is ongoing. Large scale development in breast radiotherapy trials and daily routine audits could lead to treatment standardization and systematization of contours quality assessment in trials involving radiotherapy ensuring a higher reliability of the results, while saving medical expert time.

PO-1723 Evaluation of outlier rejection in 4D-MRI for motion estimation of subsequent sessions

D. Den Boer[1], J. Veldman[1], A. Bel[1], Z. Van Kesteren[1]

[1]Amsterdam UMC- location AMC, Radiotherapy, Amsterdam, The Netherlands

Purpose or Objective

4DMRI is gaining popularity in radiotherapy of the upper abdomen, as it provides good soft tissue contrast and depicts the target motion during free breathing. Irregular breathing (sighs, hiccups) during acquisition can lead to imaging artefacts and a non-representative estimation of the reconstructed motion amplitude. Outlier rejection has been shown to increase image quality and here, we studied its influence on the estimated motion amplitude and how this correlates to the actual motion in subsequent (treatment) sessions. An overestimation of the motion may potentially lead to exposure of normal tissue, an underestimation to underdosage of the target.

Material and Methods

Ten volunteers were scanned with our 4DMRI sequence for 2-10 weekly sessions, consisting of three consecutive 4D acquisitions (total 108 acquisitions), mimicking a planning MRI (first acquisition, denoted PlAcq, Planning Acquisition) and treatment fractions (the subsequent acquisitions, TrAcq, Treatment Acquisitions). In these scans, 11 2D coronal slices were acquired repetitively; the position of the diaphragm was assessed using a 1D navigator acquisition and used for amplitude binning (10 bins).

To account for outliers, we employed a strategy (Min95) that sets the thresholds such that 95% of the diaphragm positions are included, while the included amplitude range is minimized. We compared this to a binning strategy that selects the maximum inhale and exhale positions as thresholds (MaxIE), not discarding outliers.

For the PlAcq the thresholds of both binning strategies were determined and applied to the TrAcqs, after matching the means of the two navigator signals (Fig 1). The percentage of the time the TrAcq was outside the thresholds is denoted as t_{out}. This was done with the PlAcq as a reference and using the First Acquisition of a Day (FAcqD) as a reference for the other two TrAcqs that day.