Clarifying the Dosimetric Impact of Autosegmentation Inaccuracies

Joëlle van Aalst, Master student Technical Medicine

Department of Radiation Oncology, University Medical Centre Groningen, The Netherlands

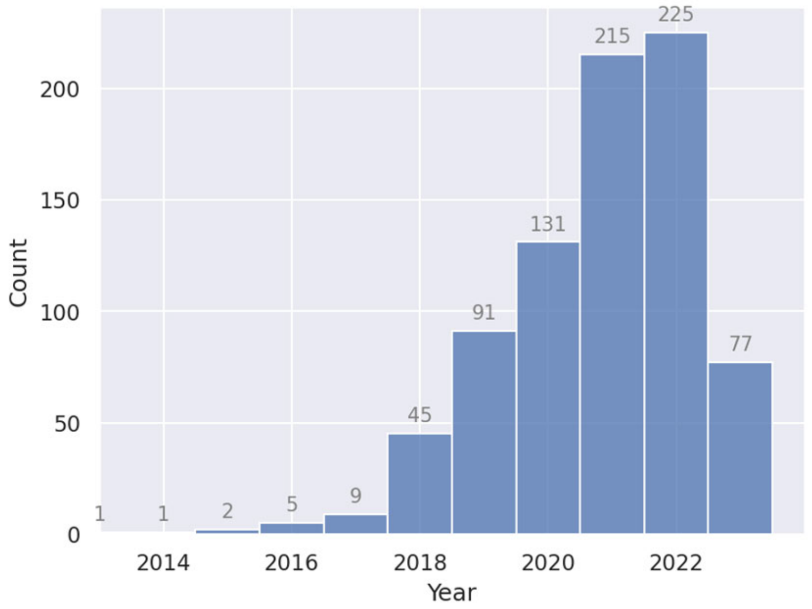
Technical Medicine, University of Twente, The Netherlands

Contributions:
Charlotte Brouwer, Tomas Janssen, Jelmer Wolterink, Roel Steenbakkers, Ilse van Bruggen, Stefan Both, Johannes Langendijk

20th International Conference on the use of Computers in Radiation Therapy (ICCR)
Lyon, France, July 8-11, 2024

Challenges in autosegmentation evaluation

- Increase in Deep learning-based autosegmentation papers
- Convention: assess geometric quality
  - Some do dose analysis
- Geometric discrepancies ≠ clinical impact [1,2]



Number of DL medical image segmentation studies published.
Isaksson et al. Sept. 2023; DOI: 10.3390/cancers15174389

[1] Sherer et al. Radiotherapy and Oncology 2021; DOI: 10.1016/j.radonc.2021.05.003
[2] Costea et al. Radiotherapy and Oncology 2022; DOI: 10.1016/j.radonc.2022.10.029

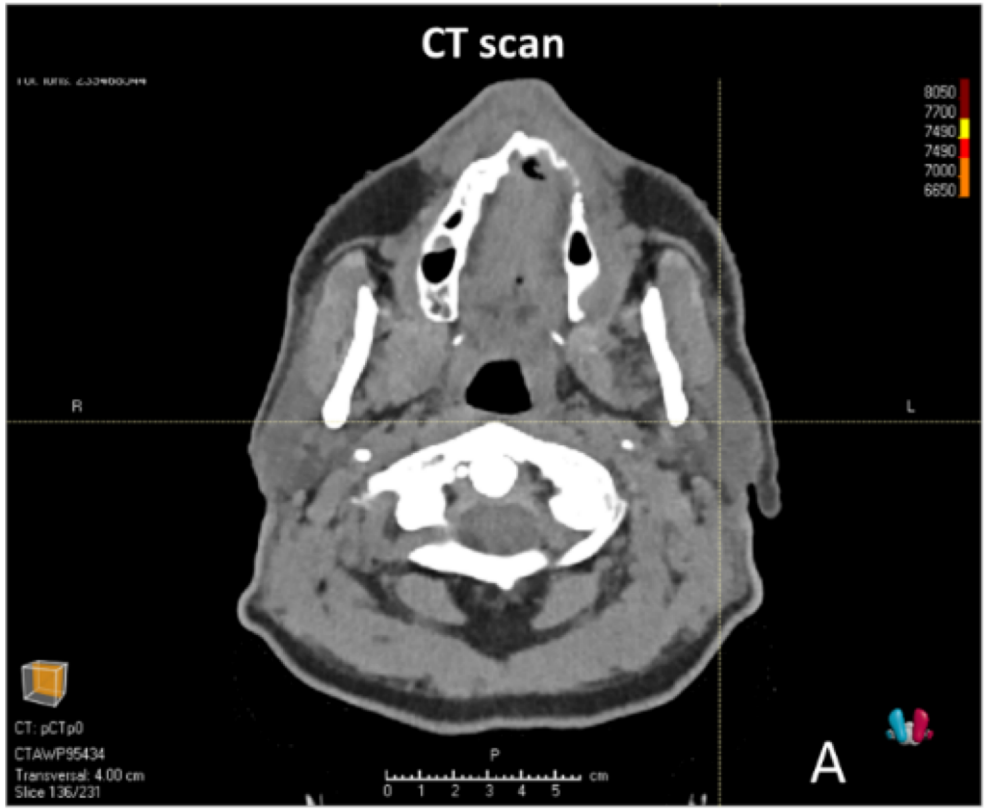
Research aim

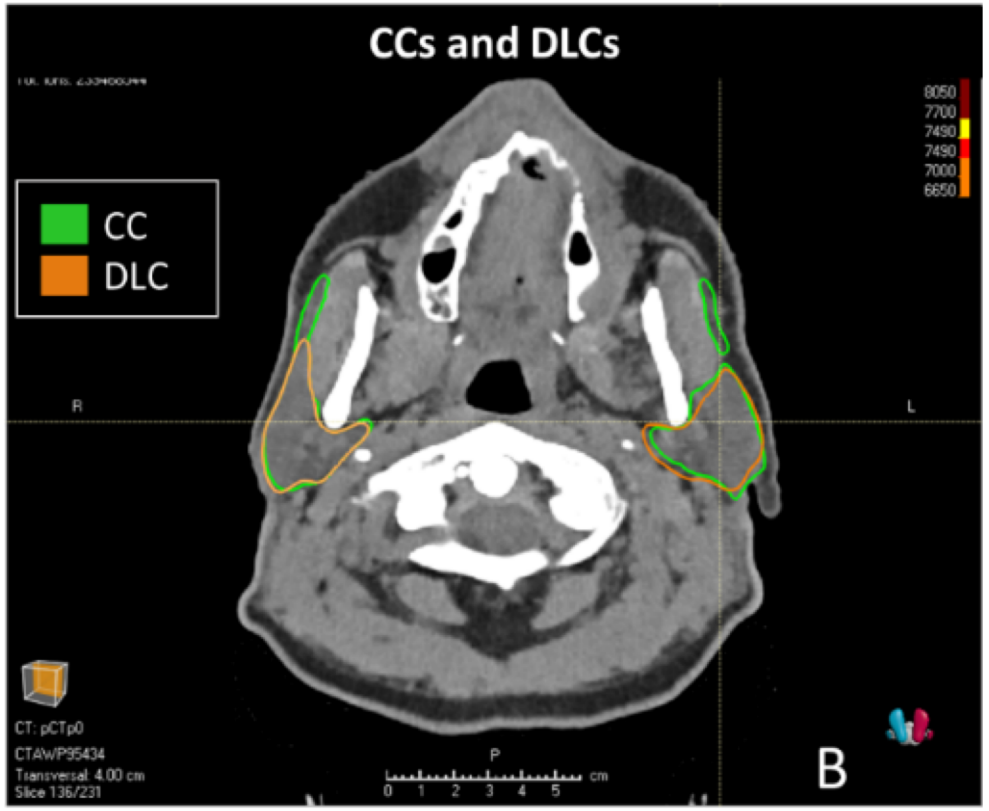
How to clarify the dosimetric impact of an autosegmentation model for Head and Neck cancer Organs at Risk?

Challenges in autosegmentation evaluation

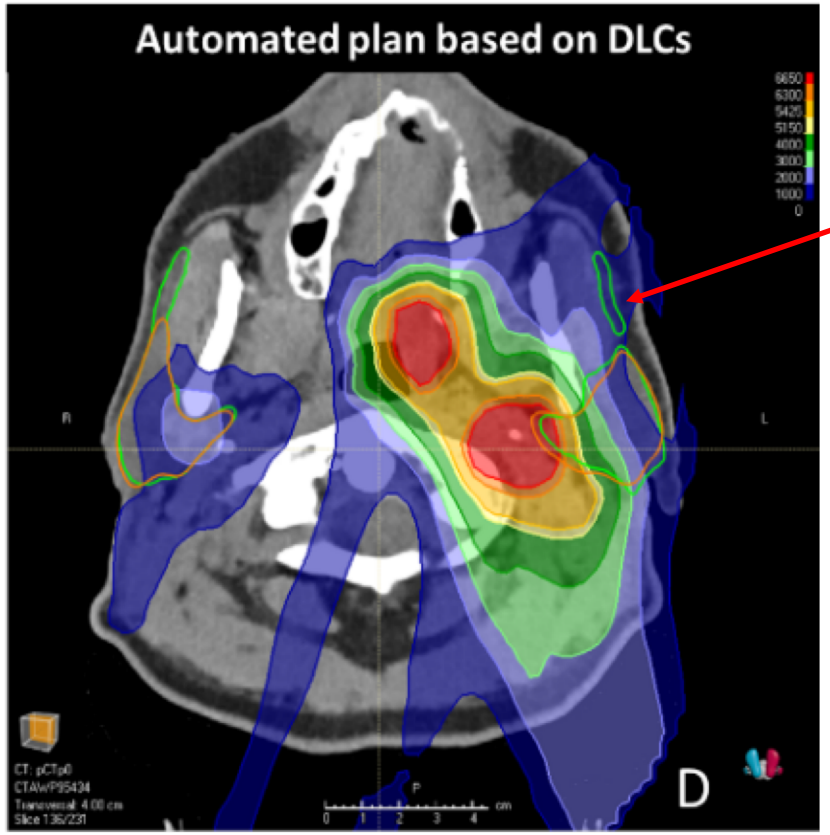
- Purpose of contouring
  - Make a treatment plan
  - Evaluate treatment plan
    - Dose is used for model-based patient selection

Case example



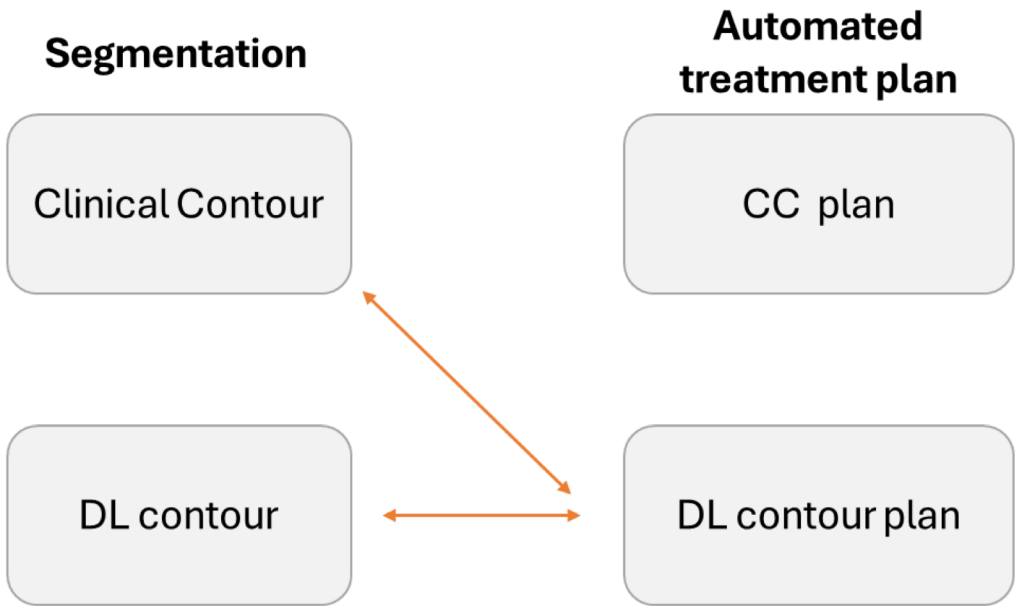


Case example

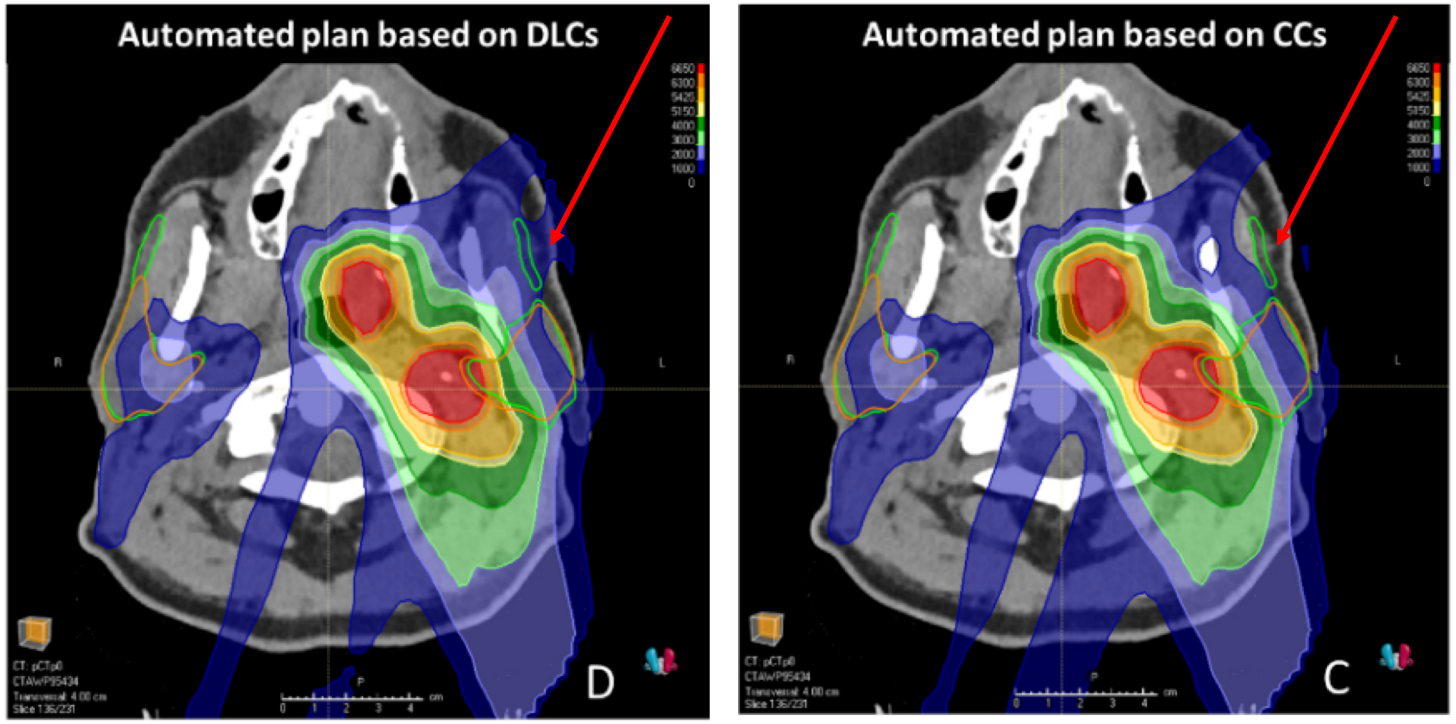


Evaluate treatment plan

Dosimetric impact on evaluation (DIE)

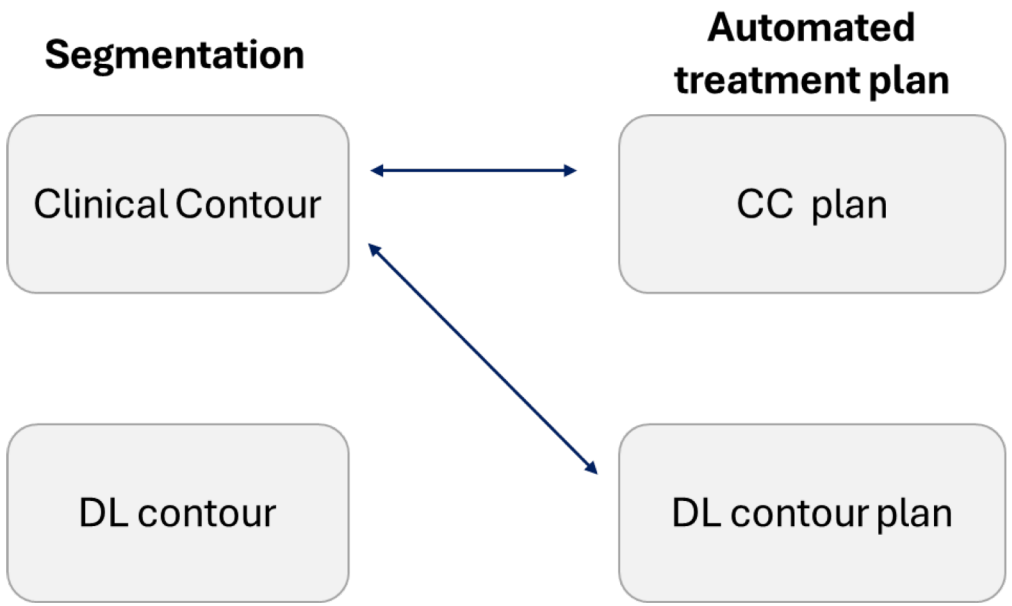


Case example



Make treatment plan

Dosimetric impact on planning (DIP)



Methods

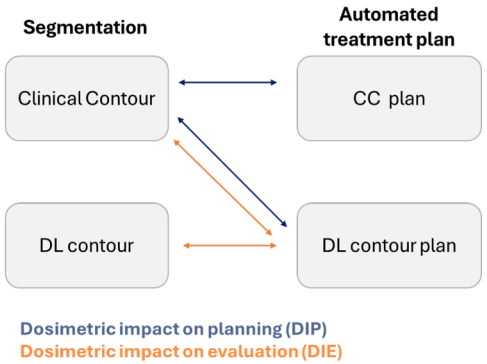
- Data
  - 68 oropharyngeal cancer patients treated with primary radiotherapy
  - CT + Clinical contours
- Autosegmentation model OARs
  - WorkflowBox 2.0, Mirada Medical Ltd.
  - Output: Deep Learning contours
- Autoplanning model
  - U-Net RayStation Development 10B
  - Output
    - Automated treatment plan based on Clinical contours
    - Automated treatment plan based on Deep Learning contours

Methods

- Autosegmentation quality assessment

Geometric comparison	Dice Score, Hausdorff distance
Dosimetric comparison	Dosimetric impact on planning (DIP), Dosimetric impact on evaluation (DIE),
Clinical impact	NTCP xerostomia (grade II and III), NTCP dysphagia (grade II and III)

Results



- Dosimetric measures
  - Larger mean dose differences for contour evaluation (DIE) than for plan evaluation (DIP)

DIP (average OAR mean dose difference):

- Glottic area: -1.5 Gy*

Others: 0.0 - 0.7 Gy

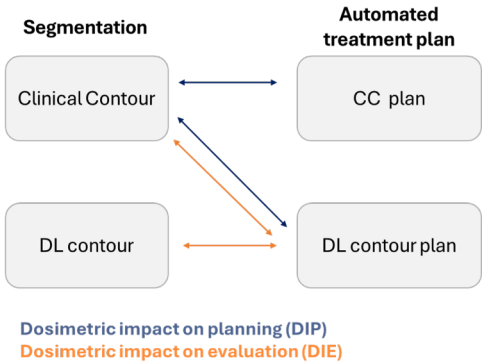
DIE (average OAR mean dose difference):

- Cervical oesophagus: 1.7 Gy*
- PCM: -1.0 Gy*
- Parotid left -1.6 Gy*
- Parotid right: -1.5 Gy*

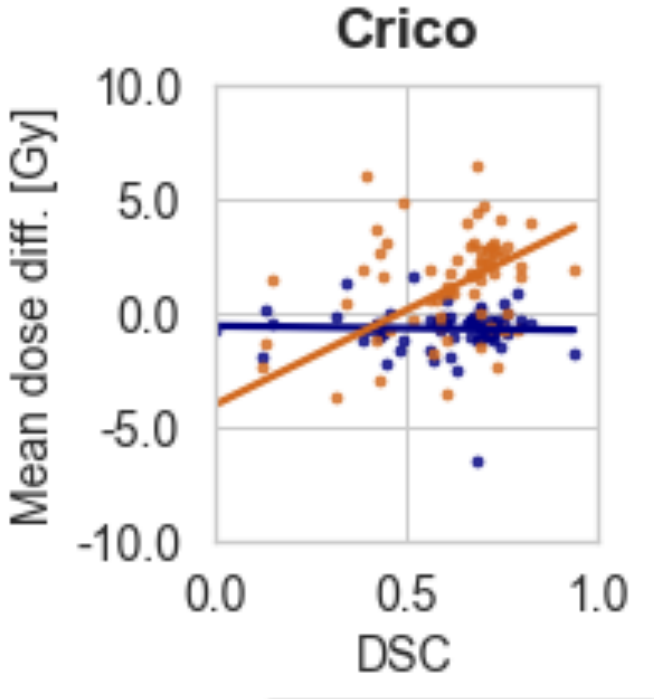
Others: 0.0 – 0.9 Gy

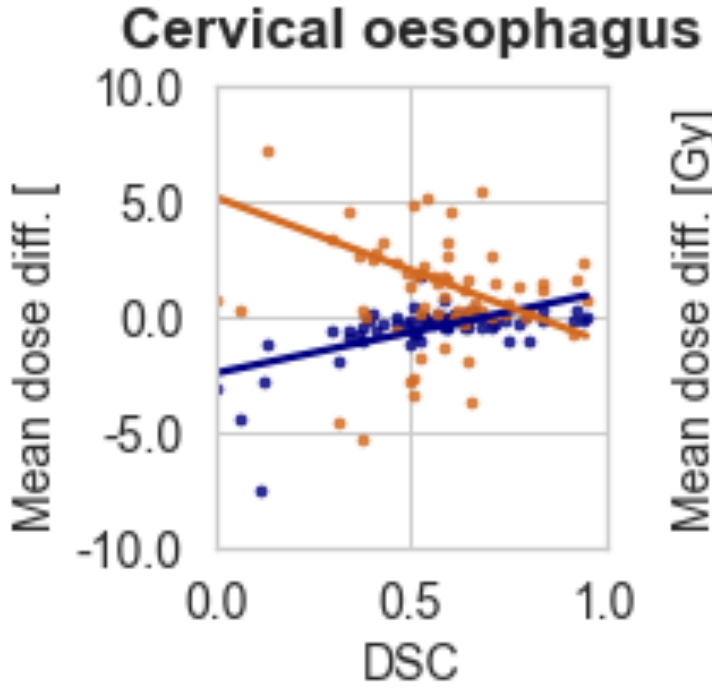
* Statistically significant: $p < 0.05$

Results



- Dosimetric measures
  - No correlation with geometric measures
  - No correlation between dosimetric measures (DIP vs DIE)





Results

- Clinical impact
  - Only xerostomia grade II ≥ 1.0 pp and significant for DIE

	Dosimetric impact on planning (DIP)	Dosimetric impact on evaluation (DIE)
Xerostomia grade II	-0.1 pp	**-1.0 pp***
Xerostomia grade III	0.0 pp	-0.4 pp*
Dysphagia grade II	0.3 pp*	-0.4 pp
Dysphagia grade III	0.1 pp*	0.2 pp

pp = percentage point
* statistically significant p <0,05

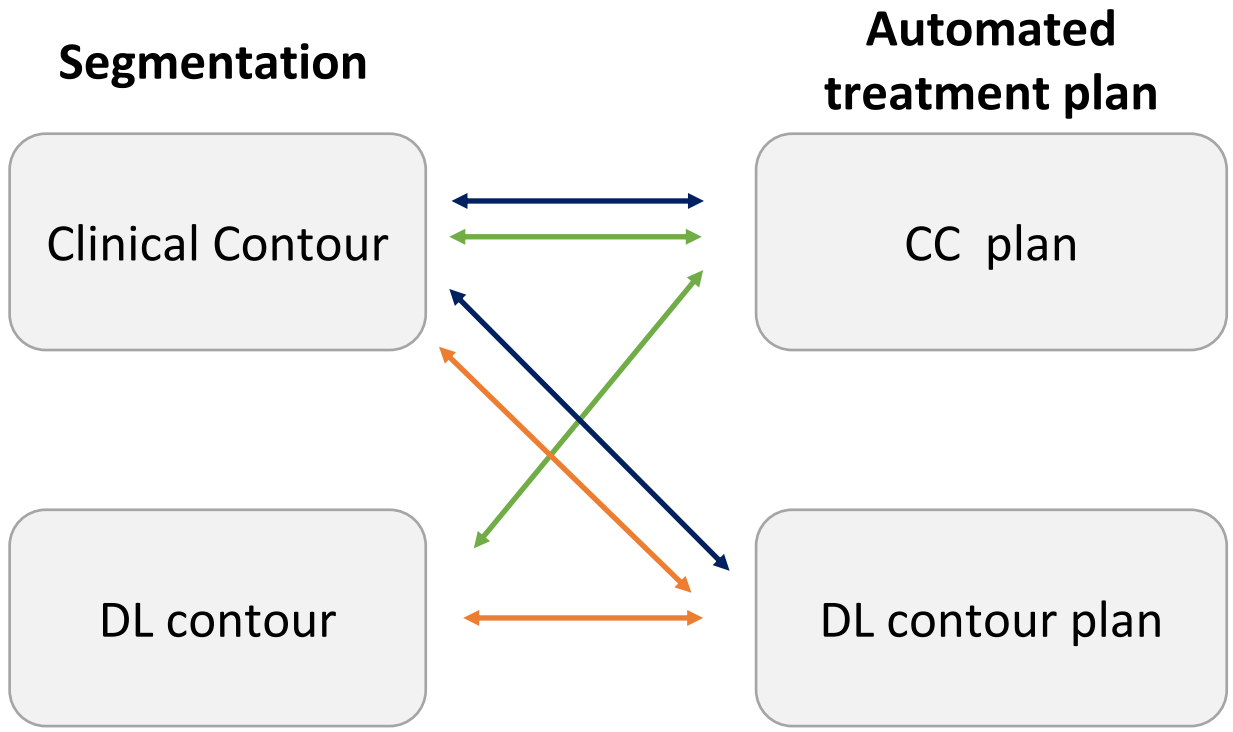
Results

- New dose plans DL segmentation = resource intensive
- Can you use clinical dose plans to approximate DIE?

DIP = dosimetric impact on planning

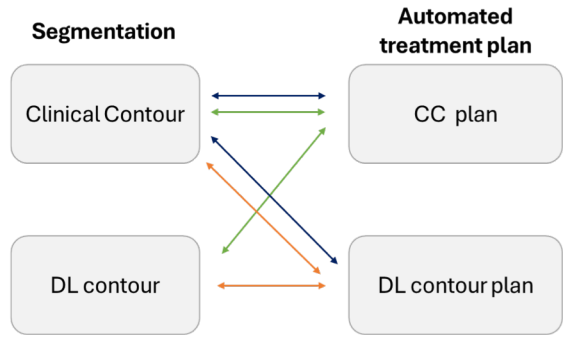
DIE = dosimetric impact on evaluation

DIA = dosimetric impact approximation

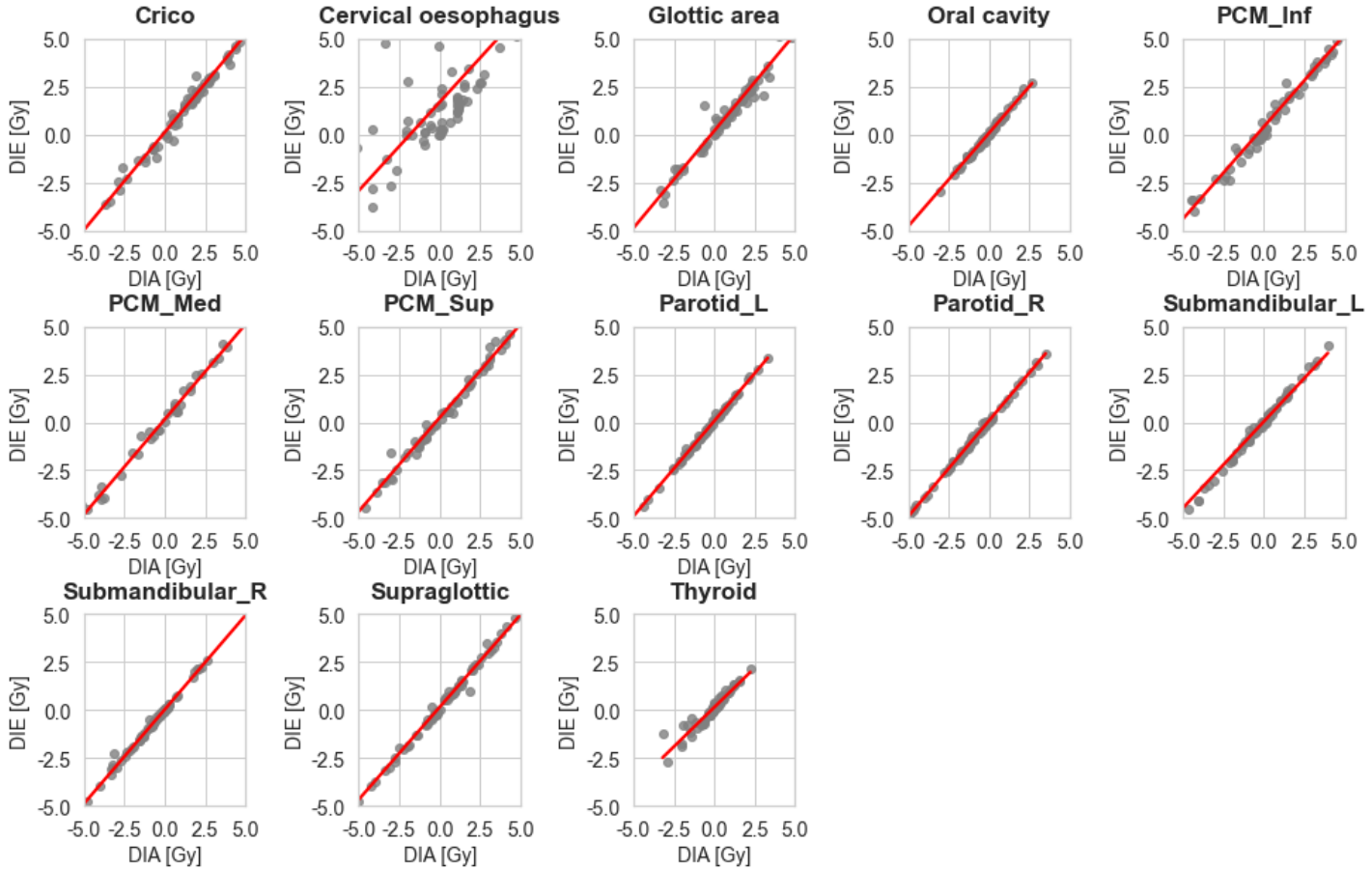


Results

- Good correlation between DIE and DIA



DIP = dosimetric impact on planning
DIE = dosimetric impact on evaluation
DIA = dosimetric impact approximation



Conclusion

- How to evaluate the dosimetric quality of an autosegmentation model for Head and Neck Organs at Risk?

1. Geometric ≠ dosimetric quality
2. Dosimetric impact on planning (DIP) ≠ dosimetric impact on evaluation (DIE)
   - NTCP **is not** significally different with DIP
   - NTCP **is** statistically different with DIE (for xerostomia grade II)
3. Dosimetric impact approximation (DIA) strong correlation with dosimetric impact on evaluation (DIE)

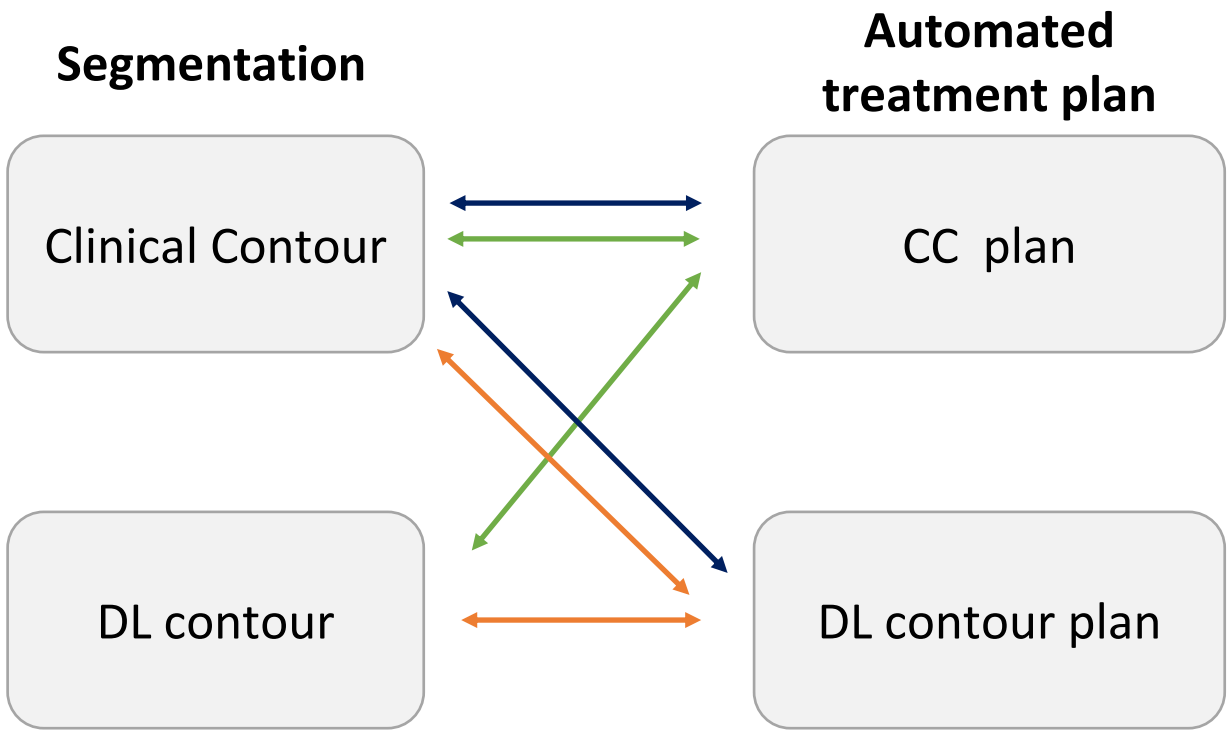
Methods

- Dosimetric comparison (mean dose difference)

DIP = dosimetric impact on planning

DIE = dosimetric impact on evaluation

DIA = dosimetric impact approximation

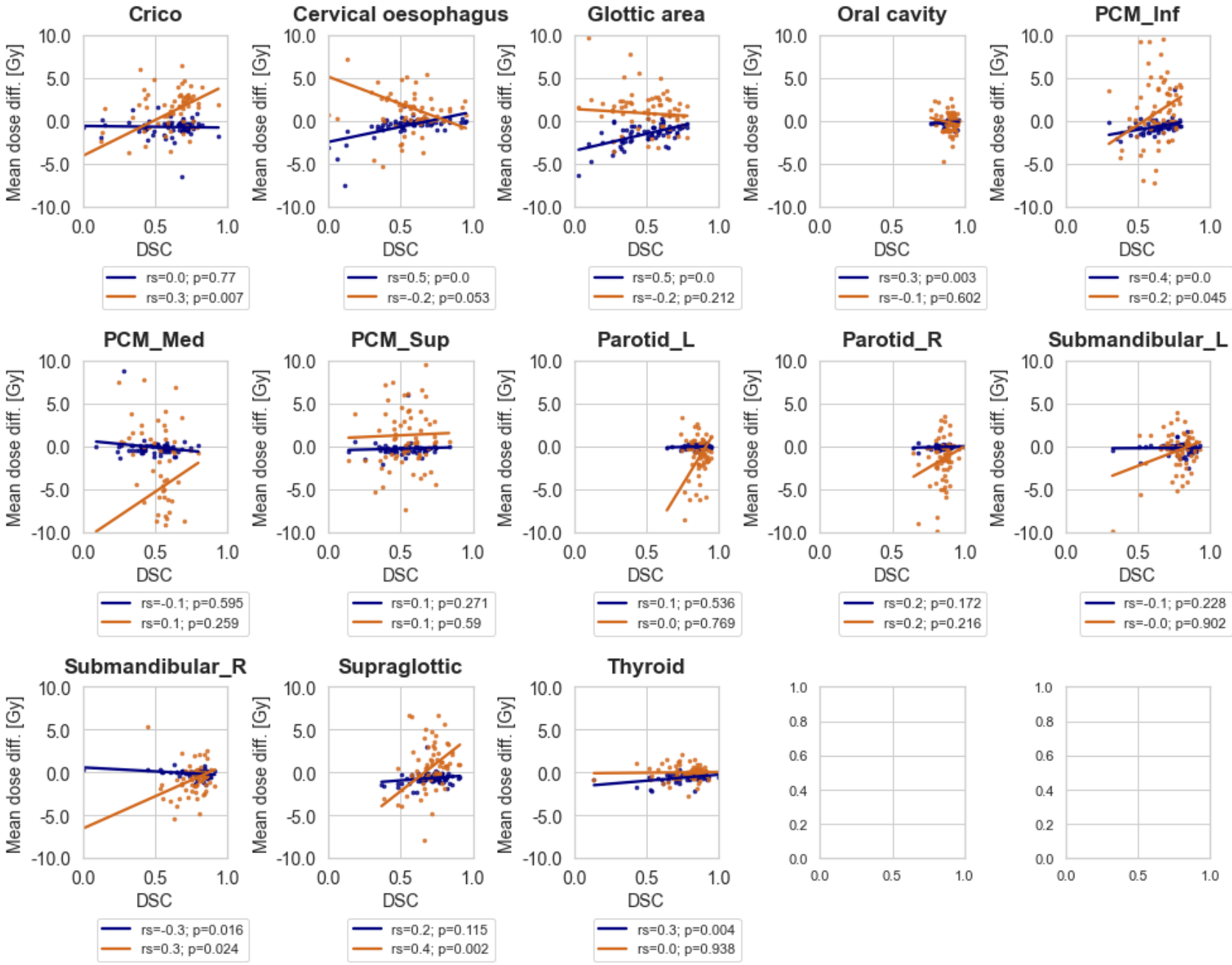


$$DIP = D_{CC}(CC) - D_{DLC}(CC)$$
$$DIE = D_{DLC}(CC) - D_{DLC}(DLC)$$
$$DIA = D_{CC}(CC) - D_{CC}(DLC)$$

Results

	DIP $D_{CC}(CC) - D_{DLC}(DLC)$		***DIE*** $D_{DLC}(CC) - D_{DLC}(DLC)$		***DIA*** $D_{CC}(CC) - D_{CC}(DLC)$		***GEOMETRIC COMPARISON***	
	Dose difference in Gy ± std	*p-value*	*Dose difference in Gy ± std*	*p-value*	*Dose difference in Gy ± std*	*p-value*	*DSC ± std*	*HD (mm) ± std*
Brainstem	0.1 ± 0.3	p < 0.001	0.4 ± 1.0	p < 0.001	0.4 ± 1.0	p < 0.001	0.92 ± 0.1	0.42 ± 0.3
Crico	-0.7 ± 1.0	p < 0.001	0.9 ± 3.4	p < 0.001	0.8 ± 3.2	p < 0.001	0.60 ± 0.2	0.85 ± 0.4
Cerv. oesophagus	-0.5 ± 1.2	p < 0.001	1.7 ± 3.7	p < 0.001	0.0 ± 3.3	p = 0.774	0.55 ± 0.2	1.26 ± 0.6
Glottic area	-1.5 ± 1.2	p = 0.001	0.9 ± 3.9	p < 0.001	0.7 ± 3.8	p < 0.001	0.50 ± 0.2	0.84 ± 0.5
Mandible	-0.3 ± 0.3	p < 0.001	-0.6 ± 0.5	p < 0.001	-0.6 ± 0.5	p < 0.001	0.95 ± 0.0	1.44 ± 0.4
Oral cavity	-0.1 ± 0.2	p < 0.001	-0.2 ± 1.2	p = 0.156	-0.3 ± 1.2	p = 0.057	0.90 ± 0.0	0.89 ± 0.3
PCM	-0.3 ± 0.9	p < 0.001	-1.0 ± 3.6	p = 0.030	-0.3 ± 5.9	p < 0.001	0.53 ± 0.1	1.07 ± 0.5
Parotid L	0.0 ± 0.2	p = 0.087	-1.6 ± 3.6	p < 0.001	-1.7 ± 3.6	p < 0.001	0.85 ± 0.1	1.21 ± 0.6
Parotid R	0.0 ± 0.2	p = 0.293	-1.5 ± 2.6	p < 0.001	-1.6 ± 2.6	p < 0.001	0.85 ± 0.1	1.19 ± 0.6
Spinal cord	0.1 ± 0.2	p < 0.001	-1.5 ± 1.8	p < 0.001	-1.6 ± 1.8	p < 0.001	0.82 ± 0.2	2.62 ± 3.4
Submandibular L	-0.1 ± 0.6	p = 0.382	-0.5 ± 2.2	p = 0.285	-0.6 ± 2.4	p = 0.231	0.77 ± 0.1	0.79 ± 0.4
Submandibular R	-0.1 ± 0.4	p = 0.423	-0.9 ± 2.1	p < 0.001	-0.9 ± 2.1	p < 0.001	0.76 ± 0.1	0.74 ± 0.5
Supraglottic area	-0.7 ± 0.7	p < 0.001	0.4 ± 3.4	p = 0.050	0.3 ± 3.6	p = 0.105	0.70 ± 0.1	0.92 ± 0.5
Thyroid	-0.6 ± 0.5	p < 0.001	0.0 ± 0.8	p = 0.662	-0.2 ± 1.0	p = 0.368	0.77 ± 0.1	1.00 ± 0.5

Results


Crico
Mean dose diff. [Gy]
DSC
rs=0.0; p=0.77
rs=0.3; p=0.007
Cervical oesophagus
rs=0.5; p=0.0
rs=-0.2; p=0.053
Glottic area
rs=0.5; p=0.0
rs=-0.2; p=0.212
Oral cavity
rs=0.3; p=0.003
rs=-0.1; p=0.602
PCM_Inf
rs=0.4; p=0.0
rs=0.2; p=0.045
PCM_Med
rs=-0.1; p=0.595
rs=0.1; p=0.259
PCM_Sup
rs=0.1; p=0.271
rs=0.1; p=0.59
Parotid_L
rs=0.1; p=0.536
rs=0.0; p=0.769
Parotid_R
rs=0.2; p=0.172
rs=0.2; p=0.216
Submandibular_L
rs=-0.1; p=0.228
rs=-0.0; p=0.902
Submandibular_R
rs=-0.3; p=0.016
rs=0.3; p=0.024
Supraglottic
rs=0.2; p=0.115
rs=0.4; p=0.002
Thyroid
rs=0.3; p=0.004
rs=0.0; p=0.938