Qualitative evaluation of auto-segmented structures for radiotherapy planning using peer review guidelines

Royal Devon and Exeter NHS Foundation Trust

Coughlan, S (MSc), Biggar, R (MSc), Stonton, C (BSc), Axelsen, A (MSc)

AIMS/OBJECTIVE: Establish the clinical acceptability of AI generated contours in a radiotherapy planning context
CLINICAL IMPACT: AI tools can accurately delineate radiotherapy contours, providing time and resource savings

Background

During treatment planning, in order to measure and control dose received by anatomical structures, they must be accurately delineated on the CT dataset - as stressed by the ICRU [1]; "Delineation of these volumes is an obligatory step in the planning process, as absorbed dose cannot be prescribed, recorded, and reported without specification of target volumes and volumes of normal tissue at risk."

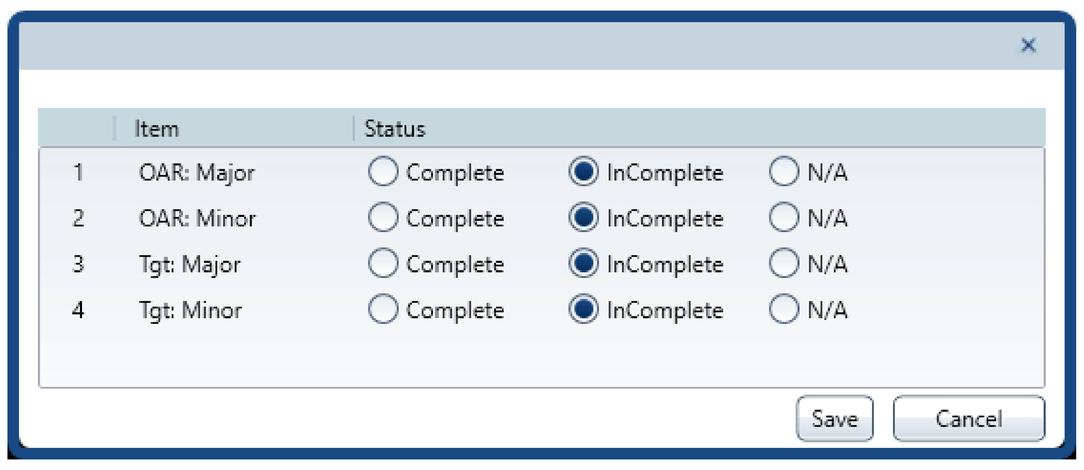
Over time, contouring tools evolved to assist this process, but efficacy varies and the level of automation is limited. Manual contouring is a meticulous and time-consuming task that requires expertise. Recent advances have seen the application of deep learning artificial intelligence algorithms to create accurate, adaptable anatomical models with the potential to yield significant resource savings. This is happening against a backdrop of significant national staffing shortages amongst the oncologist workforce.

Method

In this study, structures produced by Limbus AI Inc. Limbus Contour (version 1.4.1) [2] were evaluated in the context of RCR 2017 peer review guidance [3]. The expert user (Clinical Oncologist) was asked to appraise auto-segmented structures using familiar peer review criteria. Guidelines recommend categorising structures as requiring "Major", "Minor", or no modification. Although these terms are not well-defined, the classification is straightforward to interpret, and practical examples are provided.

Study design took the form of a simple e-questionnaire (Figure 1), forming part of the local pre-treatment Care Path workflow, driven by Varian ARIA® OMS. Data collection was automated and ensured compliance, since completion of the questionnaire was mandatory and is rate-limiting in the pre-treatment pathway. In adapting Aria task checklists for this purpose, it should be noted that the text definitions under "Item" are user definable, while the "Status" responses are not. In effect, this led to a local definition where "Complete" = Yes (i.e. modification required) and "N/A" = No (modification required). The response data was queried using a bespoke SQL query report against the Aria database.

The Limbus Contour application is configurable, and has been setup locally using 13 pre-defined anatomical templates. The template is automatically selected using Import Rules which query DICOM metadata contained within the planning CT dataset.



	Item	Status		
1	OAR: Major	Complete	InComplete	N/A
2	OAR: Minor	Complete	InComplete	N/A
3	Tgt: Major	Complete	InComplete	N/A
4	Tgt: Minor	Complete	InComplete	N/A

Save Cancel

Figure 1. Aria e-checklist ^

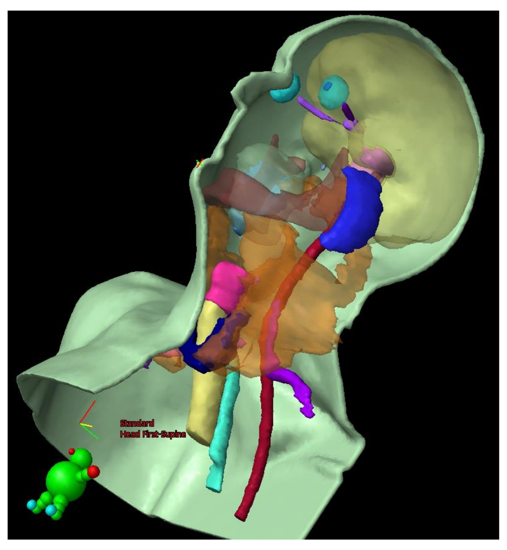


Figure 5. Example Limbus contours for a head and neck case >

Discussion

Quantitative metrics rely on benchmarking against a 'gold standard' which can never be definitive and ignores acceptable variance that will always occur. RCR peer review guidance acknowledges this and establishes a grading system for edits. Our survey method chosen, promotes clinician evaluation and streamlines the process.

Barriers to AI adoption, include IT issues, information & clinical governance, research governance and quality assurance. The Limbus Contour application has low hardware requirements, is easy to install and generally works passively with little administration needed. It operates as a local stand-alone client, which avoids the security concerns of cloud-based systems, making it easier for IT departments to implement.

Informal feedback has been provided to the Canadian developers during evaluation to help improve and develop new (inc. UK-centric) models.

Outcomes

Data for 145 subjects was collected over a 2 month period (Figure 2). Contours for normal tissue and target structures were evaluated by 12 Clinical Oncologists. Structures were reviewed and edited, if necessary, using the Varian Eclipse™ v15.5 Contouring module. At the end of the evaluation period, the data was compiled via SQL reporting.

Some Limbus target structures will necessarily be modified. For example, truncating nodal CTVs to only the involved levels, per patient. In such cases, polling data relates to judgement of the truncated structure, so not to unfairly bias outlining accuracy results.

Reviewers found that 70.3% of the structure sets required none, or "minor" modification (Figure 3). Looking at the OaRs in isolation, then the percentage of structure sets requiring "minor" or no modification rises to 95.2%. Figure 4 shows the (n=1517) auto-segmented OaR structures subdivided by treatment site.

It takes the Limbus Contour application <3 minutes to auto-segment using the most complex template [head+neck] (Figure 5), compared to ~1.5 hours of clinician time required for manual outlining. It should be noted that this is based on the processing power of a standard office desktop PC – no special servers or GPU is required. For the 125 patients in this study, we estimate this equates to a total saving of ~200 hours of consultant time - equivalent to 50 PA's in job planning terms. More comprehensive time and motion studies have been planned to better evaluate the savings.

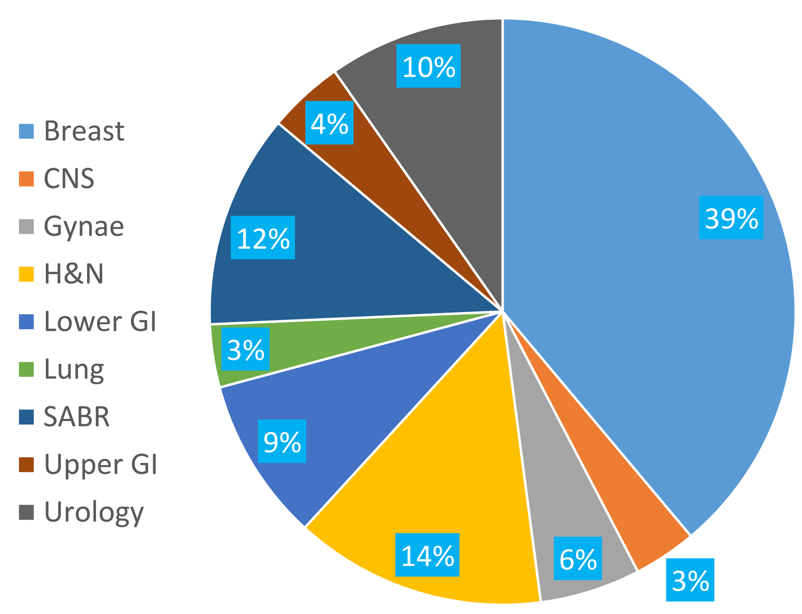


Figure 2. Limbus AI - Cases Reviewed by Site

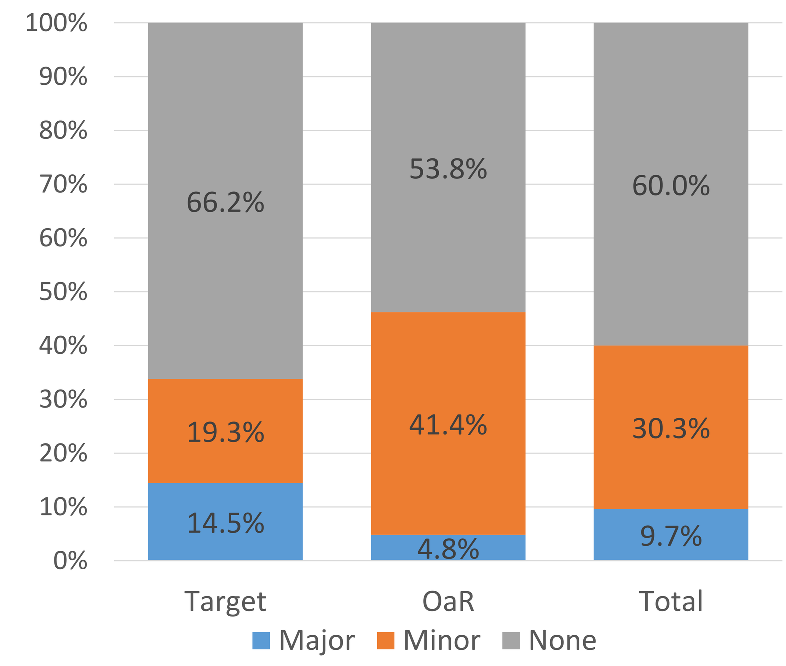


Figure 3. Peer review of structure sets

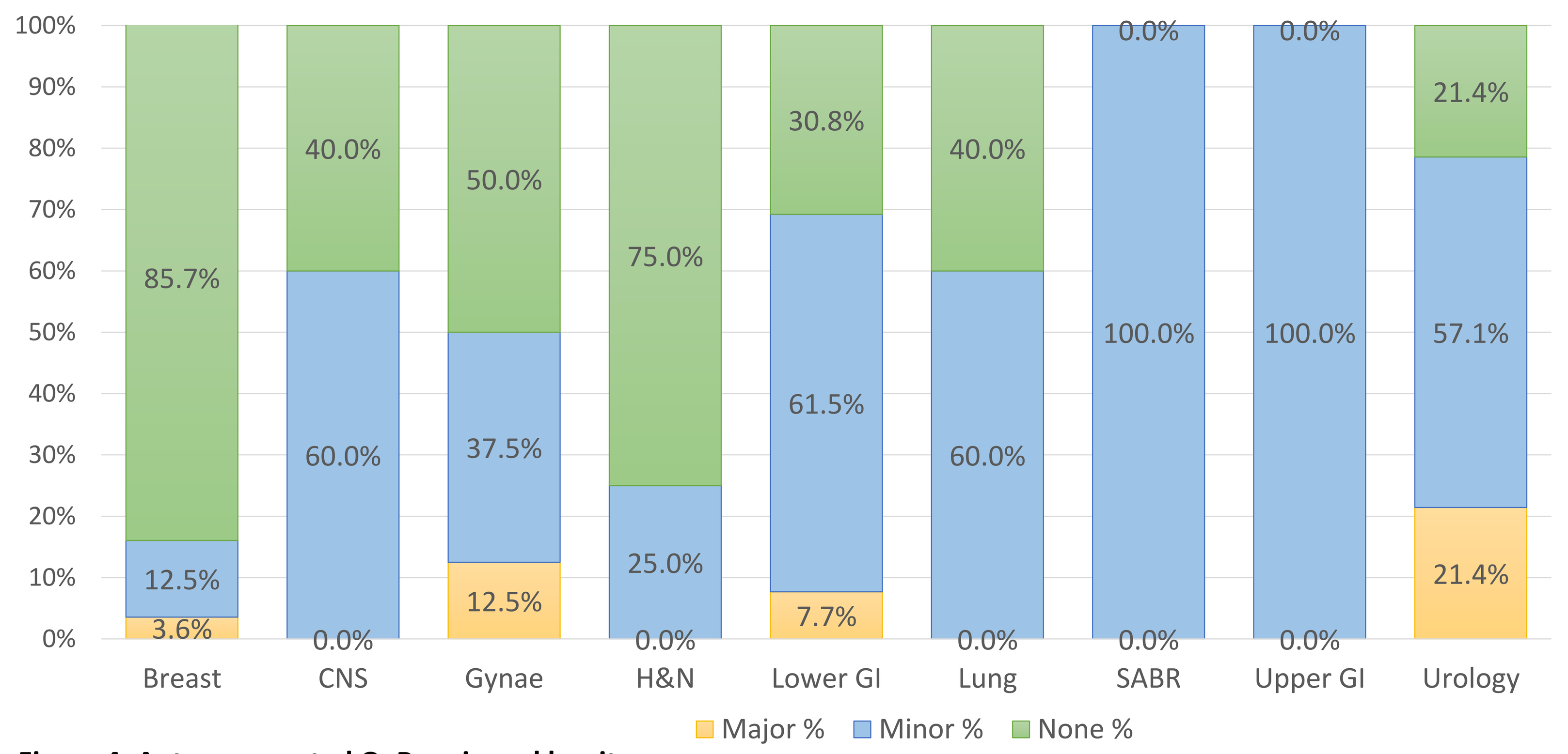


Figure 4. Auto-segmented OaR reviewed by site

Conclusion

To our knowledge, this is the first study to evaluate the quality of auto-segmented structures according to RCR (2017) peer review guidelines. This is as valid a method as any for evaluating the clinical relevance of auto-contoured structures, being pragmatic and based on current practice and national guidance.

The performance of Limbus Contour, and the quality of structures produced, has resulted in significant resource savings locally. The package has been implemented clinically for all treatment sites. This has allowed the radiotherapy planning team to undertake the majority of normal tissue contouring by mitigating overheads. As a result, clinician time has been liberated to pursue other activities.

[1] ICRU Report 83 Prescribing, Recording, and Reporting Photon-beam Intensity-modulated Radiation Therapy (IMRT). ICRU Report 83, J. ICRU Vol. 10 (1) 2010 (Oxford University Press, Oxford).

[2] Limbus AI Inc. (2021). Limbus Contour v1.5.0 Software User Manual. Limbus AI Inc.

[3] The Royal College of Radiologists. Radiotherapy target volume definition and peer review RCR guidance 2017. pp. 1-35.

NHS Royal Devon and Exeter NHS Foundation Trust

BIR The British Institute of Radiology

Simon Coughlan
Principal Dosimetrist
simon.coughlan@nhs.net