In-Silico Assessment of Automated CTV Contour Quality Assurance Integration In a Gastric Cancer Clinical Trial

Phillip Chlap[1], Mark Lee[2], Trevor Leong[3], Karen Olden[3], Matthew Field[1], Jason Dowling[4], Hang Min[4], Julie Chu[3], Jennifer Tan[3], Phillip K Tran[3], Tomas Kron[3], Annette Haworth[5], Martin A Ebert[6], Shalini K Vinod[2] and Lois Holloway[2]

[1]UNSW and Ingham Institute for Applied Medical Research, Australia, [2]Liverpool and Macarthur Cancer Therapy Centres, Australia, [3]Peter MacCallum Cancer Centre, Australia, [4]CSIRO Australian e-Health Research Centre, Australia, [5]University of Sydney, Australia, [6]Sir Charles Gairdner Hospital, Australia

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

ABSTRACT

Purpose: Radiotherapy quality assurance (RTQA) of target volume delineation is critical in clinical trials to ensure protocol adherence. Manual real-time review is resource-intensive and often limited to a subset of patients. This study evaluates the feasibility of automating contour QA to augment peer review and expand QA coverage across the entire trial cohort.

Methods: An in-silico trial was simulated using 203 cases from the AGITG/TROG TOPGEAR gastric cancer trial. Auto-segmentation models were trained using a Probabilistic U-Net to estimate uncertainty bands reflecting the acceptable range of the Clinical Target Volume (CTV). Metrics describing the fit of the manual contour to the uncertainty band (distance-to-band, volume difference, band overlap fraction) were extracted and used to train logistic regression classifiers. These classifiers were trained on successive batches of 10 cases to simulate a prospective trial rollout, iteratively updating the model while tuning for high sensitivity to minimize false negatives.

Results: Across 193 prospective simulation cases, the automated QA model achieved a sensitivity of 0.89, correctly flagging 65 of 73 violations for review. A distinct learning curve was observed, with lower sensitivity (0.63) in the initial 20 cases before the model stabilized to consistent performance as it adapted to the trial distribution. A specificity of 0.54 was observed, with 65 of 120 acceptable contours correctly passed. Implementation of this triage mechanism would have resulted in a 38% reduction in manual QA workload (73/193 cases passed), reducing the need for manual review while maintaining high sensitivity for detecting protocol violations.

Conclusion: Automated contour QA can feasibly streamline RTQA in clinical trials. Applied to the TOPGEAR trial, our approach demonstrated the potential to significantly reduce manual review workload. Crucially, the in-silico simulation exhibited an initial convergence phase, indicating the need for supervised operation before safe deployment of autonomous triage in multi-center trials.

CONTACT

Phillip Chlap
University of New South Wales
1 Campbell St, Liverpool, 2170, NSW, AUSTRALIA
phillip.chlap@unsw.edu.au

INTRODUCTION

- **RTQA is essential** to protocol adherence and trial validity
- **Centralized peer-review is resource-intensive:** rarely feasible for every patient within trial
- **TOPGEAR** (AGITG/TROG, NCT01924819), phase III trial for patients with resectable gastric cancer, restricted real-time review to a subset of patients
- **Gastric CTV contouring is challenging**: 28% of TOPGEAR cases required resubmission for protocol deviations, motivating rigorous QA
- **Automated contour QA** offers a path to full cohort coverage while reducing manual burden
- **Study Aim:** test the feasibility of automated CTV QA across the *entire* TOPGEAR cohort, in silico

METHODS AND MATERIALS

Dataset
- 203 TOPGEAR cases, simulating prospective trial deployment in silico

Phase 1: Model Initialization (Pre-Trial)
- First 10 patients contoured per protocol, confirmed by consensus workshop
- Two-stage model: nnU-Net segmentation → Probabilistic U-Net uncertainty band ("band of acceptability") (Fig. 1)

Phase 2: Prospective QA Simulation (During Trial)
- Contours scored against the uncertainty band: *distance-to-band, overlap fraction, volume difference*
- Logistic regression classifiers flagging violations retrained every 10 cases, tuned to **0.9 sensitivity**
- Flagged cases sent for peer review, passing cases can proceed to treatment

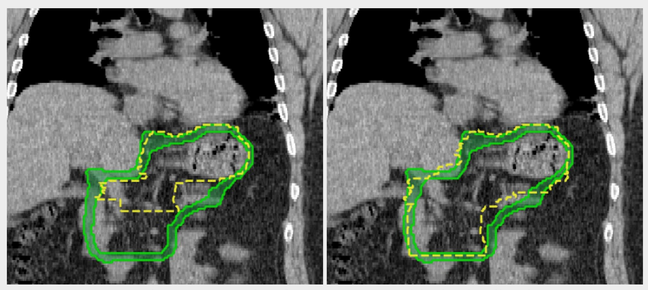

Figure 1. Coronal view example of TOPGEAR CTV (dashed yellow): a common violation (left) vs. corrected contour (right), against the model's predicted band of acceptability (green).

RESULTS

Cohort: 193 prospective cases: 73 violations, 120 acceptable (manual CTV)

Performance
- **Sensitivity 0.89:** 65 of 73 violations correctly flagged
- **Specificity 0.54:** 65 of 120 acceptable CTVs correctly passed

Workload
- **73 cases** cleared automatically, **120** triaged for peer review (Fig. 3)
- **38% reduction** in manual review workload, sensitivity preserved

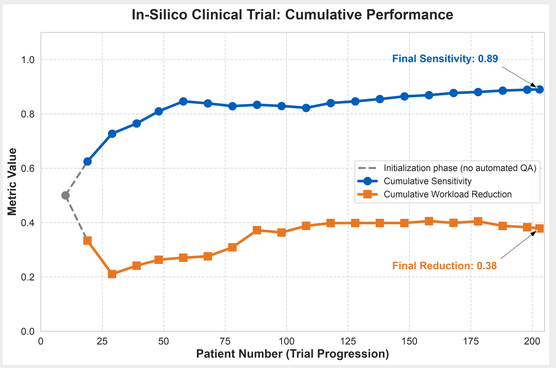

Figure 4. Cumulative sensitivity and potential workload reduction across batches of 10 cases, in-silico simulation.

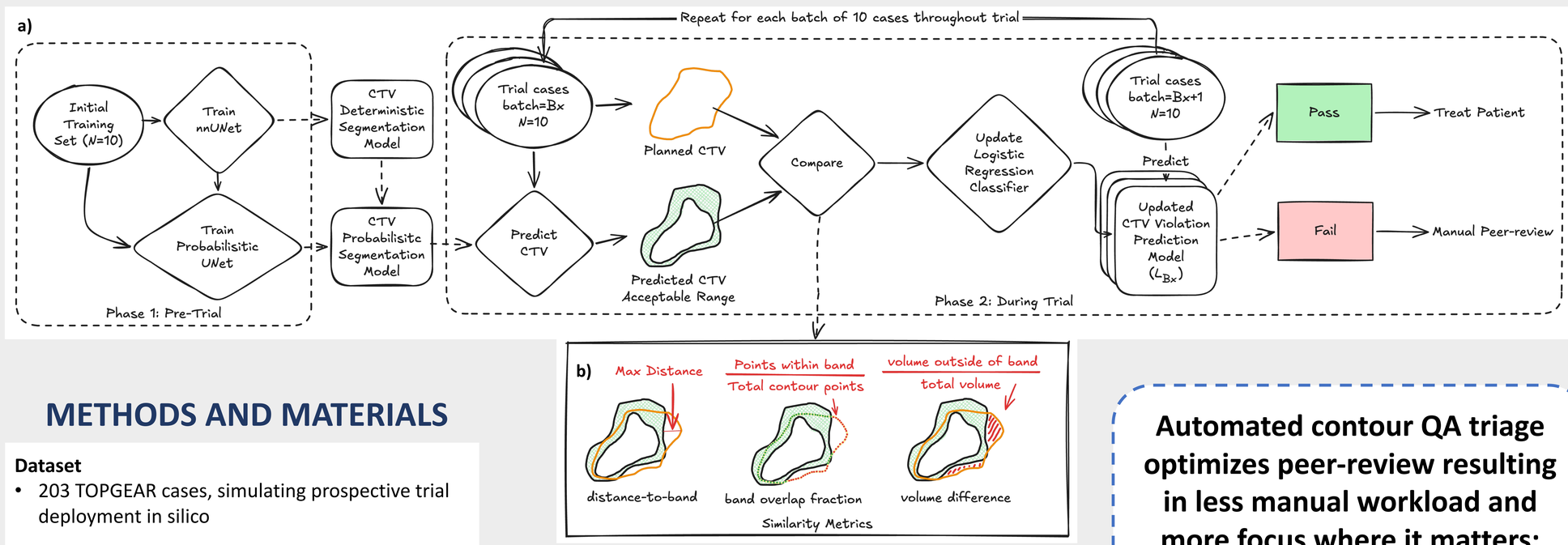

Figure 2. a) In-silico trial QA workflow. **b)** Similarity metrics between manual CTV and predicted band.

	Manual Review Violation	Manual Review Acceptable
Auto QA flagged for review	65	55
Auto QA Acceptable	8	65

Figure 3. Confusion matrix for automated contour QA, all in-silico cases.

DISCUSSION

Model Tuning
- Target sensitivity **0.9:** hard to hit early, converged by trial end (**0.89**, Fig. 4)
- **0.85** cumulative sensitivity after 5 iterations
- → early trial stages likely benefit from added manual oversight

Automated vs. Standard QA
- Standard TOPGEAR sampling: **71%** of patients manually reviewed
- Automated triage: **62%** reviewed with **100%** coverage
- 8 violations missed automatically vs. an estimated **~16 missed** under standard sampling (29% unreviewed × 28% violation rate)

Automated contour QA triage optimizes peer-review resulting in less manual workload and more focus where it matters: protocol violations.

REFERENCES

1. Leong T, Smithers BM, Michael M, Haustermans K, Wong R, Gebski V, et al. Preoperative Chemoradiotherapy for Resectable Gastric Cancer. New England Journal of Medicine 2024;391:1810–21. doi.org/10.1056/NEJMoa2405195
2. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18:203–11. doi.org/10.1038/s41592-020-01008-z
3. Chlap P, Min H, Dowling J, Field M, Cloak K, Leong T, et al. Uncertainty Estimation using a 3D Probabilistic UNet for Segmentation with Small Radiotherapy Clinical Trial Datasets. Computerized Medical Imaging and Graphics. Published online June 2024:102403. doi:10.1016/j.compmedimag.2024.102403

UNSW THE UNIVERSITY OF NEW SOUTH WALES
Ingham Institute Applied Medical Research
NSW GOVERNMENT Health South Western Sydney Local Health District

Scan to find out more
github.com/pchlap/thesis