How reliably can AI-generated OAR contours be used for prostate radiotherapy without compromising clinical dose constraints?

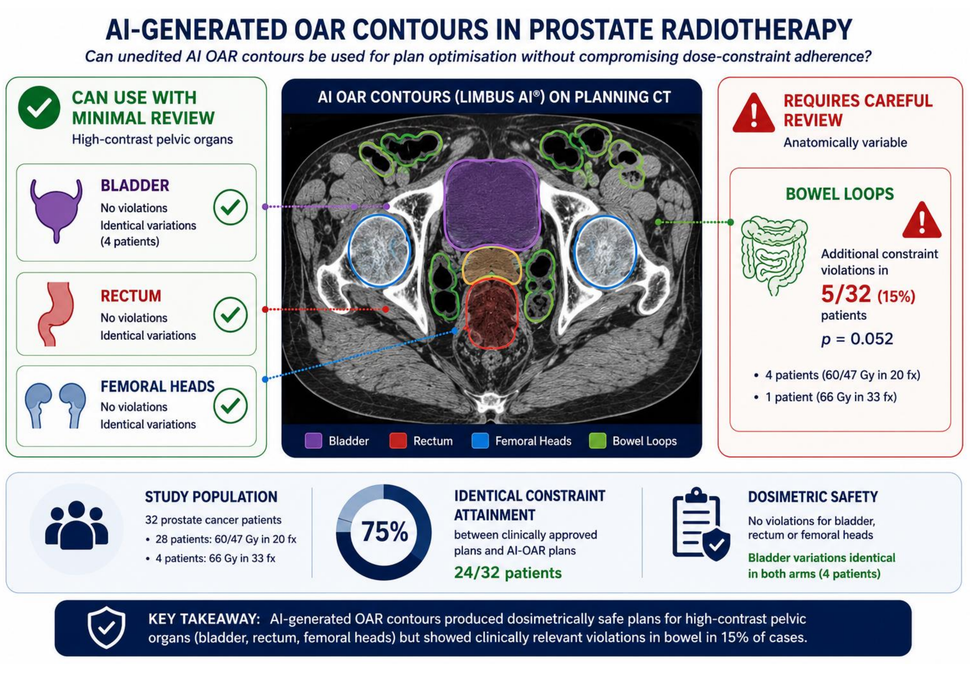
A. Nanda[1], D. Sandys[1], G. Bell[1], D. Brand[1,2]
[1]University College London Hospitals NHS Foundation Trust
[2]University College London University

Background

- Auto-contouring using artificial intelligence (AI) is increasingly adopted in radiotherapy (RT) planning to reduce contouring workload and variability.
- Most validation studies rely on geometric similarity metrics (e.g. Dice coefficient, Hausdorff distance) but geometric agreement does not reliably predict dosimetric safety.
- Few studies assess whether AI generated OARs can be used directly for plan optimisation without compromising dosimetric safety.
- This study evaluates whether unedited AI-generated OAR contours from a commercial system (Limbus AI®) can be used directly for prostate radiotherapy plan optimisation while maintaining institutional dose-constraint adherence.

Methods

- All prostate cancer patients treated between 16/09/2024 and 21/02/2025 who received photon therapy and had a suitable RapidPlan™ model were included.
- For each patient, clinician-defined and approved target volumes (TVs) were retained.
- Retrospective plans were optimised using RapidPlan™, based on Limbus AI-generated OAR contours, and evaluated against institutional dose constraints which were calculated on the original clinical (clinician-defined) contours.
- Violations (failure to meet mandatory constraints) and variations (failure to meet optimal constraints) were recorded for rectum, bladder, femoral heads, bowel and anal canal using institutional dose constraints specific to each prescription regimen (60/47 Gy in 20 fractions or 66 Gy in 33 fractions).
- The proportion of patients with violations or variations in the new AI-OAR group was compared with the original clinically approved plans using Fisher's exact test.



Conclusion

- AI-generated OAR contours produced dosimetrically safe plans for bladder, rectum and femoral heads but showed clinically relevant violations in bowel in 15% of cases. All were higher dose constraints but the absolute volumes were low.
- These findings suggest that efficiency gains with AI could be maximised with minimal review for high-contrast pelvic organs and full review for anatomically variable bowel loops.
- Human review remains essential for mobile, low-contrast or anatomically complex OARs such as bowel loops.
- The findings support a "hybrid AI–human" contouring model:
  - minimal review for reliable high-contrast structures
  - full review for high-risk or error-prone regions.
- Full autonomous contouring without clinician oversight is not currently supported by the available evidence.

Results

- Thirty-two patients were included (28 receiving 60/47 Gy; 4 receiving 66 Gy).
- Across all plans, 24/32 (75%) showed identical constraint attainment between clinically approved and AI-OAR plans.
- **No violations** occurred for rectum, bladder or femoral heads in either planning method, and bladder variations (7 variations in 4 patients) were identical in both arms. Table1.
- Differences arose predominantly in bowel structures:
  - AI-OAR plans produced one additional bowel constraint violation in 5/32 patients (15%) compared with none in the clinically approved plans (p = 0.052).
  - See Table 2 for details.

	Human Approved No. of violations	AI Optimised	P-value Fisher's exact
RECTUM	0 variations 0 violations	0 variations 0 violations	1.0 1.0
BLADDER	7 variations (in 4 patients) 0 violations	7 variations (in 4 patients) 0 violations	1.0 1.0
FEMORAL HEAD	0 variations 0 violations	0 variations 0 violations	1.0 1.0
BOWEL	5 variations (in 4 patients) 0 violations	5 variations (in 5 patients) **5 variations (in 5 patients)**	0.77 0.052

Table 1 – Violations and variations per organ at risk

Bowel Violation Cases	Prescription	Metric	Actual (cm^3)
1	66 Gy	V65 Gy < 0.1 cm^3	1.86
2	60/47 Gy	V52 Gy ≤ 0.0 cm^3	0.08
3	60/47 Gy	V52 Gy ≤ 0.0 cm^3	0.26
4	60/47 Gy	V52 Gy ≤ 0.0 cm^3	1.51
5	60/47 Gy	V52 Gy ≤ 0.0 cm^3	0.13

Table 2 – Bowel Violations in AI-optimised cases at risk