Geometric Accuracy is a Poor Predictor of Dosimetric Reliability: A Multi-Modality Analysis of Inter-Observer Variability in Pelvic Radiotherapy

Brendan Wright, MSc[1]; Jake Kendrick, PhD[1]; Lois Holloway, PhD[2,3], Dale Roach, PhD[4], Sandie Smith, B Appl Sci[2], Michael Jameson, PhD[5], Sean Bydder, FRANZCR[6], Jeremy Croker, FRANZCR[6], Rohen White, FRANZCR[6], Martin A Ebert, PhD[1,6]

[1]University of Western Australia; [2] Liverpool & Macarthur Cancer Therapy Centres, New South Wales, Australia; [3]University of New South Wales, Australia; [4]North Coast Cancer Institute, New South Wales, Australia; [5]GenesisCare, Australia; [6]Sir Charles Gairdner Hospital, Western Australia

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

INTRODUCTION

- Conventionally, radiotherapy treatment planning is grounded on binary definitions of anatomical structures.
- Binary definitions do not readily accommodate geometric uncertainty.
- This study examined the relationship between conventional binary structure overlap assessments and dosimetry, with a view to combining the greatest source of geometric uncertainty – inter-observer uncertainty – into the plan assessment process to enable more robust planning and plan assessment.
- Plans for 5 exemplar prostate patients were created for each of 10 observer contour sets.
- Dose metrics across all contour/plan combinations were extracted to quantify the potential impact of observer variation on dose reporting and predictive modelling.
- Variations in dosimetry were correlated with observer variations quantified via spatial overlap metrics.
- Note that results for Cyberknife treatments, originally identified in the submitted abstract, are now pending a review.

CONCLUSIONS

- Evaluation metrics, when assessed against plans for another observer, generally show considerable variation, exceeding objectives and capable of predicting very different treatment outcomes.
- This has implications for the accuracy and robustness of dose-volume objectives and prediction models derived from treated cohorts.
- Methods are required to incorporate observer variability in both derivation of plan parameters and for robust plan optimization.

CONTACT

Dr Martin A Ebert
Sir Charles Gairdner Hospital,
Western Australia
Martin.Ebert@health.wa.gov.au



METHODS AND MATERIALS

- CT and MR data from five pelvic radiotherapy cases were utilized, selected to be exemplars representing the anatomical diversity in a 750-patient cohort, with contours for target and OARs generated by multiple expert observers.
- VMAT plans were prescribed and planned for each observer's contours for each exemplar patient using Pinnacle autoplanning.
- Plan objectives included:
 - CTV $D_{99\%} \geq 78.0$ Gy
 - PTV $D_{95\%} \geq 78.0$ Gy
 - Rectum $V_{70Gy} \leq 10\%$
 - Bladder $V_{60Gy} \leq 30\%$
- Dosimetric indices were extracted for each observer/plan combination.
- An isotropic 7 mm CTV-to-PTV margin was intended to accommodate geometric uncertainties.
- The robustness of plans to inter-observer uncertainty was assessed through extraction of plan objective metrics across combinations.
- Geometric overlap measures – surface Dice Similarity Coefficient (sDSC), 95th percentile Hausdorff Distance (HD95) – were correlated with differences in dosimetric indices in pairwise comparisons.
- The correlation between geometric agreement and radiobiological metrics – tumour control probability (TCP) and normal tissue complication probability (NTCP) – was also assessed.

RESULTS

- Results represent approximately 650 comparisons across the 5 exemplar patients and 10 observers. Majority VOTE contour and STAPLE (Simultaneous Truth and Performance Level Estimation) contours, generated across all observers, were also included in the evaluation where applicable.
- Figure 1 indicates the stability of optimization objectives when compared across pairs of plans/observers. Per-patient results are summarised in Table 1. The large range of dosimetric indices well beyond the objective values indicates the lack of plan robustness to inter-observer uncertainty.
- There was a weak correlation between contour overlap metrics and all dose metrics extracted in cross-wise pairs of plans/observers. Table 2 shows Spearman correlation (ρ) between considered metrics, and Figure 2 shows plots of correlated metrics when $|\rho| > 0.3$.
- It was found that absolute volume of a structure and the spatial distribution of contouring variability each contribute independently to dosimetric uncertainty.
- For CTV TCP, rectum NTCP and bladder NTCP, there were 0.4%, 3.3% and 0.3% median pairwise differences between each observer's plans.
- Figure 3 indicates the variation in radiobiological metrics evaluated on pairwise plan/observer combinations. Table 3 provides a patient level summary. NTCP is impacted by observer and reflects the potential "smearing out" of a predictive model.
- Table 4 shows correlation between geometric and radiobiological metrics, indicating weak associations.

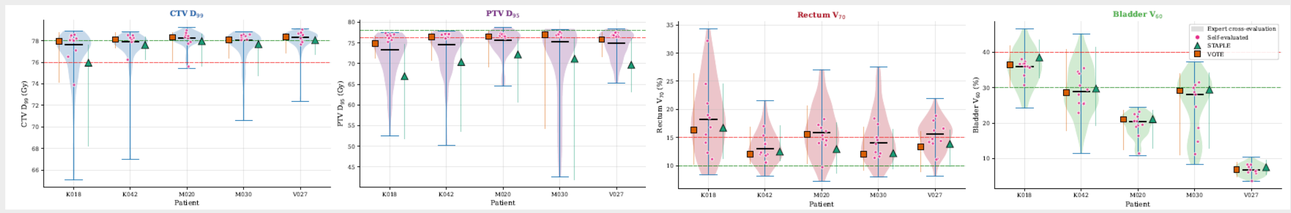


Figure 1. DVH uncertainty distributions per patient across all plan–contour combinations. Violin bodies show expert cross-evaluations only; pink dots indicate self-evaluated values; triangles and squares indicate STAPLE and VOTE evaluation contours respectively. Protocol thresholds are shown as dashed lines.

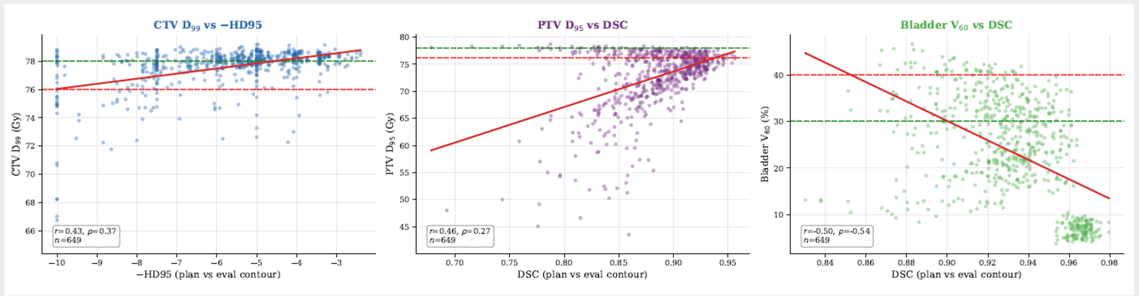


Figure 2. Best-predicting geometric overlap metric vs raw DVH value (panels shown only where Spearman $|\rho| \geq 0.3$).

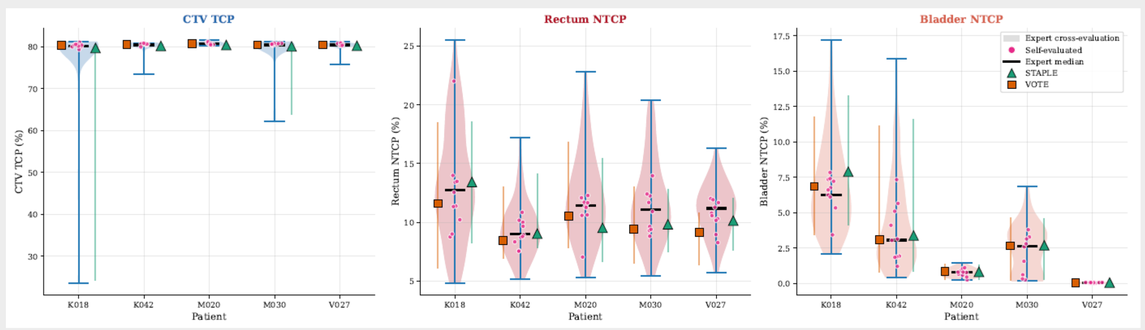


Figure 3. TCP and NTCP distributions per patient across all plan–contour combinations. Violin bodies show expert cross-evaluations; pink dots indicate self-evaluated values; triangles and squares indicate STAPLE and VOTE evaluation contours respectively.

Table 1. Variation in dose metrics for each patient across all plan/observer combinations. Dose in Gy, volume in %.

Patient	Metric	Median	Range	Span
K018	CTV D_{99}	77.60	77.6 [65.1–78.9]	13.80
K018	PTV D_{95}	73.20	73.2 [51.8–77.4]	25.60
K018	Rectum V_{70}	17.80	17.8 [8.4–34.3]	25.90
K018	Bladder V_{60}	36.00	36.0 [24.2–46.6]	22.40
K042	CTV D_{99}	77.90	77.9 [67.0–78.8]	11.80
K042	PTV D_{95}	74.50	74.5 [50.2–77.8]	27.60
K042	Rectum V_{70}	12.60	12.6 [8.1–21.6]	13.40
K042	Bladder V_{60}	28.90	28.9 [11.4–45.2]	33.80
M020	CTV D_{99}	78.20	78.2 [75.4–79.2]	3.80
M020	PTV D_{95}	75.70	75.7 [60.7–78.7]	17.90
M020	Rectum V_{70}	15.50	15.5 [7.3–27.1]	19.80
M020	Bladder V_{60}	20.60	20.6 [10.7–24.4]	13.70
M030	CTV D_{99}	78.00	78.0 [70.6–78.8]	8.20
M030	PTV D_{95}	75.20	75.2 [41.9–78.3]	36.40
M030	Rectum V_{70}	13.60	13.6 [8.0–27.6]	19.60
M030	Bladder V_{60}	28.30	28.3 [8.4–37.3]	28.90
V027	CTV D_{99}	78.30	78.3 [72.4–79.1]	6.70
V027	PTV D_{95}	74.70	74.7 [63.1–78.3]	15.30
V027	Rectum V_{70}	15.00	15.0 [8.2–21.9]	13.70
V027	Bladder V_{60}	6.80	6.8 [3.7–10.4]	6.70

Table 2. Spearman ρ between contour overlap metric and dose metric for all plan/observer combinations. *p < 0.05.

Endpoint	DSC	sDSC	−HD95	n
CTV D_{99}	0.27*	0.28*	0.37*	649
PTV D_{95}	0.27*	0.28*	0.31*	649
PTV D_{50}	0.11*	0.10*	0.12*	649
PTV D_{2}	0.15*	0.13*	0.06	649
Rectum V_{60}	-0.19*	-0.13*	-0.04	638
Rectum V_{65}	-0.19*	-0.12*	-0.04	638
Rectum V_{70}	-0.18*	-0.12*	-0.05	638
Bladder V_{50}	-0.55*	-0.14*	-0.35*	649
Bladder V_{60}	-0.54*	-0.14*	-0.34*	649

Table 3. Variation in radiobiological metrics for each patient across all plan/observer combinations. Values in %.

Patient	Metric	Value	Span
K018	CTV TCP	80.2 [23.6–81.2]	57.60
K018	Rectum NTCP	12.6 [4.8–25.5]	20.70
K018	Bladder NTCP	6.4 [2.1–17.2]	15.10
K042	CTV TCP	80.5 [73.5–81.0]	7.50
K042	Rectum NTCP	9.0 [5.2–17.2]	12.10
K042	Bladder NTCP	3.1 [0.4–15.9]	15.40
M020	CTV TCP	80.7 [80.1–81.6]	1.50
M020	Rectum NTCP	11.3 [5.3–22.8]	17.50
M020	Bladder NTCP	0.8 [0.2–1.5]	1.20
M030	CTV TCP	80.5 [62.2–81.2]	18.90
M030	Rectum NTCP	10.9 [5.4–20.4]	15.00
M030	Bladder NTCP	2.7 [0.2–6.9]	6.70
V027	CTV TCP	80.5 [75.8–81.2]	5.40
V027	Rectum NTCP	10.7 [5.7–16.4]	10.60
V027	Bladder NTCP	0.1 [0.0–0.1]	0.10

Table 3. Spearman ρ between contour overlap metric and dose metric for all plan/observer combinations. *p < 0.05.

Endpoint	DSC	sDSC	−HD95	n
CTV TCP	0.26*	0.26*	0.31*	649
Rectum NTCP	-0.16*	-0.11*	-0.05	638
Bladder NTCP	-0.55*	-0.15*	-0.35*	649

© POSTER TEMPLATE BY GENIGRAPHICS® 1.800.790.4001 WWW.GENIGRAPHICS.COM