

Clinical Impact of Contouring Variability for Prostate Cancer Tumor Boost

Authors: Allison Y. Zhong¹; Asona J. Lui, MD, PhD¹; Svetlana Kuznetsova, PhD¹; Karoline Kallis, PhD¹; Christopher Conlin, PhD²; Deondre D. Do³; Mariluz Rojo Domingo³; Ryan Manger, PhD¹; Patricia Hua, CMD¹; Roshan Karunamuni, PhD¹; Joshua Kuperman, PhD²; Anders M. Dale, PhD^{2,4,5}; Rebecca Rakow-Penner, MD, PhD^{2,3}; Michael E. Hahn, MD, PhD²; Uulke A. van der Heide, PhD⁶; Xenia Ray, PhD¹; Tyler M. Seibert, MD, PhD^{1,2,3}

¹Department of Radiation Medicine and Applied Sciences, UC San Diego School of Medicine, La Jolla, CA, USA

²Department of Radiology, UC San Diego School of Medicine, La Jolla, CA, USA

³Department of Bioengineering, UC San Diego Jacobs School of Engineering, La Jolla, CA, USA

⁴Department of Neurosciences, UC San Diego School of Medicine, La Jolla, CA, USA

⁵Halicioğlu Data Science Institute, UC San Diego School of Medicine, La Jolla, CA, USA

⁶Department of Radiation Oncology, The Netherlands Cancer Institute (NKI-AVL), Amsterdam, The Netherlands

Abstract:

Purpose: The focal radiotherapy (RT) boost technique was shown in the FLAME trial to improve prostate cancer outcomes without increasing toxicity. This technique relies on the accurate delineation of prostate tumors on MRI. The ReIGNITE RT Boost study evaluated radiation oncologists' accuracy when asked to delineate prostate tumors on MRI and demonstrated high variability in tumor contours. We sought to evaluate the impact of contour variability and inaccuracy on predicted clinical outcomes. We hypothesized that radiation oncologists' contour inaccuracies would yield meaningfully worse clinical outcomes.

Materials & Methods: 44 radiation oncologists and 2 expert radiologists contoured prostate tumors on 30 patient cases. Of these cases, those with CT simulation or diagnostic CT available were selected for analysis. A knowledge-based planning model was developed to generate focal RT boost plans for each contour per the FLAME trial protocol. Probability of biochemical failure (BF) was determined using a model from the FLAME trial. The primary metric evaluated was delta BF ($\Delta\text{BF} = \text{Participant BF} - \text{Expert BF}$). An absolute increase in BF $\geq 5\%$ was considered clinically meaningful.

Results: 8 patient cases and 394 target volumes for focal RT boost planning were included in this analysis. In general, participant plans were associated with worse predicted clinical outcomes compared to the expert plan, with an average absolute increase in BF of 4.3%. 37% of participant plans were noted to have an absolute increase in BF of 5% or more.

Conclusion: Radiation oncologists' attempts to contour tumor targets for focal RT boost are frequently inaccurate enough to yield meaningfully inferior clinical outcomes for patients.

Introduction:

In the FLAME phase III randomized controlled trial, adding a focal radiotherapy (RT) boost to prostate tumors visible on multiparametric MRI improved disease-free and regional/distant metastasis-free survival for patients with intermediate- and high-risk prostate cancer, without increasing toxicity [1,2]. This trial demonstrated that the probability a patient would experience biochemical failure (cancer recurrence) was predicted by the RT dose delivered to the visible tumor. Higher tumor doses yielded better treatment outcomes, and lower tumor doses yielded more treatment failures.

The focal RT boost technique relies on the accurate delineation of prostate tumors on MRI. A recent prospective study, ReIGNITE RT Boost, evaluated radiation oncologists' accuracy when asked to delineate prostate tumors on MRI [3]. High variability was observed in tumor contours. When using conventional MRI alone, participants completely missed the tumor (zero overlap with the true target) in a median of 13.6% of attempts (IQR 9.1-23.6%). Accuracy and reliability of tumor contours improved considerably when participants were given advanced diffusion MRI maps (called the Restriction Spectrum Imaging restriction score, or RSIrs) that improve prostate cancer conspicuity. Complete misses, for example, were much more uncommon when using RSIrs (median 0.0%, IQR 0.0% - 4.3%). Nonetheless, contours

still varied considerably between radiation oncologists. Given smooth radiation dose plans and RT penumbra, it is unclear how near misses and partial overlap will affect patient outcomes.

Here, we evaluate the impact of contour variability and inaccuracy on predicted clinical outcomes. We measured the RT dose delivered to the true tumor if a ReIGNITE participant's contour were used to generate the focal RT boost plan instead of using the guidance of an expert radiologist, as was done in the FLAME trial. Using the model from the FLAME trial [2], we then calculated the probability of biochemical failure for each RT plan. We hypothesized that radiation oncologists' contour inaccuracies would yield meaningfully worse clinical outcomes.

Methods & Materials:

ReIGNITE Study

44 radiation oncologists with a range of experience treating prostate cancer were enrolled as participants. All study recruitment materials, communications, and procedures were approved by the Institutional Review Board (IRB). Participants were asked to contour tumors on 20 patient cases of clinically localized intermediate- or high-risk prostate cancer in each of two sessions at least 1 month apart. In each case, participants were provided either conventional MRI alone or conventional MRI with RSIs overlay. 10 cases were repeated between the two sessions, with RSIs either added or removed; the participants were not made aware of this. Provided MRI sequences included T2-weighted, ADC, and DWI ($b=0$ and $b=2000$ s/mm²). In addition,

participants were given clinical information regarding each case: including the patient's age, PSA at time of MRI, number of positive biopsy cores, location of positive biopsy cores, and radiologic description of the tumor. Contouring was performed on the MIM Zero Footprint™ (ZFP) platform (MIM Software, Cleveland, OH).

Expert-defined lesions were created by the consensus of a radiation oncologist (with three years of experience) and two board-certified, subspecialist genitourinary radiologists (with five years and seven years of experience, respectively). This was done for all patient cases on conventional MRI alone, as the clinical standard per the FLAME trial is to target the tumor visible on MRI.

For this present study, we included all patient cases from ReIGNITE for which either a CT simulation or a diagnostic CT was available.

OAR and Target Volume Definitions

Where a CT simulation was available, organs at risk (OARs) had been contoured per clinical routine. For those with only a diagnostic CT, OARs were contoured manually. T2-weighted MRI for each case was rigidly co-registered to the CT to optimize registration specifically of the prostate, and the corresponding transformation was applied to the expert-defined tumor target, as well as to each participant's target contour.

Knowledge-based RT Planning

To ensure RT planning was unbiased, we developed a knowledge-based planning (KBP) automated algorithm to generate RT plans with focal tumor boost per the FLAME trial protocol: 77 Gy in 35 fractions to the whole prostate and an integrated boost up to 95 Gy to the focal target, provided no normal tissue constraints were violated. To facilitate the development of the KBP algorithm, we used Varian RapidPlan (Varian Medical Systems, Palo Alto, USA), a vendor-based solution for knowledge based planning [4] that is widely used in the field [5–10]. A total of 62 clinical plans with contoured targets (prostate and boosted lesion) and OARs (bladder, rectum, bowel, sigmoid, urethra, penile bulb, right femur, and left femur) were added to the training set for the model. RapidPlan then uses geometrical features extracted between the set of targets and OARs and the OAR dose-volume histograms (DVH) to build models that can predict the achievable OAR DVHs for a new patient. Additionally, the set of OAR DVH models is linked to a template of weighted optimization metrics that use the patient-specific predicted dose values to optimize and calculate dose for new plans. This allows for fully automated, high-quality planning. The plan optimization metrics for our model (**Table 1**) reflect those used in the FLAME trial as well as an additional urethra constraint proposed in a subsequent normal tissue complication probability analysis of FLAME participants [11].

We tested the KBP model on 10 unrelated patient cases from our institution that were treated with focal RT boost. A dosimetrist manually generated optimal plans using the FLAME protocol for each of the 10 cases, using the focal tumor target actually treated for that patient. KBP plans were generated for the 10 cases, and the resulting

dose metrics and 3D dose distributions were compared to the clinical plan used for those patients to validate the model (**Table 2**). To further validate the KBP model, we confirmed that it gave expected results for focal RT plans using the expert contours (**Table 3**). We then applied the algorithm to each participant's tumor contour and compared dosimetric parameters to those achieved when using the expert-defined tumor.

RT Plan Evaluation & Statistical Tests

For each plan, we calculated the dose covering 98% of the expert-defined tumor (D98), which was associated with probability of biochemical failure (BF) at 7 years in a model from the FLAME trial (**Figure 1**). The primary metric we evaluated was delta BF probability (Δ BF), which was calculated as (Participant BF – Expert BF). A contour that led to a $\geq 5\%$ increase in BF was considered clinically meaningful. The percentage of participants whose contours yielded a clinically meaningful BF increase was also calculated for each case. Secondary metrics included D98, BF, delta D98 (Δ D98 = Participant D98 – Expert D98), and the OAR dose metrics.

Case-wise t-tests were performed for Δ BF and Δ D98 for all patient cases. For cases that were contoured both with and without RSIs, paired t-tests were performed to compare Δ BF and Δ D98 with and without RSIs.

Results:

8 patient cases from the ReIGNITE study had available pelvic CT scans and were included for analysis (**Table 4**). 2 cases had CT simulation available, while 6 cases had diagnostic CT only. 3 cases were contoured on conventional MRI only, 2 cases were contoured on conventional MRI + RSIs only, and 3 cases were contoured on both with and without RSIs. In total, 394 participant target volumes and 8 expert target volumes were used to generate RT plans. All plans had adequate coverage of the prostate and met all key normal tissue constraints.

In general, using radiation oncologist participants' target volumes yielded worse BF and D98 values compared to the expert target volume (**Figure 2**). Across all participants' plans, BF increased by an average of 4.3%. A clinically meaningful absolute increase in BF of $\geq 5\%$ was seen in 37% of the participant plans.

Case-wise t-tests comparing participant plans to expert plans demonstrated a statistically significant difference for Δ BF and Δ D98 in 6 of 8 patient cases ($p < 0.01$) (**Table 5**). Among the 3 cases that had contours with and without RSI (Patients 6-8), paired t-tests showed a statistically significant improvement in both Δ BF and Δ D98 with RSIs for Patients 7 and 8.

Discussion:

Overall, using radiation oncologist participants' target volumes yielded plans with worse predicted clinical outcomes compared to the expert target volume. On average, participants' target volumes led to a predicted 4.3% absolute increase in BF. In more

than one-third of participant plans (37%), the absolute increase in BF exceeded our *a priori* threshold for clinical significance ($\geq 5\%$ absolute increase in BF). Thus, the high tumor contour variability demonstrated in the ReIGNITE study is expected to adversely impact clinical outcomes for patients.

ReIGNITE showed that one way to improve radiation oncologists' contouring accuracy is to provide RSIs maps [3]. The present study was not powered for full evaluation of the impact of RSIs on treatment plans, with only 3 cases included. Nonetheless, of the 3 cases that allowed for comparison of contours with and without RSIs, 2 cases showed a statistically significant improvement in ΔBF and $\Delta D98$ with RSIs. Other strategies to improve contouring accuracy for MRI-visible prostate tumors include involvement of expert radiologists in target delineation and dedicated training for radiation oncologists [12,13]. Alternative imaging modalities like PSMA PET may also prove useful [14–18]. Automated target delineation using artificial intelligence has potential, though these currently still require considerable expert oversight [19,20]. For less experienced or resourced settings, one could consider adding a margin around the gross tumor volume to account for contouring uncertainties. Future studies should measure both the feasibility of these strategies and their impact on predicted clinical outcomes of focal RT boost. Meanwhile, it is nonetheless noted that all calculated BFs in the present study were lower than that if the prostate were treated only with a uniform dose of 77 Gy, illustrating that even somewhat inaccurate contours may still benefit the patient.

In some instances, an imprecise contour led to a higher dose than that seen in the expert plan. This was especially prevalent for Patients 5, 6, and 8_{RSI}. Upon

inspection of these cases, it appeared that participants tended to draw their contours larger than the expert-defined lesion, while encompassing the true target (**Figure 3**), which likely spread out the high dose away from OARs. In clinical practice, a dosimetrist would recognize the potential for improving a plan with the correct target, e.g., by inserting an optimization structure to better target the lesion. This indicates that the KBP model used in this study has room for further improvement to achieve more optimal planning.

The present study inherits some limitations from the ReIGNITE study that serves as the source of data. Images used in ReIGNITE were from a single institution, which may limit generalizability. Radiation oncologist participants, on the other hand, were from 9 countries and many institutions. Additionally, only 8 of the 30 patient cases from ReIGNITE were analyzed here. Still, we observed statistically and clinically significant results, suggesting realistic variability in target volumes can impact patient outcomes.

Conclusion:

Radiation oncologists' attempts to contour tumor targets for focal RT boost are frequently inaccurate enough to yield meaningfully inferior clinical outcomes for patients. Further studies are warranted to investigate the impact of RSIs on predicted clinical outcomes of focal RT boost.

References:

1. Groen VH, Haustermans K, Pos FJ, Draulans C, Isebaert S, Monninkhof EM, Smeenk RJ, Kunze-Busch M, de Boer JCJ, van der Voort van Zijp J, Kerkmeijer LGW, van der Heide UA. Patterns of Failure Following External Beam Radiotherapy With or Without an Additional Focal Boost in the Randomized Controlled FLAME Trial for Localized Prostate Cancer. *European Urology*. 2022 Sep;82(3):252–7.
2. Kerkmeijer LGW, Groen VH, Pos FJ, Haustermans K, Monninkhof EM, Smeenk RJ, Kunze-Busch M, de Boer JCJ, van der Voort van Zijp J, van Vulpen M, Draulans C, van den Bergh L, Isebaert S, van der Heide UA. Focal Boost to the Intraprostatic Tumor in External Beam Radiotherapy for Patients With Localized Prostate Cancer: Results From the FLAME Randomized Phase III Trial. *JCO*. 2021 Mar 1;39(7):787–96.
3. Lui AJ, Kallis K, Zhong AY, Hussain TS, Conlin C, Digma LA, Phan N, Mathews IT, Do DD, Domingo MR, Karunamuni R, Kuperman J, Dale AM, Shabaik A, Rakow-Penner R, Hahn ME, Seibert TM. ReIGNITE Radiation Therapy Boost: A Prospective, International Study of Radiation Oncologists' Accuracy in Contouring Prostate Tumors for Focal Radiation Therapy Boost on Conventional Magnetic Resonance Imaging Alone or With Assistance of Restriction Spectrum Imaging. *International Journal of Radiation Oncology*Biography*Physics*. 2023 Jul;S0360301623076290.
4. Fogliata A, Belosi F, Clivio A, Navarria P, Nicolini G, Scorsetti M, Vanetti E, Cozzi L. On the pre-clinical validation of a commercial model-based optimisation engine:

Application to volumetric modulated arc therapy for patients with lung or prostate cancer. *Radiotherapy and Oncology*. 2014 Dec;113(3):385–91.

5. Ray X, Kaderka R, Hild S, Cornell M, Moore KL. Framework for Evaluation of Automated Knowledge-Based Planning Systems Using Multiple Publicly Available Prostate Routines. *Practical Radiation Oncology*. 2020 Mar;10(2):112–24.
6. Kaderka R, Mundt RC, Li N, Ziemer B, Bry VN, Cornell M, Moore KL. Automated Closed- and Open-Loop Validation of Knowledge-Based Planning Routines Across Multiple Disease Sites. *Practical Radiation Oncology*. 2019 Jul;9(4):257–65.
7. Li N, Carmona R, Sirak I, Kasaova L, Followill D, Michalski J, Bosch W, Straube W, Mell LK, Moore KL. Highly Efficient Training, Refinement, and Validation of a Knowledge-based Planning Quality-Control System for Radiation Therapy Clinical Trials. *International Journal of Radiation Oncology*Biography*Physics*. 2017 Jan;97(1):164–72.
8. Hussein M, South CP, Barry MA, Adams EJ, Jordan TJ, Stewart AJ, Nisbet A. Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy. *Radiotherapy and Oncology*. 2016 Sep;120(3):473–9.
9. Chua LMH, Pang EPP, Master Z, Sultana R, Tuan JKL, Bragg CM. Dosimetric comparison of RapidPlan and manually optimised volumetric modulated arc therapy plans in prostate cancer. *J Radiother Pract*. 2021 Sep;20(3):257–64.

10. Scaggion A, Fusella M, Roggio A, Bacco S, Pivato N, Rossato MA, Peña LMA, Paiusco M. Reducing inter- and intra-planner variability in radiotherapy plan output with a commercial knowledge-based planning solution. *Physica Medica*. 2018 Sep;53:86–93.
11. Groen VH, van Schie M, Zuithoff NPA, Monnikhof EM, Kunze-Busch M, de Boer JCJ, van der Voort van Zijp J, Pos FJ, Smeenk RJ, Haustermans K, Isebaert S, Draulans C, Depuydt T, Verkooijen HM, van der Heide UA, Kerkmeijer LGW. Urethral and bladder dose–effect relations for late genitourinary toxicity following external beam radiotherapy for prostate cancer in the FLAME trial. *Radiotherapy and Oncology*. 2022 Feb;167:127–32.
12. Zhong AY, Lui AJ, Katz MS, Berlin A, Kamran SC, Kishan AU, Murthy V, Nagar H, Seible D, Stish BJ, Tree AC, Seibert TM. Use of focal radiotherapy boost for prostate cancer: radiation oncologists’ perspectives and perceived barriers to implementation. *Radiat Oncol*. 2023 Nov 11;18(1):188.
13. Dornisch AM, Zhong AY, Poon DMC, Tree AC, Seibert TM. MRI-guided focal radiotherapy tumor boost for prostate cancer: a systematic review. *World Journal of Urology*.
14. Zamboglou C, Spohn SKB, Ruf J, Benndorf M, Gainey M, Kamps M, Jilg C, Gratzke C, Adebahr S, Schmidtmayer-Zamboglou B, Mix M, Bamberg F, Zschaeck S, Ghadjar P, Baltas D, Grosu AL. PSMA-PET- and MRI-Based Focal Dose Escalated Radiation Therapy of Primary Prostate Cancer: Planned Safety Analysis

of a Nonrandomized 2-Armed Phase 2 Trial (ARO2020-01). *International Journal of Radiation Oncology*Biology*Physics*. 2022 Aug;113(5):1025–35.

15. Bettermann AS, Zamboglou C, Kiefer S, Jilg CA, Spohn S, Kranz-Rudolph J, Fassbender TF, Bronsert P, Nicolay NH, Gratzke C, Bock M, Ruf J, Benndorf M, Grosu AL. [68Ga-]PSMA-11 PET/CT and multiparametric MRI for gross tumor volume delineation in a slice by slice analysis with whole mount histopathology as a reference standard – Implications for focal radiotherapy planning in primary prostate cancer. *Radiotherapy and Oncology*. 2019 Dec;141:214–9.
16. Draulans C, De Roover R, Van Der Heide UA, Kerkmeijer L, Smeenk RJ, Pos F, Vogel WV, Nagarajah J, Janssen M, Isebaert S, Maes F, Mai C, Oyen R, Joniau S, Kunze-Busch M, Goffin K, Haustermans K. Optimal 68Ga-PSMA and 18F-PSMA PET window levelling for gross tumour volume delineation in primary prostate cancer. *Eur J Nucl Med Mol Imaging*. 2021 Apr;48(4):1211–8.
17. Eiber M, Weirich G, Holzapfel K, Souvatzoglou M, Haller B, Rauscher I, Beer AJ, Wester HJ, Gschwend J, Schwaiger M, Maurer T. Simultaneous 68Ga-PSMA HBED-CC PET/MRI Improves the Localization of Primary Prostate Cancer. *European Urology*. 2016 Nov;70(5):829–36.
18. Zamboglou C, Thomann B, Koubar K, Bronsert P, Krauss T, Rischke HC, Sachpazidis I, Drendel V, Salman N, Reichel K, Jilg CA, Werner M, Meyer PT, Bock M, Baltas D, Grosu AL. Focal dose escalation for prostate cancer using 68Ga-HBED-CC PSMA PET/CT and MRI: a planning study based on histology reference. *Radiat Oncol*. 2018 Dec;13(1):81.

19. Tsui JMG, Kehayias CE, Leeman JE, Nguyen PL, Peng L, Yang DD, Moningi S, Martin N, Orio PF, D'Amico AV, Bredfeldt JS, Lee LK, Guthier CV, King MT. Assessing the Feasibility of Using Artificial Intelligence–Segmented Dominant Intraprostatic Lesion for Focal Intraprostatic Boost With External Beam Radiation Therapy. *International Journal of Radiation Oncology*Biography*Physics*. 2024 Jan;118(1):74–84.

20. Holzschuh JC, Mix M, Ruf J, Hölscher T, Kotzerke J, Vrachimis A, Doolan P, Ilhan H, Marinescu IM, Spohn SKB, Fechter T, Kuhn D, Bronsert P, Gratzke C, Grosu R, Kamran SC, Heidari P, Ng TSC, Könik A, Grosu AL, Zamboglou C. Deep learning based automated delineation of the intraprostatic gross tumour volume in PSMA-PET for patients with primary prostate cancer. *Radiotherapy and Oncology*. 2023 Nov;188:109774.

Table 1. Knowledge-based planning (KBP) optimization metrics.

Structure	Limit	Vol (%)	Dose (cGy)	Priority
GTV_xxx	Upper	0.0	9394	80
	Lower	100.0	9548	70
PTV78_DVH_xxx	Lower	100.0	7854	200
PTV78_Opt_xxx	Upper	0.0	7931	200
	Lower	100.0	7854	200
Bladder	Upper	1.1	7315	100
	Upper	3.7	3850	100
	Upper	8.5	1925	30
	Upper	0.0	7700	125
	Line			50
Femur L	Line			50
Femur R	Line			50
Penile Bulb	Line			50
Rectum	Upper	4.3	7315	100
	Upper	10.6	5775	90
	Upper	19.9	3850	90
	Upper	30.6	2695	90
	Upper	0.0	7700	125
	Line			50
Sigmoid	Upper	0.0	5775	70
	Upper	0.0	3850	70
	Upper	0.0	5159	125
Urethra	Upper	0.0	8085	125

Table 2. Knowledge-based planning (KBP) model validation with an independent set.

Parameter	Dose Constraint	Mean±Std
Prostate V100%	V100% ≥ 94-95%	94.84±0.30 %
Rectum D _{1cc}	D _{1cc} ≤ 77 Gy	78.29±0.27 Gy
Bladder D _{1cc}	D _{1cc} ≤ 80 Gy	79.33±1.27 Gy
Urethra D _{0.1cc}	D _{0.1cc} ≤ 80 Gy	79.91±1.05 Gy
Lesion D98	D98 > 95 Gy	85.42±3.00 Gy
Lesion D98	D98 > 120% Rx	110.93±3.89%Rx

Table 3. Results of knowledge-based planning (KBP) model on validation plans. The generated validation plans were based on expert contours. Clinical plans were generated by an experienced dosimetrist using clinical target and organ-at-risk volumes. All plans were normalized to meet the Prostate V100% metric, leading to no difference between Clinical and KBP plans. Other parameters were allowed to vary.

Parameter	Dose Constraint	Clinical Mean±Std	KBP Mean±Std	Difference (KBP-Clinical)
Prostate V100%	V100% ≥ 94-95%	95±0%	95±0%	0%
Rectum D _{1cc}	D _{1cc} ≤ 103-105 %	100.46±1.10%	100.85±0.99%	-0.39±1.74%
Bladder D _{1cc}	D _{1cc} ≤ 105%	101.56±0.34%	102.44±0.35%	-0.88±0.66%
Urethra D _{0.1cc}	D _{0.1cc} ≤ 105-107	101.88±0.51%	102.33±0.34%	-0.50±0.80%
Lesion Expert D98	D98 > 120% Rx	113.90±1.52% Rx	114.78±7.57%Rx	-0.88±6.10%

Table 4. Characteristics of 8 patient cases included in this study.

Median age, in years (IQR)		70 (67 - 78)
Median PSA at time of MRI, in ng/mL (IQR)		8.7 (7.7 – 18.8)
Clinical stage	T1c	6
	T2a	1
	T2c	1
PI-RADS v2.1 score per lesion	3	2
	4	3
	5	3
Biopsy status	Systematic	3
	Systematic & targeted	5
Worst Gleason score per lesion	3+4	4
	4+3	2
	4+4	1
	4+5	1
Underwent prostatectomy		2
Pathological stage	T2a	1
	T3b	1

Table 5. Case-wise t-tests of $\Delta D98 = \text{Participant D98} - \text{Expert D98}$ and $\Delta BF = \text{Participant BF} - \text{Expert BF}$. *Significant with $p < 0.05$.

Patient	Mean ($\Delta D98$) \pm Standard Deviation ($\Delta D98$)	p	Mean (ΔBF) \pm Standard Deviation (ΔBF)	p
Patient 1 _{RSI}	-2.22 \pm 1.14 Gy	$p << 0.01^*$	3.05 \pm 1.72%	$p << 0.01^*$
Patient 2	-6.04 \pm 3.64 Gy	$p << 0.01^*$	7.12 \pm 4.92%	$p << 0.01^*$
Patient 3	-8.90 \pm 5.33 Gy	$p << 0.01^*$	9.51 \pm 6.27%	$p << 0.01^*$
Patient 4 _{RSI}	-4.75 \pm 0.48 Gy	$p < 0.01^*$	7.31 \pm 0.90%	$p << 0.01^*$
Patient 5	-1.02\pm3.62 Gy	p=0.07	1.42\pm3.79%	p<0.05*
Patient 6	-2.61 \pm 4.62 Gy	$p << 0.01^*$	2.90 \pm 5.02%	$p < 0.01^*$
Patient 6 _{RSI}	-4.05 \pm 4.24 Gy	$p << 0.01^*$	4.11 \pm 4.92%	$p << 0.01^*$
Patient 7	-3.68 \pm 1.67 Gy	$p << 0.01^*$	5.38 \pm 2.50%	$p << 0.01^*$
Patient 7 _{RSI}	-2.48 \pm 1.33 Gy	$p << 0.01^*$	3.43 \pm 2.12%	$p << 0.01^*$
Patient 8	-1.41 \pm 1.33 Gy	$p << 0.01^*$	2.09 \pm 2.12%	$p << 0.01^*$
Patient 8 _{RSI}	-0.24\pm2.97 Gy	p = 0.67	0.22\pm3.97%	p=0.77

Figure 1. Biochemical failure model used in the FLAME trial [2].

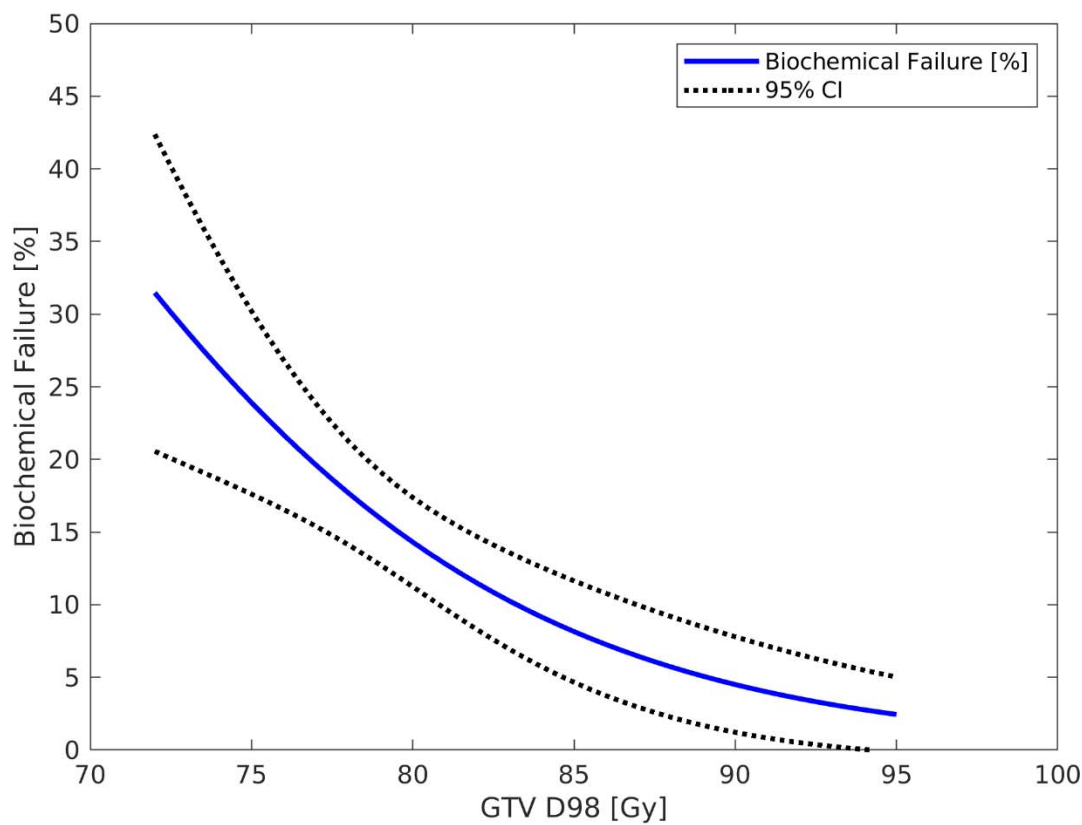


Figure 2. Violin plots for 8 patient cases for A) biochemical failure (BF), B) Δ BF = Participant BF – Expert BF, C) D98 of expert-defined tumor, and D) Δ D98 = Participant D98 – Expert D98. Predicted outcomes for the plans based on the expert contours are shown in red.

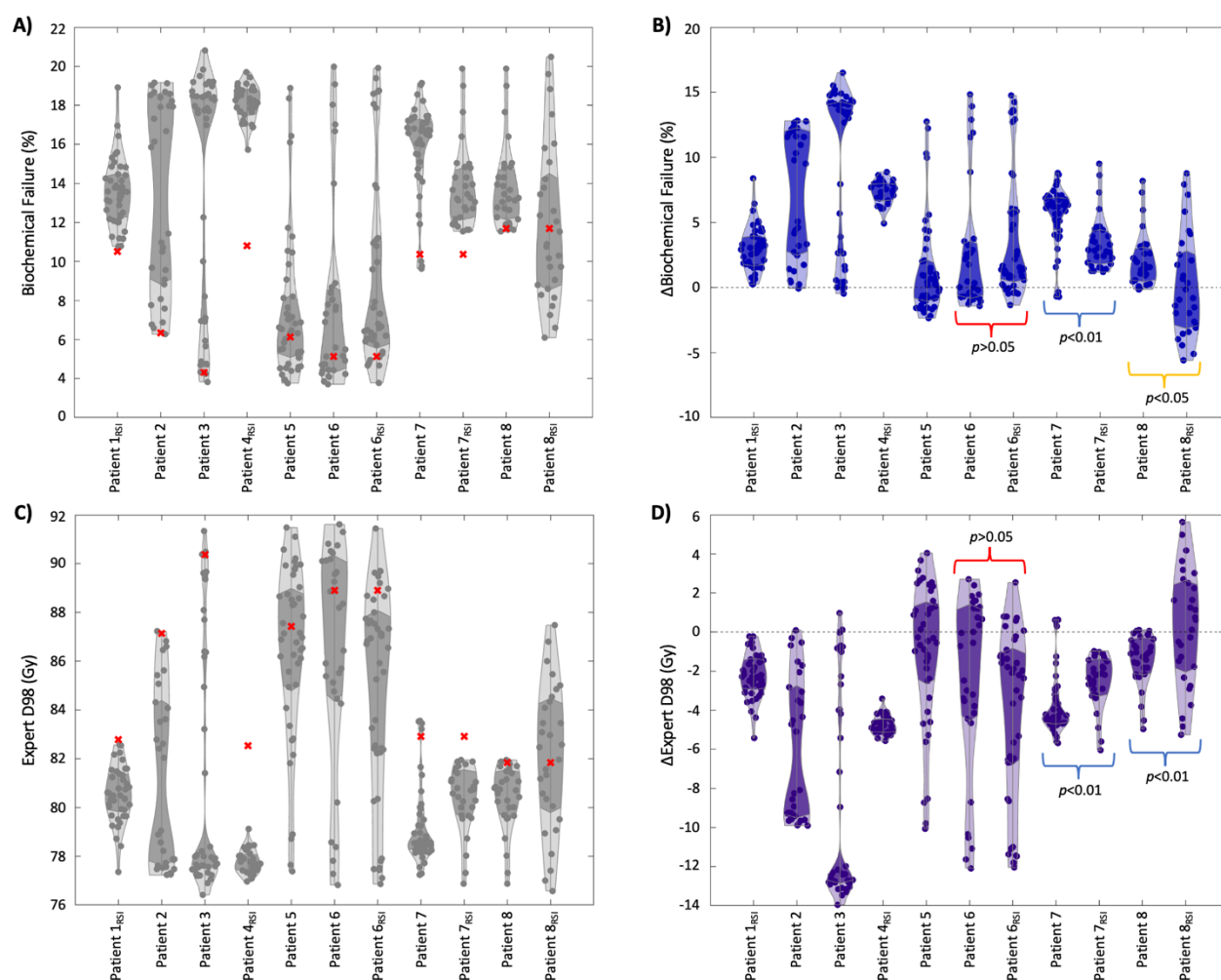


Figure 3. Example contours with dose overlays for 3 patient cases: A) Expert, B-C) Participants. Expert contours are shown in white, and participant contours are shown in blue. The greater the accuracy of the participant contour, the higher the dose delivered to the true tumor. When the participant drew the contour larger than and including the expert contour (such as in Patient 8(C)), this led to a higher dose to the true tumor than that seen in the expert plan.

