The Consequences of Organ At Risk Outline Variations in the Radiotherapy Treatment Planning for Patients with Prostate Cancer

Project Lead:
Michael Davies
(Trainee Clinical Scientist)
Swansea University Medical School
NHS SBUHB, Singleton Hospital

Supervisors:
James Williams
Douglas Etheridge
Dr. Sarah Gwynne
NHS SBUHB, Singleton Hospital
Why Does Inter-observer Outline Variation (IOV) Matter?

- Inter-observer ROI variation imparts the greatest uncertainty in the radiotherapy planning & delivery process for most tumour sites (Segedin et al. 2016).
  - IMRT: Steep dose gradients demand high accuracy.
  - Outlining impacts plan efficacy & interpretation of potential efficacy.
  - Future decisions: E.g. dose prescription & PTV margins.
  - Undermines meaningful comparison of treatments within/across centres & interpretation of clinical studies.
Introduction

What has been investigated so far?

- Past works emphasise quantitative comparison of *geometric* and *positional* variations.
  - ‘Before-and-After’ comparisons of variation after implementation of additional training/seminars/guidelines.
  - No standard metrics of comparison ⇒ Difficult to compare results across investigations.

- *Dosimetric* impact of IOV on OARs is relatively under represented in literature (only 12/31 studies: Vinod et al. 2016).
  - Those which have, observe a significant impact of IOV on ROI dose distributions (e.g. Lobefalo et al. 2013).
Key reasons for interest:

- We know IOV is significant *geometrically*. How much does it matter *dosimetrically*?
- Outlining to very high accuracies/ concordance with the considered ‘GS’ takes time!
  - Can good results be achieved with lower degrees of outlining accuracy for *some* OARs?
    → If so, patient outcome may be improved due to shorter waiting times.
Methods

Data:

- 5 Patients (duplicated).
- 3 OARs: Bladder, Bowel, Rectum.
- 7 OAR outline sets by Dosimetrists (A to G). Gold Standard (GS) outlines derived by Clinician and Senior Dosimetrist.
  - Dosimetrists were blind to duplications.
- IMRT & VMAT plans produced for 10 cases using GS outline sets by Senior Dosimetrist in Phillips Pinnacle\(^3\) version 16.0 at Singleton Hospital, Swansea.
- Prescription: 60 Gy to Prostate PTV, 48 Gy to Seminal Vesicle PTV, 20#.
Methods

Metrics of Comparison:

- Variation in mean dose to OAR across outlines for one case.
- Comparison of OAR volume proportions reported at dose constraints.
  - Instances of Mandatory or Optimal Dose Constraint failure?

- NTCP modelling
  - Lyman-Kutcher-Burman model
  - Indication of significance of impact of OAR IOV on likelihood of patient side effects.

Statistical Comparisons:

- t-test: Used to compare the mean results of one metric across outlines A to G with that of the GS.
- Coefficient of Variation: $\frac{\sigma}{\bar{x}}$, compare metrics across OARs and cases.
### Results

#### Mean Doses to OARs:

<table>
<thead>
<tr>
<th>Mod.</th>
<th>OAR:</th>
<th>Bladder</th>
<th>Bowel</th>
<th>Rectum</th>
</tr>
</thead>
<tbody>
<tr>
<td>VMAT</td>
<td>Range $D_{mean}$ (Gy)</td>
<td>13.1 to 20.3</td>
<td>1 to 6.1</td>
<td>21.3 to 34.6</td>
</tr>
<tr>
<td></td>
<td>Range CoV</td>
<td>0.013 to 0.032</td>
<td>0.039 to 0.49</td>
<td>0.051 to 0.116</td>
</tr>
<tr>
<td></td>
<td>Range t-stat</td>
<td>0.3 to 10.43</td>
<td>0.42 to 3.3</td>
<td>0.89 to 3.21</td>
</tr>
<tr>
<td>IMRT</td>
<td>Range $D_{mean}$ (Gy)</td>
<td>15.8 to 23.2</td>
<td>1.3 to 7</td>
<td>25 to 38</td>
</tr>
<tr>
<td></td>
<td>Range CoV</td>
<td>0.011 to 0.026</td>
<td>0.049 to 0.46</td>
<td>0.052 to 0.115</td>
</tr>
<tr>
<td></td>
<td>Range t-stat</td>
<td>0.45 to 11.20</td>
<td>0.42 to 3.4</td>
<td>0.91 to 2.89</td>
</tr>
</tbody>
</table>

*Note: $t_{critical} = 2.45$*

- $D_{mean}$: Bowel < Bladder < Rectum.
- t-stat results imply evidence for systematic error in $D_{mean}$ due to IOV for some cases for all OARs.
- Bladder: CoV << Bowel & Rectum (highest bladder Cov = patient who had TURP.)
  - Small $\sigma$ of $D_{mean}$ ‘amplifies’ t-stat.
- For all OARs $D_{mean}$: VMAT < IMRT.
### Results

#### Mandatory or Optimal Dose Constraint Failures?

<table>
<thead>
<tr>
<th>OAR</th>
<th>Mand. D.C. Fail.?</th>
<th>Comments:</th>
<th>Opt. D.C. Fail.?</th>
<th>Comments:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder</td>
<td>No</td>
<td>All well below</td>
<td>Yes</td>
<td>All failures: V5676 cGy ≤ 5%; V6000 cGy ≤ 3%</td>
</tr>
<tr>
<td>Bowel</td>
<td>Yes (just!)</td>
<td>All failures: V5270 cGy ≤ 0 cm³ (VMAT: Vol. ≤ 0.114 cm³) (IMRT: Vol ≤ 0.17 cm³)</td>
<td>Yes (few)</td>
<td>All failures: V4865 cGy ≤ 0.5 cm³ Vol. Excess &lt; 0.6 cm³</td>
</tr>
<tr>
<td>Rectum</td>
<td>Yes (One case pair)</td>
<td>All failures: V5676 cGy ≤ 15% (greatest failure: +1.74%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

Rectum exhibits widest variation in proportion of volume exposed depending on outliner:

- Discrepancies higher at lower D.C.s
- Greatest magnitude differences in variation: V2432 cGy
  - VMAT variation ~ 20% (GS vol. ~ 53%)
  - IMRT variation ~ 29% (GS vol. ~ 76%)

Figure: Rectum Cumulative DVH, Case 10 (VMAT)
# Results

c.f. Bladder & Bowel...

<table>
<thead>
<tr>
<th>OAR</th>
<th>D.C. (widest variation observ.)</th>
<th>GS Result (VMAT/ IMRT)</th>
<th>Range in variation (VMAT)</th>
<th>Range in variation (IMRT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder</td>
<td>V5676 cGy ≤ 35%</td>
<td>8.42 / 8.86 (%)</td>
<td>2.896%</td>
<td>2.892%</td>
</tr>
<tr>
<td>Bowel</td>
<td>V3649 cGy ≤ 158 cc</td>
<td>0/ 0 (cc)</td>
<td>3.154 cc (~2% GS vol.)</td>
<td>3.533 cc (~2.3% GS vol.)</td>
</tr>
<tr>
<td>Rectum</td>
<td>V2432 cGy ≤ 80%</td>
<td>53/ 76 (%)</td>
<td>20%</td>
<td>29%</td>
</tr>
</tbody>
</table>
Choosing Cases for NTCP Modelling:

- Subset of case pairs selected for each OAR to perform NTCP modelling investigation.
- Generally, the case pairs exhibiting the widest range in organ volume proportions at DCs chosen.

**Bladder:**
- Patient exhibiting a TURP prior to EBRT.

**Bowel:**
- Only case pair exhibiting bowel dose exposures >3649cGy for all outlines. (Case 1 & 7).
- ‘Worst’ case (highest bowel doses observed across all patients for one delineation) (Case 4 & 10).

**Rectum:**
- Case pair exhibiting mandatory DC failures (2 & 6).
- Widest differences between DVHs across outlines at DCs (4 & 10).
Results

NTCP Modelling: Bladder

- 4 endpoints: Urinary urgency; acute urinary incontinence; acute increased frequency; nocturia.
- $NTCP_{VMAT} < NTCP_{IMRT}$
  - E.g. 4.90% vs. 5.38%, urinary incontinence (GS)
- All $t$-statistics $< t_{critical}$; $|CoVs| \ll$ those of Bowel & Rectum
  - E.g. 0.022 urinary urgency vs 0.14 colon perforation/obstruction (IMRT)

⇒ • **Consequences of IOV are small for bladder, even when more complex anatomy is present.**
  • **Adjustment to current protocols/ more time spent outlining does not appear to offer significant benefit to patient side effect reduction.**

NTCP Modelling: Bowel

- 2 endpoints: Small bowel obstruction/perforation; Colon obstruction/perforation.
- All $t$-statistics $< t_{critical}$.
- $NTCP_{VMAT} < NTCP_{IMRT}$ E.g. 1.14 vs. 1.46%, Colon obstruc./ perf. (GS).
- $|CoVs|$ for Bowel similar to Rectum.
**Case 1 (IMRT)**

<table>
<thead>
<tr>
<th></th>
<th>Delin. E</th>
<th>GS</th>
<th>Delin. F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Bowel</td>
<td>0.09</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>NTCP (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon</td>
<td>1.81</td>
<td>1.46</td>
<td>1.22</td>
</tr>
<tr>
<td>NTCP (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Highest: \( NTCP_{\text{Delin} \ E} \)**
- Rectum outline over-extended superior to recto-sigmoid flexure.
- Bowel Vol. E < Bowel Vol. GS
- Therefore, proportion of bowel exposed to high/intermediate doses is exaggerated \( \Rightarrow NTCP_{\text{Delin} \ E} \uparrow \)

**Lowest: \( NTCP_{\text{Delin} \ F} \)**
- Small sub-portions of bowel close to PTV & near beam periphery excluded from Delin. F.
- Hence, proportion of bowel subject to high/intermediate doses is under-represented \( \Rightarrow NTCP_{\text{Delin} \ F} \downarrow \)
Greatest deviation in Bowel NTCPs observed...

- Rect. C terminated too inferior.
- Bowel:
  - Vol. C >> Vol. GS
  - $Vol_{High\ Dose_C} > Vol_{High\ Dose_{GS}}$

<table>
<thead>
<tr>
<th>Case 4</th>
<th>VMAT</th>
<th>IMRT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GS</td>
<td>Delin. C</td>
</tr>
<tr>
<td>Small Bowel NTCP (%)</td>
<td>0</td>
<td>0.056</td>
</tr>
<tr>
<td>Colon NTCP (%)</td>
<td>0</td>
<td>1.375</td>
</tr>
</tbody>
</table>

a) Differing opinions regarding the position of the superior limit of the rectum around the recto sigmoid junction &
b) Exclusion of sub-portions of bowel in intermediate/ high dose regions was responsible for NTCP result variation with outlining.
Greatest deviation in Bowel NTCPs observed…

This delineation choice also impacts rectum NTCP predictions. Bowel & Rectum are therefore, inter-related w.r.t. NTCP.

<table>
<thead>
<tr>
<th>Case 4</th>
<th>VMAT</th>
<th>IMRT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GS</td>
<td>GS</td>
</tr>
<tr>
<td>Small Bowel NTCP (%)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Colon NTCP (%)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

a) Differing opinions regarding the position of the superior limit of the rectum around the recto sigmoid junction &
b) Exclusion of sub-portions of bowel in intermediate/ high dose regions was responsible for NTCP result variation with outlining.
**NTCP Modelling: Rectum** (cases exhibiting mandatory DC failures)

Endpoint: Rectal Bleeding ≥ G2:
- Low probability: IMRT: <0.65%, VMAT: <0.6% (all outlines & param. sets)
- $t$-statistic > $t_{critical}$ for all param. sets $\Rightarrow$ mean(NTCP A to G) differs significantly from GS NTCP.
  - C.f. all bladder & bowel endpoints (IMRT & VMAT) $t < t_{critical}$.
- Large CoV c.f. Bladder (e.g. 0.15 vs. 0.032 Urinary Incontinence, IMRT).
  - Comparable CoV magnitudes with bowel (e.g. colon perforation/obstruction: 0.14, IMRT)
Case 6 Rectum Results Comparison (QUANTEC LKB params.):

<table>
<thead>
<tr>
<th>Delin:</th>
<th>GS</th>
<th>B</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTCP VMAT (%)</td>
<td>0.47</td>
<td>0.36</td>
<td>0.28</td>
</tr>
<tr>
<td>NTCP IMRT (%)</td>
<td>0.49</td>
<td>0.35</td>
<td>0.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Delin:</th>
<th>A</th>
<th>C</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTCP VMAT (%)</td>
<td>0.45</td>
<td>0.47</td>
<td>0.46</td>
<td>0.41</td>
</tr>
<tr>
<td>NTCP IMRT (%)</td>
<td>0.46</td>
<td>0.50</td>
<td>0.48</td>
<td>0.43</td>
</tr>
</tbody>
</table>

**Good concordance around PTV, discordance elsewhere:**

- **t-stat:** 2.54 (All) → 1.30 (GS, A, C, F, G) (IMRT), 2.56 (All) → 1.27 (VMAT)
- **Range:** 0.23% (All) → 0.07% (GS, A, C, F, G) (IMRT), 0.19% (All) → 0.07% (VMAT)
Results

NTCP Modelling Rectum Conclusion:

• Overall, prediction of occurrence is low (<1 %). However: NTCP variation exhibits high sensitivity to IOV in higher dose regions.

• Deviations in rectal outline distal to PTV do impact predictions; but gross outline deviation may be present here and good agreement with NTCP result of different outline is still observed if the outlines agree well in high dose regions.
Thanks for listening!
Lyman-Kutcher Burman NTCP Model

- NTCP assumed to exhibit sigmoidal response to Δdose:

**Probability of complication**

As partial volume irradiated ↓, Dose at which NTCP > 0 ↑ exponentially!

Therefore, outline volumes can have great effect on predictions!
Lyman-Kutcher Burman NTCP Model

- 3 parameters: $n (=1/a)$, $m$, $TD_{50\%}$
- $n$ describes volume effect
  - $0 < n \leq 1$ (typically. However, some researchers have curve fit to clinical data using $n>1$)
  - Small $n$ ⇒ ‘serial’ organ ⇒ NTCP correlates with max dose
  - Large $n$ ⇒ ‘parallel organ’ ⇒ NTCP correlates with mean dose.
- $m$ represents steepness of Dose/Response curve
  - Lower $m$ ⇒ $\Delta NTCP / \Delta Dose$ whilst $0 < NTCP < 1$
- $TD_{50\%} = $ Dose to organ leading to complication in 50% of population