KING'S College LONDON

Division of Imaging Sciences and Biomedical Engineering

Department of Medical Engineering and Physics

Assessing the impact of introducing the Acuros dose calculation algorithm into clinical use

Author: Robin Cole

Clinical Supervisors: Mohammad Hussein, Chris South & Catharine Clark

Academic Supervisor: Tony Greener

A dissertation submitted to the Faculty of Life Sciences & Medicine, King's College London in partial fulfilment of the requirements for the degree of Masters of Science in Clinical Sciences Medical Physics

22 March 2017

Abstract

The recently released Acuros XB (AXB) dose calculation algorithm is reported to have comparable accuracy to the current gold standard Monte-Carlo calculations but with significantly reduced computation times. In particular AXB is reported to be accurate in regions of high heterogeneity (such as the lungs) where the current clinical algorithm AAA is known to perform relatively poorly. For these reasons AXB is being considered for use as the standard algorithm for clinical use with the Eclipse treatment planning system (TPS) in the near future. However studies in the literature report that doses calculated with AXB and AAA can have small but statistically significant differences, yet a review of the literature does not yield a clear picture of the magnitude and direction of the differences. Furthermore AXB calculates dose to medium (Dm) and can optionally report Dm or dose-to-water (Dw), whilst AAA calculates and reports only Dw. Therefore this study investigates the dosimetric differences between AXB Dm/Dw and AAA with the aim of identifying any differences which could have clinically significant consequences.

A three-ways comparison of AXB and AAA calculated doses and measured percent-depth-dose (PDD) in a water phantom showed clear differences between AXB and AAA/measurements with AXB reporting lower doses. However the comparison of AXB with measurements is known to be challenging and requires further investigation. Next, a study of the dose differences between AXB Dw and Dm with medium (material) showed statistically significant differences, particularly in bone. A comparison with doses calculated using AAA and Monte-Carlo showed that AXB Dm was in closest agreement, and further studies in the report use AXB Dm. Finally investigations are performed into the differences to clinical plans for lung and head and neck (H&N) clinical sites. These investigations show that the magnitude and direction of dose differences between AAA and AXB varies with clinical site and organ, and therefore need to be assessed on a site-by-site basis. Lung plans showed fewer differences between AAA and AXB but with significant differences for OAR. For example the foramen received a maximum dose that was on average 0.44 Gy lower when calculated with AXB. This could have clinically significant consequences and this report recommends that lung OAR tolerances be reviewed before AXB is adopted for clinical use. Likewise the H&N plans showed statistically significant differences and tolerances should be reviewed.

Acknowledgements

I am grateful to my primary supervisors Dr Chris South and Dr Mohammad Hussein for their direction, advice and support on this project. I also wish to thank Dr Catherine Clark for many helpful conversations and her assistance in arranging my elective at the National Physical Laboratory (NPL) during which I performed some of the calculations presented in this report.

Nomenclature

- AAA Anisotropic-analytical-algorithm
- AXB Acuros XB
- BA Bland-Altman
- CCC Collapsed-cone-convolution algorithm
- 3D CRT 3D conformal radiotherapy
- CTV Clinical target volume
- D Dose
- Dm Dose-to-medium
- Dw Dose-to-water
- DVH Dose-volume histogram
- FFF Flattening filter free
- HI Homogeneity index
- H&N Head and neck
- IMRT Intensity-modulated radiotherapy
- Linac Radiotherapy linear accelerator
- LQ Linear-quadratic model
- MLC Multi-leaf collimator
- MV Mega-voltage
- MeV Mega-electron-volt
- MU Monitor units
- NPL National Physical Laboratory
- OAR Organs at risk
- PBC Pencil-beam-convolution dose calculation algorithm
- PARSPORT Parotid-sparing intensity modulated radiotherapy
- PTV Planning target volume, consisting of the CTV and a margin

- PDD Percent-depth-dose
- QA Quality-assurance
- QUANTEC Quantitative Analyses of Normal Tissue Effects in the Clinic
- RSCH Royal Surrey County Hospital
- ROI Region of interest
- SABR Stereotactic Ablative Body Radiotherapy
- SD Standard deviation
- SSD Source to surface distance
- SF Surviving fraction
- TPS Treatment planning system
- TR Therapeutic Ratio
- VMAT Volumetric arc therapy
- VOI Volume of interest

Contents

1 Introduction

1.1 The goals of radiotherapy

The goal of radical radiotherapy is to reduce the number of a tumour cells to a level that achieves permanent tumour control or reduces the symptoms of a disease. Healthy and cancerous cells have a subtly different response to radiation and it is this difference which allows the preferential removal of cancer cells whilst limiting damage to healthy tissue.

1.2 The LQ model

A cells response to radiation is characterised by a cell survival curve which is described by a linear-quadratic (LQ) model(Khan, Faiz M. 2014). This model states that cell death occurs due to a single or double interaction with radiation, with the response characterised by the coefficients alpha (α) and beta (β) respectively. For a given dose D the surviving fraction (SF) of cells is given by:

$$SF = e^{-(\alpha D + \beta D^2)}$$

The parameters α and β determine the slope of the cell survival curve, and the ratio α/β has the unit Gy. It is the different values of α/β for tumour and healthy tissue which give rise to the response curves shown in Figure 1.

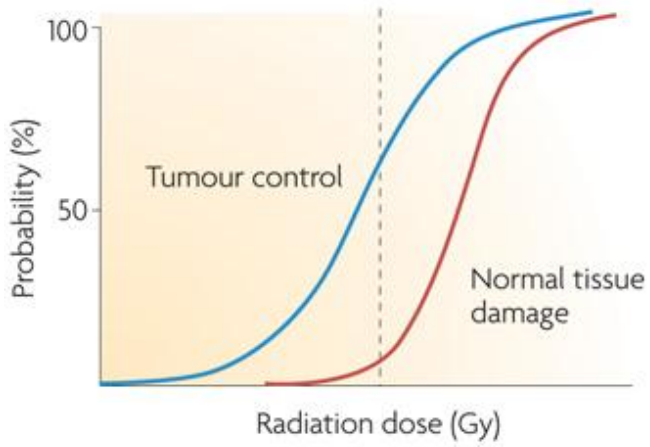

Figure 1. The probability of tumour cure, and tissues damage, increase with dose(Barnett et al. 2009)**.**

The effectiveness of a course of radiotherapy can be quantified though the Therapeutic Ratio (TR) which is the ratio of tumour response for a fixed level of normal-tissue complication. In Figure 2 the TR is better for case A than case B, where case A offers 95% tumour control for 5% incidence of complications corresponding to a TR of 19, whilst case B has a TR of 2.4.

Given the quadratic relationship between dose and response, small changes in dose at steep points on the dose-response gradient can have significant consequences such as reduced tumour control or increased normal tissue complications, and it is these small dose differences (with dose calculation algorithms) which are the focus of this report.

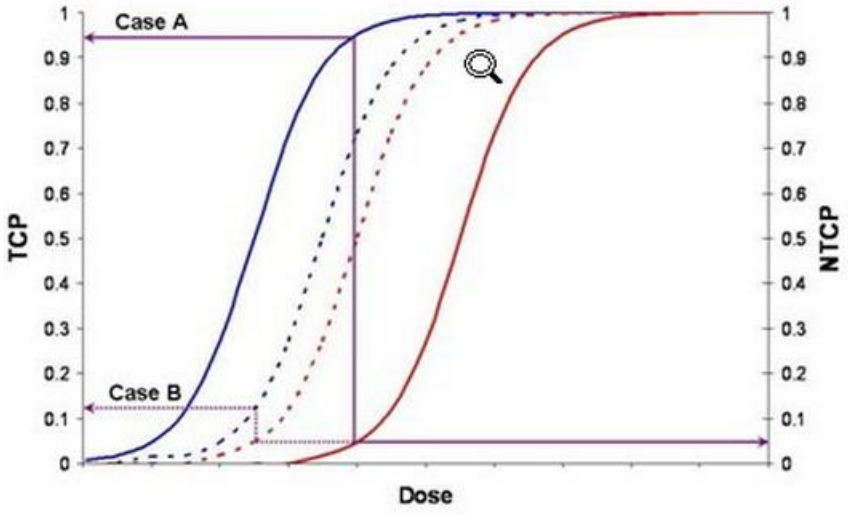

Figure 2. Dose-response curves for tumour control (blue) and normal tissue damage (red)(Hall, E.J. & Giaccia 2017)**. Dashed lines represent different conditions, e.g. different fractionation regime.**

1.3 Clinical trials

It is not possible to predict how individual patients will respond to radiotherapy, therefore clinical trials form the evidence basis for guidance on dose prescriptions and fractionation regimes. The two clinical sites investigated in this report are the lung and the head and neck (H&N) which at RSCH are planned according to the SABR UK Consortium(Consortium 2014) lung guidance and the PARSPORT(Institute of Cancer Research; 2016) trial recommendations respectively. A common source of organ at risk (OAR) tolerances is the QUANTEC publications(Marks et al. 2010) which collate data from multiple clinical trials.

2 Treatment planning systems & algorithms

2.1 Introduction

Treatment planning systems (TPS) implement one or more algorithms to compute dose deposition. Accuracy is critical if the therapy is to deliver the optimal dose to the tumour whilst minimising the dose to nearby organs at risk, and The American Association of Physicists in Medicine (AAPM) advises that the uncertainty in calculated dose distributions should be less than 2%(Rock et al. 2004). Increased computing power over time has enabled more accurate modelling of dose deposition, with the evolution illustrated in Figure 3. The most accurate algorithms implement the Monte-Carlo method but require significant calculation times, making them often impractical for routine clinical use. Other algorithms make a range of approximations and simplifications to improve calculation speed at the cost of reduced accuracy.

In general terms algorithms can be classified into three groups; correction based, model based and principle based. Model and principle base algorithms have further been classified as type 'a', 'b' or 'c'. Type 'a' do not account for changes in electron transport, whilst type 'b' do account for electron transport, and in type 'c' the full physics of the dose absorption process is modelled(Fogliata & Cozzi 2016).

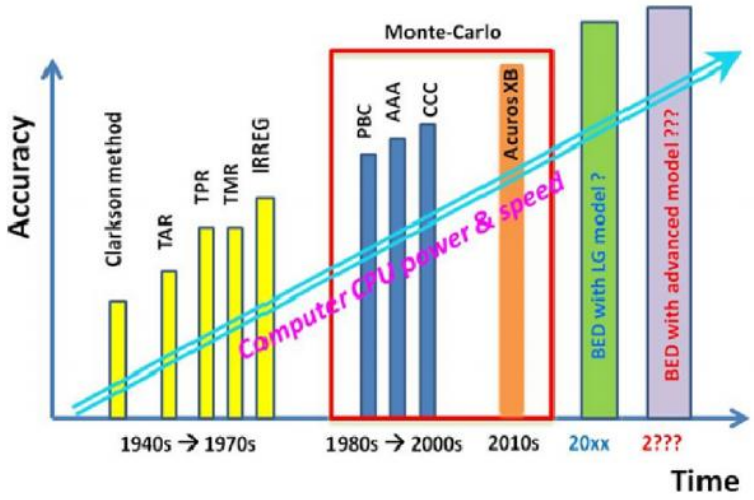

Figure 3. Evolution of dose calculation algorithms, accuracy vs. time(Lu 2013)

2.2 Correction-based algorithms

These are semi-empirical and interpolate or extrapolate dose from measured beam data in water, for example percent-depth-dose (PDD) data for different field sizes and source-to-surface - distance (SSD). Example algorithms are Clarkson's technique and IRREG which are used for

manual calculations and implemented in dose-check software, for example in the RadCalc software (RadCalc 2017). Correction-based algorithms are sufficiently accurate to be clinically useful for calculations in homogenous media (i.e. uniform water) but are not suitable for calculations in regions of heterogeneity since lateral scattering is not considered.

2.3 Model-based algorithms

These use a physical model to simulate radiation transport and compute a two-step dose deposition process. Primary photons interact with the medium and impart kinetic energy to charged particles, primarily through Compton scattering for mega-voltage (MV) beams. These charged particles deposit energy through ionisation along a finite track. A class of model based algorithm that has been implemented widely is the convolution/superposition algorithm. This algorithm does not model individual particles but models the dose deposition process using two components:

1. Terma (total energy released per unit mass) – represents the energy imparted by the interactions of *primary photons*. The Terma is obtained by multiplying the primary photon fluence with the mass attenuation coefficient.
2. Dose kernel – represents the energy deposited by *secondary particles* about a primary photon interaction site. The dose kernel is the analytic representation of the dose distribution of a narrow beam in water, obtained by measurement or Monte-Carlo simulations.

Dose is calculated from the convolution of the Terma with the dose kernel, as illustrated in Figure 4. The broad clinical beam is divided up into beamlets and the convolutions performed for all beamlets. The dose distribution is given by the *superposition* of the individual beamlet contributions.

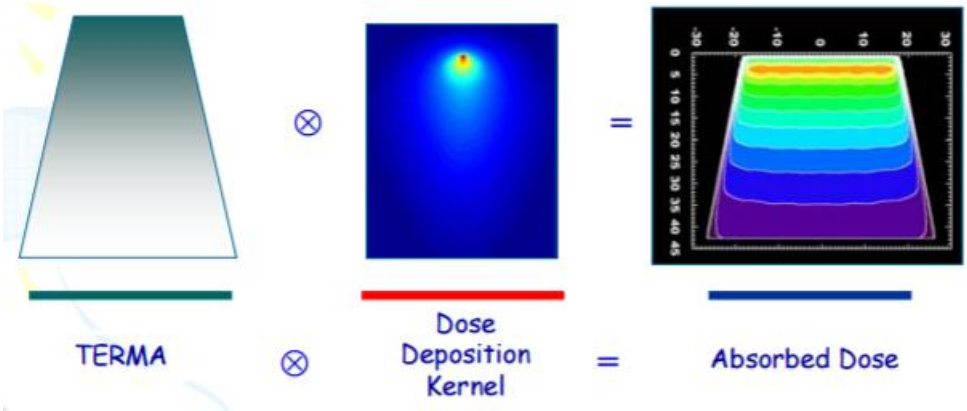

Figure 4. Convolution using FFT techniques give the absorbed dose(Knoos et al. 2008)**.**

2.3.1 Pencil-beam-convolution

The Pencil-beam-convolution algorithm (PBC) is a type 'a' algorithm in which heterogeneity is accounted for by longitudinally (1D) scaling the Kernels by effective distances calculated from the electron densities (relative to water) around an interaction point. This algorithm assumes that adjacent planes are water and therefore does not accurately model lateral scatter. Lateral scatter is usually a second order affect but can be significant for small fields in regions of low density or high heterogeneity (e.g. lung), leading to an overestimation of dose in the low-density volumes. Deviations up to 5% for 4MV beams were found relative to Monte-Carlo simulations, which increased with energy(Knoos et al. 1995).

2.3.2 Collapsed-cone-convolution

The Collapsed-cone-convolution algorithm (CCC) is a type 'b' algorithm which improves on PBC to more accurately handle inhomogeneity. In order to reduce computation time the dose spread kernel function is discretised into ~ 100 discrete cones from the interaction point, shown in Figure 5. The dose within a cone is collapsed onto a ray along the central axis of the cone. An *adaptive* CCC algorithm is implemented in Pinnacle TPS, where the resolution of the dose grid is adaptively varied to reduce computation time.

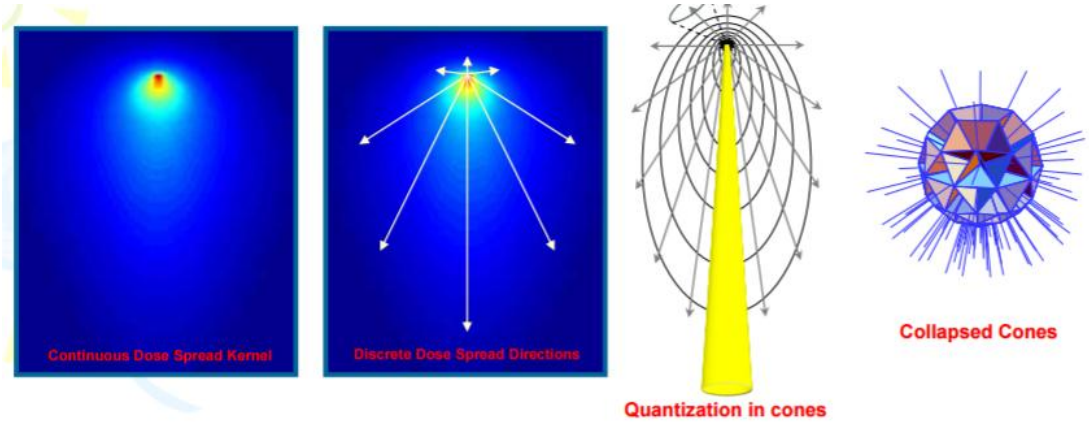

Figure 5. The continuous Kernel is discretised into cones, and the dose within a cone collapsed onto a ray along the central axis of the cone(Knoos et al. 2008)**.**

2.3.3 Anisotropic-analytical algorithm

Like CCC, the Anisotropic-analytical algorithm (AAA) is a type 'b' algorithm which aims to more accurately calculate dose distributions in regions of complex heterogeneity, but uses a different approach. It separately models dose contributions from three sources; primary photons, scattered extra-focal photons and contamination electrons originating from the flattening filter, jaws etc. Each has an associated fluence, energy deposition density function and scatter

kernel. Tissue heterogeneity is accounted for in the *full* 3D neighbourhood by anisotropically rescaling dose kernels (calculated in water) depending on the material density (relative to water) in lateral directions, shown in Figure 6. Compared to PB, AAA more accurately models the dose in low density regions (i.e. lung) shown in Figure 7. The Varian Eclipse TPS implements the AAA algorithm and this is the algorithm currently used for clinical planning at RSCH.

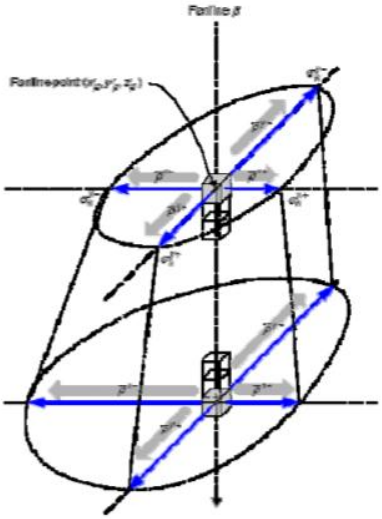

Figure 6. AAA scales the dose kernel anisotropically based on the material density relative to water (Knoos et al. 2008)**.**

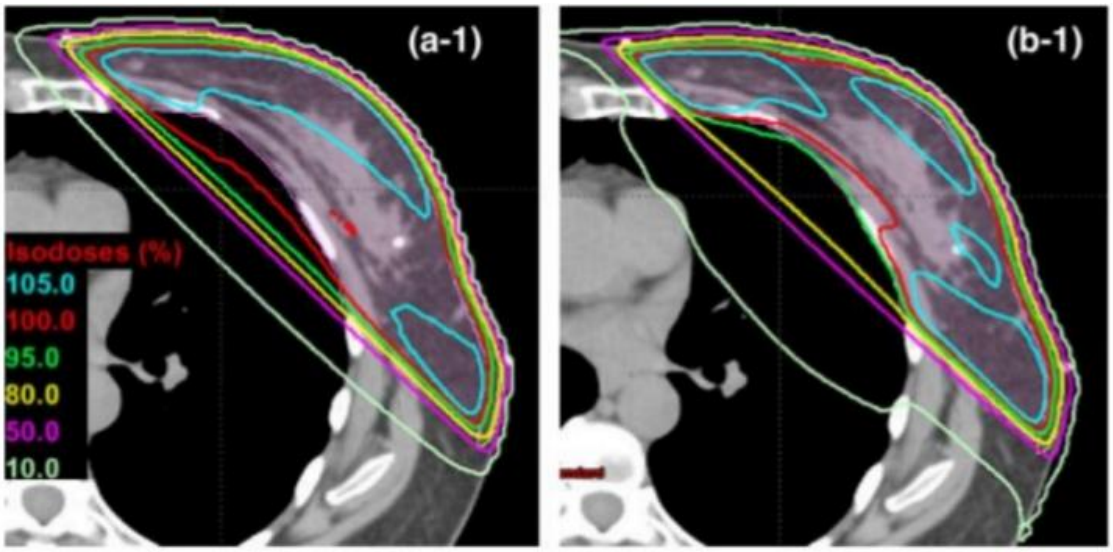

Figure 7. Comparison of (left) Pencil beam and (right) AAA(Gohar 2017)**. PB over-estimates the dose in the low density regions.**

2.4 Principle-based algorithms

In this category are AXB and Monte-Carlo, and these algorithms accurately model the physical processes responsible for dose deposition. Monte-Carlo simulates the track of individual particles whilst AXB solves coupled equations which accurately describe the dose deposition process. Whilst Monte-Carlo is considered the gold-standard (by comparison with measurements), given sufficient computation time both AXB and Monte-Carlo approach close to 100% agreement(Lu 2013). The largest differences between these algorithms and type 'b' algorithms are in low-density media, particularly in non-equilibrium conditions such as small fields or at boundaries.

2.4.1 Acuros-XB

Whereas AAA handles heterogeneities as density-based corrections to dose kernels calculated in water, Acuros-XB (AXB) solves deterministically a group of Boltzmann transport equations that describe all of the physical processes involved in dose deposition using actual material properties (Rana & Pokharel 2014). Given sufficient computation time AXB approaches the dose accuracy of the 'gold standard' Monte-Carlo algorithms (illustrated in Figure 8) but with significantly reduced computation times. Like Monte-Carlo, AXB requires a material map of the calculation volume, and in Eclipse the material database consists of 5 biological materials and 13 non-biological materials. The HU value of a voxel is converted to a CT number using the CT calibration curve, and from the CT number the material is assigned, illustrated in Figure 9. For material densities below 3.0 g/cc the material assignment is automatic and by default only biological materials are assigned. Voxels with a density higher that 3.0 g/cc must be assigned by the user, and any material assigned can be over-ridden by the user. Once the fluence is calculated using the material map, AXB calculates the dose to a voxel *i* using an energy dependent fluence-to-dose response function:

$$D_i = \int_0^{\infty} dE \int_{4\pi} d\hat{\Omega} \frac{\sigma_{ED}^{e}(\vec{r}, E)}{\rho(\vec{r})} \Psi^{e}(\vec{r}, E, \hat{\Omega})$$

- σ_{ED}^{e} is the macroscopic electron energy deposition cross section (MeV/cm)
- ρ is the material density (g/cm^3)
- Ψ^e is the angular electron fluence
- E is the energy
- $\vec{r}$ and $\hat{\Omega}$ are position and direction vectors

AXB can report dose to medium or dose to water. When dose to medium is reported, $\sigma_{ED}^{e}$and ρ are based on the material properties of the voxel. When dose to water is reported, σ_{ED}^{e} and ρ are based on water. A comparison of the calculated doses using AXB Dm/Dw is shown in Figure 10. Varians documentation(Gregory A. Failla et al. 2010) states that reporting dose to water Dw is equivalent to reporting the dose to a small volume of water embedded within the non-water material, such that the water volume is small enough not to perturb the electron fluence. Since this volume is typically much smaller than the volume of detectors used to experimentally measure dose, comparison of AXB Dw with measurements is challenging and it is recommended to '*explicitly model the volume representing the detector'*.

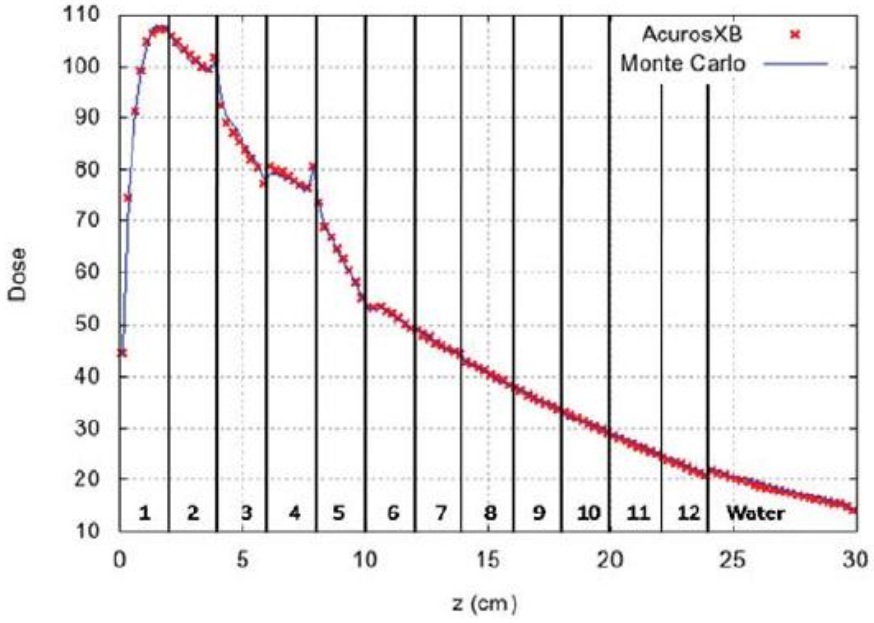

Figure 8. Percent depth dose (PDD) for a 6MV 10x10 field calculated with AXB and Monte-Carlo for a 13 slab phantom(Gregory A Failla et al. 2010) **consisting of non-biological materials. Slab materials are as follows: (1) Polystyrene – 1.05 g/cc, (2) Epoxy – 1.04 g/cc, (3) Aluminium – 2.7 g/cc, (4) PMMA – 1.19 g/cc, (5) Titanium alloy – 4.42 g/cc, (6) Radel – 1.30 g/cc, (7) Wood – 0.70 g/cc, (8) PEEK – 1.31 g/cc, (9) PVC – 1.38 g/cc, (10) Acetal – 1.42 g/cc, (11) PVDF – 1.77 g/cc, (12) PTFE – 2.20 g/cc**

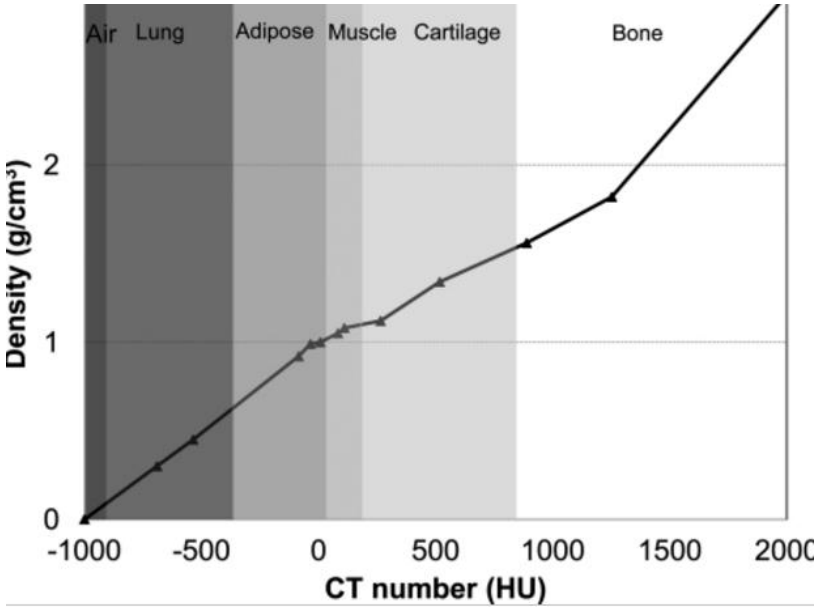

Figure 9. Graph of the CT number material assignment (five biological materials only) implemented in Eclipse. The values and ranges can be adjusted by the user(Han et al. 2011)**.**

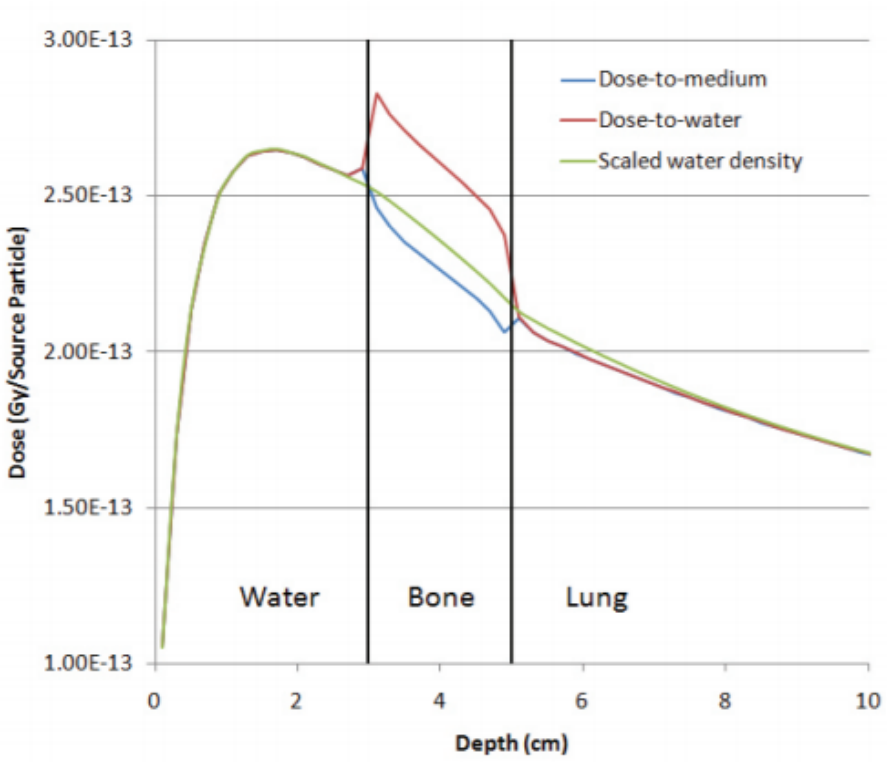

Figure 10. PDD for 5x5 6 MV field on a water-bone-lung slab phantom, from Varians documentation(Gregory A. Failla et al. 2010)**. AXB calculates dose to medium (blue line) and performs a calculation to convert this to dose to water (red line). In contrast, AAA calculates dose to water with density scaled by that the medium the dose is actually deposited in (green curve))**

2.4.2 Monte-Carlo

Monte-Carlo algorithms simulate the transport of particles through matter using a complete physical model where probability distributions are used to determine the outcome of individual interactions. The dose distribution is calculated by summing the ionising events within individual voxels. The material composition of voxels is accounted for and the dose to the medium is reported (Dm). The more histories N modelled the more accurate the simulation, and dose uncertainty scales with $1/\sqrt{N}$. The transport of billions of particles must be simulated to achieve the accuracy required for treatment planning. However increasing N also increases the computation time. Different Monte-Carlo codes will make different assumptions to reduce computation time, resulting in a trade-off between speed and accuracy. The Monaco(Elekta 2017) TPS implements a Monte-Carlo algorithm.

3 Literature survey

A literature survey using Google Scholar has been undertaken to determine which clinical sites could be most affected by the choice of dose calculation algorithm AXB vs. AAA. The studies all compare the dose metrics calculated by the algorithms, and dose metrics are discussed in the appendix. A summary of several significant publications is presented here.

1. Huang *et al* studied 10 patient plans for lung stereotactic treatments(Huang et al. 2015). Wilcoxon signed-rank tests were performed to evaluate the significance of differences in dose metrics between AAA/AXB. Energy dependent differences were observed and AXB calculated lower doses.

2. Fogliata *et al* studied 10 patient plans for non-small-cell lung cancer treatments (Fogliata et al. 2012). Paired Student's t test was used to evaluate the significance of observed differences to dose metrics. Differences were of order 0.5% but varied with tissue and energy. Mean doses to OAR presented differences of up to 3% in the worst case. No systematic differences were found between 3DCRT and IMRT.

3. Rana *et al* studied 14 patient plans for lung stereotactic treatments(Rana et al. 2014). AXB predicted lower values for various dose metrics than AAA and showed better agreement to measurements in a phantom study.

4. Zifodiya *et al* investigated the implications of selecting AXB Dm/Dw dose reporting in 10 patient plans in four clinical sites(Zifodya et al. 2016). The overall dose differences were found to be small (<1.5%), but in non-water media differences of up to 4.6% were observed (prostate plan). Overall calculated Dw values were higher than calculated Dm values. However no tests of the statistical significance of the differences were performed.

In summary, AXB has been shown to better agree with measurement in phantoms, but no consistent picture of the differences between AAA/AXB can be identified based on this survey. Dose differences are also reported between AXB Dm and Dw dose reporting.

4 Initial validation of AXB

4.1 Introduction

AXB had been configured for use at RSCH and basic commissioning performed prior to this study. In order to validate the configuration, several calculations were performed during this study to validate the performance of AXB and test the methodology.

4.2 Comparison of PDD in a simple phantom

The percent depth dose (PDD) to a water phantom was calculated under reference conditions of a 10x10 field at 100 cm SSD. Measured golden beam data previously acquired with a MicroDiamond detector (not by the author) was available for these conditions, and is also included for comparison.

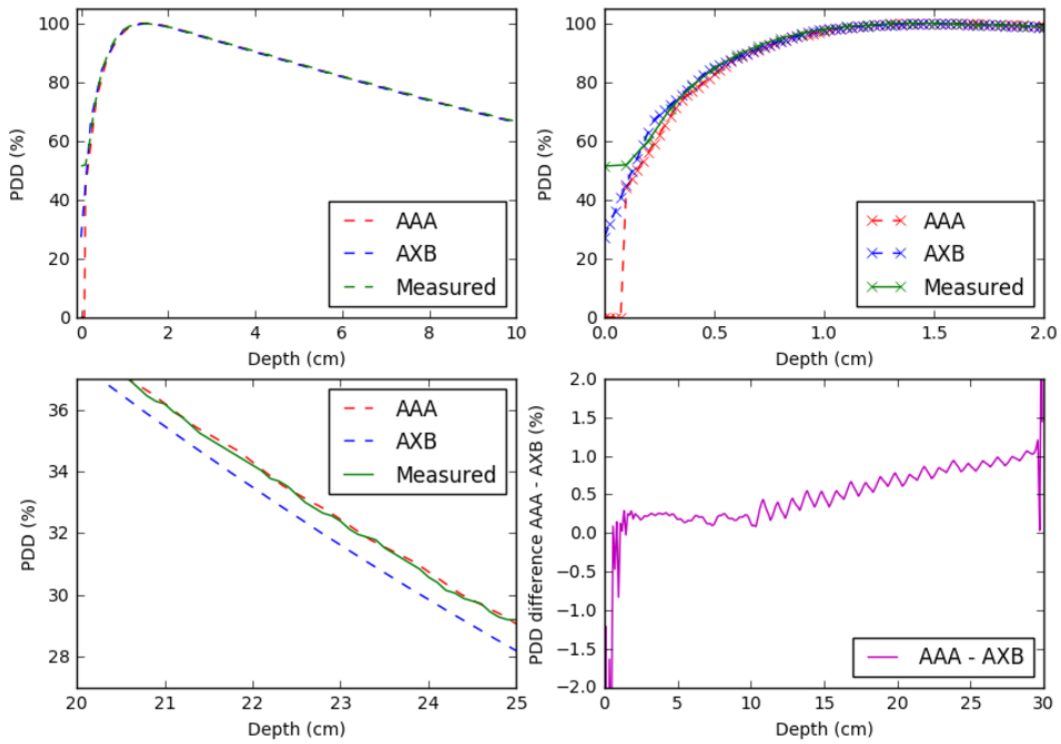

Figure 11. (Top & bottom left) Comparison of PDD, at 30 cm depth there is approximately a 1% difference in dose. (Bottom right) the difference AAA - AXB. Note that the measured data plateaus above 0.1 cm depth as a water meniscus forms over the detector.

The calculated and measured data is shown in Figure 11. AXB (Dm reporting) and AAA show good agreement with measured data below 10cm, except at the air-water interface where there is high heterogeneity and AXB appears to more accurately model the ideal PDD. Interestingly at depths beyond 10 cm AXB diverges from the AAA/measured data. As discussed in section

2.4.1, comparison of AXB with measurements is complicated, and therefore the simplistic comparison presented here is probably not adequate to explain the differences observed. However further investigation is beyond the scope of this report.

4.3 Comparison of doses due in complex plans

Next an investigation was performed to validate AXB by comparison with measured doses to an IPSM phantom (not measured by the author), shown in Figure 12. Dose data was available for plans covering different beam arrangements at 6 MV (9 plans) and 10 MV flattening filter free (FFF, 11 plans). The dose distribution from these plans were calculated with AAA/AXB Dm, and a three-ways Friedman test indicates that the differences between the AAA/AXB/measured doses are statistically significant, with $\chi^2 = 18.2$ and $p < 0.05$ for 6MV data and $\chi^2 = 20.2$ and $p < 0.05$ for 10 MV FFF.

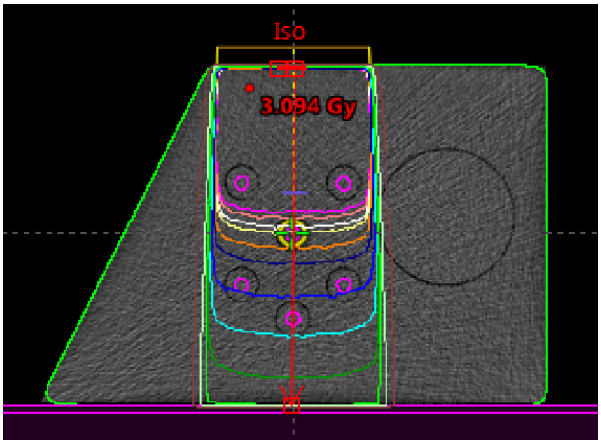

Figure 12. AXB calculated dose distribution to a phantom. The yellow circle marks the measured/calculated dose point.

The difference Δ (%) between a calculated and measured dose is given by:

$$\Delta\,(\%) = \frac{100 \times (D_{Calculated} - D_{Meas})}{D_{Meas}}$$

The mean (%) and standard deviation (SD) of the dose differences Δ are extracted in Table 1. AXB has a larger mean difference to the measurement data. However AXB also has a lower standard deviaiton than AAA, indicating that the calculation may be more consistent but with a constant offset.

Comparison	Mean	SD
6 MV AAA vs. measured Δ (%)	-0.98	0.48
6 MV AXB vs. measured Δ (%)	-1.35	0.31
6 MV AXB vs. AAA Δ (%)	-0.42	0.36
10 MV AAA vs. measured Δ (%)	-1.64	0.53
10 MV AXB vs. measured Δ (%)	-2.17	0.35
10 MV AXB vs. AAA Δ (%)	-0.53	0.62

Table 1. Mean and standard deviation of dose differences Δ between AAA/AXB/measurement.

4.4 Test of normality

Prior to selecting a statistical test to compare the significance of the observed dose differences Δ a test of the normality of data should be performed. However since there are too few samples to perform a normality test, subsequent analysis makes no assumptions about the normality of the data and uses the non-parametric Wilcoxon signed-rank test.

4.5 Significance of dose differences

The significance of dose differences Δ(%) was assessed using the non-parametric Wilcoxon signed-rank test, with the test results in Table 2. The test results $p < 0.05$ in both cases, indicating that the dose differences are statistically significant. Note that for sample sizes of less than 10 (true for the 6MV data) the p-value is only approximate.

Comparison	p-value
6MV AAA vs. AXB Δ(%)	0.009
10MV FFF AAA vs. AXB Δ(%)	0.016

Table 2. Wilcoxon test results.

4.6 Sample size

In order to determine the number of plans required to observe significant differences Δ between AAA/AXB dose metrics a bootstrap procedure was performed as described in reference(Chaikh et al. 2014). The bootstrap procedure consisted of taking 100 random samples of size n (where n ranges from e.g. 3 to 9 where at 6MV there were 9 plans) and performing the Wilcoxon test on each random sample. The mean p-value for each sample size n is then calculated. Figure 13 shows that 6 samples (plans) or more should be sufficient to observe differences of order ~ 0.5% between algorithms using the Wilcoxon test.

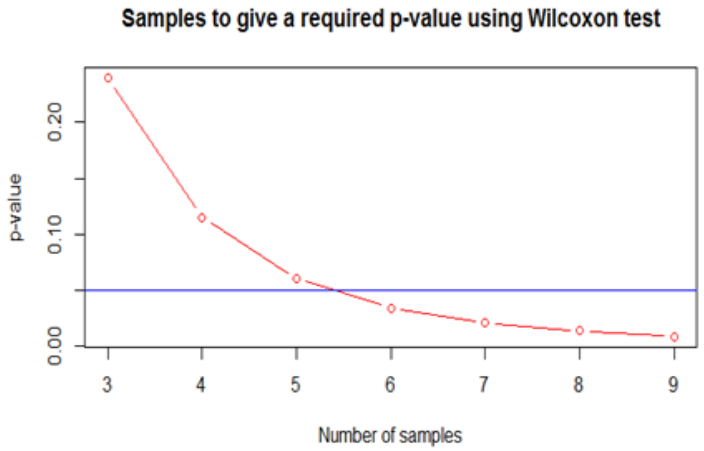

Figure 13. Mean p-value with sample size n as calculated using the Wilcoxon test.

4.7 Summary

The three-ways Friedman statistical test indicates that the differences in dose between AAA/AXB/measurement do not occur by chance, and are statistically significant. Interestingly AAA shows better agreement to measurement than AXB, although a more rigorous analysis could resolve these differences, but is beyond the scope of this report. Calculations of plans delivered to the IPSM phantom confirm the significance of the dose difference between AAA/AXB, and indicate that the magnitude of the difference is of order 0 – 1 %, consistent with reports in literature. This study concludes that AXB reports a mean dose that is lower than AAA by -0.42% at 6 MV and -0.53% at 10 MV FFF.

5 Comparing dose-to-medium and dose-to-water

5.1 Introduction

Currently treatments at RSCH are planned using the AAA algorithm which calculates dose to water (Dw). In contrast AXB calculates dose to medium (Dm) and optionally reports Dw, as discussed in Section 2.4.1. There are two views expressed in the literature over whether it is generally preferable to report Dw or Dm. Those in favour of Dw argue that since current clinical practice is based upon experience with Dw, and that tissues of interest are primarily composed of water, the best clinical outcomes can only be guaranteed by continuing to report Dw. Those in favour of Dm argue that Dm is potentially more clinically relevant since it more accurately reflects the true dose deposition within a patient, but acknowledge that further work is required to establish the clinical significance of switching to Dm reporting. In addition to these arguments it is important to take into account that AXB and Monte-Carlo calculate Dm, and reporting Dw with these algorithms introduces an additional source of uncertainty.

5.2 Methodology

In order to compare Dw and Dm a simple plan to a virtual phantom was developed by the author for this report. The phantom plan consists of four open isocentric 10x10cm fields at orthogonal gantry angles which deliver a fixed 25 monitor units (MU the calibrated unit of Linac output) to a 30x30x30 cm virtual phantom of water enclosing a 5x5x5 cm VOI, shown in Figure 14. Both the phantom and the VOI were centred to the field isocentre. For this study the dose to the isocentre (ISO) was first calculated with water as the VOI material. This formed a relative baseline (normalisation) dose for subsequent calculations where the material of the VOI was varied, changing from air (0 g/cm^3) to Titanium alloy (4.42 g/cm^3) using a manual override, and the dose to the ISO calculated for each material. Plots of the normalised ISO dose vs. VOI material are shown in Figure 15 and Figure 16. In addition to AAA/AXB, this study included the EGSnrc (National Research Council Canada 2017) Monte-Carlo code, which is considered the ‘gold standard’ for dose calculations by the NPL.

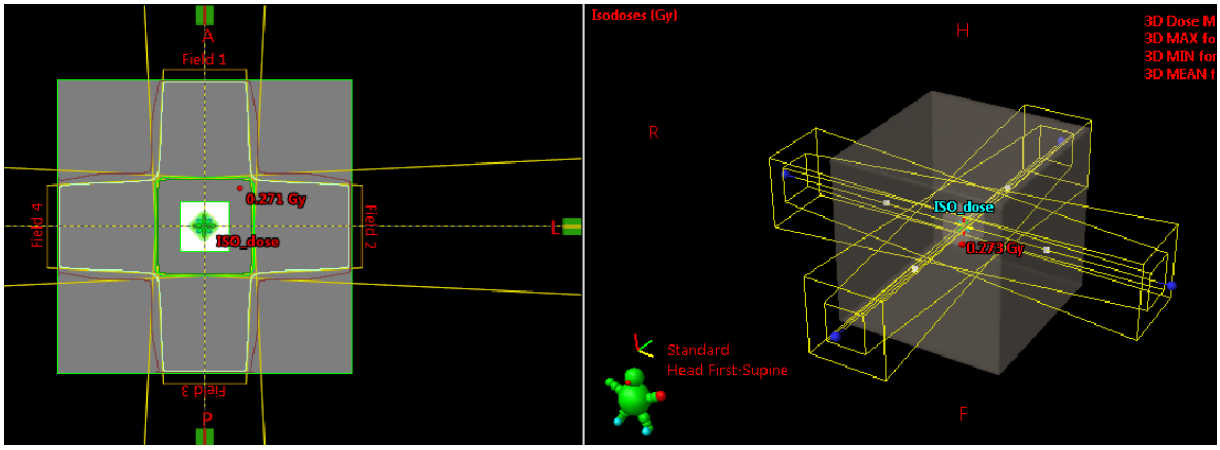

Figure 14. Virtual phantom within Eclipse TPS.

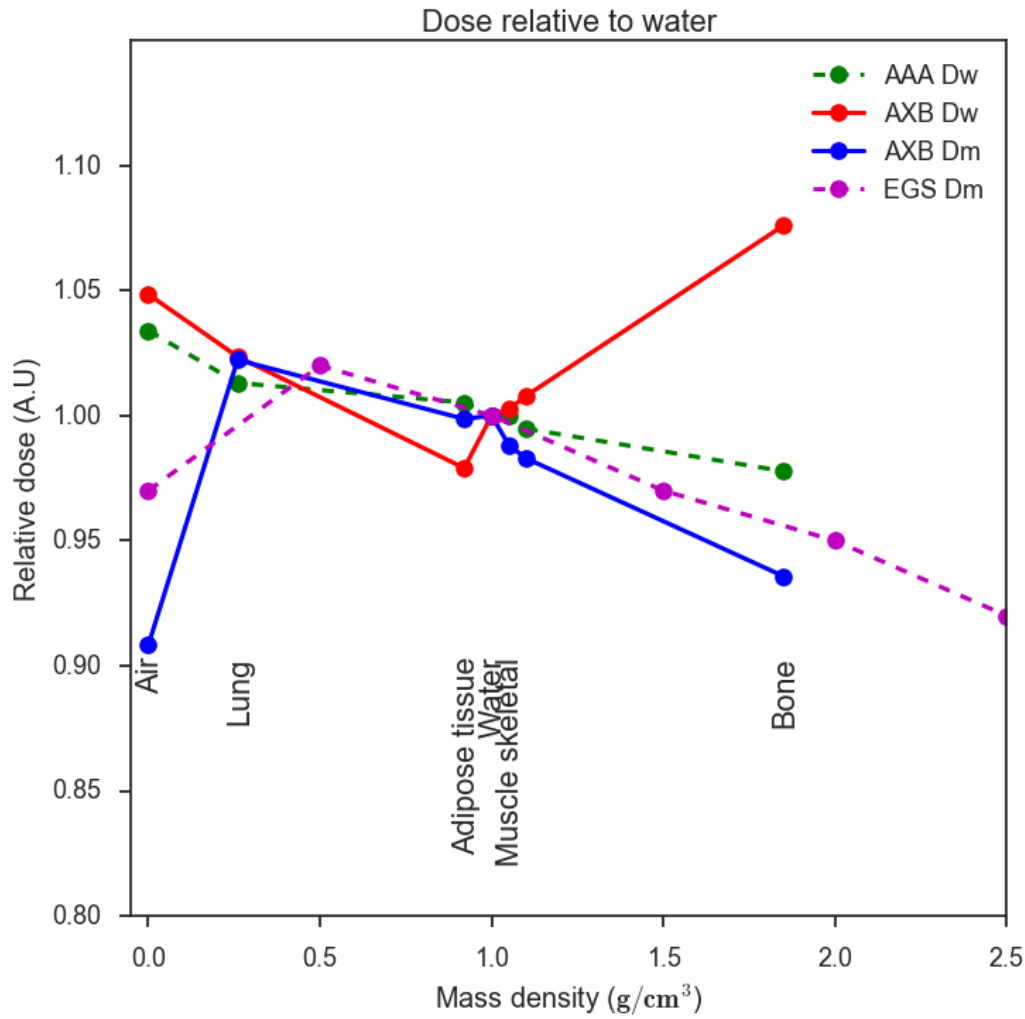

Figure 15. Plots of point dose to the ISO as the material of the VOI is varied - data for only the biological materials in the Eclipse library.

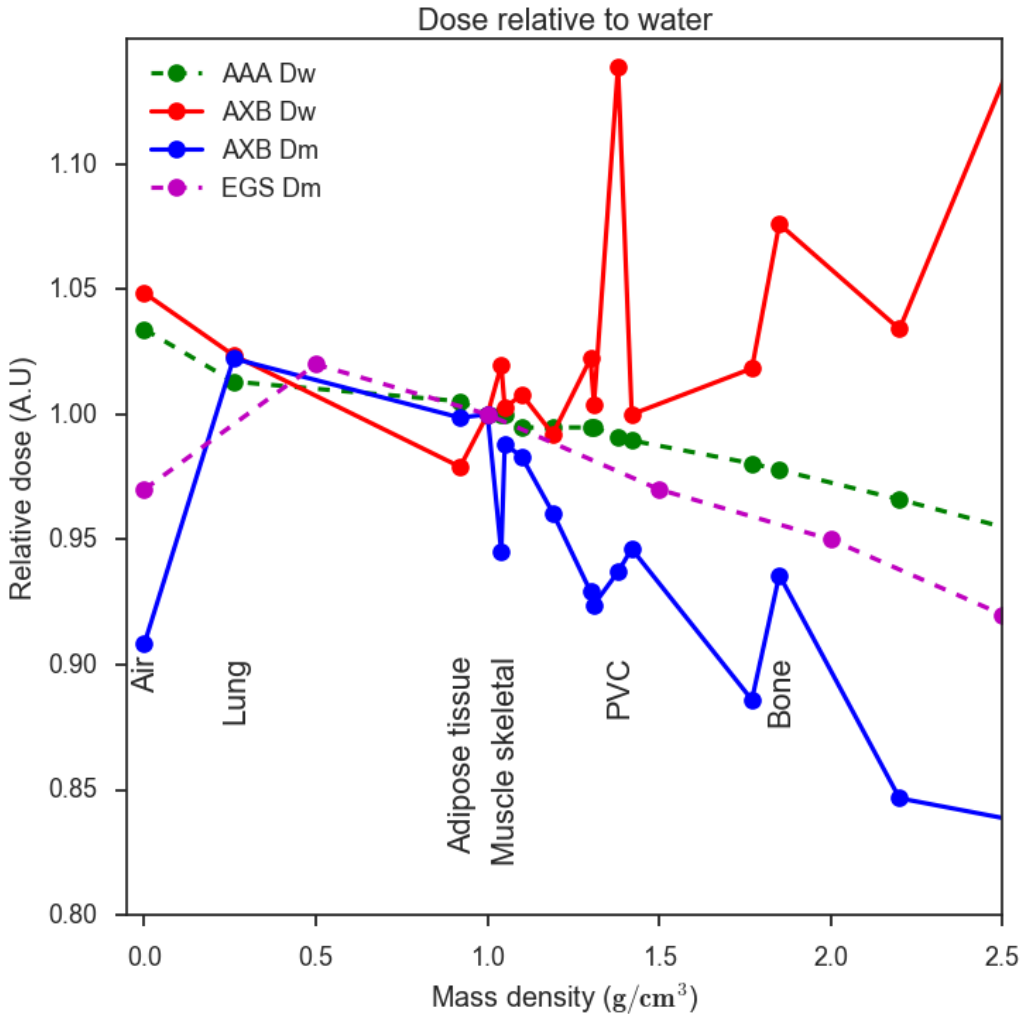

Figure 16. ISO dose data for all materials in the Eclipse material library.

5.3 Results and discussion

The dose calculation data, density, stopping power ratio ($S_{w,med}$) and associated CT number of materials are tabulated in Table 3. Figure 15 shows the ISO dose data when only the biological materials are included and shows that the variation of dose due biological materials is within 5% for most materials, with the exception of bone. AAA and EGS show the closest agreement, with AXB Dm and Dw showing significantly different behaviour for bone and air.

Figure 16 shows the ISO dose data where all materials in the Eclipse material library are included. Clearly large differences exist for the non-biological materials for AXB, and the differences increase with the density of the material. Overall AXB Dm shows a closer agreement with AAA/EGS and therefore the studies in sections 6 and 7 use AXB Dm. The difference between AXB Dm and Dw is particularly large for PVC, and there is no obvious

cause for this difference given the values for the material density and stopping power as tabulated in Table 3. Unfortunately a full analysis of this data is beyond the scope of this report, but this topic forms the basis of an ongoing investigation being performed jointly by the RSCH/NPL and other UK institutions.

5.4 Summary

The choice of dose calculation algorithm and dose reporting mode can result in dose differences for all biological materials, but most significantly for bone. Overall, AXB Dm shows the closest agreement to other algorithms, in particular to the 'gold standard' EGSnrc algorithm. The debate between using Dm or Dw clinically is likely to continue for the foreseeable future, but as far as AXB is concerned the use of Dm is preferable since reporting Dw introduces additional sources of uncertainty into the calculation of dose.

VOI Material	Mass density	Stp PWR	CT number	AAA	AXB Dm	AXB Dw
Air	0.00	0.01	-993	1.03	0.91	1.05
Lung	0.26	0.27	-734	1.01	1.02	1.02
Adipose tissue	0.92	0.94	-105	1.01	1.00	0.98
Water	1.00	1.01	0	1.00	1.00	1.00
Epoxy	1.04	1.04	38	1.00	0.94	1.02
Muscle skeletal	1.05	1.05	48	1.00	0.99	1.00
Cartilage	1.10	1.09	128	0.99	0.98	1.01
PMMA	1.19	1.16	282	0.99	0.96	0.99
Radel	1.30	1.25	470	0.99	0.93	1.02
PEEK	1.31	1.26	487	0.99	0.92	1.00
PVC	1.38	1.32	605	0.99	0.94	1.14
Acetal	1.42	1.35	671	0.99	0.95	1.00
PVDF	1.77	1.66	1283	0.98	0.89	1.02
Bone	1.85	1.73	1430	0.98	0.94	1.08
PTFE	2.20	2.03	2041	0.97	0.85	1.03
Aluminium	2.70	2.46	2913	0.95	0.83	1.20
Titanium alloy	4.42	3.81	5908	0.88	0.81	1.30

Table 3. Complete data from Figure 16.

6 Lung Study

6.1 Introduction

The lung protocol at RSCH is based on the SABR UK Consortium(Consortium 2014) lung guidance. The 55Gy/5# protocol requires that at least 95% of the PTV receive the prescribed dose, and plans are normalised to give D95% = 55 Gy. Since a lung tumour may move significantly with respiratory motion, 99% of the PTV should receive > 49.5 Gy (90% of the prescription dose). The D1CC should be between 118 to 134% of the prescription dose (i.e. 64.9 to 73.7Gy), and the DMAX < 135% (DMAX < 74.25 Gy). There are multiple OAR in the lung region as shown in Figure 17. Organs at risk (OAR) must be protected against hot-spots and there are a number of constraints on OAR. The full protocol requirements are given in the appendix.

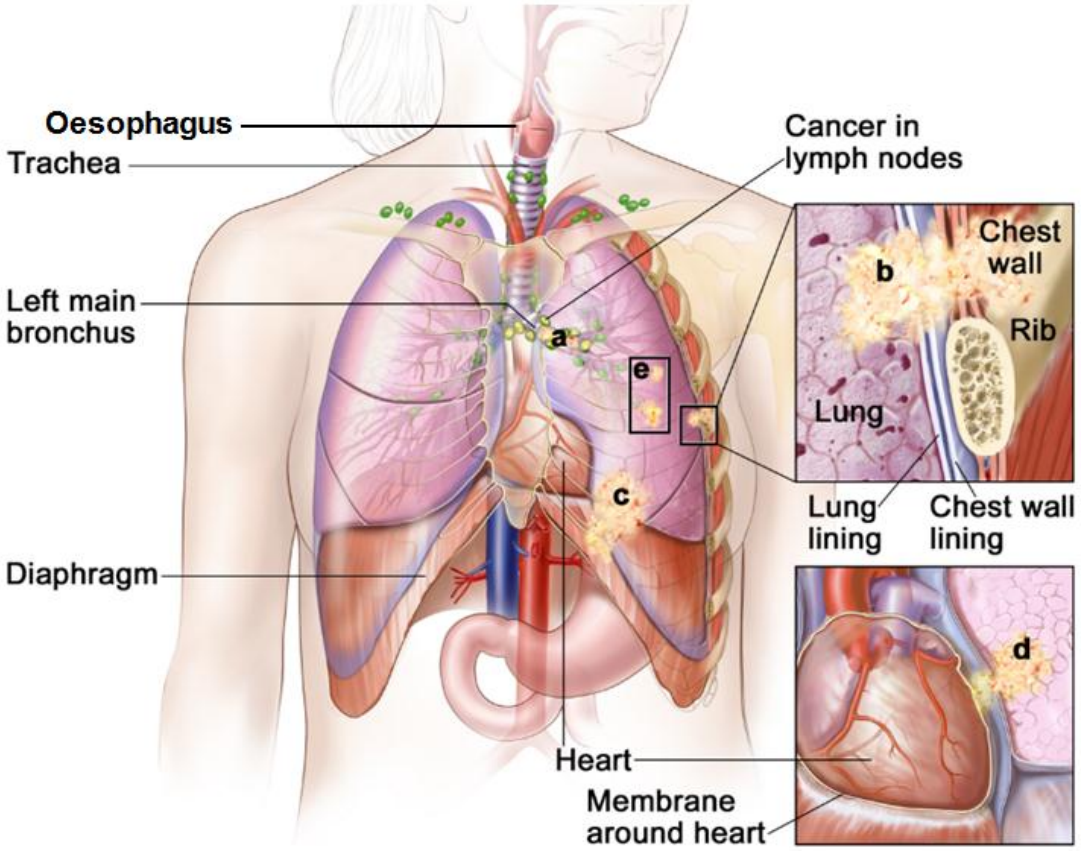

Figure 17. Major structures in the lung region, modified from reference(PDQ Adult Treatment Editorial Board 2002)**.**

6.2 Results

Six plans were included in this study. Each plan was produced previously by an experienced planner using AAA and then doses here recalculated using AXB (Dm) with fixed MU. The statistical significance of dose differences between AAA/AXB was established in Section 4 and is not repeated here. Publications typically focus on the PTV differences, but the PTV adds a 5mm margin to the Internal Target Volume (ITV) which is the volume that encompasses the gross tumour volume (GTV) motion over the full breathing cycle. The PTV may include non-tumour structures such as bone, and therefore in this study the dose differences to the ITV are considered the most clinically relevant and are investigated alongside the PTV.

A single plan is shown in Figure 18. The ITV and PTV are contoured on the dose distribution and the dose difference plots. The difference plot of Figure 18(top right) is colour coded, with purple corresponding to regions where AAA calculates the higher dose, and green to regions where AXB calculates the higher dose. The mean and standard deviation of the AAA/AXB DVH for all six plans in the study are shown in Figure 18(bottom). The mean DVH are plotted as dashed lines with the standard deviation shown as a shaded region.

In order to establish the significance of differences in metrics, statistical analysis was performed as described in the appendix. Six plans provide too few samples to test the normality of the data, and therefore no assumptions about the normality of the data can be made and the Wilcoxon paired test was the appropriate statistical test of the significance of differences. Bland-Altman (BA) plots (described in the appendix) are here used to illustrate the difference between metrics, and a BA a plot is shown for the PTV D50% metric in Figure 19. The mean difference and its significance can be displayed for many structures and metrics using a colour coded table, shown in Table 4. Only metrics with a statistically significant difference ($p < 0.05$) have data displayed in the table.

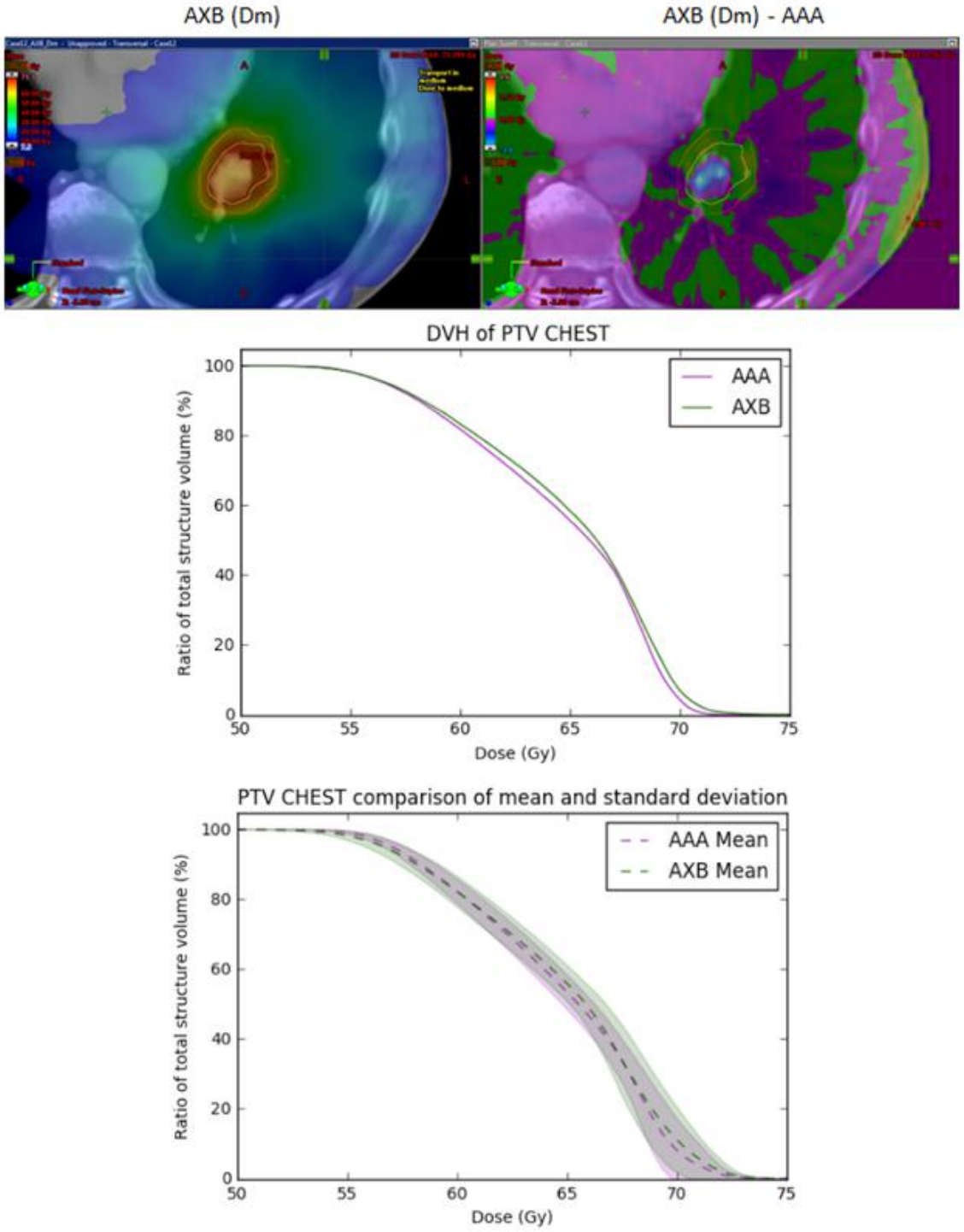

Figure 18. (Top left) dose distributions for a single plan calculated with AXB. The ITV and PTV contours are shown. (Top right) plan sum with AXB – AAA. Where AXB > AAA the colour is green. (Middle) PTV DVH for this single plan. (Bottom) PTV mean and standard deviation (shaded region) across all six plans in the study.

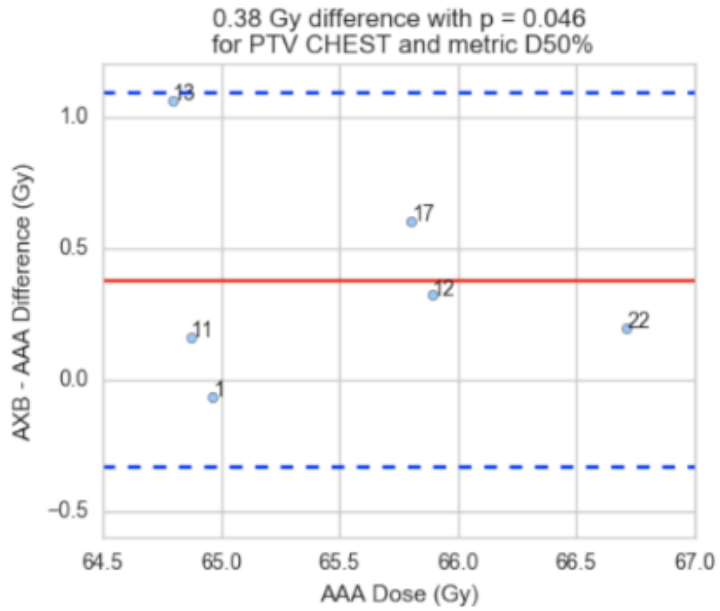

Figure 19. Bland-Altman plot showing the PTV D50% metric. Each point indicates: (horizontal axis) the AAA calculated dose, and (vertical axis) the dose difference (AXB – AAA in Gy) for that case, with the case ID number labelled. The red line indicates the mean difference, and the blue dashed lines indicate +/- 1.96 SD.

	DMAX	D0.1CC	D1CC	D5%	D50%	D95%	D99%
PTV CHEST					0.38		-0.68
ITV							
Foramen	-0.44	-0.39	-0.34	-0.32	-0.087		
Oesophagus	-0.3	-0.23	-0.17	-0.15			
L Brachial plex							
Rt Brachial plex							
Heart				-0.17			0.036
Trachea							
Bronchial tree						-0.1	-0.079
Ribs				-0.28	-0.069		
Liver							
Skin				0.51	-0.028	0.0065	0.0013

metric

Table 4. Statistically significant (p < 0.05) mean metric differences AXB-AAA (Gy).

6.3 Discussion

Referring first to the DVH of Figure 18 it appears that the differences between AAA/AXB are subtle. The data in Table 4 shows that the majority of metrics do not differ between AAA/AXB, and that those differences that are present are of the order of approximately 0.5 Gy (0.9% of the prescription dose) or less. The metrics are next discussed individually:

- In clinical practice plans are normalised so that the PTV D95% matches the prescription dose and the data in Table 4 shows that the PTV D95% is not significantly different between the algorithms.

- The D99% should receive greater than 90% of the prescription dose (D99% > 49.5 Gy), and the data shows that the PTV D99% metric is 0.68 Gy lower for AXB. This difference is the largest in magnitude for any metric, but the BA plot of the PTV D99% data in Figure 20 shows that no plans would fall below the dose constraint due to this difference, and that this target is generally achieved with a comfortable margin. Therefore, this difference is unlikely to have any impact on clinical practice.

- The PTV D50% (median dose) is significantly different between AAA/AXB and higher with AXB than AAA. There are no constraints on the PTV D50% so this difference would not affect the clinical use of the plan.

- The clinical protocol contains dose maximum (D0.1cc, DMAX etc.) constraints for the OAR, but Table 4 shows that the only OAR with significant differences are the Foramen and Oesophagus, which are had lower metrics with AXB. This is potentially clinically significant since the foramen dose is often a limiting factor to the dose escalation to the PTV. Since AXB will calculate a ~ 0.5 Gy lower dose to the foramen than AAA, a planner may *escalate* the dose to the PTV higher than they would have if they had been planning with AAA, and this could have clinical consequences such as increased toxicity and/or complications.

- Whilst the skin does not have dose constraints, the D5% metric is higher by 0.51 Gy for AXB. This difference is consistent with the improved accuracy of AXB observed at the air-water interface in section 4.2.

- The only other metric which could have clinically significant consequences is the PTV D99%, which is lower as calculated by AXB. In an extreme scenario, a plan which would be clinically acceptable if calculated by AAA may require more MU in order to meet the minimum dose requirement when calculated using AXB. However Figure

20(top) indicates that this scenario is unlikely since this metric is typically well within tolerance.

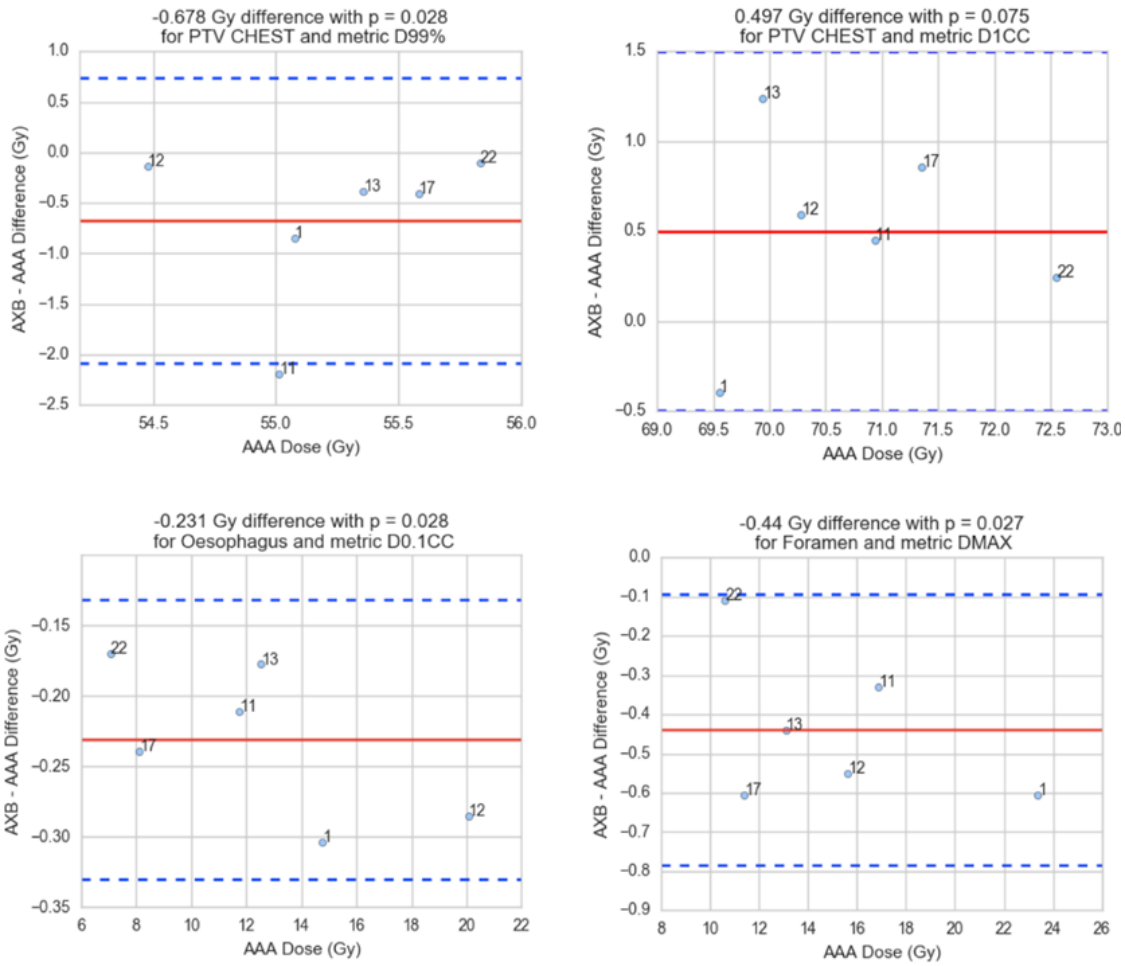

Figure 20. (Top left) PTV D99% metric – constraint of D99% >49.5 Gy, (Top right) PTV D1CC constraint 64.9 to 73.7Gy, (Bottom left) Oesophagus D0.1CC - constraint of D0.1CC < 24 Gy, (bottom right) Foramen DMAX – constraint of DMAX < 18 Gy.

6.4 Summary & recommendations

Overall the majority of dose metrics show no significant difference due to the algorithms. Significantly, the PTV D95% to which plans are normalised does not differ between algorithms. Therefore the recommendation of this study is that the overall guidance on planning lung treatments should remain as it currently is. However this study indicates that the Foramen and Oesophagus metrics do differ with algorithm. Therefore a follow up study should be performed to investigate whether significant differences in doses to these structures are observed in lung plans produced and optimised using AXB, and whether it might be necessary to adjust OAR constraints.

7 Head & Neck Study

7.1 Introduction

At RSCH head and neck (H&N) treatments follow the recommendations of the PARSPORT(Institute of Cancer Research; 2016) trial. The standard prescription for H&N treatments is 65Gy/30# to the primary PTV (PTV1, median dose D50% = 65 Gy) and 54Gy/30# to nodal areas (PTV2, D50% = 54 Gy). There are multiple OARs in the H&N region, shown in Figure 21. Many of the OARs are serial organs therefore have constraints to the maximum dose. The full protocol requirements are given in Appendix – H&N protocol.

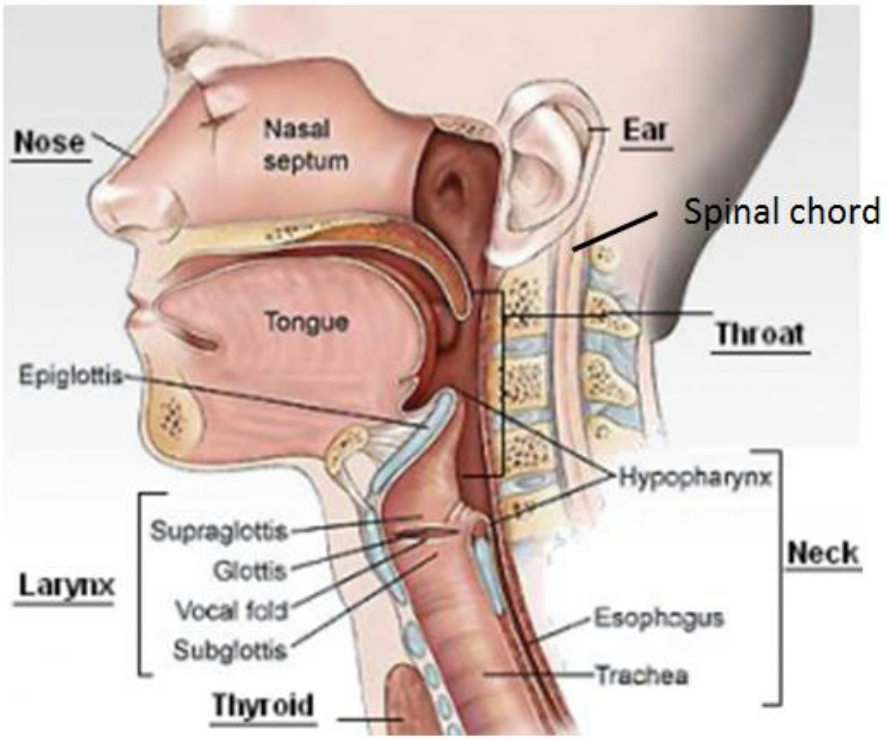

Figure 21. Anatomy of the H&N region, adapted from reference(Hull and East Yorkshire Hospitals NHS Trust 2017)**.**

7.2 Results

Six patient plans were included in this study, all VMAT RapidArc (Varian Medical Systems 2017) with 2 or 3 arcs at 6MV. As with the lung study, each plan was produced previously by an experienced planner using AAA and then doses here recalculated using AXB (Dm) with fixed MU. A single plan is shown in Figure 22(top) and the mean DVH for all six plans is shown in Figure 22(bottom). The PTV2 DVH has a large spread since the position and volume

of the PTV2 can vary significantly with respect to the PTV1.Table 5 shows the significant metrics and Figure 23 shows several BA plots.

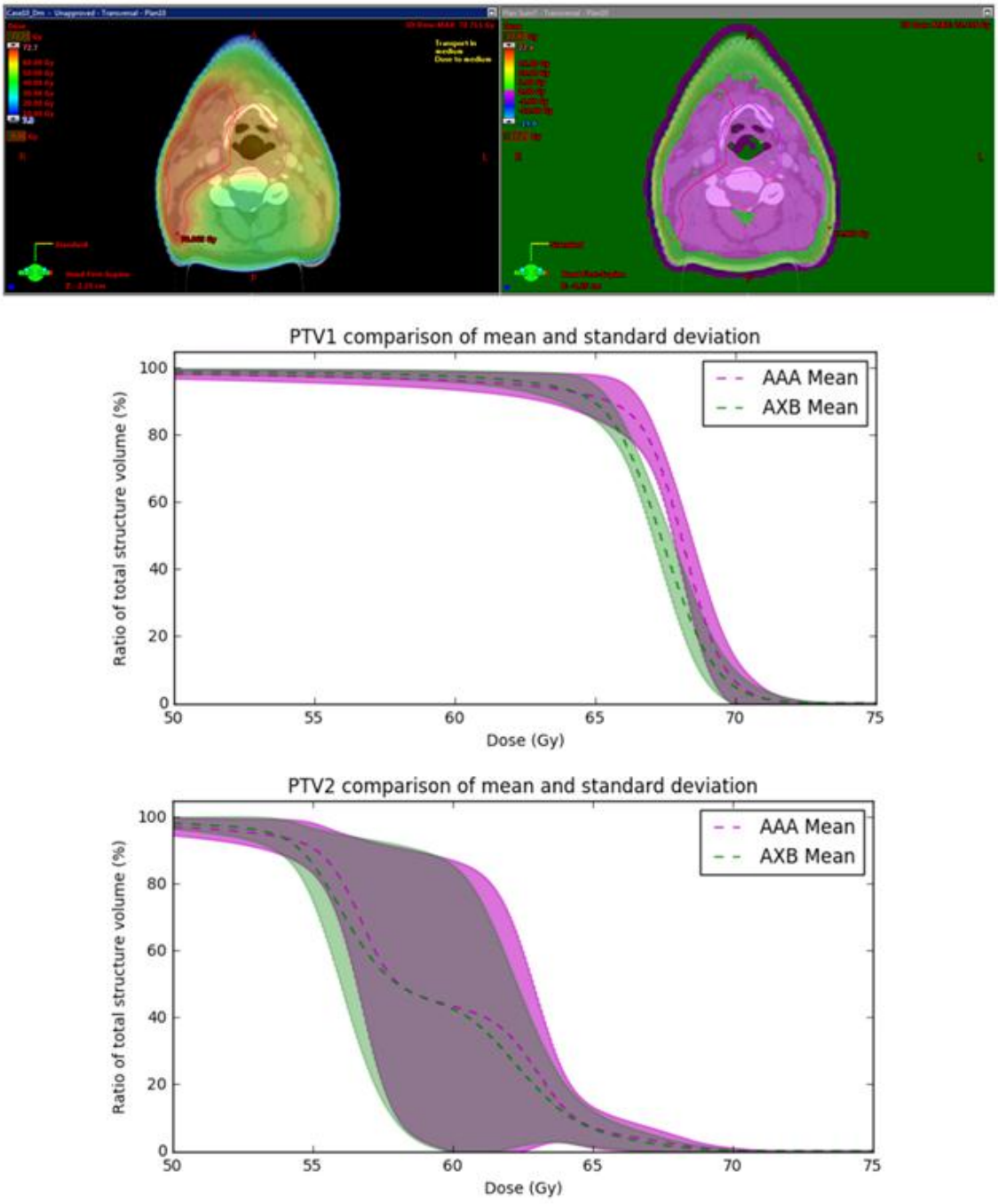

Figure 22. (Top left) Dose distribution of a single plan calculated with AXB. (Top right) plan sum with AXB – AAA. Where AXB > AAA the colour is green. (Middle) the mean and standard deviation of the DVH of all six plans for the PTV1 and (bottom) PTV2.

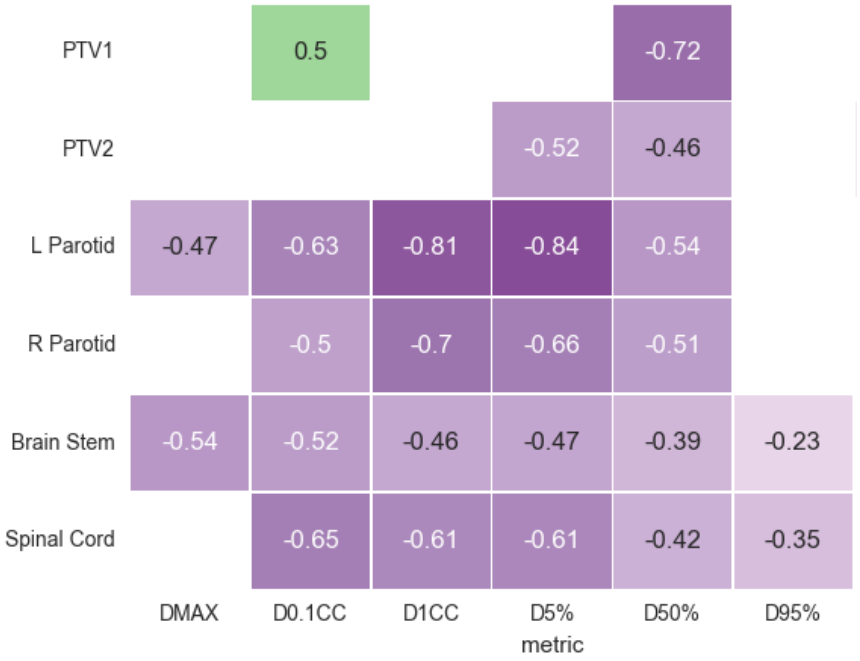

Table 5. Significant dose differences AXB-AAA (Gy) for the H&N plans. AXB consistently calculates a lower dose.

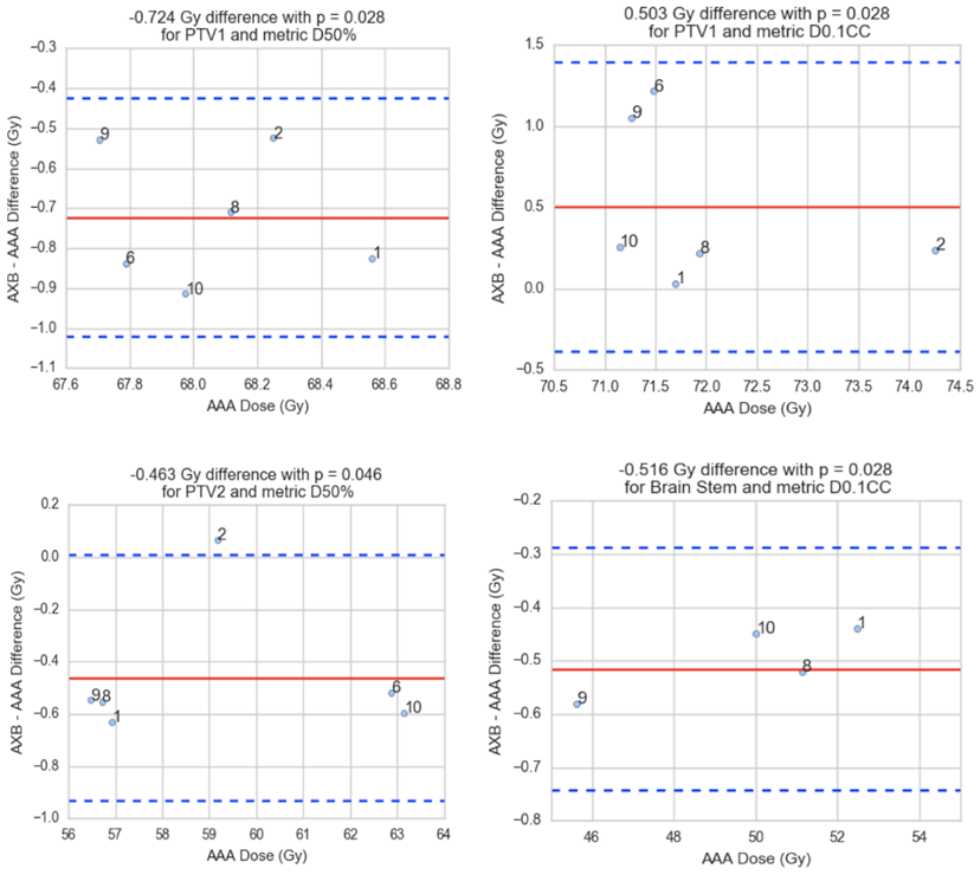

Figure 23. BA plots of the (top left) PTV D50% and (top right) D0.1CC metric. (Bottom left) The PTV2 D50% and (bottom right) the brain stem D0.1CC (showing only the hottest four cases).

7.3 Discussion

Compared to the lung results in Section 6, the H&N study appears to show a very clear trend with AXB calculating lower doses for most structures. Clinical plans are normalised to the PTV D50% and referring to Table 5 this metric is 0.72 Gy lower (1.1% of the prescription dose) with AXB. The implication is that when AXB plans are normalised the majority of the dose metrics will rise accordingly, but since the majority are lower with AXB the net result will be small. One exception is the PTV1 D0.1CC which is 0.5 Gy *higher* with AXB. Therefore the net effect of normalisation will be that the PTV1 D0.1CC will be *higher* by approximately (0.5 + 0.72)/65 (Gy) = 1.9 % compared to an AAA calculated plan. However there is no tolerance for the PTV1 D0.1CC so this should not influence clinical practice directly, but may cause the planner to make adjustments to mitigate PTV1 D0.1CC hot spots.

The spinal cord and brain stem have D0.1CC limits of 45 and 54 Gy respectively. The spinal cord D0.1CC is 0.65 Gy lower with AXB, corresponding to a difference of -1.44%, whilst the brain stem difference of 0.52 Gy corresponds to a difference of -0.96%. Therefore the normalisation of the PTV D50% would result in a lower AXB calculated dose for the spinal cord (-1.44% + 1.1% = -0.34%) and the planner may be tempted to heat the plan more than they would have had they used AAA. Conversely the spinal cord calculated dose would be higher (-0.96% + 1.1% = 0.14%) so a plan that would have been perfectly on tolerance if calculated with AAA may be out of tolerance with AXB.

7.4 Summary & recommendations

Since the vast majority of the differences are strongly correlated and plans are normalised, dose metrics will be largely unaffected by the choice of algorithm AAA/AXB. However this study indicates that dose distributions calculated with AXB may show more pronounced hot spots such as to the PTV1 D0.1CC. These could cause a planner to optimise a plan differently with AXB than had they used AAA. Therefore the recommendation is that overall guidance on planning H&N treatments be maintained as it currently is, but that a follow up study be performed to investigate hot spots and whether it might be necessary to adjust OAR constraints.

8 Report summary, conclusions & recommendations

The LQ model discussed in Section 1 predicts that small changes in dose can have significant consequences for radiotherapy treatments, including reduced tumour control or increased normal tissue complications. Section 2 described several of the algorithms in common use, highlighting the different methods they use to calculate dose. As computing power increases more accurate algorithms are being used clinically, and there is currently a shift taking place from the common use of model based (type b) algorithms to the more accurate principle based (type c) algorithms. Furthermore model based algorithms differ significant from principle based algorithms in that they calculate and report dose-to-medium by default. This creates challenges when comparing Dm calculations to measurements of Dw (reported in section 14), and results in statistically significant dose differences in calculated clinical plans.

Focussing on the clinical consequences of the dose differences between AAA/AXB, this report has confirmed that algorithms can report significantly different doses with differences typically in the range 0 – 1% for most tissues. However the magnitude of the differences increases with the density of the material, and are of order 4% for bone. Non-biological materials show an even larger difference between algorithms, and a study into these differences is ongoing. However in clinical practice, only biological materials could be selected for calculation, minimising the potential impact of differences.

Investigations were performed for two clinical sites in this report, the lung and the H&N. Differences between algorithms were observed for both sites, with few significant differences for lung but many differences for the H&N. For both sites the differences are unlikely to have significant clinical consequences, as discussed in the respective summaries. Therefore the overall recommendation of this report is that guidelines on planning should remain as they currently are. However a follow up study should be performed to investigate the differences in plans that have been produced and optimised using AXB. Complementary to this a study could be performed to investigate the differences between AXB Dm calculated doses and measured doses, as discussed in section 14.

9 References

Barnett, G.C. et al., 2009. Normal tissue reactions to radiotherapy: towards tailoring treatment dose by genotype. *Nature reviews. Cancer*, 9(2), pp.134–42. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2670578&tool=pmcentrez&rendertype=abstract.

Chaikh, A. et al., 2014. The choice of statistical methods for comparisons of dosimetric data in radiotherapy. *Radiation oncology (London, England)*, 9, p.205. Available at: /pmc/articles/PMC4261592/?report=abstract.

Consortium, S.U., 2014. Stereotactic Ablative Body Radiation Therapy (SABR). *Version 5.0, January 2015*, (April), pp.1–99. Available at: http://www.actionradiotherapy.org/wp-content/uploads/2014/12/UKSABRConsortiumGuidellinesv5.pdf.

Elekta, 2017. Monaco®. Available at: https://www.elekta.com/software-solutions/treatment-management/external-beam-planning/monaco.html.

Failla, G.A. et al., 2010. Acuros ® XB advanced dose calculation for the Eclipse ™ treatment planning system. , p.32.

Failla, G.A. et al., 2010. *Acuros XB advanced dose calculation for the Eclipse treatment planning system*, Available at: https://www.varian.com/sites/default/files/resource_attachments/AcurosXBClinicalPerspectives_0.pdf [Accessed August 31, 2016].

Fogliata, A. et al., 2012. Critical appraisal of acuros XB and anisotropic analytic algorithm dose calculation in advanced non-small-cell lung cancer treatments. *International Journal of Radiation Oncology Biology Physics*, 83(5), pp.1587–1595. Available at: https://www.researchgate.net/profile/Luca_Cozzi/publication/221801435_Critical_Appraisal_of_Acuros_XB_and_Anisotropic_Analytic_Algorithm_Dose_Calculation_in_Advanced_Non-Small-Cell_Lung_Cancer_Treatments/links/54009cc40cf2c48563ae56d7.pdf.

Fogliata, A. & Cozzi, L., 2016. Dose calculation algorithm accuracy for small fields in non-homogeneous media: The lung SBRT case. *Physica Medica*. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1120179716309887.

Gohar, R., 2017. Comparative study of AAA and PBC. Available at: http://www.slideshare.net/ramahunzai/comparative-study-of-aaa-and-pbc-1 [Accessed January 23, 2017].

Hall, E.J. & Giaccia, A.J., 2017. EMITEL. Available at: http://preview.emitel2.eu/emitwwwsql/encyclopedia.aspx [Accessed March 20, 2017].

Han, T. et al., 2011. Dosimetric comparison of Acuros XB deterministic radiation transport method with Monte Carlo and model-based convolution methods in heterogeneous media. *Medical Physics*, 38(5), p.2651. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21776802 [Accessed August 22, 2016].

Huang, B. et al., 2015. Dose calculation of Acuros XB and Anisotropic Analytical Algorithm in lung stereotactic body radiotherapy treatment with flattening filter free beams and the potential role of calculation grid size. *Radiation Oncology*, 10(1), p.53. Available at: https://ro-journal.biomedcentral.com/articles/10.1186/s13014-015-0357-0.

Hull and East Yorkshire Hospitals NHS Trust, 2017. Radiotherapy to the Head and Neck. Available at: https://www.hey.nhs.uk/patient-leaflet/radiotherapy-head-neck/ [Accessed March 15, 2017].

Institute of Cancer Research;, 2016. PARSPORT. Available at: http://www.icr.ac.uk/our-research/our-research-centres/clinical-trials-and-statistics-unit/clinical-trials/parsport [Accessed March 15, 2017].

International Commission on Radiation Units &, M., 2010. Prescribing, Recording, and Reporting Intensity-Modulated Photon-Beam Therapy (IMRT)(ICRU Report 83).

Journal of the ICRU, 10(1), pp.1–2. Available at: http://www.icru.org/testing/reports/prescribing-recording-and-reporting-intensity-modulated-photon-beam-therapy-imrt-icru-report-83.

Khan, Faiz M., and J.P.G., 2014. *Khan's the physics of radiation therapy*, Lippincott Williams & Wilkins.

Knoos, T. et al., 1995. Limitations of a pencil beam approach to photon dose calculations in lung tissue. *Physics in Medicine and Biology*, 40(9), pp.1411–1420. Available at: http://stacks.iop.org/0031-9155/40/i=9/a=002?key=crossref.5939bba3b7e0b78b2771316c6bd6240f [Accessed January 23, 2017].

Knoos, T., Mcclean, B. & Carlo, M., 2008. Dose calculation algorithms in 3DCRT and IMRT Fluence to dose Convolution. , pp.1–22.

Lu, L., 2013. Dose calculation algorithms in external beam photon radiation therapy. *International Journal of Cancer Therapy and Oncology*, 1(2), pp.01025–01028. Available at: http://ijcto.org/index.php/IJCTO/article/view/Lu/ijcto.0102.5pdf.

Marks, L.B., Ten Haken, R.K. & Martel, M.K., 2010. Guest Editor's Introduction to QUANTEC: A Users Guide. *International Journal of Radiation Oncology Biology Physics*, 76(3 SUPPL.), pp.S1-2. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20171501 [Accessed August 2, 2016].

National Research Council Canada, 2017. EGSnrc: software tool to model radiation transport - National Research Council Canada. Available at: http://www.nrc-cnrc.gc.ca/eng/solutions/advisory/egsnrc_index.html [Accessed March 16, 2017].

PDQ Adult Treatment Editorial Board, 2002. Non-Small Cell Lung Cancer Treatment. *PDQ Cancer Information Summaries*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26389355 [Accessed March 15, 2017].

RadCalc, 2017. RadCalc. Available at: http://www.lifelinesoftware.com/index.php/radcalc.

Rana, S. et al., 2014. Evaluation of Acuros XB algorithm based on RTOG 0813 dosimetric criteria for SBRT lung treatment with RapidArc. *Journal of applied clinical medical physics / American College of Medical Physics*, 15(1), p.4474. Available at: http://www.jacmp.org/index.php/jacmp/article/view/4474/html_4 [Accessed February 2, 2016].

Rana, S. & Pokharel, S., 2014. Dose-to-medium vs. dose-to-water: Dosimetric evaluation of dose reporting modes in Acuros XB for prostate, lung and breast cancer. *International Journal of Cancer Therapy and Oncology*, 2(4), p.20421. Available at: http://www.ijcto.org/index.php/IJCTO/article/view/0204.21.

Rock, L. et al., 2004. *Tissue Inhomogeneity Corrections for Megavoltage Photon Beams AAPM REPORT NO. 85*, Available at: www.medicalphysics.org [Accessed June 30, 2016].

Scipy.stats, 2017. Friedman Rank Sum Test. Available at: https://docs.scipy.org/doc/scipy-0.15.1/reference/generated/scipy.stats.friedmanchisquare.html.

scipy.stats, 2017a. scipy.stats.jarque_bera. Available at: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.jarque_bera.html [Accessed January 27, 2017].

scipy.stats, 2017b. Wilcoxon Signed-Rank Test. Available at: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wilcoxon.html [Accessed February 3, 2017].

Thompson, R.F., 2014. RadOnc : An R Package for Analysis of Dose-Volume Histogram and Three-Dimensional Structural Data Software. *JROI Journal of Radiation Oncology Informatics*, 6(1), pp.98–110. Available at: http://jroi.org/index.php/jroi/article/viewFile/25/25.

Varian, 2017. Eclipse™ Treatment Planning System | Varian Medical Systems. Available at:

https://www.varian.com/en-gb/oncology/products/software/treatment-planning/eclipse.

Varian Medical Systems, 2017. VMAT | RapidArc. Available at: https://www.varian.com/en-gb/oncology/treatment-techniques/external-beam-radiation/vmat [Accessed March 21, 2017].

Zifodya, J.M., Challens, C.H.C. & Hsieh, W.L., 2016. From AAA to Acuros XB-clinical implications of selecting either Acuros XB dose-to-water or dose-to-medium. *Australasian Physical and Engineering Sciences in Medicine*, pp.1–9. Available at: http://link.springer.com/article/10.1007/s13246-016-0436-z.

10 Appendix – DVH & dose metrics

Dose-volume-histograms (DVH) and dose metrics and are used to evaluate and compare radiotherapy plans. A DVH is a visual representation of the relationship between a structures volume and the calculated dose to the volume, where the relative volume (%) of the structure is on the vertical plot axis and dose is on the horizontal axis. Two types of DVH are in common use; the cumulative and the differential DVH, shown below in Figure 24.

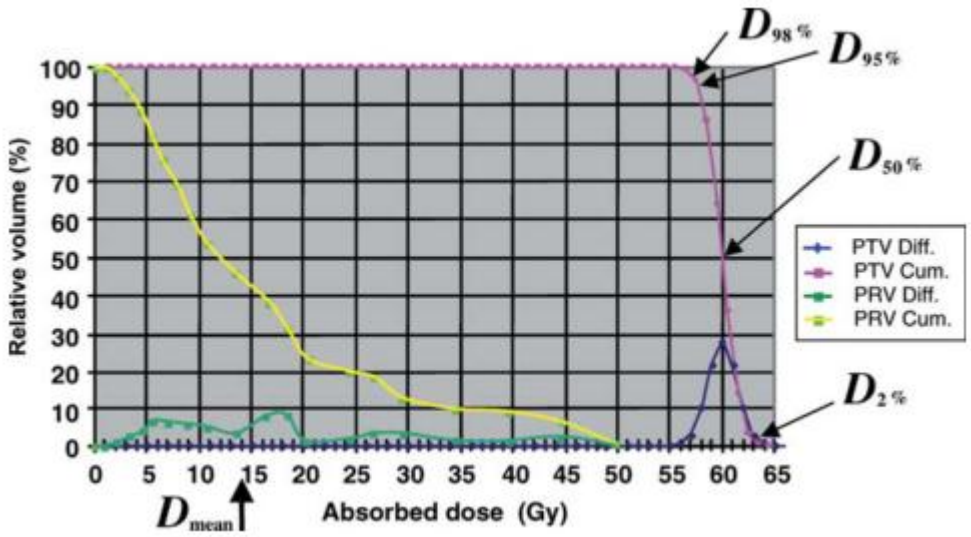

Figure 24. Related cumulative and differential DVH. Several dose metrics are indicated. Figure from ICRU report 83 (International Commission on Radiation Units & 2010)**.**

This report makes exclusive use of the cumulative DVH, where each point on the DVH represents the *minimum* dose that the volume would receive. Cumulative DVH are used to calculate dose metrics, some of which can be read directly from the DVH such as the D50% which is the dose received by 50% of the volume. Other dose metrics do not refer to relative volumes but to absolute volumes, such as the D1CC which is the dose that the hottest 1 cm^3 of a volume receives. Care must be taken when comparing dose metrics as publications sometimes use different nomenclature to refer to common metrics, for example D50% and D_{Median} are equivalent and may be used interchangeably. Other examples include the D98% and $D_{Near\text{-}min}$, as well as D2% and $D_{Near\text{-}max}$.

11 Appendix – Methodology of statistical tests

Both the AAA (V 11.0.31) and AXB (V 13.6.23) algorithms are implemented in the Eclipse TPS (Varian 2017) at RSCH. Plans in this study were previously produced by experienced planners within the department and optimised using AAA. For this study the author recalculated AAA plans using AXB Dm with fixed MU. Various dose metrics (D50% etc.) were calculated from exported DVH text files using the RadOnc R package (Thompson 2014). The metrics were analysed using python packages listed in the text below. There are numerous publications describing the comparison of dose metrics as calculated by different algorithms. From this survey a 'best practice' analysis process was developed and applied in this report. The process is as follows:

1. The absolute differences between AXB/AAA DVH dose metrics are first calculated using RadOnc.
2. A statistical test is then performed using python to assess whether the differences either occurred by chance or are statistically significant. The Friedman's test of repeated measurements is suitable for this purpose (Scipy.stats 2017). Note however that this test is only reliable for > 10 samples per algorithm.
3. Assuming a positive result, a test of the normality of the data would then determine whether the appropriate statistical tests for subsequent analysis are parametric or non-parametric. Parametric tests are considered more powerful than equivalent non-parametric tests and can detect differences with smaller sample sizes. The Jarque-Bera test of normality is suitable, but only works for > 2000 samples (scipy.stats 2017a). Since all of the investigations in study yield far fewer than 2000 samples per metric it is impossible to determine the normality of the data or otherwise. Therefore non-parametric statistical tests are used throughout.
4. To determine which dose metrics are statistically differnet between algorithms, the Wilcoxon signed-rank test is used (reference (scipy.stats 2017b), non-parametric, described in the Appendix).

It is worth noting that testing for homogeneity of variance is not necessary when comparing the metrics of two idential plans calculated using different algorithms since this constitues a *repeated* measurement. However to compare more than two algorithms a two-step analysis process would be required, consisting of an omnibus test followed by multiple paired t-tests(Chaikh et al. 2014).

12 Appendix – Wilcoxon signed rank test

A useful description of the Wilcoxon signed rank is available online(scipy.stats 2017b).

In summary:

1. The differences are ranked in order from smallest to largest.
2. The sign of the difference determined.
3. The signed rank is computed and summed giving the test statistic W.
4. The number of pairs (N) is used to calculate sigma_x
5. From W and sigma_x the z_score the significance level (p-value) of the difference is determined.

Here the test is applied to data for PTV chest structure, D1CC metric.

#	Case	AAA (Gy)	Dm (Gy)	diff (%)	abs(diff)	Sign	Rank	Signed-rank
1	Case31	66.191717	66.270579	0.119	0.119142	1	1	1
2	Case14	74.288008	74.380407	0.124	0.12438	1	2	2
3	Case22	72.554205	72.797202	0.335	0.334918	1	3	3
4	Case1	70.189127	69.751186	-0.624	0.623945	-1	4	-4
5	Case11	70.937631	71.39079	0.639	0.638813	1	5	5
6	Case12	70.280686	70.872441	0.842	0.841989	1	6	6
7	Case32	65.31573	66.01578	1.072	1.071794	1	7	7
8	Case25	71.85036	72.703666	1.188	1.187615	1	8	8
9	Case17	71.353276	72.208572	1.199	1.198678	1	9	9
10	Case13	69.938674	71.175986	1.769	1.769138	1	10	10
11	Case16	71.090959	72.614765	2.143	2.14346	1	11	11
12	Case30	65.769778	64.195821	-2.393	2.393132	-1	12	-12
13	Case18	68.26195	69.930371	2.444	2.444145	1	13	13
14	Case23	69.869227	71.868429	2.861	2.861348	1	14	14
			Mean diff =	0.837			Sum (W) =	73
							N =	14
							Sigma_x =	31.859
							z_score =	2.291
						From lookup table		p < 0.025

13 Appendix – Bland-Altman plots

In this report, tabulated data of calculated dose metrics and their differences are displayed using a Bland-Altman (BA) plot. The purpose of a BA plot is to allow visualisation of the data in a way which highlights the distribution of metrics and their differences about their mean values, and aids the identification of plans with outlier characteristics. Several plans were excluded from both the lung and the H&N studies as result of examining BA plots showed out of tolerance metrics in the AAA plan.

Case	structure	metric	AAA	Dm	Dm-AAA
Case1	PTV CHEST	D50%	64.965374	64.900700	-0.064674
Case11	PTV CHEST	D50%	64.871795	65.032941	0.161147
Case12	PTV CHEST	D50%	65.889287	66.211549	0.322262
Case13	PTV CHEST	D50%	64.793739	65.854182	1.060442
Case17	PTV CHEST	D50%	65.804513	66.406596	0.602083
Case22	PTV CHEST	D50%	66.714681	66.914450	0.199769

Table 6. PTV D50% metric raw data.

Take case 12, the data of which is in Table 6 and the BA plot in Figure 25. The AAA metric is 65.89 Gy, and the mean difference between AXB and AAA (AXB Dm - AAA) is 0.322 Gy. In the BA plot the AAA value is on the horizontal axis, and the difference on the vertical axis, with the plan numbers as labels. The mean of the difference column (mean(Dm – AAA) = 0.38 Gy) is plotted as a red line, and the blue dashed lines represent the mean +/- 1.966 standard deviations from the mean, where statistically we expect 1 in 20 values to lie at our outside this range if the data is normally distributed.

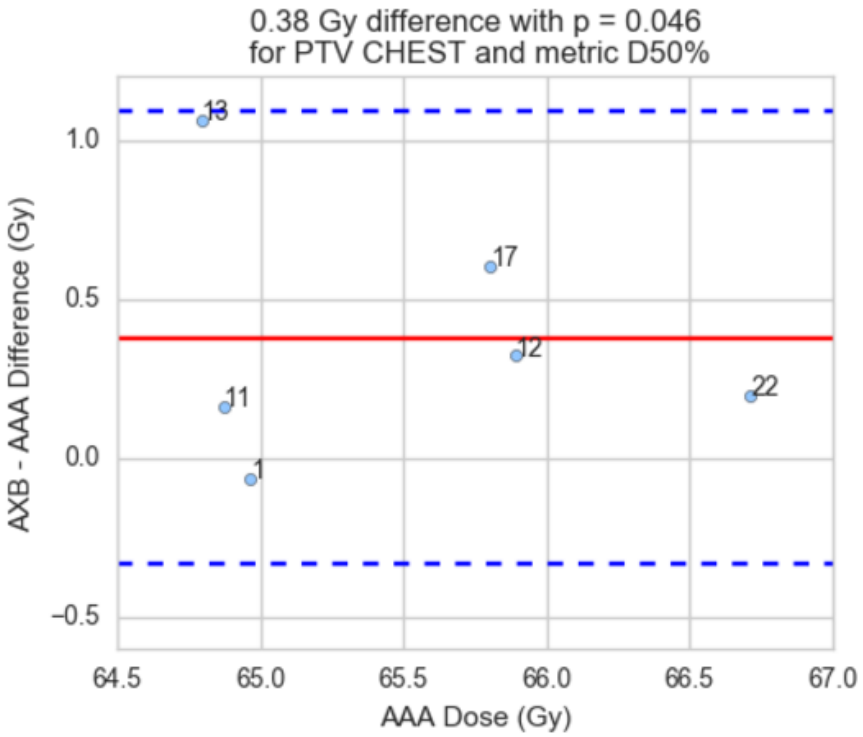

Figure 25. BA visualisation of the data in Table 6.

14 Appendix – Lung SABR protocol

Protocol document RT-PT-56 v1

CONTROLLED DOCUMENT

STEREOTACTIC ABLATIVE BODY RADIOTHERAPY (SABR) CLINICAL PROTOCOL
Non-Small Cell Lung Cancer

1.12 PTV Dose Constraints and Conformality of plan

At least 95% of the PTV should receive 100% of prescribed total dose. 99% of PTV should receive a minimum of 90% of the prescribed dose. 1cc maximum dose within the PTV must be between 118 to 134% of prescribed dose, with a point max dose not exceeding 135%.

1cc maximum dose 2cm from PTV must be < 65% of total dose. Ratio of Volume of 100% isodose to volume of PTV must be <1.25. Ratio of 50% isodose to volume of PTV must be <12 but ideally less than 4.5

1.9.2 Dose constraints for OAR

OAR		dose constraint 3#	5/8#
Foramen	Max	<18Gy	<25Gy
Oesophagus	D0.1cc	<24gy	<27gy
Brachial Plexus	D0.1cc	<24Gy	<27Gy
Heart	D0.1cc	<24Gy	<27Gy
Trachea	D0.1cc	<30Gy	<32Gy
Bronchial Tree	D0.1cc	<30Gy	<32gy
Ribs	D0.1cc	<35Gy	
	V30Gy	<30cc	
Lungs-ITV	V20Gy	<10%	
	V12.5Gy	<15%	
	V5Gy	<26%	
	D1500cc	<10.5Gy	<12.5Gy
	D1000cc	<11.4Gy	<13.5Gy

Liver	3#	5#	8#
	700cc <15Gy	700cc <15Gy	V27 <30%
	V21 <33%	V30 <60%	V24 <50%
	V15 <50%	Mean Dose <20Gy	

15 Appendix – H&N protocol

From protocol RT-PT-45 v3

Primary RT	
SIB-IMRT	
High dose PTV (PTV1)	Elective PTV (PTV2)
65Gy/30#	54Gy/30#

Organ	Parameter	Dose (Gy) in 2Gy/#
Spinal cord	Dmax	46Gy
	0.1cc	45Gy
Spinal cord PRV	Dmax	50Gy
	0.1cc	48Gy
Brain stem	Dmax	55Gy
	0.1cc	54Gy
Brain stem PRV	Dmax	59Gy
	0.1cc	55Gy
Temporal lobe	Mean	< 22Gy
Brachial plexus	Dmax	60Gy
Cochlea	Mean	48Gy
Optic nerves and chiasm	Dmax	50Gy
Retina	Dmax	50Gy
	< 60% Volume	45Gy
Lens	Dmax	8Gy
Lacrimal gland	Dmax	30Gy
Cornea	Dmax	40Gy
Pituitary gland	Dmax	45Gy
Parotid gland	Unilateral mean	20Gy
	Bilateral mean	25Gy
Sumbandibular gland	Mean	35Gy
Larynx	Mean	44Gy
	</= 27% volume	>/= 50Gy