Evidence based assessment of the clinical impact of dose variations arising in the clinical radiotherapy dosimetry chain

by

Matthew Andrew Bolt

Submitted for the degree of Doctor of Philosophy

August 2018

Department of Chemical and Process Engineering
Faculty of Engineering and Physical Sciences
University of Surrey

© Matthew Bolt 2018
Statement of Originality

This thesis and the work to which it refers are the results of my own efforts. Any ideas, data, images or text resulting from the work of others (whether published or unpublished) are fully identified as such within the work and attributed to their originator in the text, bibliography or footnotes. This thesis has not been submitted in whole or in part for any other academic degree or professional qualification. I agree that the University has the right to submit my work to the plagiarism detection service TurnitinUK for originality checks. Whether or not drafts have been so-assessed, the University reserves the right to require an electronic version of the final document (as submitted) for assessment as above.

© Matthew Bolt 2018
Abstract

Objectives

The accuracy of delivered dose depends directly upon initial beam calibration and subsequent maintenance of this beam output. The uncertainty associated with these measurements and its impact on clinical outcomes is not well documented. This work gives an evidence based approach to determining this variation and its clinical impact.

Novelty

This work will quantify for the first time the variations present in the routine maintenance of beam output on a national scale. The novel application of these dosimetric uncertainties to radiobiological models is then employed to predict the variation in clinical outcome due to the quantified dosimetric variations for specific clinical cases, including both tumour control and associated treatment complications on both individual and patient populations.

Results

The linear-quadratic and Lyman Kutcher Burman models have been implemented to allow flexibility in the modelling of individual patient doses on a fraction by fraction basis. The variation in delivered doses due to beam output variations is seen to be normally distributed with a standard deviation of 0.7%. These variations may lead to a typical patient experiencing a range in treatment outcome probabilities of over 10% for cancers with a steep dose response curve such as head and neck in both the case of an individual patient and for a patient population.

Conclusions

The precise control of beam output is shown to be a major factor in the overall uncertainty for dose delivery in modern treatment techniques. With reductions in other uncertainties in radiotherapy treatments, now may be the time to consider reduction of tolerance levels to allow optimal patient treatment and outcomes.

Key Words: Radiotherapy, Dosimetry, Audit, Linac, Radiobiological Modelling, Clinical Trial, Beam Output, Calibration, Dose Volume Histogram.
Acknowledgements

I would like to thank my supervisors Dr Tao Chen, Professor Andrew Nisbet and Dr Catharine Clark for their guidance throughout this PhD. Without their experience, and wealth of knowledge many of the ideas explored within this work would not have come to fruition, I couldn’t have asked for better supervisors for this research.

The National Physical Laboratory (NPL) provided the funding (through the National Measurement System) for this work and so made it possible for me to embark on this PhD journey and hopefully make a meaningful impact. I must thank all the members of the radiation dosimetry team at the NPL for welcoming me and providing advice whenever required. In particular I owe a great deal to Russell Thomas and Graham Bass for sharing their extensive knowledge and wisdom on the dosimetry audits conducted by NPL on which a large portion of this work is based.

I am thankful to all my colleagues in the Radiotherapy Physics Department at the Royal Surrey County Hospital, particularly my managers Tom Jordan, Liz Adams, Melanie Cunningham and Chris South for allowing me protected time away from my clinical role to allow me to carry out my research.

Audit is very much a collaborative process and I would like to extend my gratitude to all staff who have worked on the NPL and regional audit programs and to clinical colleagues who have welcomed them into their departments. All have contributed to the continued improvement of radiotherapy within the UK and this research has relieved heavily upon their valuable work.

I would like to extend my thanks and gratitude to all the staff involved in collating and supplying the local data for this work, or for allowing visits to gather the data, taking time out of their busy clinical schedules to accommodate me to enable this research to progress.

Finally I would like to thank my wife, Emma. Without her ongoing support and love I would not have been able to dedicate the time and energy required to complete this work. I also want to thank my family, friends and fellow karateka for all their years of support before and during the PhD, which without I would not be where I am now.
# Table of Contents

List of Figures ........................................................................................................ vi

List of Tables ........................................................................................................... xv

Common Abbreviations ............................................................................................ xviii

1 Introduction ............................................................................................................. 1

1.1 Overview .......................................................................................................... 1

1.1.1 Objectives of this work .............................................................................. 2

1.1.2 Novelty of work ......................................................................................... 3

1.1.3 Overview of thesis ...................................................................................... 3

1.2 Radiotherapy .................................................................................................... 4

1.2.1 Radiotherapy treatment techniques ......................................................... 5

1.2.2 Uncertainties within radiotherapy treatments ...................................... 7

1.3 Quality assurance in radiotherapy ................................................................. 7

1.3.1 Beam output calibration .......................................................................... 9

1.3.2 Maintenance of beam output .................................................................. 10

1.4 Audit in radiotherapy ...................................................................................... 11

1.4.1 Purpose of dosimetry audit ..................................................................... 11

1.4.2 Brief history of dosimetric audit ............................................................... 12

1.4.3 Dosimetry audit within the UK ................................................................. 13

1.4.4 Continued development of dosimetry audit ......................................... 15

1.4.5 Dosimetry audit within clinical trials ..................................................... 16

1.4.6 Evidence of the impact of audit ............................................................... 17

1.5 Radiobiological modelling .............................................................................. 18

1.5.1 Target theory ............................................................................................ 19

1.5.2 Tumour control probability: Linear quadratic model ............................ 21

1.5.3 Alternative empirical models investigated ............................................ 25

1.5.4 Normal tissue complication probability: Lyman-Kutcher-Burman model 31

1.6 Implementation of tools for data analysis and modelling ............................ 33
### Table of Contents

1.6.1 Implementation of modelling process .................................................. 33

1.7 An area of open investigation ................................................................. 35

2 Regional audit data .................................................................................. 36

2.1 Collation of regional audit data ................................................................. 37

2.2 Previously published regional audit results .............................................. 38

2.2.1 Previously published results from south west regional audit group D .... 38

2.2.2 Previously published results from south east central audit group E ....... 38

2.3 Audit schedules ..................................................................................... 39

2.4 Summary of regional audit data ............................................................... 40

3 Analysis of NPL audits ............................................................................ 41

3.1 Methods employed in NPL dosimetry audits ......................................... 41

3.1.1 General procedure ............................................................................ 41

3.1.2 Audit setup ....................................................................................... 42

3.1.3 Beam quality ................................................................................... 44

3.1.4 Beam output .................................................................................... 45

3.1.5 Reporting of results and treatment of data ........................................ 47

3.1.6 Statistical comparisons .................................................................... 47

3.1.7 Measurement uncertainties ............................................................... 47

3.2 Results of NPL audits ............................................................................ 48

3.2.1 Beam output results ....................................................................... 49

3.2.2 Beam quality results ....................................................................... 61

3.2.3 Ionisation chamber calibration intercomparison results ..................... 62

3.3 Discussion of NPL audit results .............................................................. 66

3.3.1 Comparison of regional and NPL audit results .................................. 69

3.3.2 Comparison to Papillon audit ........................................................... 70

3.3.3 Comparison to national rotational audit results .................................. 70

4 Machine output measurement .................................................................... 73

4.1 Data collection and formatting .............................................................. 73

4.2 Data analysis ......................................................................................... 74

4.2.1 Data recording method .................................................................... 74
Table of Contents

4.2.2 Grouping by machine age ................................................................. 74
4.2.3 Grouping by centre size ................................................................. 74
4.2.4 Drift in beam output with time ......................................................... 75
4.2.5 Measurement device comparison .................................................... 75
4.2.6 Statistical analysis ........................................................................ 75
4.3 Results ............................................................................................... 75
  4.3.1 Summary of collected data ............................................................. 76
  4.3.2 Typical beam output data ............................................................... 78
  4.3.3 Distribution of treatment machine install years ......................... 79
  4.3.4 Variation in recording methods ...................................................... 80
  4.3.5 Drifts in beam output ..................................................................... 81
  4.3.6 Variation in centre size ................................................................. 82
  4.3.7 Inter machine variation ................................................................. 83
  4.3.8 Comparison of Farmer chamber and constancy device measurements ...... 87
4.4 Analysis of long-term beam output data from RSCH ................................. 93
  4.4.1 Data summary ............................................................................. 94
  4.4.2 Correlation of beam output for different beam energies ............. 94
  4.4.3 Seasonal output trends ................................................................. 96
  4.4.4 Effect of treatment start date on delivered dose ...................... 97
4.5 Discussion of beam output measurements ........................................... 99

5 Analysis of dose variation within the PARSPORT trial ................................. 102
  5.1 Summary of included treatment plans .............................................. 103
  5.2 Development of tool for automated extraction of dosimetric statistics .......... 103
    5.2.1 Data extraction process ............................................................ 105
  5.3 Comparison of doses from conformal and IMRT plans .................. 105
    5.3.1 Reduction in dose to parotid glands with IMRT ..................... 105
    5.3.2 Reduction in dose variation to target (PTV) structure with IMRT ........ 107
  5.4 Summary of dose variation within PTV and OAR structures ............... 109

6 Radiobiological model implementation and characterisation .......................... 110
  6.1 Implementation of radiobiological models ....................................... 110
    6.1.1 Model development and input parameters ............................... 110
6.1.2 General overview of modelling and simulation approach ...................................... 111
6.1.3 Inclusion of dose variation ..................................................................................... 112
6.1.4 Linear quadratic model .......................................................................................... 113
6.1.5 Lyman Kutcher Burman model .............................................................................. 116
6.2 Sensitivity analysis .................................................................................................... 117
6.2.1 Linear quadratic model .......................................................................................... 117
6.2.2 Lyman Kutcher Burman model .............................................................................. 120

7 Clinical impact of dose variation .................................................................................. 124
7.1 Summary of dosimetric uncertainties ......................................................................... 124
7.2 Summary of clinical cases considered ........................................................................ 125
  7.2.1 Cases considered for evaluation of dose variations on tumour control probability  125
  7.2.2 Cases considered for evaluation of dose variations on normal tissue complication probability ................................................................. 125
7.3 Variation in outcome due to dose differences ............................................................. 126
  7.3.1 Variation in outcome due to a known systematic dose difference ...................... 126
  7.3.2 Variation in outcome due to machine scheduling .............................................. 132
  7.3.3 Variation in outcome due to drift in dose over course of treatment ................... 137
  7.3.4 Variation in outcome due to daily output variations ......................................... 140
  7.3.5 Assessment of impact in systematic dose differences on population of patients  145
7.4 Comparison with previous work based on European TLD audits .............................. 147
  7.4.1 TLD audit data used ............................................................................................ 147
  7.4.2 Prediction of clinical outcome using NPL audit results .................................... 147
  7.4.3 Discussion and comparison to Bentzen et al. results ........................................ 148
7.5 Discussion of radiobiological modelling results ......................................................... 149

8 Conclusions and further work ...................................................................................... 152
8.1 Conclusions ................................................................................................................. 152
8.2 Further work ................................................................................................................. 155
  8.2.1 Inclusion of beam output uncertainty into clinical trials .................................. 155
  8.2.2 Improving the prediction of toxicity for individual patients ............................. 156
  8.2.3 Assessment of impact of tolerances used ......................................................... 157
  8.2.4 Financial assessment of the change in beam output tolerances ....................... 158
8.2.5 Produce automated user tools for analysis of beam output and DHV parameter extraction .......................................................... 159
8.2.6 Impact of further QC tests .................................................................................. 159
8.2.7 Multivariate analysis ......................................................................................... 160

Bibliography ............................................................................................................ 161

Appendices ............................................................................................................... 171
Appendix A  Publications and presentations arising from this work ...................... 171
  Appendix A.1  Poster at UK Radiation Oncology Conference 2016: Impact of Linac Output
  Variation on Clinical Outcomes: ........................................................................... 171
  Appendix A.2  Poster at ESTRO 35 (2016): Impact of standardised codes of practice and related audit on radiotherapy dosimetry over 20 years. .......................................................... 172
Appendix A.3  Poster at ESTRO 36 (2017): Variation in mean dose from 204 UK linacs (Jan-June 2015) and its potential clinical impact. .................................................................................. 174
Appendix A.4  Journal Article: Radiotherapy reference dose audit in the United Kingdom by the national Physical Laboratory: 20 years of consistency and improvements .............. 176
Appendix A.5  Journal Article: A multi-centre analysis of radiation beam output measurements ................................................................................................. 177
Appendix B  Details of availability of python code used within this thesis ............... 180
  Appendix B.1  Dosimetric parameter extraction from DICOM RT files .................. 180
  Appendix B.2  Implemented radiobiological models used to within chapters 6 and 7: ................................................................. 180
Appendix C  Availability of anonymised data collated for use within this thesis .......... 180
Appendix D  Description of boxplots as used within this thesis .............................. 181
List of Figures

Figure 1.1: Varian TrueBeam Linac located at the Royal Surrey County Hospital. As well as generating a radiation beam which emerges from the linac head for treatment it also has kV and MV imaging capabilities to aid patient positioning. ................................................................. 5

Figure 1.2: Comparison of dose distributions achievable with 3D conformal techniques (left) and IMRT techniques (right). Conformation around the right parotid is achieved with IMRT, however it is not possible to the same extent with 3DCRT techniques increasing the probability of side effects. ....................................................................................................................... 6

Figure 1.3: Hub and Spoke model initially setup by IPEM for UK audits. The NPL would perform an audit in each of the original eight regions (A-H) annually to tie in audit results performed within each audit group. .................................................................................................................. 14

Figure 1.4: Number of citations on PubMed for “audit” and “radiotherapy” between 1990 and 2017. There is no sign of a reduction in publications on audit in radiotherapy. ................................. 15

Figure 1.5: Schematic showing how both TCP and NTCP increase with dose. The difference between the incidence of TCP and NTCP is known as the therapeutic ratio. Radiotherapy techniques in general will aim to maximise the therapeutic ratio, thus increasing the probability of treatment success whilst minimising associated side effects. ................................................................. 19

Figure 1.6: Examples of single and multi-target single hit models of cell killing based on target theory. Note that the multi target model contains a ‘shoulder’ commonly observed in mammalian cells studies. In this case D_0 = 1.5Gy and n=30. ................................................................. 20

Figure 1.7: Surviving fraction of cells predicted by the linear-quadratic model. The \(\alpha/\beta\) ratio is given at the point where contributions of the linear and quadratic components are equal and describes the shape of the survival curve. The shape of the curve is dependent upon the type of radiation as well as the \(\alpha/\beta\) ratio. .................................................................................................................. 23

Figure 1.8: Production of a population TCP curve constructed from the combination of individual patients. Note that the population curve is shallower than the individual patient curves. In this work the distribution of \(\alpha\) values is assumed to be normally distributed. This example modelled 1000 individual patients and used a dose per fraction of 2Gy, mean \(\alpha/\beta\) of 10 with a 20% standard deviation, and \(N_0\) of \(10^6\). ............................................................................................... 25

Figure 1.9: Plot of Poisson (solid) and Logistic (dashed) model fits to ‘TCP’ outcomes data for prostate cancer form the RT01 (Red), RMH (Blue), MD Anderson (Green) and Fox Chase Trials (Grey). ................................................................. 29
List of Figures

Figure 2.1: Example of two styles of audit conducted between a group of five centres. The left shows the ‘round-robin’ style audit cycle in which each centre conducts an audit and is audited. The right shows a single centre auditing the four other centres..........................39

Figure 3.1: Diagram of setup used for MV photon absolute dosimetry audits performed by NPL. The secondary standard ionisation chamber is positioned centrally in a 10x10cm field at a depth of 5, 7, 10 or 20cm in water......................................................43

Figure 3.2: Measured beam output ratios separated into the eight individual regional audit groups. Mean value for each beam type in each group is indicated by a horizontal line (−).........49

Figure 3.3: Geographical spread of the number of individual beams audited by NPL at each centre. Note the large number of beams audited at a single Scottish centre. The location of NPL is indicated by the black cross (x). .................................................................51

Figure 3.4: Plot of distribution of MV beam output audit results. The results are normally distributed with mean difference between host and NPL measurement of -0.06% and standard deviation of 0.68%. ..................................................................................................................52

Figure 3.5: Plot of MV output results in order of completion. The standard deviation of the previous 20 audits is shown by the solid line. Bars indicate the standard uncertainty as given in the NPL report of the audit produced for the department. Colour differentiates each regions result. ................................................................................................................53

Figure 3.6: Rolling standard deviation of NPLs MV audit results plotted against audit date. Rolling standard deviation calculated from previous 20 audits. A linear fit gives a decrease in standard deviation of 0.3% per decade with Spearman’s correlation coefficient of -0.881 (p<0.001) and $R^2 = 0.785$ indicating strong correlation.................................................................53

Figure 3.7: Plot of the destitution of the 142 MeV beam output audit results. All results were within ±3% with 6 outside of the ±2% action level usually used..............................................54

Figure 3.8: Plot of MeV output results in order of completion. The standard deviation of the previous 20 audits is shown by the dashed line. Bars indicate the standard uncertainty as given in the NPL report of the audit produced for the department. Colour differentiates each regions results. ................................................................................................................55

Figure 3.9: Boxplots of the measurement results of NPL MeV audits separated by nominal beam energy. No statistically significant difference was observed between the mean of results from each energy. Note this does not include the results from regional group 9. See Appendix D for description of boxplot display. The whiskers show the total extent of the data in this case........55
Figure 3.10: Histogram of the NPL MeV audit results with results from NHS (blue) and private (orange) centres. An abundance of private centre results is seen between -1 and -2%. NHS results are normally distributed with a mean of approximately 0%. .......................... 56

Figure 3.11: Box and whisker plots of the NPL MeV beam output results. Group 9 appears visually as an outlier and has the only mean and median outside of ±1% despite having a similar IQR to other centres. A statistical analysis also indicates a significant difference in the mean of the group 9 results (p<0.05). See Appendix D for description of boxplot display. The whiskers indicate the entire range of results in this case.................................................................................. 57

Figure 3.12: Plot of kV beam output results in order of completion. There was a total of 34 beams audited during the data period. No significant variations over time or between groups were observed...................................................... 60

Figure 3.13: Plots of distribution of number of kV results of beam output audits. All results were within ±3% and approximate a normal distribution even with a reasonably small sample........ 60

Figure 3.14: Plot of output difference against chamber intercomparison difference measured in NPL MV beam audits. The least squares fit excludes group 9 data. The Spearman’s rank correlation coefficient is 0.471 (p<0.01) with $R^2 = 0.22$ indicating weak correlation......................... 64

Figure 3.15: Plot of MeV beam results for measured output difference against intercomparison difference highlighted by regional group. Group 9 is visibly offset in comparison to the remaining results and has not been included in the linear regression. The fit has a Spearman’s rank correlation coefficient of 0.687 (p<0.01) and $R^2 = 0.338$ indicating a reasonably strong correlation........................................................................................................ 65

Figure 3.16: No correlation is observed with the kV results. There is greater uncertainty associated with these measurements than MV or MeV beams and there is also fewer results within the available dataset. $R^2 = 0.0038$, Spearman’s correlation coefficient = 0.095 (p=0.63). .................. 66

Figure 4.1: Locations of the 52 radiotherapy centres which supplied beam output measurement data for this work showing a wide geographic spread of data sources. ................................. 76

Figure 4.2: Distribution of measured machine output. Data has a mean of 0.0%, with a standard deviation of 0.8%. There are 75.7% of measurements within +/- 1%................................. 77

Figure 4.3: Probability plot (Q-Q plot) comparing the probability distribution of the output data collated and a theoretical normal distribution with mean and standard deviation as determined from the output data. The points lie along the line of unity indicating for the purpose of this work that the data can be assumed to be normally distributed. ........................................ 78
List of Figures

Figure 4.4: Typical set of beam output data showing weekly Farmer chamber and daily constancy measurements for a single linac. This dataset includes a period of upward trend from March 2015 to June 2015 followed by a calibration. ................................................................. 79

Figure 4.5: Distribution of treatment machine installation year for data collected. The oldest machine was installed in 1999 making it 16 years old at the time of data collection. A total of 204 treatment machines were included in the dataset; 96 Varian, 92 Elekta, 12 Siemens and 4 Accuray (3 Tomotherapy and 1 CyberKnife). ......................................................................................... 80

Figure 4.6: Recording methods in place for centres designated as large (left), medium (centre) and small (right) based on the number of treatment machines at a centre. The small centres have a lower proportion of paper only records within this dataset at this time. ................ 81

Figure 4.7: Variation in the number of treatment machines at each centre. Centres operated by private radiotherapy providers are generally smaller, with most having a single treatment machine. ................................................................................................................. 83

Figure 4.8: Plot showing the mean output from each treatment machine over the data collection period plotted in order of mean beam output. The machines from the centre with the greatest intra-centre variation are highlighted and have a range of 2.2% (range 0.7% to 1.5%). This centres machines showed no statistically significant difference between measurement devices used (p>0.05) and in the case of the highlighted machine two separate types of constancy device were in use. ................................................................................................................. 84

Figure 4.9: Plot showing the variation in mean dose (standard deviation) delivered by each machine for a range of time periods from 1 day to 6 months. It is seen that for typical treatment length of ~6 weeks the standard deviation is 0.68%. ................................................................. 85

Figure 4.10: Plot showing the range of the mean outputs recorded on different treatment machines within each centre. The greatest difference in the mean output between machines at a single centre was 2.2%. There were 4 centres which had a difference of greater than 2%. ...... 86

Figure 4.11: A selection of observations when performing linear regression on a Farmer chamber (orange triangles) and constancy device (blue circles) on four different machines. In a) the devices are very well matched with almost identical regression results b) indicates that the constancy device might have a different sensitivity to changes in beam output than the Farmer c) shows the Farmer chamber measuring consistently 0.5% higher than the constancy device d) shows a discrepancy of greater than 1% between Farmer and constancy device. ..................... 89

Figure 4.12: Difference between the mean of the Farmer chamber and constancy device measurements on each treatment machine. A positive result indicates the Farmer chamber measurements were greater. There were 65 of the 95 machines in which the Farmer chamber
had a mean of measurements greater than the constancy device. The mean difference was +0.23\% and this difference was statistically significant (p<0.05) to a mean of zero.

Figure 4.13: Plot of mean Farmer chamber measurements vs mean consistency device measurements on each treatment machine. A strong correlation is seen between Farmer chambers and constancy devices on each machine with $R^2$ of 0.73 and Spearman’s correlation coefficient of 0.86 (p<0.001).

Figure 4.14: Plot of standard deviations of beam output measurements on each treatment machine as measured by Farmer chamber or constancy device. There is only a very weak correlation between the results with $R^2$ of 0.14 indicating measurements made with different devices can be considered independent. Spearman’s correlation coefficient is 0.347 (p<0.001).

Figure 4.15: Plots of 10MV (left) and 12MeV (right) against 6MV results measured on the same day for one linac. The measurements have been separated into time periods determined by the point at which a calibration or equipment adjustment occurred. A banded structure appears which indicates the beam output of different energies on the same linac are correlated. It has been observed that the MV beams have greater correlation than the MeV beams.

Figure 4.16: Plot of mean monthly outputs (from four years of data) for LA3 and LA4. A clear seasonal variation is observed on both linacs.

Figure 4.17: Plot of the range of output drifts measured on each RSCH linac for each energy. As expected the newest linac (LA6) has the greatest output drift. See Appendix D for description of boxplot display. The whiskers on the plots indicate the total extent of the data and the green triangles represent the mean value for each linac in this case.

Figure 4.18: Plot of 6MV beam output and the 30 and 60 day rolling mean of these values. This simulates the mean dose received by a patient dependant on the start date and treatment duration.

Figure 4.19: Variation in standard deviation of rolling mean dose for different treatment durations. Variation exists dependant on start date due to drift in treatment machine output. Treatment courses of shorter duration inherently have a greater range of possible doses delivered to the patient.

Figure 5.1: Boxplots showing the distribution of the mean doses to the left (white) and right (grey) parotid glands separated by treatment type. The whiskers represent the range between the 5th and 95th percentiles. The IMRT plans show a statistically significant reduction in mean dose (p<0.01) when compared using a t-test. This is one of the published trial findings (Nutting et al., 2011). See Appendix D for description of boxplot display. The green triangles show the mean result. A summary of the doses is given in Table 5.1.
Figure 5.2: Plot showing the calculated standard deviation as a percentage of the 65Gy prescription dose to the primary target structure (primary PTV) for multiple dosimetric parameters. This is split into IMRT and conformal plans from the PARSPORT trial showing the difference in variation arising from different planning techniques. Each of the parameters showed a statistically significant reduction in the variance across the population (P<0.05 using the Levene test which compares the variances).

Figure 6.1: Diagram showing the features of determination of dose per fraction used within the modelling within this work. Included are a systematic offset of -1% at the start of the time period shown, a drift in beam output of +3%/year and random fluctuations around this which are drawn from a normal distribution with standard deviation of 0.2% of the nominal dose (2Gy/fraction).

Figure 6.2: Variation in the simulated TCP values at a nominal dose of 65Gy and 74Gy for head and neck and prostate cases respectively for varied population size used during optimisation of the standard deviation of α/β. The optimisation procedure to determine the value of alphabeta_sd_use in the model was repeated 20 times whilst varying the size of the simulated population (n). Using the returned optimised values, the population TCP was calculated for each to allow the variation to be assessed. The plot shows the standard error in the mean of these simulated population TCP values. The SEOM falls below 0.1% at approximately n=1000. This uncertainty was deemed acceptable for this work and so n=1000 was used for all further modelling work.

Figure 6.3: Variation in parameter sensitivity for the different parameters of the LQ model for typical prostate and head and neck cases. The change in TCP was calculated due to an increase in each parameter separately from -10% to +10%. In each case the delivered dose showed the greatest sensitivity to simulated outcome as a percentage of the mean parameter value and with N_0 had the smallest.

Figure 6.4: Plot showing the sensitivity of the calculated NTCP for the clinical cases detailed in Table 6.3 for each parameter. The parameters are the dose per fraction D50(1), m, and n as detailed in section 1.5.4. The prostate case is for Grade 2 rectal bleeding and the head and neck case is xerostomia.

Figure 7.1: Variation in calculated individual patient TCP (%) caused by a systematic shift in dose (%) throughout treatment. The shaded region indicates the most clinically relevant dose shifts which may occur. The shaded regions indicate the 95% CI for a convolution of the uncertainties in beam calibration (1.4%) alone and then combined with mean beam output (1.9% total) and
beam output fluctuations (2.0% total) (see section 7.1). The values for this plot between a dose shift of ±3% are given in Table 7.2 ................................................................. 127

Figure 7.2: Simulated population TCP change (%) based on a systematic dose shift (%) for the population. The shaded region indicates the most clinically relevant dose shifts which may occur. Fluctuations in the values arise from the random uncertainties in generation of the population of patients during the simulations (see section 6.1.4.1). The shaded regions indicate the 95% CI for a convolution of the uncertainties in beam calibration (1.4%) alone and then combined with mean beam output (1.9% total) and beam output fluctuations (2.0% total) (see section 7.1). The values for this plot for a dose shift of ±3% are given in Table 7.2 .................. 128

Figure 7.3: Plots showing the impact of a systematic dose shift on the modelled NTCP for the three clinical cases. The relationship is approximately linear over the clinical range. Variations of up to 10% in predicted NTCP are observed within the range corresponding to the shaded clinical uncertainties quantified within this work ................................................................. 130

Figure 7.4: Plot showing the modelled individual patient TCP for the three clinical cases. The solid line indicates the mean values and the shaded region represents the 5th to 95th percentiles around this. The variation modelled arises from the mean dose delivered based on the machine to which the patient is assigned. The range of individual patient TCPs is similar for each case as indicated by the arrows at the appropriate doses of interest. The machines modelled had a normal distribution of beam output with mean of 0% and standard deviation of 1.0% matching the clinical situation of combined calibration and beam output variation uncertainty............... 133

Figure 7.5: Plot showing the modelled population TCP for the three clinical cases. The solid line indicates the mean values and the shaded region represents the 5th to 95th percentiles around this. The variation modelled arises from the mean dose delivered based on the machine to which the patient is assigned. The range of population TCPs is varies for each case as indicated by the arrows at the appropriate doses of interest. The machines modelled had a normal distribution of beam output with mean of 0% and standard deviation of 1.0% matching the clinical situation of combined calibration and beam output variation uncertainty.............. 134

Figure 7.6: Boxplots indicating the relative variation in TCP for both prostate and head and neck cases due to the combined calibration and beam output uncertainty for individual patients and a simulated population. As a relative change from the nominal population TCP each of these cases is very similar with approximately 4% SD in TCP introduced by 1% SD in dose. The individual patients show a greater deviation ranging from 6.4% to 7.6% SD for the same variation in dose. See Appendix D for description of boxplot display ......................... 135
Figure 7.7: Plot showing the modelled NTCP for the assessed rectum and parotids clinical cases. The solid line indicates the mean value and the shaded region is the 5th to 95th percentile. The variation arises from a range of delivered doses arising from the calibration and beam output uncertainties (mean of 0% with a standard deviation of 1.0%). The rectal dose response curve is the greatest at the dose of interest and so has a greater range of NTCP values (indicated by the arrows).

Figure 7.8: Boxplots showing the variation in NTCP for the clinical cases at the end of treatment. This variation is caused by the combined 1.0% standard deviation of beam output and calibration. See Appendix D for description of boxplot display. The relative variation of the rectum case is greatest due to the steeper dose response as seen in Figure 7.7.

Figure 7.9: Plot showing the effect of beam output drift on predicted TCP on an individual. The shaded range indicates the 95% CI of drifts measured in this work. The Fox Chase case shows the greatest variation. A 5% annual increase in beam output gives 2% increase in TCP for the Fox Chase case and 1.5% for the RT01 and PARSPORT cases.

Figure 7.10: Plot showing the effect of beam output drift on predicted population TCP. A linear fit is shown for each to indicate the general trend. The shaded region indicates the 95% CI of drifts measured in this work. The larger the variation in \( \alpha/\beta \) values the more ‘noise’ within the simulations assessing dose drift, and so large numbers of patients would be required to clinically observe differences due to drift in beam output through the course of a treatment.

Figure 7.11: Plot showing the effect of beam output drift on NTCP. The shaded region indicates the 95% CI of drifts measured in this work. The impact on the parotid gland NTCP is less than 0.5% even for drifts of up to ±10%/year. For the rectum the variation in NTCP is larger, exceeding 0.5% for a drift of 5% per year.

Figure 7.12: Plot showing the standard deviation of individual patient TCP produced due to random variation in daily delivered dose as measured by the standard deviation of the beam outputs. Clinically these random fluctuations have a smaller impact than those caused by a systematic shift in delivered dose. A range of the plotted values are given in Table 7.5. The clinically determined standard deviation of daily delivered dose was 0.2% and the 95th percentile was 0.7% and these are indicated by the shaded regions.

Figure 7.13: Plot showing the standard deviation of population TCP produced due to random variation in daily delivered dose as measured by the standard deviation of the beam outputs. Clinically these random fluctuations have a smaller impact than those caused by a systematic shift in delivered dose. A range of the plotted values are given in Table 7.5. The clinically
determined standard deviation of daily delivered dose was 0.2% and the 95th percentile was 0.7% and these are indicated by the shaded regions. ................................................................. 142

Figure 7.14: Plot showing variation of NTCP produced (standard deviation) due to random daily dose variations with different uncertainties (standard deviation). Within the clinical range these variations introduced are relatively small. ................................................................. 144

Figure 7.16: Plot showing the predicted difference for NPL absolute dosimetry audit data. The proportion of beams exhibiting a given change in clinical outcome is shown on the y axis for change in TCP (blue), mild NTCP (orange) and severe NTCP (green). ................................................................. 148

Figure A.D.1: Schematic of the boxplot representation used within this thesis unless otherwise specified within the text. The box is bounded by the 25th and 75th percentiles (forming the interquartile range (IQR)). The whiskers represent the 5th to 95th percentile range. This particular example shows the boxplot for the results of 1000 samples drawn randomly from a normal distribution with mean value of zero and standard deviation of one. ................................................................. 181
List of Tables

Table 1.1: Example simple daily run up schedule for Linac and associated tolerances based upon local practice at the RSCH and IPEM 81 guidance. Note this is not an exhaustive list of the tests which might be performed, but an indication only. ................................................................. 9

Table 1.2: Summary of trials from which published data was used to perform initial model fitting to clinical data. The data extracted from the publications is given in Table 1.3 and the model fit results are presented in Figure 1.9. ................................................................. 28

Table 1.3: Summary of data used when fitting Poisson and Logistic models shown in Figure 1.9. Further details of each study are detailed within the text and Table 1.2. ................................................................. 29

Table 1.4: Extracted $\gamma$-values from the fitting of Poisson and Logistic models to the clinical prostate outcome data described in Table 1.3 and plotted in Figure 1.9. The largest difference in the gradient of the dose response curve at 74Gy was 1.67%. ................................................................. 30

Table 1.5: List of Python packages used heavily within this work with brief description of usage. Those marked with an asterisk (*) are part of the Scipy stack. The version numbers are correct at the end of the work and all code used within the thesis used these versions. ................................................................. 34

Table 2.1: Summary of MV data received from regional audit groups (Regions A-I have been anonymised to groups 1-9). Note: no data was received from groups 2, 6, 8 or 9. The table summarises the mean difference in output as measured by the visitor relative to the host centres. ................................................................. 37

Table 3.1: Summary of NPL audit visits during the period 1994 to February 2015. To allow comparison, data from the two MeV CoPs has been included both separately and combined. ................................................................. 48

Table 3.2: Summary of all absolute beam output measurement results from NPL audits carried out between 1994 and February 2015. The two MeV CoPs are separated for comparison. ................................................................. 50

Table 3.3: Summary of all absolute beam output measurement results from NPL audits carried out between 1994 and February 2015 for groups 1–8 and group 9 separately. There is a statistically significant difference ($p<0.05$) between the mean difference for MeV measurements as measured by group 9 when compared to other groups. The results compared are highlighted with an asterisk (*). ................................................................. 58

Table 3.4: Summary of data used in comparison of 1996 and 2003 MeV CoPs. The data from group 9 was excluded from this analysis to avoid skewing results with potentially unreliable data. A statistically significant change in the mean difference of 0.55% as measured with each CoP was observed ($p < 0.01$) using Welch’s t-test. ................................................................. 59
Table 3.5: Summary of NPL beam quality measurements as measured during audit visits. MV are shown as percentage difference of TPR, MeV are shown as difference in the depth of $R_{50,0}$ and kV are shown as percentage difference of HVL as a range of energies is included..........................61

Table 3.6: Summary of NPL ionisation chamber intercomparison measurement results from audit visits. ..........................................................62

Table 3.7: Summary of beam measurements taken during the national rotational IMRT audit. The standard deviation of results more than doubles when measuring with an ionisation chamber within a phantom when compared to reference conditions. There is no statistical difference between the corrected and uncorrected phantom measurements (p>>0.05) .....................71

Table 4.1: Summary of the measured drift of beam output over time for each manufacturer after correction for calibrations as determined using linear regression. There was no statistically significant difference observed between manufacturers at the 95% level (p>0.05). .................82

Table 4.2: Summary of data set for treatment machines containing both Farmer and consistency device measurements grouped by device type. There were three other reported devices: Sun Nuclear DoseCheck, Victoreen DoubleCheck and diodes which were all used on a single machine.................................................................................88

Table 4.3: Summary of measurements variation of devices which were used on more than 10 machines. There was no statistically significant difference observed between the consistency of the measurements taken with these devices (p>0.05)...................................................93

Table 4.4: Summary of linacs used to analyse long term trends and correlation between beam energies..................................................................................94

Table 5.1: Summary of doses delivered to parotid glands for patients within the PARSSPORT trial. Both left and right sides were very similar. IMRT plans resulted in a significant reduction in the doses received (p<0.01).........................................................................................94

Table 5.2: Standard deviation of dosimetric parameters for the primary PTV from the patients treated with conformal and IMRT plans. These results are plotted in Figure 5.2 Each of these statistics was found to be statistically different for the high dose target volume (p < 0.05) for the conformal and IMRT planning techniques.................................................................107

Table 6.1: Parameters used during TCP simulation to determine the influence of the population size on the variation in simulated TCP results caused by the optimisation of the parameter describing the standard deviation of $\alpha/\beta$. The population size (n) was varied between 5 to 5000. ........................................................................................................................................115

Table 6.2: Parameters used for modelling prostate and head and neck patients within this work using the LQ TCP model. Based on parameters published within the literature the values of the
standard deviation of the α/β value were determined through optimisation using the model. The N₀ values used for each set of simulations are based on the range of values used within the literature and were optimised to fit the clinical data.

Table 6.3: Parameters used for modelling prostate and head and neck patients within this work using the LKB NTCP model. The parameters are based upon the published literature listed and represent a range of different normal tissue complications observed clinically.

Table 7.1: Summary of dose uncertainties as assessed in this thesis. The magnitude of these variations will be used to assess the impact of dose uncertainties on patient outcomes due to beam calibration and output variation. Convolution of parameters is indicated by the ‘∗’ symbol.

Table 7.2: Calculated TCP change based on a systematic dose shift throughout the entire treatment for an individual and population of patients. The 95th percentile is ±1.9% for combined calibration and beam output maintenance uncertainties which gives rise to variation in TCP of over 10% for an individual patient and over 6% for the population TCP changes. These values are plotted in Figure 7.1 and Figure 7.2.

Table 7.3: Calculated NTCP change based on a systematic dose shift throughout the entire treatment. The 95th percentile is ±1.9% for combined calibration and beam output maintenance uncertainties. This gives rise to differences of 8% for the rectum case and up to 3% for the parotid cases. These values are plotted in Figure 7.3.

Table 7.4: The standard deviation of TCP results produced for a range of beam calibration uncertainties. The initial uncertainty in the MV beam calibration as determined through the NPL audits was 0.8%. It is observed that the results are not linear over this range. Note: values in the table are rounded to 1.d.p.

Table 7.5: Quantified values of the variation in standard deviation of the TCP due to daily beam output fluctuations arising from the simulations. Values are plotted in Figure 7.12 and Figure 7.13. The clinically determined standard deviation of daily delivered dose was 0.2% and the 95th percentile was 0.7%.

Table 7.6: Values of the variation in standard deviation of the NTCP introduced by variations in daily delivered dose (quantified by the standard deviation). The standard deviation of the clinical machines analysed in this work was 0.2% and these are all less than 0.1%.

Table 7.7: Modeled number of patients with predicted TCP of less than 10% and greater than 90% for a fixed systematic dose shift to the population. Included are the cases for prostate and head and neck cancers which have a total number of patients modelled of 26318 and 8093 respectively.
# Common Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>bNED</td>
<td>Biochemically no evidence of disease</td>
</tr>
<tr>
<td>bPFS</td>
<td>Biochemical progression free survival</td>
</tr>
<tr>
<td>cDVH</td>
<td>Cumulative Dose Volume Histogram</td>
</tr>
<tr>
<td>CoP</td>
<td>Code of Practice</td>
</tr>
<tr>
<td>D99</td>
<td>The minimum dose received by 99% of the volume of interest.</td>
</tr>
<tr>
<td>dDVH</td>
<td>Differential Dose Volume Histogram</td>
</tr>
<tr>
<td>DICOM</td>
<td>Digital Imaging and Communications in Medicine</td>
</tr>
<tr>
<td>DVH</td>
<td>Dose Volume Histogram (Commonly when referring to simply a ‘DVH’ it is the cumulative DVH (cDVH) to which this refers).</td>
</tr>
<tr>
<td>EORTC</td>
<td>European Organization for Research and Treatment and Cancer</td>
</tr>
<tr>
<td>ESTRO</td>
<td>European Society for Radiotherapy and Oncology</td>
</tr>
<tr>
<td>FFF</td>
<td>Flattening Filter Free</td>
</tr>
<tr>
<td>Gy</td>
<td>Gray (Derived unit for absorbed radiation dose, 1Gy = 1J/kg)</td>
</tr>
<tr>
<td>HVL</td>
<td>Half Value Layer</td>
</tr>
<tr>
<td>IAEA</td>
<td>International Atomic Energy Agency</td>
</tr>
<tr>
<td>IMRT</td>
<td>Intensity Modulated Radiotherapy</td>
</tr>
<tr>
<td>IPEM</td>
<td>Institute of Physics and Engineering in Medicine</td>
</tr>
<tr>
<td>kV</td>
<td>Kilovoltage photon</td>
</tr>
<tr>
<td>Linac</td>
<td>Linear Accelerator</td>
</tr>
<tr>
<td>MeV</td>
<td>Megavoltage electron</td>
</tr>
<tr>
<td>MLC</td>
<td>Multi Leaf Collimator</td>
</tr>
<tr>
<td>MU</td>
<td>Monitor Unit</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>MV</td>
<td>Megavoltage photon</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service</td>
</tr>
<tr>
<td>NPL</td>
<td>National Physical Laboratory</td>
</tr>
<tr>
<td>NTCP</td>
<td>Normal Tissue Control Probability</td>
</tr>
<tr>
<td>OAR</td>
<td>Organ at Risk</td>
</tr>
<tr>
<td>OSLD</td>
<td>Optically Stimulated Luminescent Dosimeter</td>
</tr>
<tr>
<td>PMMA</td>
<td>Poly(methyl methacrylate). A commonly used plastic.</td>
</tr>
<tr>
<td>PTV</td>
<td>Planning Target Volume</td>
</tr>
<tr>
<td>QA</td>
<td>Quality Assurance</td>
</tr>
<tr>
<td>QC</td>
<td>Quality Control</td>
</tr>
<tr>
<td>RSCH</td>
<td>Royal Surrey County Hospital</td>
</tr>
<tr>
<td>RTTQA</td>
<td>Radiotherapy Trials Quality Assurance</td>
</tr>
<tr>
<td>TCP</td>
<td>Tumour Control Probability</td>
</tr>
<tr>
<td>TLD</td>
<td>Thermo-luminescent Dosimeter</td>
</tr>
<tr>
<td>TPS</td>
<td>Treatment Planning System</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>V100</td>
<td>The volume of the structure of interest receiving at least 100% of the prescribed dose.</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Overview

In the UK around 1 in 2 people will develop cancer within their lifetime (Ahmad, Ormiston-Smith and Sasieni, 2015). Radiotherapy may be beneficial to over half of those diagnosed with cancer (Gilbert et al., 2009; Borras et al., 2015) within the UK and is a common cancer treatment behind only surgery in terms of its effectiveness (Ahmad, Ormiston-Smith and Sasieni, 2015). Radiotherapy involves the precise delivery of radiation to the treatment site (further described in section 1.2); and as such there are numerous standards which must be met to ensure patient safety and treatment effectiveness. Each treatment centre will have a quality assurance (QA) program in place which includes quality control (QC) checks on each treatment machine ensuring the treatment is accurate and precise.

As well as locally performed QC, independent measurements are undertaken by external parties which may be from a neighbouring department, a national institution (e.g. the National Physical Laboratory (NPL) in the UK) or a clinical trials group such as the National Cancer Research Institute Radiotherapy Trials Quality Assurance Group (RTTQA).

In the UK, measurements of dose are traceable to a primary standard held by NPL and a traceable measurement taken within a radiotherapy centre is one which can be followed back through a chain of documented calibrations to a primary standard. This helps ensure all radiotherapy doses delivered are consistent between centres.

The calibration of absolute dose within the UK is implemented following published Codes of Practice (CoPs) (Lillicrap et al., 1990; Burns et al., 1996; Klevenhagen et al., 1996; Thwaites et al., 2003; Aukett et al., 2005) which detail the technical setup requirements, how to calculate the delivered dose, and allows users to calibrate their treatment machines consistently. The UK has a rich history of audit within the radiotherapy community (Clark et al., 2015) and the NPL have been checking the absolute calibration of treatment machines for over 20 years providing the ‘Gold Standard’ of dosimetry audits (Palmer et al., 2011). Within the context of radiotherapy an audit generally refers to an independent measurement of one or more aspect of the radiation delivery. These NPL audits
concern fundamental dosimetry parameters such as the absolute dose delivered (beam output) and 
the beam quality (measure of beam energy).

The work within this thesis will primarily be based on UK audits and data, however similar 
procedures exist across the world (Andreo et al., 2000; Veres, 2009; Koniarova et al., 2014; Izewska, 
Lechner and Wesolowska, 2018) and so the findings are not limited to UK radiotherapy, but are 
widely applicable.

The total dose variation as measured locally and on a national scale is to be assessed. This will 
include quantification of any variation over time which may exist and determination of treatment 
machine specific results. In general an increase in dose delivered to the target (cancer cells) will 
increase the probability of eradicating the disease, however higher doses may be to the detriment of 
surrounding healthy tissues. Different cancers respond differently with some being much more 
sensitive to dose variation than others (Selvaraj et al., 2013).

Using radiobiological modelling it is possible to estimate the variation in predicted clinical outcomes 
from a variation in delivered dose. These variations in dose can arise through different sources 
including: the implementation of the beam calibration, maintenance of that calibration and patient 
positioning. An evidence based prediction of the impact of QC tests allows tolerances to be set 
appropriately depending on their purpose. For example, cancers with greater radiation sensitivity 
may warrant tighter tolerances, and similarly a clinical trial may wish to impose strict tolerances to 
enhance the trial’s statistical power by reducing the uncertainty.

The impact of uncertainty in dose associated with routine QC measurements of beam output has not 
systematically been applied to a range of cancers, and this work attempts to begin this process. This 
will give a greater understanding of the impact of QC results on different treatments.

1.1.1 Objectives of this work

The primary objectives of this work are as follows:

1. Quantify the variation in dose as measured during absolute dosimetry audits conducted by 
   NPL for all beam types: Megavoltage photon (MV), Megavoltage electron (MeV), and 
   kilovoltage photon (kV).
2. Conduct an analysis of routine measurements of beam output from UK centres and assess 
   the variation between them.
3. Apply radiobiological modelling methods to quantify the impact of dose variations on the change in TCP and NTCP for selected cancers.

Additional secondary objectives are:

4. Development of an automated tool to extract dosimetric statistics from radiotherapy treatment plans in DICOM format.
5. Develop a flexible implementation of radiobiological models for TCP and NTCP which allow the assessment of the variation of dose on a fraction by fraction basis.

1.1.2 Novelty of work

The novelty of this work lies in the application of radiobiological models to the evidence based dose variation as gathered through audits and routine QC. To gather this evidence base an analysis of beam output measurements made across the UK has been carried out and is the first of its kind on this scale. This data combined with audit data from NPL allows an accurate depiction of dose variation due to the calibration of treatment machines to be constructed and then applied to the established radiobiological models in a previously unexplored way.

Radiobiological modelling is now beginning to be included within the treatment planning system (TPS) to allow comparison of plans, however the variation in dose due to machine calibration is seldom considered and this work aims to highlight this area of dose uncertainty which departments may be able to control to a greater degree of accuracy than they currently do. With the advances in technology over time, it is perhaps worth re-evaluating the control of different parameters within the QC frameworks employed.

1.1.3 Overview of thesis

This thesis begins with a brief introduction to radiotherapy followed by a broad overview of QC involved in maintaining the delivered dose to within known limits in the remainder of chapter 1. Dosimetric audit is a vital part of this QA framework which ensures safety and consistency and its development is discussed in section 1.4. To assess the impact of dose variations, radiobiological modelling was used and the background to this and development of the models is presented in section 1.5 followed by brief details on its use in this work in section 1.6.

Collated audit data is analysed and comment given in chapters 2 and 3 which build upon the publication arising from this work for which the abstract is presented in Appendix A.4. These
chapters will quantify the dosimetric uncertainty due to calibration and addresses **objective 1** as the analysis is based upon data derived from dosimetric audits.

An additional source of variation in delivered dose arising from maintenance of the beam calibration (referred to as **beam output** variation) is explored in **chapter 4** which further builds upon published work (see Appendix A.5) and addressed **objective 2** through this large scale analysis.

Using the information on the dose variation form these chapters this is put into the wider context and compared with doses arising from treatment planning in **chapter 5** which details the development of a tool for **automated extraction of dosimetric parameters** from DICOM radiotherapy plan files. This addresses **objective 4** through the development of an automated tool for extraction of dosimetric parameters.

Following this **chapter 6** details the **implementation of radiobiological models** which allow fine control over the fraction by fraction dose variation which covers **objective 5** and begins the work to meet **objective 3**.

**Chapter 7** then uses the gathered evidence of dosimetric deviations and assessed the **impact of these for prostate and head and neck cancers and their clinical outcomes**. This is the work which completes the required work for **objective 3**. This is then discussed and concluded in **chapter 8** and the need for further work is addressed.

A complete **list of publications** and presentations arising from this work is given in Appendix A. The code used for modelling and analysis is given in Appendix B and the availability of source data is given in Appendix C.

### 1.2 Radiotherapy

Radiotherapy is a cost effective treatment and plays a vital role in around 50% of cancer cases (Gilbert *et al.*, 2009) with approximately 40% of patients who are cured of cancer receiving radiotherapy (NHS England, 2016).
Introduction

Figure 1.1: Varian TrueBeam Linac located at the Royal Surrey County Hospital. As well as generating a radiation beam which emerges from the linac head for treatment it also has kV and MV imaging capabilities to aid patient positioning.

Radiotherapy aims to deliver the correct dose of radiation to the tumour while minimising damage to surrounding healthy tissues. The most common treatment method involves the use of a linear accelerator (linac) which can rotate 360 degrees around the patient to deliver dose precisely to the target. A typical modern linac configuration is shown in Figure 1.1 which also shows the additional imaging capabilities which are now commonplace.

Radiotherapy covers a wide range of treatment techniques including ‘external beam radiotherapy’ (EBRT) in which a radiation beam is generated externally to the patient and then these beams are directed towards the target within the patient, brachytherapy in which a source of radiation is placed within the body, and radionuclide therapy in which a radioactive substance may be ingested or injected into the body. This work considers only external beam radiotherapy as this is the most commonly used method of treatment.

1.2.1 Radiotherapy treatment techniques

There are a number of treatment techniques used within EBRT to create the final treatment plan. Originally only rectangular radiation fields could be used with beams coming from fixed directions. This resulted in an area of high dose which was cuboidal in shape (rectangular when viewing a 2D
axial slice). Later additional shaping of the high dose region was possible with the introduction of beam shaping devices such as the multi-leaf collimator (MLC), which allows creation of irregular field shapes and thus dose could be conformed around the target to a greater degree. Following from this was Intensity Modulated Radiotherapy (IMRT) which introduced a modulation of beam intensity across the field and thus allowed still greater conformity of the high dose regions around the target volume with concave dose distributions possible. The current state of the art is Volumetric Modulated Arc Radiotherapy (VMAT) which, as well as allowing modulation of the beam, continually rotates the gantry around the patient and may vary in speed. This speeds up treatments and often achieves the greatest conformity as it is similar to increasing the number of fixed IMRT fields used.

An example of the change in conformity from using 3D conformal techniques to IMRT is shown in Figure 1.2 which demonstrates the conformation of dose possible around organs at risk using IMRT techniques which often are not possible with 3DCRT techniques.

![Image of dose distributions](image_url)

**Figure 1.2:** Comparison of dose distributions achievable with 3D conformal techniques (left) and IMRT techniques (right). Conformation around the right parotid is achieved with IMRT, however it is not possible to the same extent with 3DCRT techniques increasing the probability of side effects.

Traditionally, and for most current treatments the aim of radiotherapy is to completely cover the target volume homogeneously with 100% of the prescribed dose. A change in the magnitude of dose delivered by the machine will systematically affect the dose delivered to all patients on that machine so it is likely that the effect on all these patients will be similar; increase in dose will lead to

---

1 Alternative techniques including brachytherapy, SABR and SRS treatments involve the use of high dose gradients to maximise the dose to the target and follow a variety of prescriptions and systems of dosimetry.
an increased probability of cure, but greater risk of complications. A number of factors contribute to this overall dose variation (Thwaites, 2013; van der Merwe et al., 2016). Treatment techniques have advanced, for example the use of IMRT is now widespread, and automated planning techniques are becoming more common and planning is becoming more consistent (Hussein et al., 2016; Zhang et al., 2017) and so this should reduce the dose variation to patients arising from the treatment planning process.

### 1.2.2 Uncertainties within radiotherapy treatments

Uncertainties can be split into two main categories: equipment and patient. The patient uncertainties include those involved with the setup and positioning of the patient as well as variations in anatomy. The equipment uncertainties may be divided into dosimetric uncertainties which are to do with the beam generation and its accuracy, and geometric uncertainties which include variation in the radiation field size, shape and position. This work focuses on the impact of dosimetric uncertainty on clinical outcomes and will assume geometric and setup uncertainties remain constant. In particular it also assumes the radiation field characteristics remain constant apart from the magnitude of the delivered dose. To ensure that doses remain tightly controlled extensive QA programs are in place in each radiotherapy department.

As radiotherapy progresses, QC testing and associated tolerances require periodic review. It is accepted that uncertainties in treatment exist, however these should be minimised as far as reasonably practical. The International Atomic Energy Agency (IAEA) published guidance on the accuracy requirements of radiotherapy in 2016 (van der Merwe et al., 2016) indicating that whilst radiotherapy has been in use for many years, requirements evolve with the techniques in use.

One of the IAEA stated recommendations is that “radiotherapy should be applied as accurately as reasonably achievable, technical and biological factors being taken into account”. This is an important point as it indicates that different clinical situations should be considered separately. While site specific tolerances are in place regarding patient positioning and imaging, this is rarely the case for variations in beam output. It is these variations in beam output as applied to specific sites which is addressed in this thesis.

### 1.3 Quality assurance in radiotherapy

QA encompasses the management of QC checks and their documentation including protocols and their appropriate review. Here a broad overview of the QC undertaken on Linacs is given. Further detail is given in reports such as IPEM 81 (Mayles et al., 1999) which gives a more complete
Introduction

coverage of all tests which may be performed. Within the UK IPEM 81 remains one of the key
sources of information on QC testing as demonstrated in a survey conducted by Palmer et al.
(Palmer, Kearton and Hayman, 2012), despite at the time being over 10 years old. Many of the
tolerances assigned to QC tests were initially derived from work which relied on clinical observations
of visible side effects such as skin erythema (Brahme, 1984; Brahme et al., 1988; Mijnheer, 1996).

An overall treatment accuracy of ±3% (one standard deviation) is recommended in IPEM 81 and thus
tolerances for individual aspects are generally tighter than this. All uncertainties within the
treatment chain must be identified and combined (in quadrature) to give an overall uncertainty in
delivered dose. The soon to be published IPEM 81 update (Patel et al., 2018) will recommend the
use of an ‘action level’ and ‘suspension level’ to describe the cases where there are two tolerance
levels used (Patel et al., 2017). If a result is outside an action level then treatment may continue but
a plan should be put in place to investigate and correct the issues identified. If the QC results are
outside a suspension level then no further treatments should take place on that treatment machine
until the issue is resolved. However in the case of audits generally only a single level is used, and this
would equate to an action level as a result outside of this level would be unlikely to cause immediate
suspension of treatments, but would be investigated. Where known the appropriate type of level
will be referred to, but if unknown it is assumed that the action level is given.

Fundamentally all the QC tests will either be measuring the stability of the beam generation and
control, or a mechanical component such as the gantry, jaw and MLC positions. Tests vary in
frequency ranging from annual checks, such as measuring absolute dosimetry under reference
conditions, to daily checks of output constancy and light field size.

A small part of the daily machine run-up procedure might include the checks shown in Table 1.1
prior to clinical use. This is based on checks recommended in the IPEM 81 guidance (Mayles et al.,
1999) and local practice at the Royal Surrey County Hospital (RSCH).
Introduction

<table>
<thead>
<tr>
<th>Quality control test</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output Constancy</td>
<td>±3%</td>
</tr>
<tr>
<td>Light field size</td>
<td>±2mm</td>
</tr>
<tr>
<td>Optical distance indicator</td>
<td>±2mm</td>
</tr>
<tr>
<td>Machine interlocks</td>
<td>Functional</td>
</tr>
<tr>
<td>[Maze entrance, operator controls]</td>
<td></td>
</tr>
<tr>
<td>MLC positioning [static pattern check]</td>
<td>±1mm</td>
</tr>
</tbody>
</table>

Table 1.1: Example simple daily run up schedule for Linac and associated tolerances based upon local practice at the RSCH and IPEM 81 guidance. Note this is not an exhaustive list of the tests which might be performed, but an indication only.

This work focusses on the impact of beam output on clinical outcomes and so will describe the general process of measuring beam output in both absolute and relative terms. Controlling beam output is fundamental to the success of radiotherapy. Errors may have dramatic effects on the patient, and may result in serious harm or death. There have been a number of notable incidents, the response to which have helped to improve the safety of radiotherapy treatments (Porter, 2013). Reporting of incidents within radiotherapy is strongly encouraged and enhances the safety of services provided (The Royal College of Radiologists et al., 2008). An unnoticed change in beam output has the potential to adversely affect a large number of patients, either by overdosing and causing irreparable damage, or by under-dosing and failing to control the disease. To this end output measurements are undertaken daily for all clinical beams in almost all centres (Palmer, Kearton and Hayman, 2012) after the initial absolute dose calibration.

1.3.1 Beam output calibration

Within the UK all measurements of beam output, and thus all doses delivered to patients are traceable to the primary standard held at NPL (DuSautoy, 1996). This in turn is inter-compared with others around the world (Allisy, Burns and Andreo, 2009) ensuring worldwide consistency. In the UK the primary standard for MV photon radiotherapy is a graphite calorimeter designed and operated by the NPL. It is from this all other devices are compared. For determining absolute dose on a linac the UK has CoPs for each beam type. The CoPs are a result of collaboration between end users in the clinical setting and dosimetry experts at NPL. Within the UK for MV photon beams the same CoP has

---

2 During the period of this work one UK centre has calibrated using the international CoP, TRS398 (Andreo et al., 2000). No data for this centre was available for this work.
been in use since it was published in 1990 (Lillicrap et al., 1990) and this describes the process of determining the absorbed dose to water in an MV beam. For MeV beams an air kerma based CoP was published in 1996 (Burns et al., 1996) and then updated in 2003 (Thwaites et al., 2003). The 2003 MeV CoP introduced an absorbed dose to water based calibration which reduced the uncertainty of the previous air kerma based calibrations described in the 1996 MeV CoP (Rosser, Lawrence and Sinclair, 2009). The effects of these changes are explored further in chapter 2.

Treatment machines are usually calibrated against a tertiary dosimeter which has been calibrated against a secondary standard ionisation chamber (the NPL designed NE 2561 ionisation chambers ³) which itself has been calibrated at the NPL through comparison with the graphite calorimeter primary standard (DuSautoy, 1996). Once calibrated, the accuracy of delivered dose is assessed through independent audit from another centre, the Radiotherapy Trials Quality Assurance team (RTTQA), or from NPL. Reference dose audit acts to verify that treatment machines have been calibrated correctly using the relevant UK CoP.

1.3.2 Maintenance of beam output

Once initially calibrated the beam output of a treatment machine must be monitored to ensure consistency. An absolute measurement following the CoP is time consuming and is not practical to perform each day and thus alternative measurement procedures are in place as recommended in IPEM 81 (Mayles et al., 1999). There is a wide variety of measurement techniques used through the different centres and a range of equipment. Palmer et al. explored the frequency of measurements of beam output and almost all centres performed a daily check of beam output (Palmer, Kearton and Hayman, 2012). Often a constancy device which is quick to setup will be used on a daily basis and a measurement with a calibrated ionisation chamber will be performed on a less frequent basis (weekly or monthly). This practice of monitoring beam output is generally not scrutinised in the same way as the absolute beam calibration and is one aspect which will be further explored in this work. Tolerances used for these beam output measurements can also vary between centres, with tolerances for daily measurements using a consistency device ranging between ±1% to ±5% (Palmer, Kearton and Hayman, 2012). Recommended tolerances for daily output constancy and monthly calibration checks are typically ±5% and ±2% respectively (Mayles et al., 1999; Dieterich et al., 2011; IAEA, 2016).

³ The original NE 2561 was designed by NPL and manufactured by Nuclear Enterprises. That design was superseded by the type NE 2611, however manufacture and repair of the NE 2611 chamber type was taken over by NPL when Nuclear Enterprises stopped production. New secondary standard chambers are designated as being of type NPL 2611. All versions are radiologically equivalent.
The beam output is monitored and if it gets towards, or passes the action or suspension level on the measurement the beam will then be recalibrated to bring the absolute delivered dose back to that of the original calibration. This change in calibration is independently checked to ensure safety.

1.4 Audit in radiotherapy

The term audit within radiotherapy has a wide ranging meaning from review of treatment outcomes to measurement of reference dose. However in the context of physics, most commonly refers to dosimetry audit. The complexity and scope of these audits can vary widely. Early audits were solely based on absolute dosimetry and measurements of beam output under reference conditions. Many recent audits examine the entire treatment chain using an end-to-end audit approach (Budgell et al., 2011; Clark et al., 2014, 2015; Ibbott and Thwaites, 2015; Izewska, Lechner and Wesolowska, 2018), however nearly all of these still contain a fundamental measurement of the beam output as this is often desirable to help determine the source of any discrepancies.

Dosimetry audits can be classified into three levels as described by Kron et al. (Kron et al., 2002) which are:

- **Level 1:** Measurement under reference conditions in a regular phantom such as a Perspex block.
- **Level 2:** Measurements including variation in the configuration of dose delivery such as IMRT within a regular phantom.
- **Level 3:** Measurement in an anthropomorphic phantom following as closely as possible the same pathway as a patient (often referred to as an end-to-end audit).

This work will predominantly focus on level 1 audit. Most level 2 and 3 audits would also include a separate level 1 measurement which might be used to apply correction factors for beam output.

1.4.1 Purpose of dosimetry audit

Within radiotherapy there is a strong emphasis on safety and audit plays an important role in this. Numerous reports state that audit is required (Izewska and Andreo, 2000; James et al., 2008; The Royal College of Radiologists et al., 2008; Palmer et al., 2011; Clark et al., 2015; Izewska, Lechner and Wesolowska, 2018). Whilst audit is a general term and may be used as part of a QA framework to ensure continual improvement, within the radiotherapy context this often refers specifically to dosimetry audit. A dosimetry audit involves an independent check of the dose delivered, which may
Introduction

involve fundamental measurements made with an ionisation chamber under reference conditions such as those by Nisbet and Thwaites (Nisbet and Thwaites, 1997) and Thwaites et al. (Thwaites et al., 1992) or a more complex validation of a more complete dose distribution such as those in IMRT by Budgell et al. (Budgell et al., 2011) and the recent VMAT audit by Clark et al. (Clark et al., 2014). Independent external dosimetry audit ensures consistency across the UK, thus ensuring safety and allowing large scale clinical trials with comparable treatments at each centre. There are many methods used for dosimetric audit, with different measurement methods including TLDs, OSLDs, either involving a site visit or a postal audit system (Lye et al., 2014; Hurkmans et al., 2016; Alvarez et al., 2017), and delivered by local or international audit groups.

1.4.2 Brief history of dosimetric audit

Early audits were conducted by the IAEA and WHO utilising a postal service using TLDs in 1969 (Izewska and Andreo, 2000; Izewska, Lechner and Wesolowska, 2018). With a postal audit system the IAEA will send a set of TLDs to a host centre. The host centre will then follow the designated protocol to deliver a known dose to the TLDs. The TLDs are then returned for analysis and the measured dose can be compared with that expected. This system is still in place and results are published periodically (Izewska et al., 2015). The benefit of audit is demonstrated through these audits with centres receiving a second audit achieving 85% of results within the ±5% tolerance compared with 77% of those at their first audit.

Also in 1969 the Radiological Physics Centre (RPC) in Houston (now known as IROC-H (Imaging and Radiation Oncology Core – Houston)) began performing audits across the USA with a TLD service beginning in 1977. In Europe the ESTRO (European Society for Radiotherapy and Oncology) developed its own audit service in the early 90’s setting up the EQUAL (Quality Assurance Network for Radiotherapy).

Since then audit has become more embedded in the radiotherapy process due to its importance in ensuring the accuracy and safety required in radiotherapy. Audit has developed and now includes a much broader spectrum of audits and measurements. Audits now often take into account the entire treatment process with end-to-end audits, and new and developing techniques are routinely audited to aid implementation. These include complex techniques such as IMRT (Clark et al., 2009), VMAT (Clark et al., 2014) and 4DCT (Palmer et al., 2017) as well as auditing the individual techniques used to measure the accuracy of dose distributions (Hussein et al., 2017). Clinical trials in radiotherapy will now all have QC associated with them, verifying contouring, plan generation and dose delivery.
1.4.3 Dosimetry audit within the UK

For a comprehensive review of UK radiotherapy audit the reader is directed towards the recent comprehensive review conducted by Clark et al. (Clark et al., 2015) which examines the progression of UK dosimetry audit over the past 30 years. Some key aspects will be discussed which are specific to the UK audit community.

Due to the size of the UK it is not unreasonable to perform audits on a national scale. In the 1980s the first national photon audit was completed by Thwaites et al (Thwaites et al., 1992). All 64 radiotherapy centres in operation at the time were involved and reference dosimetry measurements were taken for all MV photon beams which included 61 Co-60 and 100 Linac beams. As well as the reference dosimetry audit, doses were compared with two planned treatment fields, in a uniform water equivalent phantom and one which included a lung insert, thus introducing an inhomogeneity. This audit demonstrated its merit by identification of a major calibration error on a Co-60 unit which resulted in patients receiving 25% over doses (Exeter Health Authority, 1998). The reference measurements had a mean difference (standard deviation) of +0.2% (± 1.4%) and +0.3% (±1.5%) for Co-60 and MV beams respectively. An important point stated in this paper refers to the accuracy of dose delivery:

“The accuracy requirements on each part of the whole process must be significantly less than the overall recommendations to achieve the final values required”

Thwaites et al, 1992 (Thwaites et al., 1992)

Recommendations by IPEM state a ±3% requirement in overall dose accuracy, and the above variation in dose calibration then forms around half of the overall allowable uncertainty. Additionally, there will be daily variation in beam output following initial calibration as well as uncertainties in treatment planning and patient setup.

In 1996 an MeV audit was performed by Nisbet and Thwaites (Nisbet and Thwaites, 1997) which included all UK centres with MeV beam facilities. This and the previously discussed MV audit by Thwaites provided a baseline dataset from which to compare future audits as well as for comparison with regional groups. This particular audit measured an overall difference of -0.6% ±1.8% with no significant variation between different energies identified. Following these initial national audits a
Introduction

UK wide dosimetry audit network was established in 1993 and is now coordinated by the IPEM and consists of 9 regional groups.

The NPL were invited by IPEM to perform reference dosimetry audits throughout the regional groups, thus linking the country’s audits through a ‘Hub and Spoke’ model as shown in Figure 1.3. Initially one MV audit was conducted annually in each region, and this was later expanded to include MeV and kV audits.

These NPL audits involve measurement of dose under reference conditions which conform to the relevant CoP. Experts in radiation dosimetry from NPL visit the host centre and perform independent dose measurements. This is in contrast to alternative methods, such as a postal dosimetry service offered by the International Atomic Energy Agency (IAEA) (Izewska and Andreo, 2000; Izewska, Vatnitsky and Shortt, 2004; Quality Assurance Team for Radiation Oncology (QUATRO), 2007; Izewska et al., 2015; Izewska, Lechner and Wesolowska, 2018). The NPL audits are on-site visits and so are valuable in assessing the entire calibration process in place at each department. Often it is not the individual measurements which are made incorrectly, but another aspect. This might include...
incorrect barometer calibration, or use of incorrect factors during the calibration which these NPL audits have helped identify (Nisbet and Thwaites, 1997; Thomas et al., 2017).

1.4.4 Continued development of dosimetry audit

Dosimetry audit continues to be prevalent within radiotherapy. As an example a search of “audit” and “radiotherapy” on PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) indicates an increasing number of publications over time as shown in Figure 1.4.

![Citation results for "audit" and "radiotherapy" on PubMed](image)

**Figure 1.4:** Number of citations on PubMed for “audit” and “radiotherapy” between 1990 and 2017. There is no sign of a reduction in publications on audit in radiotherapy.

Whilst it is not an area with the largest number of publications, audit and the measurements involved are fundamental to safe and effective radiotherapy.

Recently more specific and complex audits such as the national rotational radiotherapy audit (Clark et al., 2014), the national lung SABR audit (Distefano et al., 2017), the national stereotactic radiosurgery audit (Dimitriadis, 2017), and the national rectal contact brachytherapy audit (Humbert-vidan et al., 2017) have been completed instilling confidence in the introduction of advanced techniques and ensuring their safe implementation.
These ‘end-to-end’ style audits appear to be increasingly common as confidence has been built on fundamental dosimetry measurements. As techniques continue to become more complex this style of audit will be necessary to allow the entire treatment chain to be checked.

1.4.5 Dosimetry audit within clinical trials

Audit is now commonplace in radiotherapy clinical trials and has been shown to play an important part in their success. With the results from the HeadSTART trial, Peters et al. (Peters et al., 2010) demonstrated that protocol compliance had a demonstrable effect on clinical outcomes. While the study focussed on the number of cases performed noting that five times more patients were predicted to receive unsatisfactory radiotherapy if fewer than five patients were enrolled within a centre compared with those in which 20 or more were included. In this case the effect observed with poor radiotherapy outweighed the potential differences between the trial arms showing that while advances in techniques, and combination treatments are welcome, at the heart of their success is high quality radiotherapy.

With quality of radiotherapy within clinical trials of vital importance, pre-trial quality assurance has now become common place. Centres must prove they are able to deliver high quality radiotherapy which conforms to the trial protocol and this helps ensure trials maintain the statistical power they are designed with.

Audit within clinical trials developed through Europe in the 1980s (Clark et al., 2015) beginning with dosimetry audit and questionnaires. This same practice continued in the UK with trials including CHART (Continuous Hypo fractionated Accelerated RT Trial for Lung and Head and Neck Cancer) which was the first UK trial to have an associated audit program utilising machine based measurements, and RT01 (Sydes et al., 2004) which included clinically realistic dose measurements within a specially developed anthropomorphic phantom at each centre. Quality assurance within trials also aids in the safe implementation of new techniques across multiple centres. The PARSSPORT trial investigated the use of IMRT in Head and Neck cancer and was influential in providing robust quality assurance techniques to validate a centre’s implementation of IMRT treatment techniques and establish suitable tolerance levels (Clark et al., 2009). This included a questionnaire, as well as contouring and planning exercises prior to a dosimetry audit visit which involved film measurement as well as ionisation measurements of absolute dose, verifying the absolute doses delivered as well as the achieved dose distribution. This practice of robust audit and quality assurance continues to provide evidence of effective and safe technique introduction.
One recent trial which has had a large impact on centres is the CHHiP (conventional or hypofractionated high dose intensity modulated radiotherapy for prostate cancer) trial (Dearnaley et al., 2016). This trial included over 3000 patients and investigated whether a shorter treatment course for prostate cancer of 60Gy over 20 fractions (4 weeks) is at least as good as 74Gy in 37 fractions (7.5 weeks) which was previously the most common treatment regime. The results of this trial showed that the proposed 60Gy treatment regime was at least as good, and so many centres now use this as standard.

Dosimetry QA was completed for nine centres which participated in the CHHiP trial (Naismith et al., 2007). The centres were sent CT images (of the CIRS pelvic phantom) and corresponding structure sets on which the centre created a treatment plan according to protocol. This plan was then delivered to the phantom and dose measured at a point within the prostate volume with an ionisation chamber and an axial film measurement was taken. The measurements were compared with the calculated doses generated by the treatment planning system (TPS). All 13 plans were within ±2% of the overall TPS calculated dose (with individual fields generally within ±3%. When examining the film measurements differences of up to ±8% were identified near to phantom inhomogeneities. It was determined that the QC measurements demonstrated that centres could deliver treatments with an acceptable level of consistency between then, but that dosimetry of inhomogeneous regions needed further investigation.

Trial QA gives a good indication of the range of doses which may be expected to be delivered to patients and indicate the overall uncertainty which may be associated through the dosimetry chain. This uncertainty in delivered dose is demonstrably more when moving away from the high dose target region.

**1.4.6 Evidence of the impact of audit**

The IAEA conduct audits worldwide as previously mentioned (Quality Assurance Team for Radiation Oncology (QUATRO), 2007; Izewska, Lechner and Wesolowska, 2018), and these audits have been demonstrated to improve consistency and safety in radiotherapy. One point of note is that repeat audit visits show an improvement in results. The IAEA have found that typically for a first time audit visit 77% of results (from TLD measurements) are within ±5% (which is the acceptance limit used). This improves to 85% for those undergoing repeat audits (Izewska et al., 2015).

This improvement may be down to a number of factors, but out of tolerance results are followed up to help resolve problems. Identifying common problems in those centres participating enables this
information to be shared and helps prevent the same issues being faced in newly setup centres, or in developing countries in which there was previously no radiotherapy provisions. Radiobiological modelling can be used to predict the clinical impact of the discrepancies identified during audit.

1.5 Radiobiological modelling

Radiobiological modelling is used to attempt to explain or predict the clinical effect of a given dose of radiotherapy. This may range from purely empirical models to those which are biologically based. Its uses are wide ranging, covering models at both the cellular and population level (Kirkby, 2007; Kirkby et al., 2007).

In a general sense radiobiological modelling in radiotherapy aims to link a variable associated with the treatment (such as delivered dose, or tumour size) to a measurable outcome (cure, or complication rates), and they range in complexity.

The aim of radiotherapy is to cure the disease whilst minimising associated side effects. The probability of controlling tumour progression is known as TCP (tumour control probability), and the risks of associated effects to the surrounding normal tissues are known as the NTCP (normal tissue complication probability). The term ‘TCP’ may be used to refer to a number of different measures of ‘cure’ and the exact meaning should be specified. In this work, a number of different measures of TCP are examined including 5 year survival and biochemical disease free progression which are common measures of treatment success.

The dose response curves describe the effect of different doses on treatment outcome and are generally sigmoid in shape. The difference between TCP and NTCP is known as the ‘therapeutic ratio’ and a schematic of this is given in Figure 1.5. An increase in therapeutic ratio may be achieved in a number of ways; using a treatment technique with higher dose conformity (e.g. using IMRT over 3D conformal treatment techniques), through moving surrounding tissues (e.g. deep inspiration breath hold techniques in breast radiotherapy (Bruzzaniti et al., 2013; Lastrucci et al., 2017)) or through the use of combined treatments (such as combining chemotherapy with radiotherapy (National Cancer Action Team, 2011; Nomiya et al., 2013; Pollard et al., 2017)) to name a few.
Figure 1.5: Schematic showing how both TCP and NTCP increase with dose. The difference between the incidence of TCP and NTCP is known as the therapeutic ratio. Radiotherapy techniques in general will aim to maximise the therapeutic ratio, thus increasing the probability of treatment success whilst minimising associated side effects.

Different tissues respond differently to radiation and so exhibit dose response curves with varying gradients and likewise different measures of NTCP may have very different magnitude of effect for the same dose.

While there are many potential mathematical radiobiological modelling methods to explore, this work will focus on that most commonly used in clinical practice, the linear quadratic (LQ) model. Some alternative models are described briefly for comparison.

1.5.1 Target theory

Many models are based on the interactions of ionising radiation at the cellular level as cells are killed through the production of free radicals in the cell nucleus which in turn causes damage to the DNA causing the cell not to be able to replicate successfully. This DNA damage has been shown to be the primary cause of radiation induced cell death (Joiner and van der Kogel, 2009). One way of modelling this cell damage is to consider the DNA as the target and model the number of DNA strand breaks caused by the radiation.
Introduction

Initial modelling efforts assumed a single hit by radiation on a single target would produce cell death. This is known as *single-target single-hit inactivation*. This type of model produces an exponential survival curve (as shown in Figure 1.6).

![Comparison of single and multi-target single-hit models](image)

**Figure 1.6:** Examples of single and multi-target single hit models of cell killing based on target theory. Note that the multi target model contains a ‘shoulder’ commonly observed in mammalian cells studies. In this case $D_0 = 1.5\text{Gy}$ and $n=30$.

Using Poisson statistics and assuming a large number of hits, but with a small number of consecutive hits in a given cell the probability of survival for each cell is given as (Steel, 2002):

$$P(\text{cell survival}) = P(0 \text{ hit}) = e^{-D/D_0} \quad (1.1)$$

where $D$ is the delivered dose and $D_0$ is the dose which gives an average of one hit per target.

With this model a dose of $D_0$ reduces the number of cells surviving to 37% (i.e. to $e^{-1}$). This model describes situations such as the inactivation of bacteria, response to very low dose rates and some very sensitive human cells. However, in general there is usually a ‘shoulder’ seen on the survival curve (see Figure 1.6). This requires a multi-target single-hit inactivation model (Joiner and van der Kogel, 2009) which considers cell death to be caused by a single hit on each of the $n$ targets. The model can again be derived through Poisson statistics as follows:
Introduction

\[ P(0 \text{ hits on specific target}) = e^{-D/D_0} \quad (1.2) \]

thus

\[ P(\text{specific target inactivated}) = 1 - e^{-D/D_0} \quad (1.3) \]

so for \( n \) targets:

\[ P(\text{all } n \text{ targets hit}) = \left(1 - e^{-D/D_0}\right)^n \quad (1.4) \]

Giving the probability of cell survival as:

\[ P(\text{Survival}) = P(\text{not all targets inactivated}) = 1 - \left(1 - e^{-D/D_0}\right)^n \quad (1.5) \]

These multi-target models work reasonably well for describing effects at high doses, however are not as successful for clinical doses seen in radiotherapy (Joiner and van der Kogel, 2009).

The problem with models based on targets is that they do not accurately predict the response of human tissues as they do not take into account the influence of DNA strand breaks on cell survival. Even with a multi-target model a very flat response is predicted for low doses, however this is not seen in experimental data. To account for this the multi-target model was expanded to include a component for single-target cell death producing the two-component model:

\[ P(\text{survival}) = e^{D/D_1} \cdot \left(1 - \left[1 - e^{-D/D_1}\right]^n\right) \quad (1.6) \]

where the parameter \( D_1 \) fixes the initial slope of the curve (the required dose to reduce survival to \( 1/e \) in the low dose region) and other parameters are as described previously. This model still has limitations and would require a fourth parameter to best fit the clinical data, and thus would become too complex to be useful in most situations.

1.5.2 Tumour control probability: Linear quadratic model

The LQ model is used to model cell survival (AAPM, 2012) due to radiation induced cell death. The LQ model is used widely within radiotherapy (Steel, 2002; Dale and Jones, 2007) and is an approximation to many proposed mechanisms of radiation induced cell death (Sachs, Hlatky and Hahnfeldt, 2001). Cell death after exposure to radiation has been conclusively linked to DNA double
strand breaks (DSBs). The survival fraction (proportion of cells killed), $S$, is well approximated by an LQ function of delivered dose ($d$):

$$S = \exp(-\alpha d - \beta d^2) \quad (1.7)$$

where $\alpha$ and $\beta$ are coefficients for the linear and quadratic components respectively.

If irradiations are repeated $n$ times and deliver different doses, $d_i$, then we will have an overall survival fraction, $S_n$, of (Wedenberg, 2013b):

$$S_n = \prod_{i=1}^{n} S_i(d_i) \quad (1.8)$$

Expanding this gives:

$$S_n = \prod_{i=1}^{n} \exp(-\alpha d_i - \beta d_i^2) \quad (1.9)$$

with total dose, $D$, given by $D = \sum_{i=1}^{n} d_i$. If all delivered doses are equal then we can simplify this to:

$$S = [\exp(-\alpha d - \beta d^2)]^n \quad (1.10)$$

Equations (1.7) - (1.10) assume there are no ongoing repair mechanisms in place, and for the purposes of this work, this will be assumed throughout. Modelled treatments will be assumed to be delivered over the same time period, and so the impact of repair mechanisms on the steepness of the dose response curves will be only due to changes in delivered dose and not an additional time factor. Often it is assumed that each fraction of radiotherapy delivers precisely the same dose to the target in which case equation (1.10) will apply. The aim of this work is to assess the potential impact of small changes in dose delivered within each fraction and so the TCP modelling will be predominantly based on equation (1.9).

The LQ model is a mechanistic biophysical model which quantifies the effects of radiation-induced damage at the cellular level. Radiation induced cell death is highly dependent on double strand breaks (DSBs) (Sachs, Hlatky and Hahnfeldt, 2001) and the LQ model is often described by assigning the $\alpha$ component to relate to single strand breaks (SSBs) and the $\beta$ component to DSBs. The model
Introduction

inherently predicts the effect of dose rate on cell survival. With a high dose rate DSBs become more probable, and so the β component of the model becomes dominant and this gives a quadratic change in cell survival (which appears as a linear on a logarithmic scale which is a common representation) and this is what is observed in clinical practice.

The characteristics of an individual cell survival curve are governed by the values of α and β. In clinical practice the values of α and β are often not known separately and so the most common approach is to discuss the ratio of the terms; α/β, which then allow different values to be assigned to tissues with different response to radiation. This ratio is the dose at which the linear and quadratic components of the LQ model become equal. The cell survival curve given by the LQ model and influence of the α and β components is presented in Figure 1.7.

![Survival Curves](image)

**Figure 1.7:** Surviving fraction of cells predicted by the linear-quadratic model. The α/β ratio is given at the point where contributions of the linear and quadratic components are equal and describes the shape of the survival curve. The shape of the curve is dependent upon the type of radiation as well as the α/β ratio.

This work will investigate the impact of variations in the dose parameter as predicted by the LQ model for TCP over the course of a treatment, investigating the impact on individual patient predictions as well as effects on a population of patients. The variation in dose present in clinical
treatments will be determined from analysis of dosimetry audit data (see chapters 2 and 3) and analysis of routine measurements of machine beam output (see chapter 4).

The LQ as described above predicts the proportion of cells which survive an irradiation. For the purposes of predicting tumour control within radiotherapy in general it is accepted that to ‘cure’ the disease then all clonogenic cancer cells must be eliminated. This can be determined from the above LQ model using Poisson statistics to predict the probability of zero clonogenic cells surviving.

The probability of events occurring for a Poisson distribution is:

\[
P(k \text{ events}) = e^{-\lambda} \frac{\lambda^k}{k!}
\]  

(1.11)

In our case, \( \lambda \) is the proportion of cells which survive the irradiation (\( S \) from (1.7) - (1.9)) and \( k \) is the total number of cells remaining. We desire \( k=0 \) and so substituting this into (1.11) gives:

\[
P(\text{zero cells remain}) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda}
\]  

(1.12)

and so we can substitute in the cell survival after a single dose of radiation, \( d \), to a number of cells \( N_0 \) from (1.7) to give:

\[
TCP = \exp(-N_0S) = \exp(-N_0\exp(-\alpha d - \beta d^2))
\]  

(1.13)

or after \( n \) irradiations of dose \( d_i \) using (1.9) gives

\[
TCP = \exp\left(-N_0 \prod_{i=1}^{n} \exp(-\alpha d_i - \beta d_i^2)\right)
\]  

(1.14)

It should be noted at this stage that this model gives a prediction for a fixed value of \( \alpha \) and \( \beta \) and this would therefore be valid for an individual patient. The predictions given for populations take into account the variability in radiosensitivity and this is often achieved through varying the value of \( \alpha \), \( \beta \) or the \( \alpha/\beta \) ratio. For a population the curves tend to flatten due to this variation with individual patient curves being steeper (Webb and Nahum, 1993; Webb, 1994; Dale and Jones, 2007; Joiner and van der Kogel, 2009). This is demonstrated in Figure 1.8 showing a wide range of individual patient responses based only on varying the \( \alpha \)-value (mean \( \alpha/\beta \) of 10 with 20% standard deviation,
Introduction

dose per fraction of 2Gy and $N_0$ of $10^6$) following the same approach as Webb and Nahum (Webb and Nahum, 1993; Webb, 1994).

![Plot of individual patient TCPs which combine to give population TCP](image)

**Figure 1.8**: Production of a population TCP curve constructed from the combination of individual patients. Note that the population curve is shallower than the individual patient curves. In this work the distribution of $\alpha$ values is assumed to be normally distributed. This example modelled 1000 individual patients and used a dose per fraction of 2Gy, mean $\alpha/\beta$ of 10 with a 20% standard deviation, and $N_0$ of $10^6$.

One way of modelling this is to take a value of $\alpha$ from a distribution such as a normal or log-normal distribution. This Monte Carlo method was chosen to assess population variance within this work. It is noted that additional terms may be added into equation (1.14) to account for tumour growth; however this will not be included within this work.

### 1.5.3 Alternative empirical models investigated

Alternative forms of the LQ model exist which do allow replication of dose response curves which match clinically observed data. These empirical formulations use different parameters to describe a sigmoid shaped curve which may be used. These models can be simpler to fit to data and have easily described parameters. Here I will describe a Poisson model and Logistic model and demonstrate that often the results from either model are indistinguishable in terms of their clinical relevance. The
models described here are parameterised making use of the normalised dose response gradient, $\gamma$, formally defined as:

$$\gamma = D \frac{dP(D)}{dD}$$  \hspace{1cm} (1.15)$$

where $D$ is the total dose and $P(D)$ is the probability of the measured effect at total dose, $D$. Thus $\gamma$ is the percentage change in a measured response associated with a 1% increase in dose and is a dimensionless quantity. This $\gamma$ value can then be used to convert a change in dose to a change in response, $\Delta P$:

$$\Delta P = \gamma \cdot \Delta D$$  \hspace{1cm} (1.16)$$

A numerical value of the gradient of the dose response curve is a simple way to equate the effect of varying the dose on any measured effect so long as the dose response curve has been adequately determined at the relevant dose level.

1.5.3.1 Poisson model
The Poisson model follows a similar form to the LQ model described above but with an alternative parameterisation. These are fully described and discussed by Bentzen and Tucker (Bentzen and Tucker, 1997) and they give a number of different forms of the models using some common parameterisations. Here some key aspects will be highlighted.

The Poisson model may take the following form, parameterising in terms of the dose required for a 37% response level (i.e. $e^\gamma$).

$$P = \exp\left(-\exp\left[ e \cdot \gamma_{37} \left(1 - \frac{D}{D_{37}}\right)\right]\right)$$  \hspace{1cm} (1.17)$$

where $\gamma_{37}$ is the gradient of the dose response curve at 37% response rate, $D_{37}$ is the dose at a 37% response rate and $D$ is the total dose.

1.5.3.2 Logistic model
The logistic model has a variety of applications and as such many statistical software packages have built in methods for determining parameters of this model. As such it has been applied to
radiotherapy due to its relative simplicity despite having no biophysical basis. For fractionated radiotherapy the model takes the form:

\[
P(u) = \frac{\exp(u)}{1 + \exp(u)}
\]  

(1.18)

where \( u \) has the form \( u = a_0 + a_1 D + a_2 D d \), with \( D \) the total dose, \( d \) the dose per fraction and \( a_{0,1,2} \) are parameters to determine from model optimisation. With a fixed dose per fraction we can reparametrize to give \( u = a_0 + (a_1 + a_2 d)D = a_0 + bD \) and this then allows the logistic model to be parameterised in terms of \( \gamma_{50} \) and \( D_{50} \):

\[
P(D) = \frac{1}{a + \exp \left( 4 \cdot \gamma_{50} \cdot \left[ 1 - \frac{D}{D_{50}} \right] \right)}
\]  

(1.19)

with \( D_{50} \) and \( \gamma_{50} \) the dose and gradient at which the probability of response is 50% respectively.

1.5.3.3 Comparison of Poisson and logistic models

During the early stages of this work the model to use had not been decided upon. Experiments in fitting the Poisson and Logistic models described in 1.5.3.1 and 1.5.3.2 to clinical data were completed to explore the methodologies required for the modelling techniques required for implementation of the models. The data used for fitting was from the clinical trials summarised in Table 1.2. Each case showed a clear dose response.

The fits were produced for each set of data individually and a summary of the data used from each trial is given in Table 1.3. The corresponding produced fit results are displayed in Figure 1.9 with solid lines indicating the Poisson fit and dashed lines indicating the logistic fit.
**Clinical Trial** | **Brief details**
--- | ---
RT01 (Sydes *et al.*, 2004; Dearnaley *et al.*, 2014) | International randomised controlled trial comparing 64Gy (control group) and 74Gy (escalated dose group) in 2Gy fractions for prostate cancer. Determined better control of disease (10% increase in progression free survival observed) outweighed increase in side effects. 74Gy in 37 fractions became the standard fractionation regime. This trial included a total of 843 patients.

RMH (Dearnaley *et al.*, 2005) | Trial comparing 64Gy and 64Gy + boost of 10Gy (total 74Gy to prostate). PSA levels were the measure of tumour control used. PSA control rates increased (59% at 64Gy to 71% at 74Gy), however not statistically significant (p=0.1). Statistically significant increase in bladder effects for boost group, however these were short lived, manageable side effects. 126 men were included in the trial.

MD Anderson (Pollack *et al.*, 2002) | Dose escalation trial comparing 70Gy to 78Gy for prostate cancer. A significant improvement of failure free survival (increase from 64% to 70%) was found with dose-escalation. Rectal side effects were also increased (from 12% to 26% for Grade 2 or higher at 6 years). A total of 305 T1-T3 patients were included in the trial.

Fox Chase (Hanks *et al.*, 2002) | A study separated into different grades of prostate cancer (separated by initial PSA levels) assessing success through biochemical control (biochemically no evidence of disease (bNED)). A range of doses were delivered, and the median of each group has been taken in this case. Six groups were created, first by splitting by PSA level; <10, 10-20 and >20ng/mL. These were then split into favourable (Stage T1/T2a, Gleason score 2-6, no perineal invasion) and unfavourable (Stage T2b/T3 and/or Gleason score 7-10 and/or the presence of perineal invasion). A total of 232 patients participated in the trial. Those in which fits were produced were:

- Low PSA (unfavourable). Labelled as ‘Low-Unf’ in this work.
- Medium PSA (all). Labelled as ‘Med’

<p>| <strong>Table 1.2:</strong> Summary of trials from which published data was used to perform initial model fitting to clinical data. The data extracted from the publications is given in Table 1.3 and the model fit results are presented in Figure 1.9. |</p>
<table>
<thead>
<tr>
<th>Shortened Study Name</th>
<th>Dose (Gy)</th>
<th>TCP (various measures)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT01</td>
<td>64</td>
<td>43</td>
</tr>
<tr>
<td>RT01</td>
<td>74</td>
<td>55</td>
</tr>
<tr>
<td>MD Anderson</td>
<td>70</td>
<td>43</td>
</tr>
<tr>
<td>MD Anderson</td>
<td>78</td>
<td>62</td>
</tr>
<tr>
<td>RMH</td>
<td>64</td>
<td>59</td>
</tr>
<tr>
<td>RMH</td>
<td>74</td>
<td>71</td>
</tr>
<tr>
<td>Fox Chase Low-Unf</td>
<td>69</td>
<td>44</td>
</tr>
<tr>
<td>Fox Chase Low-Unf</td>
<td>71</td>
<td>51</td>
</tr>
<tr>
<td>Fox Chase Low-Unf</td>
<td>76</td>
<td>64</td>
</tr>
<tr>
<td>Fox Chase Med</td>
<td>70</td>
<td>19</td>
</tr>
<tr>
<td>Fox Chase Med</td>
<td>72</td>
<td>31</td>
</tr>
<tr>
<td>Fox Chase Med</td>
<td>76</td>
<td>84</td>
</tr>
</tbody>
</table>

Table 1.3: Summary of data used when fitting Poisson and Logistic models shown in Figure 1.9. Further details of each study are detailed within the text and Table 1.2.

Figure 1.9: Plot of Poisson (solid) and Logistic (dashed) model fits to ‘TCP’ outcomes data for prostate cancer form the RT01 (Red), RMH (Blue), MD Anderson (Green) and Fox Chase Trials (Grey).
Intr
oduction

From the resulting fits the $\gamma$ values at 74Gy were extracted (taking this as a current widely used prescription dose for the treatment of prostate cancer) and are given in Table 1.4. One aspect to note from these fits is that at both high and low doses the Poisson model gives a shallower gradient. This is discussed at length by Bentzen (Bentzen and Tucker, 1997) and is due to the varying characteristic sigmoid shapes of the dose response curves produced by each model. In clinical situations response to these low doses of radiation is generally not well known. At the doses of interest both models provide adequate predictions to clinical data.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$\gamma$-value at 74Gy</th>
<th>Difference (%/Gy)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poisson Model</td>
<td>Logistic Model</td>
</tr>
<tr>
<td>RT01</td>
<td>1.13</td>
<td>1.20</td>
</tr>
<tr>
<td>RMH</td>
<td>1.04</td>
<td>1.09</td>
</tr>
<tr>
<td>MD Anderson</td>
<td>2.37</td>
<td>2.41</td>
</tr>
<tr>
<td>Fox Chase – Low (Unfavourable)</td>
<td>2.64</td>
<td>2.76</td>
</tr>
<tr>
<td>Fox Chase – Med</td>
<td>9.72</td>
<td>11.39</td>
</tr>
</tbody>
</table>

Table 1.4: Extracted $\gamma$-values from the fitting of Poisson and Logistic models to the clinical prostate outcome data described in Table 1.3 and plotted in Figure 1.9. The largest difference in the gradient of the dose response curve at 74Gy was 1.67%.

The ‘Fox Chase Med’ data had the steepest dose response curve and thus the greatest $\gamma$-values of 9.72 (Poisson) and 11.39 (Logistic) at 74Gy. Whilst these sets of data do not describe identical patient populations, the reason behind this exercise was to compare the Poisson and Logistic models and develop familiarity with some of the techniques required in computational radiobiological modelling. With exception of the Fox Chase – medium risk data, the $\gamma$-values of both models agreed within 0.2%/Gy. This modelling exercise allowed the development of an understanding of the fundamental datatypes used within the Python language and has demonstrated the ability to perform suitable types of modelling and optimisation within this framework.

Whilst this exercise used the models parameterised by $\gamma_x$ and $D_x$ values, the intention of this work is to produce a model which is applicable for both individual patients, populations and be flexible in its use. This is the reason which has determined the use of the LQ model described in section 1.5.1 for modelling of TCP outcomes as these have a biophysical basis and there are a number of studies which provide values for the range of model parameters. In principle a fit produced by the LQ model could be produced, data points extracted and then both the above Poisson and Logistic models
could be fitted to these LQ model points allowing estimation of $\gamma_x$ values if necessary, however this is beyond the scope of this work.

It is noted that in the cases examined above, the same dose per fraction was used within each dataset. If alternative dose fractionations are used, then the physical dose may not be the most useful metric for which to produce a fit. There are a number of options which include EQD2 (biological equivalent dose in 2Gy fractions), BED (biological effective dose), and EUD (equivalent uniform dose). These metrics attempt to covert different fractionation regimes to a measure of the biological impact of the delivered dose. Probably the most common in clinical practice is the use of EQD2 to convert doses required with different fractionations as historically radiotherapy was delivered in 2Gy fractions and so this is where the initial experience has its foundations. Whilst this will not be performed as part of this work, it has been considered when developing the modelling techniques to allow an alternative ‘Dose’ to be input into the models.

1.5.4 Normal tissue complication probability: Lyman-Kutcher-Burman model

In many cases the aim in radiotherapy is to deliver a homogenous dose to the target volume. For the surrounding tissues there is often large dose gradients present across the various organs of interest. With this in mind a number of dose-volume models have been developed for predicting the response of normal tissues to radiotherapy. The most widely used is the Lyman-Kutcher-Burman (LKB) model (Lyman, 1985; Gulliford et al., 2012; Wedenberg, 2013a), and this is chosen to be used within this work to model NTCP. This model takes the following form for predicting the NTCP for a given dose, $D$, in a partial organ volume, $V$ (Steel, 2002; Gulliford et al., 2012):

$$
NTCP(D, V) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u(D,V)} \exp\left(-\frac{x^2}{2}\right) dx = \frac{1}{2} \left(1 - \text{erf}\left(\frac{u(D,V)}{\sqrt{2}}\right)\right)
$$

where the dependence of volume and dose is in the upper limit of the integral:

$$
u(D,V) = \frac{D - D_{50}(V)}{m \cdot D_{50}(V)}
$$

and the volume dependence of $D_{50}$ is assumed to follow the relationship:
Introduction

$$D_{50}(V) = \frac{D_{50}(1)}{V^n} \tag{1.22}$$

$D_{50}(1)$ denotes the uniform total dose to the entire organ producing a 50% incidence of the chosen NTCP endpoint. The model depends upon three parameters; $m$, $D_{50}(1)$, and $n$.

The volume exponent, $n$, lies between 0 and 1 with larger values corresponding to a larger volume effect. The parameter, $m$, is inversely related to the steepness of the dose-response curve. Often the complications of interest in radiotherapy are those which impact the long-term function of an organ, and organs can be classified to some degree in terms of having a parallel and serial architecture as an analogy to an electronic circuit. In a serial organ, damage to one section will have a direct affect to all following sections. The prime example of this is the spinal cord. If one section of the cord is damaged, then a correct nerve signal is not received to the remainder of the cord and so this localised damage has an effect to all parts of the body below this. A parallel organ may better tolerate a large amount of damage to a small area, but a lower dose to a large volume of the organ may have an increased affect. The lung is an example of an organ with parallel architecture. Damage to a small volume of the lung can generally be compensated for by the remainder of the lung volume. Hence the seriality of an organ relates to the value of $n$ within the LKB model in (1.22).

As well as the LKB model there are many other proposed models, however the LKB model is often the model of choice. Studies indicate that the LKB model is often statistically the most robust model to use (Wedenberg, 2013a, 2013b), and recent clinical trials often take this approach (Semenenko and Li, 2008; Gulliford et al., 2012) and so this has been adopted in this work to ensure wide applicability. Work by Wedenberg (Wedenberg, 2013b) compared the LKB model with a Poisson based model for outcome data for spinal cord and lung from the published QUANTEC (Quantitative Analysis of Normal Tissue Effects in the Clinic) dose-response relations (Kirkpatrick, van der Kogel and Schultheiss, 2010). The LKB model achieved a more robust prediction in over 90% of cases lending evidence towards using the LKB model for this work.

Within this work the models used and the details of assumptions will be specified alongside the corresponding results. The models have been developed to allow all the parameters to be fitted with any combination of remaining parameters provided.
1.6 Implementation of tools for data analysis and modelling

A large volume of data has been collated for this work and a suitable method of storing and using this data had to be determined to ensure good management throughout the project and allow reproducibility and dissemination of results as appropriate. Early work was completed using Microsoft Excel (Microsoft, 2010) however at an early stage it was felt more appropriate to move to an environment better suited to modelling and larger scale data analysis.

1.6.1 Implementation of modelling process

The Python programming language (Python Software Foundation, 2016) was selected due to it being readily available with no issues surrounding licensing which MatLab (The MathWorks Inc., 2015) may have in certain environments (such as within an NHS hospital setting). Within the radiotherapy physics group at the RSCH they had also recently began using tools which are built upon Python and so developing skills with this programming language will potentially have a wider benefit than using MatLab which is not heavily utilised within the department currently.

The approach to using Python is mixed, often utilising different tools for initial development and then final simulation purposes. Python was installed for use through installing the Anaconda distribution (Anaconda Software Distribution, 2016). One major decision was to use Python 2.x or 3.x. In this work 3.x (all code tested to run on python version 3.5.2) was selected as this is increasingly becoming the standard for newly developed python packages including those developed specifically for radiotherapy such as PyLinac (James, 2017). Within the Anaconda distribution is contained the Jupyter notebook format (‘Juyter Notebook’, 2016), which has been predominantly used for development. With this, both formatted text and blocks of code can be integrated into a single file which runs through a web browser. For simulations the IDE, Spyder (The Spyder Project Contributors, 2017), is included which has a similar feel to MatLab and allows editing and running of standard python (.py) files. Finalised code for the models was transferred to the standard python file type to allow greater portability between different environments in the future and these can be relatively easily produced from the Jupyter files used for development. All of these tools are freely available and open source.

Within python, as well as the standard library there are several available packages and these have been used for this work. Those used have been listed in Table 1.5 and the majority are from the ‘SciPy’ stack which is considered the standard set of packages used throughout Python’s scientific community.
## Python Package

<table>
<thead>
<tr>
<th>Python Package</th>
<th>Usage description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numpy (1.11.0) (van der Walt, Colbert and Varoquaux, 2011) *</td>
<td>Contains support for arrays and matrices and mathematical functions.</td>
</tr>
<tr>
<td>Scipy (0.17.1) (Jones, Oliphant and Peterson, 2001) *</td>
<td>Used for linear algebra and statistics and is built upon Numpy datatypes.</td>
</tr>
<tr>
<td>Matplotlib (2.0.0) (Hunter, 2007) *</td>
<td>Used for producing visualisation of data (i.e. production of plots).</td>
</tr>
<tr>
<td>IPython (4.2.0) (Perez and Granger, 2007) *</td>
<td>Provides interactive use of python commands and is used within the Jupyter notebook.</td>
</tr>
<tr>
<td>Pandas (0.18.1) (McKinney, 2010) *</td>
<td>Provides data structures suitable for multiple datatypes and which can be act as a database type system.</td>
</tr>
<tr>
<td>Pydicom (1.0.0a1) (Mason, 2011)</td>
<td>Package developed for working with DICOM files.</td>
</tr>
<tr>
<td>Dicompyler-core (0.5.2) (Panchal, 2017)</td>
<td>Library of modules which interact with DICOM objects for use in radiotherapy such as DVH analysis. The stand-alone software ‘dicompyler’ was also used and this provides visualisation of DICOM data as well as export and anonymization. This is built from the dicompyler-core library.</td>
</tr>
</tbody>
</table>

**Table 1.5:** List of Python packages used heavily within this work with brief description of usage. Those marked with an asterisk (*) are part of the Scipy stack. The version numbers are correct at the end of the work and all code used within the thesis used these versions.

All these packaged are freely available and open source and thus the work could be replicated and independently verified if required using these packages. The version numbers given were correct at the end of the project and all code used within this thesis was run on these versions but may have changed through the course of the work. The python language has thus been chosen to develop the models, run simulations as well as analyse the data collated and produced.
1.7 An area of open investigation

Whilst there is a wealth of publications on radiobiological modelling in radiotherapy there is little recent work addressing the impact of routine QC tests. Work is ongoing to include radiobiological modelling into treatment planning systems, however there is little to link to the dose variations measured daily on most treatment machines.

Treatment plans are often created to fulfil dosimetric targets to within a few tenths of a percent (despite known limitations of the algorithms involved), however the absolute dose may not be this accurate and should be included in the radiobiological modelling as this develops.

The variation in delivered dose between not only individual centres, but on individual treatment machines may be a determining factor in the patient’s outcome. For example, an allowable tolerance on beam output of ±2% gives rise to a potential 4% difference in dose delivered to patients treated on neighbouring machines. This is of the same order of magnitude as the differences in dose assessed in some trials examining dose escalation such as the CHHiP trial which aimed to assess a 5% difference in outcome (Dearnaley et al., 2012, 2016).

Gaps in knowledge exist in the clinical impact of the variation in dose as measured through dosimetry audits as well as routine daily QC. This work aims to fill some of those gaps, assessing the variation of dose as measured by audit and also the systematic dose variation associated with daily treatment deliveries. The dose variations determined will then be used within radiobiological models to determine the potential clinical impact of these. The aim is also to develop the modelling tools required in an open-source way which should not then limit its further development and use which often comes with work developed in a specialist computing environment.

This work will begin by assessing the variation in dose due to the beam calibration. This will be assessed through analysis of absolute dosimetry audits performed by regional UK groups and the NPL. Further variation in dose delivered due to variations additional to the calibration variations will be assessed through analysis of routine beam output measurements from UK centres. These potential variations in delivered dose will be applied to radiobiological models to predict the impact of this on clinical outcome for prostate and head and neck cancers.
2 Regional audit data

As discussed in section 1.4.3 the UK is split into regional dosimetry audit groups. Originally these were linked by audits conducted in each region by the NPL. If audits are conducted regularly each region should remain relatively consistent with all others assuming no undetected systematic shift in results within a single region as all results could technically be traced back to the time when an original audit was conducted by the NPL. In this instance it is worthwhile examining the regional audit data available to see if this is the case prior to examining the data from the NPL audits.

There are currently nine groups consisting of approximately 6-13 centres within each group. The groups as of 2018 exist as follows, eight of which are the original geographic groups shown in Figure 1.3:

A. Scottish and Northern  
B. Trans Pennine  
C. Midlands  
D. South West  
E. South East Central  
F. North East Thames  
G. South East Thames  
H. Anglia  
I. Genesis Care UK (A group of privately run radiotherapy centres formally known as Cancer Partners UK, established in 2009)

Within this work the regions have been anonymised and have been randomly assigned a group number between 1-9, which will be used to refer to the data collated and analysed within this work. This work is predominantly concerning the use of MV treatment beams, and so discussion of the regional audit groups will focus on the MV audit results. MV audits are performed at a greater frequency due to their higher proportion of clinical use.

Many groups include a mix of NHS and private radiotherapy providers and representatives from each group meet regularly to ensure similar objectives are met across the country. Regular meetings are held to discuss audits conducted within each regional group. This includes discussion of audits
completed and proposed audits allowing ideas and methods to be shared. This helps ensure consistency is maintained across the UK.

2.1 Collation of regional audit data

A request was sent via email (and in person at national interdepartmental audit meetings) for access to the audit data collected by the regional audit groups. Many of the groups acknowledged that there was not a single location in which the results were stored, and so there is a wide range in the number of results received from individual groups. Some groups had all measurements available down to the details of equipment used, whereas others were only able to send through a summary of the reference dose audit results. Storage of results is beginning to become more centralised, making use of the IPEM online document store, and so going forward it is expected that future results will be more readily available which may be useful for future work similar to this. A selection of data was received from five of the nine groups and a summary is given in Table 2.1. No data was received from groups 2, 6, 8 or 9 and the data supplied from groups 1, 3, 4, 5 and 7 varied considerably in its scope.

<table>
<thead>
<tr>
<th>Regional group</th>
<th>1</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. centres</td>
<td>8</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>No. beams</td>
<td>24</td>
<td>19</td>
<td>18</td>
<td>12</td>
<td>148</td>
</tr>
<tr>
<td>Mean (%)</td>
<td>0.2</td>
<td>0.11</td>
<td>-0.08</td>
<td>-0.21</td>
<td>0.01</td>
</tr>
<tr>
<td>Min (%)</td>
<td>-1.1</td>
<td>-1.2</td>
<td>-0.8</td>
<td>-1.5</td>
<td>-2.6</td>
</tr>
<tr>
<td>Max (%)</td>
<td>1.8</td>
<td>1.6</td>
<td>0.8</td>
<td>1.5</td>
<td>2.7</td>
</tr>
<tr>
<td>Standard deviation (%)</td>
<td>0.84</td>
<td>0.74</td>
<td>0.48</td>
<td>0.90</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Table 2.1: Summary of MV data received from regional audit groups (Regions A-I have been anonymised to groups 1-9). Note: no data was received from groups 2, 6, 8 or 9. The table summarises the mean difference in output as measured by the visitor relative to the host centres.

Group 7 provided data spanning 20 years and so this was assessed for variation over time. A decrease in the standard deviation of the measurement results was seen. A reduction from approximately 0.6% to 0.3% was observed between the first 20 and last 20 audit measurements.
2.2 Previously published regional audit results

The South West (Blake and Casebow, 2002) and South East Central (Palmer et al., 2011) audit groups had previously published results of their region’s audits and key points have been highlighted in sections 2.2.1 and 2.2.2 and compared to the data supplied for this work.

2.2.1 Previously published results from south west regional audit group D

The audits conducted were designed to take a maximum of two hours on a particular machine, and at the time NPL performed audits across the regional groups to act to tie together all regions audit results.

The MV audits conducted required the delivery of a three field plan (2 open and one wedged) to a phantom, 2Gy at the central point. An important point to note is that centres were asked to process the plans in the same way they would for a normal patient. Additionally the host and visiting centre’s chambers were compared at 5cm depth within the phantom to provide a measurement of the difference in beam output. It is this beam output variation which I will focus on in this work.

Four rounds of audit were reported on and means and standard deviations in the beam output varied between -0.2% to +0.3% and 0.4% to 0.7% respectively. Results of the measurements taken at the central point of the three field plan had larger standard deviations (as would be expected) from 1.3% to 1.8%. All audits were within the specified ±5% tolerance used.

2.2.2 Previously published results from south east central audit group E

The south east central group published an analysis of results obtained between 1993 and 2008. Prior to 2005 a ±3% tolerance was used and this was subsequently reduced to ±2%. In this set of published data they noted that without the NPL audits, which are considered the ‘gold standard’, assigning a positive or negative sign to comparisons may not be meaningful as neither the host nor visiting centre can be deemed the ‘correct’ centre. In this case the paper considers only the absolute magnitude of the discrepancies.

The maximum deviation was reported to be 2.7% for beam output measurements (which was within the ±3% tolerance used at the time). The mean of all deviations from each year’s audits completed shows a general reduction. Estimations from Figure 1 within the report show that in 1993 and 1995 the mean deviation was approximately 1%. This reduced to a mean of approximately 0.6-0.8% in 2006-2007). The 2008 results had a mean deviation of approximately 0.3% but there was only three MV audits performed in this year.
Some discrepancies were identified, particularly when a more complex audit technique was introduced in 2005, through the introduction of fields defined by MLCs. Errors identified included minor inaccuracies of the treatment planning system model close to an MLC edge, an incorrectly calibrated mercury thermometer (due to calibration at sea level rather than correcting for geographical location).

Based on their results they recommended an investigation level of ±1.5% with maximum acceptable deviation of ±2%. As well as this it was recommended that only centres which had had a recent NPL audit should perform audits on other centres. At the time of publication NPL were routinely auditing UK centres, however this is now only available at a cost to the host centre. This means that routine NPL audits are more unlikely to be performed unless additional funding is available.

### 2.3 Audit schedules

There are two primary methods of conducting audits within a group: 1) a single auditor and 2) round-robin style audits. The principle of these is demonstrated in Figure 2.1.

![Diagram of audit schedules](image)

**Figure 2.1:** Example of two styles of audit conducted between a group of five centres. The left shows the ‘round-robin’ style audit cycle in which each centre conducts an audit and is audited. The right shows a single centre auditing the four other centres.

Different schedules may be used when conducting dosimetry audits. Often audits checking the introduction of new techniques will take the approach of a single group (or individual) travelling to all sites and performing the audits, whereas regional groups may use a ‘round robin’ schedule where each centre will visit one other to perform an audit. Both these scheduling methods have benefits, but also limitations.
Regional audit data

Round robin schedules ensure travelling and time required is shared throughout the group, which is often important when there is limited funding and time available for staff from a single centre to perform audits. A single auditor approach (for example the NPL dosimetry audits and recent IMRT audits and SRS audit) has the benefit of consistency, and so all relative results are directly comparable to each other.

A brief discussion of the impact of including NPL within audit schedules to link groups together was given by Palmer et al. (Palmer et al., 2011). This analysis indicated that the uncertainty in audit results is increased (up to double) when audits conducted by the NPL are not used to link different regions.

2.4 Summary of regional audit data

The previously published regional audit data discussed in section 2.2 provides the best overview of regional audit results. The additional data collated through the request as part of this work has not greatly expanded on this.

The standard deviation of results varied between a minimum of 0.48% (18 beams) and maximum of 0.90% (12 beams). Group 7 supplied the largest set of results (148 beams) and had a standard deviation of 0.79%. The additional collated data and associated communication with departments has demonstrated that each region broadly follows a similar approach and has comparable results. With the groups linked by previously performed NPL audits, this means that all UK regions are delivering doses which are comparable to one another with no obvious outliers detected from the supplied regional results.
3 Analysis of NPL audits

The NPL provide a dosimetry audit service to both NHS and private radiotherapy centres which acts to verify the correct implementation of the relevant code of practice. The first of these dosimetry audits was conducted in June 1994 and they are ongoing and measure all beam types (MV, MeV and kV) following the CoP used within the department at the time of measurement. As previously mentioned the NPL conducted audits in each UK regional group annually. Audits are now performed when requested by individual centres and funding would need to be obtained for this.

Key aspects of the following work have been presented as a poster at ESTRO (Thomas et al., 2016) and published in the European Physics and Imaging in Radiation Oncology (PhIRO) journal (Thomas et al., 2017) (see Appendix A for publications arising from this thesis). The published work focussed solely on NHS hospitals. Here private centres data are also included along with additional discussion and analysis which was originally unpublished. It should be noted that I collated and analysed all available data presented in this chapter.

This chapter reviews the dosimetry audits for reference conditions which the NPL has completed, from June 1994 to February 2015. The analysis completed gives an overview of the state of the accuracy of absolute dose calibrations in radiotherapy across the UK throughout this time. The data included is for audits which were completed within NHS and private radiotherapy centres and has been collated from a combination of the audit reports sent to the host centres and the original data as recorded during the audit.

3.1 Methods employed in NPL dosimetry audits

3.1.1 General procedure

The audits conducted by NPL are performed according to the published UK CoPs (the 1990 CoP for MV photon dosimetry (Lillicrap et al., 1990), 1996 (Burns et al., 1996) and 2003 (Thwaites et al., 2003) CoPs for MeV, and the 1996 kV CoP (Klevenhagen et al., 1996) and its 2005 addendum (Aukett et al., 2005) relevant to the centre being audited at that time. NPL use their own calibrated chambers, electrometers, barometers, thermometers and phantoms. A member of the NPL dosimetry group visits the department and first takes measurements of the beam quality index (tissue phantom ratio $TPR_{20,10}$ for MV, $R_{50,D}$ for MeV, Half Value Layer (HVL) for kV) from which
calibration factors for the secondary standard ionisation chamber are derived. An output measurement in reference conditions for each beam quality under test was then made using both the NPL equipment and the tertiary standard supplied by the host department allowing a check of the calibration factor on the tertiary standard. These measurements are then compared to those given by the host centre, which may be a result measured on the day as part of a daily check, or calculated from a data table. As well as the absorbed dose to water, the calibration of the local thermometer, barometer and calculation of ion recombination are also checked, providing a complete independent verification of the delivered dose and all intermediate steps following the CoP, thus ensuring each aspect of beam output measurement is verified. This contrasts with postal audits which may use TLDs or Alanine in which only the dose delivered can be compared. Within a given audit cycle, the same NPL secondary standard chamber was used for all measurements wherever possible.

In the clinical environment the chamber to be calibrated may be positioned using a side-by-side phantom in the same beam as that with a calibration. The measurements can then be compared using the geometric mean to determine the calibration factor for the field instrument. A method of direct replacement or substitution, where only one chamber is positioned within the beam at a time, and then the other is placed in the same location and further measurements taken, to transfer the calibration, may also be used and was the method employed for these audits.

3.1.2 Audit setup

3.1.2.1 MV setup

Dose measurements were made in a specially constructed water-filled polymethyl methacrylate (PMMA) phantom that was the same as that used to calibrate the secondary standard chambers when sent from the hospitals to NPL (in the time period considered for these audits). This was a 30x30x30 cm open top box made from PMMA of wall thickness of 1.0 cm (considered as 1.2 cm water equivalent). The phantom contains inserts which allow the NE2561/NE2611 ionisation chamber to be positioned at water equivalent depths of 5.0, 7.0, 10.0 and 20.0 cm on the beam axis. A schematic of this setup is shown in Figure 3.1.
Figure 3.1: Diagram of setup used for MV photon absolute dosimetry audits performed by NPL. The secondary standard ionisation chamber is positioned centrally in a 10x10cm field at a depth of 5, 7, 10 or 20cm in water.

The linac was then set to produce a horizontal beam with the collimators set to give a 10.0 x 10.0 cm field at the radiation isocentre. The phantom was positioned in a horizontal beam using the light field and cross hairs on the linac. A plane flat mirror was used to align the phantom surface perpendicular to the beam direction by reflecting the light field back on to the head of the machine and ensuring the shadow of the crosshair was coincident. The reference point of the chamber was positioned on the central axis of the light field at the required reference depth in water. The focal distance was set according to local practice and the physical pointer, projected scale readings and laser agreement were compared.

3.1.2.2 MeV setup

Measurements were taken in NPL-owned WTe Solid Water phantom material (St Bartholomew’s Hospital, London, UK) using either a Roos or NACP-02 type ionisation chamber with matching WTe chamber holder plate. A selection of WTe sheets were used, ranging in thickness from 1 to 40 mm to enable the build-up depth to be changed to match the requirements of the beam quality under test. The WTe phantom was set up on the patient couch. A vertical beam (gantry = 0°) was used. The same arrangement was used for beam quality, beam output and tertiary standard calibration.
3.1.2.3 kV setup
For the measurement of kV beam quality (HVL), an NPL-owned jig was used that enabled a narrow beam geometry arrangement for a horizontal beam. Radiographs of the NPL chamber used in this geometry were taken to confirm correct alignment of the chamber. High-purity sheets of aluminium (or copper, as appropriate) covering a range of (pre-measured) thicknesses were used to enable the assessment of the HVL.

For beam output and tertiary standard calibration, an NPL-owned jig was used to place either the NPL secondary standard or the tertiary standard in air at the end of the required applicator for a horizontal beam.

3.1.3 Beam quality
The beam quality is defined in different ways depending on the type of beam; MV, MeV or kV. Full details of each are given in the relevant Code of Practice.

3.1.3.1 MV beam quality
For MV beams the beam quality is the TPR_{20,10} (Lillicrap et al., 1990). This was measured using a NE2561/NE2611 chamber and taken to be the ratio of ionisation measurements at 20.0 cm and 10.0 cm deep in water with a fixed focus to chamber distance of 100cm and constant field size (10.0 x 10.0 cm). Readings were corrected for temperature, pressure and ion recombination. In practice any differences in ion recombination with depth (dose per pulse) make only a small change in TPR, which in turn gives only a small change in absorbed dose calibration e.g. a change of 1.0% in the TPR leading to a change of 0.16% in the calibration (Klevenhagen et al., 1996). It should be noted that this may not be the case in Flattening Filter Free (FFF) beams and this issue is addressed in the recent IPEM publication (Budgell et al., 2016), however no FFF beams were audited during the time period this work covers.

3.1.3.2 MeV beam quality
For MeV beams there are two codes of practice discussed within this work, referred to as the 1996 (Burns et al., 1996) and 2003 CoPs (Thwaites et al., 2003). Major differences between them will be noted here, but a detailed discussion is contained within the 2003 CoP. The 1996 CoP derives absorbed dose to water from a 2 MV or Co-60 Air Kerma calibration of the NE2561/NE2611 secondary standard chamber. E₀ and E₂ are required to select physical data to convert to absorbed dose to water and are determined from a measurement of R_{50,UD}. The 2003 MeV CoP greatly simplifies the required steps for the user, giving the absorbed dose to water calibration of the
chamber traceable to the therapy electron primary standard calorimeter (McEwen, Williams and DuSautoy, 2001). The calibration factor for a secondary standard is given over a range of beam qualities defined by the quality index $R_{50,D}$ which is the depth at which the dose in water falls to 50% of its maximum along the central axis. This is measured at 100cm SSD in a field large enough to ensure the result is independent of field size (the required field size will generally increase with energy). Measurements were taken in WTe Solid Water using either Roos or NACP-02 chambers.

3.1.3.3 $kV$ beam quality
For $kV$ beams the beam quality is specified using the concept of the half value layer (HVL) which, in the case of the calibration of the secondary standard chamber at NPL, is defined in terms of the thickness of Aluminium or Copper required to reduce the air kerma rate by 50%. This is then used to derive the calibration factor for the secondary standard. Measurements were taken in air using NE2561/NE2611 chambers using narrow beam geometry with a setup to minimise scatter conditions which were dependent upon the host centre’s equipment location.

3.1.4 Beam output
Radiation output is defined as the absorbed dose to water per Monitor Unit (MU) (or on some units in terms of time, usually seconds) measured for a given set of reference conditions. The 1990 MV photon code of practice did not explicitly state what conditions should be used for reference dosimetry. The situation has therefore arisen where different criteria are in use; for example, either 100 cm source to surface distance (SSD) or to isocentre is used, with various measurement depths ranging from 5 – 10 cm. Hence for MV photon dosimetry there exist a number of different ways in which the reference conditions have been implemented. Care must always be taken in ensuring the interpretation of the reference conditions within a department is fully understood.

3.1.4.1 MV beam output
Radiation output is measured at a reference depth in water of 5 cm, or 7 cm for beams with quality index greater than 0.75. The output is measured at the beginning and end of each visit to ensure consistency and detect any drift, the average of these is taken for comparison with the locally provided value. The locally provided value was generally measured on the same day and the results may also have been derived from the centre’s tabulated data. Dependent on the reference conditions on occasion a percentage depth dose correction from centre-measured data was required.
3.1.4.2  MeV beam output

3.1.4.2.1  1996 code of practice
The output in terms of absorbed dose to water of the electron beam is determined based on a chain of measurements from an air kerma calibration of the NE2611 secondary standard. The transfer of calibration is first by means of a cross calibration of the secondary standard to a “Farmer” type chamber in a photon beam, then transfer from the Farmer type chamber to plane parallel electron chambers in a high energy electron beam (note whilst the Farmer chamber may be used for calibration of higher energy electron beams it is not recommended for use below an $R_{50,D}$ of 4 cm of water). The reference measurement depth ($Z_{ref}$) of $0.6 R_{50,D} - 0.1$ cm is calculated for calibration of beams, from the beam quality index. Derivation of the calibration factor for other electron energies may then be calculated as described in the CoP.

3.1.4.2.2  2003 code of practice
The 2003 MeV CoP specifies absorbed dose to water calibration factors given as a function of $R_{50,D}$ from the NPL calibration service (McEwen, Williams and DuSautoy, 2001). It is recommended that parallel plate chambers are used to measure the depth dose curves which will be used in determining the calibration of the beam. Absorbed dose should be determined in water at the calculated reference depth. Unless a significant difference was found between the value given by the host and that derived by NPL for $R_{50,D}$, and hence $Z_{ref}$, the value supplied by the host was used by NPL for the audit measurements.

3.1.4.3  kV beam output
The kV CoP (including its addendum in 2003) is split into three sections; medium (0.5 - 4 mm Cu HVL), low (1.0-8 mm Al HVL), and very low energy (0.035 - 1 mm Al HVL) x-rays. Methods for calibrating at the surface or at a depth of 2 cm in water are provided for medium energy x-rays, and the method chosen will depend upon the clinical need of a particular department. For calibration based on dose at the surface an in-air method is used, whereas dose at depth may be more accurately determined using measurements in water. The addendum included revised values for $K_{ch}$, extension to the backscatter factors, mass energy absorption coefficients, and a methodology to allow for the determination of absorbed dose to water either at the surface or at 2 cm deep dependent on clinical need. Measurements during audits were always performed in air with the chamber positioned centrally at the end of the applicator. The known thickness of the chamber is used to correct the response of the chamber back to the end of the applicator using the inverse square law. The CoPs are then used to convert the air kerma measurements to dose at the surface of a phantom or at 2 cm deep in water dependent on the requirements of the centre.
3.1.5 Reporting of results and treatment of data

Results from each audit were reported as a ratio of the value determined by NPL to that of the host. The beam quality, radiation output and calibration factor of the field instrument are derived during the audit and compared with the values in use within the department. These measurements were all performed following the recommendations of the relevant CoP in use within the department. Audit regions were able to select which centre was audited. Some had the same centre audited each time and they in turn would audit the remaining regional centres (see chapter 2). Others rotated the NPL audit amongst the group. Due to this there are a number of repeat visits to centres within this dataset, with the number of repeat visits varying between regional groups.

Measurements have been categorised by beam type, electron audits are then further divided according to which CoP was used. For each beam type the mean difference in the dose between the NPL and host centre was determined. Within this work the set of audit measurement for an individual beam will be referred to as a measurement set. The NPL report the results as a ratio between the NPL and host centre. In this work I have converted this ratio to a percentage difference to be consistent with the remainder of this work. For example a reported NPL:Host ratio of 1.015 is converted to +1.5%.

3.1.6 Statistical comparisons

The mean output ratio, standard deviation and standard deviation of the mean of results have been compared. Output ratios have been compared using an independent t-test where equal sample variance can be assumed. A Welch’s t-test has been used when the assumption of equal variances does not hold. Where multiple comparisons have been the use of ANOVA has been detailed and p-values have been adjusted using the Bonferroni method.

3.1.7 Measurement uncertainties

The measurement uncertainties associated with implementing a dosimetric calibration are detailed in each of the CoPs. These indicate that the uncertainty in the calibration of a secondary standard chamber is ±1.5% at the 95% level for MV and MeV beams and ±3% (1 standard deviation) for kV beams.

Uncertainties for each audit are provided in the audit reports supplied to the host centre. These vary between audits but are approximately as follows. The standard uncertainty for MV beam quality is 0.15%, and intercomparison is 0.4% (neglecting common systematic errors). The uncertainty in beam
output is then generally seen to be within the uncertainty of setting the focal distances and so this value varies with the spread of the two results which are averaged. This is of the order 0.1-0.2%.

3.2 Results of NPL audits

In general, an ‘audit result’ will refer to the measurement of the beam output as this is in most cases considered the most crucial aspect of the audits in a clinical situation. The results are split into the different measurements within the audit. Beam output results are discussed in section 3.2.1. The beam quality results are then presented in section 3.2.2 and results of ionisation chamber intercomparisons in section 3.2.3. Following this a concluding discussion is given in section 3.3. Each section is further split by beam type as each follows a different calibration chain and different measurement techniques are used.

A summary of the audits visits included in the analysis is given in Table 3.1. Of the 100 total visits, 56 were for MV beams. This is unsurprising as these are most commonly used clinically. In terms of the number of beams audited the MeV beams were most common with a total of 142 beams whereas only 99 MV beams were audited. Most linacs will have two or three MV beams, but often five or more MeV beams, and so often both low and high energy beams are checked during a visit to audit MeV beams. Fewer centres offer kV treatments and so accordingly fewer audits were performed on these beams. A total of 34 kV audits were undertaken in 13 visits.

<table>
<thead>
<tr>
<th>Centre Type</th>
<th>NHS</th>
<th>Private</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MV</td>
<td>47</td>
<td>81</td>
<td>9</td>
</tr>
<tr>
<td>MeV (1996)</td>
<td>6</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>MeV (2003)</td>
<td>17</td>
<td>84</td>
<td>8</td>
</tr>
<tr>
<td>MeV (All)</td>
<td>23</td>
<td>98</td>
<td>8</td>
</tr>
<tr>
<td>kV</td>
<td>12</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>All Beams</td>
<td>82</td>
<td>209</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 3.1: Summary of NPL audit visits during the period 1994 to February 2015. To allow comparison, data from the two MeV CoPs has been included both separately and combined.
3.2.1 Beam output results

A summary of the beam output results for all beams including the mean, standard deviation, minimum and maximum differences is given for these audits in Table 3.2. There were no statistically significant differences between the modalities observed apart from the results from regional group 9 which are discussed in sections 3.2.1.1 and 3.2.1.2.1. A mean difference of < 1% (<0.5% for absorbed dose to water based CoPs) between the Host and NPL for the measured values of output was found. The standard deviation for all beams was 0.89% (range 0.41% - 0.98% for each beam type). A plot of all the measured output ratios is given in Figure 3.2 which has been split by audit region and beam type. It is noted that there were no MeV audits completed within audit region 4, no kV audits in region 9, and there was a large range in the number of audits performed between the different regions (range of 11 to 60). All MV results were within ±2%. All but six of the MeV results were within ±2%. Three of these were measured during the same audit visit on the same linac and the other three were all from the same regional group (group 9).

Figure 3.2: Measured beam output ratios separated into the eight individual regional audit groups. Mean value for each beam type in each group is indicated by a horizontal line (---).
<table>
<thead>
<tr>
<th>Beam Type</th>
<th>No. Beams</th>
<th>Mean Difference (%)</th>
<th>Standard Deviation (%)</th>
<th>Min (%)</th>
<th>Max (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MV</td>
<td>99</td>
<td>-0.06</td>
<td>0.68</td>
<td>-1.6</td>
<td>1.99</td>
</tr>
<tr>
<td>MeV (1996)</td>
<td>14</td>
<td>0.75</td>
<td>0.41</td>
<td>0.1</td>
<td>1.70</td>
</tr>
<tr>
<td>MeV (2003)</td>
<td>128</td>
<td>-0.26</td>
<td>0.98</td>
<td>-2.4</td>
<td>2.7</td>
</tr>
<tr>
<td>MeV (All)</td>
<td>142</td>
<td>-0.16</td>
<td>0.98</td>
<td>-2.4</td>
<td>2.7</td>
</tr>
<tr>
<td>kV</td>
<td>34</td>
<td>0.37</td>
<td>0.91</td>
<td>-2.4</td>
<td>2.1</td>
</tr>
<tr>
<td>All Beams</td>
<td>275</td>
<td>-0.06</td>
<td>0.89</td>
<td>-2.4</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Table 3.2: Summary of all absolute beam output measurement results from NPL audits carried out between 1994 and February 2015. The two MeV CoPs are separated for comparison.

The most common MV beams audited were 6 and 10MV which accounted for 82 of the 99 MV results. MeV beams which had more than 10 audits performed were 4, 6, 9, 12, 16, and 18MeV with 6MeV being most commonly audited (29 of the 142 MeV audits).

Initially NPL’s role was to perform an audit within each region annually. Each region was free to determine how it would spread these audits amongst its local centres. This link combined with the regional audits would ensure consistency across the UK. Most regional groups opted to spread the NPL audits across the different centres. The Scottish and Northern group instead generally would have the same centre audited each year. This results in a greater number of measurements performed at that centre and this is seen clearly (for MV and MeV beams) in Figure 3.3.

For the other regional groups the audits are fairly evenly spread amongst the groups, and it is not thought that any differences in regional results would arise due to the spread of centres audited due to the consistency observed within the regional groups themselves.
Figure 3.3: Geographical spread of the number of individual beams audited by NPL at each centre. Note the large number of beams audited at a single Scottish centre. The location of NPL is indicated by the black cross (x).

3.2.1.1 MV beam output

The distribution of results is shown in Figure 3.4. Results have a mean difference of -0.06% between the NPL and host measured values with a standard deviation of 0.68%. All results were within the ±2% action level applied to measurements. The results are normally distributed (p>0.05 based on the Shapiro-Wilk test (Shapiro and Wilk, 1965) implemented within the Scipy statistics module (Jones, Oliphant and Peterson, 2001)).

The MV results are all based upon the same 1990 CoP with the same core group of staff performing the audits over the entire data period. It is interesting to examine any changes in the results over this time period.

An ANOVA and subsequent post-hoc analysis performed on the beam output measurements of each regional group indicates that group 9 has a statistically significant difference in the mean measured result (p<0.05 with Bonferroni correction). The mean of the groups 1-8 is +0.04% and the mean of group 9’s result is -0.63% (difference of 0.67%). This difference is within 1% and generally considered acceptable and so all results are included within this analysis.
Figure 3.4: Plot of distribution of MV beam output audit results. The results are normally distributed with mean difference between host and NPL measurement of -0.06% and standard deviation of 0.68%.

3.2.1.1.1 MV audit results variation with time

Figure 3.5 shows the results of the MV audits conducted by NPL in order of date completed with a superimposed rolling standard deviation of the previous 20 audits.

On these plots the rolling standard deviation of the previous 20 results shows that the variation in the MV results decreased over time. After the first 20 audits the standard deviation of results was approximately 0.8% which has reduced to around 0.4% through the period of this analysis. The variation over time is shown in Figure 3.6 with a linear fit to the data. The standard deviation is seen to decrease by 0.3% per decade. The Spearman’s correlation coefficient is -0.881 (p<0.001) with $R^2$ of 0.785 indicating a strong correlation.
Figure 3.5: Plot of MV output results in order of completion. The standard deviation of the previous 20 audits is shown by the solid line. Bars indicate the standard uncertainty as given in the NPL report of the audit produced for the department. Colour differentiates each regions result.

Figure 3.6: Rolling standard deviation of NPLs MV audit results plotted against audit date. Rolling standard deviation calculated from previous 20 audits. A linear fit gives a decrease in standard deviation of 0.3% per decade with Spearman’s correlation coefficient of -0.881 (p<0.001) and $R^2 = 0.785$ indicating strong correlation.
The demonstrable improvement over time is an indication of the success of this CoP and the consistency achieved across UK radiotherapy centres. This improvement over time is mirrored in the results of the group E’s regional audits (Palmer et al., 2011) discussed in section 2.2.2.

3.2.1.2 MeV beam output
A total of 142 MeV beams were audited and had their beam output measured by NPL. The distribution of results is shown in Figure 3.7. The mean of all result was -0.16% between host and NPL measured values with a standard deviation of 0.98%. There were no results available for MeV audits completed within regional group 4.

![NPL MeV beam output results](image)

**Figure 3.7:** Plot of the destitution of the 142 MeV beam output audit results. All results were within ±3% with 6 outside of the ±2% action level usually used.

Figure 3.8 shows the audit results along with a rolling standard deviation of the previous 20 audits for MeV beams in order of date completed. Examining the beginning and end of the measurement period the standard deviation is approximately the same between 0.5-0.7%. There were large variations observed in the rolling standard deviation over time with a maximum at 1.4%. This was a period where there were three audits conducted at the same centre which were outside of the +2% action level. This was then followed closely by a number of results from regional group 9 which were grouped around the lower -2% action level. Removing group 9 results (see section 3.2.1.2.1) produces a much more consistent standard deviation over time at around 0.6%.
Analysis of NPL audits

Figure 3.8: Plot of MeV output results in order of completion. The standard deviation of the previous 20 audits is shown by the dashed line. Bars indicate the standard uncertainty as given in the NPL report of the audit produced for the department. Colour differentiates each regions results.

Figure 3.9: Boxplots of the measurement results of NPL MeV audits separated by nominal beam energy. No statistically significant difference was observed between the mean of results from each energy. Note this does not include the results from regional group 9. See Appendix D for description of boxplot display. The whiskers show the total extent of the data in this case.
Analysis of NPL audits

There was no significant difference in measurements seen between energies of any modality and a plot of the results showing no significant trend with MeV beam energy is given in Figure 3.9.

As well as grouping by region, the results have been separated into NHS and private centres to ensure there is consistency between patients treated in different healthcare systems. A stacked histogram of the MeV results is presented in Figure 3.10.

![NPL MeV beam output results](image)

**Figure 3.10:** Histogram of the NPL MeV audit results with results from NHS (blue) and private (orange) centres. An abundance of private centre results is seen between -1 and -2%. NHS results are normally distributed with a mean of approximately 0%.

Examining the NHS and private results reveals a greater abundance of results between -1 and -2% for the private centres. On closer examination this discrepancy can be attributed to the results arising from regional group 9 and is discussed further in the following section.

3.2.1.2.1 Variation observed in the beam output measurements of regional group 9

The MeV beam output results are displayed as box and whisker plots in Figure 3.11 for each regional group. Visually group 9 appears to be an outlier when compared to groups 1-8 (group 4 had no MeV results). Group 9 is the only group with a mean and median result outside of ±1% and its interquartile range does not intersect the expected beam output difference of 0% by a large margin.
Figure 3.11: Box and whisker plots of the NPL MeV beam output results. Group 9 appears visually as an outlier and has the only mean and median outside of ±1% despite having a similar IQR to other centres. A statistical analysis also indicates a significant difference in the mean of the group 9 results (p<0.05). See Appendix D for description of boxplot display. The whiskers indicate the entire range of results in this case.

A numerical summary of the results for groups 1-8 and group 9 alone is given in Table 3.3. The difference in the mean of all results from group 9 and groups 1-8 is 1.44% and this is statistically significant (p<0.05 following ANOVA and post-hoc analysis with Bonferroni correction applied). The standard deviation is the same at 0.73%. This discrepancy identified within the data is the subject of an investigation by NPL. If the dose discrepancy does exist then the radiobiological modelling described in this work could be used to assess the impact for specific clinical situations.
Analysis of NPL audits

### NPL audit radiation output measurements summary for regional group 9, and groups 1-8

<table>
<thead>
<tr>
<th>Beam Type</th>
<th>Groups 1 - 8</th>
<th>Group 9 Only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. Beams</td>
<td>Mean Dif. (%)</td>
</tr>
<tr>
<td>MV</td>
<td>83</td>
<td>0.05</td>
</tr>
<tr>
<td>MeV (’96)</td>
<td>14</td>
<td>0.75</td>
</tr>
<tr>
<td>MeV (’03)</td>
<td>84</td>
<td>0.20</td>
</tr>
<tr>
<td>MeV (All)</td>
<td>98</td>
<td>0.27*</td>
</tr>
<tr>
<td>kV</td>
<td>34</td>
<td>0.37</td>
</tr>
<tr>
<td>All Beams</td>
<td>215</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Table 3.3: Summary of all absolute beam output measurement results from NPL audits carried out between 1994 and February 2015 for groups 1 – 8 and group 9 separately. There is a statistically significant difference (p<0.05) between the mean difference for MeV measurements as measured by group 9 when compared to other groups. The results compared are highlighted with an asterisk (*).

With the same staff group and NPL equipment used across all groups the most likely cause is some aspect of the host centres’ equipment, setup or methodologies. From discussions with NPL dosimetry staff the most likely cause is the use of a different type of solid water within this group for performing measurements. The solid water houses a Roos type ionisation chamber when measurements are taken and this is used across the group. The water equivalency may differ in some way to that of the solid water used by NPL (however this has currently not been shown to definitively be the case) or the chamber holder manufactured within the solid water is of a slightly different size. This may cause the ionisation chamber to be slightly compressed causing a change in sensitive volume and thus changing the magnitude of charge collected.

This observed statistically significant difference emphasises the strength of looking at an entire set of audit measurements together, rather than simply examining the individual results in isolation. The MV results for group 9 were also seen to be statistically different; however those results were deemed to be within the acceptable range, but could be caused through the same systematic effect present.
3.2.1.2.2 Comparison of 1996 and 2003 MeV codes of practice

For the electron beams the measured output ratios obtained from audits following the 1996 and 2003 MeV CoPs have been compared. With the different CoPs an improvement in measurement consistency was expected, and so unequal variance in the measurement sets has been assumed. In this case a Welch’s t-test was used to perform the comparison. A summary of the beam output results taken with each CoP (with group 9 excluded from the analysis) is given in Table 3.4.

<table>
<thead>
<tr>
<th>MeV CoP</th>
<th>No. Beams</th>
<th>Mean difference (%)</th>
<th>Standard Deviation (%)</th>
<th>Min (%)</th>
<th>Max (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>14</td>
<td>+0.75</td>
<td>0.42</td>
<td>+0.10</td>
<td>+1.65</td>
</tr>
<tr>
<td>2003</td>
<td>84</td>
<td>+0.20</td>
<td>0.75</td>
<td>-1.50</td>
<td>+2.70</td>
</tr>
</tbody>
</table>

Table 3.4: Summary of data used in comparison of 1996 and 2003 MeV CoPs. The data from group 9 was excluded from this analysis to avoid skewing results with potentially unreliable data. A statistically significant change in the mean difference of 0.55% as measured with each CoP was observed (p < 0.01) using Welch’s t-test.

A statistically significant difference (p<0.01) between the mean of beam output ratios was found between each CoP. A change in the means of 0.55% was observed (see Table 3.3) with the 2003 CoP being closer to the desired value.

3.2.1.3 kV beam output

A total of 34 kV beam output audit measurements were taken by NPL and these are shown plotted in order of completion in Figure 3.12 and a histogram showing the distribution of results is given in Figure 3.13. There were no kV audits performed in regional group 9 as they did not have any kV equipment.
Figure 3.12: Plot of kV beam output results in order of completion. There was a total of 34 beams audited during the data period. No significant variations over time or between groups were observed.

Figure 3.13: Plots of distribution of number of kV results of beam output audits. All results were within ±3% and approximate a normal distribution even with a reasonably small sample.
Analysis of NPL audits

The energies audited ranged from 30kV to 300kV. The most commonly measured energies were 80kV which had 5 beams audited and 300kV which had 4 beams audited. No significant difference was identified between beams within any specified energy range.

3.2.2 Beam quality results

As part of the audits beam quality was often measured, however occasionally host centre values were used and so this data is not included. A summary of the beam quality measurement is shown in Table 3.5 demonstrating good consistency between centres.

<table>
<thead>
<tr>
<th>Beam Type</th>
<th>Measure of Beam Quality</th>
<th>No. Results</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>MV</td>
<td>TPR$_{20,10}$ (NPL : Host)</td>
<td>99</td>
<td>-0.13</td>
<td>- 1.5%</td>
<td>+ 1.2%</td>
</tr>
<tr>
<td>MeV</td>
<td>$R_{50,0}$ (NPL - Host)</td>
<td>105</td>
<td>- 0.02 cm</td>
<td>- 0.28 cm</td>
<td>0.3 cm</td>
</tr>
<tr>
<td>kV</td>
<td>HVL (NPL : Host)</td>
<td>30</td>
<td>+ 0.24%</td>
<td>- 12.7%</td>
<td>+ 7.7%</td>
</tr>
</tbody>
</table>

Table 3.5: Summary of NPL beam quality measurements as measured during audit visits. MV are shown as percentage difference of TPR, MeV are shown as difference in the depth of $R_{50,0}$ and kV are shown as percentage difference of HVL as a range of energies is included.

It was not possible to give measurement specific uncertainty values for individual measurements performed by the host centre based on the audit reports and data available.

3.2.2.1 MV beam quality

For the MV results, the NPL measurements have a standard uncertainty of ±0.2%, equating to a dose difference of less than 0.1%. The variations in measurements of the beam quality give rise to less than 0.5% variation in energy for MV beams. The measurement of TPR$_{20,10}$ determines the calibration factor used to then determine the absorbed dose. The variation of calibration factor with beam quality is relatively small for MV beams as measured with a secondary standard NE2611 type chamber. The calibration curve is a 4$^{th}$ degree polynomial to convert the measured beam quality into a calibration factor:

$$\text{Calibration factor} = a + b(TPR) + c(TPR)^2 + d(TPR)^3$$  \hspace{1cm} (3.1)

where a, b, c and d have approximate values: a=15.7×10$^7$, b=-27.7×10$^7$, c= 47.6 x 10$^7$ and d=-27.5×10$^7$ and TPR ranges between 0.568 to 0.800 for Co-60 and 25MV beams (Thomas, 2008). In this case
using equation (3.1) a change of 2% in measured beam quality for a TPR$_{20,10}$ of 0.668 (6 MV), 0.739 (10MV) and 0.763 (15MV) beam gives rise to a change in calibration factor of 0.13%, 0.38% and 0.5% respectively. From the NPL audit results all measured differences in beam quality result in a difference of less than 0.4% in the dose delivered due to calibration.

### 3.2.2.2 MeV beam quality

The variations in measurements of the beam quality give rise to less than 0.5% variation in energy for MeV beams. As for the MV beams, the measured output is relatively insensitive to small changes in beam quality measurement for MeV beams. The maximum difference was 0.3cm in R$_{50,0}$.

### 3.2.2.3 kV beam quality

It should be noted that for the kV results, the NPL measurements have a standard uncertainty of ±2%, however this equates to a dose difference of less than 0.1% in the calculated calibration factor. The kV beams show the largest variation of all beam types, however due to the nature of the beams the variation in dose is small and within the clinically acceptable range. This includes kV beams which have the largest uncertainty and largest measurement deviations. These beams deliver dose over a much shorter range, and the planning and setup in general carries a greater uncertainty than MV beams which are used for a majority of clinical treatments.

### 3.2.3 Ionisation chamber calibration intercomparison results

A summary of the ionisation chamber intercomparison results from all beam types is presented in Table 3.6. These results indicate any additional error involved in transfer of the calibration from a centre’s secondary standard instrument to a field chamber.

<table>
<thead>
<tr>
<th>Beam Type</th>
<th>No. Results</th>
<th>Mean (%)</th>
<th>Standard Deviation (%)</th>
<th>Min (%)</th>
<th>Max (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MV</td>
<td>98</td>
<td>-0.03</td>
<td>0.68</td>
<td>-4.2</td>
<td>+1.2</td>
</tr>
<tr>
<td>MeV</td>
<td>111</td>
<td>+0.02</td>
<td>0.71</td>
<td>-2.2</td>
<td>+1.7</td>
</tr>
<tr>
<td>kV</td>
<td>28</td>
<td>+0.10</td>
<td>0.56</td>
<td>-0.8</td>
<td>+1.1</td>
</tr>
</tbody>
</table>

Table 3.6: Summary of NPL ionisation chamber intercomparison measurement results from audit visits.

The ionisation chamber intercomparison measurements show good agreement, with only one MV and one MeV beam result outside of the ±2% action level. These measurements give an indication of
the accuracy of the transfer within a single department of calibration between different ionisation chambers. The ionisation chambers which the calibration is transferred to are often used at a much greater frequency than the secondary standard and so in clinical practice this calibration transfer has potential to have a large impact. When re-calibration of a machine is performed, more than a single measurement is relied upon.

If a systematic error is introduced to all ionisation chambers within a department, but not related to the calibration procedure itself then there may be a direct correlation between measured differences in beam output and field chamber calibration if the beam was calibrated using the secondary standard. This relationship is explored in the following sections for each beam type. This correlation has potential to identify spurious results if both are looked at together and may allow identification of sources of error.

3.2.3.1 MV ionisation chamber intercomparisons compared with beam output

A plot of the measured output difference against the chamber intercomparison difference is given in Figure 3.14 for MV beams. Each region is highlighted with different colours. With this data the correlation has been assessed using the Spearman’s rank correlation coefficient and the corresponding $R^2$ value.

There is a weak correlation present with Spearman’s rank coefficient of 0.471 ($p<0.01$) and $R^2=0.22$. If there was no correlation then it would indicate that measurement uncertainties involved in transferring calibration dominate the differences measured during the audits. The fact that there is a possible correlation indicates there may be small systematic uncertainties present when transferring calibration, however these differences appear small and are unlikely to have a significant clinical effect.
3.2.3.2 MeV ionisation chamber intercomparisons compared with beam output

The MeV result showed a far stronger correlation between the intercomparison and the beam output differences. A plot of output vs intercomparison shown in Figure 3.15 with a linear regression line shown.

As group 9’s MeV results are known to be under investigation and may all require a correction these were removed when performing the regression analysis. The linear regression has a reasonably strong positive correlation with Spearman’s correlation coefficient of 0.687 (p<0.01) and $R^2 = 0.338$. The gradient of the fit has a value of 0.88 which is close to unity as expected. MeV can be sensitive to small changes in depth of measurement (or distance from measurement point to source) and systematic errors such as those observed may be introduced through small setup errors due to the dose gradients which may be present across the ionisation chambers used.
Figure 3.15: Plot of MeV beam results for measured output difference against intercomparison difference highlighted by regional group. Group 9 is visibly offset in comparison to the remaining results and has not been included in the linear regression. The fit has a Spearman’s rank correlation coefficient of 0.687 (p<0.01) and $R^2 = 0.338$ indicating a reasonably strong correlation.

The intercomparison results may indicate the introduction of a systematic error across calibrations within a department, however the results are excellent with a mean of 0.02% and all results within ±1.5% when group 9 is excluded. This demonstrates a good consistency in calibration transfer between measurement instruments across UK centres. There was no significant difference observed between the 1996 and 2003 CoP.

3.2.3.3 $kV$ ionisation chamber intercomparisons compared with beam output
Results from $kV$ intercomparisons are good. There was no significant correlation observed between the intercomparison and beam output measurements as shown in Figure 3.16.
3.3 Discussion of NPL audit results

A total of 100 audit visits were made during the specified time period, with 275 individual beams being audited. There were only nine results greater than or equal to the ±2% tolerance. Three of these were from kV measurements, and six from MeV measurements. None were from MV beams which are by far the most commonly used during treatments. Of the MeV results which were out of tolerance, three were from group 9 and all gave low readings and so may be adjusted upon completion of the investigation into these results if it is found that the measurements made by NPL have a systematic error identified. The remaining three MeV results which exceeded +2% for the output were from a single centre during a single visit. After further investigation at the time of the audit it transpired that the host centre’s output measurements were derived using the 1996 CoP, whereas the audit was carried out using the 2003 CoP. Correcting for this reduced all but one of the
measurements to within the tolerance, with a single result remaining unresolved at +2.7%. Generally any issues identified during the audit were investigated and resolved during the visit.

The standard deviation of the measured MV output ratios reduced from approximately 0.8% to 0.4% over the period the audits were conducted (Figure 3.5 and Figure 3.6). The initial standard deviation of 0.8% was already excellent, and for this to reduce by half is testament to the success of the process of dissemination that exists in the UK for dose from the primary standard to the clinic. This is achieved through the combination of the use of the CoP, dedicated secondary standard instrumentation and UK regional audit network. NPL reference dosimetry audits also make a valuable contribution in ensuring the continued development and improvement in the traceability, accuracy and precision of dose delivered to patients receiving external beam radiotherapy.

This improvement in audit results can be compared to similar audits, both within the UK and internationally. The audit completed by Thwaites et al. (Thwaites et al., 1992) as discussed in section 1.4.3 had a standard deviation of 1.5% and these NPL audits show improvement in consistency since then. The MeV national audit conducted by Nisbet in 1996 (Nisbet and Thwaites, 1997) (see section 1.4.3) again showed a greater variation than those measured by the NPL. Audits that use dosimeters such as TLDs rather than ionisation chambers have larger uncertainties, for example the IAEA postal audits use a ±5% tolerance (Izewska and Andreo, 2000) to account for this. The improvement seen by the IAEA following repeat audits is reflected in the improvement seen in the MV beam outputs as measured by NPL which demonstrated a reduction in standard deviation from 0.8% to 0.4%.

The combined uncertainty for the secondary standard dosimeter calibration factor is ±1.5% at the 95% confidence level (Lillicrap et al., 1990). This equates to a standard deviation of approximately 0.8% which closely matches the 0.68% standard deviation measured for the MV beams (see Table 3.2). An even closer match is seen for the 2003 MeV CoP which provides an uncertainty of 1.5% at the 95% level, corresponding to a standard deviation of 0.8%, which is only 0.07% different from that measured for the MeV results following the 2003 CoP when the group 9 results are excluded for reasons discussed in section 3.2.1.2.1.

It can be seen from the analysis of the 1996 and 2003 MeV CoP data that changing from an air kerma to an absorbed dose to water based CoP has reduced the difference in dose measurements in electron beams between clinical radiotherapy centres and NPL (section 3.2.1.2.2). The mean variation of the measured NPL:Host outputs reduced by 0.55% (p < 0.01). A common misconception is that there is greatest uncertainty for low energy electron beams. There was no significant
difference observed between measured values, or their variation for different beam energies within any beam type which could be derived from this work (see Figure 3.9 for plot showing results for MeV beams).

In the UK the implementation of a new code is recommended to be taken up by users within 3 years of implementation. During NPL audit visits it has been found that this was not achieved in all cases, including two cases where the 1990 MV code and the 2003 electron code were not implemented for 10 years. On another occasion one department implemented the 2003 electron code of practice in good time but admitted they had never got round to implementing the 1996 code. It is sometimes cited that this is due to a desire to remain consistent within the department, but this is then at the detriment of consistency on a national perspective. Time pressure on staff and availability of machine time is also given as a barrier to prompt implementation. Whilst it is imperative that patient treatments are the priority, there is on occasion a lack of understanding between staff groups regarding the needs of all the professions in ensuring the delivery of best practice in patient care.

These audits have been carried out by a small group of NPL staff over a 20 year period. As a result the procedures and practices have remained consistent over that time. Further to measuring the beam quality and absolute dose, which are reported to the host centre through a written audit report, there are additional benefits worth noting such as being able to check the host’s temperature and atmospheric pressure readings against NPL’s own calibrated instruments and these on occasion have been found to be some of the larger errors introduced into the calibration.

A more subtle benefit of “on-site” audit is that there is time for informal discussion of practices, in which potential problems may be identified and these issues clarified or pre-empted. This is not easy through postal forms of audit. While not formally provided in the results, these discussions on general practice are a significant contributory factor in ensuring the robustness of the complete dosimetry chain from primary standard to patient dose delivery. These discussions also provide a valuable opportunity to observe and understand how the implementation and interpretation of recommendations have been conducted within individual departments. There is also great value in receiving feedback from the end user community regarding the clarity (or lack of) in the CoP or other recommendations regarding complexities and issues that may not have been identified when a document was written. For example the 1990 MV photon code of practice was undoubtedly significant in being the world’s first absorbed dose based protocol, further, it is often held up as an example of how simply a code can be written. However this is a double edged sword, as recommendations of reference conditions were open to interpretation, and variations therefore
exist in the adopted reference conditions particularly regarding source to detector distance and field size at the chamber. These differences in implementation may add to the uncertainties of audit measurements.

The requirement for more clinically relevant quantities to define dose has led to the development and improvement of primary standards away from air kerma based protocols to absorbed dose to water. There is a continued requirement to drive further improvement through the development of absorbed to dose to medium protocols and through primary standards that closer reflect the effect of radiation on the tissue at the cellular or even DNA level (Galer et al., 2011). For this work to have impact there must be a rigorous, consistent and thorough method of dissemination from the primary standard via dedicated CoPs and designated secondary standard instrumentation. On occasion issues have arisen between mismatched labels on equipment and calibration certificates; better practice is to check the manufacturer’s unique serial number stamped on to the instrument. As part of this dose traceability chain, audit is a crucial part in ensuring the accurate and consistent implementation of these systems.

This continued improvement in relevant quantity and the precision and accuracy of dose measurement feeds into clinical trials helping to give a clearer picture of patient outcome. With the increasing complexity of treatment techniques both from linacs and indeed particle therapy there is a need to ensure ever more consistency across departments and between techniques and modalities in ensuring clear and beneficial results from patient clinical trials. Consistency and comparability is also very relevant with the increased interest in interrogating large data sets.

Continued development of UK primary standards for radiotherapy, implementation of specific guidance through recommendations and codes of practice combined with an extensive regional and national audit network provide a foundation for the implementation of safe and consistent radiotherapy dosimetry and ultimately for patient benefit.

3.3.1 Comparison of regional and NPL audit results
The NPL audit results, those published by the south west, south east central, and additional regional data supplied were all largely similar in its variation in measured MV beam output. Both the NPL MV audits and the south east central MV audits indicate an improvement in consistency over time, and both sets of measurements have identified errors in the processes which have subsequently been rectified indicating the impact which audit can have on safety and consistency.
The measurement standard deviation of results (excluding the published south east central group results which reported only absolute deviations as described in 2.2.2) lies within the range 0.4% - 0.9%. The standard deviation measured by NPL during their audits compares well with this at 0.8% for MV beams. This measurement consistency across regional groups (for which data was available) indicates the audit network as a whole is functioning consistently.

3.3.2 Comparison to Papillon audit

Results of a dosimetric audit of the Papillon kV unit used for contact rectal brachytherapy was published in 2016 (Humbert-vidan et al., 2017). The unit is a 50kV x-ray which generates circular fields. It is estimated that electronic brachytherapy devices account for 15% of UK kV treatment units (Palmer et al., 2016) and so it is prudent to compare the results of this small audit with the NPL audits.

The audit of the Papillon machines included the four machines (all at different centres) which were clinical at the time. NPL provided dosimetry support and the audits were performed by on-site visits.

At this stage centres were still in the early stages of clinical use and techniques were developing and so this audit provided a valuable comparison to ensure consistent implementation across each centre. The audit measurements included radiation output, beam quality (the first half value layer, HVL₁), radiation field size and uniformity.

The range of output differences was -0.8% to +2.2%. Beam quality varied by up to 6.5% (0.04mmAl). The beam quality differences only amount to a difference in delivered dose of 0.2% (Humbert-vidan et al., 2017).

The results of this audit largely match those of the NPL kV audits, however there were only 4 results in this case, so statistical comparisons are not possible. The audit indicated consistent implementation of a new treatment technique with calibration following the CoP applied to within accepted limits.

3.3.3 Comparison to national rotational audit results

The national rotational IMRT audit (Hussein et al., 2013; Clark et al., 2014) was conducted in the UK and included all centres which were already clinical or ready to deliver clinical treatments in March 2013. The NPL, RTTQA and IPEM were all involved in the setup of the audit as it also acted to provide credentialing of institutions for use of rotational IMRT techniques in clinical trials. The audit involved
each centre creating two treatment plans, one of which was common to all centres with pre-
delineated volumes to allow comparison. These plans would then be delivered to the same phantom
at each centre (the PTW Octavius II with seven29 array) and the measured dose planes compared to
the TPS at six different points spanning different dose levels. A total of 413 point dose differences
were calculated and 34 centres participated in the audit.

As well as the gamma pass rates examined extensively within the publication ionisation chamber
measurements were taken at a specified point within the phantom. Alongside these a beam output
was measured to allow correction of the ionisation chamber measurements. It is this aspect which
will be explored briefly here.

The beam output was measured at all centres and at 12 centres ionisation chamber measurements
were taken at a specific point within the phantom. The results are not statistically different from a
normal distribution and a summary of the results is given in Table 3.7.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Beam output (%)</th>
<th>Ionisation chamber measurement in phantom (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>uncorrected</td>
</tr>
<tr>
<td>Mean</td>
<td>+0.29</td>
<td>0.11</td>
</tr>
<tr>
<td>Max</td>
<td>+2.0</td>
<td>2.52</td>
</tr>
<tr>
<td>Min</td>
<td>-1.84</td>
<td>-3.19</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.85</td>
<td>1.87</td>
</tr>
<tr>
<td>No. of measurements</td>
<td>34</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 3.7: Summary of beam measurements taken during the national rotational IMRT audit. The
standard deviation of results more than doubles when measuring with an ionisation chamber
within a phantom when compared to reference conditions. There is no statistical difference
between the corrected and uncorrected phantom measurements (p>>0.05)

The variation seen in the beam output measurements matches closely the regional audit results
(approximately 0.8% standard deviation) and this is largely to be expected as very similar protocols
would have been followed. When measuring within the phantom as part of the delivered plan
however the standard deviation is approximately doubled. This increase in variation gives an insight
into the increase in variation of delivered dose when moving from reference conditions with simple
fields to more complex, clinically realistic plans. In this case assuming a normal distribution and
standard deviation of 1.7% this would result in 5% of measurements exceeding a 3.3% deviation.
Note this is based upon measurements taken within a phantom and the inclusion of a patient would further increase the uncertainties. These values can be used to give an indication of the possible variation in dose which may be expected in a more clinically realistic situation.
4 Machine output measurement

Radiotherapy machines are calibrated to deliver a known dose under a set of standard conditions and the variation in this was assessed in chapter 3. This chapter analyses the additional variation in beam output following the initial calibration as described in section 1.3.2 and there has been no large multi-centre study of these daily measurements to date. Previous work examining the variation in beam output data over time has focussed on an individual centre’s treatment machines (Hossain, 2014; O’Daniel and Yin, 2017), whereas this work aims to give a broad representation of the multi-centre variation in machine beam output measurements. The presence of trends over time, variation between measurement devices, and recording methods were investigated. This work acts as a benchmark for the current variation in dose due to variations in beam output.

This chapter builds upon the paper published in the PhiRO journal during this work (Bolt et al., 2017), expanding on a number of areas within the publication and including a more complete analysis of the use of different measurement devices.

4.1 Data collection and formatting

A request for data was sent by email to all UK radiotherapy centres through the regional audit network (Thwaites et al., 2002). The data requested was for 6MV beam output data covering the 6 month period from January 2015 to June 2015. This time period was chosen as a balance between being a manageable dataset for centres to collate, and being representative of variations in measurements seen over time. Data requested included; treatment machine model, machine install date, measurement device, data recording method, and measured beam output. A quantitative statistical analysis of the data was performed to evaluate the variations which exist in measurement of beam output.

Statistics including the mean output were calculated for each treatment machine for different measurement devices allowing comparison of delivered dose between machines. The Radiotherapy Services in England 2012 report (National Radiotherapy Implementation Group (NRIG), 2012) states that there were 265 linacs in use across 58 centres. Based on this it was estimated that there was approximately 300 linacs in operation at the time of data collection for this work.
As previously the beam output is represented as a percentage deviation from initial calibration. To remove outliers, results greater than ±5% were excluded. Centres would generally not continue treatment if output deviated to this extent (Palmer, Kearton and Hayman, 2012), and a majority of these values were actually erroneous results as stated by the centres during submission. In total 59 data points (0.22%) were removed from the dataset (53 of these were over 10% and were most often accompanied by repeat measurements which were within the normal range, thus indicating erroneous results), leaving a total of 24,501 measured beam outputs to be included in the analysis.

4.2 Data analysis

The data has been grouped and analysed in a range of ways as described in the following sections. These grouping and associated analysis allow the impact of different aspects of the radiotherapy setups to be assessed separately to see how they may affect the delivered dose.

4.2.1 Data recording method

Data was received in a variety of formats including paper records (scanned or manually transcribed), database systems (in-house or commercial) and spreadsheets, with many centres using a mix of recording techniques. A common method was to transcribe paper records into an electronic system to allow trend visualisation and additional analysis. There were eight centres which solely relied upon paper records for beam output checks.

4.2.2 Grouping by machine age

Using the install year of the machines any variation in output measurement associated with this has been investigated. This was particularly focussed on whether there was any significant difference observed for machines deemed to be beyond their optimal lifespan which is generally considered to be 10 years (NHS supply Chain, 2015; European Coordination Commitee of the Radiological Electromedical and Healthcare IT Industry, 2016).

4.2.3 Grouping by centre size

The centres were designated as small, medium or large based on the lower and upper quartiles of the total number of centres housing a particular number of treatment machines. This resulted in approximately even sized groups ensuring meaningful comparisons could be made. A small centre contained 1 or 2, medium had 3 or 4, and a large centre had 5 or more machines. For the purposes of this work satellite centres were treated separately from their parent centre.
4.2.4 Drift in beam output with time

It is known that the beam output will tend to drift over time, often at differing rates dependant on the age of the installed in-head ionisation chamber. In many cases it was observed that a treatment machine had been recalibrated which was clear from step changes within the data (see Figure 4.4 for an example). A step change in the data usually indicates that either the treatment machine or the measurement device was recalibrated. To allow assessment of the drift these step changes were removed producing a ‘corrected’ set of data. The data was corrected by subtracting the magnitude of the identified step change from all subsequent data for that machine.

The observed variation in beam output over the six-month period was assessed for each treatment machine and results were linearly extrapolated to give the annual trend. The results have then been grouped by manufacturer and treatment machine age (in-head ionisation chamber age was generally not known for this dataset).

4.2.5 Measurement device comparison

Often a range of measurement devices are used to track the machines’ beam output (Mayles et al., 1999; Palmer, Kearton and Hayman, 2012) with constancy devices used daily\(^4\) and a Farmer chamber used less frequently (weekly or monthly). Where data was supplied for both constancy device and ionisation chamber (Farmer type chambers were the only types used within this dataset) measurements the mean output as measured by each of these devices has been compared for each machine as well as the spread of measurement results. The variation in measurements taken between each device has been used to determine the reliability of each device in monitoring the beam output.

4.2.6 Statistical analysis

Where appropriate results were grouped as detailed above to allow comparison. Results presented are from data which was normally distributed and so comparison was performed by means of a t-test where appropriate. When repeated t-tests were performed a Benferroni correction was used to adjust for multiple comparisons as appropriate.

4.3 Results

The variation of delivered dose to patients due to variation in beam output following initial calibration is presented here and is grouped into a variety of sections as described above.

\(^4\)A single centre which supplied data stated they did not measure output on a daily basis.
4.3.1 Summary of collected data

Data was received from 204 treatment machines from 52 centres (see Figure 4.1). Machines included Varian (n=96), Elekta (n=92), Siemens (n=12), Tomotherapy (n=3) and Cyberknife (n=1).

The data included 41 National Health Service (NHS) centres (190 treatment machines) and 11 privately funded radiotherapy centres (14 treatment machines). Measurement frequency varied between daily and monthly and most commonly results were recorded in a mix of paper based and electronic formats. A discussion the variation in recording method is given in section 4.3.4.

Figure 4.1: Locations of the 52 radiotherapy centres which supplied beam output measurement data for this work showing a wide geographic spread of data sources.

The number of beam output measurements exceeding ±1% was 5947 (24.3%) with 47 (1.9%) exceeding ±2% and 45 (0.2%) exceeding ±3%. The measurements were normally distributed with a mean of 0.0% and a standard deviation of 0.8%. This distribution of measured machine output is given in Figure 4.2.
Figure 4.2: Distribution of measured machine output. Data has a mean of 0.0%, with a standard deviation of 0.8%. There are 75.7% of measurements within +/- 1%.

A normal probability plot ("Q-Q" plot) has been used to assess the normality of the distribution of beam outputs. Quantitative statistical tests are extremely sensitive to very small changes with a dataset this large, and so this method is more appropriate. The probability plot is shown in Figure 4.3.

The probability plot indicates that the data is normally distributed and so for the remainder of this thesis a normally distributed data set is assumed.
4.3.2 Typical beam output data

Described here is one set of beam output measurements from a single machine which demonstrates typical aspects observed within the data. A plot of the results is presented in Figure 4.4. In this case a constancy device and Farmer chamber are used daily and weekly respectively.
Figure 4.4: Typical set of beam output data showing weekly Farmer chamber and daily constancy measurements for a single linac. This dataset includes a period of upward trend from March 2015 to June 2015 followed by a calibration.

In this case the Farmer and constancy device measurements are considered well matched with less than 0.5% difference in mean results between the measurement devices. Trends observed in this data are a period of slightly unstable output measurements followed by a period of upward drift between March 2015 and June 2015. This is then followed by a calibration (in this case it is known that the machine was recalibrated) and the output is measured at around 0% for the last couple of weeks in June.

4.3.3 Distribution of treatment machine install years

The installation date was provided for 187 treatment machines. The spread of installation year for the treatment machines included within this work are shown in Figure 4.5 separated by manufacturer. There were 47 machines which were older than 10 years at the time of data collection which is beyond the recommended lifetime (NHS supply Chain, 2015; European Coordination Commitee of the Radiological Electromedical and Healthcare IT Industry, 2016).
Machine output measurement

Figure 4.5: Distribution of treatment machine installation year for data collected. The oldest machine was installed in 1999 making it 16 years old at the time of data collection. A total of 204 treatment machines were included in the dataset; 96 Varian, 92 Elekta, 12 Siemens and 4 Accuray (3 Tomotherapy and 1 CyberKnife).

No statistically significant differences in measurements of beam output were identified with machine age in this work, however often the age of the machine is not the main driver behind machine replacement. Advancement of other technologies such as imaging advances will often drive decisions on machine replacement rather than fundamental changes in the beam production. Also the primary monitor ionisation chamber age was not known for almost all machines and this is likely to have the greatest effect on the output stability than the age of the machine itself.

4.3.4 Variation in recording methods

A variety of recording methods are used across different centres. When the data was collected many centres were in the process of developing ‘paperlight’ QC techniques, and so it would be interesting to re-assess the distribution of recording methods in the future as a comparison. Many centres included within this work were completely dependent on paper records for beam output measurements. In Figure 4.6 the spread of recording methods split by centre size is shown.
The only notable difference was that small centres were least reliant on paper only records. However, it is also noted that that these centres had often been setup more recently. This is likely to contribute to the increase in electronic recording methods as no transition from paper records would have been required. Small centres may also be able to more flexibly adapt their processes due to the decreased burden of staff training and fewer records to transfer between systems.

### 4.3.5 Drifts in beam output

Within this 6-month data set 35 treatment machines were seen to have a single calibration and 7 treatment machines had two calibrations performed during this period. After correction for calibration a linear regression was performed on the data for each machine. This was linearly extrapolated to give the predicted annual drift. The overall observed drift in beam output for all treatment machines was $+0.9\%$/year. The 5th and 95th percentiles were $-2.2\%$ and $+5.1\%$ respectively indicating the large inter treatment machine variation possible.
No significant difference was seen at the 95% level (p>0.05) between the beam output trends of different manufacturers or different models. A summary for each manufacturer is given in Table 4.1.

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Number of machines</th>
<th>Mean drift (%/year)</th>
<th>Median drift (%/year)</th>
<th>Drift standard deviation (%/year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varian</td>
<td>96</td>
<td>1.22</td>
<td>0.78</td>
<td>2.27</td>
</tr>
<tr>
<td>Elekta</td>
<td>92</td>
<td>0.71</td>
<td>0.44</td>
<td>2.03</td>
</tr>
<tr>
<td>Siemens</td>
<td>12</td>
<td>-0.06</td>
<td>0.98</td>
<td>3.31</td>
</tr>
<tr>
<td>Tomotherapy</td>
<td>3</td>
<td>0.72</td>
<td>-0.26</td>
<td>2.11</td>
</tr>
<tr>
<td>CyberKnife</td>
<td>1</td>
<td>-1.06</td>
<td>-1.06</td>
<td>N/A</td>
</tr>
<tr>
<td>All</td>
<td>204</td>
<td>0.89</td>
<td>0.56</td>
<td>2.27</td>
</tr>
</tbody>
</table>

**Table 4.1:** Summary of the measured drift of beam output over time for each manufacturer after correction for calibrations as determined using linear regression. There was no statistically significant difference observed between manufacturers at the 95% level (p>0.05).

A previous study of three Varian linacs by Hossain (Hossain, 2014) concluded an annual trend in output of 2-4%, and this compares reasonably with that found for the Varian machines in this study of 1.2% (± 2.3%)/year (1 SD). A study by Grattan (Grattan and Hounsell, 2011) indicated that the variation may be model specific and may vary with age. No statistically significant differences with machine age or model were found within this dataset, however longer term data would provide a more robust insight. As not all details of calibrations of machines and devices were known in this dataset it was not possible to apply reliable corrections to constancy device measurements and so the data was used as it was supplied. It is however felt that the large volume of data will reduce any influence of mis-calibrated measurement devices to some extent. If a similar study was repeated, it would be beneficial to obtain details of both machine and device calibrations.

### 4.3.6 Variation in centre size

The data contained 20 small, 13 medium and 19 large centres. 11 centres contained a single machine and the largest centre housed 12 machines. The spread of centre sizes is presented in Figure 4.7.
Figure 4.7: Variation in the number of treatment machines at each centre. Centres operated by private radiotherapy providers are generally smaller, with most having a single treatment machine.

The data is separated into NHS and privately-operated centres and indicates that private centres are more likely to be smaller in size, most commonly having only a single treatment machine. Only two NHS centres had a single machine, whereas only 2 private centres had two or more machines.

4.3.7 Inter machine variation

In many radiotherapy centres patients will often be assigned to a single treatment machine. This might be due to technical limitations of the machines, or to improve the experience for the patient who will become familiar with the staff and surroundings through their treatment. When this is the case it is worth considering the case of the range of doses which may be delivered due to machine selection and also the start date and duration of the treatment which is discussed in section 4.4.4.

4.3.7.1 Mean machine output

The mean output of each machine over the data collection period ranged between a maximum of +1.6% and a minimum of -2.1% with an overall mean of 0.0%. The 5th and 95th percentiles were -1.1% and +0.9% respectively. A plot of the mean output for each machine is presented in Figure 4.8.
Figure 4.8: Plot showing the mean output from each treatment machine over the data collection period plotted in order of mean beam output. The machines from the centre with the greatest intra-centre variation are highlighted and have a range of 2.2% (range 0.7% to 1.5%). This centres machines showed no statistically significant difference between measurement devices used (p>0.05) and in the case of the highlighted machine two separate types of constancy device were in use.

It was perhaps surprising to observe a centre with a linac with a mean output over a six-month period of -2.1% considering the most common action level used was ±2%. This particular machine did not have a calibration performed during the data period.

Figure 4.8 also has highlighted the machines from the centre which exhibited the greatest range of mean outputs between its machines. The variation in outputs within a centre is explored further in the next section.

4.3.7.2 Variation in mean dose over patient treatment course

The mean dose for each machine above was determined based on the complete 6-month dataset. As a typical patient treatment (for treatments modelled in this thesis) may be around 6-7 weeks in length, the above analysis was also repeated for shorter timescales as this is likely to increase the
possible variation. The standard deviation in the mean dose from each machine has been determined for time periods from 1 day to 6 months to visualise this variation. A plot showing the results is presented in Figure 4.9.

**Figure 4.9:** Plot showing the variation in mean dose (standard deviation) delivered by each machine for a range of time periods from 1 day to 6 months. It is seen that for typical treatment length of ~6 weeks the standard deviation is 0.68%.

For this work the typical treatment length is 6-7 weeks. Based on the results above, a standard deviation in the mean delivered dose is taken to be 0.7% and assuming a normal distribution this results in a 95th percentile of 1.4%.

### 4.3.7.3 Inter centre variation

One aspect which affects the dose delivered to individual patients is the exact machine which they are treated on. The mean beam output has been calculated for each treatment machine and then the range of values within each of the 41 multi machine centres has been determined. These results are presented in Figure 4.10.
Figure 4.10: Plot showing the range of the mean outputs recorded on different treatment machines within each centre. The greatest difference in the mean output between machines at a single centre was 2.2%. There were 4 centres which had a difference of greater than 2%.

Within a single centre the greatest difference in mean output was 2.2% (range -0.7% to +1.5%). The values for this centre are highlighted in Figure 4.8 which shows the mean output for each machine. In the case highlighted, two different types of constancy device were used across the machines, however there was no statistically significant difference between measurements made with the devices when compared with ionisation chamber measurements. An intra-centre variation of > 2% was observed for 4 of the 52 centres (7.6%), and a variation of > 1% for 23 centres (44.2%). The mean intra-centre variation was 0.87%.

Generally, the larger the centre the larger the range of mean outputs between its machines, however the centre with the greatest range was not the largest and the maximum range within a two machine centre was 1.6%.

It should be noted that the data captures a snapshot in time, and so the distribution of ranges within centres is representative of the overall picture, the centres with the greatest ranges between machines will change with time as machines are re-calibrated.
4.3.8 Comparison of Farmer chamber and constancy device measurements

As a range of devices was used to measure and monitor beam output it is worth exploring the relationship between their measurements on each machine and to compare the devices across all machines. Those machines which supplied data for more than a single measurement device have been compared in this section. To allow comparison of the devices the output measurements after correction for calibration have been used. Using this corrected data linear regression of beam output with respect to time has been performed for each device separately on each machine. For a perfectly calibrated set of devices with zero measurement uncertainty the regression results should be equal for the separate Farmer and constancy device datasets, as should the mean result of all measurements with each device.

Of the 204 treatment machines for which data was received a total of 99 of these had data provided for both Farmer chamber and consistency device measurements. To ensure meaningful statistics treatment machines which had less than five results for each measurement device have been excluded from the analysis. This excluded four treatment machines leaving 95 machines containing data included in this analysis.

A variety of constancy devices were reported in use, which range from those having a single ionisation chamber or diode, to those with arrays of measurement points which are also capable of measuring additional aspects of the beam such as energy, symmetry and flatness. There are now also implementations using the treatment machine’s imaging panel which allow checks of multiple aspects of the beam and these are also included within the dataset. As summary of how often each device was used is given in Table 4.2.
Machine output measurement

<table>
<thead>
<tr>
<th>Device Manufacturer</th>
<th>Device Model</th>
<th>Number of centres</th>
<th>Number of Treatment Machines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Various</td>
<td>Farmer Chamber</td>
<td>29</td>
<td>95</td>
</tr>
<tr>
<td>SNC</td>
<td>Daily QA3</td>
<td>15</td>
<td>34</td>
</tr>
<tr>
<td>PTW</td>
<td>Linacheck</td>
<td>8</td>
<td>42</td>
</tr>
<tr>
<td>PTW</td>
<td>QuickCheck</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Sun Nuclear</td>
<td>CheckMate</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Standard Imaging</td>
<td>Beam Checker</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>Varian</td>
<td>MPC</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>ScandiDos</td>
<td>Delta 4</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4.2: Summary of data set for treatment machines containing both Farmer and consistency device measurements grouped by device type. There were three other reported devices: Sun Nuclear DoseCheck, Victoreen DoubleCheck and diodes which were all used on a single machine.

Figure 4.11 shows results of the linear regression from 4 machines, observing different possible situations. Figure 4.11a shows a very well matched Farmer and constancy device with almost no difference in the linear regression fit. Figure 4.11b has a steeper gradient for the constancy device which might indicate different sensitivity to changes in radiation dose than the Farmer chamber. Figure 4.11c shows an offset between intercepts of the Farmer chamber and constancy devices. This offset indicates an offset in the calibration of the devices and can be observed if there is a drift in output (as in Figure 4.11c) or if output is relatively stable (as in Figure 4.11d).
Figure 4.11: A selection of observations when performing linear regression on a Farmer chamber (orange triangles) and constancy device (blue circles) on four different machines. In a) the devices are very well matched with almost identical regression results b) indicates that the constancy device might have a different sensitivity to changes in beam output than the Farmer c) shows the Farmer chamber measuring consistently 0.5% higher than the constancy device d) shows a discrepancy of greater than 1% between Farmer and constancy device.

4.3.8.1 Systematic offset between measurement devices

While centres aim to maintain the calibration of consistency devices which match their Farmer chambers this is not always the case and constancy devices are likely to drift over time. The difference in the mean results of each device on each machine has been used to assess the offset between device calibrations. Centres often have a tolerance on this difference of 0.5-1.0% (Palmer, Kearton and Hayman, 2012). The spread of these results is shown in Figure 4.12 and shows that 65 of the 95 (68%) had a mean Farmer chamber reading greater than the mean consistency device reading. The mean of the results was +0.23% difference between the Farmer and consistency device with a standard deviation of 0.45%. A positive mean indicates that measurements with a Farmer chamber are systematically greater than those using a constancy device. When comparing the mean of the results with a value of zero using a t-test the result is significantly different (p<0.01).
Machine output measurement

Figure 4.12: Difference between the mean of the Farmer chamber and constancy device measurements on each treatment machine. A positive result indicates the Farmer chamber measurements were greater. There were 65 of the 95 machines in which the Farmer chamber had a mean of measurements greater than the constancy device. The mean difference was +0.23% and this difference was statistically significant (p<0.05) to a mean of zero.

On five machines (not all within the same centre) the mean difference in measured output between the consistency device and ionisation chamber measurements was greater than 1% and 28 exceeded 0.5%. All but one of these was deemed statistically significant (p<0.05). The maximum differences were +1.3% and -0.9%.

The results from each treatment unit can also be visualised and compared through plotting the mean Farmer chamber measurement against the mean of the consistency device measurements for each machine. This produces the plot in Figure 4.13 which shows that the data is positively correlated indicating that both devices track output changes in a similar way.
Machine output measurement

Figure 4.13: Plot of mean Farmer chamber measurements vs mean consistency device measurements on each treatment machine. A strong correlation is seen between Farmer chambers and constancy devices on each machine with $R^2$ of 0.73 and Spearman’s correlation coefficient of 0.86 ($p<0.001$).

Linear regression has been used to produce the fit shown in Figure 4.13. The correlation coefficient, $R^2$, is 0.73 indicating a strong correlation between the Farmer and consistency device measurements as expected. The Spearman’s Rank coefficient for this dataset is 0.86 ($p<0.001$), further confirming strong positive correlation between measurement devices as expected.

Centres may only examine the difference between devices on a single day during dedicated QC time, however it may be better visualised and assessed over a longer period. Taking a treatment machine with a difference of 1% between Farmer chamber and consistency device measurement it is often clear to see the offset if the data is plotted over time. This can be seen in Figure 4.11d which shows the data for a linac with an offset between devices of $>1.0\%$. Simply by differentiating the points measured with each device allows clear visualisation of large differences. In this case the Farmer chamber measurements are closer to unity which might be the determining factor for altering machine calibration.
4.3.8.2 **Measurement consistency**

As well as examining the absolute values of the measurements the variation on results from each measurement device may yield useful information. A plot of the standard deviations of results measured with consistency devices vs measurements taken with Farmer chambers indicates that there is not a strong link between them (see Figure 4.14). The $R^2$ value is 0.14 which indicates only very weak correlation. The Spearman’s correlation coefficient is 0.347 ($p<0.001$). If there was strong correlation between the measurement variations between devices then that may indicate that the treatment machine itself is the dominating factor in determining measurement uncertainty, however this does not appear to be the case. Due to the difference in variance observed between measurement devices statistical tests to compare devices have been performed using the Welch’s t-test rather than a standard t-test which assumes equal variance.

![Plot of standard deviations measured on each treatment machine with Farmer chambers and constancy devices](image)

**Figure 4.14:** Plot of standard deviations of beam output measurements on each treatment machine as measured by Farmer chamber or consistency device. There is only a very weak correlation between the results with $R^2$ of 0.14 indicating measurements made with different devices can be considered independent. Spearman’s correlation coefficient is 0.347 ($p<0.001$)
Machine output measurement

The standard deviation has been used as a measure of the uncertainty of measurement with each device to allow quantitative comparison between devices. A statistical comparison was performed on the variation in measurement consistency (as measured by the standard deviation of the residuals of the linear regression) of the devices which were used on more than 10 machines and no statistically significant difference was found. A summary of these results is given in Table 4.3 which shows closely matched sets of results for each device.

<table>
<thead>
<tr>
<th>Device</th>
<th>No. of machines</th>
<th>Summary of machine standard deviations (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean Standard deviation Min Max 5th percentile 95th percentile</td>
</tr>
<tr>
<td>Farmer Chamber</td>
<td>95</td>
<td>0.45 0.18 0.16 1.18 0.23 0.69</td>
</tr>
<tr>
<td>Beam Checker</td>
<td>14</td>
<td>0.45 0.14 0.27 0.78 0.28 0.69</td>
</tr>
<tr>
<td>Daily QA3</td>
<td>34</td>
<td>0.43 0.12 0.24 0.75 0.27 0.63</td>
</tr>
<tr>
<td>QuickCheck</td>
<td>13</td>
<td>0.41 0.17 0.18 0.75 0.23 0.72</td>
</tr>
<tr>
<td>Linacheck</td>
<td>42</td>
<td>0.42 0.15 0.17 0.85 0.23 0.68</td>
</tr>
</tbody>
</table>

Table 4.3: Summary of measurements variation of devices which were used on more than 10 machines. There was no statistically significant difference observed between the consistency of the measurements taken with these devices (p>0.05).

Devices which included fewer than five results on a machine were not included as this indicates measurement frequency was less than monthly and so may not be as statistically reliable. Taking the Farmer chamber measurements as generally the most relied upon, the standard deviation of measurements is 0.18%. When compared to the standard deviation in the mean output (see section 4.3.7) which was 0.7%, these random fluctuations are seen to be smaller, however not insignificant in their magnitude and was included in the modelling as discussed in chapter 7.

4.4 Analysis of long-term beam output data from RSCH

The above work focussed on 6MV data supplied from UK centres spanning 6 months. The following section uses beam output data from the RSCH spanning a greater period (up to four years) and including all available beam energies. This discusses generalising the results to apply to all beam energies and also the potential impact which treatment start date might have on the delivered dose.
4.4.1 Data summary

The data has been collated from 8 Varian linacs with install dates ranging from 2005 to 2015. Six of these are on the main site, and two are at a satellite centre. A summary of the energies on each linac is given in Table 4.4.

<table>
<thead>
<tr>
<th>Linac Name</th>
<th>Install Year</th>
<th>Photon Energies</th>
<th>Electron Energies</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA1</td>
<td>2008</td>
<td>6, 10</td>
<td>6, 9, 12, 16, 20</td>
</tr>
<tr>
<td>LA2</td>
<td>2010</td>
<td>6, 10</td>
<td>6, 9, 12, 16, 20</td>
</tr>
<tr>
<td>LA3</td>
<td>2006</td>
<td>6, 15</td>
<td>None</td>
</tr>
<tr>
<td>LA4</td>
<td>2007</td>
<td>6, 15</td>
<td>None</td>
</tr>
<tr>
<td>LA5</td>
<td>2005</td>
<td>6, 10</td>
<td>6, 9, 12, 16, 20</td>
</tr>
<tr>
<td>LA6</td>
<td>2015</td>
<td>6, 10, 6FFF, 10FFF</td>
<td>6, 9, 12, 16, 20</td>
</tr>
<tr>
<td>Red A</td>
<td>2013</td>
<td>6, 10, 15</td>
<td>6, 9, 12, 16, 20</td>
</tr>
<tr>
<td>Red B</td>
<td>2013</td>
<td>6, 10, 15</td>
<td>6, 9, 12, 16, 20</td>
</tr>
</tbody>
</table>

Table 4.4: Summary of linacs used to analyse long term trends and correlation between beam energies.

Measurements are taken daily using a PTW Linacheck device (each linac has its own assigned unit) and weekly using a Farmer chamber in solid water (two units may share a chamber). The data collated was that which was stored electronically between November 2012 and December 2016 and so the maximum data range is just over 4 years.

4.4.2 Correlation of beam output for different beam energies

With access to data from all energies and taking 6MV as the baseline the correlation of beam output of the other energies on each machine was examined. This was performed qualitatively by graphically visualising the correlation of the measurements taken of different beams on each day. A plot for 10MV and 12MeV against 6MV for the Red A linac (data spanning 1.5 years) is shown in Figure 4.15.
Figure 4.15: Plots of 10MV (left) and 12MeV (right) against 6MV results measured on the same day for one linac. The measurements have been separated into time periods determined by the point at which a calibration or equipment adjustment occurred. A banded structure appears which indicates the beam output of different energies on the same linac are correlated. It has been observed that the MV beams have greater correlation than the MeV beams.

The data has been split into periods between points in time when a known calibration of the linac or equipment was performed. When plotting this data a banding structure is revealed. This is similar to that seen by Hossain (Hossain, 2014). Hossain postulated that the banding structure was due to the response of electrical circuitry within the dose monitoring circuits which may switch between different ranges causing steps to appear in the beam output. In this case separating the measurements into distinct time periods seems to reveal the origin of the band structure clearly. In the case plotted in Figure 4.15 correlation coefficients are between 0.83 – 0.95 for each period for the 10MV beam. The correlation for 12MeV is lower at around 0.4-0.6. This trend of MV beams being better correlated than MeV beams was reflected in each linac and reflects the greater variation seen in daily MeV beam output measurements taken in the centre. Particularly for the MV beams this strong correlation of beam output lends itself to the possibility of reducing the amount of QC performed on each beam with perhaps a selection of energies measured daily to use time more efficiently. This has potential to increase the available clinical time, potentially allowing an additional patient to be treated each day.
4.4.3 Seasonal output trends

Another aspect observed in the analysis by Hossain was that there was a seasonal variation in beam output after corrections had been applied. The six months of UK data indicated a small increase in mean output when all machines were included over this time however this was small (of the order of 0.2%). Examining individual machines gives a much better idea of any seasonal variations and indeed these were observed in the long term RSCH data. The mean output for each month has been taken from LA3 and LA4 (the most stable machines, thus errors in output correction will have least effect) are plotted in Figure 4.16.

![Figure 4.16: Plot of mean monthly outputs (from four years of data) for LA3 and LA4. A clear seasonal variation is observed on both linacs.](image)

This effect was most strongly observed on the oldest linacs which has been reported in the literature (Grattan and Hounsell, 2011). These linacs had the most stable output and the least calibrations over this time period, and so it may be that seasonal variations are masked by larger drifts during the ‘bedding in’ period. A boxplot of the spread of the measured beam output drift for each energy on each linac is shown in Figure 4.17 and does indicate that the most recently installed linacs do have a faster rate of drift when compared with older linacs.
Machine output measurement

Figure 4.17: Plot of the range of output drifts measured on each RSCH linac for each energy. As expected the newest linac (LA6) has the greatest output drift. See Appendix D for description of boxplot display. The whiskers on the plots indicate the total extent of the data and the green triangles represent the mean value for each linac in this case.

Analysis of this long-term data matches well with data previously published, and identifies that banding within correlation plots of the beam output for different energies is likely due to calibration. At these calibration points it is highly likely that not all energies will be calibrated by the same magnitude, thus giving an offset on the plot. Even if an equal calibration change was desired, on machines in which the calibration is adjusted through mechanical means this would be very difficult to achieve.

This analysis was completed for a set of Varian linacs, however it is considered likely that this correlation between energies exists on for other models and manufacturers, thus making the work on 6MV beams more widely applicable.

4.4.4 Effect of treatment start date on delivered dose

As beam output drifts the start date of a patient’s treatment may impact on the total dose received. Here this is examined by looking at the mean beam output from LA2 for different treatment start dates and treatment lengths. At the RSCH an action level of ±2% is used and beam output will be
Machine output measurement

adjusted when it reaches this level. We can consider simply a treatment the day prior to and the day following re-calibration and it is clear a different dose will be delivered due to this.

A more realistic method of measuring doses delivered to patients is to examine the mean dose they would receive given a known start date. To do that the rolling mean output was calculated for 6MV beams on LA2 for 30 and 60 days. A plot of these results is shown in Figure 4.18.

![Mean 6MV output based on start date with various treatment durations](image)

**Figure 4.18:** Plot of 6MV beam output and the 30 and 60 day rolling mean of these values. This simulates the mean dose received by a patient dependant on the start date and treatment duration.

For longer treatment course duration, the possible range of the mean dose received reduces. Quantitatively the reduction in dose variation can be assessed by calculation of the standard deviation of the rolling means for different treatment durations. This has been completed and results are plotted in Figure 4.19.
Variation in standard deviation of rolling mean dose for different treatment durations.

Variation exists dependant on start date due to drift in treatment machine output. Treatment courses of shorter duration inherently have a greater range of possible doses delivered to the patient.

Dependant on the treatment given the duration of the treatment course may be from a single day which will have the greatest variation in delivered dose due to beam output up to around 40 days. On LA2 in this case a single day treatment has a variation in delivered dose dependant on start date of approximately 0.74%, whereas a treatment delivered over 40 days is 0.56%. The frequency of recalibration and tolerances used will impact on this variation in delivered dose to patients.

4.5 Discussion of beam output measurements

This section contains the first analysis, to the best of knowledge of the investigator, of local beam output measurements from multiple centres which is in contrast to previous studies based within a single centre (Luketina and Greig, 2004; Grattan and Hounsell, 2011; Hossain, 2014). The mean output of each machine within a centre was assessed to give an indication of the overall variation in delivered dose due to beam output to patients within that centre (section 4.3.7.1). This indicated differences in delivered dose of over 2% are possible due to beam output alone. The variation of measurements taken with different devices on the same machines was examined. While most had good agreement, there were some which had differences in the mean measured value of over 1% in which case it would be recommended to recalibrate the constancy device to better match the ionisation chamber measurements if possible.
Machine output measurement

Obtaining the data from the centres was often troublesome, either due to resource implications, or due to systems not being able to export the required data. With a continual drive to modernise services, to introduce collaborative networks between centres and move records to an electronic format this study has highlighted the importance of the initial stages of implementing an electronic QA system including full commissioning of all documented features including access to data.

It is known that measurement practice varies between centres (Palmer, Kearton and Hayman, 2012). One centre stated that beam output measurements taken during their morning run-ups had an asymmetric action level of -2% to 0%. This was to take into account the increase in beam output through the day. The aim was to deliver their high dose treatments (usually in the afternoon) with a beam output closer to the desired calibration value (tolerance effectively shifting to ±1%).

Comparing with national recommendations of monthly beam output checks to be within ±2% (Mayles et al., 1999) there was only 1.9% of all individual measurements outside of this range within this dataset. A single machine had a mean measured output outside of this range over the entire 6-month data period at -2.1% which was perhaps surprising considering the national recommendations are to be within ±2% for monthly checks.

There is potential for large intra-centre variation on dose received by the patient depending only on the treatment machine they are assigned to. The largest range within a single centre was 2.2% (see Figure 4.8) which is potentially a significant difference (as discussed in chapter 7).

Whilst many centres were moving towards electronic data storage methods, eight centres relied fully upon paper records, and so may not be able to readily quantitatively assess trends within datasets. It is noted that of the 20 centres with two or fewer treatment machines, only one of these centres relied solely on paper records for recording beam output. This may be because of the increased flexibility of small centres to adapt practices and conversely the larger initial burden of transferring between record keeping systems at larger centres. There was no significant difference in the beam output measurements identified between centres with different recording methods, however comments were made by clinical staff that electronic records are often easier to access and collate. When requesting the data for this study, those with paper records were often more reluctant to provide them due to the potential perceived time burden.

The data received from each centre varied; providing all measurements, or only a subset (such as daily, monthly or weekly). Differences in measurement device would result in differing uncertainties
Machine output measurement

on each individual set of data and it is acknowledged that this may affect comparisons between machines, however the analysis comparing different devices indicates there is no significant difference between the accuracy of each device (see section 4.3.8.2). A comparison of measurements taken with different devices on each machine indicated differences of over 1% on some machines. Tolerances for this comparison are determined locally and depend on many factors including the device used and the frequency of alternative measurement techniques. It has also been noted that the mean of measurements made with a Farmer chamber was greater than those made with consistency devices on two thirds of machines. One explanation is that often consistency devices are battery powered and battery depletion is most likely to result in reduced measurement sensitivity, thus reduced readings. This may warrant further investigation and may be a more dominant effect for different devices.

Manufacturers have slightly different technology within their treatment machines, and so there is potential for different rates of change in beam output. For instance, Varian uses a sealed ionisation chamber within the linac head whereas the chamber for Elekta is unsealed. It has been noted through communication with clinical staff that the beam output of a Varian linac will tend to mirror a change in atmospheric pressure indicating chambers may not be completely sealed units. Excluding CyberKnife (which only had a single machine within the dataset) all manufacturers showed both positive and negative trends in beam output drift within their treatment machines. There was no significant difference (p>0.05) in the trends in beam output observed between machines from different manufacturers within this dataset.

This dataset highlights the possible differences in beam output between machines over time. This variation directly relates to delivered dose and so also to patient outcomes (Bentzen et al., 2000) (to be discussed in chapter 7), further highlighting the importance of routine monitoring and maintenance. There was no major cause for concern when examining the machines included within this study, and overall, they are all closely matched given the tolerances they are working within.

This work presented the range of practices in measurement variation across different centres and gives quantitative results for the variation in beam output on different machines within a single centre. The variation over time of beam output on each machine was quantified, and again a large possible variation between machines is observed. This work may act as a baseline for further work. Whilst considered a routine measurement, control of beam output should remain a high priority to ensure the high precision and accuracy required in radiotherapy is met.
5 Analysis of dose variation within the PARSPORT trial

The use of IMRT treatment techniques continues to increase (Public Health England, 2017), offering a more precise method of dose delivery (Staffurth et al., 2015). As well as this, standardisation of techniques is increasing and the introduction of automated planning is likely to reduce the differences between patient plans (Voet et al., 2014; Hansen et al., 2016; Hussein et al., 2016). This reduction in dose variation due to the planning techniques will result in dose variation due to the beam output becoming of greater overall significance in the treatment pathway.

This chapter explores the variation in planned dose from patients included in the PARSPORT trial (Nutting et al., 2011) and the development of a method of automated extraction of dosimetric statistics from multiple DICOM plan files. Quantifying the variation in dose due to planning technique allows the variation in dose due to machine variations to be seen in the wider context of dose uncertainties.

The PARSPORT trial was a phase III multi-centre randomised control trial examining dose sparing of the parotid glands in head and neck radiotherapy treatments using 3D conformal and IMRT treatment techniques. The trial was designed to demonstrate the reduction in the induction of xerostomia (dry mouth) at 1 year when using IMRT techniques to reduce dose to the parotids (Nutting et al., 2011).

The advantage of using data from patients included within a clinical trial is that patients treated at multiple centres followed the same protocol. The centres will also have had to achieve specific criteria and have undergone QA prior to participation. This trial had comprehensive dosimetry QA completed at each of the six participating centres which included checks on the treatment planning system (including checks of basic IMRT field modelling), as well as measurements of delivered doses (Clark et al., 2009). The importance of trial QA has been demonstrated previously indicating that the power of a trial is greatly increased when it has associated QA (Pettersen, Aird and Olsen, 2008; Peters et al., 2010). The trial QA showed that there was no significant difference between step and shoot and dynamic IMRT delivery techniques (Clark et al., 2009), and so these have not been separated for this work.
5.1 Summary of included treatment plans

Dose cubes, CT images and structure sets were supplied for 79 patients through application to the ICR-CTSU (Institute of Cancer Research Clinical Trials and Statistics Unit). For the purposes of this analysis patients with the same prescribed dose of 65Gy in 30 fractions to the primary target (primary PTV) were selected. Conformal and IMRT plans aimed to deliver 50Gy and 54Gy respectively to the elective nodes (nodal PTV). This resulted in 70 patients being included in this work, 33 of which had treatment with a 3D conformal plan, and 37 which had an IMRT treatment.

Structures which are analysed in this work are the PTVs (primary PTV and nodal PTV) and the left and right parotid glands. With the IMRT treatment plans a mean dose constraint of 24Gy to the whole parotid gland was used and it is this which resulted in lower doses to the parotids than when using conformal techniques.

5.2 Development of tool for automated extraction of dosimetric statistics

During this work I encountered a number of tools but which did not allow a simple extraction of dose statistics from multiple treatment plans in an automated way. Many of the commonly used programs such as VODCA (Visualisation and Organisation of Data for Cancer Analysis) (MSS Medical Software Solutions GmbH, 2008) and CERR (A computation environment for radiotherapy research) (Deasy, Blanco and Clark, 2003) are limited in their use by licencing or cost implications (CERR is MATLAB based, and thus is not free to use). These also require manual extraction of dosimetric statistics from individual patients, making them less suited to this investigational work.

Early work with the DICOM files for this work was largely performed manually with a view of collating the DVHs for each structure. It was envisioned that this would be the most appropriate method of collating the required data for further analysis. Using this approach, several problems were encountered. As plans were created across multiple centres there was often inconsistency in structure names. For example, the left parotid structure might be named ‘l_parotid’, ‘lparotid’, ‘left_par’, or many other combinations. This resulted in a lot of manual renaming of DVHs as the data was exported and collated to try and ensure consistency in naming.

The DVHs were exported using dicomiler (Panchal, 2015) to Excel files and compiled using the python pandas package (McKinney, 2010), however due to the large number of structures, memory issues were encountered when working with this volume of data on available hardware (each patient had multiple structures and the DVH data for each of these structures was stored within a
single file). It was at this point that a new approach was decided upon to limit this issue and a more flexible approach was developed allowing future expansion beyond the 70 plans which required analysis for this work.

This tool is designed to allow retrospective analysis of treatment plans when the plan data is available and can be run on an ad-hoc basis. This situation will often be the case for smaller scale local planning studies. With a view of also developing a tool which had use outside of this work the following were the proposed requirements:

- Extract dosimetric parameters from supplied DICOM files which include the structure set, dose grid and plan file.
  o Automated extraction of all required parameters should be possible from these files.
- Allow the user to specify which dosimetric statistics to extract for each structure (e.g. mean, D95, V100)
  o Specification should be relatively flexible. For example, “D99” should be the same as “d99.0”
- Should work with data exported from common systems.
  o For the work, only the PARSPORT files originally exported from VODCA were required, but compatibility with other systems was highly desirable.
- Should have minimal hardware requirements.
  o It was desired not to have more than a single treatment plan loaded into memory at one time.
- Output the dosimetric parameters in a readily useable format for further analysis.
  o CSV files are a common format.
- Should allow a common structure name to be included for the purposes of analysis.
  o E.g. include a name such as ‘l_parotid’ for all similarly named structures.
- Allow selection of a directory of DICOM plans from which to extract dosimetric statistics.
  o The patient data was organised within individual folders already. Patient A would have all files within one folder, and Patient B all files within another.

To achieve the aims of this and to allow the tool to be freely available to others the Python environment was again chosen. The developed tool has been built using the dicompyler-core and pydicom packages as detailed in Table 1.5 (see Appendix B for code details).
The code has been tested with files exported from VODCA (MSS Medical Software Solutions GmbH, 2008), *dicompyler* (Panchal, 2015) (files were run through *dicompyler* to ensure consistent patient IDs and to identify the patient files for inclusion in the analysis), Varian’s Eclipse (Varian Medical Systems, 2016a) and Variseed (Varian Medical Systems, 2016b) systems, and Oncentra Brachy (Elekta Instrument AB Stockholm, 2015) from Elekta. The dosimetric statistics are seen to be consistent with those from the systems from which the data was originally exported where it was possible to compare these values.

### 5.2.1 Data extraction process

Data extraction using the developed code proceeds as follows.

1. User is prompted to select directory containing files for analysis.
2. User must supply list of statistics required (can be from a file, or manually typed).
3. Each folder in turn within the specified directory will be loaded and the required dosimetric statistics extracted.
4. Extracted statistics are collated internally and exported in the specified format (e.g. CSV or Excel).
   a. If it is known that there are variations in structure names, then these can be supplied, and a common name is included within the output.

The exported file contains a row for each structure which contains each extracted dosimetric statistic. This format allows simple data filtering and analysis, both using Excel and other packages.

### 5.3 Comparison of doses from conformal and IMRT plans

The plans were separated into the conformal and IMRT groups and the planned doses in each group have been compared to identify key differences. The change in mean dose to the parotid glands is first considered (as this is one of the key outcomes of the trial) and then the change in variation in dose to the primary target structure (PTV) is considered as these demonstrate different aspects of the change from conformal to IMRT planning.

#### 5.3.1 Reduction in dose to parotid glands with IMRT

Initial analysis acted to verify that the IMRT plans did indeed reduce the dose delivered to the parotid glands as detailed in the published study findings (Nutting *et al.*, 2011). This helped to ensure the data extraction was robust and suitable to analyse unknown aspects of the dose distributions and allowed visualisation of the variation within plans to one of the key organs studied within this
Analysis of dose variation within the PARSSPORT trial

trial. Figure 5.1 presents boxplots of the mean dose to both the left and right parotids for conformal and IMRT plans. A summary of the doses delivered is given in Table 5.1.

Figure 5.1: Boxplots showing the distribution of the mean doses to the left (white) and right (grey) parotid glands separated by treatment type. The whiskers represent the range between the 5th and 95th percentiles. The IMRT plans show a statistically significant reduction in mean dose (p<0.01) when compared using a t-test. This is one of the published trial findings (Nutting et al., 2011). See Appendix D for description of boxplot display. The green triangles show the mean result. A summary of the doses is given in Table 5.1.

The distribution of results for left and right parotid glands is remarkably similar for both treatment techniques however the magnitude of planned doses is significantly different between the conformal and IMRT plans (p<0.01). The reduction in dose to the parotids when using IMRT is clearly seen in Figure 5.1, with the conformal plans having a median mean dose of over 60Gy reducing to approximately 30Gy when using IMRT. In the case of the parotids the spread of doses has increased when using IMRT techniques, however clinically this would have been justified at the time due to the large reduction mean parotid dose possible.
Analysis of dose variation within the PARSPORT trial

<table>
<thead>
<tr>
<th>Plan type</th>
<th>Laterality</th>
<th>Mean</th>
<th>Median</th>
<th>Standard deviation</th>
<th>5th percentile</th>
<th>95th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conformal</td>
<td>Left</td>
<td>60.8</td>
<td>61.8</td>
<td>5.3</td>
<td>51.7</td>
<td>67.5</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>60.6</td>
<td>62.9</td>
<td>7.5</td>
<td>52.6</td>
<td>66.2</td>
</tr>
<tr>
<td>IMRT</td>
<td>Left</td>
<td>35.2</td>
<td>30.7</td>
<td>12.9</td>
<td>20.6</td>
<td>57.5</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>34.8</td>
<td>30.7</td>
<td>13.1</td>
<td>20.9</td>
<td>58.8</td>
</tr>
</tbody>
</table>

Table 5.1: Summary of doses delivered to parotid glands for patients within the PARSPORT trial. Both left and right sides were very similar. IMRT plans resulted in a significant reduction in the doses received (p<0.01).

The clear reduction in the doses to the parotid glands gives rise to the potential of dose escalation, or of reducing toxicities. The variation in doses to the PTV structures is of interest as this will be linked to the measures of success of the treatments (Nutting et al., 2011). The doses for the parotid extracted from these plan will be used in the NTCP modelling using the LKB model within this work.

5.3.2 Reduction in dose variation to target (PTV) structure with IMRT

The variation in doses to the target volumes has not been explored to the author’s knowledge, and so the variation in planned dose to the target volumes is investigated. The standard deviation of the mean, median, maximum doses as well as the D95 (the minimum dose received by the maximally irradiated 95% of the volume), and D99 for the primary PTV is shown in Figure 5.2 for both IMRT and conformal plans. The numerical values are given in Table 5.2. Each of these parameters was found to be statistically different (P<0.05 using the Levene test which compares variances (Snedecor and Cochran, 1989)) dependant on the treatment technique. A clear reduction in variation in dose to the target volumes is observed when using IMRT indicating the greater dose homogeneity achievable.
Figure 5.2: Plot showing the calculated standard deviation as a percentage of the 65Gy prescription dose to the primary target structure (primary PTV) for multiple dosimetric parameters. This is split into IMRT and conformal plans from the PARSPORT trial showing the difference in variation arising from different planning techniques. Each of the parameters showed a statistically significant reduction in the variance across the population (P<0.05 using the Levene test which compares the variances).

A similar observation to that of the primary PTV was seen for the nodal PTV. For the nodal PTV the absolute variation observed was greater using conformal planning techniques for all parameters except the D95. The variation in mean, median and D95 were found to be significantly different between treatment techniques (p<0.05). Whilst the maximum and D99 were not statistically significantly different they did still show a reduction in the variation when IMRT was used.

<table>
<thead>
<tr>
<th>Plan type</th>
<th>Standard deviation of dosimetric parameter (%) of primary PTV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Conformal</td>
<td>2.0</td>
</tr>
<tr>
<td>IMRT</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 5.2: Standard deviation of dosimetric parameters for the primary PTV from the patients treated with conformal and IMRT plans. These results are plotted in Figure 5.2 Each of these statistics was found to be statistically different for the high dose target volume (p < 0.05) for the conformal and IMRT planning techniques.
The purpose of IMRT planning is not only to reduce the delivered dose to surrounding organs by conforming the dose around them, thus reducing OAR dose (as demonstrated in section 5.3.1), but also to achieve a more homogeneous dose to the target volumes (as demonstrated in Figure 5.2). The mean dose to the PTV structures will directly scale with beam output, and so the variation in mean dose from treatment planning can be compared to this. The standard deviation of mean dose to the primary PTV when using conformal techniques is 2%. This falls to 0.5% for the patients planned with IMRT (see Figure 5.2 and Table 5.2).

Comparing this variation in dose arising from the treatment plans to that of the variation due to uncertainties in beam calibration is useful at this stage. With the standard deviation of daily beam output measurements being around 1% this compares well with the deviation in mean doses due to the planning techniques and is larger than the uncertainty arising from the IMRT plans. Thus, with the introduction of IMRT the dose variation due to beam output is now greater than that introduced in the planning process, whereas previously the planning process was dominant. The impact of changes in dose to the primary targets and OARs will be assessed through the implementation of the radiobiological models discussed in section 1.5 and with parameters detailed in the following chapter.

**5.4 Summary of dose variation within PTV and OAR structures**

The use of IMRT is primarily focussed on reducing the toxicity of treatments through the reduction in dose to surrounding OARs. When extracting the dose statistics from the PARSPORT trial this was observed for the parotid glands for which the mean dose significantly reduced which agreed with the published results (Nutting et al., 2011). This work has also found that the variation in the dosimetric parameters also reduced for the targets with a smaller spread in planned doses to the PTV structures. Both the reduction in dose to OARs and the reduction in the spread of doses to the target volumes are good achievements which are possible due to the implementation of IMRT which was heavily supported by audits. This reduction in variation due to planning means that the variation in delivered dose due to beam output is now larger than that due to planning. Thus, investigating the impact of beam output on patient outcomes is increasingly valid with modern treatment techniques.
Due to their common usage throughout radiotherapy the LQ model will be used to assess TCP and the LKB model will be used to assess NTCP variations. The LQ model is used widely within radiotherapy, such as determining the recruitment numbers required for clinical trials, and correcting for time gaps in treatment (The Royal College of Radiologists, 2008; Fowler, 2010). To date the LQ model is still the most accepted model (Brenner, 2008; Lindblom et al., 2014). In this chapter the implementation of the models for this work is detailed and the effect of each parameter is characterised. The background of radiobiological modelling is given in section 1.5. Details of the LQ and LKB models are given in sections 1.5.2 and 1.5.4 respectively.

6.1 Implementation of radiobiological models

The radiobiological models have been implemented using the Python programming language (Python Software Foundation, 2016) and largely use the Numpy (van der Walt, Colbert and Varoquaux, 2011) and Scipy (Jones, Oliphant and Peterson, 2001) packages to manipulate the data. Each of the TCP and NTCP models can be run separately or combined to share a common set of delivered doses.

The primary aim of this modelling was to allow the delivered dose to be varied on a fraction by fraction basis for individual patients. As well as this, parameters should be allowed to be specified (with appropriate variance as required) or fitted where possible if not supplied. Fitting is based upon supplying a set of data points corresponding to a TCP or NTCP at a known dose.

6.1.1 Model development and input parameters

For the LQ model the most common usage will be to fix the value of $N_0$ and $\beta$ parameters and then optimize the model by fitting the variation of the $\alpha/\beta$ (as a percentage variation of the mean value). Once this is determined for each case this can be fixed and the dose parameter varied (both varying the absolute value and standard deviation to simulate different scenarios). This follows the method
commonly employed by others to represent the variation in population TCP (Webb and Nahum, 1993; Webb, 1994). The LKB model contains three parameters which may be determined, \( m \), \( n \) and \( D_{50} \). Dose will be explored in the same way for the TPC model.

This work focuses on Head and Neck cancer and prostate cancer. When modelling Head and Neck cancer the PARSPORT trial (Nutting et al., 2011) is the primary source of data, particularly for NTCP modelling for which the 3D dose distributions were available (see chapter 5). This trial was influential in increasing the use of IMRT for head and neck cancer, demonstrating the improved dose sparing achievable for the OARs. Assessing the treatment planning data allows the variation in dose due to machine beam output to be placed in context with other sources of variation in dose received by the patient. Whilst there will understandably be large differences in doses to OARs due to the anatomical variation between patients, the dose to the target volumes should be largely comparable between plans.

The treatment for prostate cancer has been modelled upon the RT01 trial (Medical Research Council Radiotherapy Working Group, 2000; Dearnaley et al., 2007, 2014) which assessed dose escalation from 64Gy in 32 fractions to 74Gy in 37 fractions. Dose escalation showed clear improvement in outcomes and this was influential in 74Gy being the most commonly delivered dose for prostate cancer for a number of years. Simulations have also been completed for the Fox Chase Medium risk patients using the results presented in section 1.5.3.3. This had TCP values of 19%, 31% and 84% at 70Gy, 72Gy and 76Gy respectively showing a much steeper dose response curve than for the RT01 trial.

### 6.1.2 General overview of modelling and simulation approach

The approach taken here has been to simulate the delivery of dose for each fraction to each patient individually. This allows parameters to be set for an individual patient and an individual fraction as appropriate. For an individual patient all parameters will remain fixed through the duration of the simulated treatment, but the dose delivered on each fraction is allowed to vary.

The TCP and NTCP curves are then determined using fixed patient parameters for the range of doses specified. For each patient simulated these determined doses can be used simultaneously to determine TCP and NTCP. Performing this multiple times with different patient parameters allows a population to be simulated. The inclusion of variation in dose per fraction allows the impact of calibration uncertainties to be explored and assessed.
6.1.3 Inclusion of dose variation

The delivered doses may either be supplied explicitly for each simulated fraction modelled (which may be useful for studies where a known change in calibration occurs), or extracted randomly from a distribution (which allows simulation of the inherent random daily variations).

To reflect both systematic and random variations in the daily delivered dose a number of variables are incorporated when determining this. To account for the random daily variations doses are drawn from a normal distribution centred on a nominal dose per fraction with a specified standard deviation. Additionally, an offset may be applied to simulate a known systematic output difference at the first fraction. Also possible is the inclusion of a drift in the delivered dose over time which is modelled as a linear variation assuming treatments are completed daily (weekends are not included). A diagram showing this is presented in Figure 6.1. Alternatively the delivered dose per fraction may be explicitly supplied by the user which would allow the inclusion of a change in calibration during the treatment.

![Modelling of delivered doses](image)

**Figure 6.1:** Diagram showing the features of determination of dose per fraction used within the modelling within this work. Included are a systematic offset of -1% at the start of the time period shown, a drift in beam output of +3%/year and random fluctuations around this which are drawn from a normal distribution with standard deviation of 0.2% of the nominal dose (2Gy/fraction).
The variation in dose delivery has been explored in chapters 2-4 and the adaptability of the model setup allows inclusion of a range of different aspects. For this work it is assumed that a calibration does not occur during a patient’s treatment.

Through development of the model the process was defined, breaking each step into its own individual function as much as possible, however it will always be possible to further refine the inner workings. The LQ model was developed first, and so many aspects of this model are relied upon by the NTCP model such as the generation of the patient doses and a population of patients. These datasets or generated within the TCP model’s functions and then passed to the NTCP model allowing an identical set of patients and doses to be modelled with each.

6.1.4 Linear quadratic model

The LQ model has been implemented for modelling of TCP as in equation (1.14) which is reproduced again here for clarity.

\[
TCP = \exp\left(-N_0 \prod_{i=1}^{n} \exp\left(-\alpha d_i - \beta d_i^2\right)\right)
\]

To produce a TCP value the values of \(\alpha/\beta\), \(n\), \(N_0\), and \(d\) (nominal dose) are supplied. An example of one possible output of the model was presented in Figure 1.8 which showed TCP values for a population of patients being produced from the supplied values. In this example a fixed value of \(N_0\) was used and the spread of \(\alpha/\beta\) values was optimised to fit the clinical data.

The steepness and position of the curve is largely determined by the values of \(N_0\) and \(\alpha/\beta\). To simulate a population the value of \(\alpha\) is allowed to vary and has been implemented with a lognormal distribution which avoids the possible production of negative values. The value of \(\beta\) is held fixed, thus giving a range of \(\alpha/\beta\) ratios to mimic the clinical situation of varied radiosensitivities (Webb and Nahum, 1993; Webb, 1994; Dale and Jones, 2007).

The primary parameters which may be supplied to the model and shown in the code snippet below, produce the data required to produce Figure 1.8. An explanation of each supplied parameter is given as a comment in the code (comments appear after a hash, '#'). The following code shows the parameters used to represent the head and neck patients within the PARSPORT trial.
The model is arranged in a module called “TCP_NTCP” and within that the “completeTCPcalc” and “completeNTCPcalc” are functions which are called in the same way as above. These functions then use additional internal functions to calculate and return the model results which can then be plotted and further analysed.

It was desirable to make the modelling more automated by determining the value of “alphabeta_sd_use” (which is the standard deviation of the $\alpha/\beta$ value as a percentage) through automatic optimisation of this parameter. A set of dose and corresponding TCP data can be supplied to the models function and from this a value of “alphabeta_sd_use” is determined for the population.

### 6.1.4.1 Optimisation of spread of $\alpha/\beta$ values from simulated population

The approach for simulation involved producing a population of patients with different $\alpha/\beta$ ratios. The spread of $\alpha/\beta$ values will determine the gradient of the dose-response curve with a fixed value of $N_0$ and this is a common method used to simulate the clinically observed dose responses when using the LQ model (Webb and Nahum, 1993; Webb, 1994). To ensure consistency in results when repeating modelling (without explicitly specifying the parameter values for the entire population) the size of the population ($n$) required to achieve the desired uncertainty in population TCP value was investigated when optimising the value of the standard deviation in $\alpha/\beta$.

Greater deviations in $\alpha/\beta$ (radio sensitivity) result in greater spread of patient TCP values at a given nominal dose. A set of population TCP data is calculated from the mean of the individual patient data generated in the simulations. For a given set of parameters, the larger the population, $n$, the greater confidence is gained upon the determined value of $\text{alphabeta}\_\text{sd}\_\text{use}$, but this comes at a cost of increased computation time.
Optimisation of the value of `alphabeta_sd_use` to match the population curve to the supplied dose and TCP data is performed using Scipy’s “minimize_scalar” function. This optimisation progresses by varying the supplied parameter to minimise the difference between the input and output values. The magnitude of the difference calculated during this optimisation is the least squares difference of the supplied TCP points to those calculated for the population at an equal dose. If there were no statistical variations then an exact value could be determined, and this would be consistent with each run, however this is not the case unless fixed doses and α/β ratios are supplied.

An investigation into the influence of the population size used when fitting has been completed. This was done by performing multiple fits using consistent parameters but varying the size of the modelled population used to produce the optimised value of `alphabeta_sd_use`. The optimisation stability was investigated for each of the clinical cases detailed further in Table 6.2, section 6.2.1. A summary of the values is given in Table 6.1 showing the range investigated.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Head &amp; Neck (PARSPORT)</th>
<th>Prostate (RT01)</th>
<th>Prostate (Fox Chase – Med)</th>
</tr>
</thead>
<tbody>
<tr>
<td>α/β</td>
<td>10</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>β</td>
<td>0.02</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Nominal dose per fraction</td>
<td>2.167Gy</td>
<td>2Gy</td>
<td>2Gy</td>
</tr>
<tr>
<td>N₀</td>
<td>1x10⁶</td>
<td>1x10⁴</td>
<td>5x10⁴</td>
</tr>
<tr>
<td>Nominal Doses</td>
<td>65Gy</td>
<td>64Gy</td>
<td>74Gy</td>
</tr>
<tr>
<td>Nominal TCPs</td>
<td>66%</td>
<td>43%</td>
<td>55%</td>
</tr>
</tbody>
</table>

Table 6.1: Parameters used during TCP simulation to determine the influence of the population size on the variation in simulated TCP results caused by the optimisation of the parameter describing the standard deviation of α/β. The population size (n) was varied between 5 to 5000.

The values of N₀, β and α/β are representative of the values within the literature (Webb, 1994; van Leeuwen et al., 2018), the values for N₀ in particular often vary by orders of magnitude. With no dose variation assumed at this stage, a range of values for n from 5 to 5000 was used to fit the population TCP curve. The fitting was repeated 20 times for each value of n and the variation in the results assessed by calculation of the standard error of the mean of the calculated population TCP at
a dose of 65Gy for the head and neck case and at 74Gy for the prostate cases. Figure 6.2 shows the results of these simulations with varying population size.

![Graph showing variation in simulated TCP at fixed nominal dose with population size used for optimisation.](image)

**Figure 6.2:** Variation in the simulated TCP values at a nominal dose of 65Gy and 74Gy for head and neck and prostate cases respectively for varied population size (n) used during optimisation of the standard deviation of α/β. The optimisation procedure was repeated 20 times whilst varying population size. The plot shows the standard error in the mean of these simulated population TCP values. The SEOM falls below 0.1% at approximately n=1000. This uncertainty was deemed acceptable for this work and so n=1000 was used for all further modelling work.

As the population size increases the variation in the population mean TCP decreases. At approximately n=1000 the SEOM of calculated population TCP falls to 0.1% and this uncertainty was deemed acceptable for this work which was to examine relative changes an order of magnitude greater than this. For the remainder of this work, when optimising the model and calculating population TCP the value of n was fixed at 1000.

### 6.1.5 Lyman Kutcher Burman model

The NTCP model implemented follows the LKB model which is widely used and has been used to analyse the PARSPORT trials data (Miah et al., 2013), and thus parameters from this analysis can be used here to predict the variation due to dosimetric uncertainties. The modelling largely follows the
same processes internally as described for the LQ model above. The LKB model implemented is
detailed in equation (1.20) and is reproduced here for clarity.

\[
NTCP(D, V) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx = \frac{1}{2} \left(1 - \text{erf}\left(\frac{u(D, V)}{\sqrt{2}}\right)\right)
\]  

As for TCP, individual sets of NTCP results can be determined to produce a population NTCP. Fitting
may be performed using the model implementation in a similar way. In this work the individual
patient NTCPs will be calculated, and then the mean of these taken to give the population NTCP.

The model is implemented within the Python language utilising the Numpy and Scipy packages as for
the TCP model described above with the function call taking a similar form. Parameters are assumed
to be normally distributed, and this is used as a basis for producing a set of population results. In this
case the model will attempt to fit all parameters which are not supplied, however the success of this
is often dependant on selecting appropriate initial values which are sensitive to small changes due to
the model’s non-linearity.

6.2 Sensitivity analysis

This section explores the sensitivity of simulated TCP and NTCP to variations in each parameter. This
has been completed for parameter sets which simulate prostate and head and neck cancers, both
for TCP (using the LQ model) and associated NTCP (LKB model). The sensitivity has been assessed
based on varying the parameters for a single patient who is representative of a ‘typical’ patient from
the population. A range of variations between -10% and +10% has been introduced to each
parameter individually whilst holding the remaining parameters fixed.

6.2.1 Linear quadratic model

The parameters presented in Table 6.2 were used to model typical prostate and head and neck
cancer patients. These parameters are used for the remainder of the model characterisation and
taken as the standard used when further assessing the variation with dose in chapter 7.

The optimised values of the \(\alpha/\beta\) standard deviation were approximately 25% of the mean value (the
model implemented took a percentage value as its parameter) for the head and neck case.

Literature values are approximately 0.08 for the standard deviation in \(\alpha\) which has a mean value of
0.3 with \(\beta\) of 0.03 (Webb, 1994; Dale and Jones, 2007; Joiner and van der Kogel, 2009). Converting
this to a percentage gives 26.6% which is comparable to the optimised value obtained during the model optimisation here and used within this work. For the RT01 case the shallower dose response curve gives rise to a greater variation in $\alpha/\beta$, and the Fox Chase case with its very steep curve has a smaller range of $\alpha/\beta$ values which arise during the optimisation.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prostate (RT01)</td>
</tr>
<tr>
<td>$\alpha/\beta$</td>
<td>2</td>
</tr>
<tr>
<td>$\alpha/\beta$ SD (%)**</td>
<td>100%</td>
</tr>
<tr>
<td>[absolute value of $\alpha_{SD}$]</td>
<td>[0.06]</td>
</tr>
<tr>
<td>$\beta$ (held fixed)</td>
<td>0.04</td>
</tr>
<tr>
<td>Dose/fraction (d) (Held fixed for optimisation)</td>
<td>2Gy</td>
</tr>
<tr>
<td>TCP results used for optimisation</td>
<td>Ten year bPFS</td>
</tr>
<tr>
<td></td>
<td>43% @ 64Gy</td>
</tr>
<tr>
<td></td>
<td>55% @ 74Gy</td>
</tr>
<tr>
<td>Dose of interest*</td>
<td>74Gy</td>
</tr>
<tr>
<td>$N_0$</td>
<td>$1 \times 10^4$</td>
</tr>
</tbody>
</table>

Relevant references for model parameters:

**RT01 Trial:** (Medical Research Council Radiotherapy Working Group, 2000; Fowler, Chappell and Ritter, 2001; Dale and Jones, 2007; Wedenberg, 2013b; Dearnaley et al., 2014)

**Fox Chase study:** (Hanks et al., 2002)

**PARSPORT study:** (Nutting et al., 2011)

**Model parameters:** (Dale and Jones, 2007; Fowler, 2010; Hawkins, 2017; van Leeuwen et al., 2018)

* The dose of interest was the point on the TCP curve at which the variation in TCP has been assessed unless otherwise specified. This dose represents the typical treatment dose.

** Determined through model optimisation.

**Table 6.2:** Parameters used for modelling prostate and head and neck patients within this work using the LQ TCP model. Based on parameters published within the literature the values of the standard deviation of the $\alpha/\beta$ value were determined through optimisation using the model. The $N_0$ values used for each set of simulations are based on the range of values used within the literature and were optimised to fit the clinical data.
6.2.1.1 Linear quadratic model parameter sensitivity

The sensitivity of the dose per fraction, \( N_0 \), and \( \alpha/\beta \) parameters has been determined for the RT01 prostate case and PARSPORT head and neck case as these represent the range of shallow and step clinical dose response curves. Each of the parameters was varied individually between -10% and +10% and the corresponding percentage change in TCP at the dose of interest was determined. The parameter values used are those listed in Table 6.2. The parameter sensitivity results are presented in Figure 6.3.

![LQ Model parameter sensitivity](image)

**Figure 6.3:** Variation in parameter sensitivity for the different parameters of the LQ model for typical prostate and head and neck cases. The change in TCP was calculated due to an increase in each parameter separately from -10% to +10%. In each case the delivered dose showed the greatest sensitivity to simulated outcome as a percentage of the mean parameter value and with \( N_0 \) had the smallest.

The dose per fraction is the parameter to which the TCP is most sensitive in both cases examined which demonstrates the importance of accurate dosimetry when using the models to predict outcomes.

6.2.2 Lyman Kutcher Burman model

The parameters for the LKB model listed in Table 6.3 have been used to model some common complications arising from radiotherapy. Three cases are considered, the first is for Grade 2 rectal
bleeding due to treatment of the prostate with parameters determined by Gulliford et al. (Gulliford et al., 2012). These parameters were based on the RT01 trial and so is the same cohort of patients as used for some of the TCP modelling discussed in section 6.1.4. The second and third cases are based on the work published by Semenenko and Li (Semenenko and Li, 2008) in which they derived parameters for the LKB model based on the incidence of xerostomia (dry mouth) as a function of the mean dose to the parotid gland from multiple institutions. The values of each parameter determined were similar to those derived by Miah et al. (Miah et al., 2013) which were based on the PARSPORT trial data. The xerostomia definition used was a reduction of salivary gland flow rate to less than 25% at less than six months. The published parameters are used within the implementation of the model for this work and evaluated at doses which represent the doses delivered to the parotid glands for conformal and IMRT patients as detailed in 5.3.1. The three cases are referred to as “Rectum”, “Parotid (Conformal)” and “Parotid (IMRT)” for analysis purposes. The value of the dose within the model for the PARSPORT cases was determined from the patient treatment plans. The mean parotid doses were extracted from each of the patients plans and these have been used to determine the NTCP for these patients. The population NTCP value is determined from the mean of each patients NTCP. For the prostate rectal case the maximum dose to the rectum has been taken to have a mean of the prescription dose (74Gy) with a standard deviation of 4% as was the case for the conformal PARSPORT plans which were also planned using a conformal technique. A population of 100 patients was simulated in this case and the planned maximum doses held fixed for each patient throughout this work.

The parameters in Table 6.3 are supplied to the model which produces the NTCP results over the range of supplied doses. The LKB model is based upon the cumulative dose which can be determined from the individual fraction doses and the doses of interest selected correspond to realistic clinical doses. For the rectum, the maximum dose is assumed to be 100% of the prescribed dose. For the head and neck, two dose points are investigated. These correspond to the mean dose to the parotid glands for conformal plans and IMRT plans as discussed in section 5.3.1.
Radiobiological model implementation and characterisation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value: Mean (standard deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prostate</strong></td>
<td><strong>Head and Neck</strong></td>
</tr>
<tr>
<td>Grade 2 rectal bleeding</td>
<td>Xerostomia (reduction in salivary flow to &lt;25% at &lt; 6 months)</td>
</tr>
<tr>
<td>Value</td>
<td>Value</td>
</tr>
<tr>
<td>$D_{50}$</td>
<td>68.5Gy (1.1Gy)</td>
</tr>
<tr>
<td>$m$</td>
<td>0.15Gy$^{-1}$ (0.01 Gy$^{-1}$)</td>
</tr>
<tr>
<td>$n$</td>
<td>0.13 (0.02)</td>
</tr>
<tr>
<td>$v$</td>
<td>0.73</td>
</tr>
<tr>
<td>Nominal dose per fraction ($d$)</td>
<td>2Gy</td>
</tr>
<tr>
<td>Reference NTCP results</td>
<td>Grade 2 rectal bleeding within 5 years</td>
</tr>
<tr>
<td></td>
<td>36% @ 64Gy</td>
</tr>
<tr>
<td></td>
<td>47% @ 74Gy</td>
</tr>
<tr>
<td>Doses of interest *</td>
<td>Dose to rectum</td>
</tr>
<tr>
<td></td>
<td>74Gy (assumed 100% dose)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relevant references for model parameters:

**Prostate:** (Medical Research Council Radiotherapy Working Group, 2000; Dearnaley et al., 2007; Gulliford et al., 2012)

**Head and Neck:** (Semenenko and Li, 2008; Nutting et al., 2011; Miah et al., 2013)

* Determined from the planned dose distribution for the Head and Neck cases and assumed to be 100% of the dose with 4% SD for the prostate case.

Table 6.3: Parameters used for modelling prostate and head and neck patients within this work using the LKB NTCP model. The parameters are based upon the published literature listed and represent a range of different normal tissue complications observed clinically.
6.2.2.1 *Lyman Kutcher Burman model parameter sensitivity*

The sensitivity of the parameters of the LKB model, dose per fraction, $D_{50}$, $m$, and $n$, has been determined using the implemented model for the RT01 prostate and PARSORT head and neck case using the parameters in Table 6.3. Each parameter was varied between -10% and +10% of its nominal value and the difference in calculated NTCP at the specific doses was determined. The results are presented in Figure 6.4.

![LKB Model parameter sensitivity](image)

**Figure 6.4:** Plot showing the sensitivity of the calculated NTCP for the clinical cases detailed in Table 6.3 for each parameter. The parameters are the dose per fraction ($d$), $D_{50}(1)$, $m$, and $n$ as detailed in section 1.5.4. The prostate case is for Grade 2 rectal bleeding and the head and neck case is xerostomia.

The change in dose per fraction ($d$) is seen to be one of the largest influences on simulated outcome for all cases with the prostate case being most sensitive here.

As with the TCP results, the NTCP is seen to be highly sensitive to changes in delivered dose and so this is explored further throughout this work, quantifying the changes in NTCP due to changes in dose arising from different sources.
7 Clinical impact of dose variation

This chapter further explores the impact on clinical outcomes of variation in delivered doses. This work continues from the previous chapter and is based on the parameters as determined from clinical studies (see sections 6.1.4 and 6.1.5).

The variation in dose will be explored in terms of a percentage difference from the nominal dose delivered per fraction. Systematic and random dose variations will be examined. The magnitude of dose variations to be examined has been determined from the analysis of systematic and random dose variations as discussed in chapters 2-5.

7.1 Summary of dosimetric uncertainties

The dosimetric uncertainties quantified included the uncertainty in initial beam output calibration (see section 3.2.1) which would be systematic to all patients treated on a given machine. In addition to this should be included the variation in the beam output which consists of a systematic difference (see section 4.3.7.2) which may include a drift with time (as plotted in Figure 6.1) and a random uncertainty on top of this (see section 4.3.8). The uncertainties in dose delivered due to beam output calibration and drift are summarised in Table 7.1.

<table>
<thead>
<tr>
<th>Source of dose deviation</th>
<th>Relevant thesis section</th>
<th>Standard deviation</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A – Initial beam calibration</td>
<td>3.2.1</td>
<td>0.7%</td>
<td>1.4%</td>
</tr>
<tr>
<td>B – Systematic beam output deviation</td>
<td>4.3.7.2</td>
<td>0.7%</td>
<td>1.4%</td>
</tr>
<tr>
<td>C – Beam output daily fluctuations</td>
<td>4.3.8</td>
<td>0.2%</td>
<td>0.4%</td>
</tr>
<tr>
<td>Combined uncertainties:</td>
<td>A * B</td>
<td>1.0%</td>
<td>1.9%</td>
</tr>
<tr>
<td></td>
<td>A * B * C</td>
<td>1.0%</td>
<td>2.0%</td>
</tr>
</tbody>
</table>

Table 7.1: Summary of dose uncertainties as assessed in this thesis. The magnitude of these variations will be used to assess the impact of dose uncertainties on patient outcomes due to beam calibration and output variation. Convolution of parameters is indicated by the ‘*’ symbol.

The uncertainties due to the initial calibration and due to beam output variations following this are of equal magnitude. The random fluctuations in dose on a day-to-day basis are smaller but will still
be included in this work as these will affect individual patients and become more important for shorter treatments. The values in Table 7.1 will be used to assess the potential impact of dose uncertainty arising from calibration and beam output maintenance on clinical outcomes in the remainder of this chapter.

7.2 Summary of clinical cases considered

The clinical cases considered aim to highlight some interesting differences. The examination of TCP takes three cases which have different dose response curves, and so the potential difference between clinical sites and disease type can be assessed. Both the individual dose response curve for a ‘median’ patient and the population TCP curve are assessed. For the NTCP, three cases are also examined; one treatment site is split based upon the variation in dose delivered by different planning techniques. This allows change in the predicted impact of dose deviations based upon planning techniques to be quantified.

The following will explore the impact of the different sources of dose variation which may arise. This includes the impact of a systematic dose difference for all treatments and a drift in delivered dose over time whilst assuming no variation on these fixed values. The impact of the same patient being treated on different machines and the random variation in daily dose is also explored.

7.2.1 Cases considered for evaluation of dose variations on tumour control probability

The TCP measures considered and the data for these is described in section 6.2.1 and briefly recapped here. For TCP the “prostate case” considered is for 10 year bPFS (biological progression free survival) based on the RT01 trial results. The dose of interest considered is 74Gy (in 37 fractions) which had a TCP value of 55%. A second case with a much steeper dose response curve is also investigated (see Figure 1.9). This is the results from the Fox Chase study which reported bNED (biochemically no evidence of disease). A dose of 76Gy is considered with a TCP of 84%. The head and neck case is two year survival based on the PARSPORT trial. The dose of interest is 65Gy (in 30 fractions) with a TCP of 66%.

7.2.2 Cases considered for evaluation of dose variations on normal tissue complication probability

The NTCP is considered for three cases as described in section 6.2.2. The first is for rectal bleeding associated with prostate treatments from the RT01 trial. The remaining two cases are based upon the induction of xerostomia in head and neck patients. The two doses considered are representative
Clinical impact of dose variation

of the mean dose which may be received by the parotid glands if a head and neck cancer is treated using either conformal or IMRT techniques. These doses are 61Gy and 35Gy respectively.

7.3 Variation in outcome due to dose differences

Several situations arise which result in a patient receiving a delivered dose which has a systematic difference to that prescribed. These cases originate from the uncertainties in beam calibration and beam output as detailed in Table 7.1. For example, patients are often scheduled to a single treatment machine for the duration of their treatment, and so the beam output will vary dependant on scheduling. With a single machine dependent upon the treatment start date the beam output will vary. The probable range of these systematic shifts in delivered dose is explored here for individual patients and a population.

7.3.1 Variation in outcome due to a known systematic dose difference

The deviation from the nominal TCP with the given prescribed dose per fraction for a range of systematic dose shifts from -5% to +5% has been determined for both the TCP and NTCP models described. The dose is assumed to be equal for each fraction and the simulations were completed for a ‘typical’ patient (see Table 6.2). The change in both TCP and NTCP has been quantified below.

7.3.1.1 TCP

The change in TCP has been evaluated at the doses of interest given in Table 6.2 and the results of this are plotted in Figure 7.1 for individual patients, and Figure 7.2 for a population with corresponding numerical values for each given in Table 7.2.
Figure 7.1: Variation in calculated individual patient TCP (%) caused by a systematic shift in dose (%) throughout treatment. The shaded regions indicate the 95% CI for a convolution of the uncertainties in beam calibration (1.4%) alone and then combined with mean beam output (1.9% total) and beam output fluctuations (2.0% total) (see section 7.1). The values for this plot between a dose shift of ±3% are given in Table 7.2.

It is observed that the change in TCP varies in these different cases, however the impact on an individual patient in each case is comparable. The Fox Chase prostate case shows a greater change in TCP with dose. The difference arises due to the steepness of the dose response curves which is visualised for each case in Figure 7.4. The values in Figure 7.1 are given for dose shifts in the range -3% to +3% in Table 7.2. A greater difference is seen between the results of the populations between each case than for individual cases. This is seen in Figure 7.2 and is due to the differences in the range of $\alpha/\beta$ values within each population combined with the position on the dose response curve. Here a population of 1000 patients has been simulated with a range of dose shifts applied across the population.
Fluctuations in the values arise from the random uncertainties in generation of the population of patients during the simulations (see section 6.1.4.1). The shaded regions indicate the 95% CI for a convolution of the uncertainties in beam calibration (1.4%) alone and then combined with mean beam output (1.9% total) and beam output fluctuations (2.0% total) (see section 7.1). The values for this plot for a dose shift of ±3% are given in Table 7.2.

For the population TCP the Fox Chase case shows the greatest variation with dose as expected due to the very steep dose response curve. The RT01 and PARSORT cases have very similar responses to systematic dose shifts. The values are given in in Table 7.2 and show that the change in population response is approximately half that of an individual patient.

These systematic differences in delivered dose may arise for a number of reasons. Often patients are scheduled to a single machine for the duration of their treatment and each machine will deliver a slightly different dose, and if machines have different calibrations then a population of patients treated on one machine will have different absolute delivered dose when compared with another machine.
Clinical impact of dose variation

<table>
<thead>
<tr>
<th>Systematic dose shift (%)</th>
<th>Change in TCP (%)</th>
<th>Prostate RT01</th>
<th>Prostate Fox Chase</th>
<th>Head and neck PARSPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.0</td>
<td>-22.2</td>
<td>-10.8</td>
<td>-28.7</td>
<td>-17.3</td>
</tr>
<tr>
<td>-2.5</td>
<td>-18.1</td>
<td>-8.5</td>
<td>-23.6</td>
<td>-14.1</td>
</tr>
<tr>
<td>-2.0</td>
<td>-14.1</td>
<td>-7.2</td>
<td>-18.6</td>
<td>-10.7</td>
</tr>
<tr>
<td>-1.5</td>
<td>-10.4</td>
<td>-7.0</td>
<td>-13.7</td>
<td>-7.9</td>
</tr>
<tr>
<td>-1.0</td>
<td>-6.7</td>
<td>-2.7</td>
<td>-9.0</td>
<td>-5.1</td>
</tr>
<tr>
<td>-0.5</td>
<td>-3.3</td>
<td>-1.8</td>
<td>-4.4</td>
<td>-2.5</td>
</tr>
</tbody>
</table>

Table 7.2: Calculated TCP change based on a systematic dose shift throughout the entire treatment for an individual and population of patients. The 95\textsuperscript{th} percentile is ±1.9\% for combined calibration and beam output maintenance uncertainties which gives rise to variation in TCP of over 10\% for an individual patient and over 6\% for the population TCP changes. These values are plotted in Figure 7.1 and Figure 7.2.

For individual patients it is seen that with a systematic dose deviation of +2\% the change in TCP varies between 11.4\% and 15.6\% in these cases. A systematic reduction in dose yields a greater absolute change in TCP than an increase in dose with a -2\% shift giving -14.1\% to -18.6\% change. This is because the gradient of the dose response curve is non-linear, reducing when moving towards higher doses and increasing at lower doses.

For the simulated population the change is smaller and a systematic dose deviation of +2\% causes a change in TCP between 6.1\% and 8.6\% in these cases. A systematic reduction in dose by 2\% yields a change in TCP from -6.2\% to -10.7\%.
Clinical impact of dose variation

7.3.1.2 NTCP
The change in NTCP values at the doses of interest (see Table 6.2) has been determined for a range of systematic dose shifts. The results are plotted in Figure 7.3 and corresponding numerical values given in Table 7.3.

![Variation in calculated NTCP with dose variation](image)

**Figure 7.3:** Plots showing the impact of a systematic dose shift on the modelled NTCP for the three clinical cases. The relationship is approximately linear over the clinical range. Variations of up to 10% in predicted NTCP are observed within the range corresponding to the shaded clinical uncertainties quantified within this work.

The change in NTCP is greatest for the rectum case, owing to the steeper dose response curve. For the parotids, the IMRT case has a slightly greater change with dose than the conformal case.

The numerical values (as shown in Table 7.3) indicate that the clinical range of uncertainties (see section 7.1) of up to 2% yield a change in NTCP of up to approximately 8% for the rectum case and up to 3% for the parotid cases.
Clinical impact of dose variation

<table>
<thead>
<tr>
<th>Systematic dose shift (%)</th>
<th>Change in NTCP (%) for clinical case</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rectum</td>
</tr>
<tr>
<td>-3.0</td>
<td>-12.8</td>
</tr>
<tr>
<td>-2.5</td>
<td>-10.6</td>
</tr>
<tr>
<td>-2.0</td>
<td>-8.5</td>
</tr>
<tr>
<td>-1.5</td>
<td>-6.3</td>
</tr>
<tr>
<td>-1.0</td>
<td>-4.2</td>
</tr>
<tr>
<td>-0.5</td>
<td>-2.1</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.5</td>
<td>2.1</td>
</tr>
<tr>
<td>1.0</td>
<td>4.1</td>
</tr>
<tr>
<td>1.5</td>
<td>6.1</td>
</tr>
<tr>
<td>2.0</td>
<td>8.1</td>
</tr>
<tr>
<td>2.5</td>
<td>10.1</td>
</tr>
<tr>
<td>3.0</td>
<td>12.0</td>
</tr>
</tbody>
</table>

Table 7.3: Calculated NTCP change based on a systematic dose shift throughout the entire treatment. The 95th percentile is ±1.9% for combined calibration and beam output maintenance uncertainties. This gives rise to differences of 8% for the rectum case and up to 3% for the parotid cases. These values are plotted in Figure 7.3.

These potentially large differences in NTCP indicate the importance in dose control, and while they are not as large as for the TCP head and neck case discussed, they may still be significant for the patients if there is an increase in side effects, and thus a reduced quality of life even if there is a greater probability of cure. It is also important to note that the IMRT case shows a greater relative change than the conformal parotid case.

7.3.1.3 Impact of change in beam calibration uncertainty

It has been noted previously that the audit process results in a reduction in the variation in delivered doses. This was shown for the NPL audits of MV beams in section 3.2.1.1.1 in which initially there was a 0.8% standard deviation which reduced to 0.4% over the period of measurements. The change in the variation (standard deviation) of TCP and NTCP values arising from this change in beam calibration uncertainty is calculated for TCP and NTCP cases.

The change in the standard deviation of TCP and NTCP produced by the change in beam calibration uncertainty of 0.8% standard deviation to 0.4% standard deviation is given in Table 7.4.
Clinical impact of dose variation

<table>
<thead>
<tr>
<th>Standard deviation in beam calibration (%)</th>
<th>Prostate: RT01</th>
<th>Prostate: Fox Chase</th>
<th>Head and neck: PARSPORT</th>
<th>Rectum</th>
<th>Parotid: Conformal</th>
<th>Parotid: IMRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8%</td>
<td>5.3</td>
<td>3.6</td>
<td>7.1</td>
<td>4.0</td>
<td>6.3</td>
<td>3.2</td>
</tr>
<tr>
<td>0.4%</td>
<td>2.5</td>
<td>2.6</td>
<td>3.4</td>
<td>2.0</td>
<td>2.9</td>
<td>2.2</td>
</tr>
<tr>
<td>Change (%)</td>
<td>-2.8</td>
<td>-1.0</td>
<td>-3.7</td>
<td>-2.0</td>
<td>-3.4</td>
<td>-1.0</td>
</tr>
</tbody>
</table>

Table 7.4: The standard deviation of TCP results produced for a range of beam calibration uncertainties. The initial uncertainty in the MV beam calibration as determined through the NPL audits was 0.8%. It is observed that the results are not linear over this range. Note: values in the table are rounded to 1.d.p.

Reducing the variation in dose from 0.8% SD to 0.4% SD reduces the variation in TCP for an individual patient by up to 3.7% and up to 2% for a population in these cases. This variation is not insignificant as often clinical trials will be looking for changes of approximately 5% and so are comparable.

At the start of the NPL audit measurements, the uncertainty in dose delivery due to beam calibration was slightly greater than the uncertainty of the beam output maintenance as measured in this work. The final results from the NPL audits indicate that the beam output is now the largest source of uncertainty in dose delivery.

7.3.2 Variation in outcome due to machine scheduling

Many departments will aim to schedule patients to a single treatment machine for their entire treatment to give a sense of familiarity with staff and surroundings. This will result in a slightly different dose being delivered to different patients due to variations in machine outputs between the machines. Here, the potential variation in outcome measures will be assessed based on a patient being treated on 1000 different treatment machines as well as a population of 1000 patients being treated on these different machines. The machines are assumed to have a combined uncertainty of initial calibration and beam output of 1.0% (standard deviation) which is normally distributed. This is performed for the TCP and NTCP cases.
Clinical impact of dose variation

7.3.2.1 TCP

A plot showing the mean TCP and the 5th to 95th percentiles for these simulated treatments is presented in Figure 7.4 for individual patients and in Figure 7.5 for a population. On these plots the range of TCPs values associated with a given nominal dose is shown and these values will be further explored.

Figure 7.4: Plot showing the simulated individual patient TCP for the three clinical cases. The solid line indicates the mean values and the shaded region represents the 5th to 95th percentiles around this. The simulated variation arises from the mean dose delivered based on the machine to which the patient is assigned. The range of individual patient TCPs is similar for each case as indicated by the arrows at the appropriate doses of interest. The machines modelled had a normal distribution of beam output with mean of 0% and standard deviation of 1.0% matching the clinical situation of combined calibration and beam output variation uncertainty.

Large differences in calculated TCP are observed based solely on machine assignment due to the potential variation in delivered dose arising from the initial calibration and beam output uncertainties. The modelled case of the individual patients shows similar spread in TCP across the machines. Cases with steeper dose response curves are impacted to a greater extent by these dose variations. The 5th and 95th percentiles of the TCP are 63.1% and 75.3% (range 12.2%) for the RT01 prostate case, 54.6% and 69.1% (range 14.5%) for the Fox Chase prostate case, and 59.2% and 72.9% (range 13.7%) for the PARSPORT head and neck case.
Clinical impact of dose variation

The simulated population results as shown in Figure 7.5 has a much greater difference in variation between the three cases, emphasizing the increased importance of the initial machine calibration if there is a known steep dose-response.

![Variation in TCP based on machine assignment: Population](image)

**Figure 7.5:** Plot showing the simulated population TCP for the three clinical cases. The solid line indicates the mean values and the shaded region represents the 5th to 95th percentiles around this. The variation modelled arises from the mean dose delivered based on the machine to which the patient is assigned. The range of population TCPs varies for each case as indicated by the arrows at the appropriate doses of interest. The machines modelled had a normal distribution of beam output with mean of 0% and standard deviation of 1.0% matching the clinical situation of combined calibration and beam output variation uncertainty.

The Fox Chase case, which has the steepest population dose response curve, shows the greatest variation with 5th and 95th percentiles of 71.0% and 80.7% (range 9.7%). The RT01 prostate case and head and neck case had 5th and 95th percentiles of 52.5% and 58.9% (range 6.4%), and 60.1% and 66.8% (range 6.7%) respectively.

It is useful to compare the spread of simulated TCPs at the dose of interest for each case for the individual and population of patients. The change in TCP compared with the mean value at the dose of interest for each has been calculated (same data as discussed in Figure 7.4 and Figure 7.5) and plotted as a boxplot to compare the variation of each. This is presented in Figure 7.6. The
Clinical impact of dose variation

uncertainty in combined beam output and calibration to cause this variation in TCP is 1.0% standard deviation.

![Boxplots indicating the relative variation in TCP for individuals and populations due to calibration and beam output variations](image)

**Figure 7.6:** Boxplots indicating the relative variation in TCP for both prostate and head and neck cases due to the combined calibration and beam output uncertainty for individual patients and a simulated population. As a relative change from the nominal population TCP each of these cases is similar with approximately 4-5% SD in TCP introduced by 1% SD in dose. The individual patients show a greater deviation ranging from 6.4% to 7.6% SD for the same variation in dose. See Appendix D for description of boxplot display.

The variation in TCP arising from the uncertainty in the delivered dose due to initial calibration and beam output is approximately 4% for each of the population results examined here. The larger range of TCP values of the Fox Chase prostate case is offset by the larger nominal TCP value. As expected the individual patient variation is greater due to the steeper dose response curves of individuals when using the LQ model and is approximately 1.5-2x greater than the relative population variation introduced by the variation in dose.

### 7.3.2.2 NTCP

The variation in NTCP calculated due to machine assignment alone for the three populations is plotted in Figure 7.7. The range of NTCP values varies with each clinical case. The range of values at
Clinical impact of dose variation

these points is further explored. Each population of patients was simulated on 100 machines with a 1% standard deviation in dose between them. The variation after the final fraction of treatment for each case is indicated in the plot.

![Variation in NTCP based on machine assignment](image)

**Figure 7.7:** Plot showing the modelled NTCP for the assessed rectum and parotids clinical cases. The solid line indicates the mean value and the shaded region is the 5\(^{th}\) to 95\(^{th}\) percentile. The variation arises from a range of delivered doses arising from the calibration and beam output uncertainties (mean of 0% with a standard deviation of 1.0%). The rectal dose response curve is the greatest at the dose of interest and so has a greater range of NTCP values (indicated by the arrows).

As with the TCP there may be a large range of NTCP values dependant only upon the machine on which the patients is treated on arising from uncertainties in beam calibration. The 5\(^{th}\) and 95\(^{th}\) percentiles of the NTCP are 55.3% and 64.2% (range 8.9%) for the rectum case. For the parotids the variation is approximately a quarter of this with a 5\(^{th}\) to 95\(^{th}\) percentile range of 2.0% and 2.7% for the conformal and IMRT cases respectively.

7.3.2.2.1 Further detail on the spread of NTCP caused by beam output uncertainty

Here the spread of the above results is further assessed for each case. A boxplot of the results at the doses of interest for each case is given in Figure 7.8.
Figure 7.8: Boxplots showing the variation in NTCP for the clinical cases at the end of treatment. This variation is caused by the combined 1.0% standard deviation of beam output and calibration. See Appendix D for description of boxplot display. The relative variation of the rectum case is greatest due to the steeper dose response as seen in Figure 7.7.

When visualised the variation introduced to the rectum case is clearly larger than for the parotid cases. The 5th to 95th percentile range is 8.7%, 2.6% and 1.7% for the rectum, parotid IMRT, and parotid conformal cases respectively. For the rectal case this is particularly large due to the steep dose-response curve.

7.3.3 Variation in outcome due to drift in dose over course of treatment

The drift in beam output over time was examined in section 4.3.5 with a mean drift of +0.9% per year and a range of -8.9% to +8.4%. The standard deviation of this drift was 2.27% per year. The distribution of output drifts is assumed to be normally distributed and this results in a 95% CI of -3.5% to +5.3% per year. This drift will cause a different dose to be delivered for each treatment and so the impact of this will be examined for each treatment case.

Assuming the initial beam output is set to deliver the exact prescription dose, a range of beam output drifts between -10% and +10% per year has been included in the calculation of the final TCP and NTCP. This also assumes that no machine calibration is performed during the treatment period.
Clinical impact of dose variation

7.3.3.1 TCP

A plot of the impact of a drift in dose over the course of the treatment is shown in Figure 7.9 for individual patients.

![Plot showing the effect of beam output drift on predicted TCP on an individual.](image)

**Figure 7.9:** Plot showing the effect of beam output drift on predicted TCP on an individual. The shaded range indicates the 95% CI of drifts measured in this work. The Fox Chase case shows the greatest variation. A 5% annual increase in beam output gives 2% increase in TCP for the Fox Chase case and 1.5% for the RT01 and PARSPORT cases.

The relationship between the drift in beam output and change in TCP is approximately linear over this range as it is relatively small. It should be noted that over the course of a treatment of 6 weeks the shift in dose for a 5% annual drift is 0.6%, and 95% of machines have a drift of smaller magnitude. As this drift is generally small over the course of a single patient’s treatment the impact of the output drift is smaller than the initial beam output which was shown in Figure 7.1. The magnitude of the TCP variation induced by output drift is similar to that of the daily beam output fluctuations.

For a population the magnitude of the impact of drift in dose is in Figure 7.10. As in this case a population is simulated there are additional uncertainties in generating this population and so there are fluctuations visible within the results, however the general trend is still apparent and indicated by a linear fit in the figure.
Clinical impact of dose variation

Figure 7.10: Plot showing the effect of beam output drift on predicted population TCP. A linear fit is shown for each to indicate the general trend. The shaded region indicates the 95% CI of drifts measured in this work. The larger the variation in \( \alpha/\beta \) values the more ‘noise’ within the simulations assessing dose drift, and so large numbers of patients would be required to clinically observe differences due to drift in beam output through the course of a treatment.

The Fox Chase prostate case with its small variation in biological parameters used during simulation shows a clear trend, however the RT01 and PARSPORT cases have clearly visible fluctuations within the results, but still the trend is visible (as indicated by the linear fits within the figure). The larger the variations in biological parameters, the greater the number of patients which would be required to be able to identify the differences introduced by a drift in dose over the course of a typical treatment. A large drift in dose would have to exist to be clinically noticeable, or a very well stratified patient sample would be required. The impact of a dose drift is much smaller than the mean beam output (see section 7.3.1.1).

7.3.3.2 NTCP

The results generated for the NTCP including different magnitudes of dose drift is shown in Figure 7.11.
Clinical impact of dose variation

Figure 7.11: Plot showing the effect of beam output drift on NTCP. The shaded region indicates the 95% CI of drifts measured in this work. The impact on the parotid gland NTCP is less than 0.5% even for drifts of up to ±10%/year. For the rectum the variation in NTCP is larger, exceeding 0.5% for a drift of 5% per year.

The impact of the drift is approximately linear as over the course of a treatment the change will be relatively small. In comparison with the impact of the mean output (section 7.3.1) the drift plays a smaller role in determining the outcome and will have less of an impact for shorter courses of treatment.

7.3.4 Variation in outcome due to daily output variations

The measurement variation on the daily delivered doses was assessed in section 4.3.8 and its impact is assessed here for individual patients and a population. A range of daily dose delivery uncertainties (quantified by the standard deviation) between 0% and 2% is modelled for an individual patient with fixed biological parameters and for 1000 individual patients with varied α/β values forming a population. The impact of this on the variation in calculated TCP and NTCP was then determined.

7.3.4.1 TCP

The range of calculated results is displayed in Figure 7.12 for individual patients, and Figure 7.13 for a population, and corresponding numerical values are given in Table 7.5. This shows the variation in TCP due to the range of daily dose delivery uncertainties.
Clinical impact of dose variation

Figure 7.12: Plot showing the standard deviation of individual patient TCP produced due to random variation in daily delivered dose as measured by the standard deviation of the beam outputs. Clinically these random fluctuations have a smaller impact than those caused by a systematic shift in delivered dose. A range of the plotted values are given in Table 7.5. The clinically determined standard deviation of daily delivered dose was 0.2% and the 95\textsuperscript{th} percentile was 0.7% and these are indicated by the shaded regions.

The variation in outcome due to the daily fluctuations is seen to be relatively small relative to a systematic offset in beam output (see section 7.3.1). The 95\textsuperscript{th} percentile of the clinical range determined gives a variation in TCP of 1\% (SD) whereas the mean beam output based upon assigned machine gives rise to a variation in TCP of 3.5\% (SD).

The results when simulating the population TCP are shown in Figure 7.13. For this simulation a population of patients was generated and the $\alpha/\beta$ values for each individual held fixed. This population was then ‘treated’ 1000 times with a varied value for the SD of daily dose variations to determine the populations TCP variation.
Clinical impact of dose variation

Figure 7.13: Plot showing the standard deviation of population TCP produced due to random variation in daily delivered dose as measured by the standard deviation of the beam outputs. Clinically these random fluctuations have a smaller impact than those caused by a systematic shift in delivered dose. A range of the plotted values are given in Table 7.5. The clinically determined standard deviation of daily delivered dose was 0.2% and the 95\textsuperscript{th} percentile was 0.7% and these are indicated by the shaded regions.

The variation in the population TCP caused by daily dose fluctuations is much smaller than for an individual patient (almost 2 orders of magnitude). This is because if you were able to repeat the irradiations of a population multiple times then on average each would be likely to receive the nominal dose if enough repeats were performed. Thus, clinically the daily variation will have very little noticeable impact on the overall measured outcome of a large population of patients.
Clinical impact of dose variation

<table>
<thead>
<tr>
<th>Daily delivered dose standard deviation (%)</th>
<th>TCP standard deviation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prostate: RT01</td>
</tr>
<tr>
<td></td>
<td>Individual</td>
</tr>
<tr>
<td>0.1</td>
<td>0.10</td>
</tr>
<tr>
<td>0.2</td>
<td>0.21</td>
</tr>
<tr>
<td>0.3</td>
<td>0.31</td>
</tr>
<tr>
<td>0.4</td>
<td>0.43</td>
</tr>
<tr>
<td>0.5</td>
<td>0.54</td>
</tr>
<tr>
<td>0.6</td>
<td>0.64</td>
</tr>
<tr>
<td>0.7</td>
<td>0.76</td>
</tr>
<tr>
<td>0.8</td>
<td>0.83</td>
</tr>
<tr>
<td>0.9</td>
<td>0.94</td>
</tr>
<tr>
<td>1.0</td>
<td>1.03</td>
</tr>
</tbody>
</table>

Table 7.5: Quantified values of the variation in standard deviation of the TCP due to daily beam output fluctuations arising from the simulations. Values are plotted in Figure 7.12 and Figure 7.13. The clinically determined standard deviation of daily delivered dose was 0.2% and the 95th percentile was 0.7%.

The 95th percentile leads to a variation in TCP of 0.9% for the head and neck case for an individual patient, whereas the 95th percentile for the dose drifts gives 1.5% (see Figure 7.9) and the 95th percentile of the mean beam output gives rise to a difference of over 10% (see Figure 7.1). The variation in a population averaged TCP is less than 0.1%.

7.3.4.2 NTCP

The results are presented for changes in daily dose variation for a range of 0 to 2% standard deviation in Figure 7.14 for each case.
Clinical impact of dose variation

Figure 7.14: Plot showing variation of NTCP produced (standard deviation) due to random daily dose variations with different uncertainties (standard deviation). Within the clinical range these variations introduced are relatively small.

Over this range of dose variation the relationship is approximately linear. The values in the plot up to a 1% standard deviation in dose are given in Table 7.6.

<table>
<thead>
<tr>
<th>Daily delivered dose standard deviation (%)</th>
<th>Rectum</th>
<th>Parotid (conformal)</th>
<th>Parotid (IMRT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.007</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>0.2</td>
<td>0.014</td>
<td>0.005</td>
<td>0.007</td>
</tr>
<tr>
<td>0.3</td>
<td>0.019</td>
<td>0.006</td>
<td>0.010</td>
</tr>
<tr>
<td>0.4</td>
<td>0.027</td>
<td>0.010</td>
<td>0.012</td>
</tr>
<tr>
<td>0.5</td>
<td>0.033</td>
<td>0.011</td>
<td>0.017</td>
</tr>
<tr>
<td>0.6</td>
<td>0.042</td>
<td>0.014</td>
<td>0.019</td>
</tr>
<tr>
<td>0.7</td>
<td>0.045</td>
<td>0.016</td>
<td>0.024</td>
</tr>
<tr>
<td>0.8</td>
<td>0.053</td>
<td>0.019</td>
<td>0.025</td>
</tr>
<tr>
<td>0.9</td>
<td>0.063</td>
<td>0.020</td>
<td>0.028</td>
</tr>
<tr>
<td>1.0</td>
<td>0.067</td>
<td>0.023</td>
<td>0.033</td>
</tr>
</tbody>
</table>

Table 7.6: Values of the variation in standard deviation of the NTCP introduced by variations in daily delivered dose (quantified by the standard deviation). The standard deviation of the clinical machines analysed in this work was 0.2% and these are all less than 0.1%.
As a proportion of the overall dose uncertainty the variation in daily dose around the mean beam output is small. This is evident in the impact on NTCP of these daily variations. A clinically relevant standard deviation of daily dose of 0.2% gives rise to a variation in NTCP of <0.1% for each case.

7.3.5 Assessment of impact in systematic dose differences on population of patients

Here an attempt is made to estimate absolute numbers of patients impacted by the variations in delivered dose. The size of the population used has been determined from the Malthus software (Malthus Programme development team, 2011) which is a tool used to model the radiotherapy demand across the UK based upon the existing evidence base (National Cancer Action Team, 2011; Jena et al., 2012, 2015; Sanderson et al., 2017). Using Malthus the predicted number of radiotherapy courses for prostate and head and neck cancer across the UK in 2015 was determined to be 26318 and 8093 respectively and so these have been utilised in these simulations to allow quantification in terms of absolute patient numbers (using the RT01 and PARSPORT models respectively). The NTCP cases are not considered here as there would be a large range of delivered dose to the OARs and the implemented models are not well suited to include this additional variation.

A set of patients for each of the prostate and head and neck cohorts was generated with a spread of $\alpha/\beta$ ratios as detailed in Table 6.2. Each of these patients has then been modelled through treatment with a fixed dose per fraction with a systematic shift in the dose between ±5%. From this the change in the number of patients which are predicted a TCP of <10% and >90% are calculated and the variation due to the dose shift assessed. The results are tabulated in Table 7.7.
Clinical impact of dose variation

<table>
<thead>
<tr>
<th>Systematic dose shift (%)</th>
<th>Number of prostate patients</th>
<th>Number of head and neck patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TCP &lt;10%</td>
<td>TCP &gt;90%</td>
</tr>
<tr>
<td>-5</td>
<td>9892</td>
<td>7930</td>
</tr>
<tr>
<td>-4</td>
<td>9178</td>
<td>8397</td>
</tr>
<tr>
<td>-3</td>
<td>8626</td>
<td>8642</td>
</tr>
<tr>
<td>-2</td>
<td>8086</td>
<td>8989</td>
</tr>
<tr>
<td>-1</td>
<td>7441</td>
<td>9403</td>
</tr>
<tr>
<td>0</td>
<td>6865</td>
<td>9668</td>
</tr>
<tr>
<td>1</td>
<td>6282</td>
<td>10039</td>
</tr>
<tr>
<td>2</td>
<td>5760</td>
<td>10433</td>
</tr>
<tr>
<td>3</td>
<td>5077</td>
<td>10849</td>
</tr>
<tr>
<td>4</td>
<td>4461</td>
<td>11463</td>
</tr>
<tr>
<td>5</td>
<td>3932</td>
<td>11745</td>
</tr>
</tbody>
</table>

Table 7.7: Modelled number of patients with predicted TCP of less than 10% and greater than 90% for a fixed systematic dose shift to the population. Included are the cases for prostate and head and neck cancers which have a total number of patients modelled of 26318 and 8093 respectively.

Over a realistic possible clinical range of ±2% the number of prostate patients with a calculated TCP of greater than 90% changes from 8989 to 10433 with a -2% and +2% systematic dose shift to all these patients. This is an increase of 1444 patients. A change in the number of prostate patients with a TCP of <10% is seen to change by a similar magnitude (8086 patients at -2% to 5760 patients at +2%). For the head and neck case, the change in the number of patients is around half that of the prostate case, but it should be noted that there is only one third the number of patients.

The variation in response due to biological factors is large, however the impact of a systematic dose difference to a population is seen to also be significant. The above simulations could represent sets of patients treated on two machines which had incorrect initial calibrations and thus on one machine the number of patients with a predicted a TCP of >90% when compared with the other (assuming patients with a similar distribution of biological response) could be significantly different. This appears to be true for cases with a steep dose response curves (such as the PARSPORT head and neck case) and shallower curves (such as the RT01 prostate case). From this it could be concluded that the initial calibration remains vitally important in the long term as all future calibrations and delivered doses are often relative to this.
7.4 Comparison with previous work based on European TLD audits

Work by Bentzen et al. (Bentzen et al., 2000) used results from postal audits conducted by the EORTC across Europe to predict the variation in clinical outcomes due to the dose deviations as measured during the audits. Here, I will present a repeat of this work using results from the NPL audits described in chapter 3 and compare to the method used in this thesis.

7.4.1 TLD audit data used

Dosimetry audits using mailed TLDs have been conducted throughout Europe since 1987 (Bentzen et al., 2000; Clark et al., 2015; Hurkmans et al., 2016; Izewska, Lechner and Wesolowska, 2018). TLDs are mailed to the host centre who then delivers a pre-determined dose, send them back for analysis and then the measured dose is compared with the expected value. The deviations in dose are currently classified as acceptable (≤4% difference), minor (between 4% and ≤7%) and major (>7%) (Bentzen et al., 2000). Initial results were published by Hansson et al. (Hansson et al., 1993). A total of 714 measurements was included within the initial results from 55 centres including 357 beams and nine results were identified as having a major deviation (Hansson et al., 1993). After a second round, major deviations were only found at 1 of these centres. Of the measurements 90% were within ±4% with a mean difference of +0.7% and standard deviation of 4%.

Based on these results, Bentzen et al. used a multiplicative factor to predict the impact of these calibration discrepancies on clinical outcomes. The factors used aimed to be representative of a wide range of cancers. To change a percentage difference in dose to a percentage change in clinical outcome, factors of 2.3, 5.2 and 0.9 were used for tumour control, mild reactions and severe complications respectively.

7.4.2 Prediction of clinical outcome using NPL audit results

Using the MV photon NPL audits data as described in Chapter 3.2 the methodology employed by Bentzen has been repeated. Using the same multiplicative factors to convert dose deviation to outcome deviation.

Bentzen split beams into two groups; <=6MV and >=8MV (note there were no 7MV beams in either dataset). Both sets were determined to be normally distributed (p>0.05) and the beams <=6MV (n=52) and >=8MV (n=47) showed no significant difference between the two groups (p=0.59) when compared using a t-test. In this case all beams were combined for result generation. A plot showing
the proportion of treatment machines predicted to display a given percentage deviation in TCP, mild and severe NTCP measures is presented in Figure 7.15.

The 10% of beams which gave the greatest under dose would result in a 2.2% reduction in TCP. The 10% of beams that gave the greatest over dose would result in an increase in mild reactions of 4.0% and 0.7% increase in severe reactions across the population. With an assumed incidence of 5% for severe complications (as in the original paper), this increase in incidence of 0.7% is actually an increase in frequency of severe complications of 14% on those machines.

Figure 7.15: Plot showing the predicted difference for NPL absolute dosimetry audit data. The proportion of beams exhibiting a given change in clinical outcome is shown on the y axis for change in TCP (blue), mild NTCP (orange) and severe NTCP (green).

7.4.3 Discussion and comparison to Bentzen et al. results
The NPL audits were based on on-site measurements taken with an ionisation chamber, whereas the measurements used by Bentzen were based on a mailed TLD service and so these carry a greater uncertainty. Within Bentzen’s paper it is stated that the TLD measurements have a combined uncertainty of approximately 2% (1 SD). They also state that this may be an over estimation and may be closer to 1% (1 SD). The NPL reports give an estimated standard uncertainty of 0.4%. The greater uncertainty in the TLD measurements will act to increase the spread of measurements and this is supported by the results.
Clinical impact of dose variation

Despite the uncertainties associated with the measurements, this simple prediction of the impact of beam calibration on clinical outcomes acts to give a powerful suggestion of the size of the effect which might be expected on clinical outcomes due to dose deviations arising from within the dosimetry chain. While these changes in dose are small, they should be considered in the wider scope of improvements which are sought through clinical trials. Bentzen suggests combined chemoradiotherapy gives an increase in survival of 4%. Taking this as a benchmark we can see the difference between the predicted TCP in the highest and lowest 10% of beams to also be approximately 4% which is comparable.

Regardless of the methodology of audit (postal, on-site, etc), improvement is consistently observed with repeated visits indicating its worth in ensuring radiotherapy continues to improve in precision, accuracy and ensures the safe implementation of new treatment techniques.

While Bentzen's analysis used fixed factors to convert the dose deviations to clinical deviations, the methodology presented in this work provides an alternative approach which can be applied to specific situations and easily adapted for a number of different purposes. Practically, once an initial set of modelling has been performed for a particular treatment type and site it may be useful to simply quantify the gradient of the dose-response curve at a particular dose point and use this to quantify any differences. Whilst the response is not completely linear, over a small range of doses a linear approximation may be all that is required to determine tolerances required for a given situation.

7.5 Discussion of radiobiological modelling results

The results of the modelling simulations performed indicate that the variation in dose due to variations in beam output alone may be significant for individual patients. The impact of these variations is highly dependent upon the steepness of the dose response curve, and the curves for individual patients may be very steep and depend upon the particular model parameters appropriate to that clinical case.

When using the LQ model the variation in TCP due to dose variations is heavily influenced by the $\alpha/\beta$ ratio, with a greater impact for higher ratios for an individual patient. For a population of patients the variation of $\alpha/\beta$ values within this population influences the steepness of the dose response curve and so the impact on each case should be assessed independently to determine any impact of dose variations. The $\alpha/\beta$ ratio values used in this work are representative of the clinically used values. Currently each patient treated is assumed to have the same tissue radiosensitivity in clinical
Clinical impact of dose variation

practice. Looking into the future there is potential that methods will be developed to better predict an individual patient’s response to treatment either before the treatment commences, or by monitoring the response during the treatment (Tariq et al., 2015; Chaudhuri et al., 2016) and thus allowing adaptations to be implemented to improve chances of survival, or minimise toxicities to acceptable levels.

If in the future it becomes possible to predict individual patient response to radiotherapy, then the next logical step is to further tailor the treatment dose based on this information. For example, if it is known that a patient requires only 30Gy to eliminate their cancer, then any additional dose will only act to increase toxicity with no further curative gain. If this becomes standard practice, then examining the TCP of the population as a whole becomes less meaningful and individual response curves must be addressed. These are steeper than that for a population (as discussed in section 1.5.2 and visualised in Figure 1.8). This results in the variation in dose having a relatively much greater influence if the inter-patient biological differences are accounted for by the modification of prescriptions, and so the control of dose becomes ever more important. Of course, some patients may be identified as requiring a very large radiation dose to achieve the desired TCP. In these cases the impact on NTCP would come into play (assuming it could also be similarly predicted on an individual basis). It may identify patients for which radiotherapy would never provide any benefit, and so would not have to be subjected to unnecessary treatment. As individual patient responses are able to be better quantified the uncertainty in this aspect of the radiobiological models will reduce allowing more certainty in predictions.

Another aspect to consider is the improvement in technology and how this may aid in controlling the delivered dose. Clearly, advances in treatment planning and delivery techniques have come on huge amounts over the recent decades, now allowing highly conformal dose distributions to be routinely delivered. One area of interest is the possibility of delivering a ‘plan of the day’ based on the patient’s anatomical variations on a day-to-day basis (Staffurth et al., 2015; Kleijnen et al., 2016; Kron, Lehmann and Greer, 2016; Sharfo et al., 2016). This is one step on from accounting for external patient movement (through techniques such as breath hold) and patient positioning (through the increased use of imaging). However, this does not address the variation in delivered dose due to output variations on these days. It is not difficult to see a situation where the treatment machine could simply scale its delivered dose based upon measurements taken on the same day. Varian’s Halcyon treatment machine will not allow treatment to commence until it has satisfactorily passed its own checks of beam output which make use of the imaging panel (Barfield, 2017). Simply applying a correction factor based on these measurements would be technically simple to
implement and may reduce the variation in delivered dose for patients treated on different machines.

An alternative to ‘plan of the day’ might be real-time tumour tracking to shape the radiation field as the treatment is delivered. Recent advances in this show its feasibility (Fast et al., 2016; Pollard et al., 2017; Keall et al., 2018) with existing machines and this would enable margins to be significantly reduced (particularly in very mobile tumours), thus further increasing the relative significance of any dose deviations which exist.

The NTCP results are dependent on cumulative dose to the organ when using the LKB model. In the clinical situations addressed in this work the variation in NTCP was in general less than the TCP as the dose response curves were shallower. These cases considered a single patient, however a group of patients would have a large range in delivered doses to the OARs (as seen in Figure 5.1), and so this would result in a large variation in the predicted clinical outcomes across the population which is likely to mask the influence of the dose variations addressed in this work. However, it is still the case that on an individual patient level the dose variations could result in large variations in induced side effects.
8 Conclusions and further work

8.1 Conclusions

The objectives of this work were to quantify the dosimetric variations arising from beam calibration and its maintenance and apply these to radiobiological models to predict the impact of this on patient outcomes. While the beam output of radiotherapy treatment machines is routinely measured it is rarely, if ever, included when considering the doses delivered to patients. This thesis has quantified the variation of doses delivered to patients within the UK based upon national dosimetry audit data as measured by the NPL which represents the uncertainty in the initial beam calibration. Further to this the beam output varies over time and this has been analysed based on a large dataset of beam outputs from 204 treatment machines which is the largest study of its kind. The combination of these uncertainties gives the overall uncertainty in delivered dose to the patients. Established radiobiological models have been used to predict the variation in outcomes due to these dose differences for ‘typical’ patients which may be considered significant.

While dosimetric audits of beam calibration are not new, the application of the results to radiobiological models as implemented in this work is. This application gives interesting insights into the impact of imperfections in the calibration chain. Improvements in dosimetric accuracy due to repeated audits have been demonstrated previously across Europe (Izewska et al., 2015) and also locally in the UK (Palmer et al., 2011). This work has quantified the improvement over time in MV beams, the standard deviation of results reducing from 0.8% to 0.4% between the first 20 audits and most recent 20 audits conducted. This is similar to the improvement observed by Palmer et al. which analysed the results from the South East regional audit group over similar time period.

The variation in dose delivered following machine calibration has been assessed based upon a larger dataset of 25,000 beam output measurements taken in 52 centres across the UK. The standard deviation in mean machine output across this six-month period was 0.7% (6MV beams) and there were large variations within a single centre. The maximum range of beam outputs within a single centre was 2.1%. This variation between treatment machines is approximately equal to the overall variation of MV beam calibrations measured (also 0.7%), but larger than the 0.4% standard deviation measured most recently in the NPL audit work, indicating that the variation in dose following initial calibration now has a greater impact on the dose delivered to the patient than uncertainties arising within the calibration chain. A few measurement sets did not include the complete set of gathered
data obtained during the QC testing, and so there will be differences in the spread of measurements due to this. For example, some centres will have included measurements which may have been excluded clinically, whereas others have filtered this from the supplied data. Overall it is felt this is unlikely to cause a significant change in any of the results presented within this thesis.

The drift in beam outputs over time was assessed for each machine and there was a large range of drifts determined. The mean drift was +0.9% per year and smaller single centre studies have previously reported results of the same magnitude (Luketina and Greig, 2004; Hossain, 2014). This work did not identify any definitive differences due to machine age, manufacturer, or model. It would have been useful to know the age of the primary ionisation chamber within each machine as this is likely to have been better correlated with the change in sensitivity over time than the age of the machine itself. As reported previously by Hossain (Hossain, 2014) there exists a seasonal variation in beam output and this was clearly identified within the long-term data from two of the linacs at the RSCH. This was masked by random fluctuations and drifts over time on the other linacs, and alternative correction methods for the data may have provided a clearer picture of the seasonal variations. The overall dataset did have a small upward trend over the period of January to June which would also be consistent with the same seasonal trends; however an additional six months of data would have been required to assess this. Over the course of a patient’s treatment the drift in beam output plays a minor role when compared with the mean output during the treatment. In practice the most significant impact of large drifts in beam output is on the resources required to maintain the machines output to within accepted limits.

As part of this work scripts were developed to extract dosimetric data from treatment plans in DICOM format. This was achieved, and data from patient plans used within the PARSPORT trial were assessed by extraction of dosimetric parameters from the dataset. The PARSPORT trial assessed the use of conformal and IMRT plans for sparing the parotid glands during head and neck radiotherapy and showed that the reduction in mean dose to the parotid glands corresponded with a reduction in toxicity. While this reduction in doses to OARs is well known when comparing IMRT to conformal techniques, a reduction in the variation in doses to the target volumes identified in this work is not. The automated extraction of dosimetric parameters and subsequent analysis revealed that the variation in doses to the primary target was reduced when using IMRT. The standard deviation in the mean dose reduced from 2% to 0.5% of the prescription dose. When considering this alongside the uncertainties in dose delivery, the switch to IMRT has resulted in the dose variation due to variations in beam output to be larger than those due to planning techniques.
The implementation of the LQ and LKB models to allow fraction by fraction dose specification has allowed the impact of dosimetric uncertainties as determined in this work to be assessed for a patient with ‘typical’ model parameters based upon the clinical cases assessed as well as for a simulated population. Based upon the PARSORT trial the upgrade to IMRT has reduced the mean parotid dose to a point on the dose response curve where changes in dose produce a larger relative percentage change in response than for the doses delivered with conformal plans. The largest variation in TCP was seen for Fox Chase prostate case due to its very steep dose response curve.

With the increasing use of image guided radiotherapy (IGRT) techniques patient positioning is becoming more accurate and precise. Based upon local protocols, changes in dose of 2% due to changes in patient anatomy or setup may result in treatment plans being recreated, yet a 2% difference in dose due to a beam output shift may persist for many weeks before being altered, potentially affecting multiple patients. The doses delivered through imaging during a course of radiotherapy are increasingly being scrutinised. If daily volumetric imaging is performed (as may be required for some IGRT) the increase in dose may increase by around 1% of the prescribed dose (Alaei and Spezi, 2015; Zhang et al., 2015). An increase in the magnitude of delivered dose by 1% or more will be experienced by over 10% of patients (based upon a normal distribution with standard deviation of 0.7%) due to beam output variations. If the imaging dose is to be considered when making clinical decisions, then it seems right that the variation in beam output should also be considered in the same way.

Using the LQ and LKB models this thesis has quantified the impact of the variation present in delivered dose arising due to beam calibration and subsequent maintenance of beam output within a clinical setting. This includes a multi-centre analysis on the routine measurement of beam output across the UK which is the first on this scale. It has been determined that these variations in beam output are now likely to be greater than the variation arising due to initial calibration. The radiobiological models were used to investigate the impact of systematic shifts, drifts over time and random fluctuations of beam output on clinical outcomes for typical clinical cases. For cancers with steep population dose response curves such as most head and neck cancer ($\alpha/\beta=10$ within LQ model), and the Fox chase prostate case examined in this work, changes of 1% in the delivered dose result in an increase in TCP of around 5% which is significant. For an individual patient the change may be approximately double this for all cases. The variation in dose has been placed in a wider context through analysis of the variation in planned dose to the primary target of head and neck patients within the PARSORT trial. The change from conformal to IMRT planning has resulted in the dose variation from beam output now being larger than the variation as measured from generated.
Conclusions and further work

Treatment plans, yet this is not considered when recruiting for clinical trials. A reduction in the tolerances in the measurement of beam output would result in greater certainty in the delivered dose to both targets and OARs. The analysis performed within this work indicates the measurement equipment used is capable of measuring to the precision required for this. The models used within this thesis, while well accepted, may turn out not to be the best suited to these particular cases. It is believed that the predictions given are likely to be reasonable estimates of the impact of dosimetric variations arising within the clinical calibration chain.

8.2 Further work

This work has highlighted the impact of the uncertainties within the chain of dosimetry leading to patient dose delivery and identified that significant changes in predicted outcome may be introduced even if all currently recommended tolerance levels are met. There are a number of different areas of this work which merit further development and investigation and these are described here. Most of the future work involves applying the identified dosimetric uncertainties to situations which would benefit from inclusion of these. This work focussed on a few distinct clinical cases. The RT01 trial had included a wide range of prostate patients, and so further separate analysis for different subsets of patients would be desirable to better quantify the impact dependent upon the local patient mix. The Fox Chase data indicates the potential wide range of results which may be obtained for similar treatment sites but with different tumour characteristics.

8.2.1 Inclusion of beam output uncertainty into clinical trials

This thesis has shown the importance of the variations in dose present due to uncertainties in beam output. It is noted that these uncertainties are not routinely included in the assessment of patient doses, either locally, or on a larger scale. It is thought that there are many areas of current research which may benefit from the inclusion, or appreciation of the variations in dose which arise.

The initial stages of a clinical trial involve the calculation of the number of patients required to achieve a given power. For a complete assessment of the uncertainties, the variation in dose should be included within these calculations. The input of these statistical calculations is obviously very dependent upon the clinical cases being investigated, but often the magnitude of the difference in effect being investigated might be similar to the magnitude of the uncertainty in the dose variations.

Considering the CHHiP trial for example, this was designed to identify a change in outcome of 5% (Dearnaley et al., 2016). Based on the RT01 results analysed within this thesis, a change of approximately 2% in dose produces an effect of this magnitude. In a large-scale clinical trial such as
Conclusions and further work

this, there is likely to be less of an impact due to the different arms of the trial having a spread of doses with mean around 0% from that prescribed. However for smaller scale trials it may be possible for the beam output to have a reasonable large effect. For example, a trial conducted within a single centre could feasibly be comparing two different planning techniques. These techniques may only be available on a selection of the treatment machines and so each trial arm may be assigned to a different machine for their treatment. In this case it is possible that a significant difference in dose may be delivered to each trial arm simply due to the beam output on each of the treatment machines which may be of a similar magnitude to the effect being measured and may be advantageous to consider when designing the trial or analysing the results.

8.2.1.1  Assessment of dose variation within pre-clinical trials.

One area of work which is often lacking in terms of precise dosimetry is that of pre-clinical radiotherapy research. The NPL has recently begun development of a pre-clinical dosimetry service (National Physical Laboratory, 2016) to aid in the traceability of delivered doses in pre-clinical research.

There are two aspects where this work may be able to play a role. The first is that the importance of the maintenance of the initial calibration should be stressed to the departments performing this research. If there are known changes in the delivered dose, these should be included within the analysis to properly quantify the uncertainties associated with the findings. Secondly the modelling approach used within this thesis could be used to determine the impact of the variations in the machine outputs.

As well as specialised irradiation facilities, often irradiations of tissue or cellular samples are performed within a clinical department on the same machines as used for patients. Often the more homogeneous the sample being irradiated the steeper the dose response curve. This leads to small changes in delivered dose potentially being significant, and so clinical departments which partake in these experiments should aim to provide as much information as possible to allow an accurate quantification of the delivered dose and associated uncertainties.

8.2.2  Improving the prediction of toxicity for individual patients

Studies are underway aiming to link the 3D dose distribution to toxicity data based upon the interfraction motion of organs throughout the course of treatment. One aspect of this is work to assess the dose delivered to the rectum based upon volumetric images taken at each treatment. The
Conclusions and further work

delivered dose from each treatment fraction is then combined to produce a cumulative dose surface map for the rectum.

The study has shown that the cumulative dose surface map which includes the effects of inter-fraction motion is a better predictor of toxicity than simply using the original volume from the planning scan (Shelley et al., 2017). This study does not include the variation in dose delivered due to changes in beam output. To enhance the study further I believe it would be desirable to also include the known dose deviations into the analysis. The inclusion of an output variation would be relatively simple as it would simply require a scaling factor to be applied to each fraction delivered dose. This should then allow a more complete understanding of the impact of these small changes in dose on clinical outcomes.

This example was from one particular study, but there are many studies which are looking at better quantifying the impact of setup variations on clinical outcomes (Krauze et al., 2016; Moulton et al., 2017). The variation in dose is often much more simply quantified, and so should be included wherever possible to provide the most powerful studies possible, particularly if the magnitude of changes in dose is similar to that possible due to beam output variation.

8.2.3 Assessment of impact of tolerances used

This work assessed the measurements based upon the recorded data without knowledge of the precise tolerances in place. It is known that many centres currently use a ±2% tolerance as a prompt for recalibration (Palmer, Kearton and Hayman, 2012). However, through discussion with radiotherapy physicists from different centres it is clear that there may be large differences in how these tolerances are applied. Tolerances for daily constancy checks vary dependent upon equipment used (this work showed there was no significant difference between commonly used apparatus).

From the work in this thesis it is not known if there is a significant difference in the variation in beam output based upon pro-active and re-active beam calibrations. Working practices will dictate whether a centre routinely recalibrates a machine in response to out of tolerance values, or aims to pro-actively recalibrate prior to the results becoming out of tolerance. It may be that pro-active centres reduce the extreme tails on the probability of high and low doses being delivered.

Alongside this with the increasing implementation of electronic data recording and storage the volume of data which is available to analyse is increasing. Trending and visualisation are one major advantage and may allow better management of results and their decisions which arise from them.
Conclusions and further work

A simple addition of determining the mean beam output from a week of results may enable clearer trending and prediction of the next required recalibration, but I have not encountered a single centre which had something like this implemented routinely. Electronic data recording would allow trending to then alert the user when a calibration might be next due to enable appropriate staff to be rostered ensuring in-tolerance results within minimal impact on staff. This is something I hope to be implemented within my local centre, and then the method to perform this within our electronic QC system can be shared to allow its simple implementation for others.

8.2.4 Financial assessment of the change in beam output tolerances

Probably the key conclusion arising from this work is that the uncertainty in beam output following initial calibration is the greatest uncertainty within the clinical calibration chain. Reduction in this uncertainty would have a number of benefits, but would likely require greater resources. Due to the individual response of different machines an impact analysis would be specific to the needs of an individual centre dependant on current working practice. There are a number of strategies which could be utilised to reduce the variation in clinical outcomes based upon beam output maintenance. Undetected changes in beam output are often regarded as one of the highest risks and so are fundamental to QC protocols (Smith et al., 2017).

Some centres may treat particular clinical sites on a set of machines. In this case the tolerances on beam output may then become machine specific. For example, if head and neck and prostate patients are treated on machines ‘A’ and ‘B’ respectively, then to achieve the same level of uncertainty in outcome for these patients different tolerances would be applicable. Machine ‘A’ treating the head and neck patients would require a tighter tolerance than machine ‘B’ and so would likely require a greater calibration frequency, thus greater resources. If all clinical sites are treated on all machines, then a tighter tolerance would be required on all machines, resulting in the greatest staff burden.

Dependent upon the protocols in place and the staff mix, the time and staff required to perform a recalibration may vary considerably. Should machines have tolerance set based upon the steepest dose response curves which may be steeper than those assessed in this thesis (non-small cell lung cancer has an $\alpha/\beta=20$ (Fowler, 2010) for example), or should a ‘typical’ value be used? These decisions would have to be made locally dependent upon the case mix and the desires of the clinicians. This would also enable a detailed financial impact assessment to be completed.
8.2.5 Produce automated user tools for analysis of beam output and DHV parameter extraction

The analysis performed on the routine beam output data was extremely time-consuming to setup owing to differences in data recording methods and formats. Now that the fundamental analysis techniques and data format have been defined it is hoped that similar analysis could be completed in the future based upon a common data format. It would be simple to distribute a spreadsheet template to allow each centre to enter their data. These could all then be imported into a common data file and the analysis completed relatively easily, either from a single machine or multiple machines and centres.

To allow others to complete similar analysis with minimal time resources the code has all been compiled within the python language, thus allowing free, simple to install software to be used to run this. Currently these are command line based; however the development of a graphical user interface would likely increase its appeal and remove the learning curve associated with learning a new coding language. With the final data structure defined, it would also be relatively simple to develop import routines which accept a variety of formats from common electronic QC systems.

The extraction of DVH parameters, while automated, currently relies upon a specific file and folder structure to produce the required output. The initial aim is to allow all files to be ‘dumped’ into a single directory, or unrelated sub-directories whilst still allowing the data extraction for each unique treatment plan, allowing the user greater flexibility. A simple file parser to compile and organise a list of patient plans and related files within the specified directories would be required to do this. As with the beam output analysis, a simple GUI to allow the user to simply select the required files and folders to include would be desirable. Addition of statistics such as homogeneity or conformity index may also be desirable and relatively simple to include.

These tools created should be made open source to allow independent checks of the analysis and processes to be completed. This would also allow customisation and enhancements to be made by individual users to suit their needs. The GitHub repositories which hold the code and data used within this thesis will be used as the initial starting point for further development and generalisation. These are listed in Appendix B and Appendix C.

8.2.6 Impact of further QC tests

This thesis addresses deviations in delivered dose due only to variations in the absolute beam output. This is determined under a set of standard conditions and is valid at a particular depth along
Conclusions and further work

the central beam axis. If there were no other deviations in the radiation beam then this would be all that was required to build a complete picture of the delivered dose. However uncertainties due to the beam energy, symmetry and variations in field size are likely to all be present to varying degrees.

With OARs often away from the central axis the variation in beam flatness and symmetry may become important as an additional 1-2% change in dose may be present due to variations in these parameters. This thesis essentially examines a ‘1D’ situation assuming overall dose scales linearly with beam output only. This could be expanded to address the 2D and 3D clinical situations through modelling of changes in the beam profiles. It should be noted that this is not a simple extrapolation of the current analysis and would be a complex piece of work. As mentioned previously, much of the work assessing the movement of organs assumes the beams are perfectly calibrated. Development of this to include instabilities in the beams would be a large step towards being able to quantify the true cumulated dose to a patient through the course of treatment.

The initial stages would be to quantify the types of variations observed within other routinely assessed parameters such as flatness, symmetry and radiation field size alignment. Historically these are all separate measurements often rarely taken, and so each would require a separate analysis to be performed to quantify the variations and trends observed.

8.2.7 Multivariate analysis
As mentioned, this work only addresses beam output. Recent developments in technology mean that a greater number of measurements can be taken on a routine basis. For example, Varian’s ‘Machine Performance Check’ system uses the imager panel to measure a variety of mechanical and dosimetric parameters. Other devices such as the PTW QuickCheck, or Sun Nuclear DoseChecker measure a number of beam parameters simultaneously.

These simultaneous measurements taken routinely may allow multivariate analysis to predict with a greater accuracy when machine faults are going to be encountered. Electronic data storage and advances in technology now allow this type of work to be completed much easier than previously, and so this appears to be a large area of research which may involve specialised ‘Big Data’ techniques to manage. I envisage that in the future all the different data sources could be combined for example taking data from machine log files, coupled with a wide variety of QC metrics to determine when a service is due, or a more accurate QC check performed. This type of work could be easily large enough for a PhD thesis, and would be of potential interest to the manufacturers who may wish to support this.


Aukett, R. J. *et al.* (2005) ‘Addendum to the IPEMB code of practice for the determination of absorbed dose for x-rays below 300 kV generating potential (0.035 mm Al-4 mm Cu HVL).’, *Physics in medicine and biology*, 50, pp. 2739–2748. doi: 10.1088/0031-9155/50/12/001.


Bibliography

10.1016/j.clon.2011.11.009.


Kleivenhagen, S. C. et al. (1996) ‘The IPEMB code of practice for the determination of absorbed dose for x-rays below 300 kV generating potential (0.035 mm Al-4 mm Cu HVL; 10-300 kV generating potential).’, Physics in medicine and biology, 41, pp. 2605–2625.


10.1371/journal.pone.0169202.


Wedenberg, M. (2013b) FROM CELL SURVIVAL TO DOSE RESPONSE – MODELING BIOLOGICAL EFFECTS IN RADIATION THERAPY. Karolinska Institutet.


Appendices

Appendix A  Publications and presentations arising from this work

The abstracts of conference proceedings and published work produced during the period of this PhD are presented here for reference.

Appendix A.1  Poster at UK Radiation Oncology Conference 2016: Impact of Linac Output Variation on Clinical Outcomes:

M. Bolt 1,3, T. Chen 3, C. Clark 1,4, R. Jena 5, A. Nisbet 1,2

1 Dept. of Medical Physics, Royal Surrey County Hospital, Guildford, UK
2 Dept. of Physics, University of Surrey, Guildford, UK
3 Dept. of Chemical and Process Engineering, University of Surrey, Guildford, UK
4 National Physical Laboratory, Teddington, UK
5 Dept. of Oncology, Addenbrooke’s Hospital, Cambridge, UK

Available from: http://dx.doi.org/10.13140/RG.2.2.15230.59209

Abstract

Aims: When delivering radiotherapy treatments, linac output is considered constant, with a ±2% tolerance during daily checks. This work aims to quantify the impact on clinical outcomes due to variation in daily linac output.

Method: Daily output data from 133 linacs in 23 UK radiotherapy centres was collected from January to June 2015, creating a national picture of dose variation. The potential impact on clinical outcomes has been assessed from the steepness of dose response curves for a variety of cancers and variation in outputs for individual linacs and within centres.

Results: There was a difference of 3.6% between the maximum and minimum mean linac outputs, with a range of 2% between the 5th and 95th percentiles. Within a single centre a maximum difference of 2.8% was found between linacs. Only a single linac had a mean output which exceeded ±2%, with 13 linacs (9.8%) outside ±1%.
Head and Neck cancers have steep dose response with 1% dose change giving 2% TCP change and 5% in side effects. Based only on the measured output variations it is predicted that TCP would vary by 7.2% across all studied linacs, and 5.6% in an individual centre, with side effects varying by 18% and 14% respectively.

Conclusions: Variations in clinical outcomes due to linac output are not insignificant which has potential implications for QA tolerances as well as QA in clinical trials. This work is part of the larger QUASAR project aiming to quantify the impact of QA and audit on clinical outcomes.

Appendix A.2  Poster at ESTRO 35 (2016): Impact of standardised codes of practice and related audit on radiotherapy dosimetry over 20 years.

R. Thomas 1, M. Bolt 2, G. Bass 1, A. Nisbet 2, C. Clark 1

2 National Physical Laboratory, Radiation Dosimetry, Teddington, United Kingdom
1 Royal Surrey County Hospital, Medical Physics Department, Guildford, United Kingdom

Radiother. Oncol. 2016 Apr 7:119:S918,
Available from: http://dx.doi.org/10.1016/S0167-8140(16)33186-3

Abstract

Purpose/Objective: Reference dosimetry audit measurements in UK radiotherapy centres have been carried out over the last 20 years. This work examines the variation in local dosimetry calibration in a network of radiotherapy centres, draws conclusions regarding the implementation of an absorbed dose based protocol for MV photon beams and includes the measured effect of a change in the nationally recommended electron code of practice (CoP) from an air kerma based to an absorbed dose based protocol.

Material/Methods: Data from reference dosimetry audits conducted in radiotherapy centres by the National Measurement Institute (NMI) for photon, electron and kV x-rays have been collated, recording the NMI:Centre ratio for reference output measurements, beam quality, and field chamber comparison. A total of 81 photon, 98 electron and 30 kV beams were measured during 68 visits between June 1994 and February 2015. The change in the national standard deviation has been assessed over time, and differences due to the change between the two electron CoPs during this
period has been quantified. The improvement in consistency for MV beams since the adoption of a CoP traceable to the world’s first MV absorbed dose calibration service is assessed.

**Results:** The mean NMI:Centre difference for radiation output calibration is less than 0.25% for all modalities. A total of 7 measurements were reported to be outside the ±2% tolerance.

There was a statistically significant difference (p=0.008) in the mean result for the respective air kermabased electron CoP,+0.75% (n=14)with the absorbed dose based protocol giving +0.20%(n=84).

The variation in MV results has decreased steadily over time (see Figure 1). The standard deviation has halved when comparing the first and last 20 results, being 0.85% (2000) and 0.35% (2015). This trend has also been noted within regional audit groups. A linear correlation was observed between the ‘NMI:Centre output ratio’ and the ‘NMI:Centre field chamber comparison ratio’. There has been no significant difference observed between regional audit and national audit for the measured NMI:Centre ratios, but some regions have had many more NMI audits than others, some having no beams audited for a particular modality, and others having more than 20.

**Conclusions:** Data has been collated from 20 years of NMI reference dosimetry audits, and key trends and changes have been noted. The introduction of the 2003 absorbed dose-based electron CoP has decreased the difference between NMI and centre measured outputs. Use of a single MV CoP over the period of the study has contributed to a consistent reduction in variation of results. This not only shows the impact of a rigorous traceability chain developed by close collaboration between NMI and end users but also demonstrates that the NMI audit programme is likely to be a contributing factor to this improvement in consistency in dosimetry nationally.
Appendix A.3  Poster at ESTRO 36 (2017): Variation in mean dose from 204 UK linacs (Jan-June 2015) and its potential clinical impact.

M. Bolt , A. Nisbet , C. Clark , T. Chen , R. Jena

1 St Lukes’s Cancer Centre, Royal Surrey County Hospital, Radiotherapy Physics, Guildford, United Kingdom
2 National Physical Laboratory, Radiation Dosimetry, Teddington, United Kingdom
3 University of Surrey, Chemical and Process Engineering, Guildford, United Kingdom
4 Addenbrooke’s Hospital, Department of Oncology, Cambridge, United Kingdom

Radiother. Oncol. 2017, 123:S399
Available from: http://dx.doi.org/10.1016/S0167-8140(17)31194-5

Abstract

Purpose: Variation in dose delivered to patients directly impacts the effectiveness of radiotherapy treatments. The drift and daily fluctuations in linac beam calibration (output) is a contributing factor to the cumulative dose received by the patient. Knowledge of the variation in measured outputs on a national scale provides an insight into the uncertainties in dose delivery and its clinical impact.
Methods: A request for 6MV output measurement data was sent to all UK radiotherapy centres. In total, data was provided for 204 linacs situated at 52 cancer centres across the UK. The data spans 6 months from January to June 2015, totalling almost 25,000 data points. Additional data collected includes: linac model, year of install, measurement equipment and recording method. The dose response parameter, gamma, is the percentage change in treatment response caused by a percentage change in dose. Gamma values of 2.3 and 5.2 (representative for Head and Neck cancers) were used to estimate the effect on TCP and NTCP respectively [1].

Results: Based on the collated data, the UK linac outputs had a mean of -0.01% with one standard deviation of 0.88%. There was a wide variety of recording methods, with 8 centres having no form of electronic record for daily checks. Measurement data for both constancy devices and ionisation chambers was provided for 29 linacs. Of these, 8 (28%) had a discrepancy between measurement devices of greater than 0.5%, with 3 linacs (10%) having greater than 1%. The greatest variation in the mean output of an individual linac was -2.1%, with 90% of linacs having a mean output within 1% of the national mean. No significant variations were observed based on the age of the linac. The maximum range within a single centre for the mean output for each linac was 2.3% (min: -1.1%, max: +1.2%)—see figure. Assuming patients are treated on a single linac for their treatment duration, this indicates a variation in TCP of 5.3% and a variation in NTCP of 12% dependent on which linac they are treated on at that centre.

Conclusions: The data collection process indicates that many departments still rely heavily on paper QC records. The variation in treatment outcomes caused by dose variation alone indicates the importance of accurate QC. Output adjustment is one of the simplest ways of maintaining treatment consistency between individual patients, and its significance should not be forgotten with the introduction of more advanced techniques. This variation in dose should be considered when participating in clinical trials. This applies both to small scale local trials in which the technique used may determine the treatment linac, and therefore the dose delivered, as well as large multi-centre trials where the dose variation should be considered for the trial power calculations.

Appendix A.4  Journal Article: Radiotherapy reference dose audit in the United Kingdom by the national Physical Laboratory: 20 years of consistency and improvements

Russel A Thomas a, Matthew A. Bolt a,b,c, Graham Bass a, Rebecca Nutbrown a, Tao Chen c, Andrew Nisbet b,d, Catharine H. Clark a,b

a National Physical Laboratory, Teddington, UK
b Department of Medical Physics, Royal Surrey County Hospital, Guildford, Surrey, UK
c Department of Chemical and Process Engineering, University of Surrey, Guildford, Surrey, UK
d Department of Physics, University of Surrey, Guildford, Surrey, UK

Available from: http://dx.doi.org/10.1016/j.phro.2017.07.005 2405-6316

Abstract

Background and purpose: Audit is imperative in delivering consistent and safe radiotherapy and the UK has a strong history of radiotherapy audit. The National Physical Laboratory (NPL) has undertaken audit measurements since 1994 and this work examines results from these audits.

Materials and methods: This paper reviews audit results from 209 separate beams from 82 on-site visits to National Health Service (NHS) radiotherapy departments conducted between June 1994 and February 2015. Measurements were undertaken following the relevant UK code of practice. The accuracy of the implementation of absorbed dose calibration across the UK is quantified for MV photon, MeV electron and kV X-ray radiotherapy beams.

Results: Over the measurement period the standard deviation of MV photon beam output has reduced from 0.8% to 0.4%. The switch from air kerma- to absorbed dose-based electron code of practice contributed to a reduction in the difference of electron beam output of 0.6% (p < 0.01). The mean difference in NPL to local measurement for radiation output calibration was less than 0.25% for all beam modalities.

Conclusions: The introduction of the 2003 electron code of practice based on absorbed dose to water decreased the difference between absolute dose measurements by the centre and NPL. The use of a single photon code of practice over the period of measurements has contributed to a reduction in measurement variation. Within the clinical setting, on-site audit visits have been shown to identify areas of improvement for determining and implementing absolute dose calibrations.
Appendices

Appendix A.5 Journal Article: A multi-centre analysis of radiation beam output measurements

Bolt, Matthew A. a,b,c, Clark, Catharine H. a,b, Chen, Tao c, Nisbet, Andrew a,d

a Department of Medical Physics, Royal Surrey County Hospital NHS Foundation Trust, Guildford, Surrey, UK
b National Physical Laboratory, Teddington, UK
c Department of Chemical and Process Engineering, University of Surrey, Guildford, Surrey, UK
d Department of Physics, University of Surrey, Guildford, Surrey, UK

Available from: https://doi.org/10.1016/j.phro.2017.12.001

Abstract

Background and Purpose: Radiotherapy requires tight control of the delivered dose. This should include the variation in beam output as this may directly affect treatment outcomes. This work provides results from a multi-centre analysis of routine beam output measurements.

Materials and Methods: A request for 6MV beam output data was submitted to all radiotherapy centres in the UK, covering the period January 2015 – July 2015. An analysis of the received data was performed, grouping the data by manufacturer, machine age, and recording method to quantify any observed differences. Trends in beam output drift over time were assessed as well as inter-centre variability. Annual trends were calculated by linear extrapolation of the fitted data. Results Data was received from 204 treatment machines across 52 centres.

Results: were normally distributed with mean of 0.0% (percentage deviation from initial calibration) and a 0.8% standard deviation, with 98.1% of results within ±2%. There were eight centres relying solely on paper records. Annual trends varied greatly between machines with a mean drift of +0.9%/year with 95th percentiles of +5.1%/year and -2.2%/year. For the machines of known age 25% were over ten years old, however there was no significant differences observed with machine age.

Conclusions: Machine beam output measurements were largely within ±2% of 1.00cGy/MU. Clear trends in measured output over time were seen, with some machines having large drifts which would result in additional burden to maintain within acceptable tolerances. This work may act as a baseline for future comparison of beam output measurements.
Appendix A.6  Oral Presentation at ESTRO 37 (2018): Modelling the clinical impact of machine specific dose variations on outcome using national data

M. Bolt 1, C. Clark 1, T. Chen 2, A. Nisbet 3

1 Royal Surrey County Hospital, Medical Physics, Guildford, United Kingdom
2 University of Surrey, Chemical & Process Engineering, Guildford United Kingdom
3 University of Surrey, Physics, Guildford, United Kingdom

Available from: https://doi.org/10.1016/S0167-8140(18)30921-6

Abstract

Purpose or Objective
The accuracy of dose delivery during radiotherapy treatments depends upon a number of patient and equipment factors, including the current treatment ESTRO 37 machine calibration and output. Modern treatment techniques and improved image guidance reduce the variation in delivered dose due to planning and setup. A result of this is that control of treatment machine calibration is of greater significance in the overall delivery uncertainty. The impact of calibration errors and uncertainties in daily dose delivery on clinical outcomes in radiotherapy has been modelled using evidence based radiobiological models.

Material and Methods
The dose variation introduced by 3D conformal planning and IMRT planning has been assessed through extraction of dose statistics from the PARSPORT head and neck trial [1]. A dataset containing >24,000 measurements of beam output from 204 UK radiotherapy treatment machines [2] was combined with a national dataset containing dosimetry audit measurements [3] to assess the potential dose variability for patients treated on individual machines, within a single centre or nationally. Radiobiological modelling was completed for prostate and head and neck cancers incorporating the daily dose variability.

Results
IMRT planning techniques within PARSPORT not only reduced dose delivered to the OARs, but also the variation in dose to the PTV across the trial population. The D99 standard deviation reduced by around 2/3 to 3% when IMRT techniques were used whilst also increasing PTV coverage (fig 1). This is of the same magnitude as dose variations due to beam calibration. Radiobiological modelling
Appendices

predicts that variation in beam calibration and output introduces a 10% variation in predicted TCP (2 year survival) and 6% in NTCP (xerostomia) for patients in the PARSPORT trial. For prostate cancer this results in variations of TCP (bPFS) and NTCP (grade 1-2 proctitis) of 3% and 4% respectively. This variation is seen to be possible within a single treatment centre.

![PTV1 - D99](image)

Figure 1: Box plots of D99 statistics for conformal and IMRT plans included within the PARSPORT trial. IMRT plans show less variation in planned target doses.

Conclusion

Control of fundamental dosimetry within the radiotherapy chain remains highly significant even with modern treatment technologies. Variation in the control of beam calibration provides a potential source of uncertainty which could lead to measurably different outcomes for patients treated on different machines. Whilst fundamental dosimetry measurement has improved over time, the advances in treatment technology now mean that the impact of fundamental dosimetry is more significant than ever. As technology improves, the tolerances should also be reduced appropriately to best make use of these advances. This variation may have a direct effect both on small local trials as well as large scale clinical trials and should be included in their considerations of uncertainties.

References

Appendices

Appendix B  Details of availability of python code used within this thesis

The python code used is available online within the GitHub repositories as follows. Brief instructions and example usage are given within the README files present at the above repositories.

Appendix B.1  Dosimetric parameter extraction from DICOM RT files

Python script available at: https://github.com/mbolt01/DICOM_dose_parameter_extraction

The script was used to generate the dosimetric statistics used within chapter 5.

Appendix B.2  Implemented radiobiological models used to within chapters 6 and 7:

Python script available at: https://github.com/mbolt01/TCP-NTCP-dose_var_model

The script was used within chapters 6 and 7.

Appendix C  Availability of anonymised data collated for use within this thesis

Online data repository: https://github.com/mbolt01/Thesis_data

The online data repository contains CSV files which hold the data used within this thesis as follows:

<table>
<thead>
<tr>
<th>Filename</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPL_audit_data.csv</td>
<td>NPL audit measurement results collated from the NPL audit reports. The data is used within chapter 3.</td>
</tr>
<tr>
<td>Routine_outputs.csv</td>
<td>The collated data obtained after a request sent to all UK radiotherapy centres via the regional audit groups. The data is for 6MV beam outputs from January to June 2015 and is used within chapter 4</td>
</tr>
<tr>
<td>RSCH_outputs.csv</td>
<td>Collated beam output data for linacs located at the RSCH (main and satellite site). The data begins from 2014 and was collated form electronic records held at the RSCH. This data was used for section 4.4.</td>
</tr>
<tr>
<td>PARSPORT_DVH_stats.csv</td>
<td>Dosimetric statistics extracted from the PARSPORT trial DICOM RT plan sets using the automated data extraction tool developed. Data is analysed in chapter 5. (Note: the original DICOM files are not publically available).</td>
</tr>
</tbody>
</table>
Appendices

Appendix D  Description of boxplots as used within this thesis

Unless otherwise specified within the associated text the boxplots within this thesis are constructed as depicted in Figure A.D.1 unless specified within the specific figure caption. The central bar represents the median, the box indicates the extent of the 25th and 75th percentiles of the data and the whiskers indicate the 5th to 95th percentile range.

Figure A.D.1: Schematic of the boxplot representation used within this thesis unless otherwise specified within the text. The box is bounded by the 25th and 75th percentiles (forming the interquartile range (IQR)). The whiskers represent the 5th to 95th percentile range. This particular example shows the boxplot for the results of 1000 samples drawn randomly from a normal distribution with mean value of zero and standard deviation of one.