A Local Assessment of the PTV Margin Calculation in Patients Treated to the Larynx with VMAT

Kristina Caruana

A Dissertation Presented to the Faculty of Health Sciences in Part Fulfilment of the Requirements for a Master's Degree in Research at the University of Malta

October 2022

Acknowledgments

I dedicate this research to all patients who have been diagnosed with cancer.

The completion of this study would not have been possible without the help of a substantial number of individuals.

First and foremost, I would like to express my appreciation to my supervisors Dr. Paul Bezzina and Prof. Francis Zarb, for their continuous dedicated guidance and support throughout this study. Their feedback throughout the research process was greatly appreciated and was essential to the study. Also, a special thanks to Mr. Guilherme Couto, for his encouragement, assistance and guidance throughout the research process, he was always ready to offer a helping hand. His expert knowledge in the field of radiotherapy was a great asset during the process of this study.

Special thanks go to Dr Nick Refalo and Ms Jessica Farrugia for offering their time and assistance. Their expert knowledge was essential to the study. I am also grateful to Professor Liberato Camilleri for his precious time and help with the data analysis of this study.

My gratitude goes out to Ms. Dorothy Aquilina for sharing her insights and expertise on the subject and enabling the completion of this project. Al-Qaisieh Bashar and Filipe Moura also contributed their expert knowledge on the methodology, for which I am grateful.

Additionally, I would like to express my gratitude to Drs. Malcolm Buhagiar, Etienne Paris, Kelly Taliana, Louisa Farrugia, Joanna Grech, and Nick Refalo for their assistance. Also, thanks to my colleagues, who have been there for me by offering their time, support and encouragement, with a special mention to Denise Spiteri, Maria Elena Gollcher, Stephanie Rapa, Claire Farrugia, Kyle Muscat, Mark Agius, Mirabel Agius, Sarah Sciberras and Christabel Calleja.

Last but not l least, I am forever grateful to all the patients who trusted me with this study and gave their permission to participate.

Table of contents

Acknowledgments ii
Table of contents iii
List of tables vi
List of figures ix
Abbreviation and acronyms xi
Definition of concept xiii
Abstract xv

1. Introduction

1.1 Introduction to the study 2
1.2 Background to the study 3
1.3 Aims and objectives 4
1.3.1 Research question 5
1.4 Study relevance 6
1.5 Methodology 6
1.6 Ethical consideration 7
1.7 Dissertation outline 8

2. Literature review

2.1 Introduction to the chapter 10
2.2 Section A – Narrative Literature Review 11
2A.1 Introduction 11
2A.1.1 Definition of Planning Target Volume (PTV) margin 11
2A.1.2 Overview of the formulas used in calculating the PTV 12
2A.2 Target volume delineation of the larynx 16
2A.3 Methods of analysis for inter-fraction set-up errors 24
2A.4 Intra-fraction error 29
2A.5 Observer variation in image matching 31
2A.6 Other errors 33
2A.7 Conclusion 34
2.3 Section B – Systematic Literature Review 35

2B.2 Rationale for the systematic literature review 36
2B.3 Methods 37
2B.3.1 Protocol and registration 37
2B.3.2 Eligibility criteria 38
2BA.3.3 Information sources 39
2B.3.4 Keywords for literature search 39
2B.3.5 Selection process 41
2B.3.6 Data collection process 42
2B.3.7 Summary measures 44
2B.3.8 Synthesis of results 44
2B.4 Results and discussion 45
2B.4.1 Study selection 45
2B.4.2 Study characteristics 47
2B.4.3 Risk of bias across studies 49
2B.4.4 Results of synthesis 52
2B.5 Discussion 62
2B.5.1 Summary of evidence 62
2B.5.2 Limitations 69
2B.5.3 Strengths 70
2B.5.4. Conclusion of the systematic literature review 71

3. Methodology

3.1 Introduction 73
3.2 Research design 73
3.3 Population, sampling and recruitment procedure 75
3.3.1 Patients receiving VMAT to the laryngeal region 75
3.3.2 Radiographers performing image verification 78
3.3.3 Clinical oncologists and HSTs contouring CTV-PTV margins 79
3.4 Local oncology department procedures 81
3.5 Data gathering tools 84
3.5.1 Data gathering tool for target volume delineation variation 84
3.5.2 Data gathering tools for intra-fraction motion error 85

3.5.3 Data gathering tools for total inter-fraction error 85
3.5.4 Data gathering tool for inter-observer variation 86
In image matching
3.6 Validity and reliability of the data gathering tools 87
3.6.1 Validity 87
3.6.2 Reliability 90
3.7 Pilot study 93
3.8 Data collection 95
3.9 Analysis of data 100
3.10 Ethical considerations 108
3.11 Conclusion 110

4. Results and Discussion

4.1 Introduction 112
4.2 Participants 112
4.3 Reliability results 113
4.3.1 Reliability of contouring range distance measurements 113
4.3.2 Reliability of set-up errors recorded on a data collection sheet 113
4.3.3 Reliability of inter-fraction errors using soft-tissue registration 116
4.4 Target volume delineation error 120
4.4.1 Results and discussions 120
4.5 Total inter-fraction errors 130
4.5.1 Results and discussion 130
4.6 Intra-fraction errors 142
4.6.1 Results and discussion 142
4.7 Observer variation in image matching errors 152
4.7.1 Results and discussion 152
4.8 PTV margin calculation 158
4.8.1 Results and discussion 158
4.9 Limitations of the study 167
4.10 Strengths of the study 170
4.11 Conclusion 172

5. Conclusions and Recommendations

5.1 Introduction 174

5.2 Conclusion 174

5.3 Recommendations 175

5.3.1 Clinical recommendations 175

5.3.2 Recommendations for further studies 176

5.4 Overall conclusion 178

List of Tables

Table 1: Research Question using PICO Framework 5

Table 2B.1: PICOS Elements for Eligibility Criteria 37

Table 2B.2: Inclusion/Exclusion Criteria 38

Table 2B.3: Summary of Characteristics of Studies 47

Table 2B.4: Outcome of the Evaluation of the Studies 52

Table 2B.5: Confounding Variables and Strategies 53

Table 2B.6: PICOS Elements of the Studies 55

Table 2B.7: PTV Margin Methods 57

Table 3.1: Inclusion and Exclusion Criteria for Participating Patients 76

Table 3.2: Sample Calculation 78

Table 3.3: Inclusion and Exclusion Criteria for participating Radiographers 79

Table 3.4: Legend for Error Directions 102

Table 4.1: Patients' Demographic 112

Table 4.2: Reasons for Re-scans 114

Table 4.3: Description of ICC Coefficients 115

Table 4.4: Description of the Cronbach's Alpha 116

Table 4.5: Cronbach's Alpha for the Superior-Inferior (Y) Direction 117

Table 4.6: ICC for the Superior-Inferior (Y) Direction 117

Table 4.7: Cronbach's Alpha for the Left-Right (X) Direction 117

Table 4.8: ICC for the Left-Right (X) Direction 117

Table 4.9: Cronbach's Alpha for the Anterior-Posterior (Z) Direction 117

Table 4.10: ICC for the Anterior-Posterior (Z) Direction 118

Table 4.11: Cronbach's Alpha for the Left-Right (X) Direction 118

Table 4.12: ICC for the Left-Right (X) Direction 118

Table 4.13: Cronbach's Alpha for the Superior-Inferior (Y) Direction 119

Table 4.14: ICC for the Superior-Inferior (Y) Direction 119

Table 4.15: Cronbach's Alpha for the Anterior-Posterior (Z) Direction 119

Table 4.16: ICC for the Anterior-Posterior (Z) Direction 119

Table 4.17: Colour Coding System 120

Table 4.18: Systematic Individual Mean Values of the Contouring Range Measurements in each Translational Axis 123

Table 4.19: Comparison of CTV for each Observer (cm^3) 123

Table 4.20: Summary of Population Translational Total Inter-fraction Errors 136

Table 4.21: Summary of Population Rotational Total Inter-fraction Errors 130

Table 4.22: Summary of Intra-fraction Errors 150

Table 4.23: Population Systematic and Random Errors of Inter-Observer Variation Errors 156

Table 4.24: Translational Population Systematic Errors 158

Table 4.25: Translational Population Random Errors 159

Table 4.26: PTV Margin Result 161

Table 4.27: PTV Margin Result without Target Volume Delineation Error 161

Table 4.28: Comparison of Studies 162

Table 4.29: Total Population Systematic Errors for Rotation 166

Table 4.30: Total Population Random Errors for Rotation 166

List of Figures

Figure 2A.1: Illustration of Determining the Magnitude of Delineation Error 23

Figure 2B.1: Example of the Search Strategy 41

Figure 2B.2: Prisma 2009 Flow Chart 45

Figure 2B.3: PTV Margin Result 58

Figure 2B.4: Population Systematic Errors 59

Figure 2B.5: Population Random Errors 60

Figure 3.1: An Example of a CBCT Image of a Patient Treated to the Head and Neck Region 82

Figure 3.2: An Example of a Clipbox Position for Radiotherapy Treatment to the Larynx 97

Figure 3.3. Example of Mask Registration for Radiotherapy Treatment to the Larynx 98

Figure 3.4: Measure Tool – Monaco TPS 101

Figure 3.5: A Demonstration of the Superior and Inferior Measurements of Distances taken in the Coronal Plane 101

Figure 3.6: A Demonstration of the Anterior and Posterior Measurements of Distances taken in the Axial Plane 101

Figure 4.1. Frequency histograms showing the distribution of contouring range measurements (cm) that were taken on each alternating CT slice for all patients 121

Figure 4.2. A scatter graph showing the variation of the contoured clinical target volumes (cm^3) obtained from the observers 124

Figure 4.3: A Demonstration of the Outlier (observer 6) for Patient Number 2 125

Figure 4.4: A Demonstration of the Outlier for Patient Number 1 126

Figure 4.5: Scatter Graphs of Individual Mean Errors (mm) 132

Figure 4.6: Scatter Graphs of Individual Random Errors (mm) 133

Figure 4.7: Scatter graphs of the individual mean of total inter-fractional errors in the rotational direction 134

Figure 4.8: Frequency Histograms Showing the Distribution of Total Inter-fraction Errors (cm) for all fractions 136

Figure 4.9: Frequency Histograms Showing the Distribution of Total Inter-fraction Errors in the Rotational Direction (^{o}) for all fractions 139

Figure 4.10: Scatter Graphs of the Individual Mean Error of Intra-fraction Errors 143

Figure 4.11: Scatter Graphs of the Individual Random Errors of Intra-fraction Errors 144

Figure 4.12: Frequency Histograms Showing the Distribution of Translational Intra-fraction Errors (cm) 146

Figure 4.13: Frequency Histograms Showing the Distribution of Rotational Intra-fraction Errors (^{o}) for all fractions 148

Figure 4.14. Residual errors in the Left-right Direction 152

Figure 4.15: Residual Errors in the Superior-inferior Direction. 153

Figure 4.16: Residual Errors in the Anterior-posterior Direction 153

Figure 4.17: Residual Errors in the Roll Direction 154

Figure 4.18: Residual Errors in the Pitch Direction 154

Figure 4.19: Residual Errors in the Yaw Direction 155

Figure 4.20: A Bar Graph Representing the Population Systematic Errors of all the Analysed Errors 161

Figure 4.21: A Bar Graph Representing the Population Random Errors of all the Analysed Errors 160

Abbreviation and Acronyms

AP	Anterior - Posterior
CASP	Critical Appraisal Skill Programme
CBCT	Cone Beam Computed Tomography
CEO	Chief Executive Officer
CT	Computed Tomography
CTV	Clinical Target Volume
DAHANCA	Danish Head and Neck Cancer Group
GTV	Gross Target Volume
EPID	Electronic Portal Imaging Device
FREC	Faculty Research Ethics Committee
ICC	Intra-class Correlation Coefficient
ICRU	International Commission on Radiation Units and Measurements
IGRT	Image Guided Radiotherapy
IM	Image Matching
IMRT	Intensity-Modulated Radiotherapy
JBI	Joanna Briggs Institute
KV	Kilovoltage
MeSH	Medical Subject Headings
MRI	Magnetic Resonance Imaging
MV	Megavoltage
NAL	Non-action Level
PCLE	Prostate Cancer Location Error
PICOS	Population, Intervention, Comparison, Outcome, and Study Design
PET	Positron Emission Therapy
PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analysis
PTV	Planning Target Volume
RTOG	Radiation Therapy Oncology Group
RX	Roll
RY	Pitch
RZ	Yaw
SBRT	Stereotactic Ablative Radiotherapy Techniques
SCC	Squamous Cell Carcinoma (SCC)
SD	Standard Deviation
SI	Superior-Inferior
SPSS	Statistical Package for the Social Sciences
SVCI	Single Vocal Cord Irradiation
TP	Thermoplastic
TPS	Treatment Planning System
UREC	University of Malta Research Ethics Committee
VHMF	Van Herk Margin Formula
VMAT	Volumetric-Modulated Arc Therapy
X	Left-right
XVI	X-ray Volumetric Imaging
Y	Superior-inferior
Z	Anterior-Posterior
2D	Two-dimensional

3D	Three-dimensional
4D	Four-dimensional
Σ	Populations systematic error
σ	Population random error

Definition of Concepts

Bone registration: A geometric comparison of the reference image with the image acquired during treatment, using the bone algorithm to superimpose the images.

Clip box: A box that surrounds the region of interest to be superimposed during image registration.

Inter-fraction error: An error observed on different treatment fractions. It has a systematic and random component.

Intra-fraction error: An error observed during treatment delivery, and mostly results from organ motion or patient movement. It also has a systematic and random component.

Mask registration: A geometric comparison of the reference image with the image acquired during treatment, using the mask algorithm for superimposition of the images.

Observer variation in image matching error: An error caused by a variation in the techniques and performance of the user performing image registration.

Offline imaging: Assessment of the patient set-up with respect to the treatment plan, after the delivery of the treatment.

Online imaging: Assessment of the patient set-up with respect to the treatment plan, immediately prior to the delivery of the treatment.

PTV: A margin around the clinical target volume to account for variations caused by patient position and set-up, physiological changes, human factors and variations in the treatment modality.

Random error: Deviations in uncertainties between different fractions

Residual error: The displacement from the initial planned position remaining after treatment corrections.

Set-up error: Geometric displacement resulting from patients' treatment position with respect to the desired reference treatment plan.

Systematic error: Deviations in uncertainties between the treatment plan and the average patient position over the entire course of the treatment.

Target volume delineation error: A systematic error introduced during the treatment preparation phase. It denotes the difference between the defined and 'ideal' CTV.

Thermoplastic mask: An immobilisation device made from plastic material commonly used in radiotherapy

Total inter-fraction errors: Inter-fraction errors measured as a combination of set-up errors and organ motion errors

Treatment planning system: The software and hardware used to simulate the irradiation geometry to be used for patient treatment and to calculate the dose distribution within the patient.

Abstract

Background: The introduction of Volumetric Arc Therapy (VMAT) in the local oncology hospital necessitated a more precise understanding of the true uncertainty and corrections for the Planning Target Volume (PTV) margin. The study's aim was to ensure that the PTV margins for patients treated to the larynx with VMAT were optimised for use in the local oncology hospital.

Methods: The PTV margin was calculated based on data from 20 patients who received VMAT to the laryngeal region in the local oncology department. Van Herk's formula was used for the calculation of the PTV by considering the population systematic and population random errors of the following errors: target volume delineation errors, inter-fraction errors, intra-fraction errors, inter-observer variation in image matching. All data were collected and analysed using both descriptive and inferential statistics using Microsoft Excel and Statistical Package for the Social Sciences (SPSS®).

Results: The margin size was most affected by target volume delineation errors and intra-fraction errors. Target volume delineation errors had the most significant influence with population systematic errors measurements of 3.47 mm in the left-right (X) and anterior-posterior (Z) directions, and 6.92 mm in the superior-inferior (Y) directions. If the target volume delineation errors were ignored, the calculated PTV margin would be 4.9, 9.4, and 7.6 mm in the X, Y, and Z directions, respectively. The PTV margin including all the measured errors in the X, Y, and Z directions was calculated to be 10.5 mm, 20.5 mm, and 12 mm, respectively.

Conclusion: The PTV margin size obtained in this study was larger when compared to similar studies, however it was the first study to consider target volume delineation errors and inter-observer variation in image matching. Addressing these sources of errors can reduce errors and achieve a smaller margin size.

Chapter 1

Introduction

1.1 Introduction to the study

In Malta, the incidence of head and neck cancer is 2.44 per 100,000 population with a 5-year survival rate of 20% (Borg Xuereb, Dimech and Muscat, 2015). This type of cancer may originate in the oropharynx, nasopharynx, oral cavity, larynx and hypopharynx (Lo Nigro et al., 2017). Laryngeal cancer represents one third of all head and neck cancers (Koroulakis and Agarwal, 2022) and is classified as supraglottic, glottic, or subglottic, with squamous cell carcinoma (SCC) being the most common histology type (Baron et al., 2015).

Treatment options for laryngeal cancer depend on the tumour staging and clinical characteristics (Denaro et al., 2014). When treating an oncology patient with radiotherapy to the head and neck region, accuracy is imperative (Malicki, 2012). Radiotherapy may be applied more conformably to the target due to advancements in techniques such as intensity-modulated radiotherapy (IMRT), volumetric-modulated arc therapy (VMAT), and stereotactic ablative radiotherapy techniques (SBRT), resulting in a better prognosis to the patient (Nyarambi et al., 2015).

To ensure precision in dose delivery, a margin is used in radiotherapy to account for the presence of geometric errors that might occur from treatment errors such as set-up, target delineation and organ motion errors (Suzuki et al., 2012). After adding margin to the Clinical Target Volume (CTV), the Planning Target Volume (PTV) is obtained.

The concept of PTV was first introduced in the International Commission on Radiation Units and Measurements (ICRU) Report 50 (1994). It is a recommended margin that should limit dose to adjacent critical structures because unnecessary radiation can cause severe radiotherapy side effects in patients. However, as a margin, it should be large enough to ensure clinically acceptable delivery of the prescribed dose to all parts of the CTV (Lo Nigro et al., 2017; Navran et al., 2019). The aim of the study was to assess the PTV margin of patients treated to the larynx with VMAT.

1.2. Background to the study

Malta has only one oncology hospital which is equipped with three linear accelerators, Superficial X-ray unit and Computed Tomography (CT) simulator which opened its doors in 2015. In September 2018, the first patient was treated using VMAT to the head and neck. This happened as for the first time, Malta now had the facility to have linear accelerators that enabled advanced radiotherapy technique to be applied into practice. Prior to the introduction of VMAT in the local centre, patients used to be treated by means of 3D conformal radiotherapy while using planar MV imaging.

The department's clinical protocol for head and neck patients treated with VMAT required that*, "PTV should be created by growing each Clinical Target by an isotropic margin of 5 mm"* (Sir Anthony Mamo Oncology Centre, 2020, p.4). This margin was derived from other departmental protocols which were used as a guide for the implementation of VMAT to the head and neck region in the local oncology centre (Velindre Cancer Centre, 2014; South Tees Hospital, 2015). The 5 mm margin size, however, was never tested as to whether it was applicable to the local settings. As VMAT was used more frequently, a more precise understanding of the true uncertainty and corrections for the PTV margin was required.

Studies recommend that PTV margins should be calculated to be specific for every institution (Lowther, Marsh and Louwe, 2020). Treatment modalities, patient set-up accuracy, patient collaboration, immobilisation devices, and the Image Guided Radiotherapy (IGRT) technique are some of the factors that influence margin size and are unique to different departments (Merlotti et al., 2014; Marnouche et al., 2019; Minniti et al., 2016; Winey and Bussiére, 2014). There is more than one way to calculate the PTV margin and this will be discussed later in the study. Also, the margin size might vary depending on the method and calculation used (Nyarambi et al., 2015; Kim et al., 2019; Marnouche et al., 2019). However, most of the authors failed to justify the selection of the method used in their study over other methods for calculation of the PTV margin (Yin et al., 2013; Norfadilah et al., 2017; Bruijnen et al., 2019). All these factors led to the formulation of the aims and objectives which will be presented in the next session.

1.3 Aim and objectives of the study

Aim:

The aim of the study was to calculate the CTV-PTV margins for patients treated to the larynx with VMAT at the local oncology hospital.

Objectives:

To achieve the aim of the study a number of objectives were outlined:

1. A systematic literature review was conducted to identify:
    I. The various methods used in research studies to calculate the PTV margin for patients receiving VMAT in the head and neck region.
    II. Whether the studies in the review considered all the relevant factors required to calculate the margin size with their chosen formula.
2. A narrative review of the literature was conducted to identify the various methods used to measure the relevant errors for PTV margin calculation.
3. Develop data record sheets to record the data of errors that are necessary for the measurement of the PTV margin.
4. Assess reliability of
    I. The contouring range distance measurements,
    II. Set-up errors recorded on a data collection sheet,
    III. Inter-fraction errors using soft-tissue registration.
5. Calculate the systematic and random errors of inter-fraction, intra-fraction, organ motion, target delineation and observer variation in image matching for patients who have received VMAT to the larynx.
6. Calculate the PTV margins based on the findings of the study.

1.3.1 Research Question

The Population, Intervention, Comparison and Outcome (PICO) framework was used as a guide to formulate the research question (Eldawlatly et al., 2018), as shown in table 1.

Table 1. Research Question using PICO Framework

P - Population	Patients treated to the larynx with VMAT
I - Intervention	Calculate the PTV margin
C - Comparison	N/A
O - Outcome	Optimisation of the PTV margin

Research question: What are the different methods of calculating the PTV margin, and how can this margin be measured at the local general hospital for patients treated to the larynx with VMAT?

1.4 Study relevance

The PTV margin should account for the measurement of errors that are specific to the area being measured (The Royal College of Radiologists, 2021). Based on an analysis of errors caused by inter-fraction and intra-fraction motion, target delineation variation, and observer variation in image matching, PTV margin can be calculated to the radiotherapy department of Sir Anthony Mamo Oncology Centre. This procedure will benefit patients being treated with VMAT because the measured PTV margin will ensure that dose to adjacent critical structures is limited, and no unnecessary radiation is given. The calculated PTV margin will also ensure that the prescribed dose is delivered to all parts of the CTV in a clinically acceptable manner. Recommendations based on the finding of this study will be proposed to the department's administration to consider any necessary amendments to the head and neck clinical protocol in relation to the PTV margin.

1.5 Methodology

This study had a quantitative, prospective and an insider research design. The target population for this study included patients treated with VMAT to the laryngeal region in the Maltese oncology hospital from June 2021 to May 2022.

An exhaustive sampling technique was used to select patients treated to the larynx with VMAT, where twenty patients were invited to participate. This was based on the recommendations of The Royal College of Radiologists (2021).

Variation in target volume delineation was analysed by asking doctors to delineate CTV of the population sample from the CT planning images. The systematic error was calculated from the resultant variation.

Daily Cone Beam Computed Tomography (CBCT) image registration results were recorded by the participating radiographers and used to calculate systematic and random inter-fraction errors.

Patients were also imaged once a week after treatment to assess intra-fraction errors. Image registration results were recorded by an intermediary radiographer, and values were used to measure systematic and random errors.

Inter-observer variation in image matching was assessed on the participating radiographers. Values were used to measure systematic and random errors.

The values of systematic and random errors obtained from target volume delineation errors, inter- and intra-fraction errors, and inter-observer variation in image matching were used to calculate the PTV margin for patients treated to the larynx with VMAT. The formula and procedures used to calculate the margin were based on a systematic literature review performed prior to the data collection.

All data were collected and analysed using both descriptive and inferential statistics using Microsoft Excel and Statistical Package for the Social Sciences (SPSS®).

1.6 Ethical considerations

Permission to conduct this study was obtained from the following entities:

1. Permission was sought and obtained from the Chief Executive Officer (CEO), the data protection officer, the quality assurance manager, participating patients treated with VMAT to the larynx, the clinical oncologists responsible for treating head and neck areas, the professional lead of radiography in the radiotherapy department, the radiographers acting as intermediaries and the radiographers who performed image matching at the local oncology hospital where the study was conducted.
2. The study was also approved by the Faculty Research Ethics Committee (FREC) and the University of Malta Research Ethics Committee (UREC).

1.7 Dissertation outline

This dissertation is composed of 5 chapters. This first chapter contains an introduction to the study. A systematic literature review related to the study and a narrative review is presented in Chapter 2. Chapter 3 outlines the methodology that was used to collect the data. The data were then presented, analysed and discussed in Chapter 4. The final chapter concludes this study while presenting a summary of the findings and proposing recommendations for the clinical department to consider as well as recommendations for further studies.

Chapter 2

Literature Review

2.1 Introduction to the chapter

This chapter is divided into two sections:

Section A is a narrative review of the literature that discusses the different methods used by various authors to analyse target delineation errors, inter- and intra-fraction errors, inter-observer variation in image matching, and other errors that influence the PTV margin size.

Section B is a systematic literature review and presents findings related to the methods for calculation of PTV margins in the head and neck region with VMAT. This systematic review was published in the Journal of Radiotherapy in Practice in May 2021 *(Caruana K, Refalo N, Spiteri D, Couto JG, Zarb F, and Bezzina P. (2021) PTV margin calculation for head and neck patients treated with VMAT: a systematic literature review. Journal of Radiotherapy in Practice page1 of 8. doi: 10.1017/S1460396921000546)* **(Appendix A)**.

2.2 Section A – Narrative Literature Review

2A.1 Introduction

The aim of this part of the chapter was to conduct a literature review on how to measure the PTV margin. The purpose was to identify the best methods for calculating target volume delineation errors, inter-fraction errors, intra-fraction errors, and observer variation in image matching in patients receiving VMAT treatment to the laryngeal region for the local oncology department.

2A.1.1 Definition of Planning Target Volume (PTV) margin

Most radiotherapy departments define the target volume in accordance with the ICRU recommendations (Lowter, Marsh and Louwe, 2020) which is an international reference that publishes reports on photon-based radiotherapy treatment prescription, recording, and reporting (Stroom et al., 2014).

The PTV is defined as the volume that contains both clinical and subclinical illness, as well as a margin to account for set-up errors and internal movements. The PTV margin should be large enough to prevent geographical miss, however ensuring that it is limiting the dose to adjacent critical structures (ICRU, 2010).

The margin size is affected by a number of factors, such as anatomical area variation, imaging frequency, immobilisation equipment type, treatment modality, patient collaboration, and set-up procedures (Kapanen et al., 2013; Anjanappa et al., 2017). In addition, set-up errors, target volume delineation, and organ motion should ideally be considered when calculating the margin size (Suzuki et al., 2012; Stroom et al., 2014). There are numerous PTV margin formulas, such as those proposed by Van Herk, Stroom, ICRU 62, Antolak, Bel, McKenzie, and Parker, and the method of measurement used influences the margin size (Gill et al., 2015; Oh et al., 2016; Namysl-Kaletka, Tukiendorf, and Wydmaski, 2015).

2A.1.2 Overview of the formulas used in calculating the PTV

All PTV margin equations make assumptions and have limitations, and the equation chosen must take these assumptions and limitations into account.

Stroom et al., 1999 formula

The margin formula proposed by Stroom et al. (1999) was:

$$2\sum + 0.7\sigma,$$

where $\sum$ stands for population systematic errors and σ stands for random population errors and ensures that at least 95% dose is administered to 99% of the CTV. The authors used data from patients receiving treatment to the prostate, cervix, and lung cancer, and demonstrated that this method was fast and accurate for these cases (Stroom et al., 1999). This margin formula implied that the effect of systematic errors was almost three times more significant than the effect of random errors.

When using 2D PTV instead of 3D PTV, the errors will increase with a conformal 95% isodose volume enclosing the PTVs. With conformal shaped fields, multiple 2D procedures might result in under dosage (Stroom, 2000). One of the limitations for this formula is that for volumes with a small diameter, the probability for the volume to be partly outside of the margin will be larger. Another limitation is that narrow CTV regions will be blurred by geometric uncertainties (Yan Li et al., 2014).

Stroom (2000) developed a formula for calculating rotational errors in the orthogonal coordinate system, but he describes it as time-consuming.

ICRU 62 formula

The ICRU 62 (1999) report proposed to separate margins into two components: an internal margin and a set-up margin. The ICRU 62 formula suggested that systematic and random errors should be added in quadrature since the two types of margins are independent from each another. This will obtain a Standard Deviation (SD) value which would then need to be used to calculate the margin. The formula to obtain a value of the SD was described as:

$$SD_{tot} = \sqrt{(\Sigma^2 + \sigma^2)},$$

where Σ is the SD of systematic error and σ is the SD of random error (Thasanthan et al., 2014). To include the 95% of the isodose in the margin, the value obtained to measure SD_{tot} needs to be multiplied by a value of 1.96 (ICRU, 1999).

This method assumes that systematic and random errors have equal effects on the dose distribution unlike Stroom et al. (1999) formula and Van Herk et al. (2000) formula, whereby the systematic errors had more weighting in margin calculation. This may lead to false pretence (Thasanthan et al., 2014). ICRU 62 (1999) also deals with conventional external beam radiotherapy therefore this margin formula was not formulated for complex treatments such as VMAT and IMRT.

Bel et al.,1996 and Antolak et al., 1999 Formulas

Bel et al. (1996) and Antolak et al. (1999) focused on random errors effects. The margin recipe for Bel et al. (1996) was:

$$PTV = 0.7\ \sigma,$$

and Antolak et al. (1999) margin recipe was that of:

$$PTV = 1.65\ \sigma,$$

where σ, for both equations, refers to the SD of random errors. Bel et al. (1996) based the margin recipe on the dose distribution of a prostate radiotherapy plan with three and four rectangular fields and a minimum dose of prescription to the CTV of 95 percent. The 0.7 value refers to a specific beam arrangement and is not applicable to other beam arrangements (Kalyankuppam Selvaraj, 2013). Antolok et al. (1999) recommended a margin of 1.65 multiplied by random errors to ensure a CTV dose greater than 95 percent of the minimum prescribed dose without assuming any specific penumbra profile or beam arrangements.

Both authors assumed that systematic errors are insignificant because they are corrected by offline imaging strategies and quality assurance processes. Because of this assumption, the margin size may be underestimated, causing the dose distribution

to shift and potentially resulting in a geographical miss of the CTV (Vos, Naiker and MacGregor, 2020; Li, 2014; Kalyankuppam Selvaraj, 2013).

McKenzie et al. (2000) formula

McKenzie et al. (2000) proposed that the margin around the CTV account for geometric errors to ensure that no part of the CTV receives less than 95% of the prescribed dose. They applied similar procedures to Van Herk et al. (2000) formula by accounting for random errors, and they determined that a margin should be drawn to account for random set-up and organ motion uncertainties during radiotherapy. The formula is written as:

$$2.5\Sigma + \beta\ (\sigma - \sigma_P),$$

where Σ is the SD of systematic errors, σ is the SD of random errors, σ_p is a Gaussian parameter that defines the width of the penumbra, and β coefficient depends on the isodose chosen to surround the PTV. The β parameter is insensitive to target shape and corresponds to the level of blurred dose (Thomas et al., 2019).

McKenzie et al. (2000) originated the margin formula to account for dose distribution as caused by 1-6 coplanar beams. This enabled the spreading out of the exit dose around the target, minimising the dose blurring effect at the edge of the beam through the use of multiple coplanar beams, which results in a smaller random coefficient than that of the van Herk margin formula (Gordon and Siebers, 2008; Ecclestone, 2012).

Van Herk et al. (2000) formula

One of the formulas that is frequently used in studies to calculate the PTV margins is the Van Herk Margin Formula (VHMF) (Namysł-Kaletka, Tukiendorf and Wydmański, 2015). This formula is expressed as:

$$M=2.5\Sigma + 0.7\sigma.$$

The symbol “Σ” represents the SD for the population systematic errors, and the symbol “σ” represents the SD for the random population errors.

The random errors have a blurring effect on the cumulative dose distribution while systematic errors shift the cumulative dose distribution (Sonke and Van Herk, 2016). This equation ensures that ≥90% of patients receive ≥95% of the prescription dose (Van Herk et al., 2000). The PTV margin derived from the VHMF excludes rotational errors and deviation in the shape of the tumour, therefore Van Herk et al. (2000) suggested that this method of margin calculation should be considered as a lower limit to ensure the delivery of safe radiotherapy. The formula assumes that the shape of the CTV is spherical (Witte et al., 2017). Other assumptions for the VHMF are that tissue is homogeneous and that the number of fractions is limitless, this causes inaccuracies in case of hypo-fractionated or adaptive therapy. VHMF also assumes that the beam penumbra is conformal (Yoda, 2017; Li, 2014; Kalyankuppam Selvaraj, 2013).

Parker et al., 2002 formula

Parker et al. (2002) determined PTV margin in Stereotactic Radiosurgery (SRS) for lesions located intracranially. The margin needed to account for uncertainties in miniature multileaf collimator position, CT scanner and CT-Magnetic Resonance Imaging (MRI) spatial localisation, and head frame repositioning. The CTV dosimetry criteria were selected for the PTV to contain 95% of the CTV dose, and at least 95% of the CTV to receive 100% of the PTV dose.

The following assumptions were made by the authors; the PTV received 100% of the prescribed dose, linear fall-off dose to zero outside the PTV, calculations were performed on a spherical CTV with a 4 mm diameter. Another assumption was on the measured uncertainties since they were assumed to have Gaussian distributions. Systematic and random errors were added linearly and in quadrature respectively with an assumption that the errors were not correlated (Parker et al., 2002).

The suggested margin formula by Parker et al. (2002) was:

$$\Sigma+\sqrt{}(\sigma^2 + \Sigma^2),$$

where Σ refers to the SD of systematic errors and σ refers to the SD of random errors.

2A.2 Target volume delineation of the larynx

Delineation of larynx

The Radiation Therapy Oncology Group (RTOG) head and neck protocol (2014, p. 32) defined the larynx as “triangular prism shaped volume that begins just inferior to the hyoid bone and extends to the cricoid cartilage inferiorly and extends from the anterior commissure to include the arytenoids. It includes the infrahyoid, but not the suprahyoid epiglottis.”

The accurate definition of tumour and normal tissues is essential for radiotherapy planning. Uncertainties in target delineation may result in under- or over-dosage, lowering the probability of tumour control or increasing normal tissue complications (Kristensen et al., 2017). The Danish Head and Neck Cancer Group (DAHANCA) developed a set of radiotherapy guidelines for laryngeal target delineation (Jensen et al., 2020). According to these guidelines, Positron Emission Therapy (PET) scan and MRI should be used for target delineation of cancers in the larynx when available, they are however not recommended for tumours that are not visible on diagnostic scans (e.g., T1N0 larynx cancer). The visibility of tumour borders affects tumour delineation and the finite imaging resolution, or lack of supplementary images causes unavoidable inter-observer and intra-observer contouring variance (Van Herk, Osorio, Troost, 2019). Guidelines and protocols should also be followed to reduce target delineation error (Kim et al., 2021).

Methods of analysis for target volume delineation

Several studies investigated differences in the delineation of tumour and normal tissues. (Jameson et al., 2010; Dewas et al., 2011; Franco et al., 2018; Chang et al., 2021), however the direct comparison of published data is difficult due to the use of a variety of methods to quantify the target delineation error (Segedin et al., 2016).

According to The Royal College of Radiologists (2021), there are two approaches to assessing this error: one is to assess variation among doctors in delineating the target

volume (inter-observer variation), and the other is to assess the mean of the margin outline drawn repeatedly by the same doctor (intra-observer variation).

Even though various methods for quantifying observer variability in target delineation have been proposed, according to Langmack et al. (2014) and Das et al. (2021), there is no single metric type that can be used to fully assess the agreement between sets of outlines. A study by Fotina et al. (2012) had the aim of looking for possible translational tools for evaluating inter-observer variability and to look for common relationships between the different parameters reported. The authors selected, calculated, and compared different metrics that are used to determine inter-observer variation on patient cases. The parameters were classified based on different formalisms into three main groups:

1. Descriptive statistics – contains factors describing the volume distribution, such as mean, SD, range, ratio of the largest drawn volume to the smallest drawn volume, and coefficient of variation (COV).
2. Overlap measure – contains measures that describe the area of overlap between contoured volumes, such as encompassing volume ratio, Dice similarity coefficient (DSC) or Jaccard index (JI).
3. Statistical measures of agreement – contains reliability analysis tools, such as κ statistics or Intraclass Correlation Coefficients (ICC).

This literature review will be discussing the most common metrics used for assessment of target volume delineation agreement.

The Dice Similarity Coefficient

The DSC is both a spatial overlap index and a metric for validating reproducibility. It is considered one of the most popular metrics used in literature relating to inter-observer variation in target volume delineation (Trignani et al., 2019; Van Der Veen, Gulyban and Nuyts, 2019; Chang et al., 2021; Mul et al., 2021), and is expressed as follows:

$$D(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|}.$$

Where |X| and |Y| are the two sets' cardinalities (i.e., the number of elements in each set). The DSC is equal to twice the number of elements shared by both sets divided by the total number of elements in both sets (Andrews and Hamarneh, 2015). A value close to one indicates minimal variation in contouring (Caravatta et al., 2014), with 0.7 considered to be a threshold for good interobserver agreement (Langmack et al., 2014).

This metric does not provide a complete picture of the variability between two sets of volumes. Trignani et al. (2019) stated that a pair of identical volumes with different positions in space exhibit the same DSC. Fotina et al. (2012) also mentioned that DSC provides false impressions of high agreement. Also, this metric does not provide measurement of delineation errors in the three dimensions; therefore, it cannot be used for PTV margin calculation.

Jaccard Index (JI)

A similar metric to the DSC is the JI (Trignani et al., 2019), and it is associated with the DSC in the following equation:

$$JI = \frac{DSC}{2\text{-}DSC}$$

This metric measures the similarity of finite sample sets and is defined as the size of the intersection divided by the size of the sample sets' union. The JI is also situated between zero and one, where a value close to zero indicated disagreement in contouring and one indicates perfect agreement. When compared to the DSC, the penalty for a false positive delineation area increases faster (Trignani et al., 2019). Large datasets can have a significant impact on the index because they can significantly increase the union while keeping the intersection constant. Similar to DSC, the JI can only be used as a conformity index when comparing two delineated volumes and therefore cannot be used for margin calculation (Jager et al., 2015).

Conformity Index general

Conformity Index general (CIgen) enables quantification of agreement between all observers (Jager et al., 2015). The CIgen index is useful for comparing any number of delineations at the same time. CIgen = 1 denotes total overlap, whereas CIgen = 0 denotes completely separated volumes. One of the limitations of the CIgen index is that as a value that can be difficult to interpret (Kristensen et al., 2017). As a metric it is useful in determining inter-observer variability, but the obtained values cannot be used to determine the systematic target delineation errors.

Mean Distance to Agreement (MDA)

The MDA is a quantitative measure of the mean distance of voxels at the outlying points of the contour that must be moved to achieve perfect conformity with the reference contour. Jena et al. (2010) developed the MDA index to compare radiotherapy target volume delineation by different observers. Two perfectly concordant volumes have an MDA of 0 mm (Trignani et al., 2019).

MDA, according to Jena et al. (2010), provides a single scoring statistic that represents the overall conformity of the two volumes being assessed. It also provides additional statistics information on whether the non-conformity is caused by over-or under-outlining, this metric however cannot be used to determine the systematic errors of target volume delineation as the obtained values are not given in LR, SI and AP directions.

Hausdorff Distance

The Hausdorff distance is a shape comparison method that is able to demonstrate the maximum distance between each voxel in the reference set and the nearest point in the comparison set. Similar to MDA, lower values (in mm) correspond to a higher agreement between the compared volumes ((Trignani et al., 2019). Also, this metric cannot be used for margin calculation as the obtained values are not given in the LR, SI and AP directions.

Target delineation errors and PTV margin

According to the findings of this literature review, most studies used a combination of metrics to assess inter-observer variability (Fotina et al., 2012; Langmack et al., 2014; Liu et al., 2021). A combination of parameters is useful when reporting inter-observer variability in delineation since they make the data more reliable (Fotina et al., 2012). Mercieca, Belderbos and Van Herk (2021) affirmed the use of more than one metric to quantify inter-observer variation since the lack of an absolute gold standard makes accurately validating the precision of a delineated contour difficult.

Fotina et al. (2012) stated the preferred metrics for assessment of inter-observer agreement in target volume delineation are the JI, Conformation Number (CN), or CIgen since they obtained similar results of agreement. They found that DSC, when compared with the aforementioned metrics, generally produces higher calculated values of inter-observer agreement. MDA results, on the other hand, were significantly lower than the other indices. Similarity in results were observed in Trignani et al. (2019) study, were DSC, JI, MDA and Hausdorff distance were evaluated for target volume delineation agreement in the head and neck region. The difference in value between DSC and JI was 0.16, with the former having the highest value, whilst the distance obtained by MDC was 8.89 mm and mean Hausdorff distance was 36.58 mm. However, the aforementioned metrics cannot be used to calculate the PTV margin. The studies that used these metrics were interested in measuring the inter-observer variation of target delineation rather than measuring the PTV margin (Fotina et al., 2012; Jager et al., 2015; Kristensen et al., 2017; Tsang et al., 2019; Trignani et al.,2019; Hammers et al., 2020). Metrics such as DSC and CI compare relative volumes rather than absolute distances, and thus the results obtained cannot be directly translated to calculate the PTV margin. Peulen et al. (2015) proposed evaluating the distances between the delineated volumes.

According to Bernstein et al. (2021) the only metric for delineation uncertainty that can be used in a traditional PTV margin recipe is to quantify the delineation errors (ΣD). The ΣD occurs when the tumour is not properly contoured, and it persists throughout the treatment. Because these errors occur during the planning phase, they are carried over to the treatment phase and are therefore considered as systematic errors (Ecclestone, 2012). Bernstein et al. (2021) study's goals were to first quantify the

delineation uncertainty for recurrent gynaecological cancer Gross Target Volumes (GTVs) and then to calculate the associated PTV margins. The ΣD is the SD of the distances between each outline and a reference outline and is calculated using the following equation:

$$\Sigma_D(d) = \sqrt{\frac{1}{N_O - 1}\sum_{i=1}^{N_O}(d_i - \bar{d})^2}$$

Where d_i is the distance between the reference outline and the ith observer's outlines, N_o is the number of observers and $\bar{d}$ is the mean distance. This study used an approach by Deurloo et al. (2005) to measure ΣD for each point in a patient whereby the 2D contour sets were first converted to 3D surfaces. From all the doctors' GTV surfaces, a reference surface for each patient was created. At a 95 percent confidence level, the simultaneous truth and performance level estimation (STAPLE) algorithm was used to create a common consensus volume. Based on the provided outlines, this resulted in a surface that contains voxels that have a 95 percent or higher probability of belonging to the GTV.

Six vectors were created determine which sector each vertex on the consensus surface belonged to. These vectors originated in the volume's centre and pointed parallel to the patient anterior, posterior, left, right, superior and inferior axes. Another vector which originated at the centre of the consensus surface and terminated at vertex was created. The angles between vector originating at the centre and each of the vectors in the axes were then measured. The geometric mean of ΣD, calculated over all vertices in that patient resulted in the overall ΣD for a single patient.

A similar method to that of Bernstein et al. (2021) was used by Peulen et al. (2015) where the researchers were interested in quantifying the variability in target delineation in peripheral early-stage lung cancer treated with SBRT and calculate the PTV margin. The delineated contours were at first triangulated. For each patient in the data set, a 3D median surface of all observers' triangulated GTVs was computed where each point inside the median surface was designated as part of the GTV by 50% of the radiation oncologists. Following that, the perpendicular distance to each individual triangulated GTV was measured for each point describing the median surface. SD was used to express the variation in distance to a single point. The variation in distance to

all points describing the median surface (root-mean-square) was expressed as an overall SD, which is a measure of target delineation variability. The triangulation technique was also used in Remeijer et al. (1999) study. This study presented a generic method for evaluating target volume delineation in three dimensions across multiple imaging modalities.

The methods proposed by Bernstein et al. (2021), Peulen et al. (2015) and Remeijer et al. (1999) are intricate. Tudor et al. (2020) proposed a simpler approach by superimposing all delineated contours to see them all at once and identifying a plane with no steep gradients in the out of plane direction. The spatial difference between multiple observers' delineations can be used to estimate delineation errors (Figure 2B.1).

Tudor et al. (2020) proposed two methods for calculating delineation errors, both of which are dependent on sample size. If the sample size is greater than 15, it was suggested to use the same sample SD equation as mentioned in the Bernstein et al. (2021) study. If, on the other hand, the sample size is less than 15, the standard deviation can be calculated from the range using the following equation:

$$S = \frac{R}{d_2(N)}$$

R represents the data range and is calculated by measuring the distance between the inner and outermost contours in each image plane along each axis of interest at a representative point with 'average' observer variation and no outliers. $d_2(N)$ is dependent on the number of samples in the range.

The greater the number of cases studied, the lower the uncertainty in the measured SD of delineation errors. To avoid bias, care should be taken to ensure that participating observers behave as they would normally as part of the clinical process, such as using standard software and being blinded to the contours of other observers. Any disagreements should be resolved before measuring the delineation uncertainties, as they will result in an incorrect estimate of the measured delineation errors (Tudor et al., 2020).

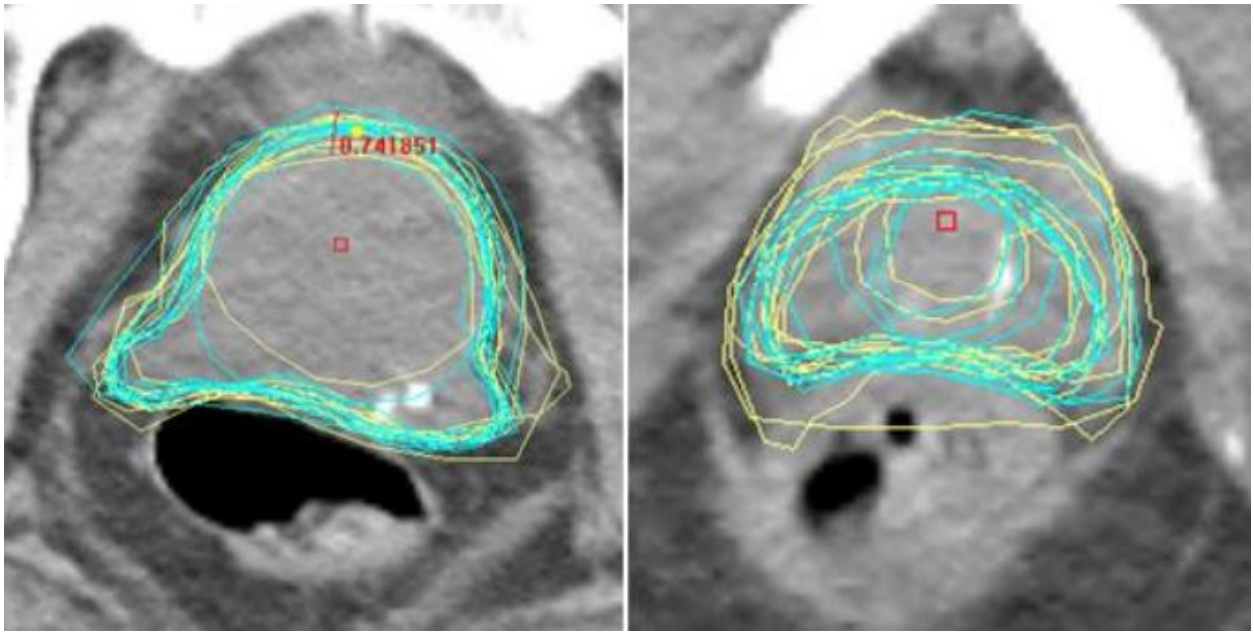


Figure 2A.1: Illustration of Determining the Magnitude of Delineation Errors. (Adapted from: Tudor, G., Bernstein, D., Riley, S., Rimmer, Y., Thomas, S., Herk, M., & Webster, A. (2020). Geometric Uncertainties in Daily Online IGRT: Refining the CTV-PTV Margin for Contemporary Radiotherapy.DO - 10.1259/geo-unc-igrt https://www.researchgate.net/publication/346717324_Geometric_Uncertainties_in_Daily_Online_IGRT_Refining_the_CTV-PTV_Margin_for_Contemporary_Radiotherapy)

Intra-observer variability

In addition to inter-observer variability, inaccuracies can occur within repeated contours produced by the same observer, a phenomenon known as intra-observer variation (Van Herk, 2004). According to Das et al. (2021), understanding intra-observer variability is critical in calculating the margin each doctor adds to their original outlines. Each doctor should be aware of his or her SD, which should be added to the original target volume during treatment planning. However, it is not a required assessment when estimating delineation error using a study of multiple delineators, since only one contour from each observer is required (Tudor et al., 2020).

2A.3 Methods of analysis for inter-fraction set-up errors

Inter-fraction variation occurs from one treatment fraction to the next during the treatment course, possibly due to daily variability in patient positioning, daily variations such as organ filling and weight changes, or volumetric changes in the tumour (Guerreiro et al., 2018).

The imaging modalities used to assess inter-fractional movement are the megavoltage portal imaging (MVPI), kilovoltage portal imaging (kVPI), the megavoltage cone-beam computed tomography (MVCBCT), and the kilovoltage cone-beam computed tomography (KVCBCT) (Zhou et at, 2018).

CBCT was initially developed for dental imaging in the early 2000s (Quereshy et al, 2008). It utilises megavoltage (MV) or kilovoltage (kV) energy emitted from an additional x-ray tube and opposing amorphous silicon flat panel imager, mounted 90˚ from the gantry head. It delivers radiation exposures at different projection angles, whilst the gantry head rotates. The software then performs a 3D reconstruction of the images for better visualisation of the patients' anatomy and target area (Srinivasan et al, 2014).

Osman et al. (2011) were among the few researchers who used CBCT imaging to estimate laryngeal inter-fraction errors and verify the quality of image registration for the vocal cords. The residual systematic and random set-up errors obtained from daily imaging were quantified in ten patients who received conventional treatment for laryngeal cancer. A clip-box containing the volume of interest was defined around the thyroid cartilage on the reference image. The CBCT was then automatically registered to the reference image (planning CT) before each treatment fraction using the XVI software's grey value match function, and manually matched by an observer. The coordinates of three representative points (front node of the thyroid cartilage, as well as the centres of the left and right arytenoids), surrounding the soft tissue region of interest were checked in each CBCT against the reference image after registration to verify the quality of the registration for the vocal cords. For each anatomical reference point, the residual inter-fractional errors were calculated. The resulting residual errors

of the anatomical structures were sub-mm in all three orthogonal directions, indicating that the automatic image matching provided suitable corrections for the vocal cords. The measured residual errors may have been introduced by observer variability. The researcher however stated that it is still vital to examine the quality of all automatic registration results as there were instances where the grey value automated match was not satisfactory for one of the patients in the study.

The effect of various CBCT-based matching methods on set-up errors for head and neck patients treated with VMAT was investigated by Mohandass et al. (2020). In this study, manual, bone, and soft tissue image registration were investigated. When the results of the three matching methods were compared, there were no statistically significant differences in systematic errors, random errors and mean set-up errors. The researcher stated that all three methods can be used without compromising the VMAT treatment. Bahig et al. (2021), however emphasised on the importance of daily imaging with a soft tissue match on patients treated to the larynx, since anatomical changes or set-up reproducibility tend to create a laryngeal shift in relation to the vertebra.

A study by Palombarini et al. (2012) investigated inter-fractional errors as a combination of set-up and organ motion errors for the prostate gland relative to the bony anatomy. To quantify organ motion, the set-up errors of the CBCT scan were first measured using a pelvic bone match with the planning CT scan. Using grey-value soft tissue matching, a manual match was later performed to superimpose the prostate on the planning CT and CBCT. The prostate motion relative to the bony anatomy was defined as the difference between the soft tissue match and the bony match in the Left-Right (LR), Superior-Inferior (SI) and Anterior-Posterior (AP) directions. The average deviation and standard deviation for bone match, soft tissue match, and organ motion were calculated for each patient in each direction. The systematic and random errors of the entire patient population, as well as the patient's standard deviation means were computed. This procedure proved to be an effective way to investigate set-up and organ motion errors.

The Royal College of Radiologists (2008) wrote an extensive report on how to calculate the inter-fraction set-up errors. One of their suggestions was that set-up

errors should be resolved into orthogonal directions for ease of analysis. Vectors quantities should be calculated to keep the correct information on direction (if shifts in the anterior direction are given a negative sign, then shifts in the posterior direction are always positive). Another suggestion made in this report was that small patient studies, such as that of Osman et al. (2011), can lead to a high level of uncertainty in the population systematic set-up error. Even for small patient numbers, the random set-up error is likely to be more accurate as long as sufficient images are acquired per patient and inter-patient variability is not excessive. As a result, it was recommended for at least 20 patients to be included in a study with at least five images per patient.

The equations for calculating systematic and random errors for inter -fraction set-up errors are given below and are all taken from The Royal College of Radiologists' report (2008).

Individual systematic error

Systematic error is a deviation that occurs in the same direction and has a similar magnitude for each fraction (The Royal College of Radiologists, 2008). The individual systematic error ($m_{individual}$) is calculated by summing the set-up errors for every imaged fraction (Δ1+ Δ2+ Δ3…), then dividing this value by the number of imaged fractions (n). This error is expressed in the following equation:

Equation 1

$$m_{individual} = \frac{\Delta_1 + \Delta_2 + \Delta_3 + \ldots\ldots + \Delta_n}{n}$$

(The Royal College of Radiologists, 2008)

Population systematic error

After obtaining the individual systematic error, the mean systematic error of the population (Mpop) is calculated. This is the mean of the individual systematic errors for the analysed patient group. A resultant value of zero indicates that there isn't a common underlying error pertaining to the sample. To calculate this error, the mean of each individual patient ($m_1+m_2+m_3+$ ...) is summed up and the resultant value is divided by the number of patients (*P*) in the group.

Equation 2

$$M_{pop} = \frac{m_1 + m_2 + m_3 + \ldots\ldots + m_p}{P}$$

(The Royal College of Radiologists, 2008)

The population systematic error is then calculated by adding up the squares of the differences between individual systematic error mean derived from equation 1 with the mean of the individual systematic error of the population derived from equation 2. The resultant sum is then divided by the number of patients minus 1 and the square root is then calculated on the obtained result. This error is expressed in the following equation:

Equation 3

$$\sum_{set\text{-}up}^{2} = \frac{(m_1 - M_{pop})^2 + (m_2 - M_{pop})^2 + (m_3 - M_{pop})^2 + \ldots + (m_n - M_{pop})^2}{(P - 1)}$$

(The Royal College of Radiologists, 2008)

Individual random error

A random error is a deviation that can vary in both direction and magnitude (The Royal College of Radiologists, 2008). To calculate the random error, first the individual

random error (σindividual) needed to be calculated. This is a SD of the corresponding mean individual set-up errors (m). This error is measured by summing the squares of the differences between the mean and set-up error from each image. This sum is then divided by the number of scans minus one. The square root of the resultant sum revealed the individual random error.

Equation 4

$$\sigma^2_{individual} = \frac{(\Delta_1 - m)^2 + (\Delta_2 - m)^2 + (\Delta_3 - m)^2 + ... + (\Delta_n - m)^2}{(n-1)}$$

(The Royal College of Radiologists, 2008)

Population random error

The Population Random error is then calculated by measuring the mean of individual random errors.

Equation 5

$$\sigma_{set\text{-}up} = \frac{\sigma_1 + \sigma_2 + \sigma_3 + ... + \sigma_p}{P}$$

(The Royal College of Radiologists, 2008)

This methodological process is also relevant when calculating intra-fraction set-up errors (The Royal College of Radiologists, 2008).

2A.4 Intra-fraction errors

Intra-fraction errors may be either due to patient movement resulting from variations in patient positioning during treatment delivery, or it could result from motion uncertainties resulting from displacements of the tumour bed and organs at risk caused by either respiration or movement (Guerreiro et al., 2018).

Real-time imaging techniques such as MRI-guided radiotherapy, fluoroscopy, and ultrasound are commonly used to assess intra-fraction errors (Bradley et al., 2011; Nonaka et al., 2019). CBCT images can also be used to determine this type of error in the absence of such modalities. To some extent, post-treatment imaging can quantify both intra-fraction motion and residual errors, but it has limitations regarding how much information is gathered throughout the treatment (The Royal College of Radiologists, 2021).

The larynx as a structure is predisposed to movement due to swallowing and respiration. Swallowing is associated with a 2 cm elevation of the larynx, and respiration can cause motion reaching 6 mm in the longitudinal direction (Bahig et al., 2021). Some authors performed swallowing control during the CBCT procedure to try and limit thyroid movement (Kwa et al. 2015; Perillo et al, 2021). Other researchers only performed swallowing control during the acquisition of the planning scan since the incidence and total duration of swallowing was small compared to the treatment time, and this reduces systematic errors (Van Herk, 2004; Osman, 2011; Bahig et al. 2021).

Durmus, Tas and Uzel (2020), where interested in determining the intra-fraction target movement through CBCT scans and in investigating the effects of laryngeal movement on the target volume. Patients were positioned with a maximum neck extension to minimise swallowing during the CT scan procedure and treatment. Two CBCT scans were performed for each fraction on sixteen patients with the thyroid cartilage used as a matching structure during image registration, and both images were matched with the reference CT. The first CBCT was obtained after setting up the patient on the treatment couch and matched with the reference CT planning images. The second

scan was also matched with the reference CT image and was obtained after treatment delivery to determine the thyroid cartilage movement during treatment.

Kwa et al. (2015) and Perillo et al. (2021) also used thyroid cartilage to match images because as a structure it adheres to the involved vocal cord. This, however, does not address the issue of residual motion caused by breathing. The purpose of the studies by Perillo et al. (2021) and Kwa et al. (2015) was to evaluate inter- and intra-fraction errors and quantify CTV to PTV margins. Perillo et al. (2021) investigated stereotactic treatment of early stage glottic cancer, whereas Kwa et al. (2015) investigated single vocal cord irradiation (SVCI) of T1a larynx tumours. In both studies, intra-fraction set-up uncertainty was assessed by acquiring post-treatment CBCT scans that were registered to the planning CT.

Perillo et al. (2021) corrected for set-up errors greater than 1 mm and up to 2 mm in any direction, and patients were treated after the couch correction was applied. After couch correction, a new CBCT image was acquired if the errors exceeded 2 mm. When the set-up errors were less than 2 mm, the CBCT before treatment delivery was considered to be the set-up CBCT. At the end of treatment, a final CBCT image was taken and this CBCT image was matched with the pre-treatment CBCT. Swallowing was always kept under control. For each patient, the displacement values indicated by CBCT imaging immediately before treatment and immediately after delivery were recorded. After that, displacement data in three directions (LR, SI, and AP) were extracted. The systematic and random errors in each direction of the three treatment sessions were calculated. The set-up variations were summarised for the entire population by taking an average of all systematic errors, obtaining the standard deviation of all systematic errors, and the root mean square of all random errors. CTV to PTV margins were calculated in each direction using the van Herk formula.

Kwa et al. (2015) used daily CBCT imaging to online correct the thyroid cartilage set-up after patient positioning with in-room lasers (inter-fraction motion correction). CBCT scans were also obtained shortly after patient repositioning and dose delivery to monitor intra-fraction motion. The margins obtained from Kwa et al. (2015) study, 1.6, 4.3, and 2.2 mm, where interestingly very similar to those estimated by Perillo et al.

(2021), 2.4, 5.1, and 2.2 mm in the X, Y, and Z directions, respectively. Both studies used the same strategy of withholding swallowing.

2A.5 Observer variation in image matching

IGRT observer variability is the variation that occurs when different observers analyse the same IGRT data set, and it should be quantified locally (The Royal College of Radiologists, 2021). This error should be considered when determining the PTV margin size, especially when using CBCT images for soft tissue-based patient positioning (Deegan et al., 2015; Hirose et al., 2020; The Royal College of Radiologists, 2021).

Data on the effects of inter-observer variation on the PTV margin size are limited because this error is frequently overlooked in studies that assess the effect of errors on the PTV margin size. A study by Hirose et al. (2020) investigated the impact of inter- and intra-observer variation of six radiographers in image matching on the CTV to PTV margin size. A total of twenty-six scans of patients who had undergone treatment to the prostate were analysed. The residual errors, which represented the difference in soft tissue and reference positioning errors for each fraction, were used to assess observers' uncertainties. Prostate Cancer Location Errors (PCLEs) of contour-based patient positioning between the reference images and pre-CBCT images were used to calculate reference positioning errors. The PCLEs indicated the centroid distance of prostate contours on reference images and pre-CBCT images. Each participating radiographer first matched the bone anatomy by performing automatic registration, followed by manual refinement of prostate position in the AP, SI, and LR directions without rotation correction. The recorded inter-observer systematic errors were measured to be 0.9, 0.9, and 0.5 mm, respectively in the AP, SI and LR direction, and 1.8, 2.2, and 1.1 mm, respectively, for random errors. Intra-observer variations were measured to be less then inter-observer variation and were found to have an insignificant effect on the PTV margin size. When incorporating both inter- and intra-observer variation, the resulting PTV margin size was measured to be 3.5, 3.8, and 2.1 mm, respectively in the AP, SI and LR direction. The resulting

values of inter-observer systematic and random errors are indicative of the importance of measuring these errors in the PTV margin calculation.

The use of a reference CT improves registration accuracy (Deegan et al., 2015). With the use of a reference CT, the radiographers are able to identify features which may aid in the refinement of soft tissue matching. Deegan et al. (2015) assessed image registration of six patients who undergone radiotherapy to the prostate. The researchers compared image matching with fiducial markers and soft tissue image registration, and the inter-observer variation was assessed using Bland-Altman analysis. The three radiographers who participated in the study were unable to use the planning CT scan as a reference during image matching, this might have led to higher inter-observer variability in soft tissue image matching.

Training and competency, as well as adherence to local and national protocols, will help to reduce inter-observer variation in image matching. These errors will generally average out over many fractions, but they can be a significant source of systematic errors in hypo-fractionated regimes (The Royal College of Radiologists, 2021). The development of site-specific image guidance protocols can help to reduce the image matching uncertainty, and aid in the reduction of the margin size through stringent IGRT methods (Deegan et al., 2015).

2A.6 Other errors

Radiotherapy errors are not limited to the ones previously listed in this review. Other uncertainties can impact on margin size, but some of these errors are difficult to quantify.

Type of uncertainties in radiotherapy include:

- Organ motion errors can occur in the delay between imaging and treatment (Van Herk, 2004);
- Phantom transfer errors: a result of the minor differences between the CT planning scan and the linear accelerator due to inaccuracies resulting from manufacturing. These errors can be measured using a rigid phantom (Liang et al., 2014);
- Errors caused by the approximation of shifts (Palombarini et al., 2012).
- Anatomic changes: patients' internal changes which might be due to tumour response, growth or oedema (Deegan et al., 2015);
- Image-guidance errors: Can result from the limited resolution of IGRT system (Deegan et al., 2015);
- Dose calculation uncertainties (Deegan et al., 2015);
- Multi-leaf collimator (MLC) motion uncertainty (Deegan et al., 2015).

Van Herk (2004) believes that it is impossible to eliminate all geometrical errors, and as a result, the PTV margin can never be reduced to zero.

2A.7 Conclusion

PTV margin assessment needs to take into consideration various errors that are present in the delivery of radiotherapy treatment.

Various metrics are available to assess inter-observer variation. A combination of parameters is useful when reporting inter-observer variability in delineation, this makes the data more reliable. However, in the assessment of the PTV margin recipe, the only metric for delineation uncertainty that can be used is to quantify the delineation errors to calculate the CTV–PTV margin required at each point (Bernstein et al., 2021).

The Royal College of Radiologists (2008) recommended that at least 20 patients should be included in a study assessing inter-fraction errors, with at least five images per patient. Soft tissue match to the larynx is preferred in inter-fraction errors assessment, especially when tight margins are applied (Bahig et al., 2021).

Intra-fraction errors are a result of patient movement and organ motion. The few studies that identified intra-fraction errors in the laryngeal region with CBCT scans suggested swallowing suppression during the CT scanning procedure to avoid introducing systematic errors (Van Herk, 2004; Osman, 2011; Bahig et al. 2021). Intra-fraction set-up errors using CBCT imaging were assessed by acquiring post-treatment CBCT scans that were registered to the planning CT. Studies recommended superimposing the position of the thyroid cartilage during image registration.

Observer variation in image matching is another source of error which needs to be accounted for in PTV margin calculation (The Royal College of Radiologists, 2021). Training and adherence to guidelines enable a reduction in this error.

In the next chapter, the methodological process used to determine the PTV margin size in patients treated with VMAT for laryngeal cancer are discussed in detail.

2.3 Section B – Systematic Literature Review

2B.1 Introduction

The systematic literature review chapter explains the search strategy procedure, the selection of literature that was employed, and the evaluation of the quality of the literature. In this chapter the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) checklist was used, where the theoretical framework, conceptual framework and critical appraisal of the literature were explained to support this study.

Evidence-based practice is critical in clinical settings (MacLure et al., 2016), but evidence must come from a reliable source (Harvey and Kitson, 2016), so a systematic literature review approach was chosen to minimise bias during research analysis and to ensure reliable findings (Saleh et al., 2014).

This systematic review was written with the goal of identifying the various methods of calculating the PTV margin in patients treated to the head and neck region with VMAT and adapting the most appropriate method for the local department with the available resources.

Advancement in imaging technology and treatment modality has enabled a reduction in the CTV-PTV margin, however there are other factors that influence the size of the margin (Liang et al., 2014). According to the ICRU 83 (2010) report, PTV margins should be determined individually for each radiotherapy centre, however this report does not specify which method should be used.

2B.2 Rationale for the systematic literature review

The previous section showed that there are several published methods of calculating the PTV margin and the selected method can have an impact on the margin size (Gill et al., 2015; Namysł-Kaletka, Tukiendorf and Wydmański, 2015).

Certain factors also influence the size of the margin such as different anatomical regions, frequency of imaging, immobilisation devices, treatment modality, patient collaboration and set-up procedures (Kapanen et al., 2013; Anjanappa et al., 2017; Djordjevic et al., 2014).

Since the PTV margin calculation influences the PTV margin result, the need to perform a systematic literature review were:

- To determine the various methods used in research studies to calculate the PTV margin for patients receiving VMAT in the head and neck region.
- To identify whether the studies in the review considered all the relevant factors required to calculate the margin size with the chosen formula.

The systematic literature review findings addressed the research question presented in chapter 1 to identify methods of calculating the PTV margin for head and neck patients treated with VMAT and adapt the most appropriate method locally based on local resources.

2B.3 Methods

2B.3.1 Protocol and registration

Before commencing the systematic literature review, an online search was conducted using The International Prospective Register of Systematic Reviews (PROSPERO) to ensure that no similar literature review had been or was being conducted. The protocol was then approved (registration number: CRD42020183573).

A detailed protocol, according to Moher et al. (2015), would facilitate the appraisal and understanding of the method chosen for the review and would be able to detect any method modifications that are required. The PRISMA checklist (Moher et al., 2019) was used as a guideline in this study to write up the protocol for the systematic literature review. PRISMA assists authors by requiring complete and transparent reporting of systematic reviews and meta-analyses (Liberati et al., 2009). The Population, Intervention, Comparison, Outcome, and Study Design (PICOS) framework (Table 2B.1) guided this review protocol (Eldawlatly et al., 2018).

Table 2B.1. PICOS Elements for Eligibility Criteria

Population	The patients receiving VMAT radiotherapy to the head and neck region
Intervention	The different methods used to calculate the PTV margin
Comparison	No comparison made
Outcomes	Determine the various methods for calculating PTV margin and their effect on the PTV margin result.
Study Design	Quantitative studies

2B.3.2 Eligibility criteria

The eligibility criteria for analysing the literature specifies which studies were included and excluded from the systematic literature review (McKenzie et al., 2019). Eligibility criteria were defined by considering the research question, the most relevant study design, and the weakest acceptable study design (Bettany-Saltikov, 2012). The eligibility criteria influence the applicability and validity of the review and ensure that the chosen articles are selected in a systematic and unbiased manner (Liberati et al., 2009).

Table 2B.2. demonstrates the inclusion and exclusion criteria used to identify studies.

Table 2B.2. Inclusion/Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Studies of patients treated with VMAT	
Image guidance prior delivery of Radiotherapy	
Patients treated to the head and neck region only	
PTV margin calculation	Studies that do not calculate PTV margins
No age restriction	
Availability of full articles	
Quantitative study	
Prospective or Retrospective	
English language studies	

The review included studies post 2007 (The rationale for excluding studies that were published pre 2007 is that VMAT was first introduced as a treatment modality in 2007 (Teoh et al., 2011).

2B.3.3 Information sources

Papers were sought through CINAHL, MEDLINE, PubMed, ProQuest (Nursing and Allied Health), Scopus and tipsRO. These databases are known to contain several articles related to health care and radiotherapy (Hummel et al., 2010; Barry and Kell, 2011; Tsang et al., 2019). Other searches were performed on Hydi which is an institutional library search engine, and on the ScienceDirect platform. White and Grey literature were both sought with the intent of reducing publication bias (Saleh et al., 2014).

The literature search was done between April and December 2020. Authors that did not have their published studies fully available for review were contacted and asked if they could provide a full text or summary of their study.

2B.3.4 Keywords for literature search

Prior to searching for relevant literature, the validity of keywords was tested by asking two experts in head and neck radiotherapy to rate the eligibility criteria and key-word search strategy. One of the experts was a clinical oncologist with over ten years of clinical experience who specialised in head and neck cancer treatment, and the other was a senior radiographer with over five years of clinical experience who was involved in the implementation stage of VMAT to the head and neck region.

Content validity was used to assess validity of the literature search strategy. Four criteria were used, originally developed by Waltz and Baussel (1981), to determine the relevance of the key words for the literature search strategy. The criteria were the following: “Not relevant, Somewhat relevant, Quite relevant and Highly relevant”, and each criterion was assigned a score from 1 to 4, with 4 being assigned to “Highly relevant”. When the experts agreed on each item, content validity was established.

Each category was aimed to reach over 80% mean agreement amongst the experts for each keyword to be included in the search strategy (Sangoseni, Hellman and Hill, 2013). There was good content validity as all the keywords were deemed highly relevant by the experts with a 100% mean agreement. The experts suggested

additional keywords to ensure that all relevant articles will be accessed. The suggestions were to include the following in the search list: Supraglottis, Subglottis, Glottis, Tongue, Sinuses, Thyroid and Lymphoma (**Appendix D**).

Other related terms that can be used as keywords were identified from the Medical Subject Headings (MeSH) thesaurus (Baumann, 2016). Boolean operators were used, such as 'AND' and 'OR' between the keywords and this allowed the combination of words and phrases to retrieve relevant literature from databases. Inverted commas were used on phrases to include all terms. The asterisks (*) was used next to some of the keywords since certain terms can be written in two ways. An example of this would be the keyword Nasophary* since the asterisk was used to look for nasopharynx and nasopharyngeal search results.

An exhaustive search for related research and studies was done through the following combination of keywords:

- PTV/Planning Target Volume
- Oropharyn*
- Hypopharyn*
- Nasopharyn*/Nasal cavity
- Laryn*/Supraglottis/Subglottis/Glottis
- Sinus*
- Thyroid
- Oral cavity/Mouth/Tongue
- Head and Neck
- Lymphoma
- Set-up/setup/set up
- VMAT/Volumetric-Modulated Arc Therapy/Volumetric Modulated Arc Therapy/RapidArc Therapy
- Error/errors

An example of a search strategy that was used when searching PubMed with a full text filter is presented in Figure 2B.1:

Query	Results
Search: **(PTV OR Planning Target Volume) AND (VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR Rapid Arc Therapy) AND (set up OR setup OR set-up) AND (error or errors) AND (Hypopharyn*)** Filters: **Full text, Humans, English, from 2007 - 2020**	1
Search: **(PTV OR Planning Target Volume) AND (VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR Rapid Arc Therapy) AND (Oral cavity OR mouth OR tongue) AND (Setup OR set-up OR set up) AND (error or errors)** Filters: **Full text, Humans, English, from 2007 - 2020**	12
Search: **(PTV OR Planning Target Volume) AND VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR Rapid Arc Therapy) AND (oropharyn*) AND (set up OR setup OR set-up) AND (error OR errors)** Filters: **Full text, Humans, English, from 2007 - 2020**	15
Search: **(PTV OR Planning Target Volume) AND VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR Rapid Arc Therapy) AND (Nasopharyn* OR Nasal cavity) AND (set up OR setup OR set-up) AND (error OR errors)** Filters: **Full text, Humans, English, from 2007 - 2020**	28
Search: **(PTV OR Planning Target Volume) AND (VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR Rapid Arc Therapy) AND (Laryn* OR supraglottis OR subglottis OR glottis) AND (Setup OR set-up OR set up) AND (error OR errors)** Filters: **Full text, Humans, English, from 2007 - 2020**	12
Search: **(PTV OR Planning Target Volume) AND (VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR RapidArc Therapy) AND (Head and Neck) AND (Setup OR set-up OR set up) AND (error)** Filters: **Full text, Humans, English, from 2007 - 2020**	55
Search: **(PTV OR Planning Target Volume) AND (VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR Rapid Arc Therapy) AND (Thyroid OR Sinus* OR lymphoma) AND (Setup OR set-up OR set up) AND (error OR errors)** Filters: **Full text, Humans, English, from 2007 - 2020**	7

Figure 2B.1. Example of the Search Strategy

2B.3.5 Selection process

As suggested by Stoll et al. (2019), a dual independent review of search results was conducted between April and December 2020. The first part of the review was to screen for the inclusion and exclusion of studies based on the title and abstract. The second part of the review was performed by reading the full text of the eligible studies selected in the first part. This process was also performed independently by the two

reviewers. Any disagreements with regards to data suitability was identified and resolved through discussions between the two until a consensus agreement was reached. The studies that fulfilled the criteria were included for the systematic review.

The full texts were then reviewed, and the search for additional relevant studies was aided by looking through the reference lists of the eligible articles, as citation text has the potential to retrieve studies that would not have been retrieved by the keyword search strategy (Papaioannou et al., 2010).

2B.3.6 Data collection process

Data extraction sheet

A comprehensive research database worksheet was created on Microsoft Excel to record the relevant extracted data from the articles selected.

The following quality measurements guided by the PRISMA checklist are the key constructs for structure and organisation purposes for the reviewed papers:

- Title and year of publication
- Geographical location where the study was performed
- Details of methods (study design, sampling procedure, length of sample follow up, risk of bias)
- Sample number (randomly assigned, withdrawal from study or exclusion with reason)
- Age range of the sample
- Anatomical region of the head and neck
- Prescribed dose
- Institution PTV margin
- Immobilisation
- Type of RT linear accelerator and other equipment used
- Imaging protocol (frequency, matching procedure, and type of imaging)
- Calculated PTV method (statistical analysis)
- Reason for choice of calculation method

- PTV margin result

Pilot study

Prior to data collection, a pilot test was performed. As suggested by Long (2014), this approach was taken to ensure that the most useful and relevant information was extracted from the studies, avoiding the need to revisit papers at a later stage.

For the pilot study, two articles were randomly selected from the pool of studies that fulfilled the eligibility criteria. The categories of the data extraction sheet were refined to extract data which was related to the review question. From the pilot test, it was noted that it would be important to add the following parameters to the data extraction sheet:

- Specific region of head and neck under investigation
- Imaging protocol
- Type of immobilisation device/s used
- Radiotherapy prescription

These modifications did not have an impact on the study design.

2B.3.7 Summary measures

The equation and the methods employed to calculate the PTV margin for head and neck patients treated with VMAT were used as outcome measures.

To explain the findings of the studies, a narrative synthesis approach was used. This approach relied primarily on the use of text to summarise and explain findings (Ryan, 2013). Relevant data from the studies were extracted and key characteristics of the studies were recorded and presented in a tabular form.

2B.3.8 Synthesis method

While the author recognised that meta-analysis enables the combination of results of individual studies and may answer questions that were not posed by separate studies (Tang, Caudy and Taxman, 2013), a narrative synthesis approach was opted since the clinical, methodological and statistical sources were too diverse to be able to perform a meta-analysis (Snilstveit, Oliver and Vojtkova, 2012).

2B.4 Results and discussion

2B.4.1 Study selection

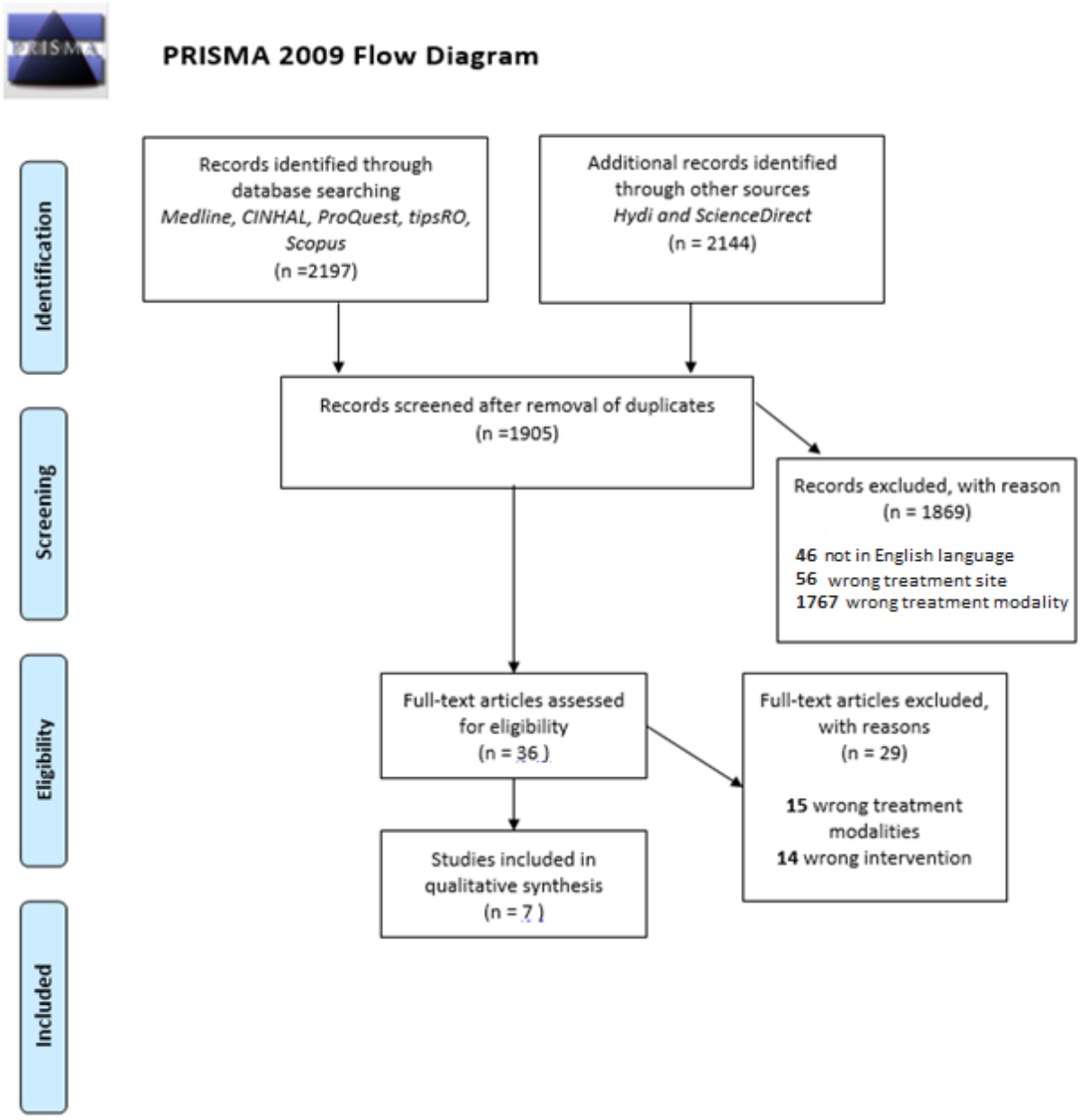


Figure 2B.2: Prisma 2009 Flow Chart *(*Adapted from*: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses).*

Using predefined sets of keywords, the search strategy aimed to maximise the amount of literature recorded by databases and other literature sources. The use of publication period (after 2007), full-text, and human species filters were applied.

Figure 2B.2. shows the PRISMA flow chart, where a total of 4341 articles were found from the search strategy. These articles were exported to a reference management software to check for duplicates. De-duplication of studies was necessary to ensure a reliable and valid pool of studies in the systematic review (Kwon et al., 2015). De-duplication was also beneficial for the first phase of the review, by reducing the workload of article selection based on the title and abstract (Bramer et al., 2016). The reference management software was able to detect and merge 2436 duplicates.

After removing duplicated articles, articles were screened for eligibility and at the end, a total of seven relevant studies were found.

2B.4.2 Study characteristics

A summary of the study characteristics is presented in table 2B.3

Table 2B.3. Summary of Characteristics of Studies

Author, Year and Country	Study Design	Head and neck region	Imaging protocol	Immobilisation device	PTV formula (errors considered)	PTV margin result
Yin et al., 2013 Southern China	Prospective Observational Analytical and Cross-sectional	Nasopharynx	Daily CBCT	5-point TP mask HR not specified	VHMF (inter- and intra-fraction errors)	***Without CBCT correction*** LR= 4.1 mm SI = 3.4 mm AP = 3.5 mm ***With CBCT correction*** LR= 1.7 mm SI = 2.2 mm AP = 2.2 mm
Oh et al., 2014 South Korea	Retrospective Observational Analytical and Cross-sectional	Not Specified	Daily CBCT	5-point TP mask Individual HR	VHMF (inter-fraction errors)	LR = 3.3 mm SI = 2.8 mm AP = 3.7 mm
Anjanappa et al., 2017 India	Retrospective Observational Analytical and Cross-sectional	Nasopharynx	Daily 2D KV imaging (KV images taken on alternate days were reviewed)	4-point TP mask HR not specified	VHMF (inter-fraction errors)	***Clivus level*** LR= 4.0 mm SI=3.2 mm AP=4.4 mm ***C3 level*** LR= 5.0 mm SI = 4.4 mm AP = 5.5 mm ***C6 level*** LR = 6.9 mm SI = 4.4 mm AP = 6.4 mm

Norfadilah et al., 2017 Malaysia	Prospective Observational Analytical and Cross-sectional	Oral cancer	Daily CBCT	5-point TP mask Mouth Bite HR not specified	VHMF (inter-fraction errors)	***HFW mouth bite*** LR =3.1 mm SI = 2.2 mm AP = 0.8 mm ***SYR*** LR= 3.8 mm SI = 6.2 mm AP= 5.1 mm
Bruijnen et al., 2019 Netherlands	Prospective Observational Analytical and Cross-sectional	Nasopharynx Oropharynx Larynx	eNAL	5-point TP mask Individual HR	VHMF (inter and intra-fraction errors)	***Nasopharynx*** S=2.8 mm I = 2.8 mm A = 2.8 mm P = 2.8 mm ***Oropharynx*** S = 3.0 mm I = 3.1 mm A = 3.0 mm P = 3.0 mm ***Larynx*** S = 4.0 mm I =3.6 mm A = 3.1 mm P = 3.1 mm ***Combined*** S = 3.3 mm I = 3.2 mm A = 3.0 mm P = 3.0 mm
Deb et al., 2019 India	Retrospective Observational Analytical and Cross-sectional	not specified	Daily imaging (eNAL for CBCT & remaining days with 2D PI)	TP mask with shoulder retraction Standard HR	VHMF (inter-fraction errors)	LR= 5.6 mm SI = 6.1 mm AP = 4.7 mm

Kukolowicz et al., 2020 Poland	Retrospective Observational Case-control	Nasopharynx and Larynx	Daily EPID	5-point TP mask Standard HR	VHMF (inter-fractional errors)	***Prior NAL protocol*** AP = 4.0 mm, SI = 6.0 mm LR = 4.0 mm ***NAL protocol*** AP = 3.0 mm SI = 2.2 mm LR = 3.0 mm

N = sample number; S = Superior; I = Inferior; A = Anterior; P = Posterior; LR = Left-Right; SI = Superior-inferior; AP = Anterior-Posterior.

C3 = Cervical Spine level 3; C6 = Cervical Spine level 6.

TP = Thermoplastic; HR = Head Rest; HFW = HeadFIX® mouthpiece; SYR = 10 ml/cc syringe barrel.

EPID = Electronic Portal Imaging Device; CBCT = Cone Beam Computed Tomography; PI = Portal Imaging.eNAL= Extended Non-action level protocol (imaging on first three fractions, followed with once weekly imaging); NAL= No action level protocol (imaging on first three fractions).

VHMF = Van Herk's Margin Formula

Table 2B.3. shows that in terms of key characteristics, there is a lot of variation between the studies. The aims and methodologies of the studies in the review are also heterogeneous. These inconsistencies made it difficult to compare studies.

2B.4.3 Risk of bias across studies

Study outcomes may be influenced by different research methods, and this might introduce biases in the results (Charrois, 2015).

Risk of bias in a systematic review is minimised in a good quality study design of the primary studies (Boutron, Ravaoud and Mohen, 2012). According to PRISMA guidelines the eligibility of potential studies should at least be checked independently by two reviewers in accordance with the eligibility criteria, to ensure a good inter-rater agreement of the selected articles (Jain and Sandhya, 2016), and this procedure was performed in this study. However, to resolve discrepancies on data searching, the study ideally should have had a third party to resolve discrepancies through discussion

(Eden et al., 2011). The researcher also performed a backward snowball technique, where references of selected articles were reviewed in search for other relevant literature to minimise selection bias of the literature (Wohlin, 2014).

Since the reviewed studies had an exclusive reliance on English-language research, this limitation may not be a representation of all the evidential research. Therefore, the exclusion of 46 articles based on language could have introduced a language bias (Konno et al., 2020).

Publication bias refers to the tendency that certain studies are more likely to be published than others since they show positive effects and would be easier to find (Bigby, 2014; Borges de Almeida and Garcia de Goulart, 2017). Sponsorship of studies will increase the risk for this type of bias (Wareham et al., 2017). An explicit thorough search on databases and other literature was performed to ensure that most data relevant to the review question was identified, however not all studies were peer-reviewed since grey literature was also sought in this review to expand the search and this could have an impact on this type of bias.

Several tools and methods are available to evaluate internal validity of the selected studies (Sullivan, 2011). For this study, the Joanna Briggs Institute (JBI) critical appraisal tools were thought to be suitable to assess individual bias in observational studies since these tools can appraise both analytical cross-sectional studies and case control studies (Moola et al., 2017). The JBI tools are particularly used in evaluation of research related to health care (Ma et al., 2020) and unlike most critical appraisal and bias tool, the JBI tool is not restricted to randomised control studies, this was essential as none of the studies specified whether random selection of participants was performed. These tools addressed the study design, quality, and provided an individual assessment for bias within each study (Moola et al., 2017).

The JBI tools have been developed by the JBI and collaborators and were designed for use in systematic reviews. Following extensive peer review, these tools were also approved by the JBI Scientific Committee (Moola et al., 2017). They are thought to be more sensitive to the validity of the evaluation tool when compared to others such as the Critical Appraisal Skill Programme (CASP) tool (Hannes, Lockwood, and Pearson, 2010).

The JBI Critical Appraisal tools for use in JBI Systematic Reviews Checklist for Analytical Cross-Sectional Studies was concerned with the following factors in the selected studies: clearly defined inclusion and exclusion criteria in a study, clear description of the population of interest, confounding factors, selection bias, reliability and validity of exposed measures and outcome measures, and statistical analysis (Moola et al., 2017). These sources of bias can threaten the validity of the results of the studies (Viswanathan et al., 2013). Six studies in this review were eligible for this tool.

The Joanna Briggs Institute critical appraisal tools for use in JBI Systematic Reviews Checklist for Case Control Studies assessed different criteria than that of the analytical cross-sectional studies. This tool was concerned with the following: comparison of groups, appropriate matching of cases and control, similar criteria for identifying cases and control, reliability and validity of exposed measures and outcomes, similarity in measurement of exposures for cases and control, confounding factors, exposure period and statistical analysis (Moola et al., 2017). Only one study in this review was found to be eligible for this tool.

The following were identified using the JBI Critical appraisal tools (**Appendix E**):

- All selected studies, except for Deb et al. study (2019), specified the inclusion and exclusion criteria in detail.
- All studies except for Deb et al. (2019) study provided sufficient detail on patients' characteristics.
- Not all studies measured the exposure in a valid and reliable way since in some studies inter-observer variability in image matching was not assessed, manual image registration was not performed, and some studies failed to identify how set-up errors were recorded.
- Selection of participants was related to both the intervention and outcomes. Participant selection bias was present in some of the studies since there were variation in the patient's characteristics and, at times, lack of information on these characteristics that have a negative effect on the validity of the results.
- Outcome measures were not always measured in a valid and reliable way. Some of the studies measured PTV margin based on inter-fractional

translational errors only and did not consider intra-fractional errors, rotation factors, organ motion, and variation in target volume delineation.

- Some of the reviewed studies had a small sample size which rendered the results to be unreliable.

Table 2B.4. Describes the outcome of the evaluated studies when using the JBI critical appraisal tool.

Table 2B.4. Outcome of the evaluation of the studies

Study	Outcome of the evaluation
Oh et al. (2014)	Very strong
Bruijnen et al. (2019)	Strong
Yin et al. (2013)	Moderate
Norfadilah et al. (2017)	Weak
Deb et al. (2019)	Weak
Anjanappa et al. (2017)	Moderate
Kukolowicz et al. (2020)	Strong

The overall evaluation of Norfadilah et al. (2017) and Deb et al. (2019) studies was determined to be weak using the JBI critical appraisal tool. The quality of information obtained from the articles was insufficient to determine the validity and reliability of the studies. Despite these studies' limitations, it was decided to keep them for further discussion in this review.

2B.4.4 Results of synthesis

Synthesis of results is a key element in a systematic review. It is the process that pools together the findings of the included studies in the review to draw conclusions based on the evidence (Verbeek, Ruotsalainen and Hoving, 2012). According to the Cochrane Consumers and Communication Review (Ryan, 2013), synthesis of data should avoid bias by being transparent and rigorous, and the methods employed should be justified and followed systematically. The authors also mentioned that narrative synthesis should ideally explore patterns in the data and include an

investigation of the differences and similarities between the findings of the studies in the review in a systematic way, with a possible logical explanation for the results of the included studies (Ryan, 2013).

The overall sample size of the seven reviewed studies was 217 patients, of which 60 were treated to the oral cavity, 47 to the nasopharynx, 43 to the oropharynx, 29 to the larynx, and 30 to either the larynx or nasopharynx, with the author not specifying the exact number of patients treated in each region.

Confounding variables

Confounding variables, which may cause confounding bias results, are one of the factors that influence a study's internal validity (Haneuse, 2016). This type of bias can cause an effect to be overestimated or underestimated (Skelly, Dettori and Brodt, 2012). The JBI tool was capable of detecting such bias.

Table 2B.5 shows the confounding variables and the stated strategy for dealing with them, as stated by the authors in the selected studies related to the head and neck region.

Table 2B.5. Confounding variables and strategies

Study	Confounding Variables	Strategy
Yin et al. (2013)	1. Weight loss	1. Examined relationship between weight loss and set-up errors and analysed the time trend of weight loss.
	2. Tumour shrinkage	2. Not specified
	3. Uncertainty in image registration	3. Not specified
	4. Not able to adjust rotational errors	4. Not specified
Oh et al. (2014)	1. Intra-fractional movement	1. Not specified
	2. Curved external anatomy	2. Not specified
	3. Loosening of fixation mask due to weight loss or tightening of mask due to swelling	3. Thermoplastic mask was remade if considerable discrepancies occurred. Rescanning and replanning were performed when necessary to reduce set-up errors

Anjanappa et al. (2017)	1. Rotation	1. Not specified
	2. Weight loss	2. Not specified
	3. Quality of kV portal imaging and DRR imaging	3. Not specified
	4. Difficulty in imaging due to superimposition of structures	4. Not specified
Norfadilah et al. (2017)	Nothing mentioned	Not applicable
Bruijnen et al. (2019)	1. Accuracy of deformable image registration	1. Not specified
	2. Left-right motion affecting image registration	2. Image acquisition of 10 mm was used. Study referred to other literature that reported that when this form of acquisition is used, the motion is minimal.
	3. Treatment modality	3. VMAT PTV margin was calculated by halving the tumour shift between the two cine MR scans
	4. Persistent tumour motion over a long period of time	4. Not specified
Deb et al. (2019)	1. Rotation	1. Not specified
	2. Weight loss	2. Not specified
	3. Tumour shrinkage	3. Not specified
Kukolowicz et al. (2020)	1. Rotation	1. The study mentions that it was very seldom to observe rotations larger than 1 degree, therefore the rotational factor was negligible
	2. Variation in treatment modality (VMAT and IMRT)	2. Not specified
	3. Anatomical changes not visible on portal imaging	3. Not specified

All the selected studies listed confounding variables in their studies. Nevertheless, there could be other confounding variables that have not been identified by the authors, such as imaging parameters, weight loss, tumour shrinkage, variation in treatment region, and inter-observer variation in image registrations. Strategies to deal with confounding factors were not always mentioned in the studies and these confounding factors could have influenced the result of the study.

PICOS elements of studies in review

To synthesise the data, studies were compared in terms of their PICOS elements. As the underlying questions for the studies differed, the studies were compared based on having homogeneous PICOS elements (McKenzie et al., 2019), as shown in table 2B.6. Studies were also compared through similarity of key characteristics that have an impact on the PTV margin calculation and result. These characteristics refer to the type of imaging protocols, head and neck region and immobilisation devices (Merlotti et al., 2014; Minniti et al., 2016; Winey and Bussiére, 2014; Marnouche et al., 2019).

Table 2B.6. PICOS elements of the studies

Study	Population	Intervention	Comparative Intervention	Outcome	Study Design
Yin et al. (2013)	Patients treated with VMAT to the Nasopharynx	Evaluated inter-fraction and intra-fraction errors	N/A	Determine the set-up errors and appropriate PTV margin	Cross-sectional, prospective, and quantitative
Oh et al. (2014)	Patients treated with VMAT to the Head and Neck, Brain, Prostate, Thorax and Abdomen	Assessed set-up errors and calculated the PTV margin	Compared set-up errors and calculated PTV margin for various tumour sites	Reduce set-up errors and optimise PTV margin	Cross-sectional, retrospective, and quantitative
Anjanappa et al. (2017)	Patients treated with VMAT or IMRT to the Nasopharynx	Evaluated inter-fraction set-up errors and derived the PTV margin	Compared the systematic error and random error of C3, C6 and Clivus	Determine the PTV margin of the Nasopharynx at three different levels	Cross-sectional, retrospective, and quantitative
Norfadilah et al. (2017)	Oral cancer patients receiving treatment with VMAT	Evaluated inter-fraction set-up errors for two different immobilisation devices	Compared HeadFIX® mouthpiece moulded with wax with syringe barrel	Determine which immobilisation device produces the least set-up errors	Cross-sectional, prospective, and quantitative
Bruijnen et al. (2019)	Patients treated with IMRT and VMAT to the Nasopharynx Oropharynx Larynx	Quantified intra-fraction motion and assessed set-up errors	N/A	Determine population based PTV margin	Cross-sectional, prospective, and quantitative
Deb et al. (2019)	Head and neck patients treated with VMAT	Assessed set-up errors and derived the PTV margin	N/A	Determine the optimal PTV margin	Cross-sectional, retrospective, and quantitative
Kukolowicz et al. (2020)	Head and neck patients treated with VMAT or IMRT	Evaluated the impact of NAL imaging protocol on treatment time and set-up errors	Compared the daily imaging protocol with NAL protocol	Reduce treatment time with an effective set-up control	Case-control, both prospective and retrospective, and quantitative

The reviewed studies agreed in terms of certain PICOS elements. The population in the review were all patients treated to the head and neck with VMAT. Some studies, however, analysed both VMAT and IMRT patients, therefore for these studies there was a variation in the treatment modality (Anjanappa et al. 2017; Bruijnen et al. 2019; Kukolowicz et al. 2020). The treatment modality main effect was the duration of the treatment which have an impact on intra-fraction motion. The intervention of the study involved the calculation of PTV margin. The studies also used a quantitative and cross-sectional design, with the exception of Kukolowicz et al. (2020) which was a case-control study.

Variations of the reviewed studies laid on the intervention, comparison, and outcomes of the studies. Oh et al. (2014), Anjanappa et al. (2017) and Deb et al. (2019) evaluated set-up errors and calculated the PTV margin whilst Norfadilah et al. (2017) focused on assessing set-up errors obtained from two sets of mouthpieces. These four studies had similar interventions since they only considered the inter-fraction motion in the PTV calculation, however, they varied in research outcomes.

Bruijnen et al. (2019) and Yin et al. (2013) also had similar interventions since these authors were interested in evaluation and quantification of inter-fraction and intra-fraction errors. These studies also had similar outcomes of determining the appropriate PTV margin size, however, they varied in terms of imaging procedures.

Kukolowicz et al. (2020) looked at the effect of the non-action level (NAL) imaging protocol on the PTV margin and treatment times. Due of the differences in study design, intervention, comparative intervention, and outcomes, the results of this study could not be compared to those of other studies.

Since they had common intervention, comparative intervention, result, and study design, Bruijnen et al. (2019) and Yin et al. (2013) were considered to have the most comparable PICOS elements.

PTV margin methods

One of the outcomes for data synthesis was to determine the different methods adopted in research studies to calculate the PTV margin. Table 2B.7. demonstrates the methods opted by the studies included in the review to calculate this margin.

Table 2B.7. PTV margin methods

Study	Target Delineation	Intra-fraction errors	Set-up errors	PTV formula
Yin et al. (2013)	x	✓	✓	$PTV = 2.5\sqrt{(\sum inter\text{-}fraction^2+\sum intra\text{-}fraction^2)} + 0.7\sqrt{(\sigma inter\text{-}fraction^2+\sigma intra\text{-}fraction^2)}$
Oh et al. (2014)	x	x	✓	$PTV= 2.5\sum+0.7\sigma$
Anjanappa et al. (2017)	x	x	✓	$PTV=2.5\Sigma+0.7\sigma$
Norfadilah et al. (2017)	x	x	✓	$PTV=2.5\Sigma+0.7\sigma$
Bruijnen et al. (2019)	x	✓	✓	$PTV = 2.5\sqrt{(\sum motion^2+\sum setup^2)} + 0.7\sqrt{(\sigma motion^2+\sigma setup^2)}$
Deb et al. (2019)	x	x	✓	$PTV=2.5\Sigma+0.7\sigma$
Kukolowicz et al. (2020)	x	x	✓	$PTV=2.5\Sigma+0.7\sigma$

All the studies in the review applied the van Herk formula to calculate the PTV margin. Bruijnen et al. (2019) and Yin et al. (2013) considered intra-fractional motion as well as set-up errors to calculate the margin. The other studies (Oh et al., 2014; Anjanappa et al., 2017; Norfadilah et al., 2017; Deb et al., 2019; Kukolowicz et al., 2020) evaluated just the set-up errors to derive the margins. None of the studies considered target delineation as part of the margin recipe.

Not all studies took rotational errors into account when calculating set-up errors. The rotational factor was not considered in Anjanappa et al. (2017) study since the review

made use of 2D KV planar imaging. Bruijnen et al. (2019) and Deb et al. (2019) made use of 3D imaging with the use of CBCT, however they also failed to account for rotational errors.

PTV margin results

Figure 2B.3 shows the variation in the PTV margin results from the studies in the review. The bar graph shows that the maximum margin result was obtained in Anjanappa et al. (2017) study whilst the minimum PTV margin result obtained was that of Yin et al. (2013) study.

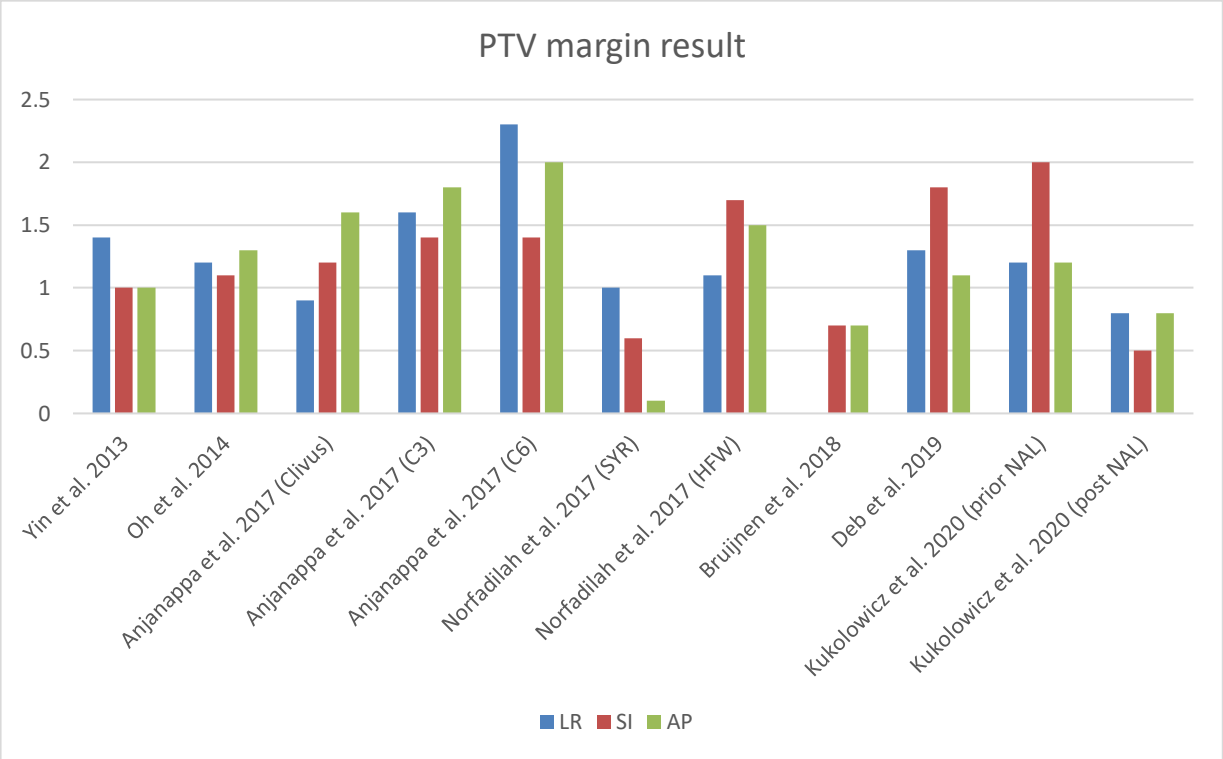


Figure 2B.3: PTV Margin Result

The mean result of the PTV margin size in the LR, SI and AP dimensions from the seven studies was 3.87 mm, 3.58 mm, and 3.91 mm, respectively, and the SD for the PTV margin in the three axes was: LR= 1.48 mm, SI = 1.61 mm, AP = 1.59 mm. The SD in the LR direction was the smallest, but it is worth noting that Bruijnen et al. (2019) did not measure the margin in the LR direction. The mean margin size and SD in the LR, SI, and AP directions of the studies reviewed are comparable.

PTV margin was calculated by Anjanappa et al. (2017) at the clivus, C3, and C6 levels. The C6 level had the highest PTV margin result, with a value of 6.9 mm in the LR dimension, 4.4 mm in the SI dimension, and 6.4 mm in the AP dimension.

Yin et al's. (2013) calculated the PTV margin for patients treated to the nasopharynx with CBCT correction to be: LR = 1.7 mm, SI = 2.2 mm, and AP = 2.2 mm. The margin size is significantly smaller when compared to the other studies in the review, but this margin was calculated using data obtained after CBCT correction. This margin was only appropriate if the daily imaging protocol was followed. Prior to CBCT correction, the PTV margin with inter- and intra-fraction errors was 4.1 mm, 3.4 mm, and 5.5 mm in the LR, SI, and AP directions, respectively.

Inter-fraction errors

Figure 2B.4. presents the population systematic set-up errors obtained by the studies in the review.

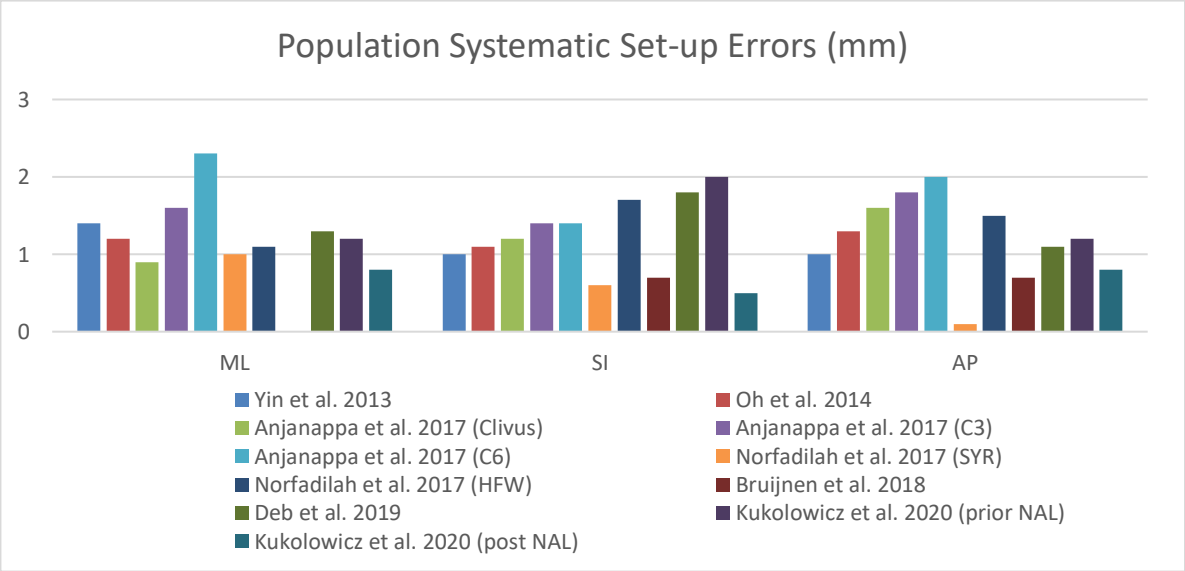


Figure 2B.4: Population Systematic Errors

The calculated SD of the population systematic errors of the reviewed studies were 0.4 mm, 0.5 mm, and 0.5 mm in the LR, SI, and AP direction, respectively. These values indicate that the reviewed studies obtained similar results for population systematic errors. The highest value of population systematic error was obtained in Anjanappa et al. (2017) study with a value of 2.3 mm in the LR direction.

Figure 2B.5. presents the population random set-up errors in all the reviewed studies.

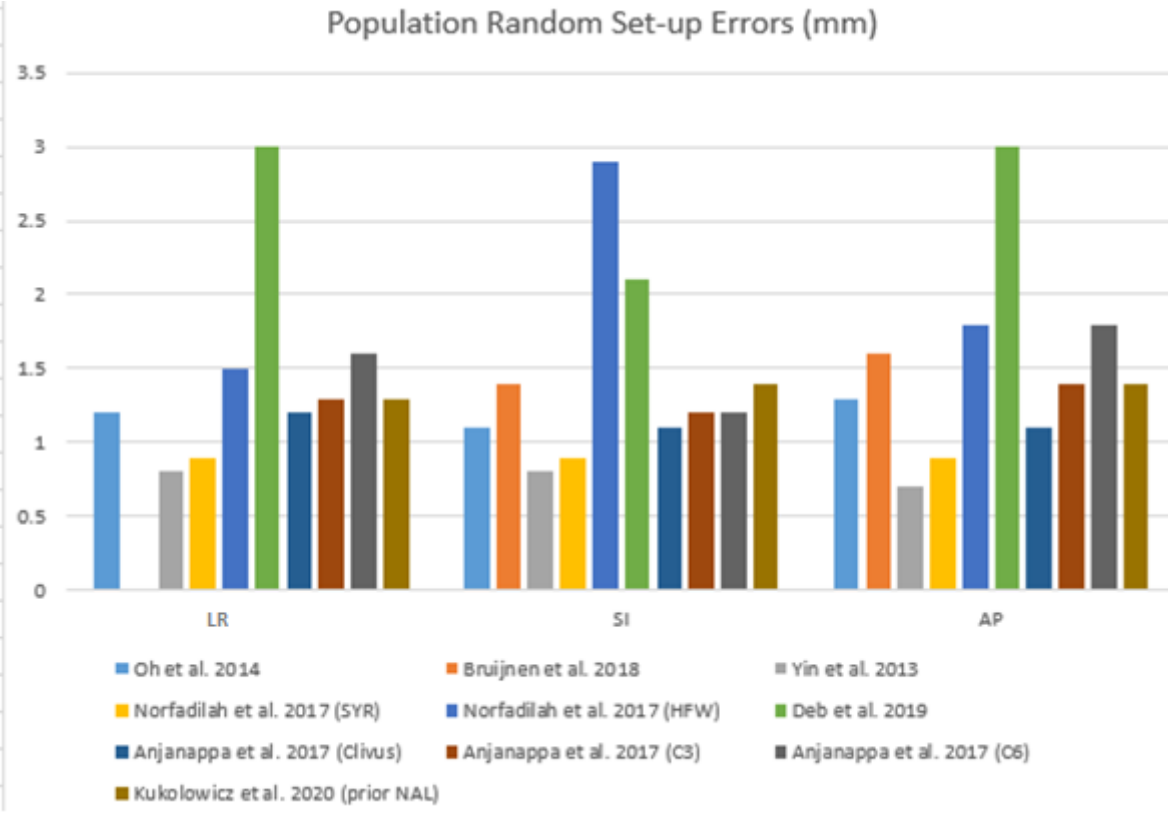


Figure 2B.5: Population Random Errors

The SD for the population random errors in the reviewed studies was slightly higher than that of the population systematic errors, with values of 0.6 mm in the LR, SI and AP direction.

Intra-fraction errors

According to Bruijnen et al. (2019), the population systematic intra-fraction errors were 0.9 mm and 0.7 mm in the superior and inferior directions, respectively, and 0.6 mm in both the anterior and posterior directions. The population random errors for motion were 0.8 mm and 0.7 mm, respectively, in the superior and inferior directions, and 0.5 mm in both the anterior and posterior directions. When the authors incorporated the

tumour motion to the margin recipe, the CTV to PTV margin expanded with, "*0.2 mm for nasopharyngeal tumours, with 0.6 mm for oropharyngeal tumours and with 1.7 mm for laryngeal tumours"*, (Bruijnen et al., 2019, p.87).

In the study by Yin et al. (2013), the population random errors varied from 0.5 mm to 0.6 mm and the intra-fraction population systematic errors ranged from 0.2 mm to 0.4 mm.

Rotational errors

Oh et al. (2014) study compared rotational errors in different anatomical regions and the rotational distribution was comparable for all the anatomical sites with the prostate area having the least rotational errors. The rotational errors for head and neck region were small since all the patients had rotational errors below 3º. Since the department had a 4º freedom couch, the yaw rotational set-up errors could have been corrected. Pitch and roll rotational set-up errors could not be corrected due to the couch limitation. In Norfadilah et al. (2017) study, rotational errors were compared and calculated for the two tongue immobilisation devices. Average rotational errors results were that of 0.00°±0.65° and 0.34°±0.59°, respectively for HFW and SYR. In Yin et al. (2013) study, the number of fractions that exceeded 2° for pitch, roll and yaw, respectively where 1 (1.1%), 0 (0.0%) and 0 (0.0%), and intra-fraction random errors were significantly smaller than inter-fraction rotation errors. Kukolowicz et al. (2020) also reported a low find in rotational errors with rotations larger than one degree seldom observed, therefore these errors were not analysed. In this study 2D imaging was performed, therefore this limits rotational results from multiple perspective since one cannot view the third dimension (Høyer et al., 2011).

2B.5 Discussion

2B.5.1 Summary of evidence

The data were analysed for trends and patterns in the method and results obtained from the studies in the review.

Based on the data collected, only seven studies were found to report patients receiving treatment to the head and neck region using VMAT. This low find in studies is in accordance with a report issued by the European Society Radiation Oncology (ESTRO), where IMRT was the most reported treatment modality for patients being treated to the head and neck region (48.5%), followed by 3D conformal radiotherapy (27.9%). VMAT procedures were reported to be less common (21.2%), whilst the least common modality was the 2D technique (Leech et al., 2017).

All the patients in the reviewed studies were treated to the head and neck region using VMAT and had their PTV margin calculated.

PTV margin equations

In radiotherapy, the procedure used to calculate the PTV margin is a point of contention. The van Herk formula is a commonly used technique for calculating PTV margins, and it was used in all seven studies, however there was no explanation for why the formula was chosen.

Namysł-Kaletka, Tukiendorf and Wydmański (2015) used three formulas to determine PTV margin outcomes for gastric cancer patients based on set-up errors: Van Herk, Stroom, and ICRU. The margin results were: 9 mm, 7 mm and 6 mm in the LR direction, 16 mm, 14 mm, and 11 mm in the SI direction as well as 8 mm, 7 mm, and 5 mm in AP direction, respectively for Van Herk, Stroom and ICRU 62 formula. This means that PTV margin results vary depending on the PTV margin formula used. The report did not specify the best method to calculate the margin however the study ultimately chose the van Herk formula since the percentage of shifts beyond the specified margin was the lowest when using this formula (Namysł-Kaletka, Tukiendorf and Wydmański, 2015).

Since the Van Herk Margin Formula (VHMF) eliminates tumour shape variance and rotational errors, it can be used as a lower limit for safely delivering radiotherapy (Van Herk et al., 2000). The CTV is spherical in shape, the tissue is homogeneous, the conformal beam penumbra is infinite, and the number of fractions is infinite, according to this formula (Witte et al, 2017; Yoda, 2017). The formula also ensures that 90% of patients receive at least 95% of the recommended dose in the CTV (Van Herk et al., 2000). As a result, the van Herk Formula appears to be adequate for calculating PTV margin in patients with head and neck cancer.

PTV margin size

The PTV margin results of all the studies in the review vary, confirming the need for departments to calculate their own specific PTV margin.

There was a large discrepancy in margin size for Anjanappa et al. (2017) and Yin et al. (2013). Both these studies evaluated patients who were treated to the nasopharyngeal region, however they varied in the use of imaging modality, outcomes, and interventions. Yin et al. (2013) assessed intra-fraction errors as well as set-up errors, whilst Anjanappa et al. (2017) evaluated just the inter-fraction errors to determine the PTV margin result. Intra-fractional errors tend to generate a greater PTV margin (Cacicedo et al., 2015), but the analysis that did not test intra-fraction errors had the smallest PTV margin of the two. The imaging modalities may be to blame for the difference in the margin result. The chosen modality has an impact on the observed set-up errors. CBCT allows for better observation of the region of interest, therefore it should be chosen modality. The CTV–PTV margins can be reduced if CBCT is used (Martins, Couto and Barbosa, 2016).

The effect of regular imaging on margin size was demonstrated in the study by Yin et al. (2013). Since the margin size was calculated based on the set-up errors obtained after CBCT correction, the resulting margin size for nasopharyngeal patients with CBCT correction was small in comparison to other studies in the review. Gupta et al. (2018) conducted a study to determine the effect of imaging frequency on PTV margin size in prostate cancer patients. This study discovered that as the frequency of image guidance decreased, the mean systematic error and SD of systematic error increased,

implying that the PTV margin would need to be increased as well to ensure adequate coverage of the CTV. Results revealed that for every 15% reduction in image guidance, a 5% increase of geographical miss could occur (Gupta et al., 2018).

Kukolowicz et al. (2020) evaluated the non-action level (NAL) procedure for patients receiving care in the head and neck area to see how effective it was at minimising set-up errors. In this study, daily online correction was found to be slightly superior to the NAL protocol. However, because rotational errors and anatomical changes were not visible on portal images, the study was unable to analyse and quantify them. According to Marnouche et al. (2019), daily online image matching was the advised image registration protocol for patients treated with tight PTV margins such as IMRT and VMAT as all set-up corrections will be adjusted daily prior treatment. Daily online CBCT usage was also justified in Oh et al. (2014) study by stating that the procedure were used to verify set-up positioning, location of the target, and assessing tumour shrinkage. Therefore, it is evident that daily imaging should be employed for tighter PTV margins.

According to Anjanappa et al. (2017), the lower neck region necessitates a larger margin in the AP and LR directions. Another study, conducted by Cheo et al. (2015), confirmed this statement by evaluating set-up errors at different levels of the neck. The largest displacement was noted in the LR direction at the C7 level of the neck, as the PTV margin prior to CBCT correction was found to be 6.52 mm. The PTV margin value in the AP and SI directions was discovered to be 2.72 mm and 4.70 mm, respectively. According to this analysis, the largest error in the SI direction was more common in upper regions of the neck. Cheo et al. (2015) study also found that the mean 3D displacement prior correction was mostly prominent in the C7 level when compared with the clivus and C4 levels. In the reviewed studies there were insufficient data to confirm these statements as the LR margin was not measured in the larynx and oropharynx. With missing data and variation in the quality evaluation of the studies, analysis of PTV margins in the three axes for different head and neck regions were not possible.

The most similar PTV margin results obtained were those by Oh et al. (2014) and Kukolowicz et al. (2020) (post NAL protocol) since these studies obtained a margin which varied with each other by a few millimetres. These studies, however varied in

imaging procedures, immobilisation devices and outcomes. Therefore, the reason for the correlation in findings is unknown.

Inter-fraction errors

According to Goyal and Kataria (2014), treatment to head and neck experience the least set-up errors when compared to other anatomical sites. The use of effective immobilisation devices such as the 5-point thermoplastic (TP) mask aid in minimising set-up uncertainties (Mandair et al., 2018).

In the reviewed studies that assessed both intra- and inter-fraction errors (Yin et al., 2013; Bruijnen et al., 2019), inter-fraction errors results were more than intra-fraction errors, and this indicates that the immobilisation devices were more effective in maintain the set-up position of the patient rather than reproducing it.

According to ICRU 62 (1999) report, numerous authors base their margin calculations on systematic and random errors. This is still evident in the reviewed studies as all studies measured the systematic and random errors to derive the PTV margin. The variation in the margin calculation laid solely on whether intra-fraction errors were being analysed. All the studies in the review assessed inter-fraction errors from the set-up errors obtained from the imaging software.

With daily imaging, the population systematic and random errors are corrected prior treatment, however in studies where daily imaging was not performed, random errors could not be compensated on the radiotherapy fractions without image guidance.

Deb et al. (2019) obtained the highest population random errors, and because ten CBCT images were performed for each patient in this study, there were days when the random errors could not be compensated for. When compared to the other studies in the review, the PTV margin was larger due to the population random errors.

Image guided radiation therapy, according to Van Kranen et al. (2013), allows adjustments of patient set-up errors by correcting such errors with an opposite shift of the treatment couch. The majority of the seven studies included in the review used on-line imaging verification, in which image acquisition, verification, and correction were performed prior to treatment delivery, with the goal of reducing random and systematic

errors. Off-line imaging was only performed in Kukolowicz et al. (2020) study. With off-line imaging, random errors cannot be corrected since the images would be verified after treatment delivery, and the aim would be to adjust for systematic errors (Goyal and Kataria, 2014).

Intra-fraction errors

In comparison to other areas, such as the gastrointestinal region and genitourinary sites, the head and neck region is thought to have the least amount of internal organ motion (Lu et al., 2012). Even though it is a region with low organ motion, swallowing during treatment may have an impact on treatment delivery (Merlotti et al., 2014).

Yin et al. (2013) and Bruijnen et al. (2019) managed to assess intra-fraction errors. Intra-fraction errors are random and are related to patient movement and internal organ motion during treatment (Michalski et al., 2012). As stated by Yin et al. (2013), longer treatment times are at a higher risk of intra-fraction motion. Few studies have assessed intra-fraction results for VMAT treatment and therefore the need to calculate this type of error was raised in Yin et al. (2013) study, since the treatment duration of VMAT is shorter when compared to other treatment modalities such as IMRT. This study obtained values for intra-fraction errors from CBCT imaging performed after treatment, while Bruijnen et al. (2019) assess organ motion and intra-fraction errors from 2D cine MRI and deformable image registration.

Bruijnen et al. (2019) was interested in investigating the contribution of respiratory tumour motion, tumour motion due to swallowing, tongue motion, and treatment set-up errors on the population based PTV margin. According to Bruijnen et al. (2019), intra-fraction movements should be quantified to determine margins based on a specific population, or personalised margins to account for internal motion. The results showed that in the head and neck region the maximum tumour motion is mostly pronounced in the larynx (Bruijnen et al., 2019). These results are comparable to a study performed by Gurney-Champion et al. (2018) where the authors assessed intra-fractional tumour motion from magnetic resonance imaging data of patients treated to the head and neck region and found that tumour motion was significantly larger for tumours in the larynx and hypopharynx than for tumours in the oropharynx (Gurney-

Champion et al., 2018). Bruijnen et al. (2019) however assessed tumour motion prior the radiotherapy treatment delivery, therefore dysphagia, which is a common side-effect in patients receiving treatment to larynx and oropharynx, was not taken into consideration. This limitation was recognised by the authors by mentioning that tumour motion over a period is unclear since incidence of tumour motion could vary over the course of treatment. In fact, it was found that laryngeal elevation is reduced in patients receiving radiotherapy to the head and neck region, and this is attributed to deficits in the posterior muscular sling (Pearson et al., 2016). Other limitations in Bruijnen et al. (2019) study was that the LR motion was not assessed, and tumour motion was processed as a linear trend.

In Yin et al. (2013) study, intra-fraction motion was assessed by acquiring a CBCT prior to treatment to calculate the initial inter-fraction errors. Post-correction CBCT was then taken to calculate the residual errors. A final CBCT was taken after treatment and the difference between the post-treatment (final) CBCT and post-correction CBCT was used to calculate the intra-fraction errors. The sample in Yin et al. (2013) study consisted of patients receiving treatment to the nasopharynx. Organ motion in Yin et al. (2013) study was not assessed, instead intra-fraction errors were based on patient motion during treatment.

Intra-fraction systematic errors, according to Duan et al. (2020), increases with time. In contrast, the findings of Yin et al. (2013) revealed that there was no significant relationship between intra-fraction errors and treatment delivery time. Yin et al. (2013) justified the obtained result by stating that there was limited data available for analysis and that they were unable to obtain statistically significant results due to the narrow range of treatment time (5.6 – 9.4 min).

Target volume delineation

According to Jameson et al. (2010), the most significant contributor to uncertainty in radiation treatment planning is inter-observer variability in anatomical contouring.

Delineation of the target volume is known for its geometric uncertainty since the procedure relies on the clinical ability of individual doctors, and this task can lead to

inter-observer variability, especially when the differentiation of the tumour from unaffected regions cannot be easily distinguished. Also, accuracy in target delineation is dependent on the quality of the imaging data (Apolle et al., 2019).

According to Stroom et al. (2014) and Suzuki et al. (2012) calculation of the PTV margin should ideally include an assessment of variation in target volume delineation, however none of the studies in the review evaluated this observer variation. In fact, Vinod et al. (2016) published a systematic review on inter-observer variation in volume delineation and this study discovered that the inter-observer factor in volume delineation is commonly ignored in studies that investigated PTV margin calculation. They identified 119 papers on target volume delineation uncertainties and all these studies showed the presence of variability between observers when it came to target volume delineation. According to Segedin and Petric (2016), one of the largest inter-observer variabilities in target delineation is that of oropharyngeal cancer. If one assumes that target delineation, organ motion and set-up errors are independent of each with a Gaussian distribution, the systematic and random errors of the mentioned factors can be combined in a quadratic sum to derive a PTV margin (Kalyankuppam Selvaraj, 2013).

Rotational errors

Even though rotational errors were not considered in van Herk's formula, four studies in the review assessed this error. According to Chang (2017), rotational errors should not be ignored in high precision treatments such as Stereotactic Radiotherapy, especially when there is a large distance between the isocentre and the target. These types of errors might cause dosimetric inaccuracies during clinical treatment and in most clinical departments are not corrected due to couch limitations (Zhang et al., 2013).

In the four studies that assessed rotational errors, the results all came out that rotational errors in the head and neck region were minimal. The magnitude of impact for rotational errors, however depends on the location of structures from the plan isocentre, and failure to correct patient's rotational error may lead to underdosing of

the target volume and unnecessary dosage to the surrounding critical structures (Arumugam et al., 2013).

2B.5.2 Limitations of the systematic literature review

One of the limitations of the study was the limited number of studies found (seven studies) after applying the inclusion and exclusion criteria. The small sample size limits the generalisability of this literature review.

There was lack of data regarding PTV margin calculation in head and neck. Also, some studies had a weak quality evaluation when evaluated with the Joanna Briggs Institute tool, therefore some of the studies were not considered to be reliable in terms of outcomes measures and statistical analysis.

The systematic review depended on pre-existing data and therefore the results from the data analysis relied on the methodology of the studies in the review. Relying on pre-existing data could introduce a self-reported data bias (Althubaiti, 2016). The studies were also subjected to confounding variables which could have had an impact on the outcome of the results.

The systematic literature review was susceptible to reporting bias since the study was limited to English language studies and this limitation resulted in language bias as other studies which were published in other languages were excluded. Another reporting bias was that of location bias since access to data was limited as the researcher was not able to go through all the resources related to health sciences due to a limitation in time and resources, however performing a dual-independent research design aided in expanding the search (Kirkham et al., 2010).

There was heterogeneity regarding the key characteristics and methodology design of the reviewed studies, therefore this limited comparison of study results.

2B.5.3. Strengths of the systematic literature review

This review represents a complete assessment of studies that calculated PTV margin on patients treated with VMAT to the head and neck region. It was able to identify gaps in existing literature that are related to determining the different methods that are used to calculate the PTV margin in the head and neck region and to identify factors that need to be considered when calculating the margin. It was opted to minimise bias during research analysis, and to ensure reliable findings.

An exhaustive search of the literature was done by two reviewers who analysed the literature independently using an appropriate number of sources that are reliable. A critical appraisal of the literature was done to evaluate the quality of the studies in the review.

The clinical implications of the study were to include the evaluation of inter-fraction motion, intra-fraction motion and target volume delineation in the margin calculation using the van Herk Formula. Based also on the systematic literature review, clinical departments should ideally opt for daily imaging as this appears to have a huge impact on the margin size.

2B.5.4. Conclusion of the systematic literature review

All the studies used the van Herk formula to measure the PTV margin and none of the reviewed studies made use of any other formulas. In view of this finding, it was concluded that formula appears to be acceptable for calculating the PTV margin for VMAT treatment for the head and neck region.

All the reviewed studies assessed inter-fraction errors from set-up errors recorded from the imaging software by considering the translational errors. The SD for the population random errors was found to be slightly higher than that of the population systematic errors. The systematic and random errors of set-up rotational errors were considered in some studies but were not utilised for the calculation of the PTV margin.

PTV margins obtained by the reviewed studies may be underestimated because systematic errors from target volume delineation were not considered, and not all studies assessed intra-fraction errors. It is recommended that further research is conducted with the scope of determining the effect of target volume variance on PTV margins for head and neck patients treated with VMAT. It is also recommended that a comparison of PTV margin results from different formulas is undertaken.

This systematic literature review served as a guide to develop the methodology for this study. Since different anatomical sites, imaging protocols, immobilisation devices, and treatment modalities may affect the PTV margin size (Kapanen et al., 2013; Djordjevic et al., 2014; Anjanappa et al., 2017), the methodology of this study was tailored to a single anatomical site of the head and neck region (the larynx), with patients undergoing treatment using the same immobilisation devices and all receiving VMAT. From the results obtained in the reviewed studies there is an indication that tumours located in the laryngeal region were found to be more susceptible to motion when compared to those found in the oropharyngeal and nasopharyngeal regions, demonstrating the significance of quantifying organ motion in the methodology section.

Chapter 3

Methodology

3.1 Introduction

This chapter discusses the data collection and analysis procedures that were used for this study. It presents a detailed description and justification of the study's design and provides a well-documented outline and discussion of the methods used, allowing any other researcher to replicate the study and test its viability. Study's limitations and ethical issues were also discussed.

To measure the PTV margin, the van Herk formula was used, and the following errors were considered in this study:

- target volume delineation variation
- intra-fraction motion errors
- total inter-fraction errors (set-up errors and organ motion)
- inter-observer variation in image matching

3.2 Research design

The study used a prospective and quantitative research approach. The research design was non-experimental and cross-sectional since there was no researcher intervention with the data, and the association of the outcomes, in this case exposures of the population, were measured and analysed in the natural setting of the population (Setia, 2016). Furthermore, this study may also be termed as *insider research* since the study was conducted in the researcher's organisation, therefore, the researcher was considered an *insider.* The researcher's role had to be balanced with the normal functional role held in the organisation and that of a researcher (Brydon-Miller and Coghlan, 2014). These two roles were kept separate through the use of intermediary persons during data collection.

Insider research provides a unique perspective on system's history and culture because it is conducted from the inside, and this allows for a depth of understanding and interpretation of data (Holian and Coghlan, 2012). Fleming (2018) identified other benefits of conducting insider research, such as access to inside knowledge, having

a pre-existing understanding that aid in data analysis and interpretation, and a generated knowledge that was intended to be relevant to the researcher's own practice. The challenge of this research design was to be able to choose when to consider oneself as a researcher and as a staff member working in the local oncology hospital. Another challenge with this type of research was the risk of disregarding certain issues in the study that an outsider would be able to perceive as important (Saidin and Khaliza, 2017).

In prospective studies, the outcome has not occurred, and the data are created after the start of the study (Ranganathan and Aggarwal, 2018). Prospective studies, as opposed to retrospective studies, allow for the tailoring of data collection to the data required to achieve the goals and are thought to be less susceptible to information bias (Taylor and Francis, 2013). This study required a once-weekly post-treatment CBCT to measure intra-fraction errors, and these data were not available retrospectively, therefore data was collected prospectively.

Data were analysed quantitatively using both descriptive and inferential statistics, with the results of the study being generalised to the population (Muijs, 2011).

The following sources of errors were included in the calculation of the PTV margin:

- variance in target volume delineation
- total inter-fraction errors (set-up and organ motion)
- intra-fraction errors
- variation in image matching

The van Herk formula was used to calculate the PTV margin size since this is an adequate formula to achieve this study's goals based on the literature review presented in Chapter 2. This is also supported by the systematic literature review evaluating the different methods to calculate PTV margins in head and neck cancer patients undergoing VMAT published by this research team[1].

[1] *Caruana, K., Refalo, N., Spiteri, D., Couto, J. G., Zarb, F., & Bezzina, P. (2021). PTV margin calculation for head and neck patients treated with VMAT: A systematic literature review. Journal of Radiotherapy in Practice, 1-8. doi:10.1017/S1460396921000546*

3.3 Population, sampling and recruitment procedure

There were three different populations included in this study:

1. Patients receiving VMAT to the laryngeal region. Their images were used to measure errors and variations.
2. Radiographers performing image verification for measurement of inter, intra-fraction errors, and inter-observer variation in image matching.
3. Oncologists and HSTs who were responsible for contouring CTVs for the measurement of target volume delineation errors.

3.3.1. Patients receiving VMAT to the laryngeal region

Patient target population

The whole extent of subjects to whom generalisation of results could be acted upon constitutes the “target population” (Banerjee and Chaudhury, 2010).

The patients’ target population referred to all patients who needed to be treated with VMAT for cancer of the larynx at a local hospital.

The inclusion and exclusion criteria for the patients’ population are detailed in table 3.1. There were no restrictions regarding gender or other therapies and all patients irrespective of their age were included in the population.

Table 3.1. Inclusion and Exclusion Criteria for Participating Patients

Inclusion Criteria	Exclusion Criteria
Treatment planned with VMAT.	
Treatment planned to the laryngeal region that was taken to include supraglottis, glottis, subglottis (Suárez-Quintanill, Fernández Cabrera and Sharma, 2020).	
Radical (curative) treatment.	
Daily online CBCT imaging through the Elekta® XVI software.	
	Patients under eighteen years of age since the disease is very rarely seen in this excluded age group.
Patients treated supine and immobilised using an Orfit 5-point thermoplastic mask.	
Ability to provide written informed consent to participate in the study.	
	Patients who did not complete their prescribed course of treatment.

Description of the accessible patients

The accessible population refers to all available patients treated with VMAT for cancer of the larynx at the local oncology hospital during the data collection period between June 2021 and May 2022 and who met the inclusion criteria. The target was a minimum of twenty patients, as recommended by The Royal College of Radiologists (2021) when determining the PTV margin size.

Patients' treatments were planned on either one of the two Elekta Versa HD™ available in the department with the dose outputs calibrated in the same way through quality assurance tests.

Patients' recruitment

A radiographer working in the local oncology hospital, acting as an intermediary person, accessed the Mosaiq Elekta® Care Management software to identify patients who fulfilled the inclusion criteria. The intermediary radiographer was responsible for providing information letters, approaching potential participants and inviting them to take part in the study as well as obtaining informed consent. This procedure was followed to protect participants' anonymity; it helped to prevent participants from feeling compelled to participate in the study if approached by the researcher (Fleming, 2018). The intermediary radiographer needed to recruit a minimum of twenty patients who received treatment during the data collection period and met the study's inclusion criteria, because smaller sample sizes could lead to large inaccuracies in the population systematic error results (The Royal College of Radiologists, 2021).

Patient sampling

Appropriate sampling is critical to ensure the external validity of the study findings (McEwan, 2020). Exhaustive sampling was used whereby all the accessible population was invited to participate in the study. Exhaustive sampling is time consuming because the searches frequently return very large data sets that are impractical to screen (Benoot, Hannes, and Bilsen, 2016), however, for this study, this was not the case as treatment to the larynx was not frequent in the local settings; therefore, the data were manageable.

To ensure that the sample size was adequate, the sample power was estimated (Table 3.2.). The type-I error was set to be 0.05, indicating a 5% chance that a significant difference is due to chance and not a true difference. This is the most common α level used by quantitative studies (Serdar et al., 2021). The power value, that is the ability to detect a difference between groups, was set to be 0.9. With these pre-determined parameters, the type-II error was measured to be 0.1, indicating a beta cut-off of 10%

that demonstrates the chance that a significant difference is missed. These values are consistent with Serdar et al. (2021) recommendation for a sufficient sample size that achieves a Type I error as low as 0.05 or 0.01 and a power as high as 0.8 or 0.9.

Table 3.2. Sample Calculation

Study Parameters	
Mean, population (*number of patients performed from the year 2018 when head and neck VMAT in the Radio Therapy department was introduced until 2021*)	60
Mean, study group	20
Alpha	0.05
Beta	0.1
Power	0.9

Calculated from: *S.P. Kane, 2019. Sample Size Calculator. ClinCalc.com https://clincalc.com/stats/samplesize.aspx*

3.3.2 Radiographers performing image verification

Radiographers' target population

Another category of participants included in this study were radiographers. The radiographers' target population consisted of six individuals working on the Elekta Versa HD™ linear accelerator which was specific to head and neck patients' treatment. Inclusion and exclusion criteria for participating radiographers are presented in table 3.3.

Table 3.3. Inclusion and Exclusion Criteria for Participating Radiographers

Inclusion Criteria	Exclusion Criteria
Trained in image matching on XVI – able to use clip-box and mask registration	
Two to seven years of clinical experience in the local oncology hospital	
Work on the head and neck treatment unit	
Available during the data collection period and willing to participate in the study	
	Radiographers who did not treat the participating patients

Radiographers' accessible population

The whole target population was accessible to the researcher, consisting of six radiographers who fit the inclusion criteria.

Radiographers' recruitment and sampling

An exhaustive sample of six radiographers was selected by the intermediary radiographer based on availability and different levels of experience in image matching.

The participants were asked to participate in procedures that enable assessment of inter- and intra-fraction errors, inter-observer variability in image matching errors and reliability assessments.

3.3.3 Clinical oncologists and HSTs contouring CTV margins

Clinical oncologists and HSTs target population

The third set of participants consisted of six doctors - three clinical oncologists and three higher specialists' trainees (HSTs). The participating doctors all worked in the radiotherapy department and were responsible for the CTV delineation of the target volume for patients receiving treatment to the larynx with VMAT.

The inclusion criteria required doctors to be trained in target delineation of head and neck tumours. Participants who could not adhere and/or complete the analysis were excluded from the study

Clinical Oncologists and HSTs accessible population

The accessible population was the same as that of the target population since it consisted of six doctors who worked in the local oncology hospital.

Sample and recruitment of participating HSTs and clinical oncologists

An exhaustive sample technique was used to recruit the HSTs and clinical oncologists for participation in this study. The sample size included all who met the inclusion criteria.

The participants were asked to delineate the target volumes from their clinical experience to assess inter-observer variation in target volume delineation.

3.4 Local oncology department procedures

Radiotherapy treatment prescription

The recommended prescribed dose according to the local clinical protocol for hypofractionated radical radiotherapy treatment for small volume glottis carcinomas (T1 and T2) was 5500 cGy in 20 daily fractions over 4 weeks. A prescribed dose of 7200 cGy in 60 fractions, 2 daily fractions over 6 weeks was also sometimes used based on the oncologists' clinical evaluation (Sir Anthony Mamo Oncology Centre, 2020).

Target volume delineation

There was no local departmental protocol for head and neck target delineation before and during the data collection period. The procedure applied at the hospital was based on published contouring guidelines such as eContour (Panjwani et al., 2019) and RTOG contouring atlas (Le et al., 2022), and training of doctors in the delineation of the target volume by senior doctors. The procedure was done once and was not counter-checked by another specialist unless the delineation was done by a clinical oncologist trainee, in which case it was done under the supervision of a specialist as required by the local protocol (Sir Anthony Mamo Oncology Centre, 2020). As a standard procedure, the whole larynx was usually contoured if the patient was receiving VMAT treatment to the larynx. Contouring was done using the Monaco® HD Treatment Planning system (TPS) (version 5.51). For patients treated with VMAT for laryngeal cancer, the standard PTV margin was 5 mm, with daily online CBCT image verification (Sir Anthony Mamo Oncology Centre, 2020).

Diagnostic CT with contrast enhancement was utilised in the local department to aid doctors in target delineation. In several cases, MRI and/or PET CT were occasionally used to enhance the CT plan for contouring.

Set-up procedure

Patients receiving treatment for larynx cancer were positioned supine, with a 5-point thermoplastic mask that immobilises the head, neck and shoulders. Three marks were placed on the thermoplastic mask during the CT scan planning procedure to identify the scan reference point. Alignment marks were also drawn on the masks to ensure that the patient was aligned straight on the treatment couch. As part of the CT planning procedure, to reduce organ motion, patients were asked to try and limit swallowing during the CT scan to avoid introducing a systematic error during this stage (Bahig et al., 2021).

Image-guided procedure

Image guidance was obtained from kV CBCT scans using the Elekta Medical Systems linear accelerator (Elekta Versa HD). The image acquisition parameters were according to the head and neck XVI protocol as follows: bow-tie filter, f1; collimator size, S20; gantry speed, 180°/minute; gantry rotation, -255° → 100° (Sir Anthony Mamo Oncology Centre, 2020). Figure 3.1 shows an example of a CBCT image of a patient treated to the head and neck region.

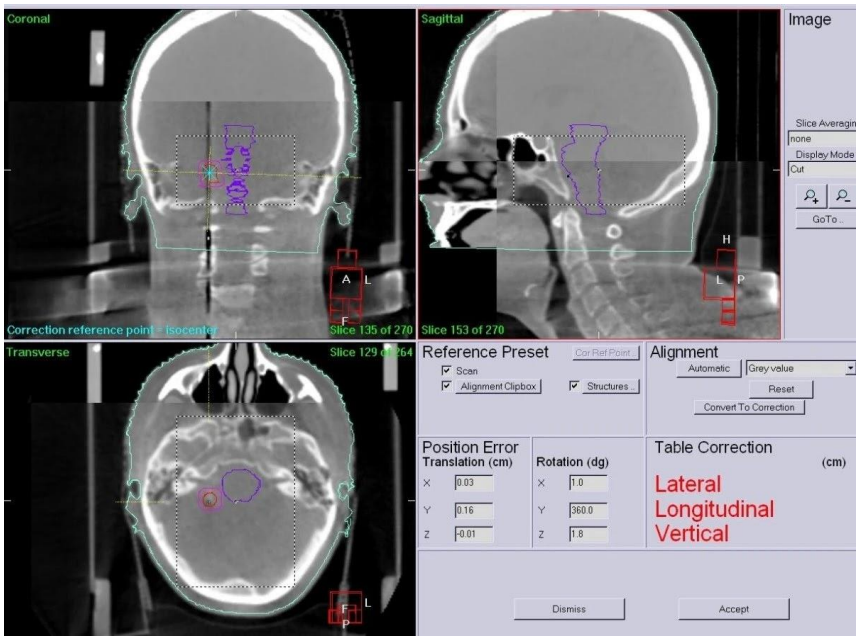


Figure 3.1: An Example of a CBCT Image of a Patient Treated to the Head and Neck Region (Ingrosso et al.,2012)

The matching protocol for CBCT imaging enabled the correction of inter-fractional errors. The clip-box was defined around the thyroid cartilage, vertebral bodies, and a portion of the cranium to encompass the volume of interest. Daily CBCTs were taken before the treatment and were matched using the Elekta® XVI software. The acquisition procedure took around two minutes. The radiographers working on the treatment unit performed online image verification using the XVI software's automatic match registration with the bone value T&R (Translation & Rotation) registration. Manual corrections of CBCTs were performed by the radiographers when necessary.

The CBCT was matched with the CT planning scan in three different planes: sagittal, coronal and transverse. As part of the routine procedure of the department, if inter-fraction set-up errors exceeded 2mm in any directional axis of the translation vector, radiographers on the treatment unit performed online corrections by shifting the treatment couch to the correct position. Patients were repositioned when translational set-up errors exceeded 1 cm and errors were not a systematic trend. When rotational errors exceeded 3°, radiographers were recommended to reposition the patient. The local department protocol required that in such cases, another CBCT scan would need to be acquired to verify the correction prior to the delivery of treatment (Sir Anthony Mamo Oncology Centre, 2020).

3.5 Data gathering tools

The following tools were used to collect the data necessary to measure the errors and variations to calculate the CTV-PTV margins.

The four tools were used to collect data to measure the following:

- target volume delineation variation
- intra-fraction motion errors
- total inter-fraction errors (set-up errors and organ motion)
- inter-observer variation in image matching

and these are discussed below.

Appendix C - Part i and **Part ii** contain the data collection tools used for this study. Part i contains the data collection tools used by the participating doctors and radiographers to record measures, whereas Part ii contains the data collection tools used for data analysis.

3.5.1 Data gathering tool for target volume delineation variation

The researcher, with the aid of the intermediary radiographer (who was responsible for anonymising the data by assigning a patient study number), prepared a data information sheet for doctors with information pertaining to the patients' diagnosis, clinical history, and tumour staging. This patient information data were provided to aid the participating doctors during the delineation of the CTV. An example of the patient information data given to the participants (participant 1) may be found in **Appendix C-Part i**.

The intermediary radiographer was also responsible for anonymising supplementary scans, such as MRI and PET scans, given to doctors as an aid for the delineation procedure as part of the department's standard procedures. Care was taken to ensure that participating doctors operated as they would ordinarily do as part of the clinical

process by using standard software and being blinded to the contours of other observers.

All target volumes were outlined slice-by-slice on the axial CT planning scans using Monaco® HD TPS (version 5.51). For each case, the CTV structures delineated by the doctors were superimposed on each other by being copied to the CT planning case. Since the doctors' and patients' information was anonymised, the researcher was able to measure the contouring range and calculate the target volume delineation errors. Data were recorded and analysed using Microsoft Excel **(Appendix C – Part ii)**.

3.5.2 Data gathering tools for intra-fraction motion errors

Intra-fraction errors were recorded from the Elekta® XVI software using a validated self-designed data record sheet **(Appendix C – Part i)** that contained the following parameters: patient demographics and clinical history – patient study number, age, sex, diagnosis, staging; PreCBCT soft-tissue registration and PostCBCT soft-tissue registration for translation and rotation parameters in the left-right (X), superior-inferior (Y), anterior-posterior (Z), Roll (Rx), Pitch (Ry), and Yaw (Rz) direction; number of repeated XVI's and reason for repetition and additional comments.

Another data record sheet, based on the Kwa et al. (2015) study, was used to collect the population systematic and random errors of translational and rotational intra-fraction errors. The data record sheet was formed using Microsoft Excel. This data record sheet facilitated data analysis **(Appendix C – Part ii)**. Intra-fraction errors values were determined by subtracting PostCBCT soft-tissue image registration results with PreCBCT soft-tissue image registration results in each direction. The data record sheet contained the corresponding values of the mean, standard deviation, population systematic, and random errors of intra-fraction errors.

3.5.3 Data gathering tools for total inter-fraction errors (set-up errors and organ motion)

To assess the total inter-fraction errors, the Elekta® XVI software's image registration findings were obtained using a validated data record sheet based on Palombarini et al. (2012) study **(Appendix C – Part i)**. It contained the following parameters: patient study number, fraction number, clip-box (bone-match) and mask registration (soft-tissue match) in the X, Y, Z, RX, RY and RZ direction; number of repeated XVI's and reason for repetition; additional comments.

Other data record sheets were created by the researcher using Microsoft Excel for ease of inter-fraction data analysis **(Appendix C – Part ii)**. One data record sheet was dedicated to translational errors, while the other was dedicated to rotational errors. Individual average deviation, individual standard deviation, population systematic errors and population random errors for inter-fraction errors were analysed and recorded in these tables.

3.5.4 Data gathering tool for inter-observer variation in image matching

A validated self-designed data gathering tool was used to obtain values for assessing inter-observer variation in image matching (**Appendix C – Part i**). The participating radiographers used this data gathering tool to record the translational (X, Y, Z) and rotational (RX, RY, RZ) errors obtained from image matching using bone registration.

To aid in the analysis of the data, the researcher transferred image matching data to another data record sheet using Microsoft Excel (**Appendix C - Part ii**). This data record sheet were based on that used by Hirose et al. (2020) and contained the corresponding population systematic and random inter- observer variation errors.

3.6 Validity and reliability of the data gathering tools

Validity and reliability are important indicators to assess the quality of the data gathered, which influences the accuracy of the findings (Sürücü and Maslakci, 2020). A study should incorporate both factors to support the validity of the findings.

All the following data gathering tools **(Appendix C – Part i)** and methodological design were tested for validity in this study:

- Target volume delineation data gathering tool
- Total inter-fraction errors data gathering tool
- Intra-fraction motion data gathering tool
- Inter-observer variation in image matching data gathering tool
- Methodological procedures for inter- and intra-fractional errors assessment

3.6.1 Validity

Heale and Twycross (2015) defined validity as the extent to which a tool can measure what it was primarily set to measure. Validity is crucial for the evaluation and development of tests (Worrell and Roberson, 2016).

Content validity is mostly used in quantitative studies (Shi, Mo and Sun 2012). This validation is based on experts' opinion on the quality of the data to be gathered and who assess if the tool measures the characteristics it was set to measure (Heale and Twycross, 2015). This process was adopted for this study. Experts in the area were appointed for the purpose of this study. These included clinical oncologists, medical physicists, and radiographers. These experts were asked to assign a score from 1 to 4 based on relevance and clarity to each category of the different tools used in this study. They were also invited to provide feedback on the data-gathering tools. When conformity of scores was established, content validity was confirmed. To be considered valid, an item or scale had to have a content validity ratio of at least 0.78 (Frey, 2018). Validity testing of each data gathering tool is discussed in the following sections. To have as large as possible a sample size, the experts who performed

validity testing where not excluded from participating in the main study. Furthermore, it was considered important to have their opinion on how to improve the data gathering tools. This opinion would not have influenced their performance in the study.

Validity of the target volume delineation data gathering tool

Two HSTs were asked to validate the data information sheet of target volume delineation errors. These experts had relevant experience and training in target volume delineation for cancer in the larynx. Additionally, they were asked to confirm that all the relevant patient information material was available for contouring.

With an individual content validity ratio of 0.9 and 0.8, respectively, and a mean ratio of 0.9, the target volume delineation data information sheet was confirmed to be valid **(Appendix D)**. In the opinion of the HSTs, PET scan images were not required for CTV delineation in larynx cancer. It was also suggested that patients' clinical examination, histology, and endoscopy results should be included when available. After making the necessary changes to the data collection sheet by taking into consideration all of the recommendations, the data collection sheet was re-validated by the same participants, yielding a content validity ratio of 1.

Validity of inter-fraction errors (set-up errors and organ motion) data gathering tool

For the validity of this tool, the experts were two radiographers with more than five years of clinical experience. These radiographers achieved their competencies in image registration and had relevant experience in image matching. The data collection tool was found to be valid, receiving a score of 1 on the validity test **(Appendix D)**. The experts advised combining the data collection tables of translation and rotational errors. Another suggestion was to combine the offline image evaluation in the same table and increase the number of rows to reflect longer treatment prescriptions. All these suggestions were included into the data collection tool.

Validity of intra-fraction motion data gathering tool

The same radiographers who acted as experts to validate the data gathering tool for inter-fraction errors were also asked to validate the intra-fraction motion data gathering

tool. This tool yielded a validity score of 1 since the experts were of the opinion that all the content was relevant to achieve the study's objectives **(Appendix D)**.

Validity of inter-observer variation in image matching data gathering tool

The data gathering tool for inter-observer variation in image matching was validated by another two radiographers with more than fifteen years of clinical experience. These radiographers had extensive clinical experience, and both underwent training in CBCT image matching. The tool for inter-observer variation, achieved a mean ratio of 0.9 of content validity.

Both experts agreed that rotational values are 'somehow relevant' for this assessment because the formula that will be used by the study to measure the PTV margin does not consider rotational errors. They did, however, mention that comparing the variation of rotational results in image matching would still be applicable for one of the objectives of this study, which was that of measuring errors present during the delivery of VMAT to the larynx, so the researcher decided to keep this variable in the data gathering tools.

External validation of the methodological design for inter- and intra-fractional errors assessment

Two foreign experts working in two different radiotherapy departments and specialised in PTV margin calculation, a medical physicist and a radiographer, were asked to independently validate the methodology design developed for this study to calculate inter- and intra-fractional errors. Both experts agreed that the methodology is appropriate to achieve the aims of this study.

The experts agreed that the clip-box (bone) displacements should be subtracted from the mask registration to measure the thyroid cartilage motion relative to the bony anatomy. One of the experts suggested calculating the displacement vector, which gives the magnitude and spatial direction of the resulting deviations. The other expert stated that when matching with the thyroid cartilage, an assumption is being made that

the thyroid cartilage motion is equal to the organ motion of the target volume. These recommendations were applied to the assessment of inter- and intra-fraction errors.

3.6.2 Reliability

Reliability is the degree to which a research method produces stable and consistent results (Sürücü and Maslakci, 2020).

The following variables were tested for reliability in this study:

- Contouring range distance measurements
- Set-up errors recorded on a data collection sheet
- Inter-fraction errors using soft-tissue registration

Reliability of the contouring range distance measurements

To determine the target volume delineation errors, the measurement of the distances of the contouring range by the researcher needed to be reliable. Therefore, inter-observer reliability was assessed. Reliability of the researcher was assessed by evaluating the consistency of the contouring range distance measurements by the researcher with those of the intermediary and medical physicists. Contouring range measurements were recorded independently on each alternating CT slice in the left, right, superior, inferior, anterior, and posterior directions for one of the five clinical cases used to assess target volume delineation errors.

SPSS software was used to measure the Intra-class Correlation Coefficient (ICC). Consistency between observers was achieved if the score was >0.7 (Koo and Li, 2016).

Reliability of set-up errors recorded on a data collection sheet

The six participating radiographers independently recorded bone-registration and mask-registration results on the data record sheet. These image registration results were collected from five CBCT scans that were saved on the XVI software. This

procedure was done to assess inter-observer reliability of data recording on a data collection sheet.

SPSS software was used to measure the ICC. Consistency between observers was achieved if the score was >0.7 (Koo and Li, 2016).

Reliability of inter-fraction errors using soft-tissue registration

The values for translational and rotational errors acquired by the participating radiographers from soft-tissue image registration were tested for inter-observer reliability. This procedure was also necessary to ensure that the radiographers were sufficiently trained to perform mask registration for the purposes of this study.

All the radiographers (six) who worked in the head and neck treatment unit were asked to perform a mask registration procedure with a manual shift when necessary to match the thyroid cartilage position with the reference CT scan. The matching of thyroid cartilage was used as a surrogate for matching the target volume due to the lack of soft-tissue contrast of the target volumes on CBCT.

The radiographers were required to match twenty-five randomly selected CBCT scans of patients who received treatment to the larynx with VMAT using the Elekta® XVI software. To ensure an adequate sample size, patients were retrieved retrospectively by the intermediary radiographer using the Mosaiq Elekta® Care Management Software and the treatment code 'Larynx VMAT'. Participating radiographers were then asked to record the registered set-up errors on the data collection sheet **(Appendix C – Part i)**.

To assess intra-observer reliability, the participating radiographers who performed image registration for the assessment of inter-observer reliability were given five of the CBCT scans previously used, after a two-week interval, chosen at random from the twenty-five CBCT scans. They were asked to repeat the image matching and record set-up errors on the data collection sheet **(Appendix C – Part i)**. The intermediary radiographer was responsible for selecting the scans and anonymisation of the

participating radiographers. This procedure was carried out to determine the radiographers' consistency when performing image matching on the same scans repeatedly.

SPSS software was used to measure Cronbach's Alpha and the ICC for both inter- and intra-observer reliability. A Cronbach's Alpha and an ICC >0.7 criteria indicated satisfactory results (Glasser, 2014).

3.7 Pilot study

Two pilot studies were conducted to test the feasibility and identify necessary modifications to the methodology before data collection. Amendments following the pilot studies improved the data gathering tools (Malmqvist et al., 2019).

For this study, the following two separate pilot studies were conducted for:

- Target volume delineation
- Inter-fraction, intra-fraction, and inter-observer variation in image matching errors

Pilot study for target volume delineation errors assessment

Since the clinical oncologists' sample size was small and participants in the pilot study could not participate in the main study, a medical physicist and a radiographer were chosen to participate in the pilot study. These two participants were asked to delineate two CT planning scans of patients treated for laryngeal cancer to ensure that they could delineate the target contours on anonymous CT scans in a timely manner. Participants' opinions were used to improve procedure and tool. A logbook was used to record the participants' opinions and assistance, and these are presented in **Appendix E**. This pilot study was beneficial in identifying and resolving issues of anonymising patients' data on Monaco TPS and CUV2 PACS software that was used to anonymise the supplementary MRI scans. It was also beneficial in identifying and resolving issues with obscuring participants' results from the other participants to avoid bias in target delineation error results.

Pilot study for inter-fraction, intra-fraction, and inter-observer variation in image matching errors assessment

Another pilot study was performed to ensure that the image registration results were recorded correctly on the self-designed data gathering sheet. Two of the six participating radiographers separately recorded the translation and rotational errors obtained from bone-registration and soft-tissue registration for each treated fraction of

a participating patient. All participants were chosen at random by the intermediary radiographer. Inter-observer variability of data was estimated to aid the researcher in identifying the necessary modifications that were required in the data collection tool. The data were later recorded on an excel sheet **(Appendix C - Part ii).**

The pilot study was useful in estimating the time required for data to be recorded on the data collection sheet by participating radiographers. The pilot study revealed that data collection would be time-consuming if it had to be done solely by the intermediary radiographer. Some data collection errors were caused by interruptions because the pilot study data were collected in between scheduled treatment sessions, but according to Dekker (2014), data collection errors could also be cause by human error through the repetitive nature of the task. As a result, the researcher decided that the radiographers working in the treatment unit record set-up errors obtained from clip-box registration on the data collection sheet and perform and record the offline CBCT image match using mask registration as soon as the treatment fraction was completed. This procedure helped eliminating errors through repetitive procedures, and in reducing the time required for data collection. The role of the intermediary radiographer was to ensure that the data were safely stored in a secured cupboard to ensure that the data collection procedure followed the guidelines for ethical approval.

Another significant point from the pilot study was that mask-registration results, for the assessment of total inter-fraction errors, should not be saved on the XVI software because doing so would change the original set-up error results obtained during patients' treatment. Intra-fraction errors obtained from post-treatment CBCT, on the other hand, could be saved.

3.8 Data collection

Data collection had to be collected after the participating patients finished their treatment. As the study was prospective and required a minimum of 20 patients, it was done over an 11-month period which was the time taken to reach this target.

The intermediary radiographer obtained demographics, diagnosis, staging, and histology information of the participating patients from patients' files in the records department and from the Mosaiq Elekta® Care Management. This information was useful in the analysis and discussion of the data. It was also important to determine whether or not the patients met the inclusion criteria.

Observers' variation in target delineation

Target delineation errors in CTV-PTV contouring were measured by assessing the inter-observer variation in contouring the CTV-PTV margin.

The six clinical oncologists/HSTs who took part in the study were required to contour the five patient cases that were randomly selected from the population sample and anonymised by the intermediary radiographer. These contours were done on the Monaco® HD TPS (version 5.51) as per departmental protocol. By assigning a different identification number to the same anonymised patient case, the doctors were blinded to the delineation of other participants (for example, case 1 was labelled as patient 11 for one doctor and patient 12 for another). The clinical procedure was carried out as usual by the participating doctors, who used the standard software to ensure the reliability of the data.

Descriptive statistics and measurement of overlap were calculated to analyse inter-observer variability.

Total inter-fraction set-up errors

Total inter-fraction set-up errors were measured using the procedure as suggested by Palombarini et al. (2012).

The total inter-fraction systematic and random errors were measured as a combination of set-up errors and organ motion errors, and it was calculated using the set-up errors observed from the daily pre-treatment CBCT. The clip-box was designed to include a portion of the cranium, cervical spine, and PTV. For the bone translation and rotation, T&R algorithm was the preferred choice for image matching since this type of algorithm provided a better bony match. As part of the standard procedure, automatic matching was initially used. When necessary, a manual match was performed afterwards by the radiographers to ensure a good superimposition of the vertebral bodies. In addition to the standard procedures, the six radiographers who worked in the head and neck treatment unit performed an offline mask registration with the mask placed over the PTV contour, later used to match the thyroid cartilage. This procedure assessed the difference between the isocentric position of the CT planning scan and the CBCT scans and was performed after the patients' treatment had been completed. Figures 3.2 and 3.3 show an example of a typical clipbox and mask registration position when using the Elekta XVI® software.

The radiographers in the treatment unit recorded the set-up errors for both the translational and rotational inter-fraction errors registered by bone-registration and those registered by mask-registration on the data record sheet (**Appendix C – Part i**). The errors were recorded by the radiographers from the XVI workstation in left-right, superior-inferior, anterior-posterior, roll, pitch, and yaw aspects, referred to as XoB, YoB, ZoB, RXoB, RYoB, RZoB for the bony match and XoS, YoS, ZoS, RXoS, RYoS, RZoS for the soft tissue mask match, respectively. The data collection sheet was kept safely in the treatment unit to ensure that patient data were safeguarded, and the intermediary radiographer was responsible for anonymising the data record sheet and handing it over to the researcher after treatment.

The researcher transferred the data recorded on the data collection sheet to Microsoft Excel. This procedure was carried out to subtract bone-registration values from mask-

registration values, allowing the analysis of population systematic and random errors for total inter-fraction errors.

Patients who required a rescan were still included in the study and the intermediary radiographer was required to identify and record on the data record sheet the day when the new plan was initiated and the reason for a re-scan. The participating radiographers were requested to record only the image registration results in which the patient was treated, and to ignore any XVI scan results if the patient required repositioning due to set-up issues.

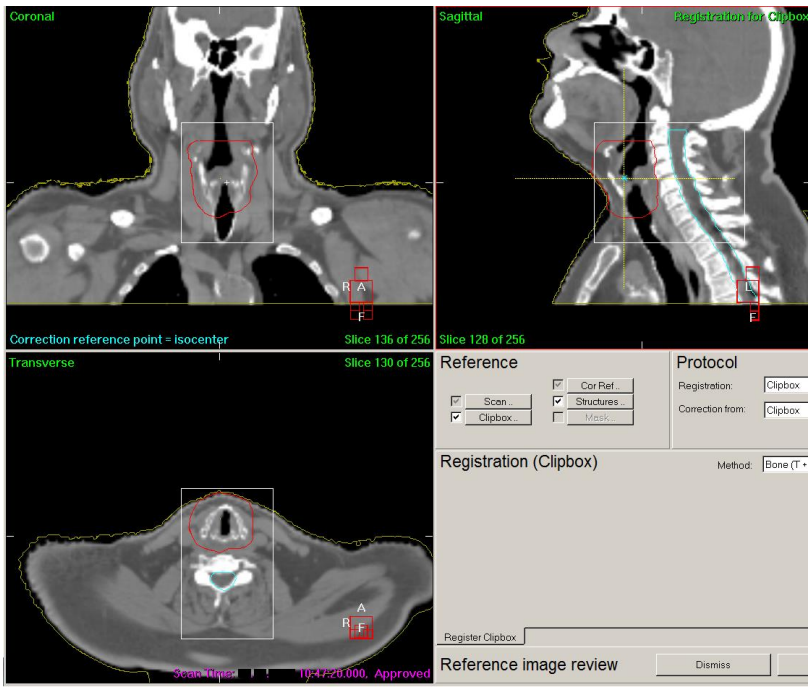


Figure 3.2: An Example of a clipbox position for radiotherapy treatment to the larynx (Elekta XVI® software)

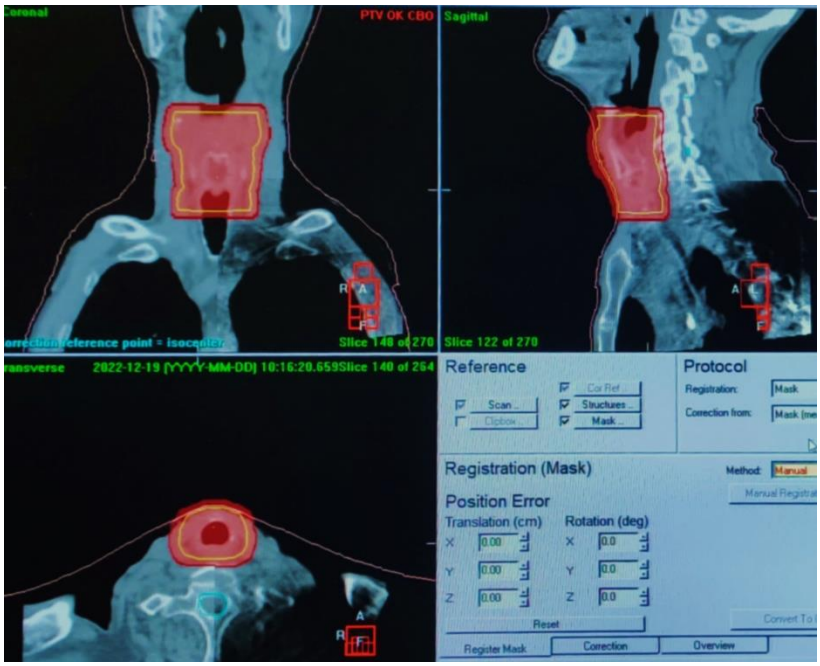


Figure 3.3. An example of mask registration for radiotherapy treatment to the larynx (Elekta XVI® software)

Intra-fraction errors

The intra-fraction errors resulting from patient positioning variation and organ motion were calculated using image verification results from weekly post-treatment CBCTs that were added to the study's accessible patient sample. Studies on intra-fraction errors, discussed in chapter 2, were used to develop this approach (Yin et al., 2013; Velec et al., 2010).

The scans were matched offline by the radiographers responsible for treatment delivery using a soft-tissue match through a mask registration around the PTV contour, ensuring that the thyroid cartilage is superimposed on the CT reference image. Image verification results of mask registration were saved onto the XVI software and collected with the aid of the intermediary radiographer using the data collection tool **(Appendix C – Part i)**. These data were later recorded by the researcher on Microsoft Excel using the data collection tool **(Appendix C - Part ii)**.

Inter-observer variation in image matching

Inter-observer variation errors of image matching were assessed only for the bone registration since the department's protocol required image matching to be based on bone registration (Sir Anthony Mamo Oncology Centre, 2020). Five patients were randomly selected from a pool of retrospective patients who received treatment to the laryngeal region with VMAT. Data were collected from retrospective patients for this procedure to ensure an adequate sample size and because these data were accessible to the intermediary radiographer.

Six radiographers who agreed to participate in the study were asked to re-analyse a total of twenty-five scans (randomly selected from the five patients) by performing automatic image registration and manual movement when required on the selected scans. The obtained shifts were recorded on a data record sheet **(Appendix C - Part i).**

Each participating radiographer was trained in image registration utilising automated and manual bone-registration and had clinical experience in image matching ranging from 3 to 7 years. The radiographers had access to views in the transverse, sagittal, and coronal planes and were able to change the window width or level whenever they wanted.

PTV margin

The population systematic and random errors of target volume delineation errors, total inter-fraction errors, intra-fraction errors and inter-observer variation in image matching were recorded on a data record sheet (**Appendix C – Part ii**) in Microsoft Excel. These translational errors were used to calculate the PTV margin using van Herk's formula.

3.9 Analysis of data

Data were analysed with the intent of presenting relevant and conclusive information to the study. Descriptive and inferential statistics were used to analyse the data. Descriptive statistics enable data summarisation using tabular, graphical, and numerical techniques (Lee, 2020) whilst inferential statistics allow one to draw a conclusion about the entire population based on an estimate from the sample (Trafimow and MacDonald, 2017).

Target Volume Delineation Errors

The method of assessing target volume delineation errors was adapted from the procedure as suggested by Tudor et al. (2020). However, the procedure was slightly modified by considering outliers in the calculation of target volume delineation errors since this was a true representation of the delineation obtained by doctors in the local department.

To analyse inter-observer variation in target volume delineation, five patients were randomly selected from the population sample by the intermediary radiographer. The patients' CT planning data were then made available to an exhaustive sample of clinical oncologists and HSTs who were available to perform target volume delineation.

All contours for each patient case were superimposed as a single structure set using the Monaco® HD TPS (version 5.51). Six perpendicular measurements were taken at specific points chosen to be visually representative of the variation around the contour on every alternating CT slice when all of the doctors' six delineated contours were visible.

The contouring range was measured as the distance taken from the outer to the innermost superimposed contours in the X, Y, and Z axes, without excluding outliers. These measurements were taken on each alternating CT slice by using the 'Measure Tool' from Monaco TPS (figure 3.4).

Figure 3.4: Measure Tool – Monaco TPS

Figures 3.5 and 3.6 show examples of measurements that were taken for patient study 2 in the coronal and axial planes, respectively. The coronal plane was used to measure the superior and inferior distances, whilst the axial plane was used to measure the anterior, posterior, left and right distances.

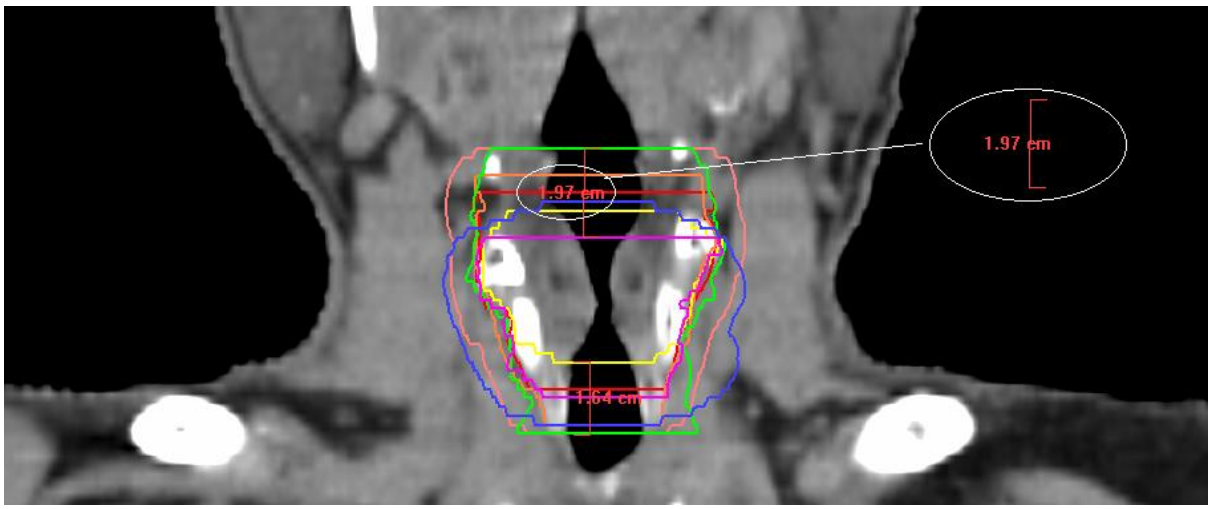


Figure 3.5: A Demonstration of the Superior and Inferior Measurements of Distances taken in the Coronal Plane.

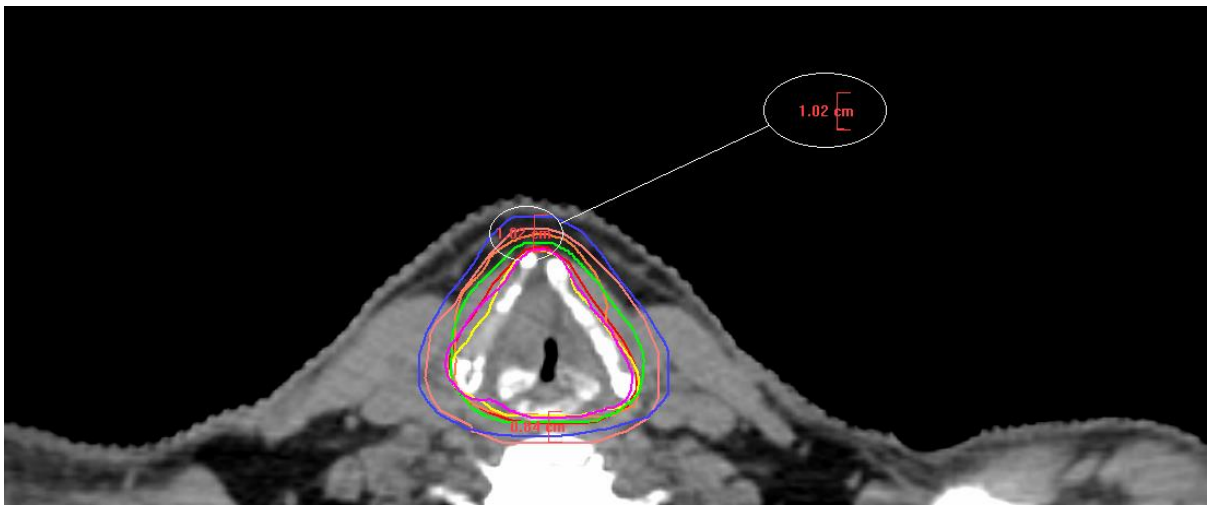


Figure 3.6: A Demonstration of the Anterior and Posterior Measurements of Distances taken in the Axial Plane.

The mean value of the measurements that were taken in each alternating CT slice was calculated in each direction, and this represented the data range of contour variation.

Since the number of observers for each case was less than 15, the standard deviation (S) was calculated using the following equation that was proposed by Tudor et al. (2020):

$$S = \frac{R}{d_2(N)}$$

Where R represents the data range and is calculated by measuring the distance between the inner and outermost contours in each image plane along each axis of interest at a representative point with 'average' observer variation. N represents the sample size, and d2 is a value that depends on the number of samples in the range. For a sample of 6 observers, the corresponding d2 value is 2.63 (Tudor et al., 2020).

The SD values from each case were combined by taking the mean value in each direction, representing the systematic error of target volume delineation.

Inter-fraction errors

The inter-fraction errors were a measure of set-up and organ motion errors (Palombarini et al., 2012). The results of the bony match were subtracted from the results of the soft tissue match to determine the total positioning errors, which were identified by the differences $X_0S – X_0B$, $Y_0S – Y_0B$, $Z_0S – Z_0B$, $RX_0S – RX_0B$, $RY_0S – RY_0B$, $RZ_0S – RZ_0B$. The result obtained indicated the motion of the larynx relative to the bony anatomy. Positive values for the translational errors (X, Y and Z) shifts indicated a left, superior and anterior displacements of the isocentre. This is demonstrated in table 3.4.

Table 3.4. Legend for Errors Directions

	X-axis	**Y-axis**	**Z-axis**
+	Left	Superior	Anterior
-	Right	Inferior	Posterior

Inter-fraction errors were recorded and calculated using Microsoft Excel using the following procedure described by The Royal College of Radiologists (2008).

The individual systematic errors ($m_{individual}$) were calculated by adding the set-up errors for each imaged fraction (Δ1+ Δ2+ Δ3...) and dividing this value by the number of imaged fractions (n).

$$m_{individual} = \frac{\Delta_1 + \Delta_2 + \Delta_3 + \ldots\ldots + \Delta_n}{n} \quad (1)$$

Following the determination of the individual systematic errors, the mean systematic errors of the population (M_{pop}) were calculated by adding the means of each individual patient (m1+m2+m3+...) and dividing the resultant value by the number of patients (P) in the group.

$$M_{pop} = \frac{m_1 + m_2 + m_3 + \ldots\ldots + m_p}{P} \quad (2)$$

The population systematic errors were then calculated by squaring the differences between the individual systematic errors mean derived from equation 1 and the population mean of individual systematic error derived from equation 2. The resulting sum was then divided by the number of patients minus one, and the square root of the result was calculated.

$$\sum\nolimits^2_{set\text{-}up} = \frac{(m_1 - M_{pop})^2 + (m_2 - M_{pop})^2 + (m_3 - M_{pop})^2 + \ldots + (m_n - M_{pop})^2}{(P - 1)} \quad (3)$$

Individual random errors ($\sigma^2_{individual}$) were measured by summing the squares of the differences between the mean and set-up errors from each image. This sum was then divided by the number of scans minus one. The square root of the resultant sum revealed the individual random error.

$$\sigma^2_{individual} = \frac{(\Delta_1 - m)^2 + (\Delta_2 - m)^2 + (\Delta_3 - m)^2 + \ldots + (\Delta_n - m)^2}{(n - 1)} \quad (4)$$

The Population Random errors (σ_{setup}) were then calculated by measuring the mean of individual random errors.

$$\sigma_{set\text{-}up} = \frac{\sigma_1 + \sigma_2 + \sigma_3 + \ldots + \sigma_p}{P} \quad (5)$$

Intra-fraction errors

Intra-fraction errors were analysed by recording the displacement values indicated by the CBCT performed immediately after treatment (postCBCT) using the mask registration. The thyroid cartilage was used as a matching structure (Durmus, Tas and Uzel. 2020). These shifts were recorded in the X, Y Z, Rx, Ry and Rz directions.

The recorded values of postCBCT soft-tissue registration were deducted from the values obtained in the preCBCT softtissue registration. This procedure measured patient movement and motion caused by tumour movement.

Microsoft Excel was used to analyse the data. The procedure used to analyse inter-fraction errors was also used to measure systematic and random errors, as described by The Royal College of Radiologists (2008).

Inter-observer variation in image matching

The residual errors, which denoted the differences between bone matching and the CT reference position, were used to evaluate observer uncertainties. The difference between the six participating radiographers in analysing the data retrospectively were used to measure inter-observer variation errors.

The procedure described by Hirose et al. (2020) was used to calculate interobserver variations. The root mean square of the residual errors along the three translational directions (X, Y, Z) and the rotational axes (Rx, Ry, Rz) by N observers was used to calculate the population systematic error (ε_{inter}) and population random error (σ_{inter}), respectively, as:

$$\varepsilon_{inter} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\varepsilon_{inter,j}^{2}} \qquad (6)$$

And

$$\sigma_{inter} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\sigma_{inter,j}^{2}}, \qquad (7)$$

N denotes the number of observers. $\varepsilon_{inter,j}$ and $\sigma_{inter,j}$ represent systematic and random residual errors for an observer *j*, respectively. The systematic error ($\varepsilon_{inter,j}$) and random error ($\sigma_{inter,j}$) for an observer *j* were measured respectively by:

$$\varepsilon_{inter,j} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(m_{inter,i,j} - \overline{m}_{inter,j}\right)^{2}} \qquad (8)$$

And

$$\sigma_{inter,j} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\sigma_{inter,i,j}^{2}}, \qquad (9)$$

Where n denotes the number of patients. $m_{inter,i,j}$ and $\dot{m}_{inter,j}$ represent, respectively, the mean residual error of a patient *i* by an observer *j* and the mean residual error of all patients by observer *j*. The SD of the residual error of a patient *i* by an observer *j* is represented by $\sigma_{inter,i,j}$. $m_{inter,i,j}$ and $\dot{m}_{inter,j}$ are given by:

$$m_{inter,i,j} = \frac{1}{F}\sum_{k=1}^{F} d_{inter,i,j,k}, \ \overline{m}_{inter,j} = \frac{1}{n}\sum_{i=1}^{n} m_{inter,i,j}, \qquad (10)$$

Where *F* is the number of fractions and $d_{inter,i,j,k}$ represents the residual error at a fraction *k* of a patient *i* by an observer *j*.

PTV margin calculation

The van Herk formula was used to calculate the PTV margin. This formula is expressed as $2.5\sum + 0.7\sigma$, where $\sum$ refers to the quadratic sum of the population systematic errors and σ denotes the quadratic sum of the population random errors (Van Herk et al., 2000).

Systematic errors were calculated for inter-observer variation in target delineation, inter-fraction errors (set-up and organ motion), intra-fractional errors, and inter-observer uncertainties in image matching.

Random errors were calculated for inter-fraction errors (set-up and organ motion), intra-fractional errors and inter-observer uncertainties in image matching. PTV margin was then calculated according to the following equation:

PTV margin = $2.5(\sum^2$ delineation + $\sum^2$ intra-fraction motion + $\sum^2$ inter-fraction [set-up and organ motion] + $\sum^2$ IM observer variation) ^{½} + $0.7(\sigma^2$ intra-fraction motion + σ^2 inter-fraction + σ^2 IM observer variation) ^{½}.

This equation was based on a study by Van Herk (2004) on errors and margins in radiotherapy, where Van Herk stated that target delineation uncertainty was purely a systematic error since it had an identical influence on all treatment fractions. Set-up errors, intra-fraction motion, organ motion and inter-observer variation in image matching errors are all subjected to random and systematic errors (Van Herk, 2004).

Justification for the use of the van Herk formula

Based on the review of the literature presented in chapter 2, the van Herk Formula was used in all published studies retrieved related to PTV margin measurement for head and neck patients treated with VMAT.

Another reason for selecting this equation for this study was that the van Herk formula ensures that ≥90% of patients receive ≥95% of the prescription dose to the CTV volume (Van Herk et al., 2000). These parameters are according to ICRU 50 recommendations which specify that the maximum and minimum doses within the PTV should be 107% and 95% respectively (ICRU, 1999) and therefore the van Herk formula ensures adequate dose coverage to the target area.

Published margin calculations formulae that distinguish between random and systematic errors, such as the ICRU 62 formula (ICRU, 1999) and Parker et al. (2002), are frequently written as a linear combination of the random and systematic errors standard deviations (SDs). Other margins such as Bel et al. (1996) and Antolak et al. (1999) which do not distinguish between random and systematic errors, are generally smaller since they tend to underestimate the impact of systematic errors.

3.10 Ethical considerations

"Ethical issues could arise in all types of studies that involve human subjects, regardless of the nature of methodological rigour" (Ignacio and Taylor, 2013, p. 60). To conduct this research, confidential medical information regarding patients' characteristics, histology and tumour staging, needed to be revealed to the intermediaries, but not to the researcher. There was the risk of having the patient's privacy compromised, for this reason, a range of ethical considerations were addressed.

Permissions **(Appendix B)** to perform the research study to ensure that it is not of harm or detriment to participants and the target population were obtained from the following entities: Professional management in radiotherapy department, Medical Physicists Area Coordinator, Quality Assurance Manager, referring Oncologists responsible for patients with head and neck cancer, the Data Protection Officers (DPO), Human Resources and Administration manager, Clinical Chairperson of Oncology, and the Chief Executive Officer (CEO) of the local general hospital where the study took place. Additionally, an application to perform the study was submitted to the University of Malta Research Ethics Committee for consideration and approval. Study commenced once ethical approval was obtained (UREC FORM V_15062020 8219).

The procedures in this study safeguarded participants' privacy during and after data collection by hiding any information that could identify participants from the researcher. Anonymity and confidentiality of the patients' sample were safeguarded by asking the intermediary radiographer to identify the eligible target population from the Mosaiq Elekta® Care Management software and record the set-up errors that were achieved from the image verification process. Patients' personal details, such as name and identity number were not recorded on the data record sheet, instead each patient was identified by a code number for the purpose of data analysis. Participants' records were secured using a password-protected computer.

All data were anonymised on the TPS with the aid of a medical physicist for the target volume delineation procedure. A new unique identifier and a new study set were

created. The original contouring delineation with which the patients were treated was obscured to the participants. The participants were also unaware of the results obtained by other participants to avoid biased results.

According to their role in the study, the patients, the radiographer who agreed to act as the intermediary person, the radiographers who were asked to perform image matching, and the clinical doctors who were asked to delineate CTV received an informational letter depending on their participation, and were required to sign an informed consent form (**Appendix B**). This procedure was done to ensure a thorough understanding of the research process.

The role of the intermediary radiographer was to identify eligible participants and invite them to participate in the study by providing information about the study and obtain their signed consent, if they accepted. Another role was to collect inter- and intra-fraction set-up errors from the XVI software and to provide the researcher with the required anonymised information for evaluation as part of the study. The intermediary radiographer was also required to liaise with the medical physicist who provided assistance during data collection of target volume delineation errors by anonymisation of patients' CT scan.

Patient consent forms and information letters were available in both Maltese and English languages. Information on the risks of ionising radiation and the benefits of the additional scan in enabling detection of movement during treatment participation in the study was clearly explained in the information sheet and consent form. The patients were told that the benefit of a once-weekly post-treatment CBCT scan, which was performed in addition to the daily pre-treatment CBCT, was to determine if there was any variation in their position from the start to the end of their treatment session on the day of imaging. If movements were detected during treatment, the radiographers could take the necessary action to try to reduce the movement for the next treatment day. Patients were also informed that this procedure was necessary to calculate the treatment margin size, which would benefit future patients. The risks were explained to the patients by informing them that the CBCT scan involved X-rays exposure but that the radiation dose would be controlled to limit the risk associated with ionising radiation. They were also informed that during the days of the weekly additional CBCT

images, the treatment procedure would also be extended by about two minutes as a result of the image acquisition.

Clinical oncologists and radiographers who were eligible for participation in this study were given consent forms and information letters which were prepared in English.

All participants were informed that their participation in the study was entirely voluntary and that they could withdraw at any time by notifying the intermediary radiographer without affecting treatment delivery. Contact details were also made available to the participants in case of any queries related to the study. The signed consent forms were stored securely in a locked cupboard by the intermediary person to ensure anonymity and will be effectively destroyed at the end of the study.

3.11 Conclusion

This chapter described the methodology and research design utilised for this research. The van Herk formula was used to calculate the CTV-PTV margin for larynx in a local oncology hospital. As opposed to most literature, which focus on inter-fraction errors, various errors were included in the calculation of this margin: target volume delineation errors, total inter-fraction errors, intra-fraction errors, inter-observer variation in image matching errors. The next chapter presents the results, data analysis, and discussion of the results.

Chapter 4

Results and Discussion

4.1 Introduction

This chapter reports, critically analyses and discusses the study's findings. The findings, discussions, strengths and limitations are presented in sections for each error that was analysed.

4.2 Participants

Patient demographics

During the data collection period (June 2021 till May 2022), 20 patients received treatment with VMAT in the local oncology hospital for cancer of the larynx, and none were excluded from participation because they all met the inclusion criteria. Patients' demographics are presented in table 4.1: gender, age, diagnosis, tumour location, staging and radiotherapy prescription.

Table 4.1. Patients' demographic

Patients' number	Sex	Age	Diagnosis	Tumour Location	Staging and/or Grading	Prescription
1	F	83	SCC	Left vocal cord with subglottic involvement	T2 N0 M0	5500cGy @275cGy in 20#
2	M	61	SCC	Left vocal cord	T1a N0 M0	5500cGy @275cGy in 20#
3	M	68	DLBCL	Thyroid involvement	Stage 1E	3000cGy @200cGy in 15#
4	M	40	DLBCL	Thyroid involvement	Stage 1E	3000cGy @200cGy in 15#
5	M	83	SCC	Both vocal cords	T1b N0 M0	5500cGy @275cGy in 20#
6	F	45	SCC	Right vocal cord	T1a N0 M0	5500cGy @275cGy in 20#

7	M	65	SCC	Right vocal cords and anterior commissure	T3 N0 M0	6600cGy @275cGy in 30#
8	M	81	SCC	Glottis and Subglottic involvement	T2 N0 M0	5500cGy @275cGy in 20#
9	F	82	SCC	Right vocal cord	T3	5500cGy @275cGy in 20#
10	M	62	SCC	Right vocal cord	T1a N0 M0	5500cGy @275cGy in 20#
11	M	62	SCC	Left vocal cord	Grade 2	5500cGy @275cGy in 20#
12	M	46	SCC	Supraglottic	T3 N0 M0	6600cGy @ 220 cGy in 30#
13	F	45	SCC	Tumour infiltration to the thyroid cartilage	Grade 4	6600cGy @ 220 cGy in 30#
14	M	72	SCC	Supraglottic	T3 N1 M0	6600cGy @ 220 cGy in 30#
15	M	68	SCC	Right vocal cord	T1a N0 M0	5500cGy @275cGy in 20#
16	M	69	SCC	Left vocal cord and anterior commissure	T1a N0 M0	5500cGy @275cGy in 20#
17	M	57	SCC	Left vocal cord	T2 N0 M0 Grade 2	6050cGy 275cGy 22#
18	M	76	SCC	Right glottis	T3/T4	5500cGy @275cGy in 20#
19	M	73	Spindle cell carcinoma	Left glottis	T1a N0 M0	5500cGy @275cGy in 20#
20	M	62	SCC	Left glottis	T1a N0 M0	5500cGy @275cGy in 20#

SCC = Squamous Cell Carcinoma
DLBCL = Diffuse Large B-Cell Lymphoma
F = Female
M = Male

The sample population's median age was 65 years, with 80% males and 20% females.

SCCs accounted for 85 percent of all diagnoses, Diffused Large B-Cell Lymphoma (DLBCL) for 10%, and Spindle Cell Carcinoma for 5%. The most common tumour staging was T1a N0 M0, accounting for 35% of the total population, and the most common prescription was 5500cGy @275cGy in 20#.

Three of the patients in the sample had a rescan at some point during their treatment. These were participants number 6, 12, and 13. To avoid interfering with the population systematic and random errors analysis, registered errors prior to re-scan were ignored for total inter-fraction errors and intra-fraction analysis. Table 4.2. lists the reasons for re-scans.

Table 4.2. Reasons for Re-scans

Patients' study number	**Reasons for re-scan**
6	Set-up issues (rotation)
12	Loss of weight, significant contour change
13	Loss of weight, significant contour change

Number of assessed CBCT scans

A total of 465 CBCT scans were examined, including 357 pre-correction scans to assess the total inter-fraction errors, 82 post-treatment scans for assessment of intra-fraction errors, and 25 CBCT scans for assessment of observers' variation in image matching.

Thirty-eight CBCTs were not analysed for the assessment of total inter-fraction and intra-fraction errors, because they had to be repeated for various reasons such as, patient rotation, chin and shoulder displacement. A patient had a total of eight repeated scans due to set-up issues caused by variation in the chin position. This was the highest number of repeated CBCTs. In contrast, there were nine patients in the sample who had no repeated scans.

4.3 Reliability results

4.3.1 Reliability of the contouring range distance measurements

The Intra-Class Correlation Coefficient (ICC) test was used to analyse the inter-observer reliability in measuring the contouring range distance of the CTVs contoured by the oncologists. Results of inter-observer variability indicated good reliability between the participating radiographer, the researcher and the medical physicists, as a good correlation of 0.88 was recorded. The ICC description of values is listed in table 4.3.

Table 4.3. Description of ICC Coefficients

ICC	Reliability
0.9 - 1.0	Excellent
0.75-0.9	Good
0.5-0.75	Moderate
0.5-0.0	Poor

Koo and Li (2016)

4.3.2 Reliability of set-up errors recorded on a data collection sheet

Inter-observer reliability of set-up errors recorded on the data collection sheet was assessed for the six participating radiographers. For this procedure, five images were randomly selected by the intermediary radiographer from the XVI software. Each radiographer recorded the image registration results on the data collection sheet.

The ICC statistical measure was used to analyse reliability. The description of ICC values is listed in Table 4.3.

An ICC value of 1 was measured, indicating an excellent reliability between the participating radiographers in recording set-up errors on the data collection sheet.

4.3.3 Reliability of inter-fraction errors using soft-tissue registration

The inter-observer reliability of set-up errors obtained by mask registration was measured to assess the reliability of the participating radiographers in performing image registration using the mask technique. Twenty-five retrospective CBCT scans of patients treated for cancer to the larynx were presented to the six participating radiographers for analysis. The radiographers were told to re-analyse the scans with a mask registration

Reliability in image matching for the mask registration was measured using Cronbach's Alpha and Intra Class Correlation Coefficient (ICC).

The Cronbach's Alpha measures internal consistency between the observers and ranges from 0 to 1; a Cronbach's Alpha closer to 1 indicates higher consistency. Cronbach's Alpha larger than the 0.7 indicated satisfactory internal consistency (Fayers and Machin, 2011). The Cronbach's Alpha description of values is listed in table 4.4.

Table 4.4. Description of the Cronbach's Alpha

Cronbach's Alpha	**Reliability**
0.9 – 1.0	Excellent
0.8 – 0.9	Good
0.7 – 0.8	Acceptable
0.6 – 0.7	Questionable
0.5 – 0.6	Unacceptable

Fayers and Machin (2011)

The ICC measured the absolute agreement between the observers, ranging from 0 to 1. The descriptions of ICC values are listed in table 4.3 and were used as a criterion to determine the degree of reliability for the analysis of inter-observer reliability in mask registration. The confidence interval was set to 95%.

Results for inter-observer variability indicated good inter-observer reliability with Cronbach's Alpha values close to 0.9 for the Y direction (Tables 4.5 and 4.6), and excellent observer reliability of > 0.9 for all the other directions (Table 4.7 – 4.10). Radiographers may have found some difficulty to match with the vertebral bodies while making sure the thyroid cartilage is inside the PTV, which may explain why the Y direction had the lowest inter-observer variability.

Table 4.5. Cronbach's Alpha for the Superior-Inferior (Y) Direction

Cronbach's Alpha	N of Items
.892	6

Table 4.6. ICC for the Superior-Inferior (Y) Direction

		95% Conf. Interval		F Test with True Value 0			
	Intraclass Correlation	Lower Bound	Upper Bound	Value	df1	df2	P-value
Single Measures	.568	.404	.737	9.249	24	120	p<0.001
Average Measures	.887	.803	.944	9.249	24	120	p<0.001

Table 4.7. Cronbach's Alpha for the Left-Right (X) Direction

Cronbach's Alpha	N of Items
.931	6

Table 4.8. ICC for the Left-Right (X) Direction

		95% Conf. Interval		F Test with True Value 0			
	Intraclass Correlation	Lower Bound	Upper Bound	Value	df1	df2	P-value
Single Measures	.678	.530	.814	14.399	24	120	p<0.001
Average Measures	.927	.871	.963	14.399	24	120	p<0.001

Table 4.9. Cronbach's Alpha for the Anterior-Posterior (Z) Direction

Cronbach's Alpha	N of Items
.955	6

Table 4.10. ICC for the Anterior-Posterior (Z) Direction

	Intraclass Correlation	95% Conf. Interval Lower Bound	Upper Bound	F Test with True Value 0 Value	df1	df2	P-value
Single Measures	.779	.659	.879	22.463	24	120	p<0.001
Average Measures	.955	.921	.977	22.463	24	120	p<0.001

Results for intra-observer variability indicated an excellent intra-observer reliability amongst the participating radiographers > 0.9, except for the Z direction which had a low Cronbach's alpha result with a value of 0.662 and an ICC average measure which showed moderate agreement with regards to reliability. Tables 4.11 – 4.16. demonstrate the intra-observer reliability results for the translational direction.

Table 4.11. Cronbach's Alpha for the Left-Right (X) Direction

Cronbach's Alpha	N of Items
.924	2

Table 4.12. ICC for the Left-Right (X) Direction

	Intraclass Correlation [b]	95% Confidence Interval Lower Bound	Upper Bound	F Test with True Value 0 Value	df1	df2	Sig
Single Measures	.865	.666	.949	13.139	16	16	<.001
Average Measures	.928	.800	.974	13.139	16	16	<.001

Table 4.13. Cronbach's Alpha for the Superior-Inferior (Y) Direction

Cronbach's Alpha	N of Items
.948	2

Table 4.14. ICC for the Superior-Inferior (Y) Direction

		95% Confidence Interval		F Test with True Value 0			
	Intraclass Correlation[b]	Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.906	.759	.965	19.128	16	16	<.001
Average Measures	.951	.863	.982	19.128	16	16	<.001

Table 4.15. Cronbach's Alpha for the Anterior-Posterior (Z) Direction

Cronbach's Alpha	N of Items
.662	2

Table 4.16. ICC for the Anterior-Posterior (Z) Direction

		95% Conf. Interval		F Test with True Value 0			
	Intraclass Correlation	Lower Bound	Upper Bound	Value	df1	df2	P-value
Single Measures	.779	.659	.879	22.463	24	120	p<0.001
Average Measures	.955	.921	.977	22.463	24	120	p<0.001

Results and analysis of data can be found in **Appendix F**.

4.4 Target volume delineation errors

To the researcher's knowledge, this is the first study to investigate target delineation errors for PTV margin calculation in patients receiving VMAT to the larynx. Several studies, including Fotina et al. (2012), Kristensen et al. (2017), Tsang et al. (2019) and Trignani et al. (2019), were conducted with the purpose of analysing observer variation in target volume delineation, with one study even analysing observer variation for target volume delineation for supraglottic laryngeal carcinoma (Jager et al., 2015). However, the statistical tests that were used in these studies, such as coefficient of variation and conformity index general, were insufficient to estimate the delineation errors to determine the PTV margin size.

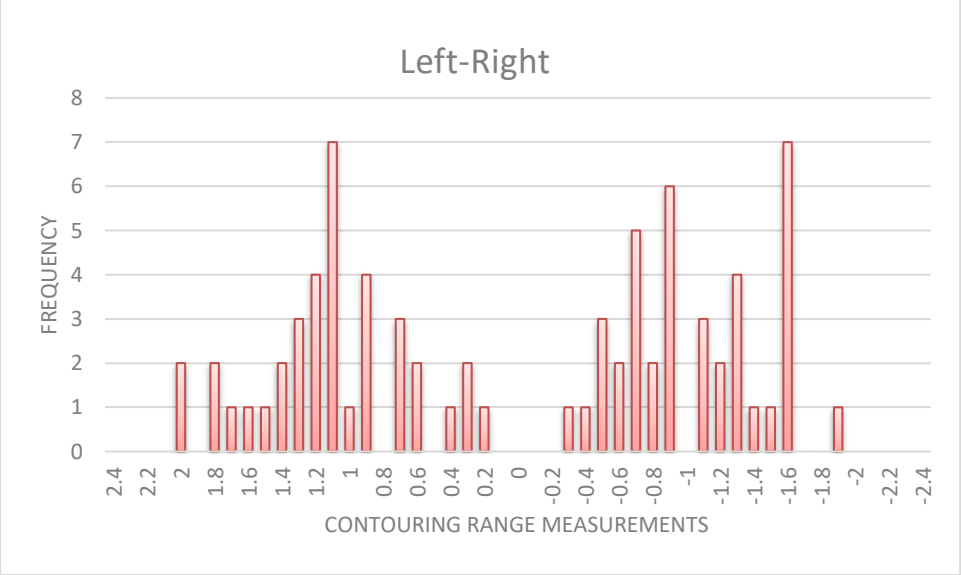
4.4.1 Results and discussions

To allow data comparison, each contour of a specific observer was assigned a specific colour. Table 4.17. demonstrates the colour coding that was used for each observer for ease of data analysis.

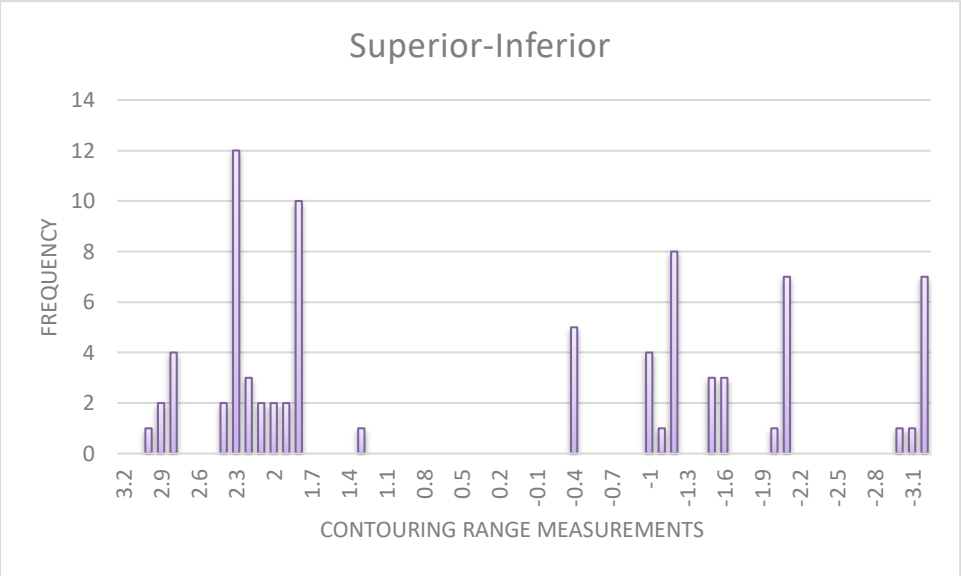
Table 4.17. Colour Coding System

Observer 1	Red
Observer 2	Orange
Observer 3	Yellow
Observer 4	Green
Observer 5	Pink
Observer 6	Purple

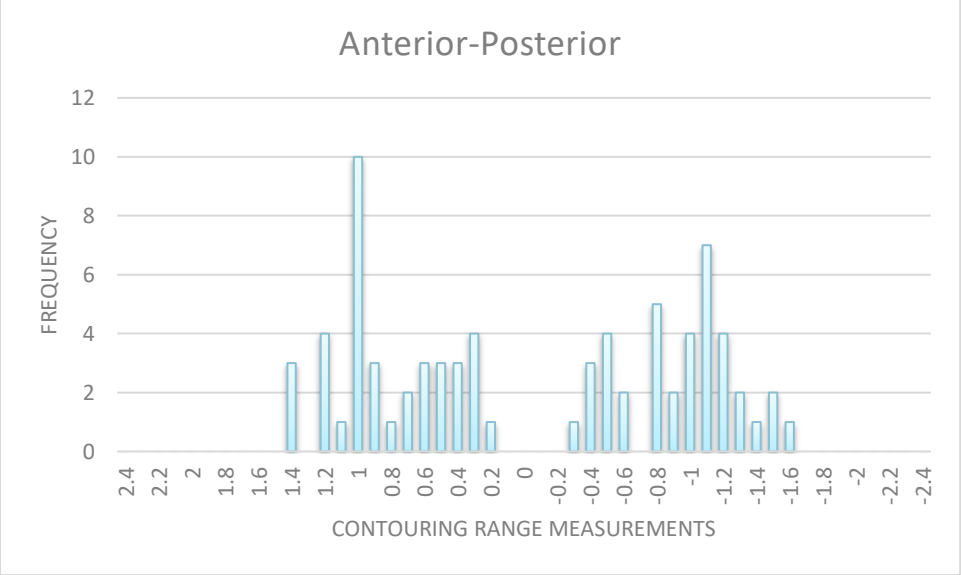
The frequency histograms of contouring range measurements are shown in figure 4.1.


Left-Right
FREQUENCY
8
7
6
5
4
3
2
1
0
2.4 2.2 2 1.8 1.6 1.4 1.2 1 0.8 0.6 0.4 0.2 0 -0.2 -0.4 -0.6 -0.8 -1 -1.2 -1.4 -1.6 -1.8 -2 -2.2 -2.4
CONTOURING RANGE MEASUREMENTS

(a)


Superior-Inferior
FREQUENCY
14
12
10
8
6
4
2
0
3.2 2.9 2.6 2.3 2 1.7 1.4 1.1 0.8 0.5 0.2 -0.1 -0.4 -0.7 -1 -1.3 -1.6 -1.9 -2.2 -2.5 -2.8 -3.1
CONTOURING RANGE MEASUREMENTS

(b)



(c)

Figure 4.1. Frequency histograms showing the distribution of contouring range measurements (cm) that were taken on each alternating CT slice for all patients. a) Left-Right (X) Target Delineation Errors. b) Superior-Inferior (Y) Target Delineation Errors. c) Anterior-Posterior (Z) Target Delineation Errors.

For the X, Y, and Z axes, the frequency histograms of the contouring range measurements all indicated a bi-modal distribution. The left direction's mode value was 1.15 cm, whereas the right direction was -1.6 cm. The superior direction's mode value was 2.3 cm and the inferior direction's mode value was -1.2 cm. The anterior direction's mode value was 1 cm, whereas the posterior value was -1.1 cm.

When compared to the other directions, the target delineation errors in the superior-inferior direction were more dispersed and less homogenous, indicating that the contouring range were the largest in these directions. The CT measurements of distances between the delineation for the superior-inferior direction ranged from -3.2 cm and 3 cm (figure 4.1, b). This was also observed by Jager et al. (2015), where the greatest delineation volume discrepancies for the epiglottic region were observed in the superior-inferior direction. Eight patients, from a total of 16, had at least one of the observers recording explicit difficulties in the delineation of target volume in the superior-inferior direction when using CT images, since the tumour borders in this direction were not clear (Jager et al., 2015)

Table 4.18. shows the mean value of the contouring range measurements for each patient from the outer to the innermost superimposed contours, obtained at each translational axis.

Table 4.18. Systematic Individual Mean Values of the Contouring Range Measurements in each Translational Axis

	Left (mm)	Right (mm)	Superior (mm)	Inferior (mm)	Anterior (mm)	Posterior (mm)
Patient 1	11.9	8.3	28.2	9.6	4.5	12.4
Patient 2	8.2	10.0	19.8	16.3	8.1	5.9
Patient 3	7.6	9.9	22.3	11.6	9.4	10.6
Patient 4	4.6	6.0	21.8	4.4	3.5	5.4
Patient 5	13.9	14.6	17.0	24.0	9.5	11.4
Population means	***9.2***	***9.8***	***21.8***	***13.2***	***7.0***	***9.1***

The superior direction had the highest overall mean discrepancy error of 21.8 mm, with patient 1 having the highest recorded average error of 28.2 mm. The anterior direction had the smallest discrepancy overall, with a mean discrepancy error of 7.0 mm, and with patient 4 having the smallest contouring range of 3.5 mm.

Table 4.19. compares the CTV for each of the six participating doctors and figure 4.2. demonstrates the resulting CTVs plotted in a scatter graph.

Table 4.19. Comparison of CTV for each Observer (cm^3)

Case	Participants					
	1	**2**	**3**	**4**	**5**	**6**
1	38.642	64.650	34.652	88.865	47.107	14.102
2	72.410	88.724	45.958	107.477	49.786	112.461
3	38.590	85.898	23.500	52.105	30.905	66.128
4	20.196	41.988	22.196	40.188	33.388	11.433
5	85.933	128.309	50.457	89.170	85.624	196.515

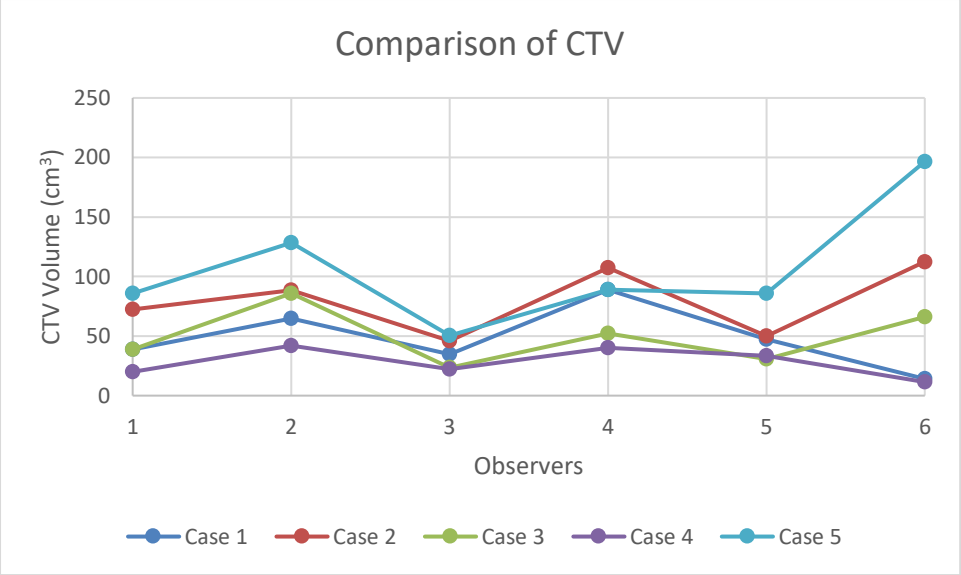


Figure 4.2. A scatter graph showing the variation of the contoured clinical target volumes (cm^3) obtained from the observers

Overall, the greatest inter-observer variation in contouring was observed in patient 5. This patient was the only one in the data set who had a tracheostomy, and by the time the patient attended the CT planning the lesion had grown significantly to almost twice the size from the initial diagnosis. The least overall inter-observer variation was observed in patient 4. This patient had a cancer in situ with a low tumour staging of T1a N0 M0.

From figure 4.2 it is evident that, when compared to the other doctors, observer 6 was identified as an outlier in the sample, who was overly generous in the delineation of three clinical cases and produced the narrowest margins in one clinical case. This observer differed from the other delineators in all directions, with discrepancy values ranging from 1 mm to 10 mm, 3 mm to 7 mm, 3 mm to 10 mm, 3 mm to 9 mm, 10 mm to 30 mm, and 3 mm to 11 mm, respectively in the anterior, posterior, left, right, superior and inferior directions.

Figures 4.2 and 4.3 demonstrate the contour discrepancy obtained by the outlier in two different planes on two different patients.

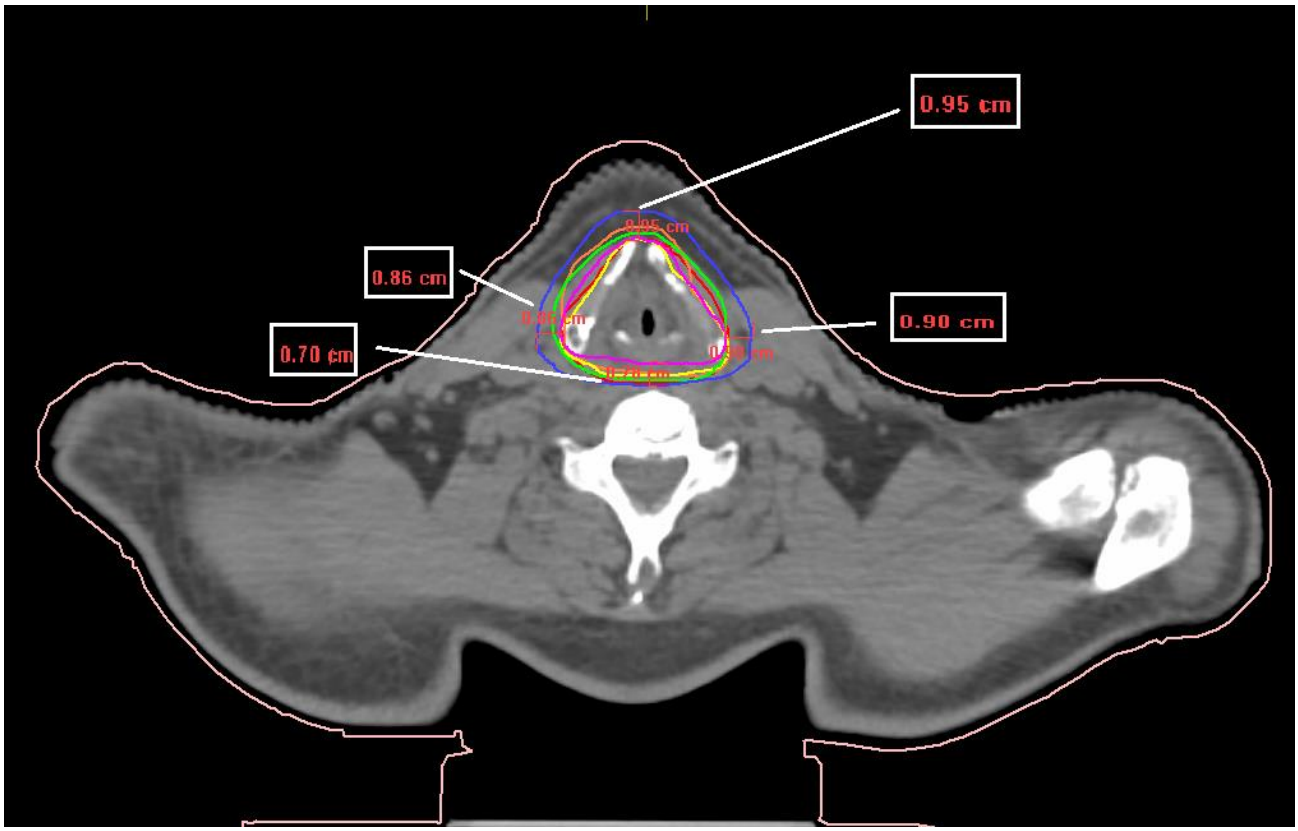


Figure 4.3: A Demonstration of the Outlier (observer 6) for Patient Number 2.

In figure 4.3, observer 6 was the outlier since it was the most generous contour in the sample for patient number 2. The contouring range distances from the outermost (the outlier) to the innermost contour are also demonstrated in this figure.

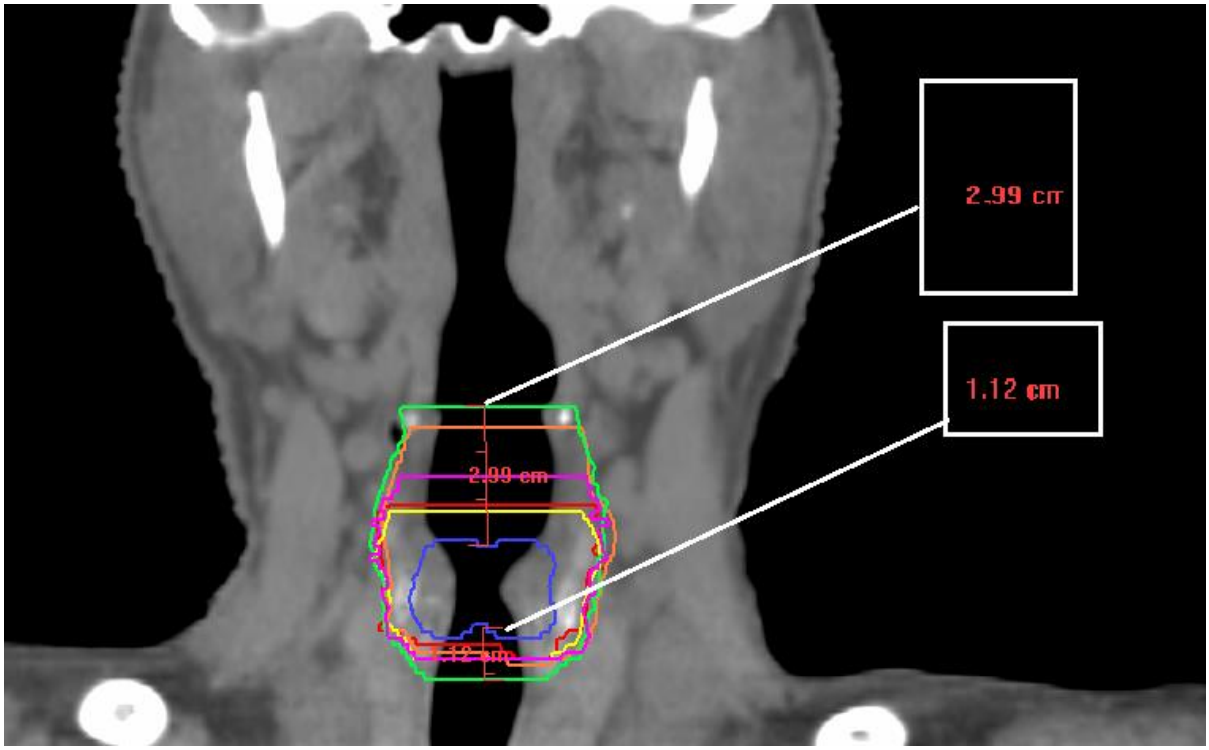


Figure 4.4: A Demonstration of the Outlier for Patient Number 1

In figure 4.4 the outlier is seen to have produced the smallest contour in patient number 1. The contouring range distances from the outermost to the innermost (observer 6) contours in the superior and inferior direction are demonstrated in this figure.

Patient 3 was the only case without any outliers. This patient had previous debulking surgery for a large T3 glottic tumour arising from the right vocal cord.

The mean distance between the outlines of six observers was 11.7 mm. However, there was a clear systematic variation in the superior-inferior direction when compared to the other directions. The population mean of the combined value of the superior and inferior directions was nearly twice as high as that obtained in the other directions, so the average range value was not considered as an isotropic delineation uncertainty (Tudor et al., 2020).

Instead, the superior-inferior direction's delineation errors were given twice the prominence of the other directions. As a result, excluding the superior-inferior direction, the average range value of six observers was 8.8 mm. The population standard deviation target volume delineation errors were estimated using the statistical method described by Tudor et al. (2020), in which the measurement of standard

deviation in small sample sized studies were dependent on the number of samples in the range calculation. In the case of this study, because six participants delineated five cases each, the sample number for each case was taken to be six, and this size of participants result in a common denominator value of 2.53. As a result, the standard deviation for the left-right and anterior-posterior directions was calculated as 8.8/2.53 = **3.47 mm**. For the estimation of target delineation in the superior-inferior direction, an average value of the contouring range of both the superior and inferior directions resulted in a value of 17.5 mm. This value was divided by 2.53 and resulted in a target delineation error of **6.92 mm**.

According to Jager (2017), observer variation in target volume delineation is higher in the head and neck region than in other tumour sites. Results obtained by 10 observer outlines in Tudor et al. (2020) for prostate delineation was an isotropic margin of 1.9mm and for fifteen observers delineation tumours located in the lungs was an isotropic margin of 2.1mm. In contrast, the delineation error results for this study were significantly higher than the other tumour regions.

Imaging quality

The accuracy of delineation is hindered if imaging modalities have a low resolution (The Royal College of Radiologists, 2021) and according to Mercieca, Belderbos and Van Herk (2021), observation variation is reduced when superimposing CT planning scan with PET scan or MRI. Simple measures, such as intravenous and/or intracavitary contrast, fiducial markers, and reproducible imaging protocols could significantly improve imaging quality. When contouring, the use of zoom levels, simultaneous viewing in multiple planes (sagittal and coronal planes), and adequate level and window settings on the planning CT reduce inter-observer variability (Segedin and Petric, 2016). In the current study, all the patients had a contrast scan fused with a non-contrast scan during CT planning. MRI diagnostic scans were also available to aid doctors in target delineation, however, the MR images were not acquired in the radiotherapy treatment position, therefore the images were not geometrically accurate and could not be superimposed on the planning scan. This limited most benefits of delineating with MRI (Schmidt and Payne, 2015).

Standardised protocols/guidelines

International guidelines, such as those of the RTOG and DAHANCA can be used by doctors for CTV delineation. Using site-specific anatomical atlases, consensus delineation guidelines, and standardised contouring protocols reduce variability between observers in various tumour sites (Kim et al., 2021). In the local oncology hospital where the study took place, there was no specific guideline which doctors follow. Selection of guidelines depended on the doctors' preference. When different or ambiguous guidelines are used for target volume delineation, this will have a significant impact on the consistency of delineated structures (Mercieca, Belderbos and Van Herk, 2021). According to Tudor et al. (2020), outliers should not be considered when measuring the range measurement of contouring because these contours would be inconsistent with clinical protocols, and one should not attempt to correct for major differences in opinion of the target volume. For this study, outliers were still considered as the department did not follow a specific clinical protocol regarding target delineation of the CTV volume for laryngeal cancer.

Since outliers were considered, the margin size may have been larger than necessary for the majority of patients, because most of the doctors' delineations were closer to each other. This could have also been the reason for the large delineation errors obtained in this study.

Specialised training

Having a diverse group of doctors with varying roles and experiences was an accurate representation of the local department. The HSTs in the local department rotate between different roles and this could cause inconsistencies in target delineation (Tudor et al., 2020).

Some publications have addressed the issue of training in target delineation. For example, Schimek-Jasch et al. (2015) reported that after a teaching session at a study group meeting, there was an improvement in overall inter-observer agreement, as evidenced by a reduction in target volumes. Khoo et al. (2012) had obtained similar results because a well-structured education programme reduced both inter- and intra-

observer prostate contouring variations. In contrast, Dewas et al. (2011), had reported no improvement among doctors following a teaching course. Reason for this could have been the high standard of the initial delineations. Furthermore, the researcher mentioned that several doctors discussed together to reach an agreement about the volumes that needed to be treated within their groups. This could have explained the homogeneity and high quality of the contours in Dewas et al. (2011) study.

4.5 Total inter-fraction errors

In this study, inter-fraction errors were a measure of set-up errors and organ motion and are being referred to as total inter-fraction errors. To our knowledge, this is the first study that assessed total inter-fraction errors for patients receiving treatment to the larynx with VMAT.

4.5.1 Results and discussions

Set-up errors using bone registration

Inter-fraction errors reported by other researchers were often focused on bone-match using CBCT or portal imaging (Yin et al., 2013). This type of match identified any errors that may occur when positioning the patient for treatment. This procedure was also followed by the radiographers as part of the standard practice in the local department (Sir Anthony Mamo Oncology Centre, 2020). In this study, the radiographers applied only the translational correction to the CBCT scans using the bone registration since the couch degree of movement was only in the translational direction, and therefore could not correct for rotational displacement.

The local departmental protocols (Sir Anthony Mamo Oncology Centre, 2020) stipulated that rotational errors should not exceed 3° and translation should not exceed 1 cm. The set-up errors recorded from the bone match ranged between -0.39 cm and 1.2 cm, -0.65 cm and 0.88 cm, and -1.02 cm and 0.67 cm in the X-, Y-, and Z-directions, respectively. Rotational set-up errors recorded from an automatic bone match ranged between -4.8º and 4.6º, -3.7º and 6.3º, -3.6º and 4.8º in the RX, RY, and RZ directions, respectively. The left-right errors exceeded the tolerance in 2 (0.5%) of the matches, but the superior-inferior, and the anterior-posterior errors did not. The rotational tolerance was exceeded in all rotational directions, with RX exceeding in 11 (2.7%) matches, RY exceeding in 33 (8.2%) matches, and RZ exceeding in 3 matches (0.7%).

The findings of this study indicated that, while radiographers generally adhered to clinical protocols, they were lenient in certain circumstances, particularly when it came to rotational errors, since the findings showed that rotational tolerance was exceeded in all rotational directions.

Lack of mask fixation due to contour loss from weight loss or tumour shrinkage, as well as mask tightening due to swelling, are also common factors that contribute to set-up errors in the head and neck region, and they may also be determined using a bone-match (Oh et al., 2014).

The efficacy of bone match can be affected by significant deformation, shrinkage, and rotation. Because changes in the shape of the tumour or patients' posture can cause misregistration and misalignment, not all structures within a clip-box can be simultaneously aligned using bone registration (Yin et al., 2013). When Gangsaas et al. (2013) investigated the relationship between primary tumour displacement and vertebral motion in patients with laryngeal cancer, the PTV margin was estimated to be 6.9 mm despite daily online vertebral repositioning due to the poor correlation between vertebrae and primary CTV displacements, which mostly occurred in the superior-inferior direction. Bahig et al. (2021) emphasised the importance of daily imaging with a soft tissue match on patients treated to the larynx, because anatomical changes or set-up reproducibility can result in a laryngeal shift in relation to the vertebra. Therefore, this study opted to match with the thyroid cartilage because the targeted area adheres to this structure.

Total inter-fraction errors

The distribution of total inter-fraction errors in each of the three translational directions and rotational axis were calculated using the 402 pre-correction CBCTs images. Set-up errors were recorded by performing an automatic and manual bone match on the Elekta XVI® software, followed by soft-tissue matching of the thyroid cartilage using a mask-registration of 0.5 cm around the PTV, and manually matching when necessary. This procedure allowed for the calculation of total inter-fraction errors, which included set-up errors and organ motion. The total inter-fraction errors were calculated by

subtracting the values obtained by mask-registration from those obtained by bone-registration.

Total inter-fraction errors in the translational direction

The individual mean (systematic) of inter-fraction errors in the X-, Y-, and Z-directions ranged between -2.1 mm (right) and 2.3 mm (left), -6.7 mm (inferior) and 4.4 mm (superior), and -0.9 mm (posterior) and 1.8 mm (anterior), respectively.

Individual mean (systematic) of total inter-fraction errors in the translational direction is presented in the following scatter graph (figure 4.5)

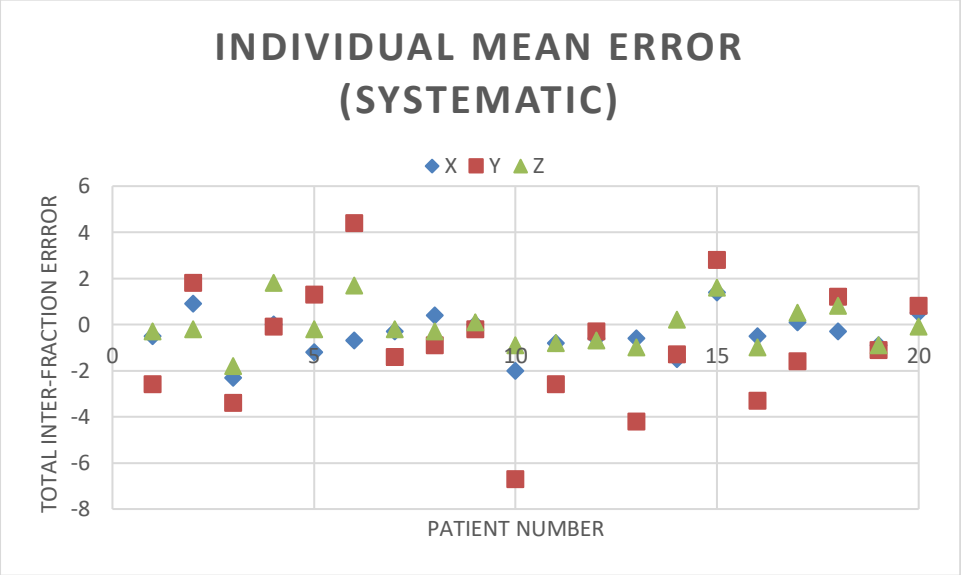


Figure 4.5: Scatter Graphs of Individual Mean Errors (mm)

The largest uncertainties of the individual mean errors were found in the superior-inferior direction and was observed in patient 10 (figure 4.4). There was a lot of variation in the thyroid cartilage position for this patient, with the soft tissue mask registration results varying from -0.13 cm to -1.14 cm in the Y direction. There were much less discrepancies in the bone registration since the superior-inferior translational errors ranged between 0.1 cm and -0.14 cm. Because the mask registration uncertainty was always in the inferior direction for patient 10, it is possible that the patient swallowed during the CT scan procedure, introducing systematic errors

at the radiotherapy treatment planning stage since swallowing is associated with a laryngeal elevation (Perillo et al., 2021).

The highest recorded individual mean errors for the remaining patients in the sample were also recorded in the Y direction, and these were present in both the superior and inferior directions. These types of errors could be caused by thyroid cartilage motion caused by breathing and swallowing during treatment (Perillo et al., 2021).

According to Laursen et al. (2012), patients with a significant amount of rotation tend to have the target shifted in the same direction. Over/under-dosage will always occur at the same position in such cases and is more harmful than random shifts. Re-scanning and re-planning the patient could correct such systematic variations. For this reason, patient 6 had a re-scan during the first two weeks of treatment due to set-up issues that resulted in rotation.

Individual random errors of total inter-fraction errors

The individual random errors of total inter-fraction errors in the X-, Y-, and Z-directions ranged between 0.1 mm and 1 mm, 0.2 mm and 1.5 mm, and 0.1 mm and 0.9 mm, respectively.

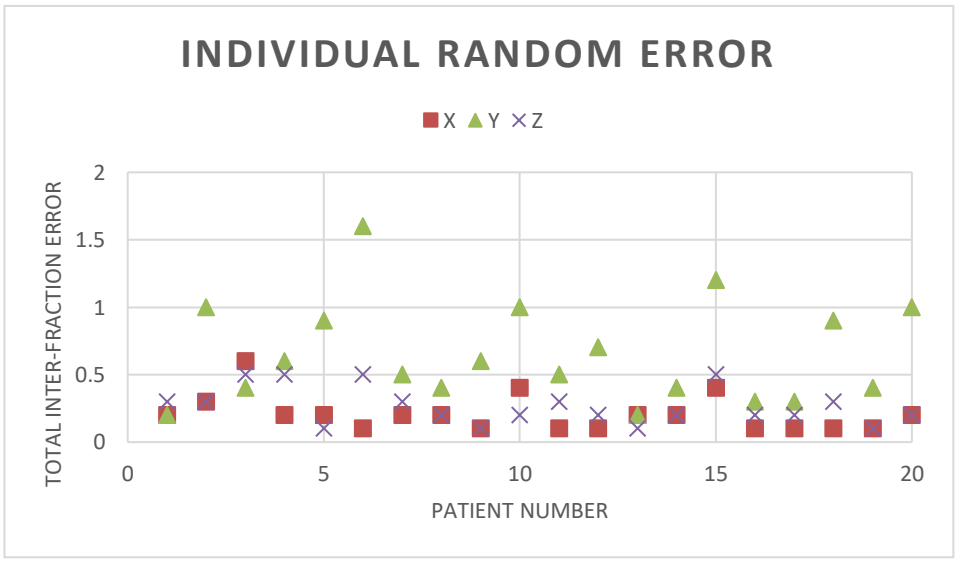


Figure 4.6. Scatter Graphs of Individual Random Errors (mm)

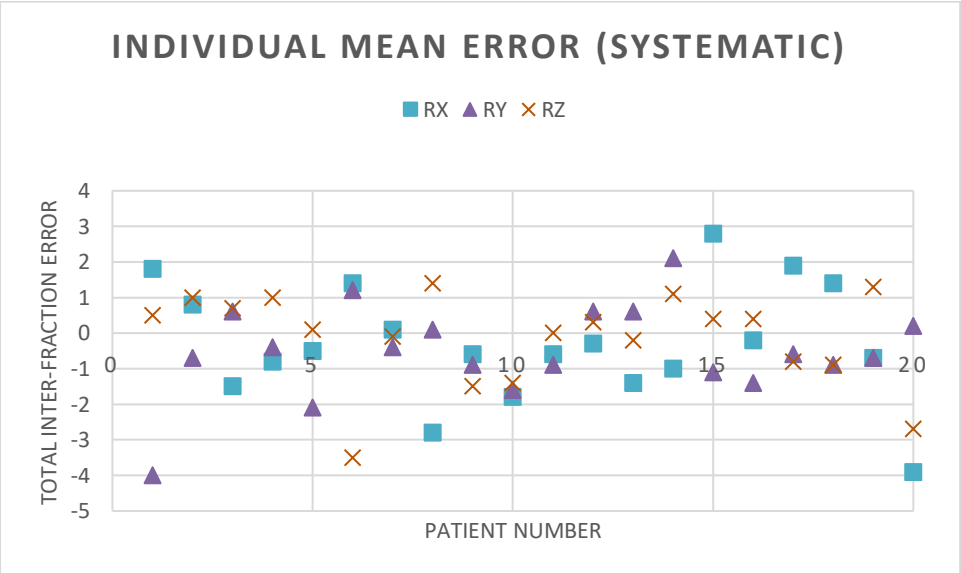
The individual random errors were overall higher compared to the individual mean errors (systematic), and the largest random errors were in the superior-inferior direction (figure 4.6). Patient 6 had the highest mean individual random error in the Y direction, measuring up to 1.6 mm. This patient had a rescan after 2 weeks of commencing treatment due to set-up issues related to rotation. After the re-scan, the radiographers used the “Grey-value T + R registration” because it provided a better image match than the bone-registration.

Individual mean errors (systematic) of total inter-fractional errors in the rotational direction

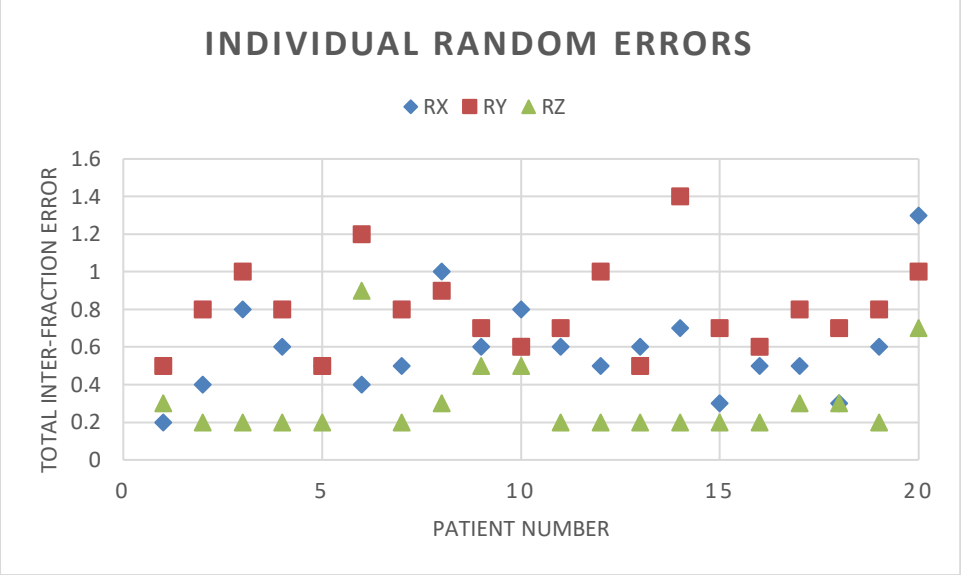
The individual mean errors (systematic) of total inter-fraction errors in the RX-, RY-, and RZ-directions ranged between -3.9 (counter-clockwise) and 2.8º (clockwise), -4.0 (counter -clockwise) and 2.1º (clockwise), and 1.4 (clockwise) and -3.5º (counter-clockwise), respectively.

The individual random errors of total inter-fraction errors in the RX-, RY-, and RZ-directions ranged between 0.2 and 1.3º, 0.5 and 1.4º, and 0.2 and 0.9º, respectively

Individual total inter-fraction errors in the rotational direction are presented in figure 4.7. a, and b.



(a)



(b)

Figure 4.7. Scatter graphs of the individual mean of total inter-fractional errors in the rotational direction (º). (a) Individual mean errors (Systematic). (b) Individual random errors.

Population systematic and population random errors of total inter-fraction errors

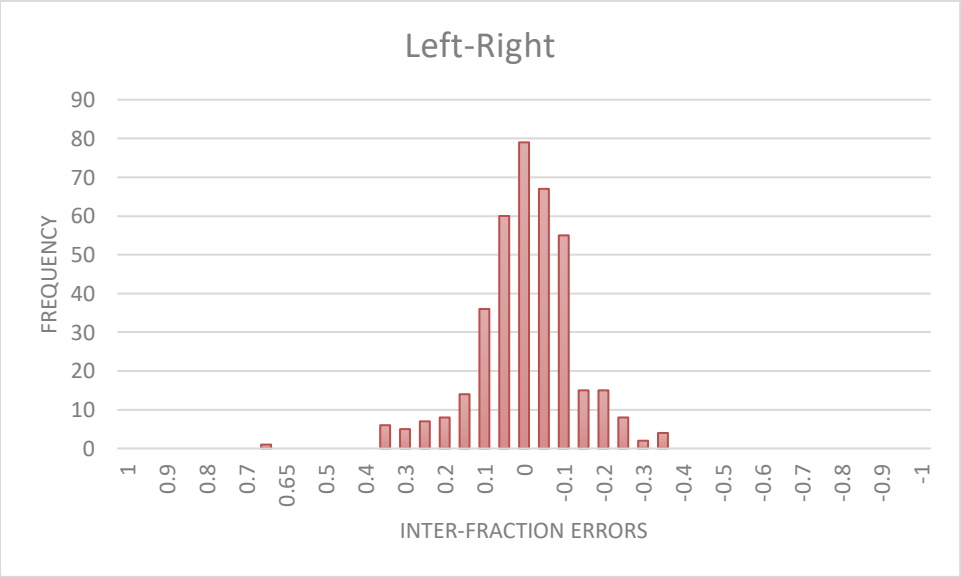
The population mean of total inter-fraction errors was calculated by averaging all individual mean errors. The population systematic errors showed the spread of individual means, whereas the population random errors were the average of all individual random errors. The population total inter-fraction errors results are summarised in table 4.20.

Table 4.20. Summary of Population Translational Total Inter-fraction Errors

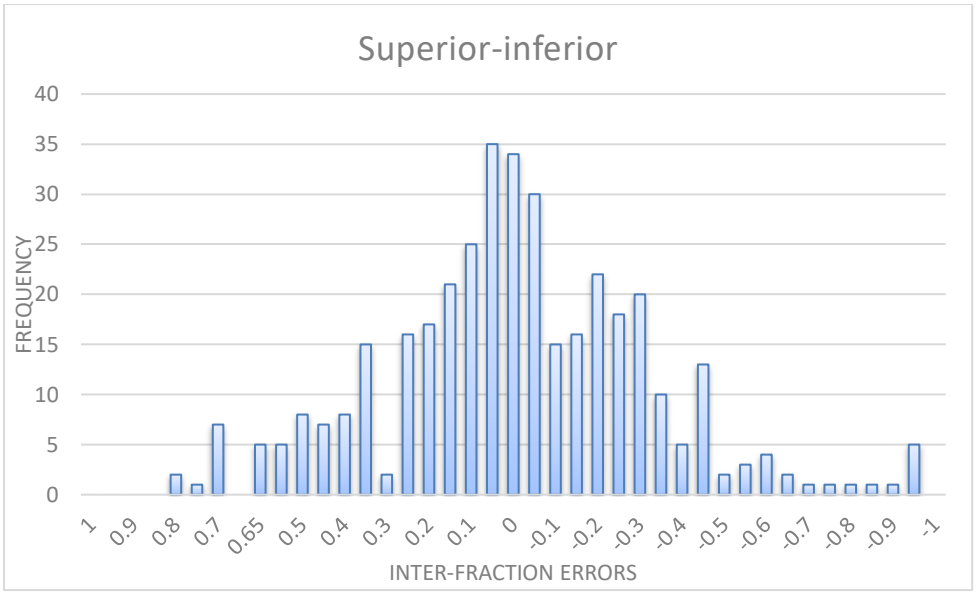
	X (cm)	Y (cm)	Z (cm)
M	-0.02	-0.09	0.02
SD	0.13	0.32	0.14
Minimum	-1.0	-0.95	-0.7
Maximum	0.65	0.8	0.65
Σ	0.1	0.26	0.1
σ	0.02	0.07	0.03

M, mean of all patients' mean; SD, Standard Deviation Σ, population systematic set-up and organ motion errors; σ, population random set-up and organ motion errors; X, left-right; Y, superior–inferior; Z, anterior–posterior

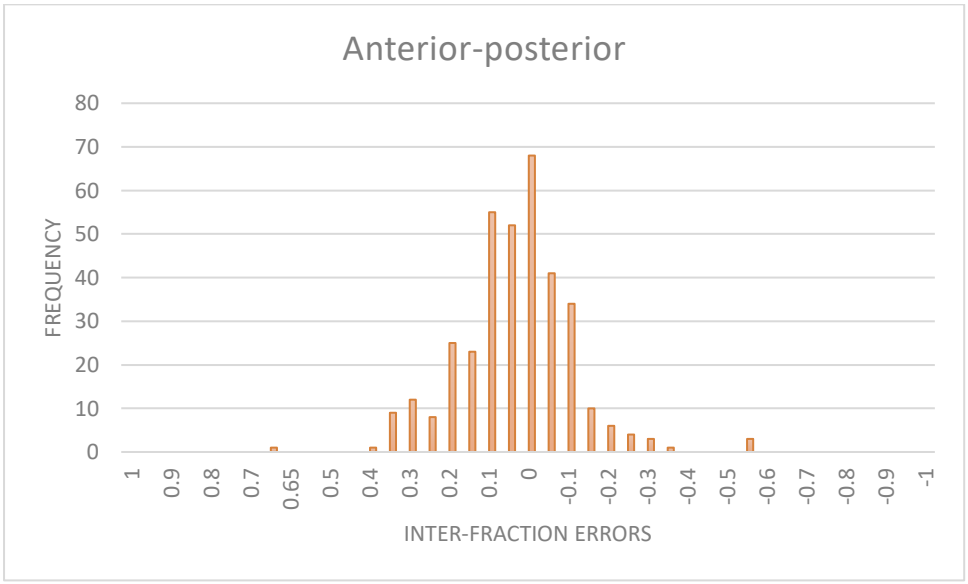
Frequency histograms of total inter-fraction errors were plotted to assess the magnitude and direction of inter-fraction errors for each translational direction and are shown in figure 4.8.a, b and c.



(a)



(b)



(c)

Figure 4.8: Frequency Histograms Showing the Distribution of Total Inter-fraction Errors (cm) for all fractions. a) Left-Right (X) inter-fraction errors. b) Superior-Inferior (Y) inter-fraction errors. c) Anterior-Posterior (Z) inter-fraction errors.

Systematic errors cause a shift in the cumulative dose and can result in a geometric miss, both of which have negative effects on tumour control and increased side effects due to higher doses to organs at risk (Hargrave & Holt, 2017). A mean that differs from 0 indicates systematic errors in the group.

The frequency histograms (figure 4.8) of the translational direction in the X direction showed a normal distribution curve with the 0 cm mark in the centre, indicating that the systematic errors did not favour any direction and were almost negligible; in fact, the population systematic errors recorded for the X directions were 0.1 cm. The population systematic error in the Z direction was also 0.1 cm, but the frequency histogram was slightly positively skewed, indicating that most errors were slightly more in the anterior direction than the posterior direction. The population systematic error in the Y direction was 0.26 cm and this error was represented on the frequency histogram as a slight positive skewness, with the median data point set around 0.05 cm. The frequency histogram of the superior-inferior direction was also wider, indicating a larger SD and, as a result, more dispersed errors in the super-inferior direction. This was also demonstrated when calculating the SD of total inter-fraction errors in the translational direction, since the results of SD were 0.13 cm, 0.32 cm, and 0.14 cm in the left-right, superior-inferior, and anterior-posterior directions, respectively.

The population random errors were 0.02, 0.07, and 0.03 cm in the left-right, superior-inferior, and anterior-posterior directions, respectively. Random errors distort the cumulative dose and are thought to have less of an impact on the planned dose than systematic errors (Hargrave & Holt, 2017). However, these must be minimised and, where possible, corrected. The most effective way to reduce random errors is to perform daily online corrections.

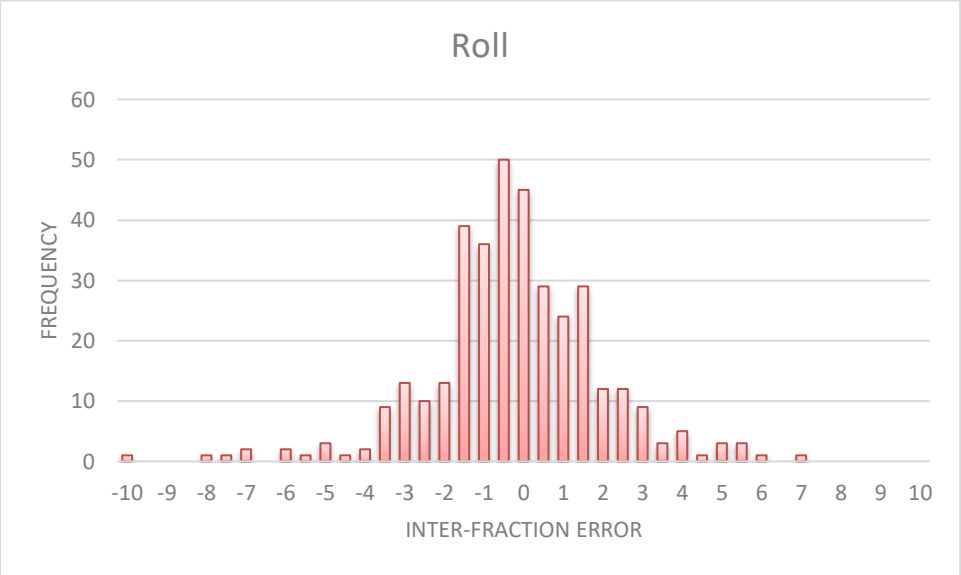
The population mean errors, population systematic errors and population random errors of rotational errors are demonstrated in table 4.21.

Table 4.21. Summary of Population Rotational Total Inter-fraction Errors

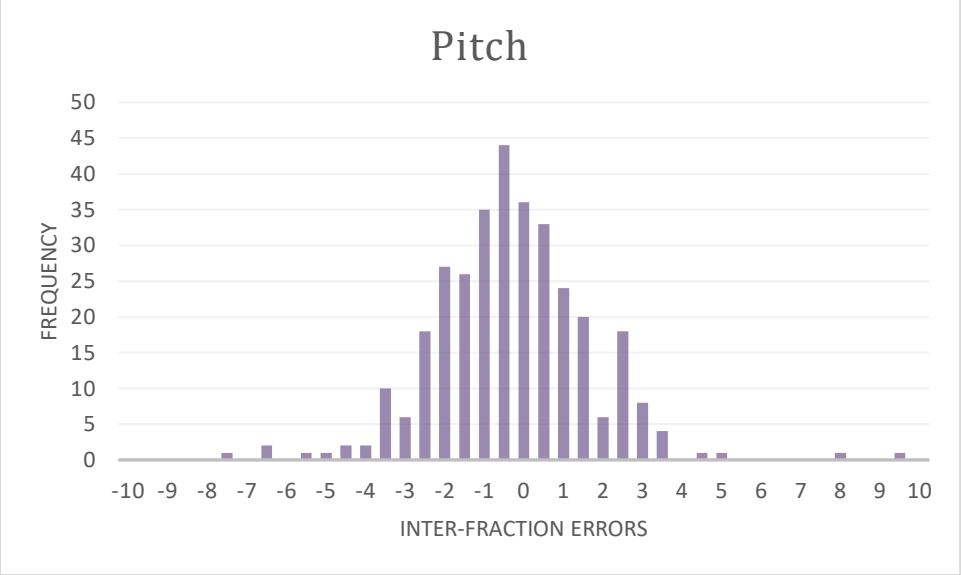
	RX (º)	RY (º)	RZ (º)
M	-0.3	-0.5	-0.08
SD	2.3	1.8	1.5
Minimum	-10.7	-7.5	-9.5
Maximum	6.6	9.5	4.0
∑	1.6	1.26	1.18
σ	0.57	0.81	0.3

M, mean of all patients' mean; SD, Standard Deviation; Σ, systematic set-up and organ motion errors; σ, random set-up and organ motion errors; RX, Roll; RY, Pitch; RZ, Yaw.

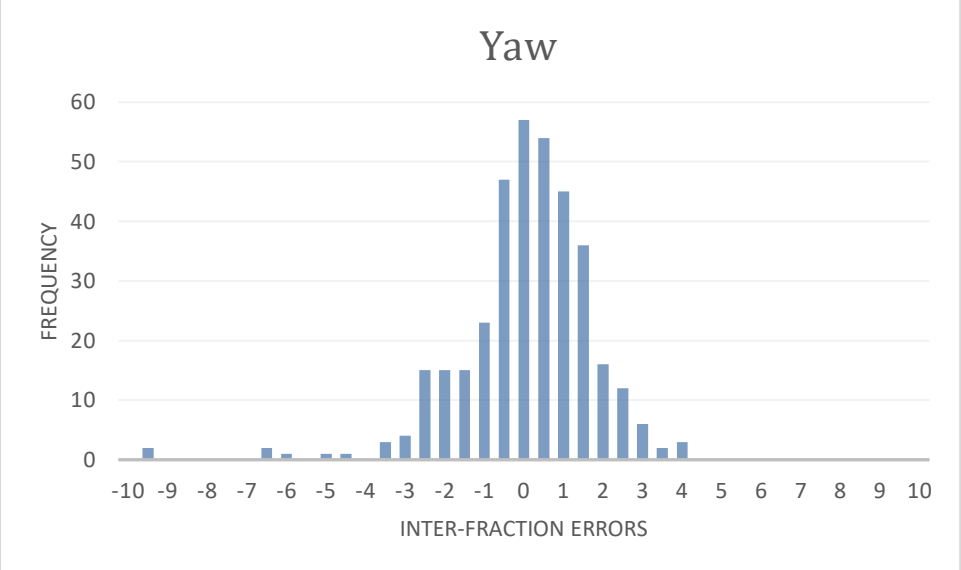
Frequency histograms of rotational total inter-fraction errors are shown in figure 4.9 a, b and c.



(a)



(b)



(c)

Figure 4.9: Frequency Histograms Showing the Distribution of Total Inter-fraction Errors in the Rotational Direction (º) for all fractions. a) Roll (Rx) inter-fraction errors. b) Pitch (RY) inter-fraction errors. c) Yaw (RZ) inter-fraction errors.

The frequency histogram of the total inter-fractional rotational errors in the RZ direction had the median data point centred around 0º, whilst the other directions had the median data points centred around -0.5º. All rotational histograms have a slight

positive skewness. They were more widely distributed when compared to the frequency histograms in the translational direction; this was also represented by SD results of 2.3º, 1.8º and 1.5º (table 4.20), indicating that most rotational errors were in the clockwise direction.

Because previous studies did not assess inter-fraction errors based on set-up uncertainties and organ motion, the population systematic and random errors of this study cannot be truly compared. However, similar results were obtained in studies that chose to match with thyroid cartilage rather than bone. The population systematic and random inter-fraction errors in Perillo et al.'s (2021) study were 0.9, 1.3, and 0.6 mm, and 1.1, 1.3, and 0.7 mm in the X-, Y-, and Z-directions, respectively. The largest error was observed in the superior-inferior direction, as in this study. Kwa et al. (2015) study also had the highest recorded error in the superior-inferior direction, as the population systematic and random errors in the X-, Y-, and Z-directions were 0.9, 2.0, and 1.1 mm and 1.0, 1.6, and 1.0 mm, respectively. According to Osman et al. (2011), the largest error in the superior-inferior direction could be due to motion caused by breathing that is averaged in the CBCT scan to a multi-slice spiral CT scan that does not precisely reflect the breathing averaged position. Another possible explanation is that the slice thickness of the reference CT used to register the daily CBCT was 2.5 mm. This could lead to registration errors in that direction when dealing with small structures with dimensions comparable to the CT slice thickness. In this study, patients were advised to suppress swallowing during the CT planning procedure to reduce the systematic errors in the superior-inferior direction.

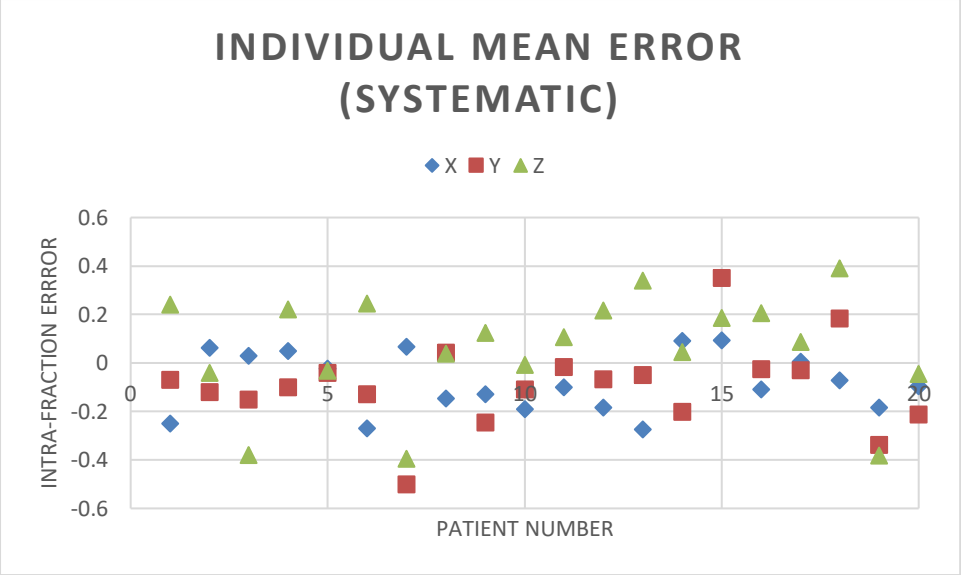
4.6 Intra-fraction errors

The larynx is a highly mobile organ since it moves during swallowing, breathing, and phonation. To assess intra-fraction errors, a total of 82 post-treatment CBCT scans were performed on the 20 participating patients. The post-treatment CBCT was taken immediately after treatment on the first treatment fraction and then once weekly. An offline match to the thyroid cartilage was performed by the participating radiographers on the treatment unit. The offline match took place on the day of the post-treatment CBCT acquisition. This strategy ensured that in subsequent fractions, when possible, the necessary advice to limit motion could be given to patients. A mask image registration was performed, with a margin of 0.5 cm surrounding the PTV considered as the region of interest, and when necessary, a manual registration was performed to match to the thyroid cartilage. The obtained registration results were deducted from the pre-treatment CBCT scan mask registration, and the resulting values were used to measure intra-fraction errors in the translational and rotational axis.

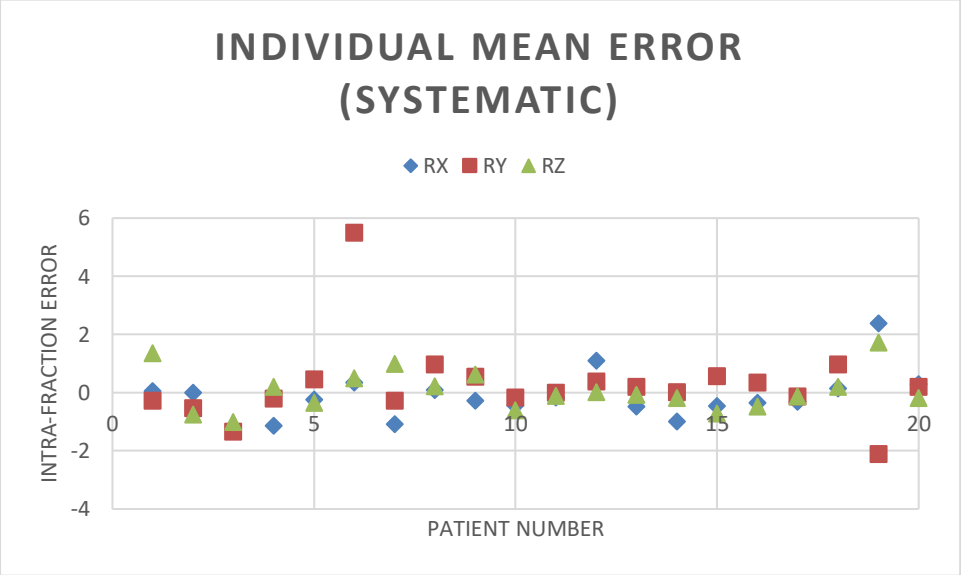
4.6.1 Results and discussion

Individual mean and individual random errors of intra-fraction errors

The scatter graphs presented in figure 4.10. a and b, and figure 4.11. a and b, demonstrate the individual mean (systematic) and individual random errors of intra-fractional errors in the translational and rotational direction.

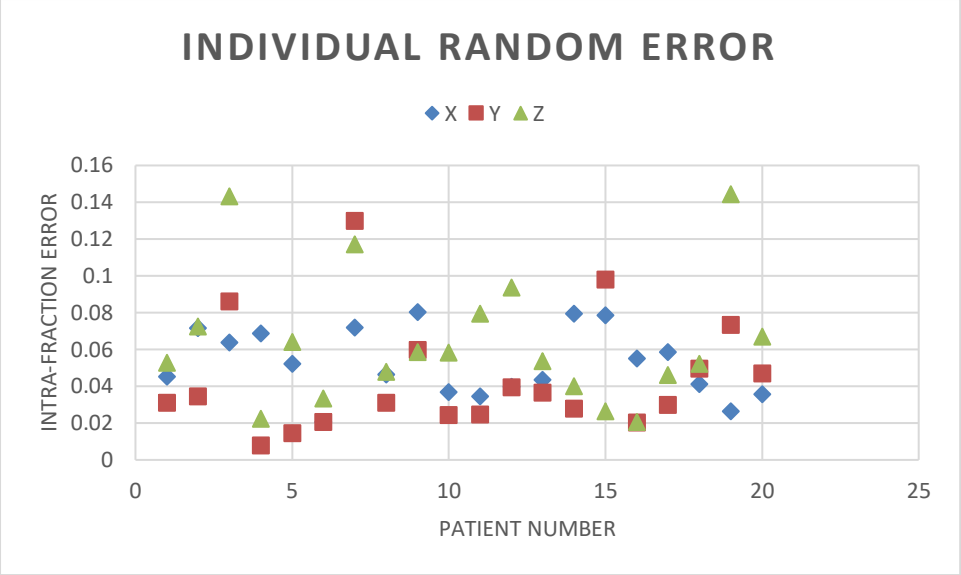


(a)

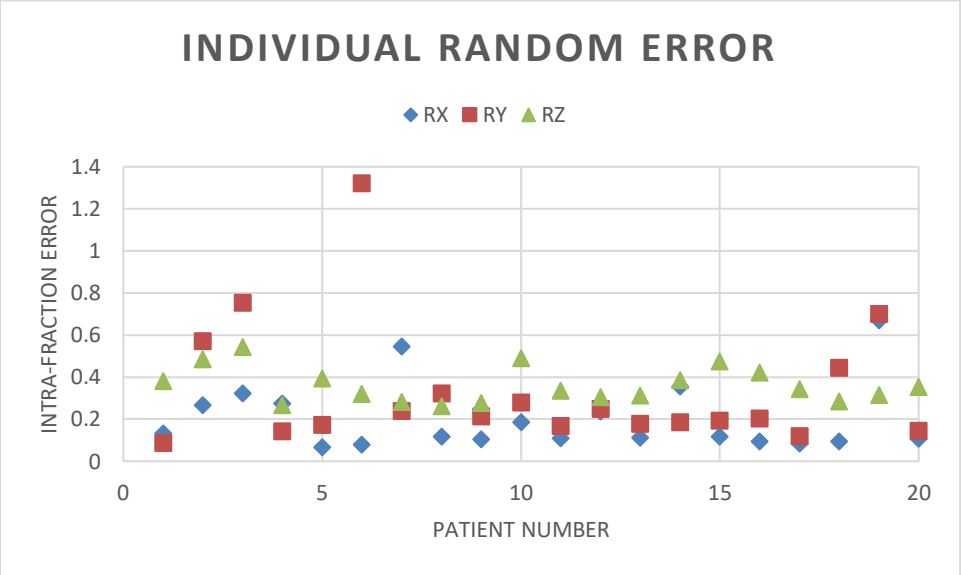


(b)

Figure 4.10: Scatter Graphs of the Individual Mean Errors of Intra-fraction Errors. (a) Translational direction (cm). (b) Rotational direction (^{o}).



(a)



(b)

Figure 4.11: Scatter Graphs of the Individual Random Errors of Intra-fraction Errors. (a) Translational direction (cm). (b) Rotational direction (^{o}).

The individual mean of intra-fraction errors in the X-, Y-, and Z-directions ranged between -2.8 mm (right) and 0.9 mm (left), -5 mm (inferior) and 3.5 mm (superior), and -3.8 mm (posterior) and 3.9 mm (anterior), respectively (figure 4.10, a). The largest uncertainty of the individual mean error in the translational direction was found in the inferior direction and was observed in patient 7. According to the radiotherapy clinical notes, this patient was very anxious throughout his treatment, and despite taking anti-

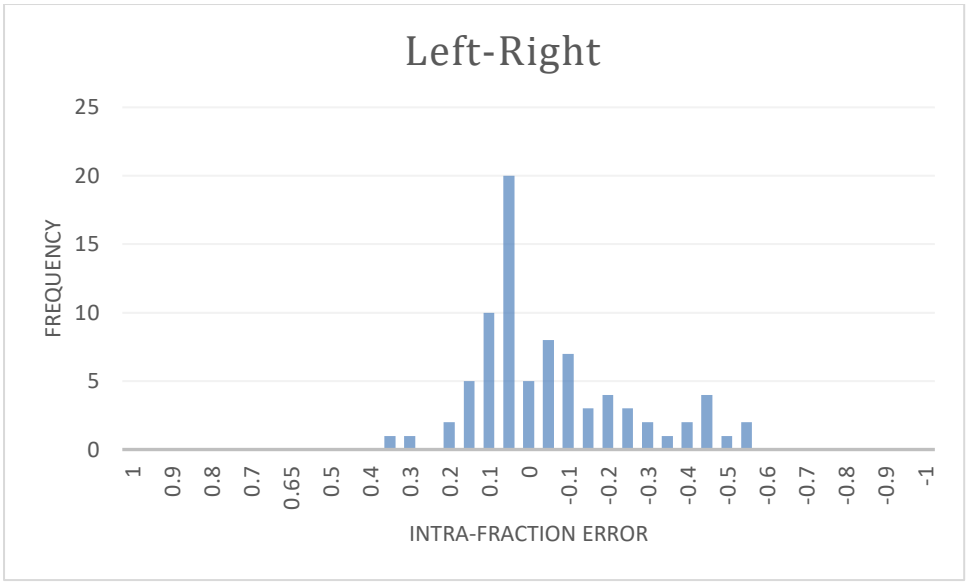
anxiety medications prior to treatment, the radiographers noticed that he swallowed frequently during CBCT exposure. To increase comfort and try to reduce anxiety, a thermoplastic mask with a large cut-out hole for the nose, which also extended to uncover the eyes, was used, and this could have reduced the strength and rigidity of the thermoplastic mask, resulting in some patient movement during treatment. This was also noted in Mulla et al. (2020) study, which assessed the setup reproducibility in the radiation treatment of patients receiving radiotherapy treatment to the head and neck using open face head and shoulder masks with customised versus standard closed head and shoulder masks. It was found that close-faced masks resulted in less set-up errors than open-faced masks.

The largest rotational error of individual mean errors was 5.5 ^{o} in patient 6 (figure 4.10, b). This patient was re-scanned due to rotation, and the data obtained after the re-scan suggested that the patient was possibly moving during treatment or holding the neck muscles stiff during CT planning and/or treatment.

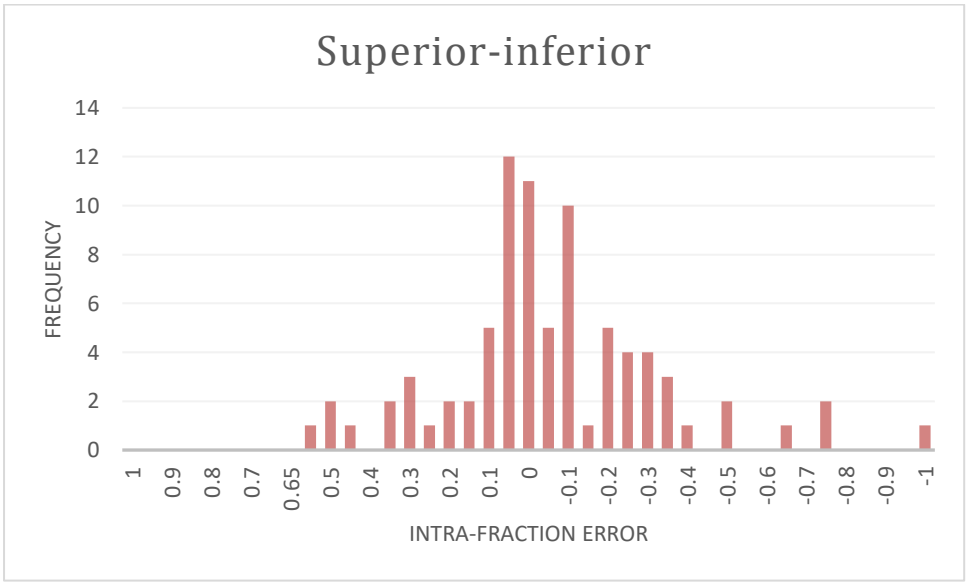
Individual random errors of intra-fraction errors

The individual random errors of intra-fraction errors in the X-, Y-, and Z-directions ranged between 0.3 mm and 0.8 mm, 0.1 mm and 1.3 mm, and 0.2 mm and 1.4 mm, respectively. There was an overall smaller discrepancy in the mean individual random errors (figure 4.10) when compared to the individual mean errors (systematic), with the left-right direction analysed as having the least random errors.

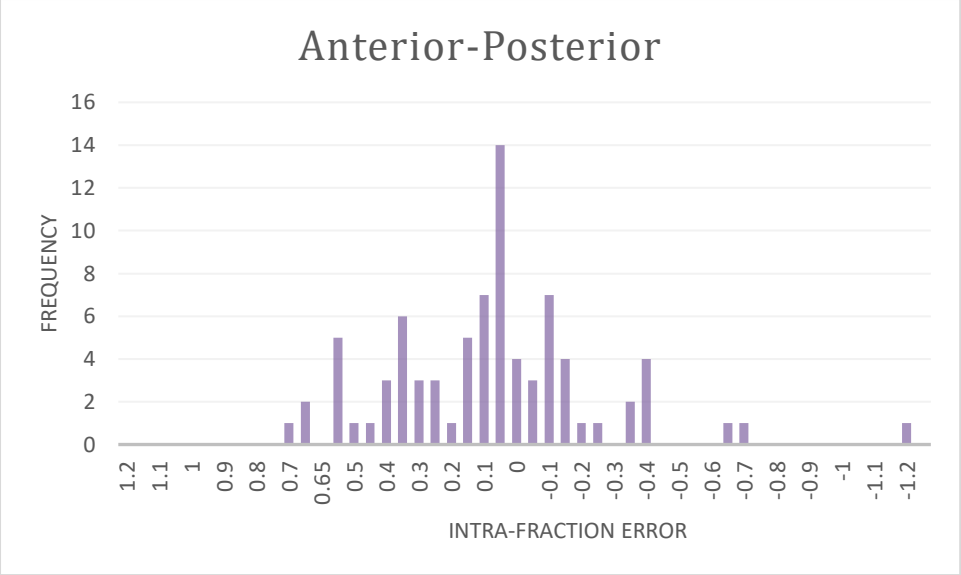
Frequency histograms of intra-fraction errors were plotted to assess the magnitude and direction of intra-fraction errors for each translational direction and are shown in figure 4.12.a, b and c.


Left-Right
FREQUENCY
25
20
15
10
5
0
1
0.9
0.8
0.7
0.65
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1
INTRA-FRACTION ERROR

(a)


Superior-inferior
FREQUENCY
14
12
10
8
6
4
2
0
1
0.9
0.8
0.7
0.65
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1
INTRA-FRACTION ERROR

(b)

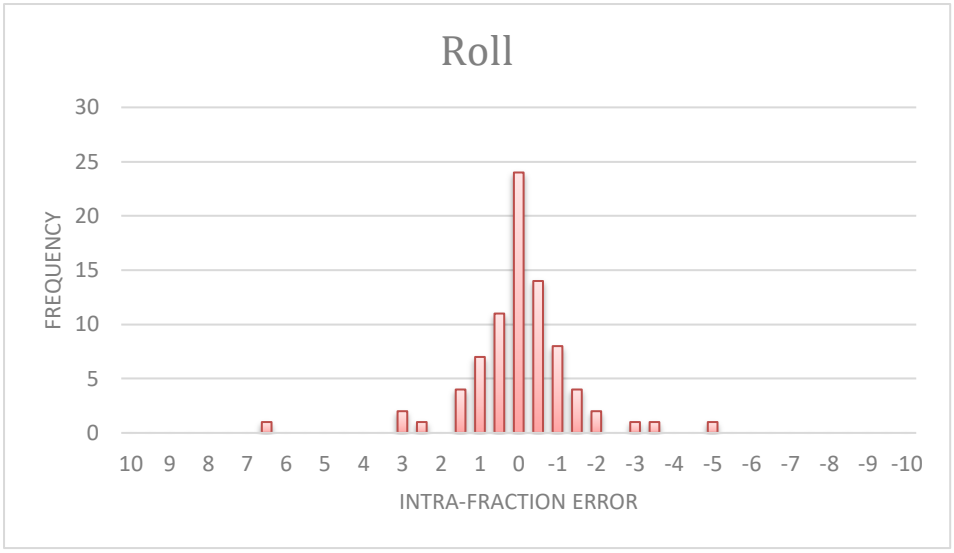


(c)

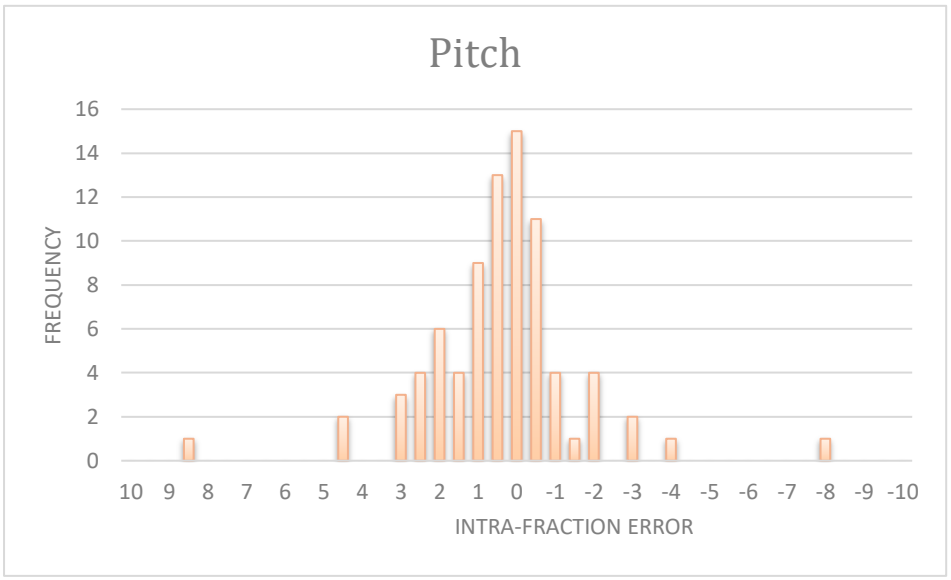
Figure 4.12: Frequency Histograms Showing the Distribution of Translational Intra-fraction Errors (cm). a) Left-Right (X) inter-fraction errors. b) Superior-Inferior (Y) inter-fraction errors. c) Anterior-Posterior (Z) inter-fraction errors.

The frequency histogram for the translational intra-fractional errors in the left-right and superior-inferior direction were negatively skewed, with the median value set around 0.05 cm. This indicated systematic errors trend in the right and inferior directions. Whereas the anterior-posterior frequency was positively skewed, with the mean value set around 0.05 cm, indicating that errors were mostly directed in the anterior direction. The left-right frequency histogram had a narrow spread when compared with the others, which indicated that it had the least random errors.

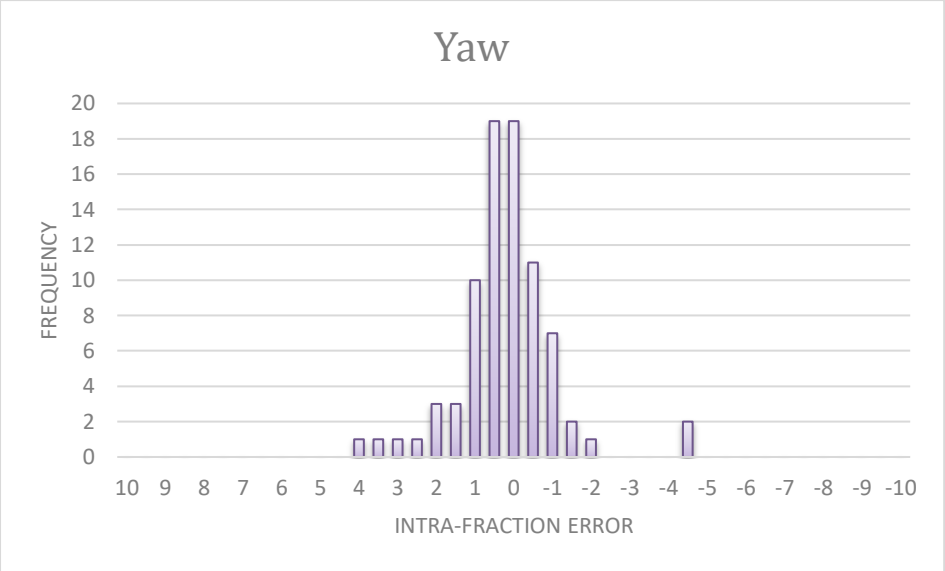
The frequency histograms of rotational total inter-fraction errors are shown in figure 4.13. a, b and c.


Roll
30
25
20
15
10
5
0
FREQUENCY
10 9 8 7 6 5 4 3 2 1 0 -1 -2 -3 -4 -5 -6 -7 -8 -9 -10
INTRA-FRACTION ERROR

(a)


Pitch
16
14
12
10
8
6
4
2
0
FREQUENCY
10 9 8 7 6 5 4 3 2 1 0 -1 -2 -3 -4 -5 -6 -7 -8 -9 -10
INTRA-FRACTION ERROR

(b)



(c)

Figure 4.13: Frequency Histograms Showing the Distribution of Rotational Intra-fraction Errors (^{o}) for all fractions. a) Roll (Rx) intra-fraction errors. b) Pitch (RY) intra-fraction errors. c) Yaw (RZ) intra-fraction errors.

The frequency histogram for the rotational intra-fractional errors in the roll and pitch direction showed a normal distribution, while the yaw was slightly positively skewed. This indicated that the rotational direction was more commonly seen in the clockwise direction.

Population systematic and random errors

The population mean (M), standard deviation (SD), population systematic errors and population random errors of intra-fraction errors were calculated for each patient and are represented in table 4.22.

Table 4.22. Summary of Intra-fraction Errors

	X (cm)	**Y** (cm)	**Z** (cm)	**RX** (º)	**RY** (º)	**RZ** (º)
M	-0.08	-0.09	0.06	-0.13	0.26	0.07
SD	0.19	0.31	0.32	1.43	1.99	1.31
Minimum	-0.59	-1.7	-1.26	-5.3	-8.31	-4.7
Maximum	0.34	0.54	0.66	6.5	8.3	3.5
∑	0.12	0.18	0.23	0.82	1.43	0.7
σ	0.05	0.04	0.06	0.2	0.34	0.36

M, mean of all patients' mean; SD, Standard Deviation; Σ, population systematic intra-fraction errors; σ, population random intra-fraction errors; X, left–right; Y, superior–inferior; Z, anterior–posterior; RX, Roll; RY, Pitch; RZ, Yaw.

Comparison of results with other studies

A study by Perillo et al. (2021) analysed intra-fraction errors for 23 patients treated for early glottic cancer with VMAT with the scope of setting treatment margins for Stereotactic Ablative Radiotherapy (SABR). Pre-treatment and post-treatment CBCTs were taken using the thyroid cartilage as a matching structure. The resulting population systematic and random errors for intra-fraction errors were 0.7, 1.6 and 0.7 mm and 1.0, 1.5 and 0.6 mm in the X, Y and Z directions, respectively. Kwa et al. (2015) also opted to measure intra-fraction errors using the thyroid cartilage as a matching structure and also obtained low values of population systematic errors of 0.4, 1.3 and 0.7 mm, respectively, in the X, Y and Z directions, and population random errors of 0.8, 1.4, and 0.8 mm, respectively in the X, Y and Z direction. These results were comparable to those of Perillo et al.'s (2021). Contrary to the results obtained in our study, both Perillo et al (2021) and Kwa et al (2015) studies found that the most significant errors were those in the superior-inferior direction. According to Perillo et al. (2021), glottic intra-fraction motion in the superior-inferior direction was linked to laryngeal intrinsic mobility, which the thermoplastic mask cannot prevent.

The lower registered errors in this study and those of Kwa et al. (2015) and Perillo et al. (2021) could be attributed to the fact that all studies suppressed swallowing during

the CT planning phase. Perillo et al. (2021) and Kwa et al. (2015) studies also went a step further by suppressing swallowing during CBCT acquisition and treatment. As a result, the intra-fraction errors were caused primarily by resting displacement, which is less than deglutition-induced displacement (Bradley et al., 2011), because the amplitude of swallowing can be measured to be 23 mm and 6 mm in the superior and anterior directions, respectively, whereas the mean breathing movement was found to be 4 mm and 2 mm in the superior-inferior and antero-posterior directions, respectively (Bahig et al., 2017). Our study did not suppress swallowing during CBCT acquisition and treatment delivery, which may have resulted in slightly higher population systematic errors than Perillo et al. (2021) and Kwa et al. (2015). However, in our study, all patients were treated with a maximum neck extension, which aids in limiting swallowing during treatment (Perillo et al., 2021). Also, swallowing during treatment is generally rare and fast; therefore, its impact on the dose distribution is minimal (Durmuş, Taş, and Uzel. 2020).

4.7 Observer variation in image matching errors

To our knowledge, this is the first study to investigate observer variation in image matching using bone-registration for cancer to the larynx. Very few studies addressed uncertainties in observer variation in using image guided radiotherapy, therefore comparison of results was limited.

4.7.1 Results and discussions

Inter-observer variation

The six radiographers who agreed to participate in the study performed image matching retrospectively on twenty-five CBCT scans. Figures 4.14 - 4.19 show the results of the recorded inter-observer residual errors when comparing the CBCT scans to the corresponding CT planning scans in each translational and rotational direction. The lack of inter-observer variation would have been demonstrated on the scatter graphs as a complete superimposition of the shapes representing the matching of the six participating radiographers.

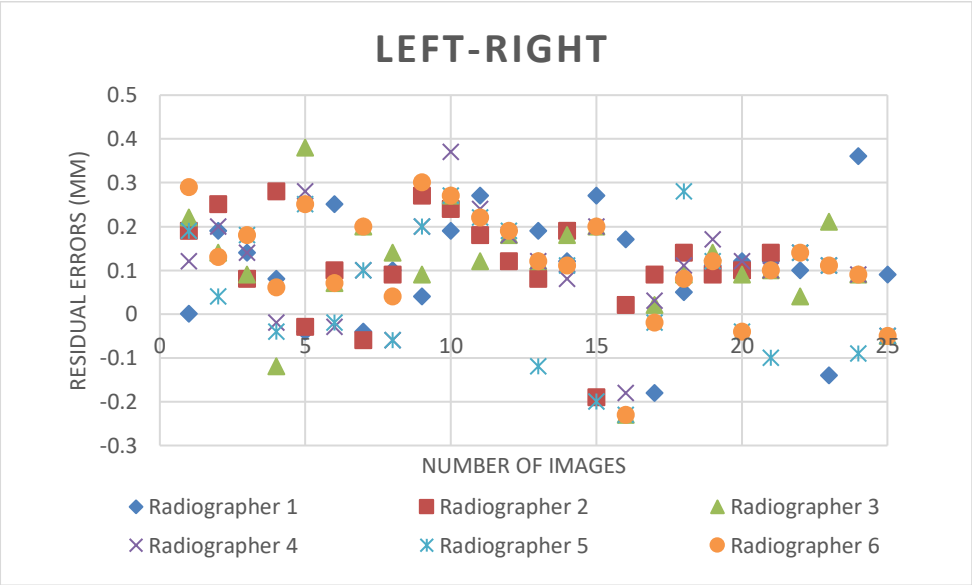


Figure 4.14. Residual errors in the left-right direction.

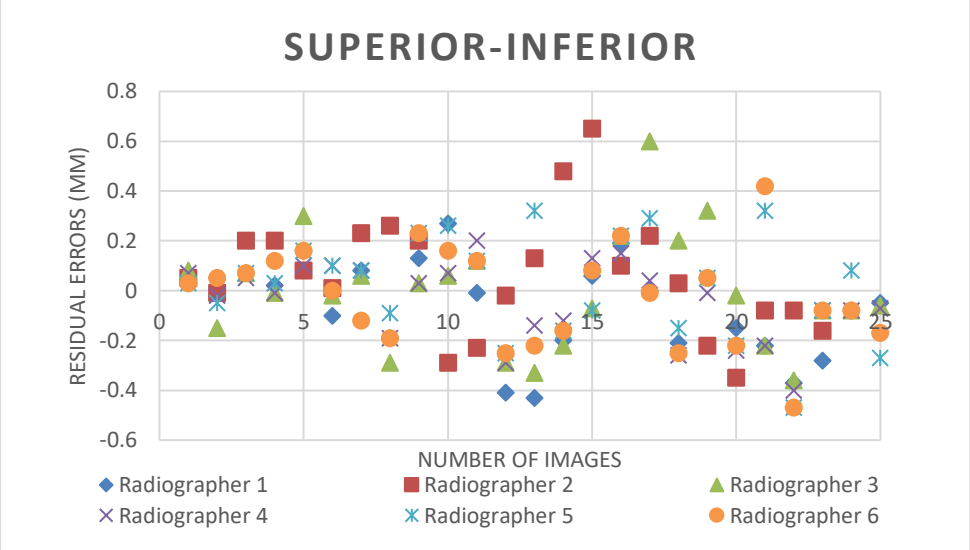


Figure 4.15: Residual Errors in the Superior-inferior Direction.

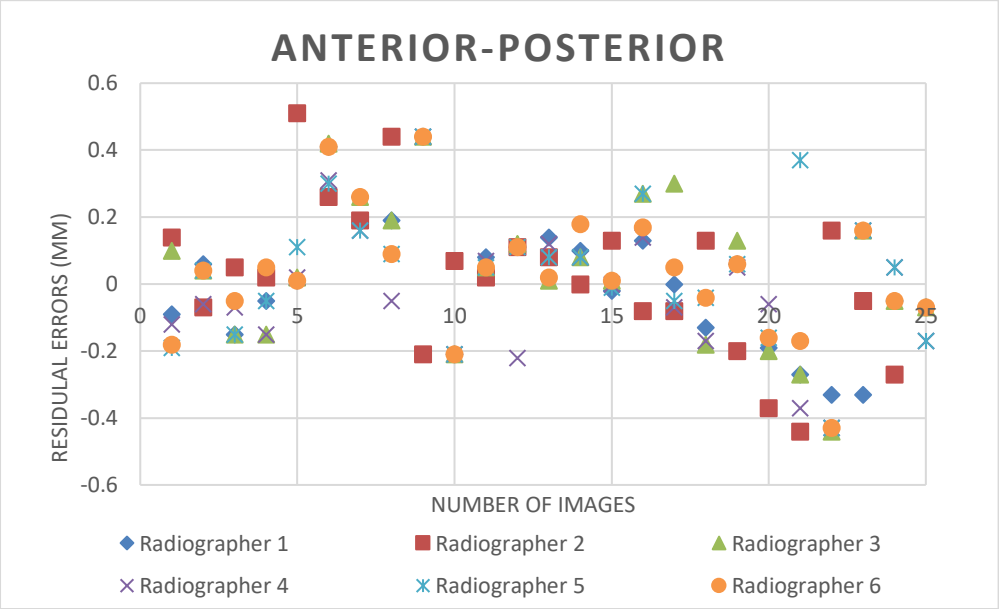


Figure 4.16: Residual Errors in the Anterior-posterior Direction

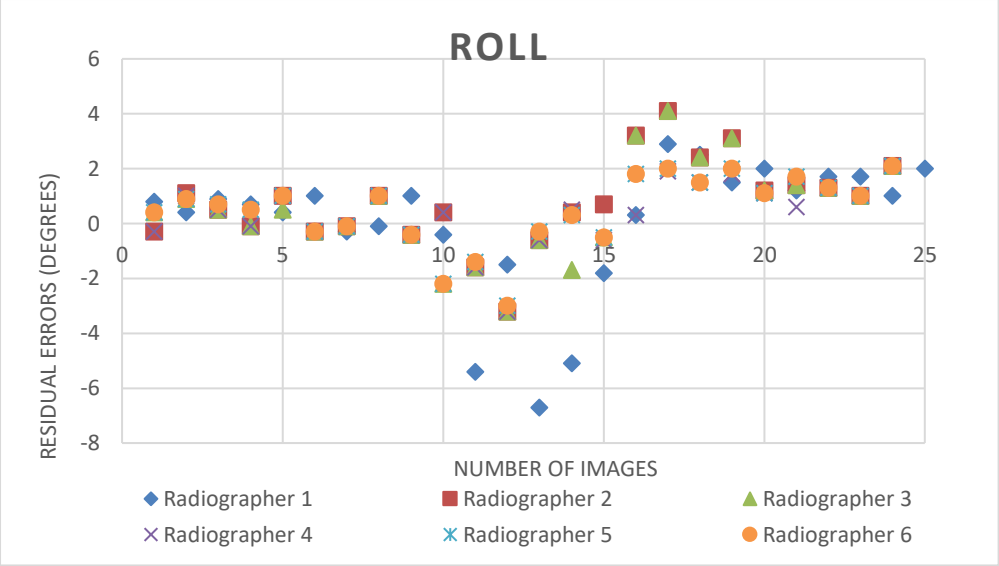


Figure 4.17: Residual Errors in the Roll Direction

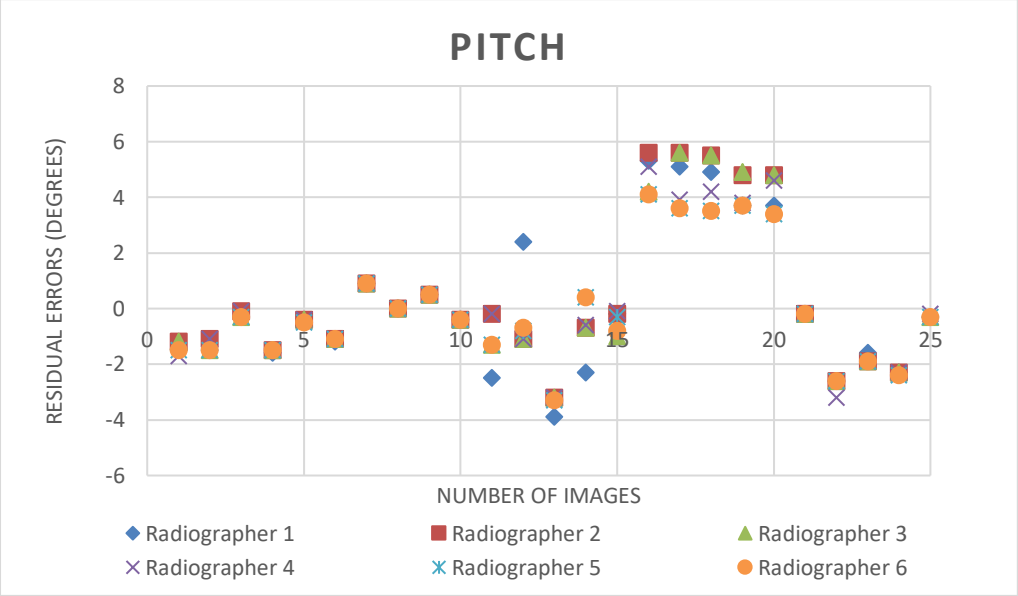


Figure 4.18: Residual Errors in the Pitch Direction

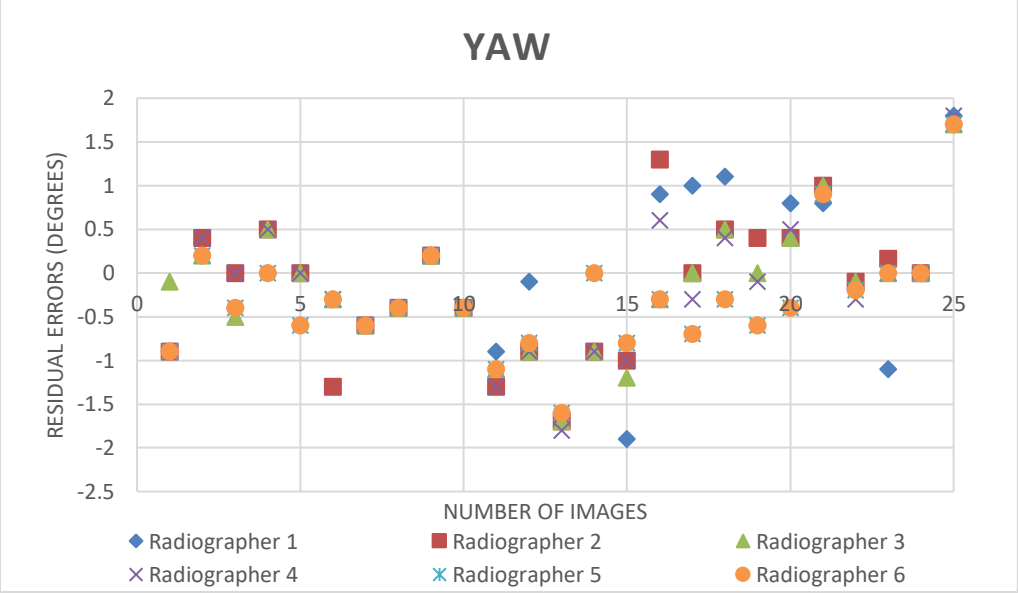


Figure 4.19: Residual Errors in the Yaw Direction

Variations of the residual error in the translational and rotational direction were overall largest in Y and RX directions, respectively, however, the largest discrepancy of residual errors were found in the Z direction, and this ranged from -0.44 to 0.37 mm. Variations in the translational and rotational directions were the smallest in the X and RZ directions, respectively.

The mean individual systematic errors of inter-observer variations in the X, Y, and Z directions ranged from 0.02 – 0.06 mm, 0.04 – 0.3 mm, 0.05 – 0.2 mm, respectively, and 0.5 – 4.6º, 3.6 – 6.2º, and 0.2 – 0.4º, respectively in RX, RY, RZ directions. The population systematic errors were analysed to be 0.02, 0.13 and, 0.2 mm, respectively in the X, Y, Z directions, and 4.6º, 5.1º and 0.3º, respectively in the RX, RY, and RZ directions.

The random errors of the inter-observer variations for each observer in X, Y, and Z directions ranged from 1.0 – 1.6 mm, 1.5 – 2.3 mm, 1.5 – 2.0 mm, respectively, and 0.8 – 1.2º, 1.0 – 1.6º and 0.4 – 0.6º, respectively in the RX, RY, and RZ directions.

The population systematic errors and population random errors of inter-observer variation errors in image matching were calculated for each patient and are represented in table 4.23.

Table 4.23. Population Systematic and Random Errors of Inter-observer Variation Errors

	X (mm)	**Y** (mm)	**Z** (mm)	**RX** (^{o})	**RY** (^{o})	**RZ** (^{o})
∑	0.02	0.13	0.2	4.6	5.1	0.3
σ	1.2	1.9	1.7	1	1.3	0.5

Σ, population systematic inter-observer errors; σ, population random inter-observer errors; X, left-right; Y, superior–inferior; Z, anterior–posterior; RX, Roll; RY, Pitch; RZ, Yaw.

For this study, the population systematic errors of inter-observer variation were 0.02, 0.13 and 0.2 mm in the X, Y, and Z directions, and 4.6, 5.1, and 0.3 ^{o} in the RX, RY and RZ directions (Table 4.23). These values were overall lower than those obtained by Hirose et al. (2020), where values of population systematic errors of inter-observer variations in X, Y and Z directions were 0.9, 0.9, and 0.5 mm. The population random errors for this study (0.4, 0.3 and 0.2 mm, in the X, Y and Z direction, respectively) were also less than those obtained in Hirose et al. (2020) study with values of 1.8, 2.2, and 1.1 mm, respectively in the X, Y, and Z direction.

The participating radiographers performed a bone match for image registration, whereas Hirose et al. (2020) assessed the inter-observer variation in the matching of the prostate target. This may explain why the current study shows less variation. Soft tissue image matching to the prostate could result in a larger image matching discrepancies because the prostate gland is subject to anatomical changes such as bladder filling and rectal changes that cannot be completely corrected, therefore, image matching subjectivity could be higher (Bell et al., 2019).

Another reason for obtaining lower values could have been attributed to the region that was investigated. A study by Mohamoud, Ryan and Moseley (2015), with the aim of understanding inter-observer variability in image matching among radiographers for various sites, found that the pelvic and head and neck regions displayed the least inter-observer variability while thorax and abdominal cases had the most when matching with CBCT. This demonstrated that observer variability was also dependable on the specific site that was investigated.

Unlike Hirose et al.'s (2020) study, this study also evaluated the population systematic and random errors of rotational direction. The findings revealed that rotational errors in observer variation in image matching were predominantly systematic. The residual rotational errors were purely from automatic matching since the radiographers in the local oncology hospital did not perform manual registration to account for rotation; therefore, the variation in rotation from one radiographer to the next could be attributed to differences in clip-box size and position since the bone matching chamfer algorithm is sensitive to the size and position of the region of interest (Sousa et al., 2021).

When using registration of bony anatomies with the planning CT, the registration accuracy could be affected by image resolution and region of interest (clip-box) employed for registration (Yin et al., 2013). It was noted that the clip-box position varied from one radiographer to the next, and that the clip-box position was not always placed as per local clinical protocols. Some radiographers placed the clip-box to cover part of the mandible, which is not considered to be a stable anatomy and therefore should not be included in the region of interest (The Royal College of Radiologists, 2021). According to Osman (2011), the effects of inter-observer variation in selecting a clip-box around the volume of interest were found to be negligible when clear guidelines for which structures to include in the clip-box were followed.

4.8 PTV margin calculation

This study accurately evaluated the PTV margin size for such patients using measurements applicable to the local oncology department. PTV margin was calculated using the van Herk Formula: Margin = $2.5\sum + 0.7\sigma$, where the $\sum$ denoted the quadratic sum of the total assessed population systematic errors (target delineation observer variation, inter-fraction errors, intra-fraction errors and inter-observer variation in image matching), and the quadratic sum of total population random errors were denoted by σ (inter-fraction errors, intra-fraction errors and inter-observer variation in image matching).

To our knowledge, this study was unique for measuring all the aforementioned errors to calculate the PTV margin for patients treated to the larynx with VMAT.

4.8.1 Results and discussions

The population systematic errors that were obtained in the translational direction and the analysed errors for the calculation of the PTV margin are shown in table 4.24. These data are presented as a bar graph for comparison of results in figure 4.20.

Table 4.24. Translational Population Systematic Errors

Population Systematic Errors (mm)			
Errors	**X**	**Y**	**Z**
Target Delineation	3.47	6.92	3.47
Inter-fraction	0.99	2.6	0.96
Intra-fraction	1.22	1.8	2.3
Inter-observer IM	0.02	0.13	0.2

X, left–right; Y, superior–inferior; Z, anterior–posterior; IM, Image matching.

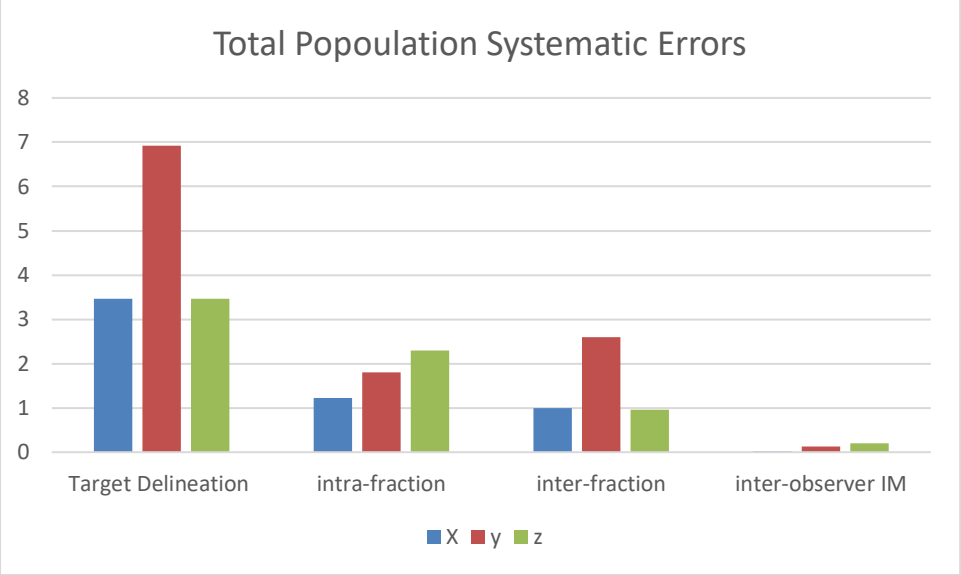


Figure 4.19: A Bar Graph Representing the Population Systematic Errors of all the Analysed Errors

Figure 4.19. demonstrates that the target volume delineation errors had the highest obtained values of population systematic errors, and the smallest error was always observed in the left-right (X) direction.

Table 4.25. displays all the analysed population random errors values, obtained in the translational direction and figure 4.20 presents these results in the form of a bar graph for comparison of values.

Table 4.25. Translational Population Random Errors

Population Random Errors (mm)			
Errors	**X**	**Y**	**Z**
Target Delineation	0	0	0
Inter-fraction	0.22	0.68	0.27
Intra-fraction	0.53	0.44	0.64
Inter-observer IM	1.2	1.9	1.7

X, left–right; Y, superior–inferior; Z, anterior–posterior; IM, Image matching.

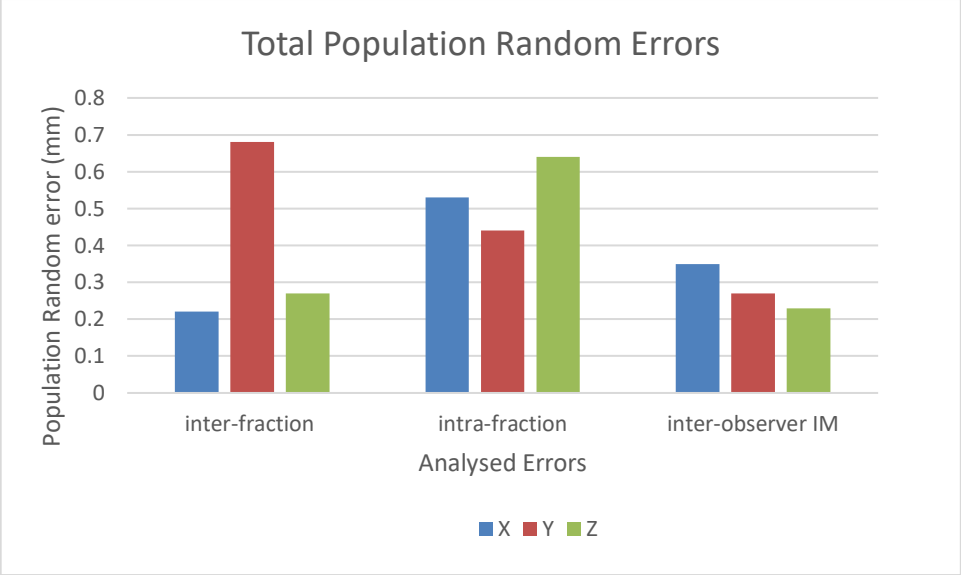


Figure 4.20: A Bar Graph Representing the Population Random Errors of all the Analysed Errors

Figure 4.20. clearly shows that inter-observer variation in image matching errors had the highest obtained values of population random errors.

Values of total population systematic and random errors where then used to calculate the PTV margin for patients treated to the larynx with VMAT in the local oncology radiotherapy department.

PTV margin result

Table 4.26. shows the total population systematic and random errors values, and the obtained PTV margin result from all the analysed errors.

Table 4.26. PTV Margin Result

	PTV margin (mm)		
	X	**Y**	**Z**
Population Systematic Errors	9.5	19.0	10.7
Population Random Errors	0.93	1.45	1.29
PTV margin	**10.5**	**20.5**	**12.0**

X, left–right; Y, superior–inferior; Z, anterior–posterior; PTV, Planning Target Volume

The margin obtained in this study was that of **10.5, 20.5** and **12 mm** in the X, Y and Z directions, respectively. This margin was large when compared to previous studies that evaluated the PTV margin size for laryngeal cancer.

Table 4.27 shows the PTV margin result without considering the Target Volume Delineation Errors.

Table 4.27. PTV Margin Result without Target Volume Delineation Errors

	PTV margin (mm)		
	X	**Y**	**Z**
Population Systematic Errors	3.9	7.9	6.3
Population Random Errors	0.9	1.5	1.3
PTV margin	**4.9**	**9.4**	**7.5**

X, left–right; Y, superior–inferior; Z, anterior–posterior; PTV, Planning Target Volume

Table 4.28. compared previous studies with the current studies to determine the reason for obtaining a larger margin size.

Table 4.28. Comparison of Studies

Study	Sample size	Treatment type	Immobilisation	Frequency of imaging	Assessed errors	PTV margin
Current study	20	VMAT	5-point TP mask (Orfit) and standard but individually selected head rest	Daily CBCT	- Target Delineation - Inter-fraction - Intra-fraction - Inter-observer variation in IM	With Target Delineation X = 10.5 mm Y = 20.5 mm Z =12 mm Without Target Delineation X= 4.9 mm Y = 9.4 mm Z = 7.5 mm
Kwa et al., 2015	42	IMRT	5-point TP mask	Daily CBCT	- Inter-fraction - Intra-fraction	X = 1.6 mm Y = 4.3 mm Z = 2.2 mm
Perillo et al., 2021	23	SBRT	5-point TP mask and a bite block	Daily CBCT	- Inter-fraction - Intra-fraction	X = 2.4 mm Y = 5.1 mm Z = 2.2 mm
Kukolwicz et al., 2021	30 (larynx and nasopharynx	IMRT or VMAT	5-point TP mask (Orfit) and standard but individually selected head rest	Daily CBCT	- Inter-fraction	X = 4 mm Y = 6 mm Z = 4 mm

IMRT- Intensity Modulated Radiation Therapy, VMAT – Volumetric Modulated Arc Therapy, SBRT – Stereotactic Body Radiation Therapy, CBCT – Cone beam CT, TP – Thermoplastic, IM – Image Matching, X – Left-Right, Y – Superior-Inferior, Z – Anterior-Posterior

All studies assessed patients treated with a 5-point TP mask, and patients who had daily CBCT imaging. The studies varied with regards to the sample size, type of treatment used, and errors analysed.

There could have been several reasons for the higher margin in this study when compared to other similar ones. As discussed in Chapter 2, the margin size is influenced by anatomical area variation, imaging frequency, immobilisation equipment type, treatment modality, patient collaboration, and set-up procedures (Kapanen et al., 2013; Anjanappa et al., 2017). However, these were similarities between the current study and the studies in table 4.28.

Similarities in the immobilisation devices was observed in all studies since all patients were treated with a 5-point thermoplastic mask, however in Perillo et al's. (2021) study the patients were also treated with a bite block. Bite blocks are considered as an aggressive type of immobilisation, and they are particularly effective in eliminating most head rotations (The Royal College of Radiologists, 2021). The combination of two immobilisation devices, such as a thermoplastic mask and a bite block, was found to be more effective for reproducibility in head and neck treatment positions (Ingrosso et al., 2012). One of the reasons for the smaller PTV margin obtained in Perillo et al's (2021) study could have been attributed to this type of set-up since the population systematic errors of inter-fraction errors were found to be 0.9, 1.3 and 0.6 mm in the X, Y and Z directions, whilst this study had a slightly larger population systematic errors of 1, 2.6 and 1 mm in the X, Y and Z directions.

The sample size was sufficient for all patients in the examined studies since every study had a sample size of at least 20 patients with more than 5 images per patient. A smaller sample size could have resulted in uncertainties in the population systematic errors (The Royal College of Radiologists, 2021).

Kukolwicz et al. (2021) did not however identify the proportion of patients who underwent treatment to the nasopharynx and larynx; instead, errors were calculated jointly for both regions, and this anatomical variation could have had a negative influence on the margin calculation (Kapanen et al., 2013).

Another possible reason for obtaining a larger PTV margin in this study was that most studies that examined PTV margin tended to focus on just the inter-fraction errors and neglected assessment of other errors (Chapter 2 – Section B). Other sources of errors such as target delineation errors and observe variation in image matching should be quantified to establish an appropriate PTV margin (The Royal College of Radiologists, 2021).

Although the traditional method of accounting for uncertainties was to generate the PTV by extending the CTV with a suitable safety margin, it was possible that in other studies, clinical oncologists delineated the CTV to account for laryngeal movement, even though organ movement should be part of the PTV and not of the CTV, as per ICRU definition. In fact, in a study by Williamson et al., (2016), the CTV was drawn to account for organ motion and deformation. In the local department, CTV was considered as just the delineation of tumour and other tissue with presumed subclinical spread (Sir Anthony Mamo Oncology Hospital, 2020).

In a study by Osman (2011), a small isotropic margin of 2 mm was found to be adequate for SVCI treated with IMRT. This tight margin was beneficial for voice-sparing and re-irradiating patients in the event of a local recurrence. This margin size was permitted due to recent technological advances in image acquisition with 4D-CT to account for organ motion in CT planning, image-guided verification with CBCT, and a TPS with Monte Carlo dose calculations. In the local oncology department, 4D-CT was not performed, and clinical oncologists did not perform SVCI procedures. Performing a 4D-CT would have been beneficial to observe the extent of the target movement, which could also have been compared to the CBCT acquisition and treatment delivery. Osman (2011), however, based the PTV margin on set-up errors, respiratory movements and inter-fraction anatomical variations and neglected to identify the effects of target volume delineation uncertainties and observer variation in image matching.

Summary of the analysed errors

The largest systematic errors that were obtained in this study were that of the observer variation in target volume delineation. Target volume delineation errors increasingly

make up the majority of total errors because they are often not rectified, unlike online daily imaging, which reduces the systematic and random errors of set-up errors (The Royal College of Radiologists, 2021). The local oncology hospital lacked clinical protocols of target delineation for head and neck patients. When there are no clinical protocols to be followed, the consistency of contouring suffers (Kim et al., 2021). If one had to ignore the target volume delineation errors from the PTV margin calculation, the resulting margin would be 4.9, 9.4 and 7.5 mm in the X, Y and Z directions, approximately half the actual margin size. Therefore, the target volume delineation errors contributed substantially to the margin size.

The second most significant errors were the intra-fraction errors. Although having the second overall largest population systematic errors and random errors, the significance of these errors on the PTV margin were much less than that of the target volume delineation errors. If these errors had to be subtracted from the margin calculation, the margin size would be 9.9, 19.9, and 10.3 mm in the X, Y, and Z directions, respectively, which is about 1 mm difference in the actual margin size.

The largest random errors were that of inter-observer variation in image matching. The significance of these errors had less of an impact on the margin size as observer variation in target volume delineation since random errors have a blurring effect on the dose distribution whilst systematic errors shift the cumulative dose distribution (Sonke and Van Herk, 2016).

Rotational errors

Systematic rotational errors might result in a dose distribution that does not coincide with the PTV as intended, whereas random rotational errors also blur the dose distribution in the PTV (Novak et al., 2021). Although the van Herk formula and other margin equations do not take the impact of rotational errors into account (Caruana et al., 2021), population systematic and random errors of rotation were still analysed for inter- and intra-fraction errors and observer variation in image matching to understand the effect and significance of such errors. Values of population systematic errors and

population random errors obtained from the analysed errors are shown in table 4.29. and 4.30., respectively.

Table 4.29. Total Population Systematic Errors for Rotation

Errors	**RX** (o)	**RY** (o)	**RZ** (o)
Total inter-fraction	1.6	1.26	1.18
Intra-fraction	0.82	1.43	0.7
Inter-observer variation IM	4.6	5.1	0.3

Table 4.30. Total Population Random Errors for Rotation

Errors	**RX** (o)	**RY** (o)	**RZ** (o)
Total inter-fraction	0.57	0.81	0.3
Intra-fraction	0.2	0.34	0.36
Inter-observer variation IM	1	1.3	0.5

The Pitch (RY) direction, which is the rotation of the patient's transversal axis, always had the largest errors in both the population systematic and random errors assessments. Since the cervical spine is flexible, the degree of pitch rotation increases with distance from the iso-centre, requiring a six-degree-of-freedom couch adjustment. Rotational errors in the pitch direction have little impact on the dose distribution of small spherical treatment volumes, but they may have an impact when treating long non-spherical volumes (Stieb et al., 2018).

4.9 Limitations of the study

Limitations of the target volume delineation errors assessment

To determine the standard deviation of a population with any reasonable precision, Tudor et al., (2020) recommended a sample size of at least 30 patients since standard deviation estimations from a small sample could be unreliable. In this study, six doctors analysed five CT scans, which resulted in a total of 30 CT scans being analysed. A bigger sample would have been preferred to obtain more reliable results, however due to time constraints and doctors' workload, the number of cases delineated by each of the six doctors had to be restricted to five.

Since the researcher was not present with the doctors during delineation, there was no way of knowing if the doctors discussed clinical cases used for analysis amongst themselves. According to Dewas et al. (2011), if doctors discuss clinical cases this could influence their interpretation of the CTV.

A possible limitation of this study could be that the doctors were not asked regarding difficulty levels in contouring targets on certain scans. This had been reported to be useful by Das et al., (2021) since it was proven that the most difficult scan to delineate was the one with the highest observer variability. This would also have been beneficial to determine whether experience and training influenced the target delineation errors.

Limitations of the total inter-fraction errors assessment

The thyroid cartilage was used as a surrogate for matching to the target volume due to the lack of CBCT contrast of the target volume when compared to the thyroid cartilage. As a result, the limitation of this method is that it does not fully determine the random and systematic errors caused by patient positioning and soft tissue motion, which varies for each patient depending on the tumour position relative to the thyroid cartilage.

The radiographers performed the mask registration procedure after each treatment fraction and recorded inter-fraction errors on the data collection sheet. Due to the high

workload in the treatment unit, these procedures may have been rushed, which could result in errors. However, the researcher has no evidence that this happened.

The inter-fraction errors were considered as a product of set-up errors and organ motion errors and were not measured using the same methods as in other peer-reviewed studies. Instead, the methodology, with the aid of two external validators, was modified to apply it to the local department. Following a different procedure from other studies that assessed inter-fraction errors made the comparison of results more challenging.

Since the use of mask-registration was not standard procedure in the department, the results could have been influenced by the participating radiographers' lack of experience in performing the mask-registration technique. As a result, the researcher felt the need to test the participants' reliability in using this tool, which they proved to be reliable.

Limitations of the intra-fraction errors assessment

Even though an exhaustive sampling was used, one of the limitations of the study was the size of the available data. Since data obtained before patients' re-scans were not used, seven off-line XVIs were not included for the study. Furthermore, post-treatment CBCTs were not performed daily, but rather on the first fraction and then once weekly in order to limit the radiation dose to the patients. Although the sample size met The Royal College of Radiologists' (2008) recommendation and was large enough to obtain clinically relevant errors data, it would have been preferable if the sample size had been larger for assessing intra-fraction errors, since according to Button et al. (2013) this would ensure greater reliability and statistically significant results.

A limitation similar to that of total inter-fraction errors assessment was that radiographers on the local treatment unit did not typically perform mask registration as part of their normal routine, so they had no prior experience with this technique.

Since the CBCT used in this study did not provide real-time online tracking data, the intra-fraction motion of the target during treatment could not be evaluated. Instead, the post-treatment CBCT only showed motion obtained after treatment administration.

Limitations of observer variation in image matching errors assessment

Although all the radiographers who worked in the head and neck treatment unit participated in this study, there was a limitation with the number of cases that were considered for the evaluation of observer uncertainties. This had to be accepted as a limitation because the researcher needed to ensure radiographer participation by not burdening them with a massive workload. More cases could have produced more accurate population systematic and random errors results.

In the clinical setting, the radiographer who was responsible for performing online matching of the XVI was assisted by another radiographer prior to delivering treatment, as opposed the offline matching for the assessment of the inter-observer variation which was done by a single radiographer without any assistance. This could have influenced the outcomes of observer variation in image matching (Bell et al., 2019). In addition, radiographers had access to case details in clinical settings, but this information was not given during the assessment of inter-observer variation in image matching.

A limited number of studies such as, Deegan et al. (2015), Mohamoud, Ryan and Moseley (2015) and Hirose et al. (2020), examined observer variation in image matching, therefore it was difficult to compare the findings of this study to those of others that used a similar methodology.

Overall limitations of the study

Although rotational errors for each analysed error were measured, the population systematic and population random errors of rotations were not considered when calculating the PTV margin using van Herk's formula because this formula does not account for rotational errors. Adding the rotation errors should result in a larger margin.

Other errors, such as image deformation, real-time organ motion, phantom transfer errors, anatomical changes, dose calculation uncertainties and approximation of shifts, were not considered when calculating the PTV margin due to time restriction

and inaccessibility of devices to measure such errors. Therefore, further studies can complement the results achieved here.

As larynx cancer treatment was infrequent, data collection was limited to 20 patients; however, despite the small sample size, the study complied with The Royal College of Radiologists' (2021) recommendations.

4.10 Strengths of the study

Strengths of the target volume delineation errors assessment

The analysis of target volume delineation errors were performed using an exhaustive sample, which included all doctors who met the inclusion criteria. Having all the doctors trained in head and neck delineation region participating in this study ensured that the target delineation errors obtained were a true representation of the local department.

The accuracy of measuring the target volume delineation errors was evaluated. The high agreement in the measured distance (Pearson correlation coefficient of 0.88) indicated good reliability in the contouring range measurement. This procedure was required to ensure the accuracy of the measured distances.

Anonymisation of participants was used in order to obscure the results of participants from each other. This procedure ensured that doctors did not use the delineation of CTVs of other participants as a reference for their own.

The researcher was aware that the doctors had a heavy workload, and participation in this study could have been perceived as a burden by participants, affecting their contouring performance. As a result, participants were not rushed in delineating CTVs. Despite the researcher's two-month time frame, all the participating doctors had completed delineating the cases earlier. Therefore, the time frame did not influence the study's outcome since the doctors were not required to contour the target faster than a normal clinical situation.

Strengths of the total inter-fraction errors assessment

Validity and reliability testing before data collection ensured validity of the methodology to estimate inter-fraction errors and the results collected in the local department.

The population size used in this study was large enough to obtain clinically relevant error data, and systematic and random errors were calculated as per The Royal College of Radiologists (2008) recommendations.

Strengths of the intra-fraction errors assessment

This study follows methods similar to those used in larger, peer-reviewed studies. It also used The Royal College of Radiologists (2008) recommended methods for calculating random and systematic errors. Therefore, ensuring that the methodology is adequate for the objectives.

Strengths of observer variation in image matching errors assessment

One of the study's strengths was that all observers were radiographers with significant image guided experience and training in using the CBCT. The participating radiographers were also a complete representation of those responsible for treating the targeted patients.

Although rotation was not used in the final PTV margin calculation, it was still assessed for inter-observer variation in image matching, because even though the radiographers did not manually alter rotational directions, rotation was still variable depending on the clip-box position.

Overall strengths of the study

A systematic literature review was conducted prior to data collection to ensure that the most ideal formula for measuring the PTV margin was chosen for the scope of this study and to identify factors that must be considered when calculating the margin.

Even though target delineation errors and observer variation in image matching were frequently ignored during PTV margin assessment in previous studies, these errors were measured in this study because they also influence PTV margin size.

Validity and reliability testing of data collection tools were done to ensure correct measurements of all the analysed errors.

Being an insider type of research proved to be favourable toward the results of this study. In chapter 2, it was found that suppressing swallowing during the CT planning procedure reduces systematic errors, therefore the researcher made sure that this procedure was implemented in the department as an automatic voice command set by the radiographers working in CT planning to every patient who was having treatment to the head and neck region. It was also noted from chapter 2 that matching with the thyroid cartilage was a good surrogate to ensure that the target volume is inside the PTV, therefore a mask registration was implemented for the scope of this study.

4.11 Conclusion

This chapter presented the study's findings and results, which are required to answer the research question and achieve the research objectives. The following chapter will discuss the implications of the results on current practise and provide recommendations for further research.

Chapter 5

Conclusions and Recommendations

5.1 Introduction

In this chapter the conclusions drawn from the results and recommendations for clinical practise and further research based on the findings of this study are presented.

5.2 Conclusion

Using the van Herk formula and accounting for target volume delineation, inter-fraction, intra-fraction and observer variation in image matching errors in the equation, a PTV margin of 10.5 mm, 20.5 mm, and 12 mm was calculated, respectively in the X, Y and Z directions. This result varied significantly from the 5 mm PTV margin used for head and neck patients in the local oncology department. Applying the resulting PTV margin into practice would increase the radiotherapy-induced side effects due to the irradiation of normal tissue. However, neglecting the calculated margin size may result in the risk of missing the treatment target area. Therefore, the cause of the large margin size that was obtained in this study should be addressed. Another point to note was that the methodology of this study differed from other studies that assessed the PTV margin for head and neck patients treated with VMAT, which may have contributed to the significant variation in the PTV margin of this study compared to other published literature.

Target volume delineation errors were found to have the largest impact on the PTV margin size. Enabling ways to reduce the target volume delineation errors in the local oncology department would improve the conformity of the PTV margin size to the target area.

The second-largest errors were intra-fractional errors. When compared to the target volume delineation, the errors' impact on margin size were lower. The volume was found to be larger than previous studies, such as Kwa et al. (2015) and Perillo et al.'s (2021), that used a similar methodology.

Total inter-fraction errors were considered as a combination of set-up and organ motion errors. The translational and rotational values obtained for total inter-fraction

errors were low, with the population systematic errors calculated to be the largest, at 0.26 cm in the superior-inferior direction. This value compared well with previous studies that used a similar methodology to measure inter-fraction errors.

Observer variation in image matching were the least significant errors on the PTV margin size. However, they were found to have the largest rotational uncertainty which was predominantly systematic. Observer variation in image matching could be attributed to the variation of the clip-box size and position by the participating radiographers.

5.3 Recommendations

The following recommendations were derived from the data of this study. They are divided into two sections: recommendations for clinical practice and research recommendations.

5.3.1 Clinical recommendations

In this section, recommendations are being made on how to reduce each analysed error.

<u>Target volume delineation errors</u>

- It is recommended for clinical contouring protocols to be implemented as this will reduce the impact of target volume delineation errors. It is also advised that doctors receive contouring training and follow published contouring guidelines.

<u>Total inter-fraction errors and intra-fraction errors</u>

- During the data collection period, new immobilisation devices were procured to treat head and neck patients with VMAT. It is being recommended to measure inter-fraction and intra-fraction errors for these new devices. This procedure would be required to assess whether the calculated PTV margin from this study would still be applicable when using the new immobilisation devices.

- Suppressing swallowing during CT planning helps to reduce the population's systematic errors (Bahig et al., 2021), therefore, it is being recommended that the local department continues to maintain this procedure which was implemented prior to the data collection period.

- The researcher suggests using real-time imaging for VMAT larynx radiotherapy. A business plan would need to be performed to assess the benefits, limitations, resources required, costs, and alternatives that would aid in determining the need for this technique and consider the implementation of this process into clinical practice.

Inter-observer variation in image matching

- Adherence to protocols and guidelines on CBCT image matching by radiographers working on the treatment unit should be reinforced with a structured learning and training programme. A regular audit of concordance is also being recommended, as suggested by the Royal College of Radiologists (2021). These procedures would assist in reducing inter-observer variation in image matching.

Other recommendations

- To strengthen and validate the results of this study, it is recommended that the measured errors are calculated for a larger group of patients.

- It is recommended that the local oncology hospital assess the PTV margin being applied for other treatment areas. Data collection for all pathologies and anatomical areas would allow to perform PTV margin calculation for all sites.

5.3.2 Recommendations for further studies

- Further research is required to collect data using the same methodology for measuring set-up errors and organ motion for other parts of the body. This would enable comparison of the data obtained in this study with that of real-time tumour motion and set-up errors measurements. This will further evaluate the validity of the methodological design used for this study.

- Future studies are required to compare PTV margins utilising the various formulas available that could be implemented. These studies will assess the variation obtained on the margin size from the different utilised formula.

- Recommendations for future work to evaluate inter-observer variation in image matching with an increased sample size, to ensure more statistical reliable results.

- Further research investigating swallowing suppression during treatment using gating techniques and its impact on intra-fraction errors is recommended.

- It is recommended that a local study is undertaken to compare the inter-observer variation of image registration performed by radiographers using clip-box with that of mask registration. This study will enable the comparison of the competency levels with the reliability of image matching using both techniques.

- Further studies to measure target delineation errors for assessing PTV margin are recommended as most studies that assessed target volume delineation errors were not done with the scope of measuring the delineation errors to calculate the PTV margin size.

- Recommendations for further research to be conducted in the local department with the scope of evaluating competency levels of doctors when delineating the target volume and how training levels influence margin size. The purpose of this study would be to see if training and adherence to local standard procedures may reduce target volume delineation errors.

- Although this study took great precautions in measuring errors that influence the margin size, other errors were not considered. Therefore, it is suggested for research to assess the impact of errors, such as dose calculation uncertainties, phantom transfer errors and anatomical changes on the PTV margin size.

5.4 Overall conclusion

Locally, this study was the first to measure the PTV margin for patients receiving treatment to the larynx with VMAT. On an international level, this study was the first to consider target volume delineation errors, as well as inter-observer variation errors, in the PTV margin calculation for patients receiving VMAT treatment to the larynx. It is also the first study to measure inter-fraction errors as a combination of organ motion and set-up errors for patients treated with VMAT to the larynx. The study is also one of the few that considered a vast number of errors in margin calculation and this highlight the importance of having more research done on this subject area.

The objectives outlined in chapter 1 were achieved. Based on the results obtained, the local clinical department may consider ways of how to reduce the PTV margin size for patients receiving treatment to the larynx with VMAT. It was also suggested by various authors that PTV margin size calculation should not be based on one error but should consider measuring a different array of errors for a more accurate assessment of the PTV margin size. The implementation of some or all of the recommendations could result in a better and safer practise for patients receiving similar treatments.

.

References

Almeida, Carlos Podalirio Borges de, & Goulart, Bárbara Niegia Garcia de. (2017). How to avoid bias in systematic reviews of observational studies. *Revista CEFAC, 19*(4), 551-555. 10.1590/1982-021620171941117

Althubaiti, A. (2016). Information bias in health research: definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare, 9*(1), 211-217. 10.2147/JMDH.S104807

Andrews, S., & Hamarneh, G. (2015). Multi-Region Probabilistic Dice Similarity Coefficient using the Aitchison Distance and Bipartite Graph Matching. https://arxiv.org/abs/1509.07244

Anjanappa, M., Rafi, M., Bhasi, S., Kumar, R., Thommachan, K. C., Bhattacharya, T., & Ramadas, K. (2017). Setup uncertainties and PTV margins at different anatomical levels in intensity modulated radiotherapy for nasopharyngeal cancer. *Reports of Practical Oncology & Radiotherapy, 22*(5), 396-401. 10.1016/j.rpor.2017.07.005

Antolak, J. A., & Rosen, I. I. (1999). Planning target volumes for radiotherapy: how much margin is needed? *International Journal of Radiation Oncology, Biology, Physics, 44*(5), 1165-1170. 10.1016/S0360-3016(99)00117-0

Apolle, R., Appold, S., Bijl, H. P., Blanchard, P., Bussink, J., Faivre-Finn, C., Khalifa, J., Laprie, A., Lievens, Y., Madani, I., Ruffier, A., de Ruysscher, D., van Elmpt, W., & Troost, E. G. C. (2019). Inter-observer variability in target delineation increases during adaptive treatment of head-and-neck and lung cancer. *Acta Oncologica, 58*(10), 1378-1385. 10.1080/0284186x.2019.1629017

Arumugam, S., Xing, A., Vial, P., Scotti, A., Stirton, R., Goozee, G., & Holloway, L. (2013). Impact of patient rotational errors on target and critical structure dose in IMRT: A 3D simulation study. *Journal of Physics. Conference Series, 444*, 12045. 10.1088/1742-6596/444/1/012045

Ashour, M. G., Shouman, T. H., Hassouna, A. H., El Din, Reem M. Emad, Mokhtar, M. H., & Khalil, M. A. (2017b). Measuring Radiotherapy Setup Errors in IMRT Treated Head and Neck Cancer Patients Requiring Bilateral Neck Irradiation, NCI-Egypt Experience. *Journal of Cancer Therapy, 8*(13), 1160-1168. 10.4236/jct.2017.813099

Bahig, H., Rosenthal, D. I., Nguyen-Tan, F., Fuller, D. C., Yuan, Y., Hutcheson, K. A., Christopoulos, A., Nichols, A. C., Fung, K., Ballivy, O., Filion, E., Ng, S. P., Lambert, L., Dorth, J., Hu, K. S., & Palma, D. (2021). Vocal-cord Only vs. Complete Laryngeal radiation (VOCAL): a randomized multicentric Bayesian phase II trial. *BMC Cancer, 21*(1), 446. 10.1186/s12885-021-08195-8

Banerjee, A., & Chaudhury, S. (2010). Statistics without tears: populations and samples. *Industrial Psychiatry Journal, 19*(1), 60-65. 10.4103/0972-6748.77642

Baron, C. A., Awan, M. J., Mohamed, A. S. R., Akel, I., Rosenthal, D. I., Gunn, G. B., Garden, A. S., Dyer, B. A., Court, L., Sevak, P. R., Kocak-Uzel, E., & Fuller, C. D. (2015). Estimation of daily interfractional larynx residual setup error after isocentric alignment for head and neck radiotherapy: quality assurance implications for target volume and organs-at-risk margination using daily CT on-rails imaging. *Journal of Applied Clinical Medical Physics, 16*(1), 159-169. 10.1120/jacmp.v16i1.5108

Barry, M., & Kell, M. R. (2011). Radiotherapy and breast reconstruction: a meta-analysis. *Breast Cancer Research and Treatment, 127*(1), 15-22. 10.1007/s10549-011-1401-x

Bartolo, C., Muscat. K. (2019). XVI for Ca Head & Neck VMAT Radiotherapy. Sir Anthony Mamo Oncology Centre.

Baumann, N. (2016). How to use the medical subject headings (Me). *International Journal of Clinical Practice (Esher), 70*(2), 171-174. 10.1111/ijcp.12767

Bel, A., van Herk, M., & Lebesque, J. V. (1996). Target margins for random geometrical treatment uncertainties in conformal radiotherapy. *Medical Physics (Lancaster), 23*(9), 1537-1545. 10.1118/1.597745

Bell, L. J., Eade, T., Hruby, G., Bromley, R., & Kneebone, A. (2019). Implementing daily soft tissue image guidance with reduced margins for post-prostatectomy radiotherapy: research-based changes to clinical practice. *Journal of Medical Radiation Sciences, 66*(4), 259-268. 10.1002/jmrs.362

Benoot, C., Hannes, K., & Bilsen, J. (2016). The use of purposeful sampling in a qualitative evidence synthesis: A worked example on sexual adjustment to a cancer trajectory. *BMC Medical Research Methodology, 16*(1), 21. 10.1186/s12874-016-0114-6

Bernstein, D., Taylor, A., Nill, S., Imseeh, G., Kothari, G., Llewelyn, M., De Paepe, K. N., Rockall, A., Shiarli, A. -., & Oelfke, U. (2021). An Inter-observer Study to Determine Radiotherapy Planning Target Volumes for Recurrent Gynaecological Cancer Comparing Magnetic Resonance Imaging Only With Computed Tomography-Magnetic Resonance Imaging. *Clinical Oncology (Royal College of Radiologists (Great Britain)), 33*(5), 307-313. 10.1016/j.clon.2021.02.003

Bettany-Saltikov, J. (2012). *How to Do a Systematic Literature Review in Nursing: a Step-By-step Guide*. McGraw-Hill Education.

Bigby, M. (2014). Understanding and Evaluating Systematic Reviews and Meta-analyses. *Indian Journal of Dermatology, 59*(2), 134-139. 10.4103/0019-5154.127671

Bölling, T., Weege, J., Eich, H. T., Timmermann, B., Meyer, F., Rübe, C., Kortmann, R., Fischedick, K., Rödel, C., Koch, R., & Willich, N. (2015). Acute and late side effects to salivary glands and oral mucosa after head and neck radiotherapy in children and adolescents. Results of the "Registry for the evaluation of side effects after radiotherapy in childhood and adolescence". *Head & Neck, 37*(8), 1137-1141. 10.1002/hed.23715

Borg Xuereb, H. K., Dimech, A., & Muscat, K. (2015). *Local attitudes in the treatment of low prognosis head and neck squamous cell carcinoma*. Malta Medical Journal.

Boutron, I., Ravaud, P., & Moher, D. (Eds.). (2012). Randomized Clinical Trials of Nonpharmacological Treatments (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b11349

Bradley, J. A., M.D, Paulson, E. S., Ph.D, Ahunbay, E., Ph.D, Schultz, C., M.D, Li, X. A., Ph.D, & Wang, Dian, M.D., Ph.D. (2011). Dynamic MRI Analysis of Tumor and Organ Motion During Rest and Deglutition and Margin Assessment for Radiotherapy of Head-and-Neck Cancer. *International Journal of Radiation Oncology, Biology, Physics, 81*(5), e803-e812. 10.1016/j.ijrobp.2010.12.015

Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L., & Bekhuis, T. (2016). De-duplication of database search results for systematic reviews in EndNote. *Journal of the Medical Library Association, 104*(3), 240-243. 10.3163/1536-5050.104.3.014

Bruijnen, T., Stemkens, B., Terhaard, C. H. J., Lagendijk, J. J. W., Raaijmakers, C. P. J., & Tijssen, R. H. N. (2019). Intrafraction motion quantification and planning target volume margin determination of head-and-neck tumors using cine magnetic resonance imaging. *Radiotherapy and Oncology, 130*, 82-88. 10.1016/j.radonc.2018.09.015

Brydon-Miller, M., & Coghlan, D. (2014). *The Sage encyclopedia of action research*. Sage reference.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafo, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience, 14*(6), 442. 10.1038/nrn3502

Cacicedo, J., Perez, J., Perez, J., Ortiz de Zarate, R., Ortiz de Zarate, R., del Hoyo, O., del Hoyo, O., Casquero, F., Casquero, F., Gómez-Iturriaga, A., Gómez-Iturriaga, A., Lasso, A., Lasso, A., Boveda, E., Boveda, E., Bilbao, P., & Bilbao, P. (2015). A prospective analysis of inter- and intrafractional errors to calculate CTV to PTV margins in head and neck patients. *Clinical & Translational Oncology, 17*(2), 113-120. 10.1007/s12094-014-1200-z

Caravatta, L., Macchia, G., Mattiucci, G. C., Sainato, A., Cernusco, N. L. V., Mantello, G., Di Tommaso, M., Trignani, M., De Paoli, A., Boz, G., Friso, M. L., Fusco, V., Di Nicola, M., Morganti, A. G., & Genovesi, D. (2014). Inter-observer variability of clinical target volume delineation in radiotherapy treatment of pancreatic cancer: a multi-institutional contouring experience. *Radiation Oncology (London, England), 9*(1), 198. 10.1186/1748-717X-9-198

Caruana, K., Refalo, N., Spiteri, D., Couto, J. G., Zarb, F., & Bezzina, P. (2021). PTV margin calculation for head and neck patients treated with VMAT: a systematic literature review. *Journal of Radiotherapy in Practice,* , 1-8. 10.1017/S1460396921000546

Chang, J. (2017). A statistical model for analyzing the rotational error of single isocenter for multiple targets technique. *Medical Physics (Lancaster), 44*(6), 2115-2123. 10.1002/mp.12262

Chang, X., Deng, W., Wang, X., Zhou, Z., Yang, J., Guo, W., Liu, M., Qi, X., Li, L., Zhang, K., Zhang, M., Shi, Y., Liu, K., Zhao, Y., Wang, H., Yu, Z., Zhang, J., Wang, L., Qiao, X., . . . Xiao, Z. (2021). Interobserver variability in target volume delineation in definitive radiotherapy for thoracic esophageal cancer: a multi-center study from China. *Radiation Oncology (London, England), 16*(1), 1-102. 10.1186/s13014-020-01691-4

Charrois, T. L. (2015). Systematic Reviews: What Do You Need to Know to Get Started? *Canadian Journal of Hospital Pharmacy, 68*(2)10.4212/cjhp.v68i2.1440

Cheo, T., Loh, Y., Chen, D., Lee, K. M., & Tham, I. (2015). Measuring radiotherapy setup errors at multiple neck levels in nasopharyngeal cancer (NPC): A case for differential PTV expansion. *Radiotherapy and Oncology, 117*(3), 419-424. 10.1016/j.radonc.2015.09.032

Das, I. J., Compton, J. J., Bajaj, A., & Johnstone, P. A. (2021). Intra- and inter-physician variability in target volume delineation in radiation therapy. *Journal of Radiation Research,* 10.1093/jrr/rrab080

Deb, J., Chaudhuri, S., Panda, D., Sinha, S. N., Kumar, S., Kumar, N., Basu, S., Kumar, N., M, A. P., & Gupta, S. K. (2019). Retrospective analysis of random and systematic errors in radiation

therapy of head and neck cancer patients and its clinical predictive implications with VMAT treatment. *International Journal of Research in Medical Sciences, 7*(7), 2758. 10.18203/2320-6012.ijrms20192914

Deegan, T., Owen, R., Holt, T., Fielding, A., Biggs, J., Parfitt, M., Coates, A., & Roberts, L. (2015). Assessment of cone beam CT registration for prostate radiation therapy: Fiducial marker and soft tissue methods. *Journal of Medical Imaging and Radiation Oncology, 59*(1), 91-98. 10.1111/1754-9485.12197

Deurloo K. E. I., Steenbakkers R. J. H. M., Zijp L. J., de Bois J. A., Nowak P. J. C. M., Rasch C. R. N., Van Herk M. Quantification of shape variation of prostate and seminal vesicles during external beam radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys*. 2005;61:228–38. doi: 10.1016/j.ijrobp.2004.09.023.

Dekker, S. (2014). *The Field Guide to Understanding 'Human Error'* (3rd ed.). CRC Press. 10.1201/9781317031833

Denaro, N., Russi, E. G., Adamo, V., & Merlano, M. C. (2014). State-of-the-Art and Emerging Treatment Options in the Management of Head and Neck Cancer: News from 2013. *Oncology, 86*(4), 212-229. 10.1159/000357712

Dewas, S., Bibault, J., Blanchard, P., Vautravers-Dewas, C., Pointreau, Y., Denis, F., Brauner, M., & Giraud, P. (2011). Delineation in thoracic oncology: a prospective study of the effect of training on contour variability and dosimetric consequences. *Radiation Oncology, 6*(1), 118. 10.1186/1748-717X-6-118

Duan, X., Zhou, Y., Dai, H., Zhao, L., Qian, J., Yang, D., Zhang, L., Luo, C. and Li, G. (2020), Evaluating the positional uncertainty of intrafraction, adjacent fields, and daily setup with the BrainLAB ExacTrac system in patients who are receiving craniospinal irradiation. *J Appl Clin Med Phys*, 21: 35-46. https://doi.org/10.1002/acm2.12909

Djordjevic, M., Sjöholm, E., Tullgren, O., & Sorcini, B. (2014). Assessment of residual setup errors for anatomical sub-structures in image-guided head-and-neck cancer radiotherapy. *Acta Oncologica, 53*(5), 646-653. 10.3109/0284186X.2013.862593

Ecclestone, G (2012). *Assessing The Clinical Application Of The Van Herk Margin Formula For Lung Radiotherapy* https://digital.library.ryerson.ca/islandora/object/RULA%3A1894

Eden, J., Levit, L. A., Berg, A. O. & Morton, S. C. (2011). *Finding what works in health care.* http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=375277

Eldawlatly, A., Alshehri, H., Alqahtani, A., Ahmad, A., Al-Dammas, F., & Marzouk, A. (2018). Appearance of Population, Intervention, Comparison, and Outcome as research question in the title of articles of three different anesthesia journals: A pilot study. *Saudi Journal of Anaesthesia, 12*(2), 283-286. 10.4103/sja.SJA_767_17

Fotina, I., Lütgendorf-Caucig, C., Stock, M., Pötter, R., & Georg, D. (2012). Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. *Strahlentherapie Und Onkologie, 188*(2), 160-167. 10.1007/s00066-011-0027-6

Franco, P., Arcadipane, F., Trino, E., Gallio, E., Martini, S., Iorio, G. C., Piva, C., Moretto, F., Ruo Redda, M. G., Verna, R., Tseroni, V., Bona, C., Pozzi, G., Fiandra, C., Ragona, R., Bertetto, O., & Ricardi, U. (2018). Variability of clinical target volume delineation for rectal cancer patients planned for neoadjuvant radiotherapy with the aid of the platform Anatom-e. *Clinical and Translational Radiation Oncology, 11*, 33-39. 10.1016/j.ctro.2018.06.002

Frey, B. (2018). The Sage encyclopedia of educational research, measurement, and evaluation. *SAGE Publications*, DOI:10.4135/9781506326139

Gangsaas, A., Astreinidou, E., PhD, Quint, S., MSc, Levendag, Peter C., MD, PhD, & Heijmen, B., PhD. (2013). Cone-Beam Computed Tomography–Guided Positioning of Laryngeal Cancer Patients with Large Interfraction Time Trends in Setup and Nonrigid Anatomy Variations. *International Journal of Radiation Oncology, Biology, Physics, 87*(2), 401-406. 10.1016/j.ijrobp.2013.06.2032

Gill, S. K., Reddy, K., Campbell, N., Chen, C., & Pearson, D. (2015). Determination of optimal PTV margin for patients receiving CBCT-guided prostate IMRT: comparative analysis based on CBCT dose calculation with four different margins. *Journal of Applied Clinical Medical Physics, 16*(6), 252-262. 10.1120/jacmp.v16i6.5691

Glasser, S. P. (2014). *Essentials of Clinical Research* (2nd ed. 2014. ed.). Springer. 10.1007/978-3-319-05470-4

Gordon, J. J., & Siebers, J. V. (2008). Evaluation of dosimetric margins in prostate IMRT treatment plans. *Medical Physics (Lancaster), 35*(2), 569-575. 10.1118/1.2826558

Guerreiro, F., Seravalli, E., Janssens, G. O., van de Ven, Cees P, van den Heuvel-Eibrink, Marry M, & Raaymakers, B. W. (2018). Intra- and inter-fraction uncertainties during IGRT for Wilms' tumor. *Acta Oncologica, 57*(7), 941-949. 10.1080/0284186X.2018.1438655

Gupta, M., Gamre, P., Kannan, S., Rokde, G., Krishnatry, R., & Murthy, V. (2018). Effect of imaging frequency on PTV margins and geographical miss during image guided radiation therapy for prostate cancer. *Practical Radiation Oncology, 8*(2), e41-e47. 10.1016/j.prro.2017.09.010

Gurney-Champion, O. J., McQuaid, D., Dunlop, A., Wong, K. H., Welsh, L. C., Riddell, A. M., Koh, D., Oelfke, U., Leach, M. O., Nutting, C. M., Bhide, S. A., Harrington, K. J., Panek, R., & Newbold, K. L. (2018). MRI-based Assessment of 3D Intrafractional Motion of Head and Neck Cancer for Radiation Therapy. *International Journal of Radiation Oncology, Biology, Physics, 100*(2), 306-316. 10.1016/j.ijrobp.2017.10.016

Hammers, J. E., Pirozzi, S., Lindsay, D., Kaidar-Person, O., Tan, X., Chen, R. C., Das, S. K., & Mavroidis, P. (2020). Evaluation of a commercial DIR platform for contour propagation in prostate cancer patients treated with IMRT/VMAT. *Journal of Applied Clinical Medical Physics, 21*(2), 14-25. 10.1002/acm2.12787

Haneuse, S. (2016). Distinguishing Selection Bias and Confounding Bias in Comparative Effectiveness Research. *Medical Care, 54*(4), e23-e29. 10.1097/mlr.0000000000000011

Hannes, K., Lockwood, C., & Pearson, A. (2010). A Comparative Analysis of Three Online Appraisal Instruments' Ability to Assess Validity in Qualitative Research. *Qualitative Health Research, 20*(12), 1736-1743. 10.1177/1049732310378656

Hargrave, C., & Holt, T. (2017). Margins Matter: current radiation therapy practice and future directions. *Journal of Medical Radiation Sciences, 64*(2), 79-81. 10.1002/jmrs.233

Harvey, G., & Kitson, A. (2016). PARIHS revisited: from heuristic to integrated framework for the successful implementation of knowledge into practice. *Implementation Science : IS, 11*(1), 33. 10.1186/s13012-016-0398-2

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing, 18*(3), 66-67. 10.1136/eb-2015-102129

Hirose, T., Arimura, H., Fukunaga, J., Ohga, S., Yoshitake, T., & Shioyama, Y. (2020). Observer uncertainties of soft tissue-based patient positioning in IGRT. *Journal of Applied Clinical Medical Physics, 21*(2), 73-81. 10.1002/acm2.12817

Holian, R., & Coghlan, D. (2012). Ethical Issues and Role Duality in Insider Action Research: Challenges for Action Research Degree Programmes. *Systemic Practice and Action Research, 26*(5), 399-415. 10.1007/s11213-012-9256-6

Høyer, M., Thor, M., Thörnqvist, S., Søndergaard, J., Lassen-Ramshad, Y., & Paul Muren, L. (2011). Advances in radiotherapy: from 2D to 4D. *Cancer Imaging, 11 Spec No A*(1A), S147-S152. 10.1102/1470-7330.2011.9036

Hummel, S., Simpson, E. L., Hemingway, P., Stevenson, M. D., & Rees, A. (2010). Intensity-modulated radiotherapy for the treatment of prostate cancer: a systematic review and economic evaluation. *Health Technology Assessment (Winchester, England), 14*(47), 1-108. 10.3310/hta14470

ICRU. (1994). *ICRU Report 50—Prescribing, Recording and Reporting Photon Beam Therapy*. American Association of Physicists in Medicine. 10.1118/1.597396

ICRU. (1999) *ICRU Report 62: Prescribing, Recording and Reporting Photon Beam Therapy (Supplement to ICRU Report 50)*. ICRU, Bethesda.

ICRU. (2010). *ICRU Report 83 Prescribing, Recording, and Reporting Photon-beam Intensity-modulated Radiation Therapy (IMRT) -Journal of the ICRU* - Vol 10 No 1 2010. Oxford University Press. https://books.google.com.mt/books?id=KZhstgEACAAJ

Ignacio, J.J., Taylor, B.J. (2013). Ethical issues in health-care inquiry: A discussion paper. *International Journal of Nursing Practice.* 10.1111/ijn.12017

Ingrosso, G., Miceli, R., Fedele, D., Ponti, E., Benassi, M., Barbarino, R., Di Murro, L., Giudice, E., Santarelli, F., & Santoni, R. (2012). Cone-beam computed tomography in hypofractionated stereotactic radiotherapy for brain metastases. *Radiation Oncology (London, England), 7*(1), 54. 10.1186/1748-717X-7-54

İsmail Faruk Durmuş, Bora Taş, & Ömer Erol Uzel. (2020). Determination of Intrafraction Laryngeal Movement in Radiotherapy. *Türk Onkoloji Dergisi, 35*(2)10.5505/tjo.2020.2185

Jager, E. A. (2017). *Towards accurate target delineation for head and neck cancer* https://www.narcis.nl/publication/RecordID/oai:dspace.library.uu.nl:1874%2F350500

Jager, E. A., Kasperts, N., Caldas Magalhaes, J., Philippens, M. E. P., Pameijer, F. A., Terhaard, C. H. J., & Raaijmakers, C. P. J. (2015). GTV delineation in supraglottic laryngeal carcinoma :

interobserver agreement of CT versus CT-MR delineation. *Radiation Oncology (London, England), 10*(1), 26. 10.1186/s13014-014-0321-4

Jain, & Sandhya. (2016). Guidelines for Systematic Reviews. *International Dental & Medical Journal of Advanced Research, 2*10.15713/ins.idmjar.48

Jameson, M. G., Holloway, L. C., Vial, P. J., Vinod, S. K., & Metcalfe, P. E. (2010). A review of methods of analysis in contouring studies for radiation oncology. *Journal of Medical Imaging and Radiation Oncology, 54*(5), 401-410. 10.1111/j.1754-9485.2010.02192.x

Jan-Jakob Sonke, & Marcel van Herk. (2016). Errors, Margins, Localization, and Correction. https://radiologykey.com/errors-margins-localization-and-correction/

JENA, R., KIRKBY, N. F., BURTON, K. E., HOOLE, A. C. F., TAN, L. T., & BURNET, N. G. (2010). A novel algorithm for the morphometric assessment of radiotherapy treatment planning volumes. *British Journal of Radiology, 83*(985), 44-51. 10.1259/bjr/27674581

Jenny Fleming. (2018). Recognizing and resolving the challenges of being an insider researcher in work-integrated learning. *International Journal of Work-Integrated Learning, 19*(3), 311. https://search.proquest.com/docview/2227915506

Jensen, K., Friborg, J., Hansen, C. R., Samsøe, E., Johansen, J., Andersen, M., Smulders, B., Andersen, E., Nielsen, M. S., Eriksen, J. G., Petersen, J. B. B., Elstrøm, U. V., Holm, A. I., Farhadi, M., Morthorst, M. H., Skyt, P. S., Overgaard, J., & Grau, C. (2020). The Danish Head and Neck Cancer Group (DAHANCA) 2020 radiotherapy guidelines. *Radiotherapy and Oncology, 151*, 149-151. 10.1016/j.radonc.2020.07.037

Joanne E McKenzie, Sue E Brennan, Rebecca E Ryan, Hilary J Thomson & Renea V Johnston. (2019). *Chapter 9: Summarizing study characteristics and preparing for synthesis.* Cochrane training. https://training.cochrane.org/handbook/current/chapter-09#_Ref531070527

Juan Suárez-Quintanilla, Alejandro Fernández Cabrera, & Sandeep Sharma. (2020). Anatomy, Head and Neck, Larynx. *StatPearls Publishing,* https://www.ncbi.nlm.nih.gov/books/NBK538202/

Kalyankuppam Selvaraj, J. (2013). *Modelling the effect of geometric uncertainties, clonogen distribution and IMRT interplay effect on tumour control probability* http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.617475

Kapanen, M., Laaksomaa, M., Tulijoki, T., Peltola, S., Wigren, T., Hyödynmaa, S., & Kellokumpu-Lehtinen, P. (2013). Estimation of adequate setup margins and threshold for position errors

requiring immediate attention in head and neck cancer radiotherapy based on 2D image guidance. *Radiation Oncology (London, England), 8*(1), 212. 10.1186/1748-717X-8-212

Khoo, E. L. H., F.R.A.N.Z.C.R, Schick, K., B.App.Sc, Plank, A. W., Ph.D, Poulsen, M., M.D, Wong, W. W. G., F.R.A.N.Z.C.R, Middleton, M., M.B.A, & Martin, J. M., F.R.A.N.Z.C.R. (2012). Prostate Contouring Variation: Can It Be Fixed? *International Journal of Radiation Oncology, Biology, Physics, 82*(5), 1923-1929. 10.1016/j.ijrobp.2011.02.050

Kim, K. S., Cheong, K., Kim, K., Koo, T., Koh, H. K., Chang, J. H., Chang, A. R., & Park, H. J. (2021). Interobserver variability in clinical target volume delineation in anal squamous cell carcinoma. *Scientific Reports, 11*(1), 2785. 10.1038/s41598-021-82541-5

Kim, S., Oh, S. A., Yea, J., & Park, J. W. (2019). Prospective assessment of inter- or intra-fractional variation according to body weight or volume change in patients with head and neck cancer undergoing radiotherapy. *PLOS ONE*, 14, e0216655. https://doi.org/10.1371/journal.pone.0216655

Kirkham, J. J., Dwan, K. M., Altman, D. G., Gamble, C., Dodd, S., Smyth, R., & Williamson, P. R. (2010). The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *Bmj, 340*(7747), 637-640. 10.1136/bmj.c365

Konno, K., Akasaka, M., Koshida, C., Katayama, N., Osada, N., Spake, R., Amano, T. (2020).Ignoring non-English-language studies may bias ecological meta-analyses https://doi.org/10.1002/ece3.6368

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155-163. 10.1016/j.jcm.2016.02.012

Koroulakis, A., & Agarwal, M. (2022). Laryngeal Cancer. *StatPearls Publishing,* https://www.ncbi.nlm.nih.gov/books/NBK526076/

Kristensen, I., Nilsson, K., Agrup, M., Belfrage, K., Embring, A., Haugen, H., Svärd, A., Knöös, T., & Nilsson, P. (2017). A dose based approach for evaluation of inter-observer variations in target delineation. *Technical Innovations & Patient Support in Radiation Oncology, 3-4*(C), 41-47. 10.1016/j.tipsro.2017.10.002

Kukolowicz, P., Mietelska, M., & Kiprian, D. (2020). Effectiveness of the No action level protocol for head & neck patients – Time considerations. *Reports of Practical Oncology and Radiotherapy,* 10.1016/j.rpor.2020.04.005

Kwa, S. L. S., PhD, Al-Mamgani, Abrahim, MD, PhD, Osman, S. O. S., PhD, Gangsaas, A., BSc, Levendag, Peter C., MD, PhD, & Heijmen, B. J. M., PhD. (2015). Inter- and Intrafraction Target Motion in Highly Focused Single Vocal Cord Irradiation of T1a Larynx Cancer Patients. *International Journal of Radiation Oncology, Biology, Physics, 93*(1), 190-195. 10.1016/j.ijrobp.2015.04.049

Kwon, Y., Lemieux, M., McTavish, J., & Wathen, N. (2015). Identifying and removing duplicate records from systematic review searches. *Journal of the Medical Library Association, 103*(4), 184-188. 10.3163/1536-5050.103.4.004

Langmack, K. A., Perry, C., Sinstead, C., Mills, J., & Saunders, D. (2014). The utility of atlas-assisted segmentation in the male pelvis is dependent on the interobserver agreement of the structures segmented. *British Journal of Radiology, 87*(1043), 20140299. 10.1259/bjr.20140299

Laursen, L. V., Elstrøm, U. V., Vestergaard, A., Muren, L. P., Petersen, J. B., Lindegaard, J. C., Grau, C., & Tanderup, K. (2012). Residual rotational set-up errors after daily cone-beam CT image guided radiotherapy of locally advanced cervical cancer. *Radiotherapy and Oncology, 105*(2), 220-225. 10.1016/j.radonc.2012.08.012

Le, Q. T., Wuthrick, E., Xiao, Y., Machtay, M., & Powell, T. (2022). RTOG. RTOG. https://www.rtog.org/

Lee, J. (2020). Statistics, Descriptive. *International Encyclopedia of Human Geography (Second Edition)*, 13-20. https://doi.org/10.1016/B978-0-08-102295-5.10428-7.

Leech, M., Coffey, M., Mast, M., Moura, F., Osztavics, A., Pasini, D., & Vaandering, A. (2017). ESTRO ACROP guidelines for positioning, immobilisation and position verification of head and neck patients for radiation therapists. *Technical Innovations & Patient Support in Radiation Oncology, 1*, 1-7. 10.1016/j.tipsro.2016.12.001

Li, W. T. Y. (2014). *Determining Rational Planning Target Volume Margins for Intracranial Stereotactic Radiotherapy* Available from ProQuest One Academic Eastern Edition https://search.proquest.com/docview/1648985804

Liang, J., Li, M., Zhang, T., Han, W., Chen, D., Hui, Z., Lv, J., Zhang, Z., Zhang, Y., Zhang, L., Zheng, R., Dai, J., & Wang, L. (2014). The effect of image-guided radiation therapy on the margin between the clinical target volume and planning target volume in lung cancer. *Journal of Medical Radiation Sciences, 61*(1), 30-37. 10.1002/jmrs.42

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS Medicine, 6*(7), e1000100. 10.1371/journal.pmed.1000100

Liu, Z., Liu, F., Chen, W., Tao, Y., Liu, X., Zhang, F., Shen, J., Guan, H., Zhen, H., Wang, S., Chen, Q., Chen, Y., & Hou, X. (2021). Automatic Segmentation of Clinical Target Volume and Organs-at-Risk for Breast Conservative Radiotherapy Using a Convolutional Neural Network. *Cancer Management and Research, 13*, 8209-8217. 10.2147/CMAR.S330249

Lo Nigro, C., Denaro, N., Merlotti, A., & Merlano, M. (2017). Head and neck cancer: improving outcomes with a multidisciplinary approach. *Cancer Management and Research, 9*, 363-371. 10.2147/CMAR.S115761

Long, L. (2014). Routine piloting in systematic reviews—a modified approach? *Systematic Reviews, 3*(1), 77. 10.1186/2046-4053-3-77

Lu, H., Lin, H., Feng, G., Chen, J., Shu, L., Pang, Q., Cheng, J., Peng, L., Wu, D., Liao, C., & Mo, Y. (2012). Interfractional and intrafractional errors assessed by daily cone-beam computed tomography in nasopharyngeal carcinoma treated with intensity-modulated radiation therapy: a prospective study. *Journal of Radiation Research, 53*(6), 954-960. 10.1093/jrr/rrs041

Ma, L., Wang, Y., Yang, Z., Huang, D., Weng, H., & Zeng, X. (2020). Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? *Military Medical Research, 7*(1), 7. 10.1186/s40779-020-00238-8

MacLure, K., Paudyal, V., & Stewart, D. (2016). Reviewing the literature, how systematic is systematic? *International Journal of Clinical Pharmacy, 38*(3), 685-694. 10.1007/s11096-016-0288-3

Malicki, J. (2012). The importance of accurate treatment planning, delivery, and dose verification. *Reports of Practical Oncology and Radiotherapy, 17*(2), 63-65. 10.1016/j.rpor.2012.02.001

Malmqvist, J., Hellberg, K., Möllås, G., Rose, R., & Shevlin, M. (2019). Conducting the Pilot Study: A Neglected Part of the Research Process? Methodological Findings Supporting the Importance of Piloting in Qualitative Research Studies. *International Journal of Qualitative Methods, 18*, 160940691987834. 10.1177/1609406919878341

Mandair, D., Ho, Q.-A., Wertheim, B., Johnson, G., Roe, D., Watchman, C., & Yi, S. (2018). Comparison of two thermoplastic immobilization mask systems in daily volumetric image guided radiation therapy for head and neck cancers. *Biomedical Physics & Engineering Express*, 4. https://doi.org/10.1088/2057-1976/aad574

Marnouche, E. A., Abdhelak, M., Lalya, I., Zaghba, N., Bazine, A., & et al. (2019). Set-up Errors and Imaging Verification Protocols for Head and Neck Cancers Radiotherapy in Morocco. *Journal of Oncology Research and Treatment, 4*

Martins, L, Couto, J G, Barbosa, B. (2016). Use of planar kV vs. CBCT in evaluation of setup errors in oesophagus carcinoma radiotherapy. *Rep Pract Oncol Radiother*; 21 (1), 57–62. doi: 10.1016/j.rpor.2015.10.005

McEwan, B. (2020). Sampling and validity. *Annals of the International Communication Association, 44*(3), 235-247. 10.1080/23808985.2020.1792793

McKenzie, A. L., Herk, M. v., & Mijnheer, B. (2000). The width of margins in radiotherapy treatment plans. *Physics in Medicine & Biology, 45*(11), 3331-3342. 10.1088/0031-9155/45/11/315

Mercieca, S., Belderbos, J. S. A., & van Herk, M. (2021). Challenges in the target volume definition of lung cancer radiotherapy. *Translational Lung Cancer Research, 10*(4), 1983-1998. 10.21037/tlcr-20-627

Merlotti, A., Alterio, D., Vigna-Taglianti, R., Muraglia, A., Lastrucci, L., Manzo, R., Gambaro, G., Caspiani, O., Miccichè, F., Deodato, F., Pergolizzi, S., Franco, P., Corvò, R., Russi, E. G., & Sanguineti, G. (2014). Technical guidelines for head and neck cancer IMRT on behalf of the Italian association of radiation oncology - head and neck working group. *Radiation Oncology (London, England), 9*(1), 264. 10.1186/s13014-014-0264-9

Michalski, A., Atyeo, J., Cox, J., & Rinks, M. (2012). Inter- and intra-fraction motion during radiation therapy to the whole breast in the supine position: A systematic review. *Journal of Medical Imaging and Radiation Oncology, 56*(5), 499-509. 10.1111/j.1754-9485.2012.02434.x

Minniti, G., Osti, M. F., & Niyazi, M. (2016). Target delineation and optimal radiosurgical dose for pituitary tumors. *Radiation Oncology,* 10.1186/s13014-016-0710-y

Mohamoud, G., MRT(T), Ryan, M., MRT(T), & Moseley, D., PhD. (2015). Inter-observer Variability in Cone Beam CT Image Matching Amongst Radiation Therapists: A Departmental Initiative. *Journal of Medical Imaging and Radiation Sciences, 46*(1), S8. 10.1016/j.jmir.2015.01.028

Mohandass, P., Khanna, D., Nishaanth, B., Saravanan, C., Bhalla, N., Puri, A., & Mohandass, B. (2020). IMPACT OF THREE DIFFERENT MATCHING METHODS ON PATIENT SET-UP ERROR IN X-RAY VOLUMETRIC IMAGING FOR HEAD AND NECK CANCER. *Reports of Practical Oncology and Radiotherapy, 25*(6), 906-912. 10.1016/j.rpor.2020.09.001

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1), 1. 10.1186/2046-4053-4-1

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement*. PLoS Med* 6(7): e1000097. doi:10.1371/journal.pmed1000097

Moola S, Munn Z, Tufanaru C, Aromataris E, Sears K, Sfetcu R, Currie M, Qureshi R, Mattis P, Lisy K, Mu P-F. (2017). *Chapter 7: Systematic reviews of etiology and risk.* ().Aromataris E, Munn Z (Editors).

Muijs, D. (2011). *Doing quantitative research in education with SPSS.* http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=509362

Mul, J., Melchior, P., Seravalli, E., Saunders, D., Bolle, S., Cameron, A. L., Gurtner, K., Harrabi, S., Lassen-Ramshad, Y., Lavan, N., Magelssen, H., Mandeville, H., Boterberg, T., Kroon, P. S., Kotte, A. N. T. J., Hoeben, B. A. W., van Rossum, P. S. N., van Grotel, M., Graf, N., . . . Janssens, G. O. (2021). Inter-clinician delineation variation for a new highly-conformal flank target volume in children with renal tumors: A SIOP-Renal Tumor Study Group international

multicenter exercise. *Clinical and Translational Radiation Oncology, 28*, 39-47. 10.1016/j.ctro.2021.03.001

Mulla, Z., Alwassia, R. K., Senan, E. M., Soaida, S., Mohamed, Ahmed Ahmed Mohamed Ameen, Almerdhemah, H., Iqbal, H. A., & Muamenah, H. M. (2020). A comparative study between open-face and closed-face masks for head and neck cancer (HNC) in radiation therapy. *Reports of Practical Oncology and Radiotherapy, 25*(3), 382-388. 10.1016/j.rpor.2020.03.009

Namysł-Kaletka, A., Tukiendorf, A., & Wydmański, J. (2015). Calculation of planning target volume margins using the van Herk, Stroom and ICRU methods in patients . *Oncology and Radiotherapy, 3 (33)*

Navran, A., Heemsbergen, W., Janssen, T., Hamming-Vrieze, O., Jonker, M., Zuur, C., Verheij, M., Remeijer, P., Sonke, J., van den Brekel, M., & Al-Mamgani, A. (2019). The impact of margin reduction on outcome and toxicity in head and neck cancer patients treated with image-guided volumetric modulated arc therapy (VMAT). *Radiotherapy and Oncology, 130*, 25-31. 10.1016/j.radonc.2018.06.032

Nonaka, H., Onishi, H., Watanabe, M., & Nam, V. H. (2019). Assessment of abdominal organ motion using cine magnetic resonance imaging in different gastric motilities: a comparison between fasting and postprandial states. *Journal of Radiation Research, 60*(6), 837-843. 10.1093/jrr/rrz054

Norfadilah, M. N., Ahmad, R., Heng, S. P., Lam, K. S., Radzi, A. B. A., & John, L. S. H. (2017). Immobilisation precision in VMAT for oral cancer patients. *Journal of Physics. Conference Series, 851*, 12025. 10.1088/1742-6596/851/1/012025

Novak, J., Du, D., Shinde, A., Li, R., Amini, A., Chen, Y., Wong, J., Han, C., Neylon, J., Liu, A., & Glaser, S. (2021). Setup Accuracy in Craniospinal Irradiation: Implications for Planning Treatment Volume Margins. *Advances in Radiation Oncology, 6*(5), 100747. 10.1016/j.adro.2021.100747

Nyarambi, I., Chamunyonga, C., & Pearce, A. (2015). CBCT image guidance in head and neck irradiation: the impact of daily and weekly imaging protocols. *Journal of Radiotherapy in Practice, 14*(4), 362-369. 10.1017/S1460396915000266

Oh, S. A., Yea, J. W., Kang, M. K., Park, J. W., & Kim, S. K. (2016). Analysis of the Setup Uncertainty and Margin of the Daily ExacTrac 6D Image Guide System for Patients with Brain Tumors. *PloS One, 11*(3), e0151709. 10.1371/journal.pone.0151709

Oh, Y., Baek, J., Kim, O., & Kim, J. (2014). Assessment of setup uncertainties for various tumor sites when using daily CBCT for more than 2200 VMAT treatments. *Journal of Applied Clinical Medical Physics, 15*(2), 85-99. 10.1120/jacmp.v15i2.4418

Osman, S. (2011). *Image-Guided Intensity-Modulated Radiotherapy for Single Vocal Cord Irradiation in Early Glottic Cancers* https://www.narcis.nl/publication/RecordID/oai:repub.eur.nl:26739

Palombarini, M., Mengoli, S., Fantazzini, P., Cadioli, C., Degli Esposti, C., & Frezza, G. P. (2012). Analysis of inter-fraction setup errors and organ motion by daily kilovoltage cone beam computed tomography in intensity modulated radiotherapy of prostate cancer. *Radiation Oncology (London, England), 7*(1), 56. 10.1186/1748-717X-7-56

Panjwani, N., Gillespie, E., Murphy, J., Lundy, S., Price, J., D'Souza, L., Sharifzadeh, Y., Shams, P., Onochie, I., Narra, L. R., Lin, D., & Sherer, M. (2019). *eContour*. eContour. https://econtour.org/

Papaioannou, D., Sutton, A., Carroll, C., Booth, A., & Wong, R. (2010). Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Information and Libraries Journal, 27*(2), 114-122. 10.1111/j.1471-1842.2009.00863.x

Parker, B.C., Shiu, A.S., Maor, M.H., Lang, F.F., Liu, H.H., White, R.A. and Antolak, J.A. (2002). PTV margin determination in conformal SRT of intracranial lesions. *Journal of Applied Clinical Medical Physics*, 3: 176-189. https://doi.org/10.1120/jacmp.v3i3.2561

Pearson, W.G. Jr., Davidoff, A.A., Smith, Z.M., Adams, D.E., Langmore, S.E. (2016). Impaired swallowing mechanics of post radiation therapy head and neck cancer patients: A retrospective videofluoroscopic study. *World J Radiol*. 8(2):192-9. doi: 10.4329/wjr.v8.i2.192. PMID: 26981228; PMCID: PMC4770181.

Perillo, A., Landoni, V., Farneti, A., & Sanguineti, G. (2021). Organ motion in linac-based SBRT for glottic cancer. *Radiation Oncology (London, England), 16*(1), 1-106. 10.1186/s13014-021-01833-2

Peter, M. Fayers, F., & David Machin, M. (2013). *Quality of life: assessment, analysis and interpretation*. Wiley-Blackwell.

Peulen, H., Belderbos, J., Guckenberger, M., Hope, A., Grills, I., van Herk, M., & Sonke, J. (2015). Target delineation variability and corresponding margins of peripheral early stage NSCLC treated with stereotactic body radiotherapy. *Radiotherapy and Oncology, 114*(3), 361-366. 10.1016/j.radonc.2015.02.011

Quereshy, Faisal A., MD, DDS, Savell, Truitt A., DDS, MD, & Palomo, J. Martin, DDS, MSD. (2008). Applications of Cone Beam Computed Tomography in the Practice of Oral and Maxillofacial Surgery. *Journal of Oral and Maxillofacial Surgery, 66*(4), 791-796. 10.1016/j.joms.2007.11.018

Ranganathan, P., & Aggarwal, R. (2018). Study designs: Part 1 - An overview and classification. *Perspectives in Clinical Research, 9*(4), 184-186. 10.4103/picr.PICR_124_18

Remeijer, P., Rasch, C., Lebesque, J. V., & van Herk, M. (1999). A general methodology for three-dimensional analysis of variation in target volume delineation. *Medical Physics (Lancaster), 26*(6), 931-940. 10.1118/1.598485

Saidin, Khaliza. (2017). Insider Researchers: Challenges & Opportunities. Proceedings of the ICECRS. 1. 10.21070/picecrs.v1i1.563.

Saleh A.A., Ratajeski M.A., Bertolet M. (2014). Grey Literature Searching for Health Sciences Systematic Reviews: A Prospective Study of Time Spent and Resources Utilized. *Evid Based Libr Inf Pract*, 9(3):28-50. doi: 10.18438/b8dw3k. PMID: 25914722; PMCID: PMC4405801.

Sangoseni, O., Hellman, M., & Hill, C. (2013). Development and Validation of a Questionnaire to Assess the Effect of Online Learning on Behaviors, Attitudes, and Clinical Practices of Physical Therapists in the United States Regarding Evidenced-based Clinical Practice. *The Internet Journal of Allied Health Sciences and Practice,* 10.46743/1540-580X/2013.1439

Schimek-Jasch, T., Troost, E. G. C., Rücker, G., Prokic, V., Avlar, M., Duncker-Rohr, V., Mix, M., Doll, C., Grosu, A., & Nestle, U. (2015). A teaching intervention in a contouring dummy run improved target volume delineation in locally advanced non-small cell lung cancer. *Strahlentherapie Und Onkologie, 191*(6), 525-533. 10.1007/s00066-015-0812-8

Schmidt, M. A., & Payne, G. S. (2015). Radiotherapy planning using MRI. *Physics in Medicine & Biology, 60*(22), R323-R361. 10.1088/0031-9155/60/22/R323

Segedin, B., & Petric, P. (2016). Uncertainties in target volume delineation in radiotherapy – are they relevant and what can we do about them? *Radiology and Oncology, 50*(3), 254-262. 10.1515/raon-2016-0023

Serdar, C. C., Cihan, M., Yücel, D., & Serdar, M. A. (2021). Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochemia Medica, 31*(1), 10502. 10.11613/BM.2021.010502

Setia, M. (2016). Methodology series module 3: Cross-sectional studies. *Indian Journal of Dermatology, 61*(3), 261-264. 10.4103/0019-5154.182410

Shi, J., Mo, X., & Sun, Z. (2012). Content validity index in scale development. *Zhong Nan Da Xue Xue Bao. Yi Xue Ban = Journal of Central South University. Medical Sciences, 37*(2), 152. https://www.ncbi.nlm.nih.gov/pubmed/22561427

Shikha Goyal, & Tejinder Kataria. (2014). Image Guidance in Radiation Therapy: Techniques and Applications. *Radiology Research and Practice, 2014*, 705604-10. 10.1155/2014/705604

Skelly, A., Dettori, J., & Brodt, E. (2012). Assessing bias: the importance of considering confounding. *Evidence-Based Spine-Care Journal, 3*(1), 9-12. 10.1055/s-0031-1298595

Snilstveit, B., Oliver, S., & Vojtkova, M. (2012). Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *Journal of Development Effectiveness, 4*(3), 409-429. 10.1080/19439342.2012.710641

Sousa, F., Jourani, Y., Van den Begin, R., Otte, F., Ridai, S., Desle, M., Ferreira, A., Ahmimed, R., van Klink - de Goeij, Moniek C.M, & Van Gestel, D. (2021). Evaluation of the XVI dual registration tool for image-guided radiotherapy in prostate cancer. *Technical Innovations & Patient Support in Radiation Oncology, 18*, 22-28. 10.1016/j.tipsro.2021.03.003

South Tees Hospital, 2015. Protocol for Head and Neck IMRT, Radiotherapy and Oncology Department.

Srinivasan, K., Mohammadi, M., & Shepherd, J. (2014). Applications of linac-mounted kilovoltage Cone-beam Computed Tomography in modern radiation therapy: A review. *Polish Journal of Radiology, 79*, 181-193. 10.12659/PJR.890745

Stieb, S., Malla, M., Graydon, S., Riesterer, O., Klöck, S., Studer, G., & Tanadini-Lang, S. (2018). Dosimetric influence of pitch in patient positioning for radiotherapy of long treatment volumes; the usefulness of six degree of freedom couch. *British Journal of Radiology, 91*(1091), 20170704. 10.1259/bjr.20170704

Stoll, C. R. T., Izadi, S., Fowler, S., Green, P., Suls, J., & Colditz, G. A. (2019). The value of a second reviewer for study selection in systematic reviews. *Research Synthesis Methods, 10*(4), 539-545. 10.1002/jrsm.1369

Stroom, J. (2000). Safety margins for geometrical uncertainties in radiotherapy. *Medical Physics (Lancaster), 27*(9), 2194. 10.1118/1.1289899

Stroom, J. C., de Boer, H. C. J., Huizenga, H., & Visser, A. G. (1999). Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability. *International Journal of Radiation Oncology, Biology, Physics, 43*(4), 905-919. 10.1016/s0360-3016(98)00468-4

Stroom, J., PhD, Gilhuijs, K., PhD, Vieira, S., PhD, Chen, W., MSc, Salguero, J., PhD, Moser, Elizabeth, MD, PhD, & Sonke, J., PhD. (2014). Combined Recipe for Clinical Target Volume and Planning Target Volume Margins. *International Journal of Radiation Oncology, Biology, Physics, 88*(3), 708-714. 10.1016/j.ijrobp.2013.08.028

Sullivan, G. M. (2011). A Primer on the Validity of Assessment Instruments. *Journal of Graduate Medical Education, 3*(2), 119-120. 10.4300/jgme-d-11-00075.1

Surucu, L., & Maslakci, A. (2020). VALIDITY AND RELIABILITY IN QUANTITATIVE RESEARCH. *Business & Management Studies: An International Journal, 8*(3), 2694-2726. 10.15295/bmij.v8i3.1540

Suzuki, J., Tateoka, K., Shima, K., Yaegashi, Y., Fujimoto, K., Saitoh, Y., Nakata, A., Abe, T., Nakazawa, T., Sakata, K., & Hareyama, M. (2012). Uncertainty in patient set-up margin analysis in radiation therapy. *Journal of Radiation Research, 53*(4), 615-619. 10.1093/jrr/rrs003

Taliana. K., Bezzina. M., Muscat. K., Bartolo. C. (2020). Clinical Protocol: Radiotherapy for Head and Neck Cancer. Sir Anthony Mamo Oncology Centre.

Tang, L. L., Caudy, M., & Taxman, F. (2013). A Statistical Method for Synthesizing Meta-Analyses. *Computational and Mathematical Methods in Medicine, 2013*, 1-9. 10.1155/2013/732989

Taylor, B., & Francis, K. (2013). *Qualitative Research in the Health Sciences*. Routledge. 10.4324/9780203777176

Teoh, M., Clark, C. H., Wood, K., Whitaker, S., & Nisbet, A. (2011). Volumetric modulated arc therapy: a review of current literature and clinical use in practice. *British Journal of Radiology, 84*(1007), 967-996. 10.1259/bjr/22373346

Thasanthan L, Piyasena WRO, Croos AMC, Dhanushia R, Narayanan PP (2014) Assessment of Three-Dimensional Set-up Errors in Pelvic Radiation Therapy. *J Cancer Prev Curr Res* 1(2): 00009. DOI: 10.15406/jcpcr.2014.01.00009

The Royal College of Radiologists. (2021). On Target 2. Updated guidance for image-guided radiotherapy. https://www.rcr.ac.uk

The Royal College of Radiologists, The Society and College of Radiographers, & Institute of Physics and Engineering in Medicine. (2008). On target. Ensuring geometric accuracy in radiotherapy. https://www.rcr.ac.uk/sites/default/files/publication/BFCO%2808%295_On_target.pdf

Thomas, S. J., Evans, B. J., Harihar, L., Chantler, H. J., Martin, A. G. R., & Harden, S. V. (2019). An evaluation of the mid-ventilation method for the planning of stereotactic lung plans. *Radiotherapy and Oncology, 137*, 110-116. 10.1016/j.radonc.2019.04.031

Trafimow, D., & MacDonald, J. A. (2017). Performing Inferential Statistics Prior to Data Collection. *Educational and Psychological Measurement*, 77(2), 204–219. https://doi.org/10.1177/0013164416659745

Trignani, M., Argenone, A., Di Biase, S., Musio, D., Merlotti, A., Ursino, S., Orlandi, E., Genovesi, D., & Bacigalupo, A. (2019). Inter-observer variability of clinical target volume delineation in definitive radiotherapy of neck lymph node metastases from unknown primary. A cooperative study of the Italian Association of Radiotherapy and Clinical Oncology (AIRO) Head and Neck Group. *Radiologia Medica, 124*(7), 682-692. 10.1007/s11547-019-01006-y

Tsang, Y., Hoskin, P., Spezi, E., Landau, D., Lester, J., Miles, E., & Conibear, J. (2019). Assessment of contour variability in target volumes and organs at risk in lung cancer radiotherapy. *Technical Innovations & Patient Support in Radiation Oncology, 10*, 8-12. 10.1016/j.tipsro.2019.05.001

Tudor, G., Bernstein, D., Riley, S., Rimmer, Y., Thomas, S., Herk, M., & Webster, A. (2020). Geometric Uncertainties in Daily Online IGRT: Refining the CTV-PTV Margin for Contemporary Radiotherapy.*DO - 10.1259/geo-unc-igrt* https://www.researchgate.net/publication/346717324_Geometric_Uncertainties_in_Daily_Online_IGRT_Refining_the_CTV-PTV_Margin_for_Contemporary_Radiotherapy

van der Veen, J., Gulyban, A., & Nuyts, S. (2019). Interobserver variability in delineation of target volumes in head and neck cancer. *Radiotherapy and Oncology, 137*, 9-15. 10.1016/j.radonc.2019.04.006

van Herk, M. (2004). Errors and margins in radiotherapy. *Seminars in Radiation Oncology, 14*(1), 52-64. 10.1053/j.semradonc.2003.10.003

van Herk, M., Osorio, E. V., & Troost, E. G. C. (2019). Is reducing irradiated margins key to improving outcomes for radiotherapy? *The Lancet Oncology, 20*(9), 1208-1210. 10.1016/S1470-2045(19)30539-X

van Herk, M., Remeijer, P., Rasch, C., & Lebesque, J. V. (2000). The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. *International Journal of Radiation Oncology, Biology, Physics, 47*(4), 1121-1135. 10.1016/S0360-3016(00)00518-6

van Kranen, S., Mencarelli, A., van Beek, S., Rasch, C., van Herk, M., & Sonke, J. (2013). Adaptive radiotherapy with an average anatomy model: Evaluation and quantification of residual deformations in head and neck cancer patients. *Radiotherapy and Oncology, 109*(3), 463-468. 10.1016/j.radonc.2013.08.007

Velec, M., B.Sc, Waldron, J. N., M.D, O'Sullivan, B., M.D, Bayley, A., M.D, Cummings, B., M.B, Kim, J. J., M.D, Ringash, J., M.D, Breen, S. L., Ph.D, Lockwood, G. A., M.Math, & Dawson, L. A., M.D. (2010). Cone-Beam CT Assessment of Interfraction and Intrafraction Setup Error of Two Head-and-Neck Cancer Thermoplastic Masks. *International Journal of Radiation Oncology, Biology, Physics, 76*(3), 949-955. 10.1016/j.ijrobp.2009.07.004

Velindre Cancer Centre Radiotherapy Treatment Protocol For Head and Neck Cancer, 2014

Verbeek, J., Ruotsalainen, J., & Hoving, J. L. (2012). Synthesizing Study Results in a Systematic Review. *Scandinavian Journal of Work, Environment & Health, 38*(3), 282-290. 10.5271/sjweh.3201

Vinod, S. K., Jameson, M. G., Min, M., & Holloway, L. C. (2016). Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. *Radiotherapy and Oncology, 121*(2), 169-179. 10.1016/j.radonc.2016.09.009

Viswanathan, M., Berkman, N. D., Dryden, D. M., & Hartling, L. (2013). *Assessing risk of bias and confounding in observational studies of interventions or exposures*

Vos, A., Naiker, T., & MacGregor, H. (2020). Determination of an optimal treatment margin for intracranial tumours treated with radiotherapy at Groote Schuur Hospital. *SA Journal of Oncology, 4*(3), e1-e5. 10.4102/sajo.v4i0.124

Waltz C. F. & Bausell R. B. (1981). Nursing research : design statistics and computer analysis. F.A. Davis.

Wareham, K. J., Hyde, R. M., Grindlay, D., Brennan, M. L., & Dean, R. S. (2017). Sponsorship bias and quality of randomised controlled trials in veterinary medicine. *BMC Veterinary Research, 13*(1), 234. 10.1186/s12917-017-1146-9

Williamson, C. W., MAS, Green, G., MD, Noticewala, S. S., MAS, Li, N., PhD, Shen, H., MS, Vaida, F., PhD, & Mell, L. K., MD. (2016). Prospective Validation of A High Dimensional Shape Model for Organ Motion in Intact Cervical Cancer. *International Journal of Radiation Oncology, Biology, Physics, 96*(4), 801-807. 10.1016/j.ijrobp.2016.08.015

Winey, B., & Bussiere, M. (2014). Geometric and dosimetric uncertainties in intracranial stereotatctic treatments for multiple nonisocentric lesions. *Journal of Applied Clinical Medical Physics, 15*(3), 122–132-132. 10.1120/jacmp.v15i3.4668.

Witte, M. G., Sonke, J., Siebers, J., Deasy, J. O., & van Herk, M. (2017). Beyond the margin recipe: the probability of correct target dosage and tumor control in the presence of a dose limiting structure. *Physics in Medicine & Biology, 62*(19), 7874-7888. 10.1088/1361-6560/aa87fe

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. Paper presented at the 1-10. 10.1145/2601248.2601268 http://dl.acm.org/citation.cfm?id=2601268

Worrell, F. C., & Roberson, C. C. B. (2016). 2014 Standards for Educational and Psychological Testing: Implications for ethnic minority youth. In S. L. Graves & J. J. Blake (Eds.), *Psychoeducational assessment and intervention for ethnic minority children: Evidence-based approaches* (pp. 41–57). American Psychological Association. https://doi.org/10.1037/14855-004

Yan Li, W. (2014). *Determining Rational Planning Target Volume Margins for Intracranial Stereotactic Radiotherapy*. Graduate Department of Institute of Medical Science University of Toronto

Yin, W., Sun, Y., Chi, F., Fang, J., Guo, R., Yu, X., Mao, Y., Qi, Z., Guo, Y., Liu, M., & Ma, J. (2013). Evaluation of inter-fraction and intra-fraction errors during volumetric modulated arc therapy in nasopharyngeal carcinoma patients. *Radiation Oncology (London, England), 8*(1), 78. 10.1186/1748-717X-8-78

Yoda, K. (2017). A Radiotherapy Treatment Margin Formula When Systematic Positioning Errors are Relatively Small Compared to Random Positioning Errors: A First-Order Approximation. *International Journal of Medical Physics, Clinical Engineering and Radiation Oncology, 6*(2), 193-196. 10.4236/ijmpcero.2017.62017

Zhang, Q., Song, Y., Chan, M., Burman, C., & Yamada, Y. (2013). Feasibility study of real-time planning for stereotactic radiosurgery. *Medical Physics (Lancaster), 40*(3), 031711-n/a. 10.1118/1.4792637

Zhou, L., Bai, S., Zhang, Y., Ming, X., Zhang, Y., & Deng, J. (2018). Imaging Dose, Cancer Risk and Cost Analysis in Image-guided Radiotherapy of Cancers. *Scientific Reports, 8*(1), 10076-8. 10.1038/s41598-018-28431-9

Bibliography

Abbas, H., Chang, B., & Chen, Z. J. (2014). Motion management in gastrointestinal cancers. *Journal of Gastrointestinal Oncology, 5*(3), 223-235. 10.3978/j.issn.2078-6891.2014.028

Alaei, P., & Spezi, E. (2015). Imaging dose from cone beam computed tomography in radiation therapy. *Physica Medica, 31*(7), 647-658. 10.1016/j.ejmp.2015.06.003

Amaoui, B., Hadaoui, A., Mouhssine, D., & Semghouli, S. (2020). Evaluation of setup errors in conformal radiotherapy for pelvic tumours: Case of the Regional Center of Oncology, Agadir. *Radiation Medicine and Protection, 1*(2), 99-102. 10.1016/j.radmp.2020.05.003

Amelio, D., Winter, M., Habermehl, D., Jäkel, O., Debus, J., & Combs, S. E. (2013). Analysis of inter- and intrafraction accuracy of a commercial thermoplastic mask system used for image-guided particle radiation therapy. *Journal of Radiation Research, 54*(suppl_1), i69-i76. 10.1093/jrr/rrt038

American Society of Radiologic Technologists. (2013). Image guided Radiation Therapy: Intra-fraction Monitoring. http://static.crowdwisdomhq.com/asrt/documents/IGRT_Previews/ACAD13_IGRT_Mod7_Preview.pdf

Aslian, H., Delana, A., Kaiser, S. R., Moretti, E., Foti, C., Bregant, P., de Denaro, M., Longo, F., & Severgnini, M. (2018). A multicenter dosimetry study to evaluate the imaging dose from Elekta XVI and Varian OBI kV-CBCT systems to cardiovascular implantable electronic devices (CIEDs). *Physica Medica, 55*, 40-46. 10.1016/j.ejmp.2018.10.013

Basu, T., & Bhaskar, N. (2019). *Overview of Important "Organs at Risk" (OAR) in Modern Radiotherapy for Head and Neck Cancer (HNC)*. IntechOpen.

Berchtold, A. (2016). Test–retest: Agreement or reliability? *Methodological Innovations, 9*, 205979911667287. 10.1177/2059799116672875

Bolsi, A., M.Sc, Lomax, A. J., Ph.D, Pedroni, E., Ph.D, Goitein, G., M.D, & Hug, E., M.D. (2008). Experiences at the Paul Scherrer Institute With a Remote Patient Positioning Procedure for High-Throughput Proton Radiation Therapy. *International Journal of Radiation Oncology, Biology, Physics, 71*(5), 1581-1590. 10.1016/j.ijrobp.2008.02.079

Buccheri, R. K., & Sharifi, C. (2017). Critical Appraisal Tools and Reporting Guidelines for Evidence-Based Practice. *Worldviews on Evidence-Based Nursing, 14*(6), 463-472. 10.1111/wvn.12258

Buzdar, S. A., Afzal, M., Nazir, A., & Gadhi, M. A. (2013). Accuracy requirements in radiotherapy treatment planning. *Journal of the College of Physicians and Surgeons--Pakistan, 23*(6), 418. https://www.ncbi.nlm.nih.gov/pubmed/23763803

Cao, X., Liu, M., Zhai, F., Li, N., Bao, C., Liu, Y., & Chen, G. (2019). Comparison of different registration methods and landmarks for image-guided radiation therapy of pulmonary tumors. *BMC Medical Imaging, 19*(1), 46. 10.1186/s12880-019-0343-3

Carlson, M. D. A., & Morrison, R. S. (2009). Study Design, Precision, and Validity in Observational Studies. *Journal of Palliative Medicine, 12*(1), 77-82. 10.1089/jpm.2008.9690

Cheng, C. S., Jong, W. L., Ung, N. M., & Wong, J. H. D. (2017). Evaluation of Imaging Dose From Different Image Guided Systems During Head and Neck Radiotherapy: A Phantom Study. *Radiation Protection Dosimetry, 175*(3), 357-362. 10.1093/rpd/ncw357

Cherry, P., & Duxbury, A. M. (2019). *Practical Radiotherapy*. John Wiley & Sons, Incorporated.

Delishaj, D., Ursino, S., Pasqualetti, F., Matteucci, F., Cristaudo, A., Soatti, C. P., Barcellini, A., & Paiar, F. (2018). Set-up errors in head and neck cancer treated with IMRT technique assessed by cone-beam computed tomography: a feasible protocol. *Radiation Oncology Journal, 36*(1), 54-62. 10.3857/roj.2017.00493

Deng, J., & Zhang, Y. (2014). Toward Personalized Imaging in Image-guided Radiotherapy. *Cancer Science & Research: Open Access, 1*(1)10.15226/csroa.2014.00105

Deshpande, S., Dhote, D., Thakur, K., Pawar, A., Kumar, R., Kumar, M., Kulkarni, M. S., Sharma, S. D., & Kannan, V. (2016). Measurement of eye lens dose for Varian On-Board Imaging with different cone-beam computed tomography acquisition techniques. *Journal of Medical Physics, 41*(3), 177-181. 10.4103/0971-6203.189481

Ding, G. X., Ph.D, & Coffey, C. W., Ph.D. (2009). Radiation Dose From Kilovoltage Cone Beam Computed Tomography in an Image-Guided Radiotherapy Procedure. *International Journal of Radiation Oncology, Biology, Physics, 73*(2), 610-617. 10.1016/j.ijrobp.2008.10.006

Ding, G. X., Alaei, P., Curran, B., Flynn, R., Gossman, M., Mackie, T. R., Miften, M., Morin, R., Xu, X. G., & Zhu, T. C. (2018). Image guidance doses delivered during radiotherapy: Quantification, management, and reduction: Report of the AAPM Therapy Physics Committee Task Group 180. *Medical Physics, 45*(5), e84-e99. 10.1002/mp.12824

Dzierma, Y., Beyhs, M., Palm, J., Niewald, M., Bell, K., Nuesken, F., Licht, N., & Rübe, C. (2015). Set-up errors and planning margins in planar and CBCT image-guided radiotherapy using three different imaging systems: A clinical study for prostate and head-and-neck cancer. *Physica Medica, 31*(8), 1055-1059. 10.1016/j.ejmp.2015.09.002

Faber, J., & Fonseca, L. M. (2014). How sample size influences research outcomes. *Dental Press Journal of Orthodontics, 19*(4), 27-29. 10.1590/2176-9451.19.4.027-029.ebo

Fitchett, R. C., Aldus, E. J., Fitchett, L. R., & Cross, J. (2018). The lived experience of head and neck cancer patients receiving curative radiotherapy: A systematic review and meta-ethnography. *Psycho-Oncology (Chichester, England), 27*(9), 2077-2086. 10.1002/pon.4790

Forker, L., Stratford, J., Whitehurst, P., Slevin, N., Anandadas, C., & Choudhury, A. (2015). Practice Change after Evaluation of an Offline Correction Protocol for Image-guided Radiotherapy in Head and Neck Cancer. *Clinical Oncology, 27*(12), 750-751. 10.1016/j.clon.2015.07.001

Girard, J. M., & Cohn, J. F. (2016). A Primer on Observational Measurement. *Assessment (Odessa, Fla.), 23*(4), 404-413. 10.1177/1073191116635807

Goldsworthy, S., Leslie-Dakers, M., Higgins, S., Barnes, T., Jankowska, P., Dogramadzi, S., & Latour, J. M. (2017). A Pilot Study Evaluating the Effectiveness of Dual-Registration Image-Guided Radiotherapy in Patients with Oropharyngeal Cancer. *Journal of Medical Imaging and Radiation Sciences, 48*(4), 377-384. 10.1016/j.jmir.2017.09.004

Gregoire, V., & Mackie, T. R. (2011). Dose prescription, reporting and recording in intensity-modulated radiation therapy: a digest of the ICRU Report 83. *Imaging in Medicine, 3*(3), 367-373. 10.2217/iim.11.22

Harrison, J. K., Reid, J., Quinn, T. J., & Shenkin, S. D. (2017). Using quality assessment tools to critically appraise ageing research: a guide for clinicians. *Age and Ageing, 46*(3), 359-365. 10.1093/ageing/afw223

Higgins, J. (2019). *Cochrane handbook for systematic reviews of interventions* (Second edition ed.). Cochrane, Wiley Blackwell.

Howlin, C., O'Shea, E., Dunne, M., Mullaney, L., McGarry, M., Clayton-Lea, A., Finn, M., Carter, P., Garret, B., & Thirion, P. (2014). A randomized controlled trial comparing customized versus standard headrests for head and neck radiotherapy immobilization in terms of set-up errors,

patient comfort and staff satisfaction (ICORG 08-09). *Radiography, 21*(1), 74-83. 10.1016/j.radi.2014.07.009

Jadon, R., Pembroke, C. A., Hanna, C. L., Palaniappan, N., Evans, M., Cleves, A. E., & Staffurth, J. (2014). A Systematic Review of Organ Motion and Image-guided Strategies in External Beam Radiotherapy for Cervical Cancer. *Clinical Oncology (Royal College of Radiologists (Great Britain)), 26*(4), 185-196. 10.1016/j.clon.2013.11.031

Jamil, N. (2020). Understanding probability sampling techniques : Simple Random Sampling, Systematic sampling, Stratified sampling and Cluster sampling. https://www.researchgate.net/publication/346311717_Understanding_probability_sampling_techniques_Simple_Random_Sampling_Systematic_sampling_Stratified_sampling_and_Cluster_sampling

Jin, Y., Chen, D., Cabay, R. J., Wang, A., Crowe, D. L., & Zhou, X. (2013). Role of microRNA-138 as a Potential Tumor Suppressor in Head and Neck Squamous Cell Carcinoma. *International Review of Cell and Molecular Biology* (pp. 357-385). Elsevier Science & Technology. 10.1016/B978-0-12-407697-6.00009-X

Kei Ma, W., Hogg, P., & Norton, S. (2014). Effects of kilovoltage, milliampere seconds, and focal spot size on image quality. *Radiologic Technology, 85*(5), 479. https://www.ncbi.nlm.nih.gov/pubmed/24806050

Kim, S. H., Oh, S. A., Yea, J. W., & Park, J. W. (2019). Prospective assessment of inter- or intra-fractional variation according to body weight or volume change in patients with head and neck cancer undergoing radiotherapy. *PloS One, 14*(5), e0216655. 10.1371/journal.pone.0216655

Ko, C., & Citrin, D. (2009). Radiotherapy for the management of locally advanced squamous cell carcinoma of the head and neck. *Oral Diseases, 15*(2), 121-132. 10.1111/j.1601-0825.2008.01495.x

Krishna Mohan, Surapaneni, Mrsc, & Mohan. (2010). Research Bias: A Review For Medical Students. *Journal of Clinical and Diagnostic Research, 4*

Kumar, A. S., Singh, I. R. R., Sharma, S. D., John, S., & Ravindran, B. P. (2016). Radiation dose measurements during kilovoltage-cone beam computed tomography imaging in radiotherapy. *Journal of Cancer Research and Therapeutics, 12*(2), 858-863. 10.4103/0973-1482.164699

Langsenlehner, T., Döller, C., Winkler, P., Gallé, G., & Kapp, K. S. (2013). Impact of inter- and intrafraction deviations and residual set-up errors on PTV margins. *Strahlentherapie Und Onkologie, 189*(4), 321-328. 10.1007/s00066-012-0303-0

Lin, C., Xu, S., Yao, W., Wu, Y., Fang, J., & Wu, V. W. C. (2017). Comparison of set up accuracy among three common immobilisation systems for intensity modulated radiotherapy of nasopharyngeal carcinoma patients. *Journal of Medical Radiation Sciences, 64*(2), 106-113. 10.1002/jmrs.189

Loganathan, T. (2014). Assessment of Three-Dimensional Set-up Errors in Pelvic Radiation Therapy. *Journal of Cancer Prevention & Current Research, 1*(2)10.15406/jcpcr.2014.01.00009

Lowther, N., Marsh, S., Louwe, R. (2020). Dose accumulation to assess the validity of treatment plans with reduced margins in radiotherapy of head and neck cancer, *Physics and Imaging in Radiation Oncology*, Volume 14, Pages 53-60, ISSN 2405-6316, https://doi.org/10.1016/j.phro.2020.05.004.

Machiels, J., Lambrecht, M., Hanin, F., Duprez, T., Gregoire, V., Schmitz, S., & Hamoir, M. (2014). Advances in the management of squamous cell carcinoma of the head and neck. *F1000 Prime Reports, 6*, 44. 10.12703/p6-44

Masa-Nga, W., Tannanonta, C., Khamfongkhruea, C., Thongsawad, S., Laebua, K., & Apiwarodom, N. (2018). Assessment of Inter-fraction Target Motion using On-board Imaging System for Prostate Cancer. *Journal of the Medical Association of Thailand, 101 (6):87*

Mozdarani, H. (2012). Biological Complexities in Radiation Carcinogenesis and Cancer Radiotherapy: Impact of New Biological Paradigms. *Genes, 3*(1), 90-114. 10.3390/genes3010090

Mukesh, M., Benson, R., Jena, R., Hoole, A., Roques, T., Scrase, C., Martin, C., Whitfield, G. A., Gemmill, J., & Jefferies, S. (2012). Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: can segmentation protocols help? *British Journal of Radiology, 85*(1016), e530-e536. 10.1259/bjr/66693547

Neubauer, E., Dong, L., Followill, D. S., Garden, A. S., Court, L. E., White, R. A., & Kry, S. F. (2012). Assessment of shoulder position variation and its impact on IMRT and VMAT doses for head and neck cancer. *Radiation Oncology (London, England), 7*(1), 19. 10.1186/1748-717X-7-19

Ong, C. L., Giaj-Levra, N., Nicosia, L., Figlia, V., Tomasini, D., Franken, E. M., & Alongi, F. (2020). Intra-fraction and Inter-fraction analysis of a dedicated immobilization device for intracranial radiation treatment. *Radiation Oncology (London, England), 15*(1), 1-200. 10.1186/s13014-020-01639-8

Pádraig Wright, Julian Stern, & Michael Phelan. (2011). *Core Psychiatry* (third edition ed.). Elsevier.

Parvathaneni, U., Laramore, G. E., & Liao, J. J. (2012). Technical Advances and Pitfalls in Head and Neck Radiotherapy. *Journal of Oncology, 2012*, 597467-13. 10.1155/2012/597467

Pastor, M., Lopez Pousa, A., del Barco, E., Perez Segura, P., Astorga, B., Castelo, B., Bonfill, T., Martinez Trufero, J., Grau, J., & Mesia, R. (2018). SEOM clinical guideline in nasopharynx cancer (2017). *Clinical & Translational Oncology, 20*(1), 84-88. 10.1007/s12094-017-1777-0

Peng, J. L., Liu, C., Chen, Y., Amdur, R. J., Vanek, K., & Li, J. G. (2011). Dosimetric consequences of rotational setup errors with direct simulation in a treatment planning system for fractionated stereotactic radiotherapy. *Journal of Applied Clinical Medical Physics, 12*(3), 61-70. 10.1120/jacmp.v12i3.3422

Radiation Therapy Oncology Group. (2014). RTOG 1016 phase III trial of radiotherapy plus cetuximab versus chemoradiotherapy in HPV-associated oropharynx cancer.

Rana, S. (2013). Intensity modulated radiation therapy versus volumetric intensity modulated arc therapy. *Journal of Medical Radiation Sciences, 60*(3), 81-83. 10.1002/jmrs.19

Roach, D., Jameson, M. G., Dowling, J. A., Ebert, M. A., Greer, P. B., Kennedy, A. M., Watt, S., & Holloway, L. C. (2018). Correlations between contouring similarity metrics and simulated treatment outcome for prostate radiotherapy. *Physics in Medicine & Biology, 63*(3), 035001. 10.1088/1361-6560/aaa50c

Ryan R; Cochrane Consumers and Communication Review Group. 'Cochrane Consumers and Communication Review Group: data synthesis and analysis'. http://cccrg.cochrane.org, June 2013

Saha, A., Mallick, I., Das, P., Shrimali, R. K., Achari, R., & Chatterjee, S. (2015). Evaluating the Need for Daily Image Guidance in Head and Neck Cancers Treated with Helical Tomotherapy: A Retrospective Analysis of a Large Number of Daily Imaging-based Corrections. *Clinical Oncology, 28*(3), 178-184. 10.1016/j.clon.2015.11.014

Sánchez-Rubio, P., Rodríguez-Romero, R., & Castro-Tejero, P. (2014). A retrospective tomotherapy image-guidance study: analysis of more than 9,000 MVCT scans for ten different tumor sites. *Journal of Applied Clinical Medical Physics, 15*(6), 30-45. 10.1120/jacmp.v15i6.4663

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., . . . Higgins, J. P. T. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. *Bmj, 366*, l4898. 10.1136/bmj.l4898

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *Bmj, 343*(jul22 1), d4002. 10.1136/bmj.d4002

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., . . . Higgins, J. P. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *Bmj, 355*, i4919. 10.1136/bmj.i4919

Stratton, S. J. (2021). Population Research: Convenience Sampling Strategies. *Prehospital and Disaster Medicine, 36*(4), 373-374. 10.1017/S1049023X21000649

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53-55. 10.5116/ijme.4dfb.8dfd

Thomson, H., Craig, P., Hilton-Boon, M., Campbell, M., & Katikireddi, S. V. (2018). Applying the ROBINS-I tool to natural experiments: an example from public health. *Systematic Reviews, 7*(1), 15. 10.1186/s13643-017-0659-4

Thorlund, K., Imberger, G., Johnston, B. C., Walsh, M., Awad, T., Thabane, L., Gluud, C., Devereaux, P. J., & Wetterslev, J. (2012). Evolution of Heterogeneity (I2) Estimates and Their 95% Confidence Intervals in Large Meta-Analyses. *PloS One, 7*(7), e39471. 10.1371/journal.pone.0039471

Tiong, A., Lao, L., MacKean, J., Goonetilleke, M., & Kron, T. (2016). Faculty of Radiation Oncology Position Paper on the use of Image–Guided Radiation Therapy. *Journal of Medical Imaging and Radiation Oncology, 60*(6), 772-780. 10.1111/1754-9485.12463

Tran, Cindy, HonBSc, MRT(T), Sae, Stephanie, MRSc, MRT(T), Moseley, D., Dr, Kassam, Z., Dr, Cho, C., Dr, Smith, K., & Wong, J., Dr. (2014). Use of 4D-CBCT to Assess Target and Organ Motion in Esophageal Cancer Patients. *Journal of Medical Imaging and Radiation Sciences, 45*(2), 185. 10.1016/j.jmir.2014.03.080

Trotti, A., Harris, J., Gillison, M., Eisbruch, A., Harari, P. M., Adelstein, D. J., Sturgis, E. M., Galvin, J. M., Koyfman, S., Blakaj, D., Razaq, M. A., Colevas, A. D., Beitler, J. J., Jones, C. U., Dunlap, N. E., Seaward, S. A., Spencer, S. A., Ridge, J. A., Phan, J., & Le, Q. T. (2018). NRG-RTOG 1016: Phase III Trial Comparing Radiation/Cetuximab to Radiation/Cisplatin in HPV-related Cancer of the Oropharynx. *International Journal of Radiation Oncology, Biology, Physics, 102*(5), 1604-1605. 10.1016/j.ijrobp.2018.08.051

Ursachi, G., Horodnic, I. A., & Zait, A. (2015). How Reliable are Measurement Scales? External Factors with Indirect Influence on Reliability Estimators. *Procedia Economics and Finance, 20*, 679-686. 10.1016/s2212-5671(15)00123-9

van Kranen, S. R. (2017). *Geometric uncertainties and mitigation strategies in radiotherapy of head & neck cancer* https://www.narcis.nl/publication/RecordID/oai:dare.uva.nl:publications%2F10419ddc-46d2-4525-941a-21e3491d545a

Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., Poole, C., Schlesselman, J. J., & Egger, M. (2014). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *International Journal of Surgery (London, England), 12*(12), 1500-1524. 10.1016/j.ijsu.2014.07.014

Von Hippel, P. T. (2015). The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC Medical Research Methodology, 15*(1)10.1186/s12874-015-0024-z

Yu, Y., Michaud, A. L., Sreeraman, R., Liu, T., Purdy, J. A., & Chen, A. M. (2014). Comparison of daily versus nondaily image-guided radiotherapy protocols for patients treated with intensity-modulated radiotherapy for head and neck cancer. *Head & Neck, 36*(7), 992-997. 10.1002/hed.23401

Zeidan, O. A., Ph.D, Langen, K. M., Ph.D, Meeks, S. L., Ph.D, Manon, R. R., M.D, Wagner, T. H., Ph.D, Willoughby, T. R., M.S, Jenkins, D. W., M.D, & Kupelian, P. A., M.D. (2007). Evaluation of image-guidance protocols in the treatment of head and neck cancers. *International Journal of Radiation Oncology, Biology, Physics, 67*(3), 670-677. 10.1016/j.ijrobp.2006.09.040

Zhang, X., Shan, G., Liu, J., & Wang, B. (2016). Margin evaluation of translational and rotational set-up errors in intensity modulated radiotherapy for cervical cancer. *SpringerPlus, 5*(1), 1-9. 10.1186/s40064-016-1796-2

Appendices

Appendix A

Journal of Radiotherapy in Practice

cambridge.org/jrp

PTV margin calculation for head and neck patients treated with VMAT: a systematic literature review

Kristina Caruana[1], Nick Refalo Dr[2], Denise Spiteri[2], José Guilherme Couto[1], Frank Zarb Dr[1] and Paul Bezzina Dr[1]

[1]University of Malta, Msida Campus, Msida, Malta and [2]Sir Anthony Mamo Oncology Centre, Msida, Malta

Literature Review

Cite this article: Caruana K, Refalo N, Spiteri D, Couto JG, Zarb F, and Bezzina P. (2021) PTV margin calculation for head and neck patients treated with VMAT: a systematic literature review. *Journal of Radiotherapy in Practice* page 1 of 8. doi: 10.1017/S1460396921000546

Received: 31 May 2021
Revised: 29 July 2021
Accepted: 17 August 2021

Key words:
Head and Neck; VMAT; Planning Target Volume; PTV; Van Herk; Margin; Error; Systematic literature review

Abbreviations:
PTV, Planning Target Volume; VMAT#, Volumetric Arc Therapy; CBCT, Cone Beam Computed Tomography

Author for correspondence:
Kristina Caruana, University of Malta, Msida Campus, Msida, Malta, MSD 2080.
E-mail: kristina.caruana@gov.mt

Abstract

Aim: The intent of the review was to identify different methodological approaches used to calculate the planning target volume (PTV) margin for head and neck patients treated with volumetric arc therapy (VMAT), and whether the necessary factors to calculate the margin size with the selected formula were used.

Materials and Methods: A comprehensive, systematic search of related studies was done using the Hydi search engine and different databases: MEDLINE, PubMed, CINAHL, ProQuest (Nursing and Allied Health), Scopus, ScienceDirect and tipsRO. The literature search included studies published between January 2007 and December 2020. Eligibility screening was performed by two reviewers.

Results: A total of seven studies were found. All the reviewed studies used the Van Herk formula to measure the PTV margin. None of the studies incorporated the systematic errors of target volume delineation in the PTV equation. Inter-fraction translational errors were assessed in all the studies, whilst intra-fraction errors were only included in the margin equation for two studies. The studies showed great heterogeneity in the key characteristics, aims and methods.

Findings: Since systemic errors from target volume delineation were not considered and not all studies assess intra-fraction errors, PTV margins may be underestimated. The recommendations are that studies need to determine the effect of target volume variance on PTV margins. It is also recommended to compare PTV margin results using various formulas.

Introduction

The planning target volume (PTV) concept was firstly introduced in the International Commission on Radiation Units and Measurements (ICRU) Report 50[1]. The PTV includes the clinical target volume (CTV) – which is the volume that encompasses the clinical and subclinical disease – plus a margin to account for internal movements and set-up errors[1]. According to the ICRU 83 report, PTV margin size should be calculated for each radiotherapy department, this is because set-up procedures, treatment modalities and imaging modalities are some of the department-specific factors that can influence the number of movements and set-up errors that must be accounted for in this margin. This report, however, did not specify the method to be used for margin calculation[2].

There are also several PTV margin formulas such as Van Herk, Stroom, ICRU 62 formula, Antolak, Bel, McKenzie and Parker, and the selection of the method of calculation could have an influence on the margin size[3].

A systematic literature review was conducted to identify the methodological approaches used to calculate the PTV margin in the head and neck region with volumetric arc therapy (VMAT) across published literature. To achieve this aim, the objectives of this review were to identify the formulas used to calculate the PTV margin and whether the reviewed studies considered the necessary factors to calculate the margin size with the formula used.

Methods

A systematic literature review was performed to answer the question: What methods are employed by studies to calculate the PTV margin for patients treated with VMAT to the head and neck region?

The Preferred Reporting Item for Systematic Review (PRISMA) checklist was used to guide the write-up of the systematic literature review protocol (PRISMA checklist, 2009). The following demonstrated the Population, Intervention, Comparison, Outcome and Study Design (PICOS) framework which was used to guide the literature search[4]:

CAMBRIDGE UNIVERSITY PRESS

https://doi.org/10.1017/S1460396921000546 Published online by Cambridge University Press

Population – Patients receiving radiotherapy in the form of VMAT to the head and neck region.

Intervention – Calculation of the PTV margin.

Comparison – No comparison made

Outcomes – Identification the PTV margin formula and the factors considered.

Study Design – Quantitative studies.

A systematic search on databases in medicine and radiotherapy: MEDLINE, PubMed, CINAHL, ProQuest (Nursing and Allied Health), Scopus, tipsRO and ScienceDirect. The Hydi search engine was also used to find related studies.

Keywords were validated by two experts in the field and content validity was used to assess the validation of the keywords. There was 100% mean agreement and minor suggestions were taken on board.

The Medical Subject Headings (MeSH) thesaurus was used to search for other related words that could be used as keywords. The asterisks (*) next to the keywords identified other terms that are written in different ways and Boolean operators were also used to allow combination of words and phrases to expand the search.

The following combinations of keywords were used to search for related studies:

- Nasopharyn*/Nasal cavity
- Oropharyn*
- Laryn*/Supraglottis/Subglottis/Glottis
- Hypopharyn*
- Oral cavity/Mouth/Tongue
- Sinus*
- Thyroid
- Lymphoma
- Head and Neck
- Set-up/setup/set up
- Error/errors
- VMAT/Volumetric-Modulated Arc Therapy/Volumetric Modulated Arc Therapy/RapidArc Therapy
- PTV/Planning Target Volume

Publication period (1 January 2007 to 30 December 2020), full-text and human species filters were used to aid in the selection process. Figure 1 is a representation of the search strategy that was used on PubMed.

Only quantitative studies related to the calculation of the PTV margin for patients being treated with VMAT to the head and neck region were included. The review was also restricted to English language studies.

A dual independent literature search was done by two researchers with over 5 years of clinical experience. The reviewers performed a separate search for the literature using the same research criteria. The search was done between April and December 2020. In the first phase of the review, the literature was screened for the inclusion criteria based only on the title and abstract. For the second phase of the review, a full-text reading of the studies was performed on the eligible studies that were selected in the first phase. Disagreements with regard to data suitability were resolved by consensus between reviewers.

Meta-analysis was not performed due to the heterogeneity in the methodological, statistical and clinical sources; therefore, a narrative synthesis approach was selected[5].

Results

A total of 4341 articles were found. After removing duplicated articles and screening the studies for eligibility, a total of seven relevant studies were found. Figure 2 shows the PRISMA flow chart and Table 1 presents a summary of the study characteristics.

Methods to calculate the PTV margin

The PTV margin was calculated for all the studies in the review using the Van Herk formula. Both inter- and intra-fraction motion were measured in Bruijnen et al.[6] and Yin et al.[7] studies, whilst the other studies derived the margin by evaluating the inter-fraction errors. None of the reviewed studies assessed and included target delineation variation in the margin formula. Table 2 demonstrates the methods opted by the reviewed studies to calculate the PTV margin.

PTV margin size

There was a discrepancy in the PTV margin results of the reviewed studies; however, this was expected since the factors influencing the margin differ from department to department.

The largest discrepancy was found in Anjanappa et al.[8] and Yin et al.[7] studies. These studies analysed the margin in the nasopharyngeal region, however, Yin et al.[7] made use of cone beam computed tomography (CBCT) imaging and assessed inter- and intra-fraction errors, whilst Anjanappa et al.[8] made use of two orthogonal images and evaluated the inter-fraction errors. Intra-fractional contributes to create a larger PTV margin[9]; however, in these two studies, the smallest PTV margin result was in the study that did not evaluate intra-fraction errors. This discrepancy in the margin result could be attributed to the imaging modalities. The chosen modality has an impact on the set-up error that is detected. CBCT should be the modality of choice, since it allows for better observation of the volumes of interest[10].

Oh et al.[11] and Kukolowicz et al.[12] (post-no action level protocol) had the most similar PTV margin result. Both studies, however, varied in immobilisation devices, imaging protocols and outcomes. The similarity of results was most likely by chance. In the reviewed studies, the medial-lateral (ML) margin was not measured in the larynx and oropharynx region. Comparison of margins in the different regions of the head and neck was also not possible, since the studies did not provide data regarding the different areas.

Inter-fraction errors

The reviewed studies obtained similar results for population systematic errors with the standard deviation (SD) of the systematic errors of the reviewed studies measured to be 0·4 mm, 0·5 mm and 0·5 mm in the ML, superior-inferior (SI) and anterior-posterior (AP) direction, respectively. The SD for population random errors resulted to be slightly higher than that of the population systematic errors, with each direction (ML, SI and AP) obtaining a value of 0·6 mm.

Deb et al.[13] study obtained the highest population random error. This study treated patients without daily imaging, instead a total of 10 CBCT images were acquired for each patient. Population systematic and random errors can be corrected prior treatment with daily imaging; however, in studies where daily imaging are not performed, random error can not be compensated. For this reason, the PTV margin of Deb et al.[13] study resulted to be measured larger when compared with other studies in the review.

Rotational errors were analysed in four studies. Oh et al.[11] study compared rotational errors in different anatomical regions. The rotational error for head and neck region was below 3° and this value was small when compared with other anatomical regions. Norfadilah et al.[14] study also calculated rotational errors with

https://doi.org/10.1017/S1460396921000546 Published online by Cambridge University Press

Query	Results
Search: **(PTV OR Planning Target Volume) AND (VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR Rapid Arc Therapy) AND (set up OR setup OR set-up) AND (error or errors) AND (Hypopharyn*)** Filters: **Full text, Humans, English, from 2007 - 2020**	1
Search: **(PTV OR Planning Target Volume) AND (VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR Rapid Arc Therapy) AND (Oral cavity OR mouth OR tongue) AND (Setup OR set-up OR set up) AND (error or errors)** Filters: **Full text, Humans, English, from 2007 - 2020**	12
Search: **(PTV OR Planning Target Volume) AND VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR Rapid Arc Therapy) AND (oropharyn*) AND (set up OR setup OR set-up) AND (error OR errors)** Filters: **Full text, Humans, English, from 2007 - 2020**	15
Search: **(PTV OR Planning Target Volume) AND VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR Rapid Arc Therapy) AND (Nasopharyn* OR Nasal cavity) AND (set up OR setup OR set-up) AND (error OR errors)** Filters: **Full text, Humans, English, from 2007 - 2020**	28
Search: **(PTV OR Planning Target Volume) AND (VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR Rapid Arc Therapy) AND (Laryn* OR supraglottis OR subglottis OR glottis) AND (Setup OR set-up OR set up) AND (error OR errors)** Filters: **Full text, Humans, English, from 2007 - 2020**	12
Search: **(PTV OR Planning Target Volume) AND (VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR RapidArc Therapy) AND (Head and Neck) AND (Setup OR set-up OR set up) AND (error)** Filters: **Full text, Humans, English, from 2007 - 2020**	55
Search: **(PTV OR Planning Target Volume) AND (VMAT OR Volumetric-Modulated Arc Therapy OR Volumetric Modulated Arc Therapy OR Rapid Arc Therapy) AND (Thyroid OR Sinus* OR lymphoma) AND (Setup OR set-up OR set up) AND (error OR errors)** Filters: **Full text, Humans, English, from 2007 - 2020**	7

Figure 1. Search strategy.

the aim of comparing two tongue immobilisation devices. Average rotational errors result for headFIX® mouthpiece and syringe mouthpiece were 0·00°±0·65° and 0·34°±0·59°, respectively. In Kukolowicz et al.[12] rotations larger than one degree were seldom observed; therefore, these errors were not taken into consideration. This study, however, performed 2D imaging, therefore, rotational results from multiple perspective were not analysed. Small values of rotations were also observed in Yin et al.[7] study, with the number of fractions rarely exceeding 2° for pitch, roll and yaw directions.

Intra-fraction translational errors

Bruijnen et al.[6] measured intra-fraction errors from 2D cine magnetic resonance imaging (MRI) and deformable image registration. In this study, respiratory tumour motion, swallowing, tongue motion and set-up errors were investigated to determine the PTV margin size. When the tumour motion was incorporated into the PTV margin formula, the margin expanded by 0·6 mm for oropharyngeal tumours, 0·2 mm for nasopharyngeal tumours and 1·7 mm for laryngeal tumours[6].

In Yin et al. study[7], the intra-fraction population systematic error during the 5–9 min VMAT period ranged from 0·2 mm to 0·4 mm, and the population random error ranged from 0·5 mm to 0·6 mm.

Discussion

PTV margin equations

In radiotherapy, there is an issue on the method selected to determine the PTV margin. The Van Herk formula is a widely used strategy for PTV margin calculation, and this equation was used in all the reviewed seven studies. The reason for selecting the formula was not specified and lack of comparison of this formula against other options for head and neck was identified as a major gap in the literature.

A study by Namysł-Kaletka, Tukiendorf and Wydmański[15] used three formulas; Van Herk, Stroom and ICRU, to assess PTV margin results based on set-up errors for gastric cancer patients. The margin results were compared, and the study

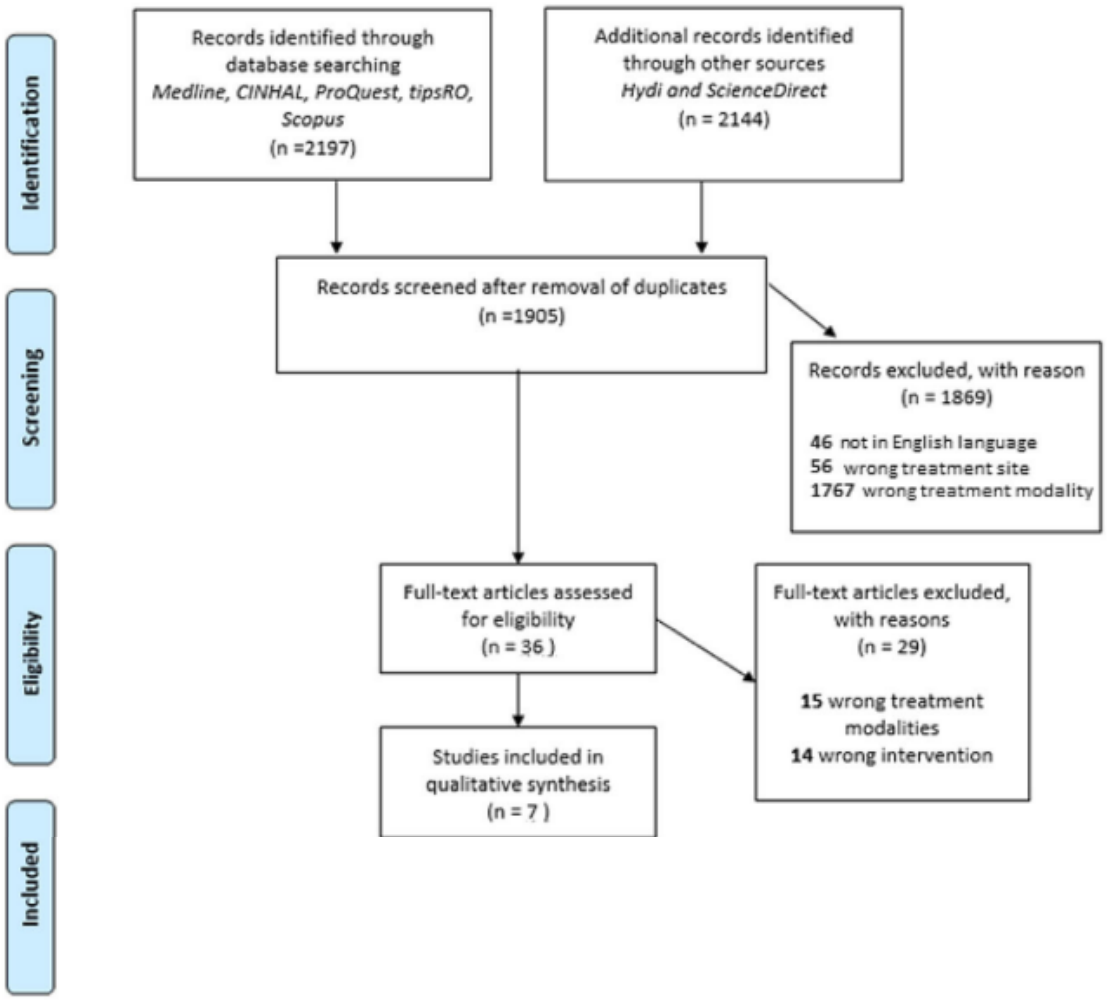


Figure 2. PRISMA 2009 flow chart *(Adapted from: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses).*

revealed that the formula being used has an impact in the PTV margin. As such, the formula should be selected carefully.

The Van Herk's formula assumes that the CTV is spherical in shape, the tissue is homogeneous, conformal beam penumbra and that the number of fractions is infinite[16]. As such, the use of this formula for the PTV margin calculation should be used with caution when these assumptions do not apply.

Due to the exclusion of rotational errors and shape variation, the Van Herk's formula should be considered a lower limit for the delivery of safe radiotherapy. The formula guarantees that 90 per cent of patients receive a minimum of 95 per cent of the recommended dose in the CTV[16]. Therefore, this formula seems to be adequate for the calculation of PTV margin in head and neck patients.

PTV margin size

The studies in the review had different PTV margin size and this continues to necessitate the importance of the departments to calculate their own specific margins. Margin sizes seem to be affected by the imaging protocols and immobilisation devices.

Yin et al.[7] study demonstrated the effect of daily imaging on the margin size for nasopharyngeal patients. The resulting margin size was small when compared to other studies in the review, since the PTV margin was calculated on the set-up errors obtained after CBCT correction. Kukolowicz et al.[12] stated that daily online correction was slightly better than the NAL (no action level) protocol for patients having treatment to the head and neck region. However, the study failed to analyse rotational errors and could not assess anatomical changes since these were not visible on portal images.

Daily imaging protocols should be adhered for patients treated with a tight PTV margin, such as intensity-modulated radiotherapy or VMAT. Daily imaging aids in verifying the set-up position, identifying the location of the target, assessing tumour shrinkage and making the necessary corrections prior to each exposure[11].

https://doi.org/10.1017/S1460396921000546 Published online by Cambridge University Press

Table 1. Summary of characteristics of studies included in the narrative synthesis

Author, year and country	Study design	Head and neck region	Imaging protocol	Immobilisation device	PTV formula	PTV margin result
Yin et al., 2013 Southern China	Prospective Observational Analytical and Cross-sectional	Nasopharynx	Daily CBCT	5-point TP mask HR not specified	VHMF (inter- and intra-fraction errors)	*Total without CBCT correction*: ML = 4·1 mm SI = 3·4 mm AP = 3·5 mm *Total with CBCT correction*: ML = 1·7 mm SI = 2·2 mm AP = 2·2 MM
Oh et al., 2014 South Korea	Retrospective Observational Analytical and Cross-sectional	Not specified	Daily CBCT	5-point TP mask Individual HR	VHMF (inter-fraction error)	ML = 3·3 mm SI = 2·8 mm AP = 3·7 mm
Anjanappa et al., 2017 India	Retrospective Observational Analytical and Cross-sectional	Nasopharynx	Daily 2D KV imaging (KV images taken on alternate days were reviewed)	4-point TP mask HR not specified	VHMF (inter-fraction error)	*Clivus level:* ML = 4·0 mm SI = 3·2 mm AP = 4·4 mm *C3 level:* ML = 5·0 mm SI = 4·4 mm AP = 5·5 mm *C6 level:* ML = 6·9 mm SI = 4·4 mm AP = 6·4 mm
Norfadilah et al., 2017 Malaysia	Prospective Observational Analytical and Cross-sectional	Oral cancer	Daily CBCT	5-point TP mask Mouth Bite HR not specified	VHMF (inter-fraction error)	*HFW mouthbite*: ML = 3·1 mm SI = 2·2 mm AP = 0·8 mm *SYR*: ML = 3·8 mm SI = 6·2 mm AP = 5·1 mm
Bruijnen et al., 2018 Netherlands	Prospective Observational Analytical and Cross-sectional	Nasopharynx Oropharynx Larynx	eNAL	5-point TP mask Individual HR	VHMF (inter and intra-fraction errors)	*Nasopharynx*: S = 2·8 mm I = 2·8 mm A = 2·8 mm P = 2·8 mm *Oropharynx*: S = 3·0 mm I = 3·1 mm A = 3·0 mm P = 3·0 mm *Larynx*: S = 4·0 mm I = 3·6 mm A = 3·1 mm P = 3·1 mm *Combined*: S = 3·3 mm

(*Continued*)

Table 1. (*Continued*)

Author, year and country	Study design	Head and neck region	Imaging protocol	Immobilisation device	PTV formula	PTV margin result
						I = 3·2 mm A = 3·0 mm P = 3·0 mm
Deb et al., 2019 India	Retrospective Observational Analytical and Cross-sectional	Not specified	Daily imaging (eNAL for CBCT an remaining days with 2D PI)	TP mask with shoulder retraction Standard HR	VHMF (inter-fraction error)	ML = 5·6 mm SI = 6·1 mm AP = 4·7 mm
Kukolowicz et al., 2020 Poland	Retrospective Observational Case-control	Nasopharynx and larynx	Daily EPID	5-point TP mask Standard HR	VHMF (inter-fractional error)	*Prior NAL protocol:* AP = 4·0 mm SI = 6·0 mm ML = 4·0 mm *NAL protocol:* AP = 3·0 mm SI = 2·2 mm ML = 3·0 mm

N = sample number; S = superior; I = inferior; A = anterior; P = posterior; ML = medial-lateral; SI = superior-inferior; AP = anteriorterior; C3 = cervical spine level 3; C6 = cervical spine level 6; TP = thermoplastic; HR = head rest; HFW = HeadFix® mouthpiece; SYR = 10 mL/cc syringe barrel; EPID = electronic portal imaging device; CBCT = cone beam computed tomography; portal imaging; eNAL = extended no action level protocol (imaging on first three fractions, followed with once weekly imaging); NAL = no action level protocol (imaging on first three fractions); VHMF = Van Herk's margin formula.

Norfadilah et al.[14] study assessed the impact of variation in tongue immobilisation on the margin size, and the results were indicative that immobilisation devices influence the PTV margin size.

According to Anjanappa et al.[8] study, the lower neck region requires a larger PTV margin in the ML and AP direction. Another similar result was obtained in Cheo et al.[17] study, where the set-up errors were evaluated in different levels of the neck and the largest displacement was found to be in the ML direction (6·52 mm). As compared to the SI direction, the PTV margin findings in the other reviewed studies do not appear to indicate any substantial difference in the margin size of the ML and AP direction. Anjanappa et al.[8] suggestion of increasing the margin size in the ML and AP direction should therefore not be considered for all clinical situations.

Inter-fraction and intra-fraction errors

All the studies in the review analysed inter-fraction errors; therefore, the method of calculating the margin varied solely on whether intra-fraction error was being assessed. Van Herk et al. suggested to include target volume delineation variation and intra-fractional errors in the margin estimation, as well as including the SD of these errors in quadrature[18].

The systematic literature found that inter-fraction errors were generally higher than intra-fraction errors, which indicates that maintaining the position during treatment leads to less errors than reproducing the set-up between treatments.
therefore intra-fraction movements should be quantified[6]. Bruijnen et al.[6] and Yin et al.[7] were the only studies that analysed both intra-fraction errors and inter-fraction errors. Intra-fraction error is related to internal organ motion and patient movement during treatment, therefore it is a random error[19].

In Bruijnen et al.[6] study, when the tumour motion was incorporated into the PTV margin formula, the margin expanded mostly in the laryngeal region. This indicates that this region is subjected to greater tumour motion due to swallowing[6]. These results were similar to a study by Gurney-Champion et al.[20] where the tumour motion analysed with MRI imaging was found to be significantly larger in the larynx and hypopharynx when compared to oropharynx.

In Yin et al.[7] study, intra-fraction errors were assessed via CBCT images. The study focused on patients receiving treatment to the nasopharynx and intra-fraction errors were based on patient movement during treatment.

Few studies obtained intra-fraction results on patients treated with VMAT and the need to calculate this error was raised by Yin et al.[7]. There was no significant correlation between treatment delivery time and intra-fraction errors. This study had some limitations since there was a limited data for analysis and statistically significant results could not be obtained due to the narrow range of treatment time. These results are contradictive to Hoogeman et al.[21], who stated that with an increment in time, intra-fraction systematic errors increase.

Van Herk's formula does not consider rotational errors[18]. Four studies from the review still investigated the rotational set-up errors on their population sample. In high-precision treatments, it is important that rotational errors are not ignored, especially when the distance from the isocentre to the target is large or when the tumour has a non-spherical shape[22]. In most clinical

Table 2. PTV margin methods

Study	Target delineation	Intra-fraction error	Inter-fraction errors	PTV formula
Oh et al. (2014)	x	x	✓	$PTV = 2{\cdot}5\Sigma + 0{\cdot}7\sigma$
Bruijnen et al. (2018)	x	✓	✓	$PTV = 2{\cdot}5\sqrt{(\Sigma motion^2 + \Sigma setup^2)} + 0{\cdot}7\sqrt{(\sigma motion^2 + \sigma setup^2)}$
Yin et al. (2013)	x	✓	✓	$PTV = 2{\cdot}5\sqrt{(\Sigma inter\text{-}fraction^2 + \Sigma intra\text{-}fraction^2)} + 0{\cdot}7\sqrt{(\sigma inter\text{-}fraction^2 + \sigma intra\text{-}fraction^2)}$
Norfadilah et al. (2017)	x	x	✓	$PTV = 2{\cdot}5\Sigma + 0{\cdot}7\sigma$
Deb et al. (2019)	x	x	✓	$PTV = 2{\cdot}5\Sigma + 0{\cdot}7\sigma$
Anjanappa et al. (2017)	x	x	✓	$PTV = 2{\cdot}5\Sigma + 0{\cdot}7\sigma$
Kukolowicz et al. (2020)	x	x	✓	$PTV = 2{\cdot}5\Sigma + 0{\cdot}7\sigma$

departments, these errors are not corrected due to couch limitations[23]. Rotational errors were minimal in the four studies.

Target volume delineation

Calculation of the PTV margin should incorporate the systematic errors obtained from target volume delineation[7,8]. Even though target volume delineation might have the largest impact in margin size since it is fixed throughout the treatment[24], none of the studies in the review evaluated this factor. This type of systematic error is commonly ignored in studies that investigate PTV margin calculation[24] leading to a potential underdosage to the tumour. This is important for all head and neck locations, but more so for the oropharyngeal cancer since interobserver variability in target delineation is greater in this region[25].

Limitation

The researchers ran a thorough search and all the body of evidence found was analysed and discussed; however, the availability of data on PTV margin calculation to the head and neck region was limited and this resulted in a small sample of studies. This could have limited the findings of significant relationships.

Also, some studies were not reliable since they obtained a weak quality evaluation after been analysed by the Joanna Briggs Institute tool, as such, the researchers were careful in drawing conclusions from these publications.

Since the review relied on pre-existing data, the obtained results were dependent on the methodology of the studies. Self-reported data bias could be introduced from relying on pre-existing data. The studies were also limited to the English language; therefore, 46 non-English language studies were excluded in the first phase of the review, this resulted in the review to be susceptible to reporting bias and language bias[26].

Comparison of study results was limited due to the studies being heterogenous in terms of the key characteristics of the studies and methodology design.

Although the findings of this study should be interpreted with caution, this review represents a comprehensive examination of studies that analysed PTV margin to the head and neck region with VMAT.

Conclusion

The Van Herk formula was used in all the studies in the review and none of the studies made use of other PTV margin formula. This margin seems adequate to calculate the PTV margin for head and neck patients treated with VMAT.

Most studies only included inter-fraction errors into the van Herk formula. However, PTV margin should incorporate target volume delineation, intra-fraction errors and inter-fraction set-up errors to ensure a well-defined margin.

The result indicated that tumours in the laryngeal region where more susceptible to motion when compared to those found in the nasopharynx and oropharynx.

Inter-fraction translational errors were assessed in all the studies from set-up errors that were registered by the imaging software. The SD of population random errors was found to be a little bit higher than that of population systematic errors. The systematic and random errors of set-up rotational errors were not considered in most studies, and the obtained values were not included in the PTV formula since this formula assumes that the CTV is spherical and, therefore, is unaffected by rotation.

The findings of the review where in line with other studies that stated that different anatomical regions, immobilisation devices, imaging frequency, treatment modality, set-up procedures and patient collaboration influence the size of the PTV margin.

Supplementary Material. To view supplementary material for this article, please visit https://doi.org/10.1017/S1460396921000546.

Acknowledgements. None.

Financial Support. None.

Conflict of Interest. The authors declare none.

References

1. Jones D. ICRU Report 50—Prescribing, Recording and Reporting Photon Beam Therapy. [Online] American Association of Physicists in Medicine; 1994, 833–834. Available from: doi: 10.1118/1.597396
2. Gregoire V, Mackie T R. Dose prescription, reporting and recording in intensity-modulated radiation therapy: a digest of the ICRU Report 83. London: Future Medicine Ltd; 2011;3(3): 367–373. Available from: doi: 10.2217/iim.11.22
3. Gill S K, Reddy K, Campbell N, Chen C, Pearson D. Determination of optimal PTV margin for patients receiving CBCT-guided prostate IMRT:

comparative analysis based on CBCT dose calculation with four different margins. United States: Wiley; 2015;16(6): 252–262. Available from: doi: 10.1120/jacmp.v16i6.5691
4. Eldawlatly A, Alshehri H, Alqahtani A, Ahmad A, Al-Dammas F, Marzouk A. Appearance of Population, Intervention, Comparison, and Outcome as research question in the title of articles of three different anesthesia journals: A pilot study. Saudi J Anaesth 2018; 12: 283–286.
5. Snilstveit B, Oliver S, Vojtkova M. Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. Informa UK Limited; 2012;4(3): 409–429. Available from: doi: 10.1080/19439342.2012.710641
6. Bruijnen T, Stemkens B, Terhaard C H, Lagendijk J J, Raaijmakers C P, Tijssen R H. Intrafraction motion quantification and planning target volume margin determination of head-and-neck tumors using cine magnetic resonance imaging. Ireland: Elsevier BV; 2019;130:82–88. Available from: doi: 10.1016/j.radonc.2018.09.015
7. Yin W-J, Sun Y, Chi F, et al. Evaluation of inter-fraction and intra-fraction errors during volumetric modulated arc therapy in nasopharyngeal carcinoma patients. England: BioMed Central Ltd; 2013;8(1): 78. Available from: doi: 10.1186/1748-717X-8-78
8. Anjanappa M, Rafi M, Bhasi S, et al. Setup uncertainties and PTV margins at different anatomical levels in intensity modulated radiotherapy for nasopharyngeal cancer. Netherlands: Elsevier Urban & Partner Sp. z o.o; 2017;22(5): 396–401. Available from: doi: 10.1016/j.rpor.2017.07.005
9. Stroom J, Gilhuijs K, Vieira S, et al. Combined recipe for clinical target volume and planning target volume margins. Int J Radiat Oncol Biol Phys 2014; 88 (3): 708–714. doi: 10.1016/j.ijrobp.2013.08.028
10. Martins L, Couto J G, Barbosa B. Use of planar kV vs. CBCT in evaluation of setup errors in oesophagus carcinoma radiotherapy. Rep Pract Oncol Radiother 2016; 21 (1), 57–62. doi: 10.1016/j.rpor.2015.10.005
11. Oh Y, Baek J, Kim O Kim J. Assessment of setup uncertainties for various tumor sites when using daily CBCT for more than 2200 VMAT treatments. J Appl Clin Med Phy 2014; 15 (2): 85–99. doi: 10.1120/jacmp.v15i2.4418
12. Kukolowicz P, Mietelska M, Kiprian D. Effectiveness of the No action level protocol for head & neck patients – Time considerations. Elsevier B.V; 2020; Available from: doi: 10.1016/j.rpor.2020.04.005
13. Deb J, Chaudhuri S, Panda D, et al. Retrospective analysis of random and systematic errors in radiation therapy of head and neck cancer patients and its clinical predictive implications with VMAT treatment. 2019;7(7): 2758. Available from: doi: 10.18203/2320-6012.ijrms20192914
14. Norfadilah M N, Ahmad R, Heng S P, Lam K S, Radzi A B A, John L S H. Immobilisation precision in VMAT for oral cancer patients. 2017;851:12025. Available from: doi: 10.1088/1742-6596/851/1/012025
15. Namysł-Kaletka A, Tukiendorf A, Wydmański J. Calculation of planning target volume margins using the van Herk, Stroom and ICRU methods in patients with gastric cancer. Oncol Radiother 2015; 3 (33).
16. van Herk M, Remeijer P, Rasch C, Lebesque J V. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. Int J Radiat Oncol Biol Phys 2000; 47 (4): 1121–1135. doi: 10.1016/s0360-3016(00)00518-6
17. Cheo T, Loh Y, Chen D, Lee KM, Tham I. Measuring radiotherapy setup errors at multiple neck levels in nasopharyngeal cancer (NPC): A case for differential PTV expansion. Ireland: Elsevier Ireland Ltd; 2015;117(3): 419–424. Available from: doi: 10.1016/j.radonc.2015.09.032
18. Van Herk M. Errors and Margins in Radiotherapy. United States: Elsevier inc;2004;14(1): 52–64. Available from doi: 10.1053/j.semradonc.2003.10.003
19. Shikha G, Kataria T. Image guidance in radiation therapy: techniques and applications. Radiol Res Pract 2014; 2014:705604–10. doi: 10.1155/2014/705604
20. Gurney-Champion O J, McQuaid D, Dunlop A, et al. MRI-based assessment of 3D intrafractional motion of head and neck cancer for Radiation Therapy. Int J Radiat Oncol Biol Phys 2018; 100 (2): 306–316. doi: 10.1016/j.ijrobp.2017.10.016
21. Hoogeman M S, Nuyttens J J, Levendag P C, Heijmen BJM Time dependence of intrafraction patient motion assessed by repeat stereoscopic imaging. Int J Radiat Oncol Biol Phys 2008; 70 (2): 609–618. doi: 10.1016/j.ijrobp.2007.08.066
22. Chang, J. A statistical model for analyzing the rotational error of single isocenter for multiple targets technique. Medical Physics (Lancaster) 2017; 44 (6): 2115–2123. doi: 10.1002/mp.12262
23. Zhang Q, Song Y,Chan M, Burman C, Yamada Y. Feasibility study of real-time planning for stereotactic radiosurgery. United States: American Association of Physicists in Medicine; 2013;40(3):031711-n/a. Available from: doi: 10. 1118/1.4792637
24. Vinod S K, Jameson M G, Min M. Holloway L C. Uncertainties in volume delineation in radiation oncology: a systematic review and recommendations for future studies. Radiother Oncol 2016; 121 (2): 169–179. doi: 10.1016/j.radonc.2016.09.009
25. Segedin B, Petric P. Uncertainties in target volume delineation in radiotherapy – are they relevant and what can we do about them? Radiol Oncol 2016a; 50 (3): 254–262. doi: 10.1515/raon-2016-0023
26. Kirkham J J, Dwan K M, Altman D G, Gamble C, Dodd S, Smyth R, Williamson P R The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. Bmj 2010; 340 (7747): 637–640. doi: 10.1136/bmj.c365

Appendix B

FREC Approval

Dear Kristina,

your recently submitted amendments have been reviewed and approval is granted oBo FREC. You may proceed with data collection.

Good luck.

Regards

[redacted]

L-Università ta' Malta

[redacted]
BSc(Hons) (Melit.), MSc(Melit.), RM

Senior Lecturer, Department of Midwifery
Chairperson, Faculty Research Ethics Committee

Faculty of Health Sciences
Office No. 48
[redacted]
[redacted]

Data Protection Clearance Decleration Form

SPTAR MATER DEI

Data Protection Clearance Declaration Form

REF: 68/2021

I hereby declare that I will respect the confidentiality and privacy of any personal data or information that I will come across at Mater Dei and will in no circumstance disclose any such information to third parties.

I confirm that information submitted for Data Protection Clearance is correct and that I will abide with conditions issued in same clearance notice.

- **This clearance does not cover ethical approval.**
- **All your potential participants must be 18 +.**
- **All documents presented to your participants must include UOM's logo.**
- **Your submitted documentation must remain unchanged.**
- **What was declared during this clearance process is what you will abide to.**
- **You and your intermediary must abide with all the articles of the GDPR (EU) 2016 / 679 throughout the data collection process and thereafter.**
- **You are requested to submit a copy of your findings to this office at the end of your study.**
- **Please communicate with the Chair's office before you start. You must also present this clearance email**

I also declare that I am aware of the provisions of the:

General Data Protection Regulation (2016)
(ref: https://idpc.org.mt/en/Pages/gdpr.aspx),
Computer misuse provisions of the Criminal Code
(ref: http://www.justiceservices.gov.mt/DownloadDocument.aspx?app=lom&itemid=8574),
and, the Professional Secrecy Act
(ref: http://www.justiceservices.gov.mt/DownloadDocument.aspx?app=lom&itemid=8844&l=1)

and that I will abide by all Government and Hospital regulations related to data, information and use of IT Systems and services (ref: http://ictpolicies.gov.mt , http://www.kura.gov.mt).

SPTAR MATER DEI

Data Protection Clearance Declaration Form

REF: 68/2021

Full Name: Kristina Caruana

ID/ Passport: [redacted]

Approval Date from DPO: 24th March 2021

Approval Date from CEO: 01st February 2021

Data Collection Period (From – To): May 2021 – May 2022

MDH Official Approval Names: [redacted]

Name of Study / Audit: A local assessment of the PTV margin calculation in patients treated to the larynx with VMAT

Applicant's Signature: [redacted] Kristina Caruana (Mar 30, 2021 20:22 GMT+2)

CEO Permission

CEO at Health-MDH
Today, 12:52
Caruana Kristina [redacted]

Dear Ms Caruana,

Kindly note that approval has been given by Ms [redacted] for you to conduct this study in line with applicable hospital protocols.

Please also be reminded that CEO's approval has to be sought before any data being shared outside of hospital locally or abroad.

Regards

[redacted]
Personal Assistant To CEO

SPTAR MATER DEI

[redacted]

Sir Anthony Mamo Oncology Centre Research Approval Form

SIR ANTHONY MAMO ONCOLOGY CENTRE

FORM : **Oncology Proposal/Approval Audit/ Research purposes**

Document Code: ONCO-GeFO-P/A-001. Ver.01 Reference SOP : ONCO-Ge-PD.AP--001.Ver.01

Clinical Chairperson (Haematology - Oncology):
Name and Surname (in block letters) and Signature:

[redacted]

Human Resources and Administration Manager:
Name and Surname (in block letters) and Signature:

[redacted]

An approval is granted to carry out the study/audit at any SAMOC Department. Patient information can be accessible only by complying with the following data protection principles, which are set out in the General Data Protection Regulation 2016. In summary these state that patient's data shall:

- ***Be obtained and processed fairly and lawfully and shall not be processed unless certain conditions are met. Therefore patient's information (including scans) should be made anonymous -by an appointed radiotherapy staff (from the Head of section)***
- ***Be obtained for a specified and lawful purpose and shall not be processed in any manner incompatible with that purpose.***
- ***Be adequate, relevant and not excessive for those purposes (in the case of a study or audit).***
- ***Be accurate and kept up to date.***
- ***Not be kept longer than is necessary for that purpose***
- ***Be processed in accordance with the data subject's rights.***
- ***Be kept safe from unauthorised access, accidental loss or destruction.***
- ***Not be transferred to any third party unlawfully.***

FORM :	**Oncology Proposal/Approval Audit/ Research purposes**
Document Code: ONCO-GeFO-P/A-001. Ver.01	Reference SOP : ONCO-Ge-PD.AP--001.Ver.01

Clinical Consultant Oncologist/s:
Name and Surname (in block letters) and Signature

[redacted]

Heads of:
(Name, Surname and Section (in block letters) and Signature)

Radiotherapy Department:

Radiography

[redacted]
Professional Lead Allied Health
Radiotherapy Department
[redacted]

[redacted]

Medical Physics

[redacted]

[redacted]
Medical Physicist Area Co-Ordinator
Medical Physics Radiotherapy Department
[redacted]
25/3/2021.

Patient Consent Form (In English)

L-Università ta' Malta

Information Letter

Dear Patient,

My name is Kristina Caruana and I am currently reading for a MSc by Research at the University of Malta. As part of my course requirements, I am conducting a research study entitled, ***'A local assessment of the PTV margin calculation in patients treated to the larynx with VMAT'***. The aim of this study is to determine the radiotherapy treatment margin size for patients being treated for larynx cancer with VMAT. Your participation in this study could help improve the service given to future patients and to update the current procedures in radiotherapy.

I hereby ask for your permission to access data from your CT planning scans and Cone Beam Computed Tomography (CBCT) scans. The CBCT scan refers to the scan that is taken prior to treatment delivery to ensure that you are in the correct position. Both CT planning scans and daily CBCT scans are performed as a standard clinical practice. As part of this study, I will also be seeking your approval to perform an additional scan once a week for the whole duration of your treatment. This procedure is only done on patients who consent to participate in this research and is an additional procedure to the standard practice at SAMOC. The benefit of this additional scan is that the radiographers would assess if there was any variation in your position from the beginning till the end of your treatment session on that particular day. If movement during treatment is detected, the radiographers may make the necessary adjustments to try and reduce this for the next treatment day. This procedure is also essential to calculate the treatment margin size which would benefit future patients. It is important to note the CBCT scan involves the use of X-rays, this radiation dose, however, will be controlled to limit the risk that results from ionising radiation. The treatment procedure will also increase by about two minutes during the days of the weekly additional CBCT images.

If you agree to participate, your approval is being sought to:

1. access your examination report,

2. access your radiotherapy information and clinical note,

3. access your imaging data.

Your consent is also being sought so that once a week an additional CBCT examination will be performed. You may keep a copy of this information sheet and consent form for your perusal. Should you wish to participate in this study, kindly sign the provided consent form and hand it back to the radiographer who gave it to you.

The signed consent form will be stored securely in a locked cupboard, by the intermediary person to ensure anonymity. Please note that you are not obliged to participate in this study, and you may withdraw from the study at any time without giving a reason. Furthermore, withdrawal from this study or refusing to participate will not have any negative repercussions

since there will be no penalty or loss of benefits to which you are entitled to. The participation in this research is completely on voluntary basis and you are free to accept or refuse to take part without giving a reason.

I would like to assure you that all data collected will be strictly anonymous. Confidentiality will be maintained throughout the entire study and your identity and personal information will not be revealed in any publications, reports or presentations arising from this research. All data collected from this study shall be used solely for the purpose of this study. The researcher, the academic supervisor/s and the examiners will have access to anonymised data only. Also, the gathered data will be erased upon completion of the study whilst the results will be kept in an anonymised format, stored on a password protected computer being only accessible to the researcher. As a participant, you have the right under the General Data Protection Regulation (GDPR) and national legislation to access, rectify, and where applicable ask for the data concerning you to be erased (or retained in anonymised form). Please note that psychological support is provided should this be required.

This study has been approved by the Research Ethics Committee of the Faculty of Health Sciences at the University of Malta.

Thank you for your time and consideration. Should you have any questions or concerns do not hesitate to contact me by e-mail on [redacted] or phone me on [redacted]. You can also contact [redacted] my supervisor.

Yours sincerely,

L-Università ta' Malta

Patients Consent Form

Title of Research Study: A local assessment of the PTV margin calculation in patients treated to the larynx with VMAT

Please Tick X

I, _____give my consent to take part in the study conducted by Kristina Caruana. The purpose of this document is to specify the terms of my participation in this research study. ☐

1. I have been given written and verbal information about the purpose of the study and all questions have been answered. ☐

2. I understand that I have been invited to take part in a study in which the researcher will investigate the PTV margin size for patients being treated with VMAT to the larynx. ☐

3. I am aware that I once a week I will have an additional CBCT scan after the treatment to assess for any variation in my position during treatment. ☐

4. I am aware that the researcher will have access to anonymised data only and therefore the researcher cannot identify the patients. ☐

5. I am aware that my identity and personal Information will not be revealed in any publications, reports or presentations arising from this research. ☐

6. I also understand that by allowing the researcher to access my report and scans will be helping to improve the service given to future patients by aiding to update current procedures in the radiotherapy department. ☐

7. I voluntarily agree to participate in this study and have been given a copy of consent form. ☐

8. Under the General Data Protection Regulation (GDPR) and national legislation I have the right to access, rectify, and where applicable erase the data concerning about myself. ☐

9. I understand that I am free to accept, refuse or stop participation at any time without giving any reason and that my confidentiality will be maintained throughout the study. Also, refusing to participate or withdrawing from the study would involve no penalty or loss of benefits to which I am otherwise entitled. ☐

Patient Consent Form (In Maltese)

L-Università ta' Malta

Ittra ta' Informazzjoni lil-Pazjenti

Għażiż/a Pazjent/a,

Jiena Kristina Caruana, fil-preżent qed insegwi l-kors ta' MSc fir-Riċerka mal-Università' ta' Malta. Bħala parti mir-rekwiżiti tal-kors, qed nagħmel riċerka bit-titlu "A local assessment of the PTV margin calculation in patients treated to the larynx with VMAT". L-għan ta' dan l-istudju hu li ninvestigaw id-daqs tal-marġni għal pazzjenti li jiħdu l-kura bir-raġġi fil-laringi b' VMAT. Is-sehem tiegħek f'dan l-istudju se jkun qiegħed jgħin kemm fit-titjib tas-servizz mogħti lil pazjenti fil-ġejieni kif ukoll sabiex jiġu aġġornati l-proċeduri kurrenti fid-dipartiment tar-raġġi.

Inti qed tiġi mistieden/mistiedna biex tieħu sehem f'dan l-istudju għalhekk qiegħda nitlob il-permess tiegħek biex naċċessa informazzjon:

1. mir-rapporti klinci tal-onkologija,
2. mir-ritratti tas-CT scan tiegħek,
3. mir-ritratti li jittieħdu qabel ir-raġġi.

Nixtieq wkoll nitlob il-permess biex jittieħdu ritratti oħra darba fil-ġimgħa wara li tingħata ir-raġġi. Dawn ir-ritratti huma meħtieġa biex naraw jekk hemmx varjazzjoni filpożizzjoni tiegħek waqt it-trattament. Jekk ċaqlieq fil-pożizzjoni tiegħek tigi osservata, ir-radjografi ikunu jistgħu jiħdu l-passi neċċesarji biex inaqsu dan it-tip ta' movument fil-ġranet ta' trattamenti l-oħra. Din il-proċedura hi wkoll essenzjali biex jiġi mkejjel ilmarġni kif suppost. Importanti li tkun taf li 'CBCT scan' jinvolvi lużu tar-raġġi (X-rays). Dawn ir-raġġi se jkunu kontrollati biex innaqsu ir-riskju li joħroġ minnhom. Il-proċedura ta' t-trattament se żżidlek wkoll b' xi żewġ minuti fil-ġranet li jkollok dawn it-tip ta' ritratti.

Jekk taċċetta li tipparteċipa, qed tiġi mitlub/a biex tipprovdi l-kunsens tiegħek. Kopja tal-ittra ta' informazzjoni u tal-formula ta' kunsens se tingħata lilek għal referenza filfutur. Jekk taċċetta li tipparteċipa f'dan lis-studju, qed tiġi mitlub/a biex tiffirma l-kopja tal-formula ta' kunsens u taddiha lir- radjografu li għaddielek din l-informazzjoni. Ilformoli ta' kunsens iffirmati ser jinżammu f'post sigur mill-intermedjarji f'armarju msakkar sabiex tiġi żgurata iktar l-anonimità. Jekk jogħġbok innota li se tkun qed tintalab tipprovdi biss data li hija meħtieġa għal-riċerka. M'intix obbligat/a li tieħu sehem u tista' tirtira mill istudju x'ħin trid mingħajr ma tagħti ebda raġuni. Barra minn hekk li tirtira mill-studju ma jħalli ebda riperkussjoni negattiva fuqek. Għalhekk il-parteċipazzjoni tiegħek f'dan l-istudju hija volontarja u inti ħieles/ħielsa li taċċetta jew tirrifjuta li tieħu sehem mingħajr ma tagħti raġuni.

L-Università ta' Malta

Nassigurak li kull data misura se tkun qed tinħażen b'mod anonimu. Kunfidenzjalita' se tinżamm matul l-istudju kollu u l-identita' tiegħek u kull informazzjoni personali miġbura m'huma se jiġu żvelata mkien fit-teżi, fir-rapporti, il-preżentazzjonijiet u/jew il-pubblikazzjonijiet li jistgħu jirriżultaw minnha. Barra minn hekk kull informazzjoni miġbura minn din ir- riċerka għandha tintuża biss għall-għan ta' dan l-istudju. Ir-ritratti u informazzjoni fuq ir-ritratti se jingħataw lili b'mod anonimu. Jiena flimkien mas-superviżur akkademiċi u l-eżaminaturi se jkollna aċċess biss għad-data anonima. Kull informazzjoni miġbura ser titħassar wara t-tmiem ta dan l-istudju, u r-riżultati ser jibqgħu miżmuma b'mod anonimu fuq kompjuter protett bil password, aċċessibli biss għar-riċerkatriċi. Skond il- "General Data Protection Regulation" (GDPR), bħala participant/a f'dan l-istudju, int għandek id-dritt li taċċesa, tirranġa, u fejn applikabli ssaqsi għal informazzjoni li tikkonċerna fuqhek biex titħassar. Int tista' wkoll twaqqaf il-parteċipazzjoni tiegħek f' dan l-istudju mingħajr ma jkun hemm xi tip ta riperkussjonijiet rigward it-trattament tiegħek. Nixtieq ninfurmak li hemm provdut servizz ta' support psikoloġiku, f' kaz ta' bżonn.

Dan l-istudju ġie approvat mill-kumitat għall- Etika fir-Riċerka fi ħdan il-Fakuta tax-Xjenzi tas-Saħħa, fl-Universita' ta' Malta.

Grazzi ħafna tal-ħin tiegħek f'dan l-istudju. Jekk ikollok xi mistoqsijiet rigward dan l-istudju tiddejjaqx tikkuntatjani fuq l-e-mail: [redacted] jew ċempilli fuq [redacted].

L-Università ta' Malta

Formula ta' Kunsens tal-Pazjenti

Titlu tal-istudju: A local assessment of the PTV margin calculation in patients treated to the larynx with VMAT.

Jien, _____nagħti l-kunsens tiegħi biex nieħu sehem fl-istudju mmexxi minn Kristina Caruana. L-għan ta' dan id-dokument hu li jiġu speċifikati ttermini tal-parteċipazzjoni

Jekk Jgħoġbok Immarka **X**

1. Jien ingħatajt informazzjoni miktuba u verbali dwar l-għan tal-istudju u l-mistoqsijiet kollha twieġbu. ☐

2 Nifhem li se nkun qed nipparteċipa fi studju, fej ir-riċerkatriċi ħa tinvestiga il- qisien tal-marġini tar-raġġi f'persuni li qed jingħataw trattament fil-laringi b' VMAT. ☐

3. Nifhem li darba fil-ġimgħa se jittieħdu ritratti b' CBCT wara t-trattament biex ir-riċerkattriċi tara jekk kienx hemm xi varjazzjoni fil-pożizzjoni tiegħi waqt it-trattament. ☐

4. Naf ukoll li r-riċerkatriċi se jkollha aċċess biss għall informazzjoni anonima sabiex ma tkunx tista t'identifika l-pazjenti. ☐

5. Naf li l-identita' tiegħi u l-informazzjoni personali m'huma se jinkixfu mkien fit-teżi, fir-rapporti, preżentazzjonijiet u/jew fil-pubblikazzjonijet li jistgħu jirriżultaw minnha. ☐

6. Nifhem wkoll li billi ngħati permess ir-riċerkatriċi tkun tista' taċċessa rapporti u imaġini radjografiċi meħuda għal għan ta' trattament. B'hekk inkun qed ngħin li ntejjeb is-servizz tal-pazjenti fil-ġejjieni sabiex jaġġornaw l-proċeduri kurrenti tar-raġġi. ☐

8. Jien volontarjament aċċettajt li nipparteċipa f'dan l-istudju, u ingħatajt kopja tal-formola tal- kunsens. ☐

9. Skond il- General Data Protection Regulation bħala participant/a f'dan l-istudju, jien għandi id-dritt li naċċesa, nirranga, u fejn applikabli nsaqsi għal informazzjoni li tikkonċerna fuqi biex titħassar. ☐

10. Jien għandi d-dritt li nirtiera minn dan l-istudju mingħajr ebda raguni f-kull ħin. Nagħmel dan mingħajr ebda riperkussjonijiet, penali, jew nuqqas ta' benefiċċi ft-trattament tiegħi. Nifhem li se tinżamm l-kunfidenzjalita' tiegħi f'dan l-istudju. ☐

L-Università ta' Malta

Re: Seeking permission to participate in the study – Clinical Oncologists and Specialists' Trainees

My Name is Kristina Caruana and I am currently registered as a student pursuing a postgraduate course leading to the award of a MSc by Research at the University of Malta. This entails that I undertake a research study which has to be submitted during the final year of the course. The proposed title of my study is *'A local assessment of the PTV margin calculation in patients treated to the larynx with VMAT'*. [redacted], Head of the Department of Radiography, is the primary supervisor. [redacted], Senior Lecturer and Mr [redacted], Assistant Lecturer are the co-supervisors. [redacted], the Clinical Chairperson of the Oncology Department is appointed as an advisor.

The aim of this study is to calculate the PTV margin on patients treated for cancer of the Larynx with VMAT at the local Oncology Centre. As part of this study, I will be assessing inter-fraction and intra-fraction errors, as well as analysing the variation in target volume delineation.

To fulfil the aim of the study I shall be recruiting about twenty patients' prospectively with the aid of an intermediary person who is a radiographer working in the clinical department. The intermediary person will be collecting set-up error data from the XVI software and is required to identify eligible patients. Patients that fit the inclusion criteria will be provided with information about the study and will be invited to participate. if the patients accept, the intermediary radiographer will need to obtain their signed consent. The intermediary person will also provide me with the required anonymised information for evaluation and will be asked to liaise with the medical physicist, assisting me through this study, to anonymise patients CT scan.

On a weekly basis those patients who consent to participate in the study will be having an additional scan after the treatment. These scans would need to be scheduled by the radiographers working in R&V and the data will be collected by the intermediary person in order to measure intra-fraction errors.

To assess the variation in target volume delineation, the CT planning scans of the patients in the sample will be distributed to the clinicians who will be asked to delineate the CTV volume. This data will then need to be exported to IMSIMQA™ to assess the systematic error in this procedure.

I am therefore asking for your consent to participate in this study. Your role, should you accept, will be to delineate the CTV volumes of five randomly selected patients from the data set. All patients in the sample will be diagnosed with a laryngeal tumour. The procedure to delineate the CTV volume for a patient would generally take around fifteen minutes.

Anonymity and confidentiality will be maintained throughout the entire study and your identity and personal information will not be revealed in any publications, reports or presentations arising from this research. All data collected from this study shall be used solely for the purpose of this study. The researcher, the academic supervisor/s and the examiners will have access to anonymised data only. Also, the gathered data will be erased upon completion of the study whilst the results will be kept in an anonymised format, stored on a

password protected computer being only accessible to the researcher. As a participant, you have the right under the General Data Protection Regulation (GDPR) and national legislation to access, rectify, and where applicable ask for the data concerning you to be erased (or retained in anonymised form).

It is my intention to share the results of my study with the radiotherapy department to optimise current PTV margin departmental protocols.

I appreciate the time taken to respond to this request. Should you require further information or clarifications do not hesitate to contact me.

Yours sincerely,

L-Università ta' Malta

Re: Seeking permission to participate in the study – Radiographers

My name is Kristina Caruana and I am currently registered as a student pursuing a postgraduate course leading to the award of a MSc by Research at the University of Malta. This entails that I undertake a research study which has to be submitted during the final year of the course. The proposed title of my study is 'A local assessment of the PTV margin calculation in patients treated to the larynx with VMAT'. [redacted], Head of the Department of Radiography, is the primary supervisor. [redacted], Senior Lecturer and Mr [redacted], Assistant Lecturer are the co-supervisors. [redacted], the Clinical Chairperson of the Oncology Department is appointed as an advisor.

The aim of this study is to calculate the PTV margin on patients treated for cancer of the Larynx with VMAT at the local Oncology Centre. As part of this study, I will be assessing inter-fraction and intra-fraction errors, as well as analysing the variation in target volume delineation.

To fulfil the aim of the study I shall be recruiting about twenty patients' prospectively with the aid of an intermediary person who is a radiographer working in the clinical department. The intermediary person will be collecting set-up error data from the XVI software and is required to identify eligible patients. Patients that fit the inclusion criteria will be provided with information about the study and will be invited to participate. if the patients accept, the intermediary radiographer will need to obtain their signed consent. The intermediary person will also provide me with the required anonymised information for evaluation and will be asked to liaise with the medical physicist, assisting me through this study, to anonymise patients CT scan.

Also, on a weekly basis those patients who consent to participate in the study will be having an additional scan after the treatment. These scans would need to be scheduled by the radiographers working in R&V and the data will be collected by the intermediary person in order to measure intra-fraction errors.

To assess the variation in target volume delineation, CT planning scans of the patients in the sample will be distributed to the clinicians who will be asked to delineate the CTV volume. This data will then need to be exported to IMSImQA™ to assess the systematic error in this procedure.

I am asking for your consent to participate in this study. Your role, should you accept, will be to perform the scan verification procedure to ensure the inter-rater reliability of the set-up errors. You will also be asked to perform image matching on a total of twenty-five randomly selected scans from the study's population sample. All patients in the sample will be diagnosed with a laryngeal tumour. This procedure is estimated to take one hour to be performed.

Anonymity and confidentiality will be maintained throughout the entire study and your identity and personal information will not be revealed in any publications, reports or presentations arising from this research. All data collected from this study shall be used solely for the purpose of this study. The researcher, the academic supervisor/s and the examiners will have access to anonymised data only. Also, the gathered data will be erased upon completion of the study whilst the results will be kept in an anonymised format, stored on a

password protected computer being only accessible to the researcher. As a participant, you have the right under the General Data Protection Regulation (GDPR) and national legislation to access, rectify, and where applicable ask for the data concerning you to be erased (or retained in anonymised form).

I appreciate the time taken to respond to this request. Should you require further information or clarifications do not hesitate to contact me.

Yours sincerely,

Appendix C

Part i

Patient information data for the measurement of target volume delineation error

Patient 16

62-year-old, heavy smoker.

Staging
Stage 1

Diagnosis
Diagnosed with left Vocal Cord SCC, extending to anterior commissure. Severe dysplasia on anterior right vocal cord. Both cords mobile.

No Endoscopy results available

MRI findings
There is focal enhancement of the left vocal cord (confirmed histologically to represent well differentiated squamous cell carcinoma). No evidence of gross pathological enhancement in the supraglottis on this side. No frank mass lesions identified in the right true and false vocal cords. There is no cervical lymphadenopathy by size criteria. Unfortunately, cartilaginous invasion cannot be assessed.

Patient 26

64-year-old presented with a one year history of hoarseness. Heavy smoker. Drinks 2 bottles of alcohol daily.

Diagnosis
Right vocal cord - Moderately differentiated squamous cell carcinoma
Left vocal cord - Moderate to focally severe epithelial dysplasia

Staging

There is effacement of the ventricle on right side
" Initial erosive changes of thyroid cartilage on right.
" No subglottis extension (staging T3)
" No lymphadenopathy by size criteria but non-specific
" Round looking lymph node right level III (9x8mm)

Biopsy results
The right vocal cord biopsies show multiple mucosal fragments which are infiltrated in places by a moderately differentiated squamous cell carcinoma that has arisen on a background of severe epithelial dysplasia with overlying hyperparakeratosis. The tumour is composed of islands of malignant squamous epithelial cells with a pleomorphic nuclei, prominent nucleoli and moderate amounts of eosinophilic cytoplasm. Keratin formation is identified but is not prominent.

The biopsies from left vocal cord are composed of mildly inflamed fibrous corium covered by non-keratinised stratified squamous epithelium in which cytologic and architectural atypia amounting to moderate to focally severe dysplasia is seen. Invasive squamous cell carcinoma is not seen in these biopsies.

CT Neck findings
CT Neck: There is pathological enhancement causing bulging of the right true vocal cord. This enhancement involves the anterior two thirds of the right vocal cord extending to the anterior commissure. Dubious involvement of the anterior third of the left vocal cord. Enhancement extends cranially to efface the ventricles on the right side.
Initial erosive change of the inner cortex of the thyroid cartilage. No subglottic extension.

Patient 36

81-year-old, non smoker. Few months history of dysphonia.

Staging
Large T3 glottic tumour arising from right vocal cord. Tumour debulked. Subglottis spared.

CT Neck Thorax findings
There is a small collection in the retropharyngeal space on the right measuring 6 x 10 mm. Subluxation of the right cricoarythenoid joint with the arytenoid cartilage mildly displaced anteriolaterally. Oedema of both aryepiglottic folds and of the paraglottic fat limiting accurate assessment. The thyroid gland is unremarkable.

Patient 46

46-year-old
(D and I mentioned treatment to the vocal cords) - No other clinical notes available.

MR findings
The larynx is unremarkable with no masses of areas of frank pathological enhancement identified. The nasopharynx, oropharynx, parapharyngeal and retropharyngeal spaces are within normal limits (save for a very small Tornwaidt cyst). Incidental note is made of a prominent accessory lobe of the right parotid gland. Otherwise, unremarkable appearance of the parotid and submandibular glands and thyroid gland. There are bilateral slightly prominent cervical lymph nodes, none exceeding 1 cm in short axis.
No significant abnormalities
cervical nodes not exceeding 1 cm.

Endoscopy results showed Erythemateous Rt cord.
No abnormality detected in the neck.

Anterior commissure mucosal biopsy: cancer in situ
right vocal cord mucosal biopsy : ca in situ, highly suspicious for invasion.

Patient 56

81 year old presented with hoarseness

Diagnosis
Differentiated SCC.

Emergency tracheostomy procedure performed (larynx dysfunctional). Was not fit for laryngectomy

MRI findings
Tracheostomy in situ with a small amount of surgical emphysema in the left lower neck.
Nasogastric tube is seen to close the pharynx.
Note is again made of a polypoid largely intraluminal mass arising from the glottic plane. This measures 2.0 (AP) x 1.5 (LL) x 2.0 (CC) cm. At the level of the subglottis the lesion almost completely occupies the airway.
The most cranial margin of the mass is the ventricle on the left side, whereas the inferior margin of the mass is just above the lower margin of the cricoid cartilage.
There is no convincing evidence of infiltration of the paraglottic space. There is no infiltration of the laryngeal cartilages.
No lymphadenopathy in the neck.

Conclusion: Disease recurrence at the level of the glottis and subglottis. Pathological tissue is located largely intraluminally, with no evidence of infiltration of the paraglottic space.

Inter-observer reliability of soft-tissue registration

	Scan	X	Y	Z	Rx	Ry	RZ
Patient 1	**1**						
	2						
	3						
	4						
	5						
Patient 2	**1**						
	2						
	3						
	4						
	5						
Patient 3	**1**						
	2						
	3						
	4						
	5						
Patient 4	**1**						
	2						
	3						
	4						
	5						
Patient 5	**1**						
	2						
	3						
	4						
	5						
Comments:							

Intra-observer reliability of soft-tissue registration

Scan	X	Y	Z	Rx	Ry	RZ
1						
2						
3						
4						
5						

Data Record Sheet for Intra-fraction errors assessment using a soft tissue match

Patient study number: _____ **Age:** _____ **Sex:** _____

Diagnosis: _____

Staging: _____

Prescription: _____

Fraction	X (mm)	Y (mm)	Z (mm)	Rx (^{o})	Ry (^{o})	Rz (^{o})
PreCBCTsoft Week 1						
PostCBCTsoft Week 1						
PreCBCTsoft Week 2						
PostCBCTsoft Week 2						
PreCBCTsoft Week 3						
PostCBCTsoft Week 3						
PreCBCTsoft Week 4						
PostCBCTsoft Week 4						
PreCBCTsoft Week 5						
PostCBCTsoft Week 5						
PreCBCTsoft Week 6						
PostCBCTspft Week 6						

Comments:

Data Record Sheet for Inter-fraction errors (Set-up errors and Organ Motion)

Patient study number _____												
	Clip-box registration						**Mask registration**					
Fraction	**X (mm)**	**Y (mm)**	**Z (mm)**	**Rx (º)**	**Ry (º)**	**Rz (º)**	**X (mm)**	**Y (mm)**	**Z (mm)**	**Rx (º)**	**Ry (º)**	**Rz (º)**
1												
2												
4												
5												
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												
23												
24												
25												
26												
27												
28												
29												
30												

Data record sheet for inter-observer variation in image matching for bone registration

Radiographer study number:							
	Scan	**X**	**Y**	**Z**	**RX**	**RY**	**RZ**
Patient 1	1						
	2						
	3						
	4						
	5						
Patient 2	1						
	2						
	3						
	4						
	5						
Patient 3	1						
	2						
	3						
	4						
	5						
Patient 4	1						
	2						
	3						
	4						
	5						
Patient 5	1						
	2						
	3						
	4						
	5						
Comments:							

Part ii

Data record sheet for target volume delineation error

Patient study number _____						
CT slices	**Anterior (mm)**	**Posterior (mm)**	**Superior (mm)**	**Inferior (mm)**	**Left (mm)**	**Right (mm)**
Individual Mean						

Data Record Sheet for Intra-fraction Errors

	$X1_s - X0_s$ (mm)	**$Y1_s - Y0_s$ (mm)**	**$Z1_s - Z0_s$ (mm)**	**$RX1_s - RX0_s$ (º)**	**$RY1_s - RY0_s$ (º)**	**$RZ1_s - RZ0_s$ (º)**
μ						
SD						
∑						
σ						

μ = Average of all errors
SD = Standard Deviation
∑ = Population systematic intra-fraction error
σ = Population random intra-fraction error
$X0_s$= Pre-treatment CBCT soft tissue image registration result in the left-right direction
$Y0_s$= Pre-treatment CBCT soft tissue image registration result in the superior-inferior direction
$Z0_s$= Pre-treatment CBCT soft tissue image registration result in the anterior-posterior direction
$X1_s$ = Post-treatment CBCT soft tissue image registration result in the left-right direction
$Y1_s$ = Post-treatment CBCT soft tissue image registration result in the superior-inferior direction
$Z1_s$ = Post-treatment CBCT soft tissue image registration result in the anterior-posterior direction
$RX0_s$ = Pre-treatment CBCT soft tissue image registration result in the Roll direction
$RY0_s$ = Pre-treatment CBCT soft tissue image registration result in the Pitch direction
$RZ0_s$ = Pre-treatment CBCT soft tissue image registration result in the Yaw direction
$RX1_s$ = Post-treatment CBCT soft tissue image registration result in the Roll direction
$RY1_s$ = Post-treatment CBCT soft tissue image registration result in the Pitch direction
$RZ1_s$ = Post-treatment CBCT soft tissue image registration result in the Yaw direction

Data record sheet of systematic and random translational errors of inter-fraction errors (set-up errors and organ motion errors)

	$X0_S$		$X0_B$		$X0_S - X0_B$		$Y0_S$		$Y0_B$		$Y0_S - Y0_B$		$Z0_S$		$Z0_B$		$Z0_S - Z0_B$	
Patient:	μ_i	σi	μ_i	σi	μ_i	σi	μ_i	σi	μ_i	σi	μ_i	σi	μ_i	σi	μ_i	σi	μ_i	σi
1																		
2																		
3																		
4																		
5																		
6																		
7																		
8																		
9																		
10																		
11																		
12																		
13																		
15																		
16																		
17																		
18																		
19																		
20																		
Errors	∑	σ	∑	σ	∑	σ	∑	σ	∑	σ	∑	σ	∑	σ	∑	σ	∑	σ

μ_i = Individual average deviation
σi = Individual standard deviation
∑ = Population systematic error
σ = Population random error
$X0_s$= Pre-treatment CBCT soft tissue image registration result in the left-right direction
$Y0_s$= Pre-treatment CBCT soft tissue image registration result in the superior-inferior direction
$Z0_s$= Pre-treatment CBCT soft tissue image registration result in the anterior-posterior direction
$X0_B$ = Pre-treatment CBCT bone image registration result in the left-right direction
$Y0_B$ = Pre-treatment CBCT bone image registration result in the superior-inferior direction
$Z0_B$ = Pre-treatment CBCT bone image registration result in the anterior-posterior direction

Data record sheet of systematic and random rotational errors of inter-fraction errors (set-up errors and organ motion errors)

	$RX0_S$		**$RX0_B$**		**$RX0_S - RX0_B$**		**$RY0_S$**		**$RY0_B$**		**$R0Y_S - RY0_B$**		**$RZ0_S$**		**$RZ0_B$**		**$RZ0_S - RZ0_B$**	
Patient	μ_i	σi	μ_i	σi	μ_i	σi	μ_i	σi	μ_i	σi	μ_i	σi	μ_i	σi	μ_i	σi	μ_i	σi
1																		
2																		
3																		
4																		
5																		
6																		
7																		
8																		
9																		
10																		
11																		
12																		
13																		
15																		
16																		
17																		
18																		
19																		
20																		
Errors	∑	σ	∑	σ	∑	σ	∑	σ	∑	σ	∑	σ	∑	σ	∑	σ	∑	σ

μ_i = Individual average deviation
σi = Individual standard deviation
∑ = Population systematic error
σ = Population random error
$RX0_s$ = Pre-treatment CBCT soft tissue image registration result in the Yaw direction
$RY0_s$ = Pre-treatment CBCT soft tissue image registration result in the Pitch direction
$RZ0_s$ = Pre-treatment CBCT soft tissue image registration result in the Roll direction
$RX0_B$ = Pre-treatment CBCT bone image registration result in the Roll direction
$RY0_B$ = Pre-treatment CBCT bone image registration result in the Pitch direction
$RZ0_B$ = Pre-treatment CBCT bone image registration result in the Yaw direction

Data record sheet for inter-observer variation in image matching

Radiographer study number _____						
	X	**Y**	**Z**	**RX**	**RY**	**RZ**
Patient 1:image number						
1						
2						
3						
4						
5						
Patient 2:image number						
1						
2						
3						
4						
5						
Patient 3: image number						
1						
2						
3						
4						
5						
Patient 4: image number						
1						
2						
3						
4						
5						
Patient 5: image number						
1						
2						
3						
4						
5						
ε_{inter}						
σ_{inter}						

X= Inter-observer variation errors in the left–right direction
Y= Inter-observer variation errors in the superior–inferior direction
Z= Inter-observer variation errors in the anterior–posterior direction
Rx = Inter-observer variation errors in the Roll direction
Ry = Inter-observer variation errors in the Pitch direction
Rz = Inter-observer variation errors in the Yaw direction
ε_{inter} = Population systematic error for inter-observer variation in image matching
σ_{inter} = Population random error for inter-observer variation in image matching

Data record sheet for the calculation of the PTV margin

Population Systematic Error			
	x	**y**	**z**
Target Volume Delineation			
Intra-fraction			
Total inter-fraction			
Inter-observer variation IM			

Population Random Error			
	x	**y**	**z**
Target Volume Delineation			
Intra-fraction			
Total inter-fraction			
Inter-observer variation IM			

Appendix D

Validity of Keywords for the Systematic Literature Review (*1*)

Keywords	Not relevant	Somewhat relevant	Quite relevant	Highly relevant
	1	2	3	4
Nasopharyn*				✓
Oropharyn*				✓
Laryn*				✓
Hypopharyn*				✓
Oral cavity/mouth				✓
Head and Neck				✓
Set-up/setup/set up				✓
Error/errors				✓
VMAT/ Volumetric-Modulated Arc Therapy/Volumetric Modulated Arc Therapy/ RapidArc Therapy				✓
PTV/Planning Target Volume				✓

(1) Add tongue with oral cavity / mouth
(2) Add supra, sub, glottis to larynx

Name and Surname: [redacted]

Signature: [redacted]

Validity of Keywords for the Systematic Literature Review (*2*)

Keywords	Not relevant	Somewhat relevant	Quite relevant	Highly relevant
Score	1	2	3	4
Nasopharyn*				✓
Oropharyn*				✓
Laryn*				✓
Hypopharyn*				✓
Oral cavity/mouth				✓
Head and Neck				✓
Set-up/setup/set up				✓
Error/errors				✓
VMAT/ Volumetric-Modulated Arc Therapy/Volumetric Modulated Arc Therapy/ RapidArc Therapy				✓
PTV/Planning Target Volume				✓

Add other treatment sites related to head and neck cancers eg. sinuses, lymphomas and thyroid.

Name and Surname: [redacted]

Signature: [redacted]

Validity of the Target Volume Delineation Data Gathering Tool (*1*)

	Not relevant	Somehow relevant	Quite relevant	Highly relevant
Score	**1**	**2**	**3**	**4**
CT Simulation				✓
Clinical History				✓
Staging and histological Diagnosis				✓
CT contrast scan (when available)				✓
MRI scan (when available)				✓
PET scan (when available)		✓		
Comments:				

Clinical Examination + Endoscopy ✓

Name and Surname [redacted]

Signature [redacted]

Validity of the Target Volume Delineation Data Gathering Tool (*2*)

	Not relevant	Somehow relevant	Quite relevant	Highly relevant
Score	**1**	**2**	**3**	**4**
CT Simulation				✓
Clinical History				✓
Staging and Diagnosis				✓
CT contrast scan (when available)				✓
MRI scan (when available)		✓		
PET scan (when available)	✓			
Comments:				

Name and Surname [redacted]

Signature [redacted]

Data gathering tool Validity for Assessing Inter-fraction error (Set-up errors and Organ Motion) (*1*)

	Not relevant	Somewhat relevant	Quite relevant	Highly relevant
Score	1	2	3	4
Patient number				4
Clipbox registration				
X				4
Y				4
Z				4
Rx				4
Ry				4
Rz				4
Mask registration				
X				4
Y				4
Z				4
Rx				4
Ry				4
Rz				4
Information with regards to CBCT imaging				
Repeated XVI's				✓
Reasons for repeated XVI's				✓
Additional Comments				

Comments ① Instead of Repeated XVIs to write Number of Repeated XVIs
② Should include note that where an XVI is repeated only the shifts of the last XVI repeat are written down.
③ In data collection form add rows to include also those patients who have more than 20 fractions

Name and Surname: [redacted]

Signature: [redacted]

Data gathering tool Validity for Assessing Inter-fraction error (Set-up errors and Organ Motion) (*2*)

	Not relevant	Somewhat relevant	Quite relevant	Highly relevant
Score	1	2	3	4
Patient number				X
Clipbox registration				
X				X
Y				X
Z				X
Rx				X
Ry				X
Rz				X
Mask registration				
X				X
Y				X
Z				X
Rx				X
Ry				X
Rz				X
Information with regards to CBCT Imaging				
Repeated XVI's				X
Reasons for repeated XVI's				X
Additional Comments				X
Comments: The data that is to be collected is very relevant to the aims of the study				

Name and Surname: [redacted]

Signature: [redacted]

Data Gathering Tool for assessment of Intra-fraction errors (*1*)

	Not relevant	Somewhat relevant	Quite relevant	Highly relevant
Score	1	2	3	4
Patient Information				
Age				X
Sex			X	
Diagnosis				X
Prescription				X
Pre-treatment CBCT soft tissue registration				
X				X
Y				X
Z				X
Rx				X
Ry				X
Rz				X
Post-treatment CBCT soft tissue registration				
X				X
Y				X
Z				X
Rx				X
Ry				X
Rz				X
Comments: Again, very relevant.				

Name and Surname: [redacted]

Signature: [redacted]

Data Gathering Tool for assessment of Intra-fraction errors (*2*)

	Not relevant	Somewhat relevant	Quite relevant	Highly relevant
Score	1	2	3	4
Patient information				
Age				✓
Sex			✓	
Diagnosis				✓
Prescription			✓	
Pre-treatment CBCT soft tissue registration				
X				✓
Y				✓
Z				✓
Rx				✓
Ry				✓
Rz				✓
Post-treatment CBCT soft tissue registration				
X				✓
Y				✓
Z				✓
Rx				✓
Ry				✓
Rz				✓

Comments:

Name and Surname: [redacted]

Signature: [redacted]

Validity of Inter-oberver variation in image matching data gathering tool (*1*)

	Not relevant	Somehow relevant	Quite relevant	Highly relevant
Score	**1**	**2**	**3**	**4**
Radiographer number				
Images				✓
X				✓
Y				✓
Z				✓
RX			✓	[redacted]
RY			✓	[redacted]
RZ			✓	[redacted]
Comments: Rotation is dependend on Automatic registeration. If we can do mannual Registration will be more accuracy. Rest others are highly relevant.				

(only Rotation)

Name and Surname: [redacted]

Signature: [redacted]

Validity of Inter-oberver variation in image matching data gathering tool (*2*)

	Not relevant	Somehow relevant	Quite relevant	Highly relevant
Score	**1**	**2**	**3**	**4**
Radiographer number				✓
Images			✓	
X				✓
Y				✓
Z				✓
RX				✓
RY				✓
RZ				✓
Comments: Even though the rotational translation will give the same error in automatic matching, it's good to compare, whereas its impossible to apply the rotational translation.				

Name and Surname: [redacted]

Signature: [redacted] 16/3/22.

Appendix E

Author Oh et al. Year 2014 Record Number 1

	Yes	No	Unclear	Not applicable
1. Were the criteria for inclusion in the sample clearly defined?	☒	☐	☐	☐
2. Were the study subjects and the setting described in detail?	☒	☐	☐	☐
3. Was the exposure measured in a valid and reliable way?	☒	☐	☐	☐
4. Were objective, standard criteria used for measurement of the condition?	☒	☐	☐	☐
5. Were confounding factors identified?	☒	☐	☐	☐
6. Were strategies to deal with confounding factors stated?	☒	☐	☐	☐
7. Were the outcomes measured in a valid and reliable way?	☒	☐	☐	☐
8. Was appropriate statistical analysis used?	☒	☐	☐	☐

Author Bruijnen et al. Year 2018 Record Number 2

	Yes	No	Unclear	Not applicable
1. Were the criteria for inclusion in the sample clearly defined?	x			
2. Were the study subjects and the setting described in detail?	x			
3. Was the exposure measured in a valid and reliable way?			x	
4. Were objective, standard criteria used for measurement of the condition?	x			
5. Were confounding factors identified?	x			
6. Were strategies to deal with confounding factors stated?	x			
7. Were the outcomes measured in a valid and reliable way?			x	
8. Was appropriate statistical analysis used?	x			

Author Yin et al. Year 2013 Record Number 3

	Yes	No	Unclear	Not applicable
1. Were the criteria for inclusion in the sample clearly defined?	[x]	[ ]	[ ]	[ ]
2. Were the study subjects and the setting described in detail?	[x]	[ ]	[ ]	[ ]
3. Was the exposure measured in a valid and reliable way?	[ ]	[ ]	[x]	[ ]
4. Were objective, standard criteria used for measurement of the condition?	[x]	[ ]	[ ]	[ ]
5. Were confounding factors identified?	[x]	[ ]	[ ]	[ ]
6. Were strategies to deal with confounding factors stated?	[x]	[ ]	[ ]	[ ]
7. Were the outcomes measured in a valid and reliable way?	[ ]	[x]	[ ]	[ ]
8. Was appropriate statistical analysis used?	[ ]	[x]	[ ]	[ ]

Author Norfadilah et al. Year 2017 Record Number 4

	Yes	No	Unclear	Not applicable
1. Were the criteria for inclusion in the sample clearly defined?	☒	☐	☐	☐
2. Were the study subjects and the setting described in detail?	☒	☐	☐	☐
3. Was the exposure measured in a valid and reliable way?	☐	☐	☒	☐
4. Were objective, standard criteria used for measurement of the condition?	☐	☒	☐	☐
5. Were confounding factors identified?	☐	☒	☐	☐
6. Were strategies to deal with confounding factors stated?	☐	☒	☐	☐
7. Were the outcomes measured in a valid and reliable way?	☒	☐	☐	☐
8. Was appropriate statistical analysis used?	☐	☒	☐	☐

Author Deb et al. Year 2019 Record Number 5

	Yes	No	Unclear	Not applicable
1. Were the criteria for inclusion in the sample clearly defined?	☐	☐	☒	☐
2. Were the study subjects and the setting described in detail?	☐	☒	☐	☐
3. Was the exposure measured in a valid and reliable way?	☐	☒	☐	☐
4. Were objective, standard criteria used for measurement of the condition?	☐	☒	☐	☐
5. Were confounding factors identified?	☒	☐	☐	☐
6. Were strategies to deal with confounding factors stated?	☐	☒	☐	☐
7. Were the outcomes measured in a valid and reliable way?	☒	☐	☐	☐
8. Was appropriate statistical analysis used?	☒	☐	☐	☐

Author Anjanappa et al. Year 2017 Record Number 6

	Yes	No	Unclear	Not applicable
1. Were the criteria for inclusion in the sample clearly defined?	☒	☐	☐	☐
2. Were the study subjects and the setting described in detail?	☒	☐	☐	☐
3. Was the exposure measured in a valid and reliable way?	☐	☐	☒	☐
4. Were objective, standard criteria used for measurement of the condition?	☒	☐	☐	☐
5. Were confounding factors identified?	☒	☐	☐	☐
6. Were strategies to deal with confounding factors stated?	☐	☒	☐	☐
7. Were the outcomes measured in a valid and reliable way?	☐	☒	☐	☐
8. Was appropriate statistical analysis used?	☒	☐	☐	☐

Author Kukolowicz et al. Year 2020 Record Number 7

	Yes	No	Unclear	Not applicable
1. Were the groups comparable other than the presence of disease in cases or the absence of disease in controls?	☒	☐	☐	☐
2. Were cases and controls matched appropriately?	☒	☐	☐	☐
3. Were the same criteria used for identification of cases and controls?	☒	☐	☐	☐
4. Was exposure measured in a standard, valid and reliable way?	☐	☒	☐	☐
5. Was exposure measured in the same way for cases and controls?	☒	☐	☐	☐
6. Were confounding factors identified?	☒	☐	☐	☐
7. Were strategies to deal with confounding factors stated?	☐	☒	☐	☐
8. Were outcomes assessed in a standard, valid and reliable way for cases and controls?	☒	☐	☐	☐
9. Was the exposure period of interest long enough to be meaningful?	☐	☐	☐	☒
10. Was appropriate statistical analysis used?	☒	☐	☐	☐

Logbook

Pilot study for target delineation error assessment

The pilot study for target delineation contours was conducted between the 4th of January and the 7th of January 2022. This pilot test was done with the goal of identifying any issues that might arise during data collection in terms of anonymity and the ability to delineate contours in a timely manner.

A medical physicist and a radiographer were asked to use the Monaco TPS to delineate two scans of patients who had completed treatment for laryngeal cancer. The intermediary radiographer selected these patients.

One issue raised was that if the patients were anonymised as patient 1, patient 2, patient 3, and so on, doctors could easily go on these contours and check how the other participants delineated the target. To try to obscure results from other participants, the medical physicist suggested making 6 copies of the same patient, naming them patient X1, X2, X3, X4, X5, X6 (where X is the patient number and the following number indicates the participant number), and assigning one of the copies to each participant. This procedure was tested on the participating radiographer and seemed to be effective. We then realised that this procedure also needed to be repeated for the supplementary MRI scans. The participating radiographer was able to anonymise the MRI scans effectively using CUV2 PACS software. These MRI images were copied on a pen drive, were labelled appropriately for each participating doctor (only identifiable by the intermediary radiographer) and using a free online DICOM viewer the images could be seen in their sequence.

Each participant spent approximately 10 minutes delineating targets; however, because none of the participants were clinical oncologists, time estimates could only be approximated. In practise, doctors would need to read the patient's clinical history and refer to previous diagnostic images. As a result, to gain a better understanding of delineation times, I asked one of the department's clinical oncologists how much time they typically spend delineating larynx cases. The doctor stated that this is highly

dependent on the clinical case, but it should not take more than 30 minutes. Knowing how long it would take to contour the clinical cases was helpful in estimating how long it would take doctors to contour all of the cases so that a deadline could be set.

After the delineations were done, these contours needed to be superimposed to measure the contouring range of the participants to measure the target delineation error. This procedure was done successfully with the assistance of the medical physicist.

The opinions of the participants were used to improve the procedure and tool. Using their expertise, this procedure was modified to be effective for data analysis.

Appendix F

Results for inter-observer reliability of set-up errors using mask registration

The following tables present the Cronbach's alpha and Intra-class correlation co-efficient (ICC) for translational set-up errors in the X, Y, Z directions. Values are also given for rotational set-up errors in the RX, RY and RZ direction.

Table 6.1. Cronbach's Alpha for the Left-Right (X) direction

Cronbach's Alpha	N of Items
.931	6

Table 6.2. ICC for the Left-Right (X) direction

	Intraclass Correlation	95% Conf. Interval Lower Bound	95% Conf. Interval Upper Bound	F Test with True Value 0 Value	df1	df2	P-value
Single Measures	.678	.530	.814	14.399	24	120	p<0.001
Average Measures	.927	.871	.963	14.399	24	120	p<0.001

Table 6.3. Cronbach's Alpha for the Superior-Inferior (Y) direction

Cronbach's Alpha	N of Items
.892	6

Table 6.4. ICC for the Superior-Inferior (Y) direction

	Intraclass Correlation	95% Conf. Interval Lower Bound	95% Conf. Interval Upper Bound	F Test with True Value 0 Value	df1	df2	P-value
Single Measures	.568	.404	.737	9.249	24	120	p<0.001
Average Measures	.887	.803	.944	9.249	24	120	p<0.001

Table 6.5. Cronbach's Alpha for the Anterior-Posterior (Z) direction

Cronbach's Alpha	N of Items
.955	6

Table 6.6. ICC for the Anterior-Posterior (Z) direction

		95% Conf. Interval		F Test with True Value 0			
	Intraclass Correlation	Lower Bound	Upper Bound	Value	df1	df2	P-value
Single Measures	.779	.659	.879	22.463	24	120	p<0.001
Average Measures	.955	.921	.977	22.463	24	120	p<0.001

Cronbach's Alpha for the X and Z set-up errors were very close to one (shown in Table 6.1 and 6.5). These values indicate an excellent inter-observer reliability amongst the radiographers since similar values of set-up error were obtained from CBCT image matching. These results are complimented by the ICC average measures (shown in Table 6.2 and 6.6).

The Cronbach's alpha and the ICC average measures in the Y direction, although slightly less then X and Z values, also showed a good inter-observer reliability with values close to 0.9 (shown in Table 6.3. and 6.4.).

Table 6.7. Cronbach's Alpha for the Roll (Rx) direction

Cronbach's Alpha	N of Items
.957	6

Table 6.8. ICC for the Roll (Rx) direction

		95% Conf. Interval		F Test with True Value 0			
	Intraclass Correlation	Lower Bound	Upper Bound	Value	df1	df2	P-value
Single Measures	.773	.651	.875	23.030	24	120	p<0.001
Average Measures	.953	.918	.977	23.030	24	120	p<0.001

Table 6.9. Cronbach's Alpha for the Pitch (Ry) direction

Cronbach's Alpha	N of Items
.992	6

Table 6.10. ICC for the Pitch (Ry) direction

	Intraclass Correlation	95% Conf. Interval Lower Bound	Upper Bound	F Test with True Value 0 Value	df1	df2	P-value
Single Measures	.951	.917	.975	122.015	24	120	p<0.001
Average Measures	.992	.985	.996	122.015	24	120	p<0.001

Table 6.11. Cronbach's Alpha for the Yaw (Rz) direction

Cronbach's Alpha	N of Items
.962	6

Table 6.12. ICC for the Yaw (Rz) direction

	Intraclass Correlation	95% Conf. Interval Lower Bound	Upper Bound	F Test with True Value 0 Value	df1	df2	P-value
Single Measures	.805	.696	.894	26.056	24	120	p<0.001
Average Measures	.961	.932	.981	26.056	24	120	p<0.001

The Cronbach's Alpha for the Rx, Ry and Rz rotational direction were all nearly one since they ranged from 0.957 to 0.992 (shown in Table 6.7, 6.9 and 6.11). This demonstrates an excellent inter-observer reliability. The ICC average measures add to these results with values ranging from 0.953 to 0.992 (shown in Table 6.8, 6.10, and 6.12).

Results for intra-observer reliability in set-up errors using mask registration

To assess intra-observer reliability, the radiographers were asked to re-analyse five scans that were pre-selected from the twenty-five CBCT scans after a two-weeks interval. Cronbach's alpha and ICC were also used for analysis and the results demonstrate the consistency of radiographers when analysing the same scans repeatedly.

Table 6.13. Cronbach's Alpha for the Left-Right (X) direction

Cronbach's Alpha	N of Items
.924	2

Table 6.14. ICC for the Left-Right (X) direction

	Intraclass Correlation [b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.865[a]	.666	.949	13.139	16	16	<.001
Average Measures	.928[c]	.800	.974	13.139	16	16	<.001

Table 6.15. Cronbach's Alpha for the Superior-Inferior (Y) direction

Cronbach's Alpha	N of Items
.948	2

Table 6.16. ICC for the Superior-Inferior (Y) direction

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.906[a]	.759	.965	19.128	16	16	<.001
Average Measures	.951[c]	.863	.982	19.128	16	16	<.001

Table 6.17. Cronbach's Alpha for the Anterior-Posterior (Z) direction

Cronbach's Alpha	N of Items
.662	2

Table 6.18. ICC for the Anterior-Posterior (Z) direction

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.503[a]	.041	.787	2.957	16	16	.018
Average Measures	.669[c]	.079	.881	2.957	16	16	.018

The Cronbach's alpha and the ICC of the X and Y directions show an excellent intra-observer reliability amongst the participating radiographers, however the Z direction had a questionable Cronbach's alpha result with a value of 0.662 (Table 6.17) and an ICC average measures of 0.669 (Table 6.18) which showed moderate agreement with regards to reliability**.**

Table 6.19. Cronbach's Alpha for the Roll (Rx) direction

Cronbach's Alpha	N of Items
.949	2

Table 6.20. ICC for Yaw (Rz) direction

	Intraclass Correlation [b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.827[a]	.616	.921	22.356	16	16	<.001
Average Measures	.937[c]	.798	.950	22.356	16	16	<.001

Table 6.21. Cronbach's Alpha for the Pitch (Ry) direction

Cronbach's Alpha	N of Items
.971	2

Table 6.22. ICC for the Pitch (Ry) direction

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.970[a]	.921	.989	64.119	16	16	<.001
Average Measures	.985[c]	.959	.994	64.119	16	16	<.001

Table 6.23. Cronbach's Alpha for the Yaw (Rz) direction

Cronbach's Alpha	N of Items
.919	2

Table 6.24. ICC for Yaw (Rz) direction

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.847[a]	.636	.941	12.356	16	16	<.001
Average Measures	.917[c]	.778	.970	12.356	16	16	<.001

The Cronbach's Alpha for the Rx, Ry and Rz rotational direction were all nearly one since they ranged from 0.919 to 0.971 (shown in Table 6.19, 6.21, and 6.23). This demonstrates an excellent intra-observer reliability. The ICC average measures add to these results with values ranging from 0.917 to 0.985 (shown in Table 6.20, 6.22, and 6.24).