The Utility of Post-Treatment FDG PET/CT Metrics In the Evaluation of Head And Neck Squamous Cell Carcinoma

by

Reem Althubaiti

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

Institute of Cancer and Genomic Sciences
College of Medical and Dental Sciences
University of Birmingham
October 2023

Abstract

^{18}F-Fluorodeoxyglucose (FDG) positron-emission tomography–computed tomography (PET/CT) imaging is a valuable imaging procedure for tumour staging, restaging, assessing the response to therapy, and in follow-up surveillance among patients diagnosed with head and neck squamous cell carcinoma (HNSCC). While qualitatively assessing PET/CT images has become widely adopted, quantitative analyses have also been introduced to overcome several limitations of visual analysis, enabling a more objective and easier examination of inter- and intra-patient variations. This thesis presents research to investigate the role of FDG PET/CT quantitative metrics in HNSCC. A systematic review and three retrospective studies were undertaken to assess the utility of PET/CT quantitative metrics, including maximum standardised uptake value (SUV_{max}) and peak standardised uptake value (SUV_{peak}) derived at three months following the completion of chemoradiotherapy (CRT) in HNSCC. These studies confirmed the continuous use of the SUV_{max} in post-treatment HNSCC and developed further insights regarding the diagnostic and prognostic value of these metrics in this setting. It also provided more insight into the impact of utilising different body size metrics as well as relative metrics in SUV normalisation. Specifically, the systematic review aimed to evaluate the evidence base in the literature on the diagnostic performance and prognostic value of post-treatment FDG PET/CT metrics derived from primary tumour and lymph node sites to discriminate residual disease and predict survival outcomes in patients with HNSCC. This review revealed inconsistent results, requiring further investigation to determine the value of these metrics in the post-treatment setting. The first study aimed to identify the normalisation method that is least sensitive to body size measurements when reporting PET/CT parameters in patients with HNSCC. It was found that variations in body size might have a limited impact

on those metrics; therefore, the use of any body size factors appears to be acceptable when analysing post-treatment FGD PET/CT quantitative metrics in HNSCC. The second study sought to assess the diagnostic performance of these quantitative metrics for distinguishing HNSCC residual disease. It was found that all metrics could discriminate between residual disease at the primary and nodal sites at certain thresholds. It was also found that the use of lean body mass (LBM) and body surface area (BSA), as well as a second normalisation of the FDG uptake in the background regions, did not improve their performance. Lastly, the third study investigated the prognostic utility of these quantitative metrics in predicting survival outcomes, as well as the effect of human papilloma virus (HPV) status on their prognostic value. Regardless of HPV status, post-treatment primary tumour and lymph node SUV metrics were found to be significant prognostic factors for predicting three-year progression-free survival (PFS). However, the analysis of five-year overall survival (OS) showed inconclusive findings. Overall, it was concluded that the use of either total body weight, LBM, or BSA in SUV normalisation is suitable for analysing post-treatment scans of HNSCC. This includes the evaluation of quantitative metrics for prognostication and discriminatory ability. Importantly, an increase in post-treatment SUV_{max} and SUV_{peak} values higher than the thresholds presented in this thesis, could potentially signify an augmented likelihood of disease progression. Patients with lesions exhibiting SUV_{max} and SUV_{peak} values that surpass the specified thresholds should thus be closely monitored.

Dedication

This thesis is dedicated to my mom for her endless love and support and to my children, Dana and Layan.

Acknowledgments

I would like to praise Allah, the Almighty, for the grace and guidance given to me in completing this PhD thesis.

I would like to express my gratitude and profound appreciation to my principal supervisor, Professor Hisham Mehanna, for his efforts, generous time, and guidance over the past four years. I am deeply appreciative of his continuing support, patience, and willingness to share his enormous expertise and knowledge in the field of head and neck cancer with me.

I would also like to extend my sincere gratitude and thanks to my co-supervisor, Dr. Paul Nankivell, for his guidance, advice, and support throughout the course of my PhD studies.

I am extremely grateful to my supervisor Dr. Bal Sanghera for his continuous encouragement and supervision of the technical aspects of my research.

I am greatly thankful to Dr. Ridhi Agarwal for her time and effort in guiding me through the statistical analyses.

I appreciate Dr. Khalid Hussain's support with PET/CT image analysis and oversight of my work in the nuclear medicine department. Many thanks to Daniel Fakhry-Darian for his invaluable help with the PET/CT scans' data processing.

I would also like to thank my colleagues and team members at InHANSE for their support and insights. Special thanks to Ahmad Abou-Foul for his contribution to the systematic review as a second reviewer. Thank you to all of my friends for their kindness and the good times we shared.

My thanks also go to King Abdul-Aziz University in Saudi Arabia for sponsoring my studies.

I would like to express my gratitude and appreciation to my family. I would like to thank my parents, who have been a source of inspiration to me throughout my life, for their continuous support and prayers. Also, many thanks to my sisters and brothers. To my lovely daughters, Dana and Layan, you gave me the strength to keep going when I felt like giving up. Finally, I would like to express my deep appreciation and love to my husband, Omar Alabbasi, for everything he has done for me throughout my PhD journey.

Table of contents

List of figures

List of tables

List of common abbreviations

American Joint Committee for Cancer (AJCC)

Area under the ROC curves (AUC)

Blood pool (BP)

Body mass index (BMI)

Body surface area (BSA)

Chemoradiation therapy (CRT)

Computed tomography (CT)

Concurrent chemoradiotherapy (CCRT)

Confidence intervals (CI)

Disease-free survival (DFS)

Disease-specific survival (DSS)

European Association of Nuclear Medicine (EANM)

European Organization for Research and Treatment of Cancer (EORTC)

Fluorine (^{18}F)

[^{18}F]-fluoro-2-deoxyglucose (FDG)

Hazard ratio (HR)

Head and neck cancer (HNC)

Head and neck squamous cell carcinoma (HNSCC)

Health Research Authority (HRA)

Histopathology (HPR)

HPV negative disease (HPV^{-ve})

HPV positive disease (HPV^{+ve})

Human Papilloma Virus (HPV)

Induction chemotherapy (IC)

Intensity-modulated radiation therapy (IMRT)

Interquartile range (IQR)

Intraclass correlation coefficient (ICC)

Lean body mass (LBM)

Locoregional control (LRC)

Locoregional response to treatment (LRT)

Lymph node (N)

Magnetic resonance imaging (MRI)

Maximum standardised uptake value (SUV_{max})

Maximum SUV normalised to body surface area (SUVmax (bsa))

Maximum SUV normalised to lean body mass (SUVmax (lbm))

Maximum SUV normalised to total body weight (SUVmax (w))

Mean standardised uptake value (SUV_{mean})

Metabolic tumour volume (MTV)

Metastatic-free survival (MFS)

Nasopharyngeal carcinoma (NPC)

National Cancer Institute (NCI)

Negative predictive value (NPV)

Neoadjuvant chemoradiotherapy (NCRT)

Oral squamous cell carcinoma (OSCC)

Oropharyngeal squamous cell carcinoma (OPSCC)

Overall survival (OS)

Peak standardised uptake value (SUV_{peak})

Peak SUV normalised to body surface area (SUVpeak (bsa))

Peak SUV normalised to lean body mass (SUVpeak (lbm))

Peak SUV normalised to total body weight (SUVpeak (w))

Positive predictive value (PPV)

Positron emission tomography/ computed tomography (PET/CT)

Primary tumour (T)

Progression-free survival (PFS)

Radiation therapy (RT)

Receiver operating characteristic (ROC)

Region of interest (ROI)

Spearman correlation coefficient (r)

Standard deviations (SD)

Standardised uptake value (SUV)

Standards for the Reporting of Diagnostic Accuracy Studies (STARD)

Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)

The Response Evaluation Criteria in Solid Tumors (RECIST)

Three- dimensional conformal radiation therapy (3D-RT)

Total lesion glycolysis (TLG)

Ultrasonography (US)

Volume of interest (VOI)

Weight (W)

World Health Organization (WHO)

CHAPTER ONE

GENERAL INTRODUCTION

1. Chapter 1. General introduction

1.1 Epidemiology

Head and neck cancer (HNC) is the seventh most prevalent malignancy (Schuttrumpf et al., 2020, Mody et al., 2021) and encompasses several types of tumours that originate in the anatomic sites of the upper aerodigestive tract (Mody et al., 2021). In 2018, approximately 890,000 new cases of head and neck squamous cell carcinoma (HNSCC) were reported worldwide, with a death rate of around 450,000 (Johnson et al., 2020). Between 2016 and 2018, Cancer Research UK estimated that 12,400 new cases of HNC were diagnosed annually in the UK, representing 3% of all new cancer cases and making HNC the eighth most prevalent cancer in the country (Cancer Research UK, 2023).

1.2 Subtypes of HNC

HNCs consist of cancers of the oral cavity, nasopharynx, oropharynx, hypopharynx, larynx, paranasal sinuses, and salivary glands (Mehanna et al., 2010, Pak et al., 2014). Of these cancers, HNSCC makes up greater than 90% of all head and neck malignancies (Sanderson and Ironside, 2002, Karam et al., 2018, UK, 2020, Hsieh et al., 2021). Incidences of HNSCC vary by country/region and may be related to tobacco-derived carcinogens or excessive alcohol intake (Johnson et al., 2020). It is usually identified in older people who are heavy smokers and drinkers (Chow, 2020).

1.2.1 Human Papilloma Virus (HPV)

HPV infection is linked to an increasing number of oropharyngeal malignancies. There are over 130 different HPV types that have been identified and classified according to their carcinogenic potential; the most common is HPV 16, which accounts for 90% of all HPV-related oral squamous cell carcinoma (OSCC). In comparison to HNSCC not associated with HPV, HPV-associated OSCC has evolved as a different head and neck subtype with unique clinical features and a molecular profile, highlighting the need for routine HPV testing for OSCC (Economopoulou et al., 2019).

1.3 Staging

Staging is a fundamental step in making treatment decisions for patients diagnosed with HNC. The American Joint Committee for Cancer (AJCC) has provided several staging schemas for each type of HNC, which follow the TNM staging criteria for malignant lesions (Stepnick and Gilpin, 2010). The T element represents the size and extent of the primary lesion, which may be different depending on the subtype of the lesion. Regional metastatic lymph node involvement can be classified with the N element, while distant metastatic lesions such as those in bones, soft tissues, or distant lymph nodes can be described by the M element (Plaxton et al., 2015). In addition, further analysis relating to prognosis can be achieved by using stage groups, which can be indicated with Roman numbers I-IV to describe early-stage to advanced-stage disease (Wong, 2018).

1.4 Management of Head and Neck Cancer

1.4.1 Surgery

Simple surgical excisions of primary tumours may be performed to remove small malignant lesions without the need for reconstruction. Obtaining pathologic analysis of resected lesions helps to predict the prognosis and guide therapy, especially in patients with increased risk of recurrent disease. In primary cancer cases with larger sized lesions, surgery is often necessary, which involves complete excision combined with reconstruction (Wong, 2018). Moreover, a neck dissection, which is defined as a surgical approach used for removing involved lymph nodes, is a standard medical intervention for managing neck lymphadenopathy. Establishing nodal level for the involved lymph node is a critical factor in determining the type of neck dissection to perform, which is mainly classified as a radical neck dissection or a modified radical neck dissection (Robbins et al., 2002, Wong, 2018).

1.4.2 Radiotherapy

External beam radiation therapy (RT) with or without adjuvant chemoradiation therapy (CRT) is a well-established standard treatment strategy that can be used as a definitive treatment or as an adjuvant to primary tumour surgery. Recent technological advances enabled the development of three-dimensional conformal radiation therapy (3D-RT). Although this technique has an improved ability to delineate tumour volume effectively compared to 2D-RT, it poses some limitations such as exposing normal tissues to unnecessary radiation during treatment (Deschler et al., 2014). A further improved type of RT is intensity-modulated radiation therapy (IMRT). One prominent advantage of this technique

is that it reduces the radiation dose to normal surrounding critical organs while providing adequate coverage of the target volume (Deschler et al., 2014, Castelli et al., 2018).

1.4.3 Chemotherapy

While chemotherapy is not effective as a primary modality to treat cancer as a sole treatment, the use of it has significantly improved cancer control, organ preservation, and delayed time until recurrence in certain conditions. Several chemotherapy approaches are typically used for cancer treatment. The most commonly used method is concurrent chemoradiotherapy (CCRT), which is a method of administering chemotherapeutic agents during RT. The goal of this treatment approach is to intensify the effect of radiation to improve locoregional control and organ preservation. This type of treatment has been shown to be effective in moderately advanced stage (III, IV) pharyngeal and laryngeal cancers, except in cases of T4 laryngeal and hypo-laryngeal malignant lesions. This method can be applied in conjunction with primary or adjuvant RT. Previous evidence has shown that this method can produce favourable outcomes in terms of locoregional control (LRC) and overall survival (OS) in patients with positive multiple lymphadenopathy, positive lesions margins, and in cases with cervical extracapsular spread (Deschler et al., 2014).

1.5 Medical Imaging

1.5.1 CT & MRI

For many years, conventional medical imaging procedures, including x-ray radiography, cross-sectional imaging with computed tomography (CT), magnetic resonance imaging (MRI), and ultrasonography (US), have been utilized for anatomical localization and characterization of lesions based on differentiation of morphological size, density, and water content (Farwell et al., 2014). In addition, obtaining functional information such as tumour metabolic activity has been shown to be useful for the detection of lesions, which can be achieved by PET/CT imaging (Sheikhbahaei et al., 2015).

1.5.2 PET/CT

Imaging with ^{18}F-fluorodeoxyglucose (FDG) positron emission tomography (PET) has emerged as a valuable functional imaging technique for providing analysis of the cellular activity of lesions in a variety of oncologic applications. With advances in hardware and software applications, integrated PET/CT systems have been developed that allow sequential image acquisition of both structural and functional information of target lesions, providing improved sensitivity and specificity with a single imaging modality compared to anatomical imaging techniques (Farwell et al., 2014).

1.6 Principles of PET

PET imaging is based on measuring radioactivity distribution *in vivo*. The principle of PET involves injecting a radiotracer, which distributes within patients' body organs and tissues (Hogg and Testanera). Positively charged particles (positrons) are emitted from radioactive atoms, these combine with negatively charged electrons in a process called annihilation, leading to the emission of two photons travelling in opposite ways. Each of these photons will have an energy of 511 keV (Van Der Veldt et al., 2013). These photons are captured within a small-time window ~ns with high-sensitivity coincidence imaging by PET detectors that surround the patients. Finally, acquisition of millions of coincident detection events, following image reconstruction techniques, generates a spatial distribution of radioactivity inside the body, producing PET/CT images (Hogg and Testanera) (Figure 1.2).

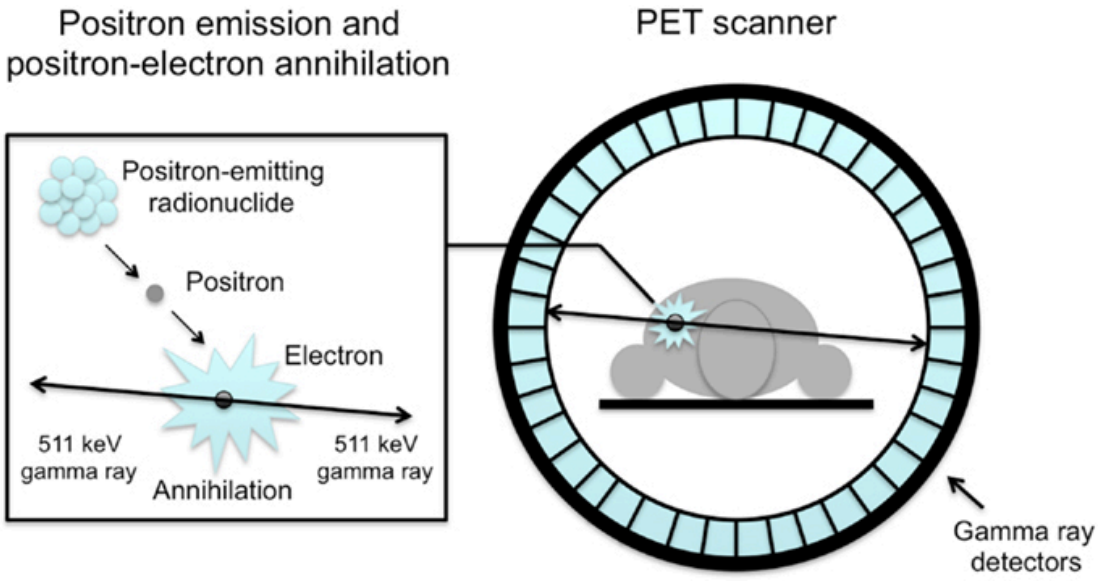


Figure 1.1: Positron emission and annihilation processes.
Demonstration of positron and electron annihilation reaction and the process of detecting coincidence imaging (Van Der Veldt et al., 2013).

1.7 PET FDG pharmacokinetics

^{18}F, a fluorine radionuclide, is synthesized by medical cyclotrons and has a short half-life of 109.7 min. It is used to label several molecular tracers for imaging different targets. Currently, FDG, a glucose analogue, is the most widely used radiopharmaceutical for diagnosis, staging, and restaging multiple tumours (Boellaard et al., 2015). Malignant, metabolically active cells, absorb the radiotracer FDG in a pattern similar to the pattern utilized by glucose. Basically, glucose is transported by facilitated transporters (GLUT) followed by undergoing a glycolytic reaction (glycolysis). The more metabolically active the cells, the greater the FDG uptake within the cells. However, unlike glucose, once absorbed by the metabolically active cells, FDG does not undergo additional metabolic reactions but remains stable, providing a mean for imaging (Kapoor et al., 2004, Abouzied et al., 2005).

1.8 Standard PET/CT Protocol

A standard PET/CT protocol, described by Boellaard et al. (2015), involves proper patient preparation, intravenous administration of the radioactive tracer, and PET/CT image acquisition. First, patients are instructed to refrain from consuming coffee or caffeinated drinks for four hours prior to FDG injection. In addition, patients are asked to avoid strenuous exercise for at least 6 hours and ideally 24 hours before the examination to minimize muscle uptake. Moreover, to allow normal biodistribution of the tracer, the patient's blood glucose must be maintained below 200mg/dL prior to FDG injection. Next, an average of 60 minutes of uptake time is advised, with an acceptable range of 55–75 minutes. This is the period between FDG injection and initiation of the PET/CT scan. To reduce the possibility of brown fat uptake and muscular uptake in the neck region, patients are recommended to keep warm, inactive,

and silent during the uptake time. After that, CT image acquisition is acquired first, followed by a PET sequence. Finally, CT non-attenuation corrected, CT attenuated corrected, PET images, and co-registered PET/CT images are produced and retrieved for analysis and reporting (Boellaard et al., 2015).

It is noteworthy to remark that ^{18}F-fluoromisonidazole is another potential PET tracer that provides quantitative assessments of hypoxia, which is one of the key factors influencing treatment resistance in HNSCC. In a study on 73 patients with HNSCC, they found that a significant number of patients (79%) exhibited hypoxia, and the degree of hypoxia and the size of the hypoxic volume were both independent factors in survival outcomes (Dirix et al., 2009). Diffusion-Weighted Magnetic Resonance Imaging (DW MRI) is another medical imaging technique that has demonstrated numerous prospective applications in the field of head and neck cancer, as evidenced by preclinical and clinical research. It can be used to determine the aggressiveness of lesions and monitor the response to therapy, providing valuable information for treatment planning and evaluation (Dirix et al., 2009). Dynamic contrast-enhanced MRI is an additional imaging technique that involves injecting a contrast agent intravenously to assess changes in signal intensity. This technique provides information about the permeability, perfusion, and interstitial pressures within lesions, which are factors that impact treatment response (Dirix et al., 2009).

1.9 FDG Uptake Assessment

1.9.1 Qualitative approach

A significant point to address is the method by which FDG uptake is assessed with PET/CT imaging. A qualitative assessment is a widely used approach to evaluate FDG uptake in PET/CT imaging. This method consists of a visual inspection of the metabolic activity of target lesions in comparison to surrounding tissues or reference organs. For instance, visually comparing FDG uptake in lesions to the uptake in the liver or mediastinal blood pool. The avid appearance of asymmetric FDG uptake in lesions compared to the background uptake in a staging PET/CT exam is usually indicative of an abnormality. In assessing post-treatment PET/CT images, FDG absence or persistence in target tumours indicate a complete response or therapy failure, respectively (Ziai et al., 2016). In addition, well-trained nuclear medicine practitioners consider the patient's clinical and family history, laboratory results, and other imaging studies when performing a qualitative evaluation (Hirata and Tamaki, 2021).

The Response Evaluation Criteria in Solid Tumours (RECIST) guidelines, created in collaboration with the European Organization for Research and Treatment of Cancer (EORTC), the National Cancer Institute (NCI) of the United States, and the National Cancer Institute of Canada Clinical Trials Group, refined and simplified the World Health Organization (WHO) standardised criteria for assessing tumour response. Tumour burden, overall, is determined by summing the size of lesions in baseline scans prior to the start of treatment. Treatment response is then determined by measuring the change in the size of these lesions (Hirata and Tamaki, 2021).

Although qualitative assessment of PET/CT images is still the most commonly used method, a major problem associated with visual inspection of PET/CT images is that it is a subjective method of interpreting the results, which makes it greatly operator-dependent (Manca et al., 2016).

1.9.2 Quantitative approach

In terms of response assessment with FDG PET, the EORTC recommendations represented a semi-quantitative approach, which is based on the proportional change in SUV value following standard patient preparation procedures. The Response Criteria in Solid Tumours is a second proposed set of quantitative criteria for FDG PET response evaluation (PERCIST) that assesses lesion metabolic response to treatment. This modified version recommends using lean body mass (LBM) normalised standardised uptake value (SUV) over the ordinary total body weight for SUV normalisation, which may demonstrate practical improvements in cancer care and clinical research (Hirata and Tamaki, 2021). In Chapter 4, we addressed factors that influence SUV in head and neck cancer and mentioned the possible benefits of LBM as a normalising factor.

While the proposed metabolic response criteria have the potential to improve clinical reporting in a variety of ways, their widespread use is currently limited by several challenges. First, the effectiveness of these criteria has not been extensively validated in clinical research. Second, some of these metrics are only used for specific types of tumours, such as the Deauville/London and Cheson criteria for lymphoma, limiting their applicability to other cancers. Thirdly, for effective evaluation of the percentage change in pre- and post-treatment SUVs, both PET/CT scans should be acquired under

roughly identical conditions. Lastly, these reporting systems are always developing, and the best criteria could change over time (Niederkohr et al., 2013).

Absolute quantification is difficult to achieve clinically, involving dynamic scans, defining arterial input functions, and tissue-time activity curves with compartmental model analysis (Hirata and Tamaki, 2021).

However, the ability of PET to provide a simple *in vivo* semi-quantification metric of metabolic and physiological markers in various malignant neoplasms, including HNSCC has been proposed as an effective tool. Although a qualitative analysis of PET/CT images is sometimes adequate for diagnosis, staging, and detection of malignant lesions, PET quantification can provide a more effective and accurate approach, especially for assessing therapy response, improving prognostic value, and in radiotherapy planning (Manca et al., 2016). One potential application of quantitative assessment of the therapy response is enabling early and accurate stratifying responder and non-responder patients (Boellaard, 2009).

Importantly, analysing lesions quantitatively can be advantageous since it may be less operator-dependent and can be automatically derived, thus enabling more objective and convenient methods for examining inter- and intra-patient variations (Castelli et al., 2016). Niederkohr et al. (2013) emphasized that the lack of a quantifiable parameter may make it difficult or impossible to compare studies among different institutions. Therefore, due to the need for a more objective method for interpreting tumour metabolic activity, research in the field of PET/CT quantification is expanding.

1.10 Quantitative metrics

PET/CT quantitative analysis may be categorized into semi-quantitative and volumetric parameters, which can be produced semi-automatically by drawing regions of interest (ROI) or volumes of interest (VOI) contouring target lesions.

1.10.1 SUV

SUV is a routinely applied semi-quantitative parameter in clinical practice, reflecting tracer radioactivity concentration within a 2D *ROI* or 3D *VOI*.

Semi-quantitative parameters consist of the mean standardised uptake value (SUV_{mean}), maximum standardised uptake value (SUV_{max}), and peak standardised uptake value (SUV_{peak}).

SUV_{mean} represents the mean of radioactivity concentration obtained from all voxels within the ROI or VOI where a decision must be made where the lesion ends. SUV_{max} corresponds to the radioactivity concentration at the most active voxel within the ROI or VOI. SUV_{peak} generally reflects the average value of radiotracer uptake, typically within a 1 cm^3 VOI surrounding the pixel with the maximum activity (Ziai et al., 2016) (Table 1.1).

To more accurately calculate the SUV, a body size metric should be applied to eliminate possible dependency on factors like patient weight and injected activity. The general formula for calculating SUV is presented in Equation 1.

$$SUV = \frac{tissue\ tracer\ activity}{injected\ dose/body\ size}$$ (Ziai et al., 2016)

Equation 1.1: The general formula for calculating the standardised uptake value (SUV) in PET/CT imaging

The most commonly used body size metric in SUV calculation is total body weight in kg. Other body size metrics that are less commonly used are LBM and body surface area (BSA).

The following are SUV definitions based on the body size metric used for normalisation:

- SUV_w is for SUV scaled to total body weight in kg.
- SUV_{lbm} is for SUV scaled to LBM. This is the same metric of SUL defined in PERCIST guidelines (Wahl et al., 2009).
- SUV_{bsa} is for SUV scaled to BSA.

The same formula is applied to calculate the quantitative metrics of SUV_{max}, SUV_{mean}, and SUV_{peak} (Ulaner, 2018).

Each of the quantitative metrics has advantages and limitations. For example, one of the advantages of SUV_{max} is that its value is independent of the size of the VOI applied provided it covers the most active voxel. This makes it less observer-dependent, hence more reproducible than SUV_{mean} (Ziai et al., 2016). On the other hand, since SUV_{max} only corresponds to the value of a single voxel/pixel within a VOI/ROI, it does not reflect FDG intensity from the whole lesion. Another limitation is that SUV_{max} is more

susceptible to image noise than other imaging parameters. This increases its sensitivity to the differences in patient factors and imaging protocols (Im et al., 2018).

Similarly, SUV_{mean} has advantages and limitations. Because radioactivity concentration is taken from many voxels, this makes SUV_{mean} less sensitive to image noise than SUV_{max}. However, SUV_{mean} is susceptible to inter- and intraobserver variability due to its reliance on the size of the ROI (Ziai et al., 2016).

SUV_{peak} might be a more robust measurement because it combines the repeatability of SUV_{max} with the reduction of image noise provided by SUV_{mean}. However, the shape and size of the selected ROI can have a significant impact on the SUV_{peak} value. In addition, there are no standard guidelines for defining its value (Vanderhoek et al., 2012, Im et al., 2018). SUV_{peak} is also difficult to evaluate in small lesions, and it requires specialist software that is not generally available (Ziai et al., 2016) and is not routinely used clinically. Therefore, in conclusion, there is no perfect imaging metric for lesion quantification.

Notably, comparison between different SUVs ideally requires SUV harmonization to avoid SUV bias. There are numerous national schemes that align phantom SUVs to an acceptable range, typically using 3D filters (Daisaki et al., 2021). However, there can be challenges with this approach when comparing historic data between different scanners if phantom data was unavailable or clinical protocols were acquired under different conditions than the phantom acquisition.

1.10.2 Volumetric Parameters

Volumetric parameters constitute the second division of PET/CT quantitative measurements, which consist of metabolic tumour volume (MTV) and total lesion glycolysis (TLG), as well as SUV_{mean} and SUV_{peak}, which were previously described (Table 1.1).

The MTV value represents the volume of the tumour with significant metabolic activity, whereas TLG is the measurement of MTV multiplied by determining the SUV_{mean} of the target tumour. To define the MTV value, several segmentation methods have been described in the literature (Im et al., 2018).

Table 1.1: PET/CT semi-quantitative & volumetric measurements.

Parameter	Name	Definition	Advantages	Disadvantages
SUV_{mean}	Mean standardised uptake value	Mean value of all voxels within the ROI/ VOI	_It is less sensitive to image noise than SUV_{max}	_It is susceptible to inter- and intraobserver variability due to its reliance on the size of the ROI
SUV_{max}	Maximum standardised uptake value	Highest voxel value within the ROI/ VOI	_It is independent of the size of the VOI. This makes it less observer-dependent, hence more reproducible than SUV_{mean}	_It does not reflect FDG intensity from the whole lesion. _It is more susceptible to image noise than other parameters.
SUV_{peak}	Peak standardised uptake value	Mean value of radioactivity concentration around the voxel with the highest activity	_It combines the repeatability of SUV_{max} with the reduction of image noise provided by SUV_{mean}	_It is affected by the shape and size of the ROI _There are no standard guidelines for defining its value _It is difficult to evaluate in small lesions
MTV	Metabolic tumour Volume	Radioactivity concentration within volume of tumour	_It covers the whole lesion	
TLG	Total Lesion Glycolysis	The result of MTV multiplied by SUV_{mean}	_It covers the whole lesion	

1.11 Segmentation methods for measuring MTV

One of the main methods utilized to segment a tumour for defining TLG is threshold-based segmentation. The thresholding technique may be subsequently divided into four approaches:

1) The first approach is measuring MTV based on a fixed absolute threshold. In this method, tumour regions and surrounding background within the images are partitioned based on a specific threshold value, thus assigning voxel values that are greater than the threshold value to the tumours, whereas lower values than the threshold apply to the background accordingly. Several fixed absolute thresholds have been used, ranging from 2.0 to 5.0, of which a threshold of 2.5 has been the most used threshold to define MTV (Im et al., 2018). However, there are known issues with using a fixed absolute threshold. If a lesion has low FDG uptake, it may not meet the absolute threshold, making MTV calculation difficult. Furthermore, if a tumour has a high FDG uptake, such as greater than an SUV of 15, MTV could be easily overestimated due to the spillover effect (Im et al., 2018).
2) A fixed-relative threshold is the second approach to calculate MTV and is applied by utilizing a fixed percentile value of SUV_{max} within a tumour ROI and rejecting pixels below this value. In this method, percentages typically ranging from 33-44% have been used to measure MTV; however, 40% and 42% of SUV_{max} are the most commonly used percentages to calculate tumour volume (Im et al., 2018). However, in post-treatment head and neck cancer, the accuracy of these segmentations should be validated.

3) The third thresholding technique is the background thresholding technique, which entails first placing an ROI over a reference region, such as the liver or mediastinal blood pool, followed by calculating a threshold value equal to SUV_{mean} + 1 or 2 standard deviations (SD) (Im et al., 2018).
4) The last thresholding method is the adaptive thresholding technique, in which the appropriate threshold value is adjusted on a case-by-case basis instead of pre-assigning specific fixed absolute or relative values (Im et al., 2018). Thus, the threshold value for each lesion may differ, depending on the visual assessment of the lesion that is being examined.

Therefore, establishing the right segmentation is essential in post-treatment head and neck cancer to enable the implementation of the thresholding technique for HNSCC lesion segmentation.

1.12 Body size metrics in SUV calculation

In clinical practice, the use of total body weight is the most common body size metric applied for normalisation (Adams et al., 2010, Aide et al., 2017). However, studies have found that the use of total body weight in SUV standardization results in an overestimation of SUV values, especially in obese patients. Obese patients usually have higher fat percentages, and white fat absorbs less FDG, causing SUV values in other tissues to rise (Sarikaya et al., 2020). In a thin patient with considerable muscle mass, there may likely be a lower lesion SUV value because muscle competes for FDG with the lesion (Adams et al., 2010).

Therefore, to minimize bias from SUV dependence on body weight and to allow adequate SUV comparisons among patients, normalisation to other body size metrics has been investigated. In normal organs, numerous studies have recommended the use of LBM (Sanghera et al., 2009, Keramida and Peters, 2019, Sarikaya et al., 2020) and BSA (Kim et al., 1994) for SUV normalisation. However, in cancer-specific lesions, a study found that normalisation of lung cancer SUVs to LBM as a function of height alone should be avoided due to the significant correlation observed between SUV_{lbm} and height (Hallett et al., 2001). In HNSCC patients, a limited number of studies evaluated the effect of different normalisation methods on PET/CT metrics on the basis of the correlations between post-treatment SUVs and these independent variables. Finding the normalizing method that is least correlated with body size parameters will most likely be the best metric to reduce SUV error due to patient size. As a result, more accurate SUV readings would be applied in the interpretation of post-treatment PET/CT scans of HNSCC.

1.13 Lesions-to-background ratios

In the absence of standardization, absolute thresholds are difficult to implement in routine clinical practice. Establishing thresholds requires the use of standardised protocols, acquisitions, and reconstructions to facilitate SUV comparability. In routine clinical practise, the use of a standard protocol can be difficult, as some patients' imaging routines can deviate from the standard protocol and would not be suitable for quantitative interpretation. One potential solution that has been proposed is by comparing lesion SUV with the SUV derived from a background region. In other words, dividing the lesion SUV by the background SUV. This might be useful in reducing the variation among

quantitative metrics. The theory is that the same parameters, such as injected activity, glucose level, and reconstruction mechanisms, have similar effects on both lesion and background uptakes and so are nullified in the quotient. Consequently, these relative measurements could be applicable in routine imaging with equivalent diagnostic performance (Helsen et al., 2020b).

An early study by Van Den Hoff et al. (2013) found that tumour-to-blood SUV ratio was preferable to tumour SUV as a surrogate imaging metric of lesions metabolism in patients with colorectal cancer (van den Hoff et al., 2013). In a subsequent study of rectal cancer patients who underwent neoadjuvant chemoradiotherapy (NCRT), SUV_{max} normalised to liver uptake was a better predictor of pathologic complete response rate than SUV_{max} and SUV_{max} normalised to blood pool uptake (Park et al., 2014). Only one study in HNSCC evaluated lesions-to-background ratios and found that the best diagnostic performance relative metric was achieved when lesions SUVs were further normalised by the liver SUV (Helsen et al., 2020b). Even though this was the first study to compare lesion-to-background ratios in HNSCC, their evaluation was restricted to SUVs derived from nodal lesions. Further evidence is therefore needed to support their findings and clarify whether the methods of normalisation to background uptake have a true effect on SUV utility in post-treatment HNSCC.

1.14 Diagnostic performance of post-treatment PET/CT

SCC has been described earlier as the most common head and neck cancer, including the primaries of the oral cavity, hypopharynx, larynx, and oropharynx (Hsieh et al., 2021).

At the time of diagnosis, the majority of patients present with advanced locoregional disease (Tang et al., 2020), and where unresectable, CRT or RT alone are commonly used.

An important prognostic factor in HNSCC is tumour response to definitive CRT; 50–70% of patients with HNSCC achieve a complete response after therapy, with 30–50% showing evidence of residual or recurrent disease either at the primary site or regional lymph node (Gupta et al., 2011). After treatment failure, the median overall survival is less than one year (Sagardoy et al., 2016). This emphasizes that early detection of recurrence is important to optimize clinical management of this malignancy, emphasizing the value of post-therapy follow-up imaging procedures used for early imaging of recurrences as they can fundamentally alter clinical management and survival (Sheikhbahaei et al., 2015). It would therefore be highly advantageous to have an accurate tool to evaluate treatment efficacy early to identify patients who require salvage-tailored treatments (Cacicedo et al., 2016).

Standard imaging techniques, such as CT and MRI, serve an important role in HNSCC management. However, inconclusive findings due to inflammation caused by chemoradiotherapy or the presence of fibrotic lesions may be apparent when performing imaging for post-treatment conventional assessment (Lell et al., 2000).

A potential approach to improve diagnostic accuracy is the use of functional imaging with PET (de Bree et al., 2009). The development of an integrated PET/CT system has greatly contributed to the management of HNSCC in terms of staging, restaging, and assessing response to therapy.

One significant advantage of PET/CT is the additional functional evaluation in terms of metabolic activity (glycolysis) of malignant lesions. Thus, it allows identification of metabolic abnormalities before morphological changes occur (Kim et al., 2021).

Although qualitative analysis of PET/CT images is the current standard for staging and restaging (Boellaard, 2009), quantitative analysis offers the prospect of a more objective and accurate approach, particularly for assessing therapy response and improving prognostic value compared to visual inspection alone (Boellaard, 2009, Manca et al., 2016). In addition, PET/CT quantification is less operator-dependent and can be more automatically derived, thus enabling easier and more accurate objective methods for examining inter- and intra-patient variability (Castelli et al., 2016).

The high negative predictive value (NPV) of PET/CT performed three-month post-treatment was confirmed in a randomized clinical trial (Mehanna et al., 2016). However, the low positive predictive value (PPV) (about 50%) results in a high rate of false positives (Cheung et al., 2017). Therefore, improving the accuracy to distinguish post-treatment inflammation from residual disease is critical to minimize the number of unnecessary invasive procedures and their associated costs and morbidities.
One potential method that has been investigated is the use of improved post-treatment PET/CT metrics to accurately detect residual lesions and predict survival outcomes. Several quantitative parameters have been examined, including SUV_{max} (Vainshtein et al., 2014, Shimomura et al., 2014, Katahira-Suzuki

et al., 2015, Sagardoy et al., 2016, Fatehi et al., 2019a, Dejaco et al., 2020, Helsen et al., 2020a), and less commonly SUV_{peak}. Despite this, no consensus has yet been reached on defining quantitative thresholds for these parameters and their diagnostic performance.

1.15 Prognostic value of PET/CT metrics

1.15.1 Pre-treatment PET/CT parameters

The utility of pre-treatment PET/CT quantitative metrics has been assessed in several studies. Bonobo et al. (2018) conducted a meta-analysis of 25 studies, including a total of 2,223 patients; approximately 95% were diagnosed with locally advanced HNSCC and received concurrent CRT for treatment. Analysis of relative risk (RR) estimates for determining associations between the prognostic value of baseline FDG parameters and OS, progression-free survival (PFS), and LRC was conducted on 11, 8, and 4 studies, respectively. Specifically, the impacts of pre-treatment SUV_{max} and MTV on patient survival outcomes were analyzed. The authors revealed that unfavourable outcomes were significantly associated with only a greater baseline MTV value, with a reported RR estimate of OS (summary RR 1.86, 95% CI 1.08–3.21), PFS (summary RR 1.81, 95% CI 1.14–2.89), and LRC (summary RR 3.49, 95% CI 1.65– 7.35). Thus, the results suggested that pre-treatment MTV was the only prognostic factor of patient outcomes, while pre-treatment SUV_{max} was not (Bonomo et al., 2018).

Similarly, in another recent meta-analysis of 28 studies, which examined the association between pre-treatment PET/CT metrics, including SUV_{max}, MTV, and TLG, in 1871 patients with biopsy-confirmed HNC and survival data. They found that higher values of both pre-treatment MTV and TLG were

associated with unfavourable event-free and overall survivals, suggesting that volumetric parameters might provide better prognostic biomarkers compared to SUV_{max} (Wang et al., 2019a). Another systematic review and meta-analysis reached similar conclusions (MTV, TLG, and patient outcomes) (Pak et al., 2014).

The prognostic value of baseline PET/CT has also been investigated in several subsites of HNC. An early prospective study evaluated the potential prognostic value of pre-treatment primary tumour TLG and nodal SUV_{max} in 126 subjects with oral cavity squamous cell carcinoma (OCSCC) who underwent radical surgery for treatment. Patients with lower TLG and nodal SUV_{max} than 71.4 and 7.5, respectively, reported favourable survival outcomes. Therefore, pre-treatment TLG and SUV_{max} were found to be prognostic factors in patients with OCSCC (Abd El-Hafez et al., 2013).

The predictive significance of primary tumour pre-treatment SUV_{max} in patients with locally advanced nasopharyngeal carcinoma (NPC) has also been investigated. The authors found that pre-treatment SUV_{max} may be a predictor of disease-free survival in patients with NPC (Xie et al., 2010), confirming the result of a previous observation (Lee et al., 2008).

In contrast, a meta-analysis by Huang et al. (2017) found that pre-treatment primary tumour metrics were associated with patient survival, while pre-treatment nodal metrics were not (Huang et al., 2017). Despite the overall promising role of pre-treatment PET/CT parameters in predicting clinical outcomes in HNSCC, additional studies showed that PET/CT metrics were not significant predictors of clinical outcomes. For example, a retrospective study by Aslan et al. (2017) investigated the prognostic

significance of pre-treatment SUV_{max}, MTV, and TLG in 47 subjects with locally advanced HNSCC in predicting survival outcomes (DFS and OS). This study found that none of the pre-treatment PET/CT metrics were significantly associated with patient outcomes, suggesting that pre-treatment metrics are not prognostic factors in HNSCC (Aslan et al., 2019). Another study that evaluated the prognostic role of baseline SUV_{max} in locally advanced HNSCC also demonstrated no correlation with OS, DFS, or LRC (Cacicedo et al., 2017b). As a result, contradictory results have been seen regarding the significance of pre-treatment PET/CT measurements in predicting treatment success and patient survival.

1.15.2 Post-treatment PET/CT parameters

Alterations in the characteristics of the tumour that occur during treatment have the potential to be predictive of treatment response and long-term prognosis (Martens et al., 2019).

Several studies have examined the utility of post-treatment PET/CT semiquantitative and volumetric measurements in patients diagnosed with HNSCC (Mayo et al., 2019, Dejaco et al., 2020, Helsen et al., 2020b, Connor et al., 2021).

Similar to pre-treatment research findings, positive and negative associations were found between SUV post-treatment metrics and patient outcomes. A retrospective study of 37 oral cancer patients which investigated the prognostic value of pre-treatment and eight-week post-treatment SUV_{max} values found that post-treatment SUV_{max} was the only parameter associated with patients' survival (Oyama et al., 2020). Similar findings were observed in 70 patients with pharyngeal SCC (Katahira-Suzuki et al., 2015)

and in 98 patients with pharyngolaryngeal SCC (Moeller et al., 2010). However, other studies showed no association between post-treatment metrics and patient outcomes (Connor et al., 2021).
A detailed evaluation of the role of post-treatment PET/CT metrics is covered in Chapter 2.

While conflicting results on the prognostic significance of PET/CT metrics for predicting prognosis were shown, studies , in general, have demonstrated that the SUV_{max} in a post-treatment setting could offer more prognostic value for predicting treatment failure and survival outcomes (Moeller et al., 2010, Katahira-Suzuki et al., 2015, Oyama et al., 2020). However, only a few studies have investigated the ability of SUV_{max} derived at three months post-treatment to predict survival outcomes in HNSCC (Ito et al., 2014, Kim et al., 2016a, Mayo et al., 2019, Connor et al., 2021), and it is not clear if post-treatment SUV_{peak} could be a better predictor of survival compared to SUV_{max} in patients with HNSCC. Identifying whether these metrics have prognostic value in HNSCC may aid to stratify the risks for disease recurrence or residual and improve survival outcomes.

In a subtype of oropharyngeal squamous cell carcinoma (OPSCC), HPV-associated OPSCC has been identified as a different group of head and neck cancer with an independent prognosis for survival (Budach and Tinhofer, 2019). The distinct histological features and tumour FDG metabolic activity of HPV-related oropharyngeal cancer (Connor et al., 2021) affect post-treatment SUV_{max} (Chan et al., 2012, Vainshtein et al., 2014, Castelli et al., 2016, Connor et al., 2021). However, little is known about the influence of HPV status on SUV prognostic ability (Helsen et al., 2020a, Connor et al., 2021).

Furthermore, as described before, the calculation of SUV requires the use of selected body size metrics for SUV normalisation. However, most of the studies in the field of quantitative imaging have focused on SUV_{max} normalised to total body weight, while limited research has investigated SUV metrics normalised to LBM or BSA. Therefore, the full clinical influence of LBM and BSA on the diagnostic and prognostic value of SUVs is unclear. Investigating this would allow us to see if such methods improved the ability of a parameter to detect residual disease and predict an outcome and, hence, could be more accurate measures than SUVs normalised to total body weight.

1.16 Aims of the study

The overarching aim of this study was to explore the utility of three months post-treatment FDG PET/CT quantitative metrics in HNSCC.

1.16.1 General Aims

Aim 1) To explore the influence of using different body size metrics on SUV calculation in post-treatment setting.

Aim 2) To investigate the diagnostic performance of post-treatment PET/CT metrics in HNSCC.

Aim 3) To investigate the prognostic significance of post-treatment PET/CT metrics in HNSCC.

1.16.2 Specific Aims

1.16.2.1 Aims of Chapter 2:

1) To evaluate an evidence-base in the literature on the diagnostic performance and prognostic value of post-treatment PET/CT metrics derived from primary tumour and lymph node sites to discriminate residual disease and predict survival outcomes in patients with HNSCC.
2) To identify the most commonly used normalisation method for SUV calculation in HNSCC.

1.16.2.2 Aims of Chapter 3

1) To describe the methodology used in this thesis, including the source of patient data, imaging protocol, image analysis, and statistical analysis tests applied in each chapter.

1.16.2.3 Aims of Chapter 4

1) To investigate the association of HNSCC lesions SUV_{max}, $SUV_{peak,}$ MTV, and TLG, normalised by different body metrics, with patient body size measurements to identify the normalisation method that is least sensitive to these body factors when reporting post-treatment PET/CT parameters in patients with HNSCC.
2) To explore whether this association differed when SUV metrics were derived from the normal reference organ of the liver (liver SUVs) versus HNSCC SUVs acquired from the same patients.

1.16.2.4 Aims of Chapter 5

1) To evaluate the diagnostic performance of SUV_{max} and SUV_{peak} normalised by total body weight obtained separately from the primary tumour and involved lymph nodes to distinguish residual disease at three months post-treatment PET/CT imaging.
2) To assess whether using LBM and BSA in SUV calculation improved the diagnostic performance of the parameters.
3) To assess whether lesion-to-background ratios (relative metrics) improved the diagnostic performance of the parameters.
4) To present optimal thresholds for primary tumours and involved lymph node SUV_{max} and SUV_{peak} to predict residual lesions.
5) To analyse the repeatability of the post-treatment SUV measurements to ensure that the same image analysis method was used in determining all examined metrics, including lesions SUV_{max} and SUV_{peak}, as well as background regions SUV_{max} and SUV_{mean} (liver, cerebellum, and blood

pool), when acquired at two different time points by a single observer with an interval of at least five months.

1.16.2.5 Aims of Chapter 6

1) To determine if three-month post-treatment SUV_{max} and SUV_{peak} obtained from primary tumour (T) and lymph node (N) sites were related to:
    - 3-year progression-free survival (PFS)
    - 5-year overall survival (OS)
2) To determine whether the prognostic value of PET/CT parameters for survival outcomes is enhanced by employing other SUV normalisation methods such as LBM and BSA.
3) To investigate differences in prognostic significance when lesions were analysed according to the lesion's HPV status.

1.16.2.6 Aims of Chapter 7

1) To summarise the primary research findings and provide a general discussion of the research, including comments on the work's strengths and limitations, a discussion of clinical implications, and suggestions for future research.

For a visual representation of the thesis plan and aims, see Figure 1.2.

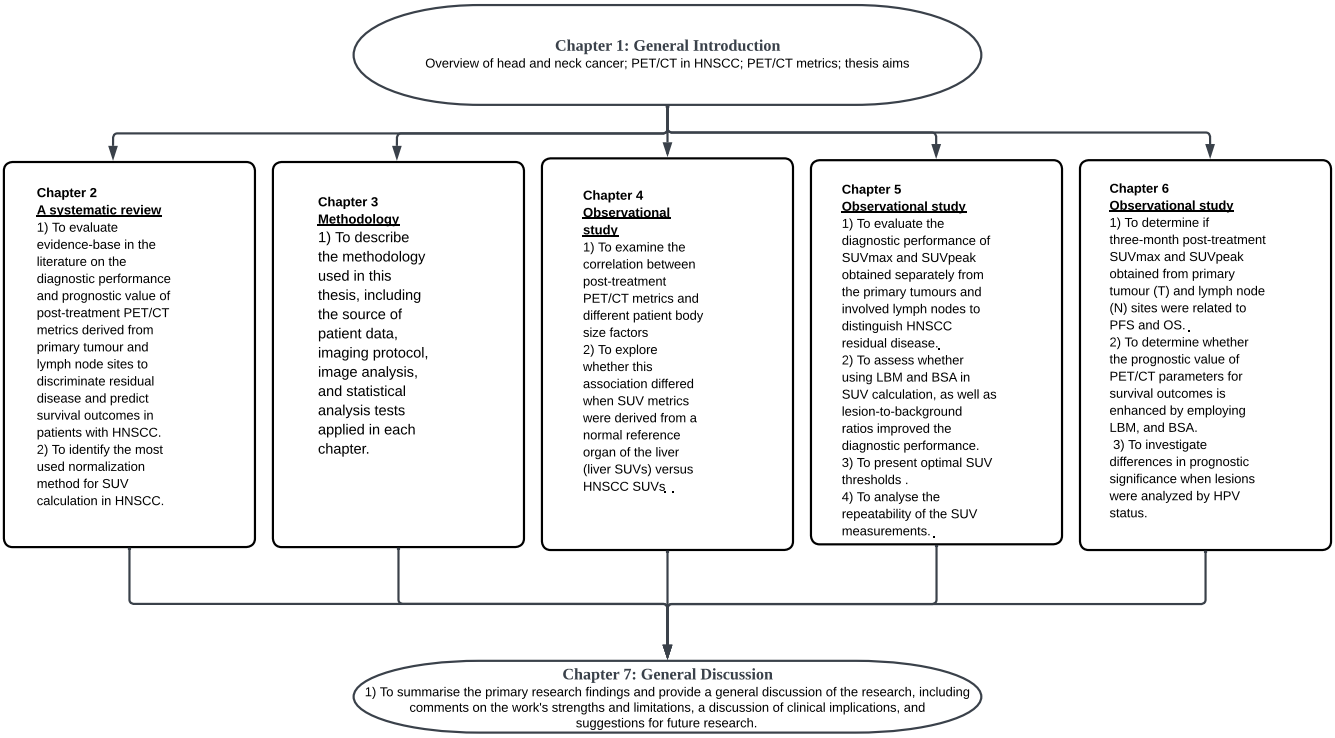


Figure 1.2: Visual representation of the thesis, including the aims of each chapter

CHAPTER 2

THE UTILITY OF POST-TREATMENT FDG PET/CT QUANTITATIVE METRICS TO DISCRIMINATE RESIDUAL DISEASE AND PREDICT SURVIVAL OUTCOMES IN PATIENTS WITH HNSCC: A SYSTEMATIC REVIEW

2. Chapter 2: The utility of post-treatment FDG PET/CT quantitative metrics in patients with HNSCC: A systematic review

This systematic review seeks to:

1) Evaluate the evidence base in the literature on the diagnostic performance and prognostic value of post-treatment PET/CT metrics derived from primary tumour and lymph node sites to differentiate residual disease and predict survival outcomes in patients with HNSCC.
2) Identify the most commonly used normalisation method for SUV calculation in HNSCC.

2.1 Materials and methods

2.1.1 Protocol and registration

This systematic review was registered at PROSPERO with registration number CRD42020193444 and was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines for reporting systematic reviews (Liberati et al., 2009).

2.1.2 Eligibility criteria

The PICO method—which stands for population, interventions, comparators, and outcomes—was used to set eligibility criteria.

Population

The target population of our analysis consisted of patients diagnosed with HNSCC—including cancers of the oral cavity, oropharynx, hypopharynx, and larynx—who received definitive treatment with radiotherapy with or without chemotherapy and then underwent follow-up FDG PET/CT imaging between two and four months after treatment. Patients who received palliative treatment, such as induction chemotherapy (IC), or who underwent curative-intent surgery, or neck dissection prior to PET/CT imaging were excluded. Also, a review and summary of previous literature was reported, which included author names, dates of publication, study design, timing of PET/CT imaging, and quantitative metrics thresholds. Studies that lacked quantitative metrics threshold data, as well as PET/CT studies that used non-FDG radiotracers were excluded. Only studies written in English were analysed. As a result, it is possible that some relevant studies have not been included. Other reasons for exclusion are listed in Figure 2.1.

Interventions

The main interventions of this systematic review were post-CRT PET/CT metrics, including SUV_{max}, SUV_{peak}, MTV, and TLG. Then, the values of the parameters obtained from three clinically relevant sites were analysed, including the following:

- Primary tumours and lymph nodes combined PET/CT quantitative metrics (Combined SUV)
- Primary tumours separate quantitative metrics
- Lymph nodes separate quantitative metrics

PET/CT parameters were defined according to reference guidelines (Boellaard et al., 2015) as previously described (see Chapter 1, Sections 1.10.1 and 1.10.2).

Comparators: None

Outcomes:

We sought to investigate the following:

1) Diagnostic performance of post-CRT FDG PET/CT primary tumour and lymph node HNSCC SUV_{max}, MTV, and TLG to discriminate residual disease
2) The prognostic value of post-CRT FDG PET/CT primary tumour and lymph node HNSCC SUV_{max}, MTV, and TLG to predict survival outcomes in terms of:
    - OS
    - PFS. PFS was defined as the time to disease progression and the time to locoregional and/or systematic relapse
3) The body size metric used in SUV normalisation in HNSCC

Study design

All types of observational studies were included. Studies were excluded if they examined pre-treatment metrics or did not provide sufficient data on post-treatment PET/CT metrics.

2.1.3 Information sources

Two reviewers (RA and AA-F) independently conducted a systematic search of the Medline, Embase, and Cochrane CENTRAL databases using comprehensive keywords and MeSH terms to identify relevant articles (Figure 2.1). Furthermore, the reference lists of the listed studies were examined to find additional relevant studies. The search was updated to identify newly published studies in various databases through February 8, 2023.

2.1.4 Data management

The initial search for identifying eligible studies was performed independently by the two reviewers (RA and AA-F). The second reviewer (AA-F) conducted a comprehensive search, screening tiles and abstracts of potential studies. Then, citation results for potential studies were uploaded into EndNote 20.0.1, and duplicates were removed. The full text of the relevant studies was then screened by the two reviewers. If both reviewers agreed, studies that matched the inclusion criteria were included. In the event of a disagreement, the viewpoint of a senior member of the review team (HM and PN) was considered.

2.1.5 Search strategy

Complete keywords and the strategy used for identifying eligible studies are provided in Appendix 1.

2.1.6 Study selection

The method used for the study selection is included in Section 2.1.4.

2.1.7 Data collection process

Two authors (RA and AA-F) extracted data separately, and any inconsistencies were resolved by consensus or by consulting senior members of the review team (HM and PN). Furthermore, contact was made with the original study's corresponding author if clarification was required or if the absence of data could compromise the study's eligibility. A follow-up email was sent if no response was received after two weeks. If no response was received within the next two months, the study was excluded.

2.1.8 Data items

Patients' demographic information, disease-related features, and survival outcome data were collected by RA and then reviewed by AA-F.

2.1.9 Risk of bias

To assess the quality of the included studies, the Quality in Prognostic Studies (QUIPS) tool was utilised to evaluate potential sources of bias (Appendix 2). This tool involves examining six study domains: study participation, study attrition, prognostic factor measurement, outcome measurement, study confounding, and statistical analysis and reporting. The Robvis tool was used to visualise the risk of bias assessments (McGuinness and Higgins, 2021).

2.2 Statistical analysis:

Data were analysed for primary tumour and lymph node combined quantitative metrics, primary tumour quantitative metrics, and lymph node quantitative metrics.

To assess the diagnostic performance of post-treatment quantitative metrics to discriminate residual disease, threshold values and their associated sensitivity, specificity, PPV, and NPV were extracted when available.

Similarly, to examine their prognostic relevance in predicting survival outcomes, we directly collected survival data consisting of the rate of survival and hazard ratios (HRs) with 95% confidence intervals (CIs) when available for all imaging parameters and each of the patient outcomes (OS, PFS, and DFS), recurrence-free survival (RFS), and disease-specific survival (DSS).

The quantitative metric threshold values were collected based on the results of the receiver operating characteristic (ROC) curve.

Several authors were contacted to obtain information on missing data. We received three raw databases (Nishimura et al., 2016, Mayo et al., 2019, Zhong et al., 2020), and one database (complete data) was subsequently locally analysed (Nishimura et al., 2016).

For this database, we used the following statistical tests to analyse the data:

- ROC curve to determine optimal thresholds and estimate their related sensitivity and specificity (Kamarudin et al., 2017)

- Survival analysis by Cox-regression to identify HRs and their CIs of OS and PFS for each examined imaging parameter (combined SUV_{max}, primary tumour SUV_{max}, and lymph node SUV_{max})

Combined SUV_{max} was based on the maximum radioactivity concentration obtained from one lesion (primary tumour or involved nodal lesion) (Kim et al., 2016a, Connor et al., 2021), or (primary tumour, involved nodal lesion, or metastasis) (Ito et al., 2014).

Furthermore, an attempt was made to perform a meta-analysis. However, due to the heterogeneity of the studies and the unavailability of complete data, overall median sensitivity and specificity values were calculated and presented to evaluate the quantitative parameters' overall diagnostic performance in discriminating residual disease in post-treatment PET/CT imaging. Forest plots were also developed to provide graphical representations of pooled estimates of prognostic value. Summary pooled estimates and their CIs were calculated based on random effects models (Bonomo et al., 2018). The metan function in Stata/SE:16.1 was used to create the forest plots.

2.3 Results

2.3.1 Data collection and analysis

Following an initial search, the titles and abstracts of potentially eligible studies were independently examined by two authors (RA and AA-F). Then, the full text of 252 articles was retrieved for further assessment. Based on the inclusion criteria, 238 studies were excluded, resulting in around 14 eligible studies. The reasons for exclusions are shown in Figure 2.1. The search was updated to include newly released articles. Two additional studies were included. Therefore, we analysed 16 studies.

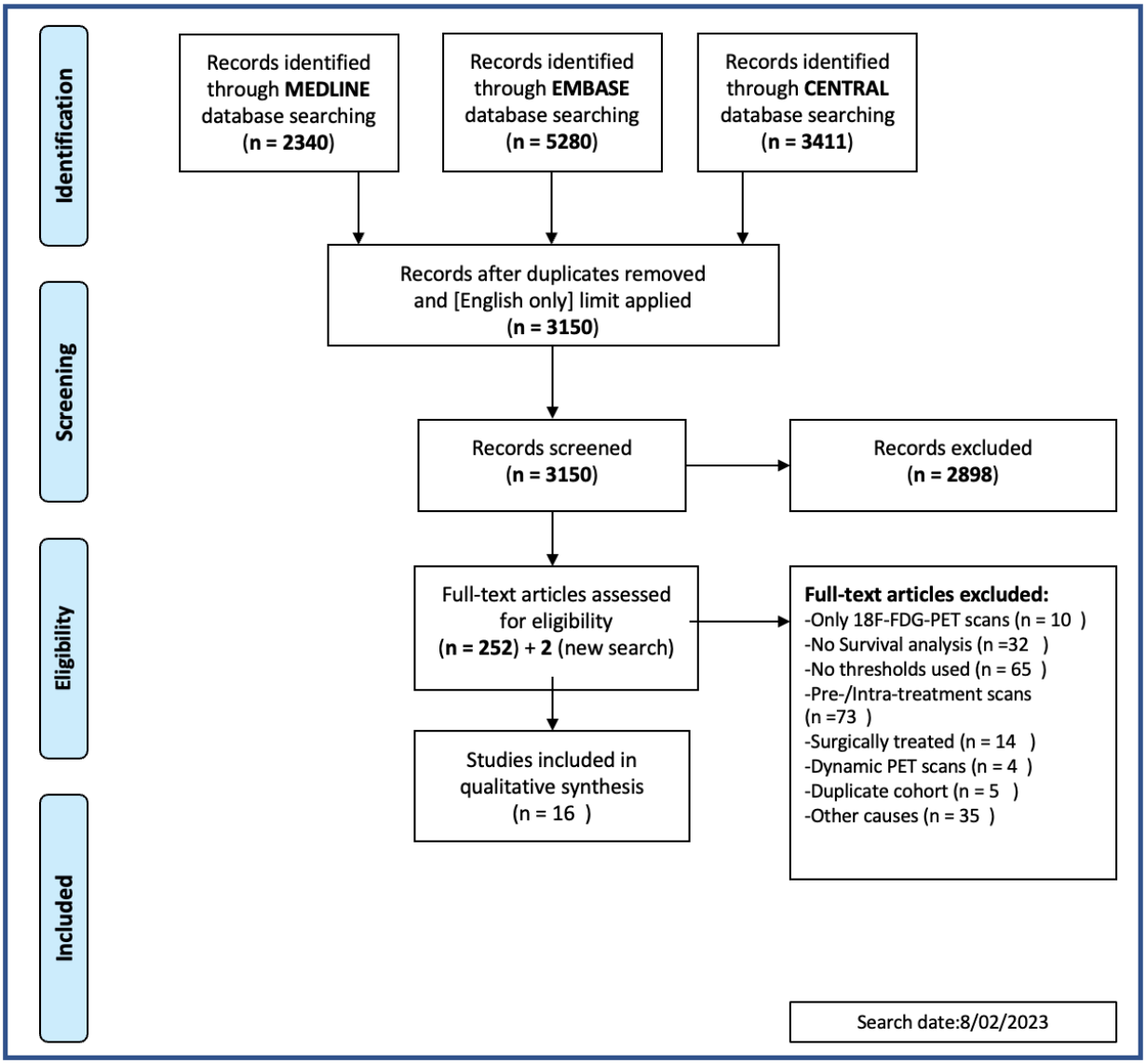


Figure 2.1: PRISMA Flow Diagram for included studies.

*The same patient cohort was used in Moeller et al. (2009) and Moeller et al. (2010). Therefore, we considered them a single study.

2.3.2 Patient characteristics and disease-related features

The overall population consisted of 1,513 patients diagnosed with HNSCC. Most patients were men (81.4%), with ages ranging from 26 to 88 years (Table 2.1). The oropharynx was the most common tumour site examined, followed by the hypopharynx, while the nasal cavity/paranasal sinus (PNS) was the tumour site least evaluated across the included studies (Table 2.2).

Table 2.1: Overview of the design of included studies and patient characteristics.

Study Number	First author	Year	Country	Study Design	No. of patients	Age (yrs)	Age range (yrs)	Male, *n* (%)
1	Oyama et al.	2020	Japan	Prospective	37	Median (65)	26-79	31 (83.8)
2	Helsen et al.	2020	Belgium	Prospective	124	Median (59)	43-79	93 (75.0)
3	Dejaco et al.	2020	Austria	Retrospective	33	Mean (65)	32-81	29 (87.9)
4	Fatehi et al.	2019	India	Retrospective	75	Mean (57)	15-81	66 (88.0)
5	Awan et al.	2017	USA	Retrospective	108	NR	50-70	88 (81.5)
6	Nelissen et al.	2017	UK	Retrospective	206	Median (59)	35- 88	157 (76.2)
7	Riaz et al.	2017	Pakistan	Retrospective	93	Mean (48.8)	NR	61 (65.6)
8	Kim et al.	2016	Korea	Retrospective	78	Median (62)	24–79	63 (80.8)
9	Nishimura et al.	2016	Japan	Retrospective	235	Median (64)	16-92	197 (83.8)
10	Sjövall et al.	2016	Sweden	Retrospective	105	Median (61)	34–89	78 (74.0))
11	Katahira-Suzuki et al.	2015	Japan	Retrospective	70	Median (65)	21–91	61 (87.1)
12	Sagardoy et al.	2015	France	Retrospective	47	Mean (56.4)	NR	40 (85.1)
13	Ito et al.	2014	Japan	Retrospective	36	Mean (62.0)	NR	31 (86.1)
14	Vainshtein et al.	2014	USA	Retrospective	101	Median (55)	34–76	93 (92.0)
15	Chan et al.	2012	USA	Retrospective	67	<60, ≥60	–	60 (90.0)
16	Moeller et al.	2009/10	USA	Prospective	98	Mean (58)	36–79	83 (85.0)

Table 2.2: An overview of tumour sites analysed.

First author	Oropharynx, n (%)	Larynx, n (%)	Hypopharynx, n (%)	Oral cavity, n (%)	Nasopharynx, n (%)	Nasal cavity, n (%)	Other
Oyama et al.	0	0	0	37	0	0	0
Helsen et al.	69 (55.6)	20 (16.1)	11 (8.9)	8 (6.5)	8 (6.5)	0	8 (6.3)
Dejaco et al.	15 (45.5)	2 (6.1)	4 (12.1)	5 (15.2)	1 (3.0)	0	6 (18.2)
Fatehi et al.	16 (15.8)	24 (27.6)	24 (27.6)	0	7 (9.3)	0	5 (6.7)
Awan et al.	79 (73.1)	10 (9.3)	11 (10.2)	2 (1.9)	0	0	6 (5.6)
Nelissen et al.	161 (78.2)	23 (11.2)	20 (9.7)	0	0	0	2 (1.0)
Riaz et al.	10 (10.8)	18 (19.4)	0	26 (28.0)	32 (34.4)	1 (1.1)	6 (6.5)
Kim et al.	47 (60.3)	3 (3.9)	19 (24.4)	5 (6.4)	0	1 (1.3)	3 (3.9)
Nishimura et al	73 (29.6)	29 (11.7)	89 (36.0)	11 (4.5)	26 (10.5)	4 (1.6)	15 (6.1)
Sjövall et al.	6 (6.0)	0	5 (4.0)	96	0	0	0
Katahira-Suzuki et al.	25 (35.7)	0	36 (51.4)	0	9 (12.9)	0	0
Sagardoy et al.	NR	NR	NR	NR	NR	NR	NR
Ito et al.	2 (5.6)	6 (16.7)	13 (36.1)	3 (8.3)	0	12 (33.3)	0
Vainshtein et al.	101 (100.0)	0	0	0	0	0	o
Chan et al.	63 (94.0)	0	0	0	0	0	3 (4.5)
Moeller et al.	77 (79.0)	12 (12.0)	9 (9.0)	0	0	0	0

2.3.3 Descriptive analysis

The majority of the included studies were retrospective in design (n=13/16). Most studies (11/16) had evaluated quantitative measurements acquired from post-treatment PET/CT done on an average of three months after CRT, whereas five studies (5/16) had undertaken PET/CT imaging prior to three months after CRT (Tables 2.3 and 2.4). The average (mean or median) follow-up time was given in all studies, which ranged from 6 to 52 months. In addition, SUV_{max} was the only evaluated parameter. A study on MTV was excluded due to incomplete data (Murphy et al., 2011). None of the studies involved TLG parameter. Various studies have analysed the diagnostic and prognostic significance of SUV_{max} derived from primary tumour and lymph node combined SUV_{max} (3/16), primary tumour only (2/16), involved lymph node only (6/16), and both primary tumour and lymph node separate SUV_{max} (5/16). Primary tumour and lymph node cutoff values also varied (Tables 2.3 and 2.4).

The results of the QUIPS tool showed an overall moderate risk of bias in the domain of study participation. This rating was due to the heterogeneity of the included head and neck cancer subtypes. Some studies evaluated combined HNSCC, while others examined specific subsites, such as oral cancer. An overall moderate risk of bias was also seen in the domains of the study attrition and the prognostic factor measurements. This was likely due to the small sample sizes in some of the included studies or the lack of complete data on participants who were lost to follow-up. Similarly, a moderate risk of bias was reported in the domain of outcome measurement. This was due to differences in the definitions or durations examined for the survival endpoints. Some studies analysed PFS, while others examined DSS or DFS. A moderate

risk of bias was observed in the domain of confounding factors. The influence of the duration of time between the end of treatment and post-treatment PET/CT imaging was one of the possible confounding factors. Some studies used FDG PET/CT imaging three months after treatment, while others used PET/CT imaging earlier. Finally, the QUIPS tool showed an overall low risk of bias in the dominance of statistical analysis and reporting. Although not all of the included studies published estimated HRs, many did report survival rates. ROC was also used to determine optimal thresholds, and this method is widely accepted in assessing diagnostic studies (Kamarudin et al., 2017).

A detailed quality assessment is provided in Figure 2.2.

Table 2.3: Selected published studies reporting the diagnostic value of post-treatment metrics in HNSCC.

Study	Year	Study design	Site	No. of patients	Timing of PET/CT	SUV metrics	Determination of thresholds	Site	Thresholds	Sensitivity	Specificity	NPV	PPV
Vainshtein et al.	2014	R	HPV-related SCC	101	(median) 13.4 weeks	SUVmax	ROC curve (taken from Moller et al.(2009)	T	6.5	33.0% (95% CI, 1–91%)	91.0% (95% CI, 83–96%)	98.0% (95% CI, 92–100%)	10.0% (95% CI, 0–45%)
								N	2.8	63.0% (95% CI, 24–91%)	70.0% (95% CI, 59–79%)	95.0% (95% CI, 87–99%)	16.0% (95% CI, 5–33%)
Sagardoy et al.	2015	R	HNSCC	47	(median) 3.3 months	SUVmax	ROC curve	T	3.7	73.3%	87.1%	–	–
								N	4.0	85.7%	100.0%	–	–
Nelissen et al.	2017	R	HNSCC	206	(median) 3 months	SUVmax	Predefined	T	4.0	–	–	–	53.0%
								N	4.0	–	–	–	93.0%
								T	6.0	–	–	–	65.0%
								N	6.0	–	–	–	100.0%
								T	8.0	–	–	–	92.0%
								N	8.0	–	–	–	100.0%
Moller et al.	2009	P	HNSCC	98	2 to 2.8 months	SUVmax	ROC curve	T	6.5	70.0%	93.7%	96.1%	58.3%
								N	2.8	75.0%	76.1%	96.2%	27.3%
*Nishimura et al.	2016	R	HNSCC	235	(mean± SD) 1.5± 0.6 months	SUVmax	ROC curve	Combined	3.9	52.2%	83.2%	–	–
								T	3.9	40.9%	82.0%	–	–
								N	2.7	47.9%	81.2%	–	–
kim et al.	2016	R	HNSCC	78	(mean± SD) 3.2 ± 1.1 months	SUVmax	ROC curve	Combined	4.4	90.0%	83.8%	98.3%	45.0%
**Chan et al.	2012	R	HPV-related SCC	67	(median) 2.9 months	SUVmax	NR	N	2	100.0%	71.0%	100.0%	21.7%
								N	2.5	60.0%	72.6%	95.7%	15.0%
Fatehi et al.	2019	R	HNSCC	75	10 to 14 weeks	SUVmax	ROC curve	N	4.62	73.5%	92.3%	–	–
Helsen et al.	2020	R	HNSCC	124	12 weeks	SUVmax	ROC curve	N	2.2	79.7% (95% CI, 59.2–100%)	80.8% (95% CI, 73.0–88.6%)	96.6% (95% CI, 92.7–100%)	37.0% (95% CI, 20.1–53.9%)
Dejaco et al.	2020	R	HNSCC	41	(median) 10 weeks	SUVmax	ROC curve	N	3.75	69.0%	100.0%	–	–
Sjovall et al.	2016	R	HNSCC	105	(median) 6 weeks	SUVmax	ROC curve	N	2.0	82.6 % (95% CI, 60.5–94.3%)	61.0% (95% CI, 49.5–71.4%)	92.6% (95% CI, 81.2–97.6%)	37.3% (95% CI, 24.5–51.9%)

Table 2.3 summarises all published studies on the diagnostic performance of post-treatment primary tumour and nodal combined SUV_{max}, primary tumour separate SUV_{max}, and involved lymph node separate SUV_{max} for discriminating HNSCC residual or recurrent lesions.

Abbreviations: T, tumour; N, lymph node; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating characteristics; SUV_{max}, maximum standardised uptake value; NR, not reported; SD, standard deviation; R, retrospective; P, prospective; HNSCC, head and neck squamous cell carcinoma; HPV, human papilloma virus.

- Primary tumour and lymph node combined SUV_{max} is highlighted in orange.
- Lymph node SUV_{max} is highlighted in grey.
- *This study was locally analysed.
- **This study examined SUV_{max} normalised to LBM.

Table 2.4: Selected published studies reporting the prognostic value of post-treatment metrics in HNSC.

Author	Year	Study design	Site	No. of patients	Timing of PET/CT	SUV metrics	Deremination of thresholds	Site	Threshold	Survival endpoints	Outcomes
Kim et al.	2016	R	HNSCC	78	(mean± SD) 3.2 ± 1.1 months	SUVmax	ROC curve	Combined	4.4	3 years OS	SUVmax ≥ 4.4: 56.9% (95% CI, 30.1–76.3 %) SUVmax < 4.4: 87.7 % (95% CI, 75.9–94.0 %) HR = 4.25 (95% CI:1.54–11.74); logrank P = 0.002
										3 years PFS	SUVmax ≥4.4: 42.9% (95 %CI, 20.4–63.6 %) SUVmax <4.4: 81.1 % (95 %CI, 67.5–89.5 %) HR = 4.79 (95 % CI: 2.02–11.32); logrank P < 0.001
Ito et al.	2014	R	HNSCC	36	8-12 weeks	SUVmax	ROC curve	Combined	6.1	2 years OS	Post-treatment SUVmax had a prognostic value to predict survival. SUVmax >6.1: mean OS of 12.1 months (95% CI, 6.3-18.0 months) SUVmax <6.1: mean OS of 44.6 months (95% CI, 39.9-49.3 months)
*Nishimura et al.	2016	R	HNSCC	235	(mean ± SD)1.5± 0.6 months	SUVmax	ROC curve	Combined	3.9	OS	HR= 3.05 (95 %CI, 1.98-4.69)
								T	3.9	OS	HR= 2.40 (95 %CI, 1.52-3.80)
								N	2.5	OS	HR= 1.72 (95 %CI, 1.10-2.69)
								Combined	3.9	DFS	HR= 3.72 (95 %CI, 2.42-5.71)
								T	3.9	DFS	HR= 2.63 (95 %CI, 1.68-4.12)
								N	2.5	DFS	HR= 2.49 (95 %CI, 1.62-3.84)
Oyama et al.	2020	P	Oral cancer	37	2 months	SUVmax	ROC curve	T	4.4	5 year OS	SUVmax <4.4: 85.2% (95% CI, 60.6-95.0%) SUVmax >4.4: 38.6% (95% CI, 13.4-63.6%); log-rank test (P=0.006)
Katahira-Suzuki et al.	2015	R	pharyngeal cancer	70	(median) 7 weeks	SUVmax	ROC curve	T	5	3 years OS	SUVmax<5 = 85.3% SUVmax ≥ 5 = 16.9%; p= <0.0001 HR= 4.24 (95 % CI,1.07–16.75); P= 0.039.
								N	2.45	3 years OS	SUVmax < 2.45 = 84.3% SUVmax ≥ 2.45= 47.1%; p=0.017 HR= 1.01 (95 % CI, 0.26–3.88) P= 0.994
								T	5	3 years MFS	SUVmax < 5= 87.0% SUVmax ≥ 5= 78.8%; P= 0.137
								N	2.45	3 years MFS	SUVmax< 2.45= 89.6% SUVmax ≥ 2.45= 66.7%; P=0.016
Riaz et al.	2017	R	HNC	93	4 to 6 months	SUVmax	NR	T	5.0	3 year DFS	SUVmax <5 = 62% SUVmax 5–10 = 42% SUVmax >10 = 6%
Moeller et al.	2010	P	HNSCC (pharyngolaryngeal)	98	2 to 2.8 months	SUVmax	ROC curve	T	6	3 year DSS	Survivorship did stratify by primary SUVmax (log-rank p > 0.005)
								N	3.4	3 year DSS	Survivorship did not stratify by nodal SUVmax
Sjövall et al.	2016	R	HNSCC	105	(median) 6 weeks	SUVmax	ROC curve	N	2	DSS and OS	SUVmax was not significantly predictor of OS and DSS
Awan et al.	2017	R	HNSCC	108	(median) 12.9 weeks	SUVmax	NR	N	2.5	3 year RFS	P16+: 3-year RFS: 89.7% with SUVmax<2.5 and 50% SUVmax >2.5 P16-: 3-year RFS: 72.0% with SUVmax<2.5 and 21.4% SUVmax >2.5

Table 2.4 summarises all published studies that reported the prognostic value of post-treatment primary tumour and nodal combined SUV_{max}, primary tumour separate SUV_{max}, and involved lymph node separate SUV_{max} for predicting survival outcomes.

Abbreviations: T, tumour; N, lymph node; NPV; ROC, receiver operating characteristics; SUV_{max}, maximum standardised uptake value; HR, hazard ratio; SD, standard deviation; R, retrospective; P, prospective; CI, confidence interval; HNSCC, head and neck squamous cell carcinoma; HPV, human papilloma virus; OS, overall survival; PFS, progression-free survival; DFS, disease-free survival; MFS, metastatic-free survival; DSS, disease-specific survival; RFS, recurrence-free survival.

- Primary tumour and lymph node combined SUV_{max} is highlighted in orange.
- Lymph node SUV_{max} is highlighted in grey.
- *This study was locally analysed.

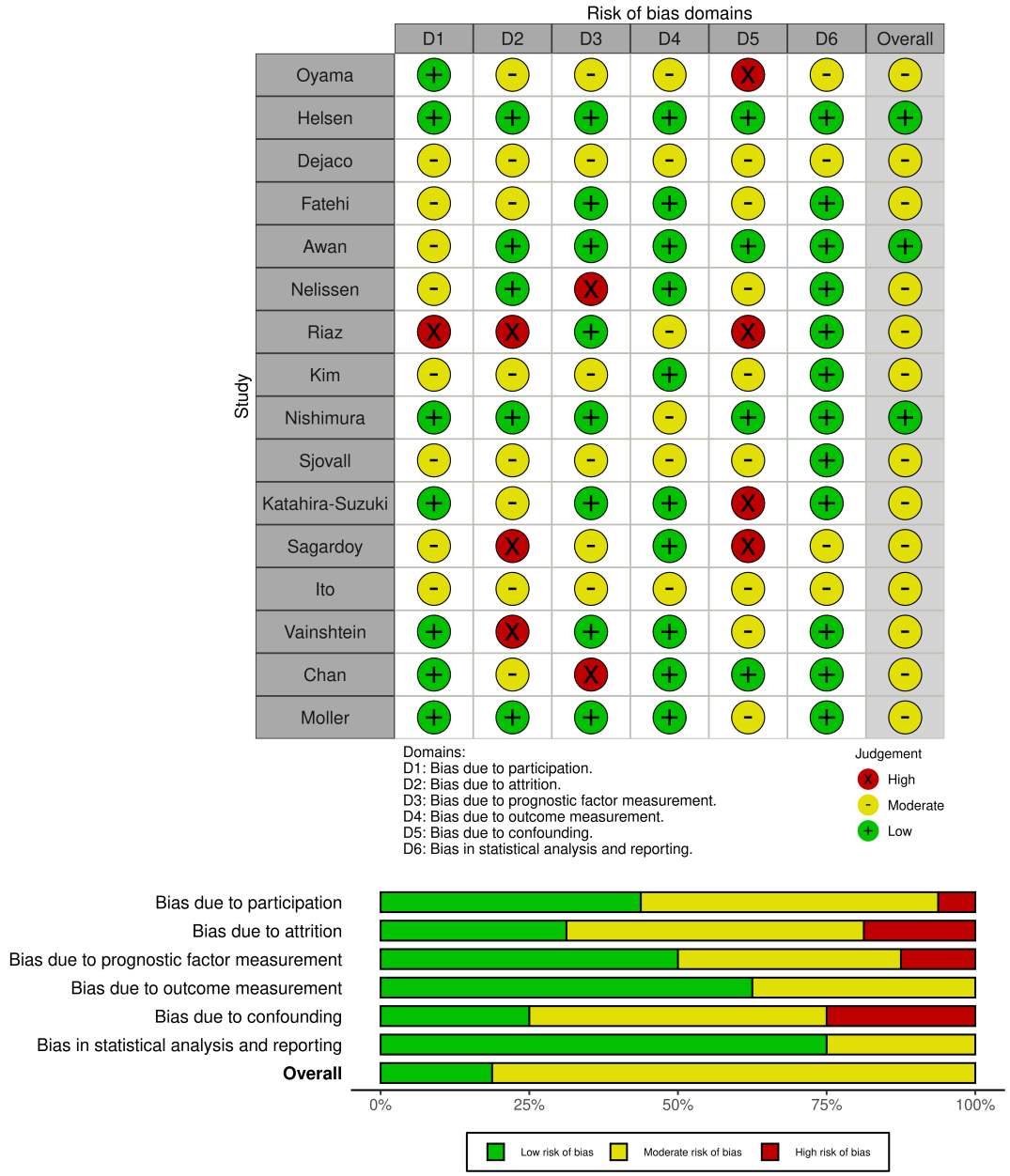


Figure 2.2: Quality in Prognostic Studies (QUIPS).
The risk of bias and applicability concerns were rated as high (red), moderate (yellow), and low (green).

2.3.4 Diagnostic performance of FDG PET/CT parameters for the detection of residual or recurrent disease

2.3.4.1 The diagnostic performance of combined SUV_{max}

One study reported the diagnostic performance of post-treatment $SUV_{max,}$ which was determined from the lesion with the highest FDG radioactivity concentration calculated from the primary tumour or regional lymph node, in predicting immediate locoregional and/or systemic failure (Kim et al., 2016a). Another study has been locally analysed to assess the diagnostic performance of primary tumour and lymph node combined SUV_{max} for distinguishing residual or recurrent post-treatment HNSCC (Nishimura et al., 2016). This provided a total of two studies with 313 patients involved in this sub-analysis. Two thresholds were reported for the combined SUV_{max}: 3.9 (sensitivity of 52.20% and specificity of 83.20%) and 4.4 (sensitivity of 90.00% and specificity of 83.80%) (Kim et al., 2016a, Nishimura et al., 2016) (Table 2.5).

The overall pooled estimate of post-treatment HNSCC primary tumour and lymph node combined SUV_{max} showed an SUV_{max} median of 4.15, with an overall sensitivity and specificity of about 71.10% and 83.8%, respectively. This could suggest that when a clinical decision is based on examining the combined SUV_{max} value from either the primary tumour or involved lymph node sites, an overall SUV_{max} ≥4.15 could be indicative of persistent disease or likelihood of recurrence after the completion of CRT with an overall specificity (true negatives) of around 83.80% (Table 2.5).

Table 2.5: Combined SUV_{max} with associated sensitivities and specificities across studies

Study	**Cutoff**	**Sensitivity**	**Specificity**
Kim et al.	4.40	90.00%	83.80%
Nishimura et al.	3.90	52.20%	83.20%
Overall (median)	4.15	71.10%	83.80%

2.3.4.2 The diagnostic performance of primary site SUV_{max}

Three studies reported the diagnostic performance of post-treatment SUV_{max} for predicting treatment response at primary tumour sites (Moeller et al., 2009, Vainshtein et al., 2014, Sagardoy et al., 2016). Another study was locally analysed to assess the diagnostic performance of primary tumour SUV_{max} for distinguishing residual or recurrent post-treatment HNSCC (Nishimura et al., 2016). This provided a total of four studies, with 481 patients included in this sub-analysis. Post-treatment primary tumour SUV_{max} thresholds ranged from 3.70 to 6.50 and their associated sensitivity differed from 33.00% to 73.30%, while the specificity varied from 82.00% to 93.70% (Table 2.6).

The overall pooled estimate of post-treatment HNSCC primary tumours showed an SUV_{max} median of 5.2, with an overall sensitivity and specificity of about 55.45% and 89.05%, respectively, suggesting that a post-treatment primary tumour SUV_{max} ≥5.2 could be indicative of persistent disease or likelihood of recurrence after the completion of CRT with an overall specificity (true negatives) of around 89.05% (Table 2.6).

Table 2.6: The diagnostic performance of post-CRT primary tumour SUV_{max} with associated sensitivity and specificity for discriminating HNSCC residual and recurrent lesion at primary sites for these studies

Study	**Cutoff**	**Sensitivity**	**Specificity**
Nishimura et al.	3.90	40.90%	82.00%
Sagardoy et al.	3.70	73.30%	87.10%
Vainshtein et al	6.50	33.00%	91.00%
Moeller et al.	6.50	70.00%	93.70%
Overall (median)	5.2	55.45%	89.05%

2.3.4.3 The diagnostic performance of neck node SUV_{max}

Six studies directly evaluated the diagnostic performance of post-treatment SUV_{max} determined from lymph node sites for predicting treatment response (Moeller et al., 2009, Vainshtein et al., 2014, Sagardoy et al., 2016, Sjovall et al., 2016, Fatehi et al., 2019b, Dejaco et al., 2020). Another study was locally analysed to assess the diagnostic performance of lymph node SUV_{max} for distinguishing residual or recurrent post-treatment HNSCC (Nishimura et al., 2016). This provided a total of seven studies, with 702 patients included in this sub-analysis. Post-treatment nodal SUV_{max} thresholds ranged from 2.00 to 4.62, and their associated sensitivity differed from 47.90% to 85.70%, while the specificity varied from 61.00% to 100.00%.

The overall pooled estimate of post-treatment HNSCC nodal lesions showed an SUV_{max} median of 2.80, which yielded sensitivity and specificity of about 73.50% and 81.20%, respectively, suggesting that a nodal $SUV_{max} \geq 2.8$ could be indicative of persistent disease or the likelihood of recurrence after the completion of CRT, with an overall specificity (true negatives) of around 81.20% (Table 2.7).

Table 2.7: The diagnostic performance of post-CRT lymph node SUV_{max} with associated sensitivity and specificity for discriminating HNSCC residual and recurrent lesions at nodal sites for these studies

Study	**Cutoff**	**Sensitivity**	**Specificity**
Dejaco et al.	3.75	69.00%	100.00%
Fatehi et al.	4.62	73.50%	92.30%
Sjovall et al.	2.00	82.60%	61.00%
Nishimura et al.	2.70	47.90%	81.20%
Sagardoy et al.	4.00	85.70%	100.00%
Vainshtein et al.	2.80	63.00%	70.00%
Moeller et al.	2.80	75.00%	76.10%
Overall (median)	2.80	73.50%	81.20%

2.3.5 Prognostic significance of post-treatment FDG PET/CT parameters with survival outcomes

2.3.5.1 Primary tumour and lymph node combined SUV_{max} and OS

Two studies reported the prognostic value of post-treatment combined SUV_{max} and OS (Ito et al., 2014, Kim et al., 2016a). Another study was locally analysed to assess the prognostic value of combined SUV_{max} for predicting OS (Nishimura et al., 2016). The time between treatment completion and post-treatment PET/CT varied from around 1.5 months to 3.2 months. Patient survival was analysed based on different combined SUV_{max} values, which ranged from 3.9 to 6.1.

In all included studies, post-treatment primary tumour and nodal combined HNSCC SUV_{max} were found to be correlated with OS. For example, Kim et al. (2016) found that an overall cutoff value of 4.4 obtained from primary tumour and lymph node HNSCC combined SUV_{max} was a discriminator between favourable and poorer OS. The three-year OS with a cutoff ≥4.4 was reported to be 56.9%, compared to 87.7% in the other group. The risk of death within three years post-treatment was increased by a factor of 4.25 (95% CI: 1.54–11.74) in patients with combined SUV_{max} ≥4.4.

Consistent with this, Ito et al. (2014) found that, at a higher cutoff value of 6.1, the two-year OS significantly differed, with an average of 12.1 months and 44.6 months, in the group of patients with combined SUV_{max} ≥ and <6.1, respectively. Using a lower combined SUV_{max} cutoff value, Nishimura et al. (2016) found that a combined SUV_{max} of 3.9 was optimal for comparing clinical outcomes. Based on a local analysis, the risk of death within three years

post-treatment was increased by a factor of 3.05 in the group of patients with higher combined SUV_{max}.

Overall, all included studies support the notion that the post-treatment combined SUV_{max} of the primary tumour and lymph nodes can predict OS; however, the optimal thresholds proposed by the studies varied considerably. Potential reasons for this are provided in the Discussion section.

In the meta-analysis, one study provided an HR for the association between post-treatment primary tumour and lymph node combined SUV_{max} and OS (Kim et al., 2016a). Another study was further locally analysed to calculate the estimated HRs for OS (Nishimura et al., 2016). This provided a total of two studies with 313 HNSCC patients, for which there were two HR estimates for OS. Using the random effect model (Bonomo et al., 2018), we found no statistically significant association between combined SUV_{max} and OS, with an overall estimated pooled HR of 1.63 (95% CI: -0.25–3.50, *P*=0.253) (Figure 2.3).

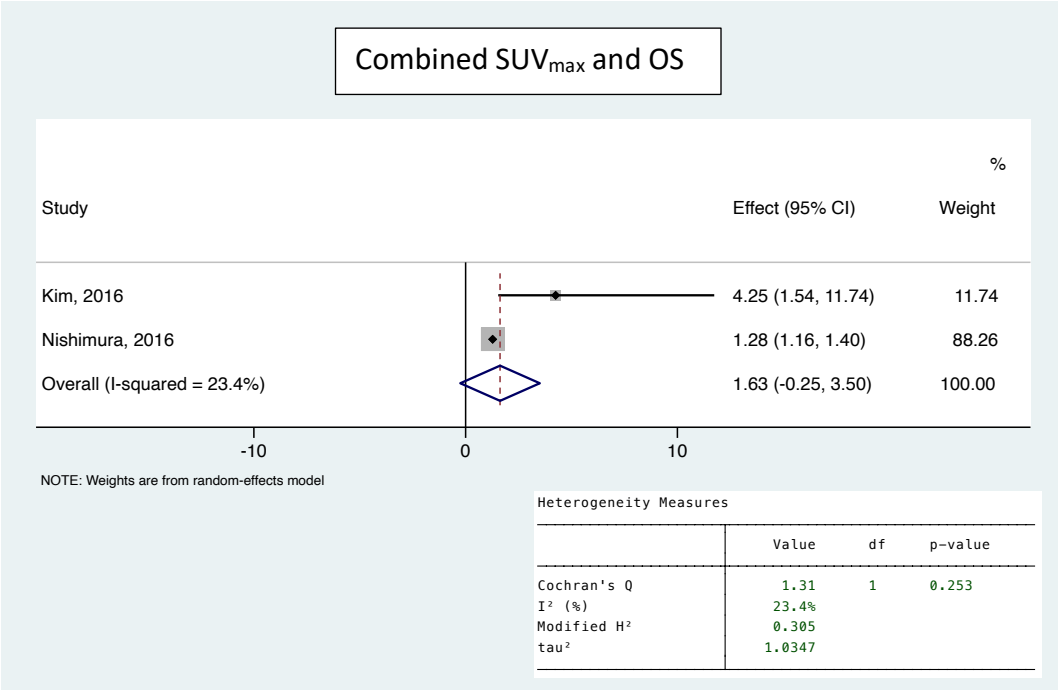


Figure 2.3: Forest plot shows visual analysis of OS HRs associated with primary tumour and lymph node combined SUV_{max}

2.3.5.2 Primary tumour site SUV_{max} and OS

Two studies reported the prognostic value of post-treatment primary tumour SUV_{max} and OS (Katahira-Suzuki et al., 2015, Oyama et al., 2020). Another study was locally analysed to assess the prognostic value of primary tumour SUV_{max} for predicting OS (Nishimura et al., 2016). These studies explored various subtypes of head and neck cancer, including combined HNSCC (Nishimura et al., 2016) , oral cancer (Oyama et al., 2020), and pharyngeal cancer (Katahira-Suzuki et al., 2015). The time difference between treatment completion and post-treatment PET/CT varied from around 1.5 months to 2 months. Patient survival was analysed based on different primary tumour SUV_{max} values, which ranged from 3.9 to 5.0.

Post-treatment primary tumour HNSCC SUV_{max} was found to be correlated with OS in combined HNSCC and in specific head and neck cancer subtypes. In 37 patients with oral cancer who underwent two months post-treatment PET/CT imaging, Oyama et al. (2020) suggested an optimal primary site SUV_{max} cutoff of 4.4. At this cutoff, the five-year OS for post SUV_{max} <4.4 was 85.2% versus 38.6% for SUV_{max} >4.4.

Similarly, in 70 patients with pharyngeal cancer, Katahira-Suzuki et al. (2015) found that patients with a slightly higher cutoff for primary tumour SUV_{max} >5.0 had an unfavourable five-year OS (16.9% vs. 85.3%). In combined HNSCC, the risk of death increased by a factor of 2.40 in patients with a primary tumour SUV_{max} of 4.4 compared to the other group.

In a meta-analysis, one study provided direct HR for the association between post-treatment primary tumour SUV_{max} and OS (Katahira-Suzuki et al., 2015). Another study was

further locally analysed to calculate the estimated HRs for OS (Nishimura et al., 2016). This provided a total of two studies with 305 HNSCC patients, for which there were two HR estimates for OS. Using the random effect model, we found no statistically significant association between primary tumour SUV_{max} and OS, with an overall HR pooled estimate of 1.22 (95% CI: 1.07–1.37, *P*=0.451) (Figure 2.4).

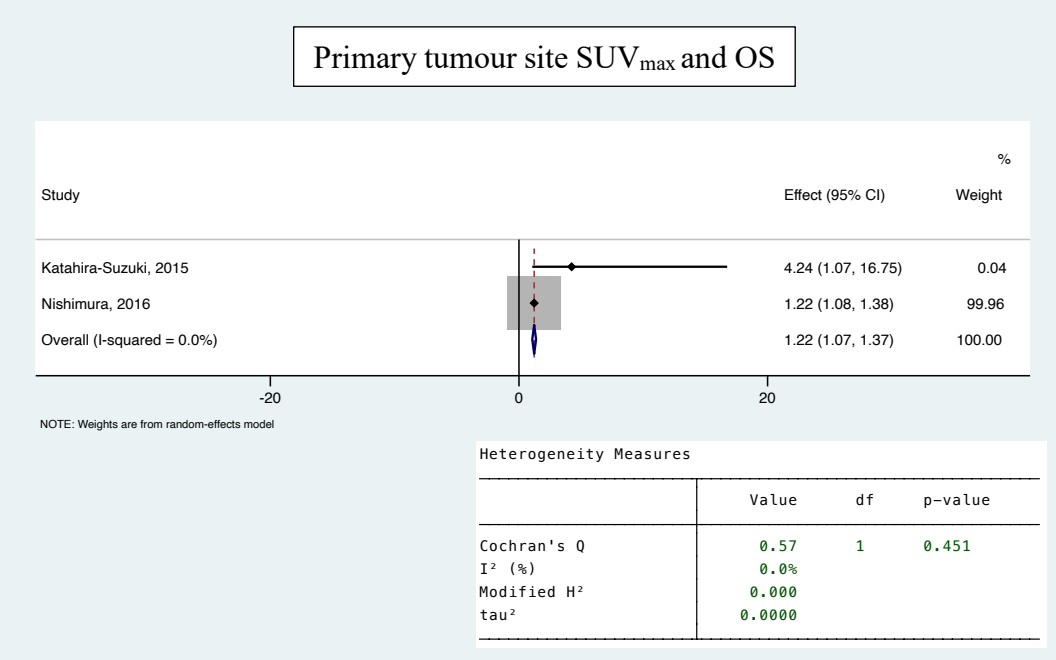


Figure 2.4: Forest plot shows visual analysis of OS HRs associated with primary tumour SUV_{max}

2.3.5.3 Lymph node site SUV_{max} and OS

Only one study reported the prognostic value of post-treatment nodal SUV_{max} and OS (Katahira-Suzuki et al., 2015). Another study was locally analysed to assess the prognostic value of lymph node SUV_{max} for predicting OS (Nishimura et al., 2016). Two types of head and neck cancer were examined in these studies: combined HNSCC (Nishimura et al., 2016) and pharyngeal SCC (Katahira-Suzuki et al., 2015). The time between treatment completion and post-treatment PET/CT varied from around seven weeks to two months. Patient survival was analysed based on two similar nodal SUV_{max} values of 2.45 and 2.50.

In 70 patients with pharyngeal SCC who underwent post-treatment PET/CT imaging at a median of seven weeks after the completion of treatment, Katahira-Suzuki et al. (2015) found that post-treatment nodal SUV_{max} was not associated with OS. However, in combined HNSCC, the risk of death increased by a factor of 1.72 (95% CI: 1.10–2.69) in patients with a nodal SUV_{max} >2.5.

In a meta-analysis of 307 patients (Katahira-Suzuki et al., 2015, Nishimura et al., 2016), we found no statistically significant association between nodal tumour SUV_{max} and OS, with an overall HR pooled estimate of 1.26 (95% CI: 1.12–1.40, *P*=0.781) (Figure 2.5).

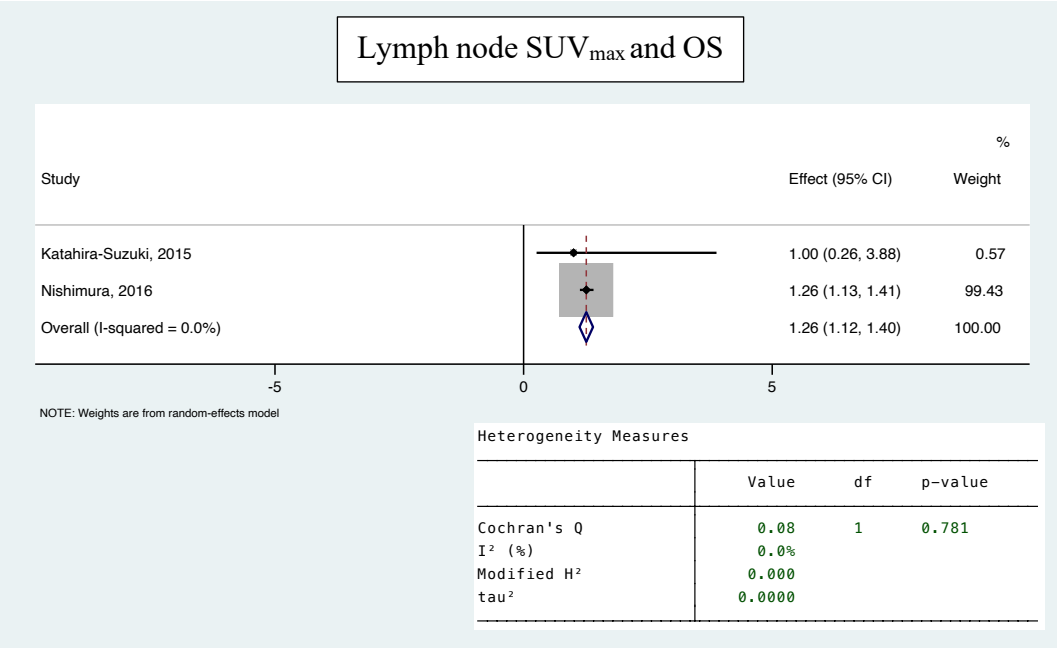


Figure 2.5: Forest plot shows visual analysis of OS HRs associated with lymph node SUV_{max}

2.3.5.4 Primary tumour and lymph node combined SUV_{max} and PFS

Only one study reported the prognostic value of post-treatment combined SUV_{max} and PFS (Kim et al., 2016a). Another study was locally analysed to assess the prognostic value of combined SUV_{max} for predicting PFS (Nishimura et al., 2016). Both studies examined combined HNSCC. The time between treatment completion and post-treatment PET/CT varied from around 1.5 months to 3.2 months. Patient survival was analysed based on two similar combined SUV_{max} values of 3.9 and 4.4.

Kim et al. (2016a) found that a threshold value of ≥4.4 obtained from combined SUV_{max} resulted in a three-year PFS of 42.9% in comparison to 81.1% in patients with a combined SUV_{max} of <4.4. The risk of disease progression within three years post-treatment was increased by a factor of 4.79 (95% CI: 2.02–11.32) (Kim et al., 2016a).

Similarly, Nishimura et al. (2016) found that a lower cutoff of 3.9 was optimal for comparing clinical outcomes. Based on a local analysis, the risk of disease progression within three years post-treatment was increased by a factor of 3.72 (95% CI: 2.42–5.71). These findings suggest that post-treatment combined SUV_{max} is a prognostic factor for disease progression.

In the meta-analysis, one study provided an HR for the association between post-treatment primary tumour and lymph node combined SUV_{max} and PFS (Kim et al., 2016a). Another study was further locally analysed to calculate the estimated HRs for PFS (Nishimura et al., 2016). This provided a total of two studies with 313 HNSCC patients, for which there were two HR estimates for PFS. Using the random effect model, we found no statistically significant

association between primary tumour and lymph node combined SUV_{max} and PFS, with an overall HR pooled estimate of 3.84 (95% CI: 2.28—5.39, *P*=0.670) (Figure 2.6).

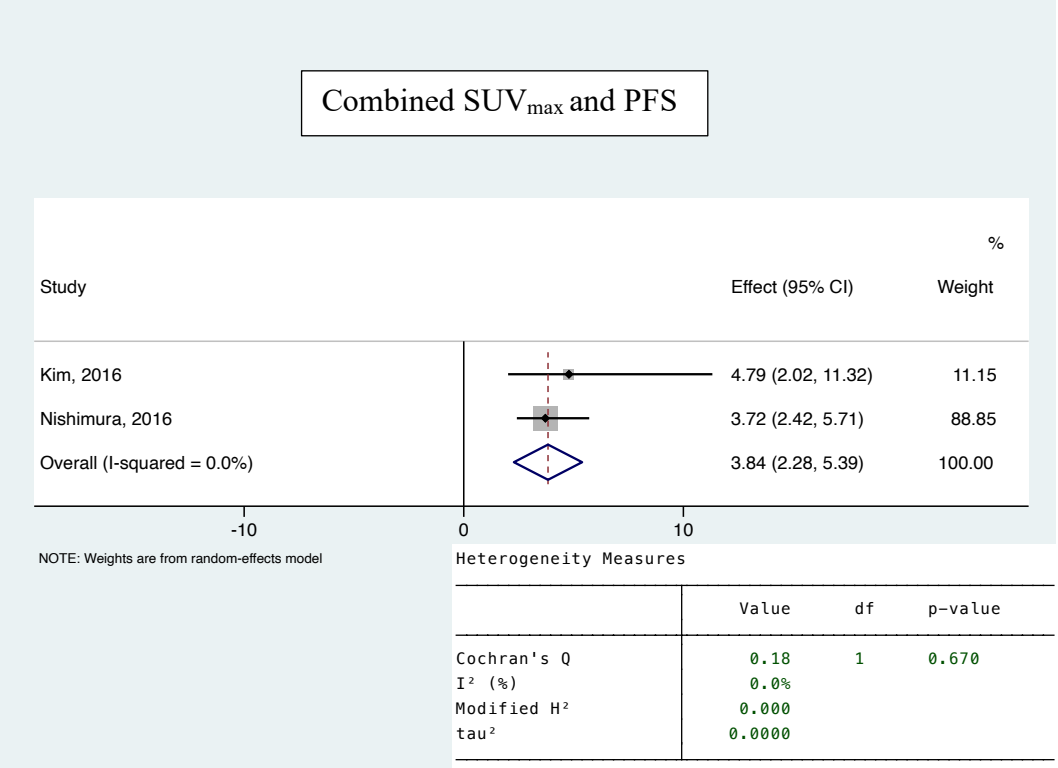


Figure 2.6: Forest plot shows visual analysis of PFS HRs associated with primary tumour and lymph node combined SUV_{max}

2.3.5.5 Primary tumour site SUV_{max} and PFS

Four studies were included to evaluate the prognostic value of post-treatment primary tumour SUV_{max} and PFS (Moeller et al., 2010, Katahira-Suzuki et al., 2015, Riaz et al., 2017, Oyama et al., 2020). Another study was locally analysed to assess the prognostic value of primary tumour SUV_{max} for predicting PFS (Nishimura et al., 2016). These studies explored various subtypes of head and neck cancer, including oral cancer (Oyama et al., 2020), combined HNSCC (Nishimura et al., 2016), pharyngolaryngeal HNSCC (Moeller et al., 2010), and pharyngeal cancer (Katahira-Suzuki et al., 2015). The time difference between treatment completion and post-treatment PET/CT varied from around 1.5 months to 6 months. Patient survival was analysed based on different primary tumour SUV_{max} values, which ranged from 3.9 to 6.3.

In 98 patients with pharyngolaryngeal cancer, Moller et al. (2010) found that patients with a post-treatment primary tumour HNSCC SUV_{max} of 6.0 had unfavourable DSS compared to the other group. Similarly, Oyama et al. (2020) found that at the primary tumour cutoff point of 6.3 patients, five-year local control differed significantly (96.3% vs. 41.7%; *P*=0.0001). At a lower primary SUV_{max} cutoff of 5.0, locoregional control rates also significantly differed (SUV_{max} <5=73.7% and SUV_{max} ≥5 = 10.0; *P*= >0.001) (Katahira-Suzuki et al., 2015).

A study by Riaz et al. (2017) found that the post-treatment primary tumour SUV_{max} of <5, 5–10, and >10 yielded three-year DFS rates of 62%, 42%, and 6%, respectively. Based on a local analysis, the risk of death within three years post-treatment was increased by a factor of 2.63 (95% CI: 1.68–4.12) in the group of patients with a higher nodal SUV_{max} cutoff. In contrast, Katahira-Suzuki et al. (2015) discovered that five-year metastatic-free survival (MFS) did not

differ significantly between patients with a higher and lower primary tumour cutoff value of 5. (87.0 vs. 78.8: *P*=0.137). Therefore, inconsistent findings were observed regarding the prognostic value of post-treatment nodal SUV_{max}.

In the meta-analysis, one study provided an HR for the association between post-treatment primary tumour SUV_{max} and locoregional control (Katahira-Suzuki et al., 2015). Another study was further locally analysed to calculate the estimated HRs for PFS (Nishimura et al., 2016). This provided a total of two studies with 305 HNSCC patients, for which there were two HR estimates for PFS. Using the random effect model, we found no statistically significant association between primary tumour and lymph node combined SUV_{max} and PFS, with an overall HR pooled estimate of 2.69 (95% CI: 1.48–3.89, *P*=0.522) (Figure 2.7).

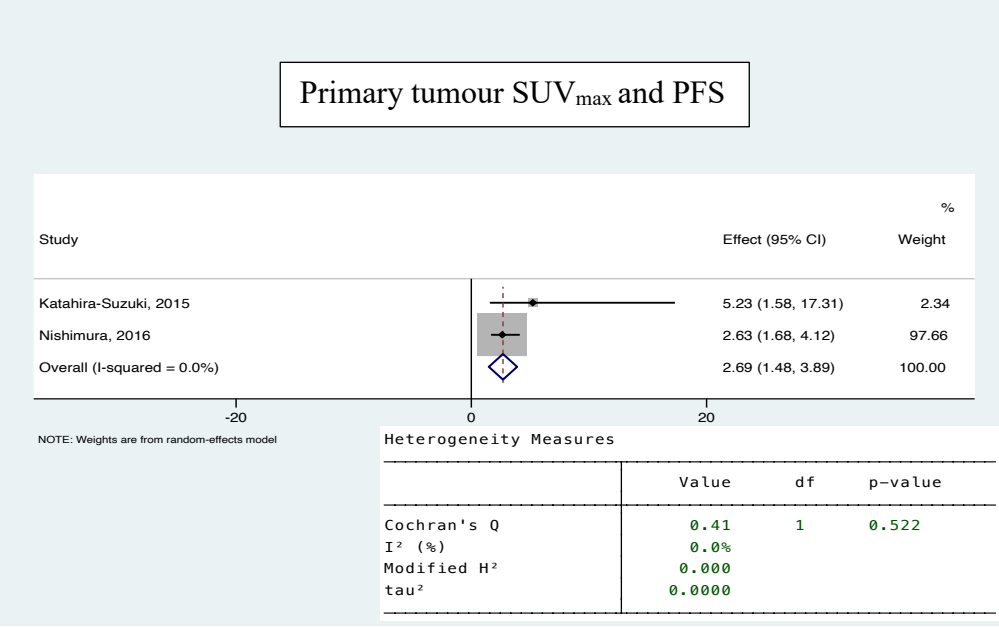


Figure 2.7: Forest plot shows visual analysis of PFS HRs associated with primary tumour and SUV_{max}

2.3.5.6 Lymph node site SUV_{max} and PFS

Four studies reported the prognostic value of post-treatment primary tumour SUV_{max} and PFS (Moeller et al., 2010, Katahira-Suzuki et al., 2015, Sjovall et al., 2016, Awan et al., 2017). Another study was locally analysed to assess the prognostic value of nodal SUV_{max} for predicting PFS (Nishimura et al., 2016). These studies explored various subtypes of head and neck cancer, including combined HNSCC (Sjovall et al., 2016, Nishimura et al., 2016, Awan et al., 2017), pharyngeal cancer (Katahira-Suzuki et al., 2015), and pharyngolaryngeal HNSCC (Moeller et al., 2010). The time difference between treatment completion and post-treatment PET/CT varied from around 1.5 months to 2.8 months. Patient survival was analysed based on different nodal SUV_{max} values, which ranged from 2.0 to 3.4.

Katahira-Suzuki et al. (2015) found that three-year locoregional control and three-year MFS did not change significantly when patients were divided based on a nodal SUV_{max} of 2.45. Two additional studies reached similar conclusions. They found that post-treatment nodal SUV_{max} of 3.4 and 2.0 were not associated with three-year DSS (Sjovall et al., 2016, Moeller et al., 2010).

Furthermore, Awan et al. (2017) analysed post-treatment imaging, stratifying individuals with P16+ and P16- statuses at a cutoff of 2.5. Post-treatment nodal SUV_{max} using a cutoff of 2.5 was correlated with RFS in both groups, in P16+ (three-year RFS of 89.7% with an LN SUV_{max} <2.5 versus 50% with SUV_{max} >2.5). In P16- (three-year RFS was 72.0% with an LN SUV_{max} <2.5, 21.4% with an SUV_{max} >2.5). This was the only study to examine the predictive value of SUV_{max} in patients with positive versus negative HPV status.

In a meta-analysis of 307 patients (Katahira-Suzuki et al., 2015, Nishimura et al., 2016), we found no statistically significant association between nodal tumour SUV_{max} and PFS, with an overall HR pooled estimate of 2.48 (95% CI: 1.82–3.14, *P*=0.846) (Figure 2.8).

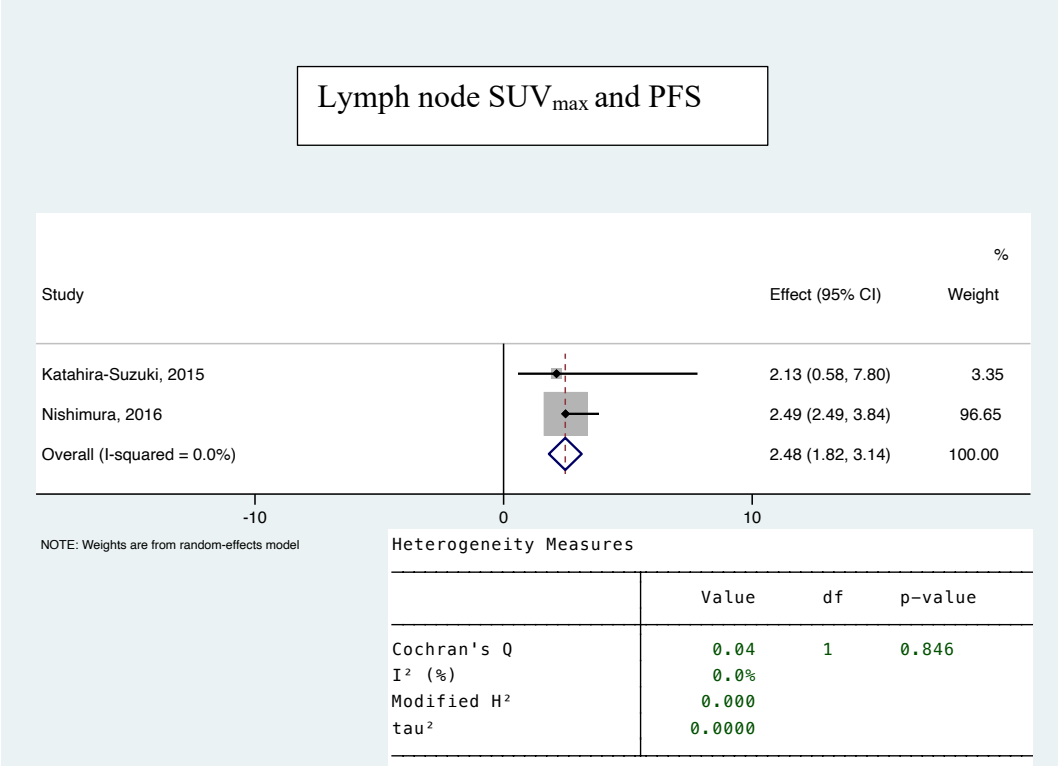


Figure 2.8: Forest plot shows visual analysis of PFS HRs associated with nodal SUV_{max}

2.4 Discussion

The purpose of this systematic review was to examine the current evidence for both the diagnostic performance and the prognostic value of post-treatment primary tumour and lymph node PET/CT metrics to discriminate residual disease and predict survival outcomes in patients with HNSCC treated by CRT. There were 16 studies that met the inclusion criteria. The present evidence shows that both post-CRT primary tumour and lymph node SUV_{max} could be used to predict residual disease on three-month PET/CT scans of HNSCC; however, the prognostic relevance was inconclusive.

Summary findings

A few studies have evaluated the diagnostic performance of post-treatment PET/CT quantitative metrics to discriminate residual disease at three months post-treatment from combined SUV_{max} and from primary tumour and lymph node separate SUV_{max} (Table 2.3).

The results of these studies agreed that both post-CRT primary tumour and lymph node metrics might be able to discriminate residual disease on three-month PET/CT scans of HNSCC. However, SUV_{max} thresholds varied significantly. One potential reason for these variations is the difference in time between the completion of CRT and post-treatment PET/CT imaging. According to Glastonbury (2020), the time interval between the end of treatment and the time to PET/CT is a critical factor. It is essential to wait enough time after therapy before assessing the treatment response to minimise treatment-induced inflammation. If the PET/CT is performed too soon, it may be difficult to differentiate between inflammation and persisting disease, and small residual lesions may go undetected. This increases the rate of false positives (Sagardoy et al., 2016). In this review, eight studies provided data on the

diagnostic performance of post-treatment primary tumours and lymph nodes performed at or after three months after therapy, while three studies contained data on PET/CT metrics performed less than three months post-treatment. This variation in timing may affect the rate of false positives and negatives, resulting in discrepancies in optimal thresholds and their related sensitivities and specificities. The optimal timing for post-treatment PET is described below.

Another factor that could have contributed to the disparity in the ideal threshold was the measures of sensitivity and specificity. Choosing a lower threshold value would improve its sensitivity (true positive rate) and decrease its specificity (true negatives) (Wang et al., 2021). The influence of sensitivity and specificity on threshold selection for post-CRT HNSCC lesions was difficult to assess due to the variability of the included studies. As a result, research into the trade-off effect for selecting the best primary tumour and lymph node SUV_{max} calculated three months post-CRT is recommended.

Furthermore, while both primary tumour and lymph node measurements were found to be capable of distinguishing residual illness, one study found that the ROC curve analysis was more accurate when performed on lymph node sites (Sagardoy et al., 2016). This could be due to higher inflammation at the primary tumour site compared to the nodal sites, reducing its accuracy at the primary sites. However, due to the small number of included studies and their heterogeneity, more research is needed to clarify the ability of the primary tumour and lymph node separate quantitative metrics to predict residual HNSCC disease in three months post-treatment PET/CT.

With regard to SUV metrics' prognostic value, we found inconsistent findings. Combined SUV_{max} was found to be predictive of OS and PFS (Ito et al., 2014, Kim et al., 2016a). However, primary tumour SUV_{max} was found to be a predictive factor in some studies, but not in others. On the other hand, lymph node SUV_{max} was not predictive in the majority of the included studies (Table 2.4). This suggests that an increase in post-treatment primary tumour SUV_{max} metrics could be more predictive of survival outcomes than nodal SUV_{max}. However, we cannot be certain because the included studies were heterogeneous in some ways. For example, some studies evaluated combined HNSCC, while others looked at specific head and neck cancer subsites, such as oral cancer, pharyngeal cancer, and laryngeal cancer. The inconsistent findings might also be due to the use of different types of scanners, parameters, and imaging protocols (Boellaard et al., 2015).

Also, as discussed in Chapter 1, HPV-associated OPSCC has been identified as a distinct subtype of head and neck cancer with a different prognosis (Budach and Tinhofer, 2019). A study by Awan et al. (2017) found that patients who had HPV^{+ve} disease with lymph node SUV_{max} >2.5 had worse three-year RFS (50%) than those who were HPV^{-ve} with nodal SUV_{max} <2.5 (three-year RFS 72%). In contrast, Connor et al. (2021) found that, regardless of HPV status, post-treatment combined SUV_{max} did not predict two-year survival (Connor et al., 2021). Therefore, the role of HPV status in SUV readings and its prognostic value remains unclear.

As mentioned previously, the time difference between therapy completion and PET/CT is a critical factor. A large clinical trial by Mehanna et al. (2016) found that PET/CT imaging at three months post-treatment was an effective modality for treatment response evaluation. Currently, post-treatment PET/CT imaging is usually performed at least three months after treatment, allowing for the reduction of CRT-induced effects, such as inflammation. Although this trial provided strong evidence for the effectiveness of post-treatment imaging at three months post-treatment, it did not examine the usefulness of PET/CT quantitative metrics. Knowing such information could be advantageous, as quantitative metrics could be used as a supplement to qualitative evaluation, providing an objective method for analysing lesions. Also, it could allow easier comparisons among patients across institutions, as previously described in Chapter 1 (Castelli et al., 2016).

SUV_{max}, as previously defined, reflects the radioactive concentration from a single pixel. This makes SUV_{max} more sensitive to technical and biological factors. Even though SUV harmonisation is sometimes difficult among different institutes, exploring the role of other metrics, such as SUV_{peak}, might provide a more accurate assessment method in comparison to SUV_{max}. In this review, all included studies analysed SUV_{max}, but none explored the utility of SUV_{peak}. According to Hiromitsu et al. (2021), using the SUV_{peak} can help eliminate disparities between scanners. Although deriving SUV_{peak} usually requires more time and effort compared with SUV_{max}, which is a simpler parameter to report, exploring the role of SUV_{peak} in HNSCC could be useful. Therefore, more research is needed to investigate the diagnostic and prognostic ability of SUV_{peak} to discriminate residual disease and predict survival outcomes in post-treatment HNSCC.

SUV normalisation, as stated in Chapter 1, involves the use of body size measures, such as total body weight, LBM and BSA, for SUV calculation. While SUV_{max} scaled to total body weight is the most commonly used parameter in clinical practice, it is possible that it is overestimated, particularly in obese patients (Sarikaya and Sarikaya, 2020). In this review, only one study examined SUV_{max} normalised to LBM (Chan et al., 2012), while the rest investigated SUV_{max} normalised to total body weight ($SUV_{max(w)}$) (Tables 2.3 and 2.4). However, none of the studies compared the effects of employing various body size measurements on SUV readings. Establishing such information could reveal whether the use of LBM and BSA reduces potential weight-related errors in SUV metrics, thus potentially improving their diagnostic or prognostic performance.

Strengths and limitations

This systematic review followed the PRISMA guidelines and a previously published protocol on PROSPERO. Also, this was the first systematic review to evaluate the diagnostic performance and prognostic value of PET/CT quantitative metrics determined on an average of three months post-CRT in HNSCC. Likewise, this review has limitations; the number of included studies to evaluate each parameter and outcome was limited. Even though a strict inclusion criterion was applied, there was still considerable study heterogeneity in terms of the timing of PET/CT imaging and the patient population. A further limitation was the lack of a uniform definition of survival outcomes. Some studies, for example, used PFS to reflect disease progression, whereas others used DFS and DSS. Some studies did not report complete data, such as the number of events in patients with a certain SUV threshold value. Treatment regimens have also been varied among the included studies. This may potentially impact

treatment response and thus, survival outcomes. A meta-analysis was not performed due to these issues.

In conclusion, despite variations observed in threshold values and their associated accuracies across studies, the findings of this analysis suggest that post-treatment primary tumour and lymph node SUV_{max} might be able to discriminate residual HNSCC lesions, but thresholds have significantly varied. With regard to the optimal thresholds and their prognostic significance, further investigation is required due to inconsistent findings in the included studies.
Regarding the utility of SUV_{peak}, the influence of LBM and BSA on SUV normalisation, and the effect of HPV status on SUV readings, all these areas seem to be gaps in the literature that should be filled by the development of high-quality evidence.

CHAPTER 3

GENERAL METHODOLOGY

3. Chapter 3: General methodology

3.1 Setting and participants

The hospital electronic database was screened to identify patients who were diagnosed with HNSCC, including cancers of the oral cavity, oropharynx, hypopharynx, and the larynx, who received (chemo)radiotherapy CRT and then underwent three-month post-treatment follow-up FDG PET/CT imaging at NHS University Hospitals Birmingham between 2016 and 2019. Exclusion criteria involved patients with PNSs, thyroid, or salivary gland cancers except for SCC of the nasal cancer; patients who only received palliative treatment, such as IC; and patients who underwent curative-intent surgery or neck dissection prior to PET/CT imaging. Furthermore, data on subjects' age, gender, follow-up time, tumour site, HPV status, TNM staging, and histology findings were retrospectively collected and analysed. Patients with insufficient data, such as histopathology results or loss of follow-up, were excluded. To avoid the possibility of image artefacts impairing quantitative PET/CT measurements, studies with severe artefacts caused by extravasation, motion, or intense brown fat uptake proximal to target lesions were excluded.

3.2 Study size

This was a retrospective study. All eligible patients were included. The size of the study was comparable to that of other published studies (Vainshtein et al., 2014, Sjovall et al., 2016, Awan et al., 2017, Helsen et al., 2020a).

3.3 Ethical approval:

This retrospective study was approved by the University of Birmingham, and ethical approval was granted by the Health Research Authority (HRA); REC reference 21/PR/1354 (21 September 2021). Due to the retrospective, anonymized nature of the study, the need for written informed consent was waived.

3.4 General scanning protocol

The imaging protocols were optimised for clinical use. Prior to imaging, the patients fasted for a minimum of four hours. After ensuring that the patient's blood glucose level was maintained between 4 and 12 mmol/L, an intravenous ^{18}F-FDG dose was administered. The injected activity was based on body weight, with an injection regime of 3.5 MBq/kg (+/- 10%). No intravenous or oral contrast was administered. Afterwards, the patients were instructed to relax in a warm area for an average uptake time of 60 min. After voiding the bladder just before the scan, the patients were positioned supine, typically with their arms down, and whole-body PET/CT imaging was performed from the skull base to the mid-thighs using one of the two scanners:

1) The first scanner was the Siemens Biograph 64-mCT. For PET acquisition: flow motion mode, slice thickness of 5 mm, pixel spacing of 4.1/4.1, scan duration of 20 min, continuous bed motion of 0.6 mm/s (head and neck region) or 0.8 mm/s (rest of the body), CT attenuation correction, and reconstruction method for AC Ultra HD (TrueX + ToF (2i, 21 s, Gaussian filter 2 mm). CT acquisition parameters included 5 mm slice thickness, 0.5 s rotation time, 0.8 pitch, automatic exposure mAs (reference value of 60 mAs) and 120 kVp.

2) The GE Medical System Discovery 710 was the second scanner used. For PET acquisition: slice thickness 3.27, recon diameter 700, FOV dimensions 700/153, pixel spacing 3.6/3.6 time per bed position 3 min/bed; attenuation correction method: smooth; reconstruction method: Qclear 400. For CT acquisition: slice thickness 3.75 mm, matrix size 512*512, automatic exposure mAs and kVp 120.

3.5 General image analysis

Medical image viewing and analysis software, Hermes Hybrid Viewer PDR version 5.0.1, was used to analyse lesions quantitatively and segment tumour volumes. Attenuation-corrected PET and CT images were analysed, and fused images were then displayed in the sagittal, coronal, and axial planes. To locate the target lesions, PET/CT imaging reports were also reviewed. Each primary and involved lymph node lesion's SUV_{max}, SUV_{peak}, TLG, and MTV were determined by positioning 3-D VOIs. When delineating a lesion boundary, the associated CT images were reviewed. In lesions with unclear boundaries, clinical advice was sought. Three different normalisation factors, including total body weight, LBM and BSA, were applied to calculate the examined parameters (Table 3.1). SUV normalised to LBM (SUV_{lbm}) corresponds to the SUL parameter stated in the PERCIST guidelines (Wahl et al., 2009); however, for a simplified comparison, the acronym SUL was not used throughout this thesis.

Table 3.1: Definitions of maximum and peak SUV metrics normalised to total body weight, LBM, and BSA

Parameter	Definition
SUVmax (w)	Maximum SUV normalised to total body weight
SUVpeak (w)	Peak SUV normalised to total body weight
SUVmax (lbm)	Maximum SUV normalised to lean body mass (LBM)
SUVpeak (lbm)	Peak SUV normalised to lean body mass (LBM)
SUVmax (bsa)	Maximum SUV normalised to body surface area (BSA)
SUVpeak (bsa)	Peak SUV normalised to body surface area (BSA)

3.6 PET/CT quantitative metrics

3.6.1 SUV_{max}

To calculate SUV_{max}, a nonthreshold VOI was placed over the highest uptake region of the lesion. The SUV_{max} of the lesion was then determined semi-automatically three times using different body size measurements for normalisation (Figure 3.1). These body size measurements consisted of total body weight (kg), LBM (kg), and BSA (m^2).

SUV was calculated based on Equation 1 (see Chapter 1, Section 1.10.1).

Applying weight in kg to calculate SUV_w, LBM for SUV_{lbm}, and BSA for SUV_{bsa}.

BSA was calculated based on the following formula (Hallett et al., 2001):

$$BSA = (W^{0.425} \times H^{0.725}) \times 0.007184 \quad (2)$$

LBM was determined based on height and weight following James' formulas for male and female (Tahari et al., 2014):

$$LBM(Female) = 1.07\ x\ W - 148\ x\ (W/H)\text{^}2 \quad (3)$$

$$LBM(Male) = 1.10\ x\ W - 128\ x\ (W/H)\text{^}2 \quad (4)$$

3.6.2 SUV_{peak}

To calculate SUV_{peak}, a 3D sphere with a diameter of around 1.2 cm was drawn (Boellaard et al., 2015). Then, SUV_{peak} values with various normalisation factors, including total body weight, LBM and BSA, were reported (Figure 3.2).

3.6.2.1 Negative post-treatment PET/CT scans (SUV_{max} and SUV_{peak})

To determine the SUV_{max} and SUV_{peak} of the lesions in patients who responded fully to treatment in which PET/CT scans were clear, a standard VOI with a diameter of 1.25 cm was placed on the location of the pre-treatment lesions using pre-treatment CT (Helsen et al., 2020b) or MRI as a reference. The VOI was carefully placed to avoid placing it over other high-uptake regions.

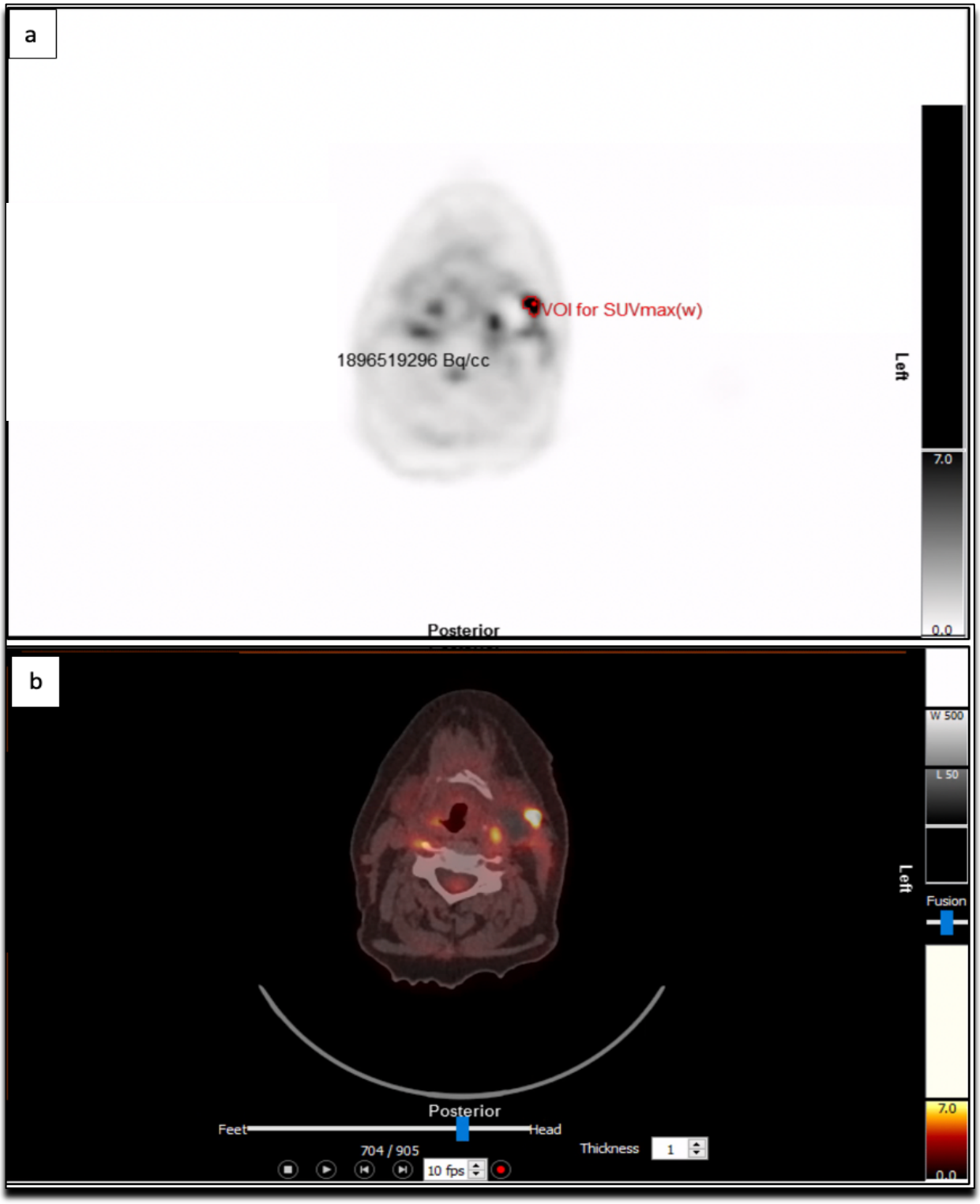


Figure 3.1: SUV_{max} of HNSCC.
Post- CRT PET/CT for patient with supraglottic cancer shows positive involvement of left lymph node with $SUVmax_w$ of 11.2 on **a)** attenuation-corrected PET image and **b)** Fused PET/CT image. This $SUVmax_w$ of 11.2 reflects the value of the radioactivity concentration at the most active voxel within the target lesion.

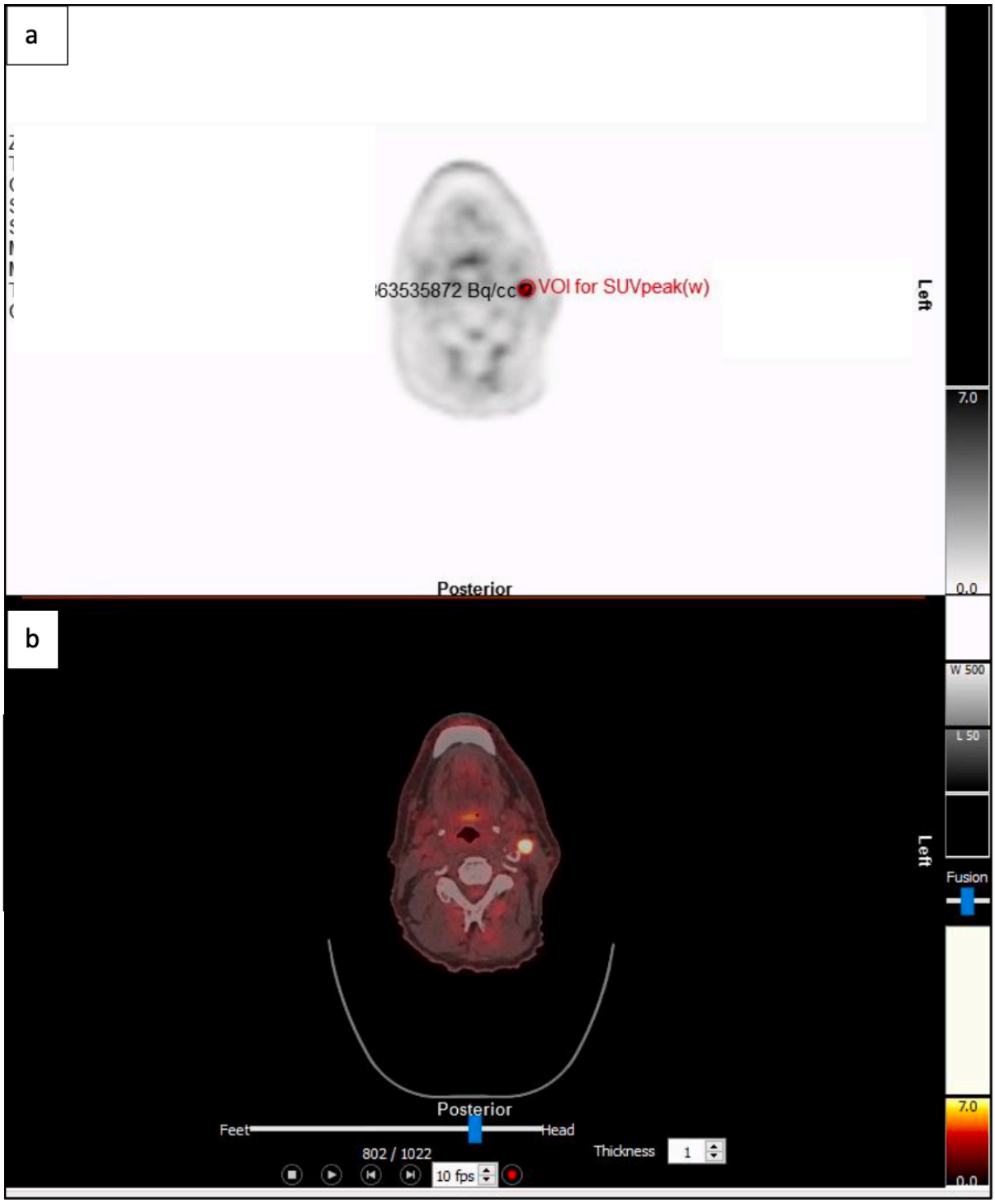


Figure 3.2: SUV_{peak} of HNSCC.
A 1.25 cm diameter VOI cantered on a persistent avid lymph node lesion shows an $SUVpeak_w$ of 8.14 in a patient diagnosed with tonsillar cancer on both **a)** attenuation-corrected PET image and **b)** fused PET/CT image. This $SUVpeak_w$ of 8.14 reflects the average value of radioactivity concentration within the VOI surrounding the pixel with the highest activity.

3.6.3 Metabolic tumour volume (MTV)

To calculate the MTV for each detectable lesion, a VOI was first drawn to determine the lesion SUV_{max}. Following this, a fixed-relative thresholding technique was applied (Im et al., 2018). In this method, a threshold percentage relative to the value of SUV_{max} was set to contour tumour boundaries. For instance, if an initial VOI was formed with a threshold set at a value of 3.0 and a tumour SUV_{max} was defined as 8.0, this value is multiplied by the percentage chosen relative to the tumour SUV_{max} (50%), providing a new threshold value of $8.0X0.5=4.0$. In this lesion, all voxels with SUV equal to or greater than 4.0 would be assigned to the tumour, while others (<4.0) would be assigned to the background.

For PET tumour segmentation, the European Association of Nuclear Medicine (EANM) suggests measuring MTV using 41% and 50% relative thresholds to SUV_{max} (Boellaard et al., 2015). Based on visual observations in our cohort, we found that these segmentations failed to accurately reflect the apparent volume of the majority of the included lesions. According to Boellaard et al. (2015), due to noise, tracer uptake heterogeneities, and occasionally low tumour-to-background ratios, these fixed thresholds do not always delineate lesions accurately and care must be taken to exclude non-lesion uptake in any VOIs to reduce bias in volume estimates.

To evaluate the quality of these approaches, a grading system, like that applied by Zwezerijnen et al. (2021), was used. This system classifies the accuracy of a segmentation method into three categories: acceptable segmentation (score 1), nonacceptable segmentation (score 2), and poor segmentation (score 3). Score 1 represented a visually accurate tumour delineation or minimal background uptake (Figure 3.3); score 2

corresponded to a segmentation that included some visibly noticeable local background areas (Figure 3.4-a); and score 3 indicated a visually extensive inclusion of background uptake (Figure 3.4-b). The results of these segmentations are shown in Appendix 3.

Therefore, we modified the methodology and investigated an alternative method, the adaptive-thresholding technique, in which the appropriate threshold value is adjusted on a case-by-case basis instead of pre-assigning specific fixed absolute or relative values when contouring lesions and calculating corresponding MTVs (Im et al., 2018). Contouring HNSCC lesions in the post-treatment setting was improved by using a relatively high percentage of 60–80%. This segmentation was then applied to contour tumour MTVs, which were analysed in Chapter 4.

3.6.4 Total lesion glycolysis (TLG)

TLG is the second type of PET/CT volumetric parameter. This is the result of MTV multiplied by SUV_{mean} (Im et al., 2018).

The formula for calculating a lesion's TLG is as follows:

$$TLG = SUVmean \times MTV \quad (5)$$ (Im et al., 2018)

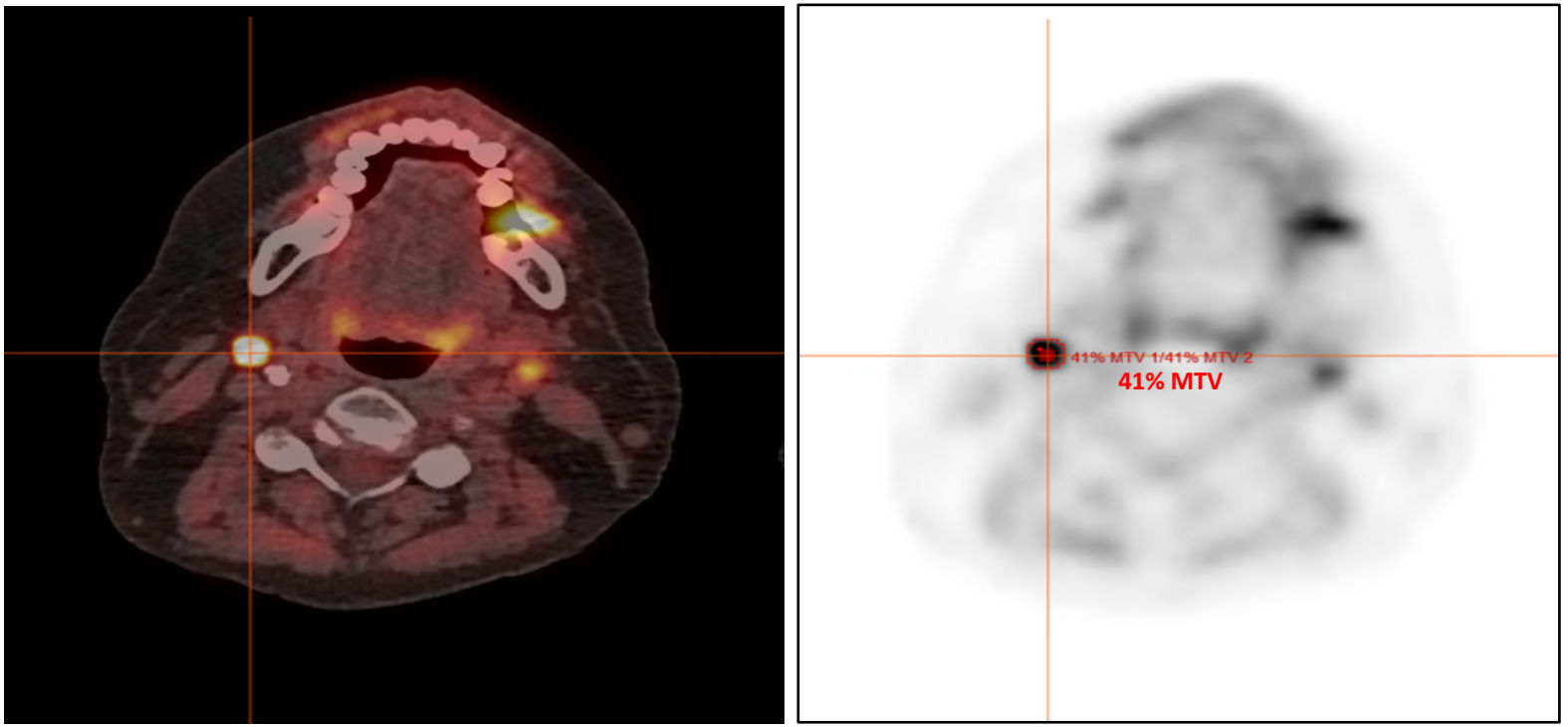


Figure 3.3: Acceptable MTV segmentation shows a visually accurate tumour delineation with using 41% segmentation.

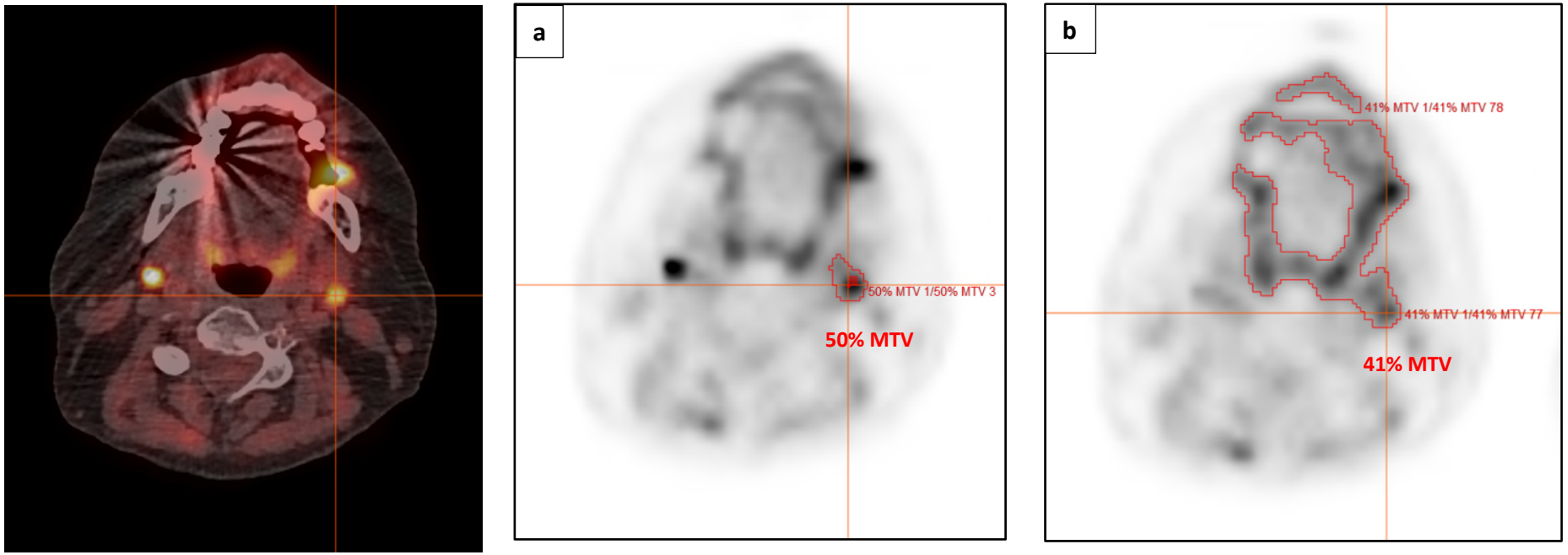


Figure 3.4: Segmentation shows a) nonacceptable MTV segmentation b) even worse poorly defined MTV segmentation.

3.7 SUV in background regions

SUV_{mean} and SUV_{max} for each patient were derived from the liver, cerebellum, and blood pool (BP).

A 3-cm spherical diameter VOI was placed over the liver's right lobe, avoiding any focal uptake (Boellaard et al., 2015) (Figure 3.5).

Over the cerebellum, a second VOI with a diameter of 2.0 cm was drawn (Figure 3.6).

The BP SUV_{mean} was estimated by drawing VOIs with diameters ranging from 1.5 to 2.1 cm over the aortic arch, excluding vessel walls (Figure 3.7).

All SUV_{mean} and SUV_{max} background reference regions normalised by weight, LBM, and BSA were recorded.

.

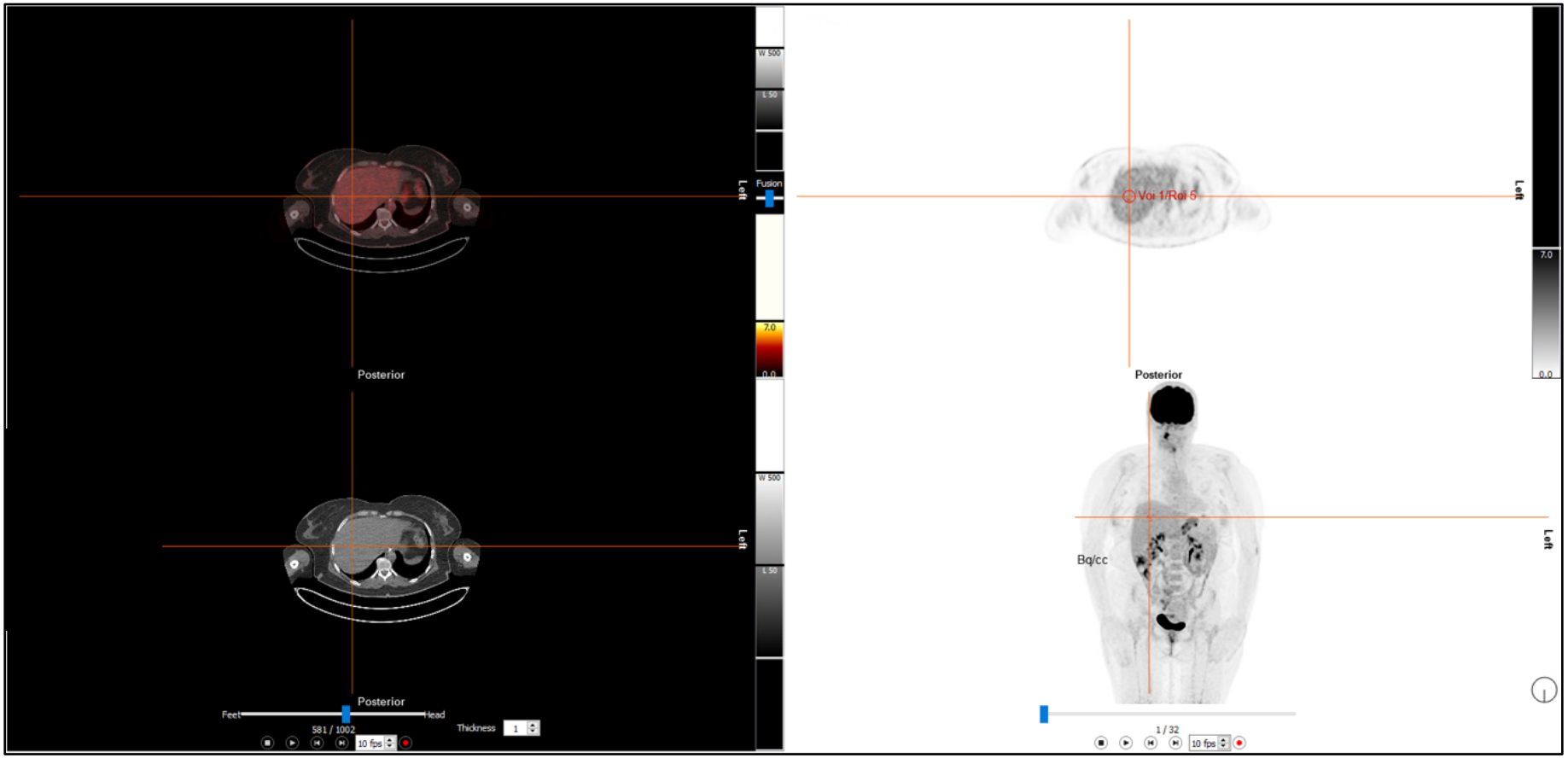


Figure 3.5: The liver SUV_{mean} and SUV_{max} derived from a fixed-size VOI drawn over the right lobe of the liver.

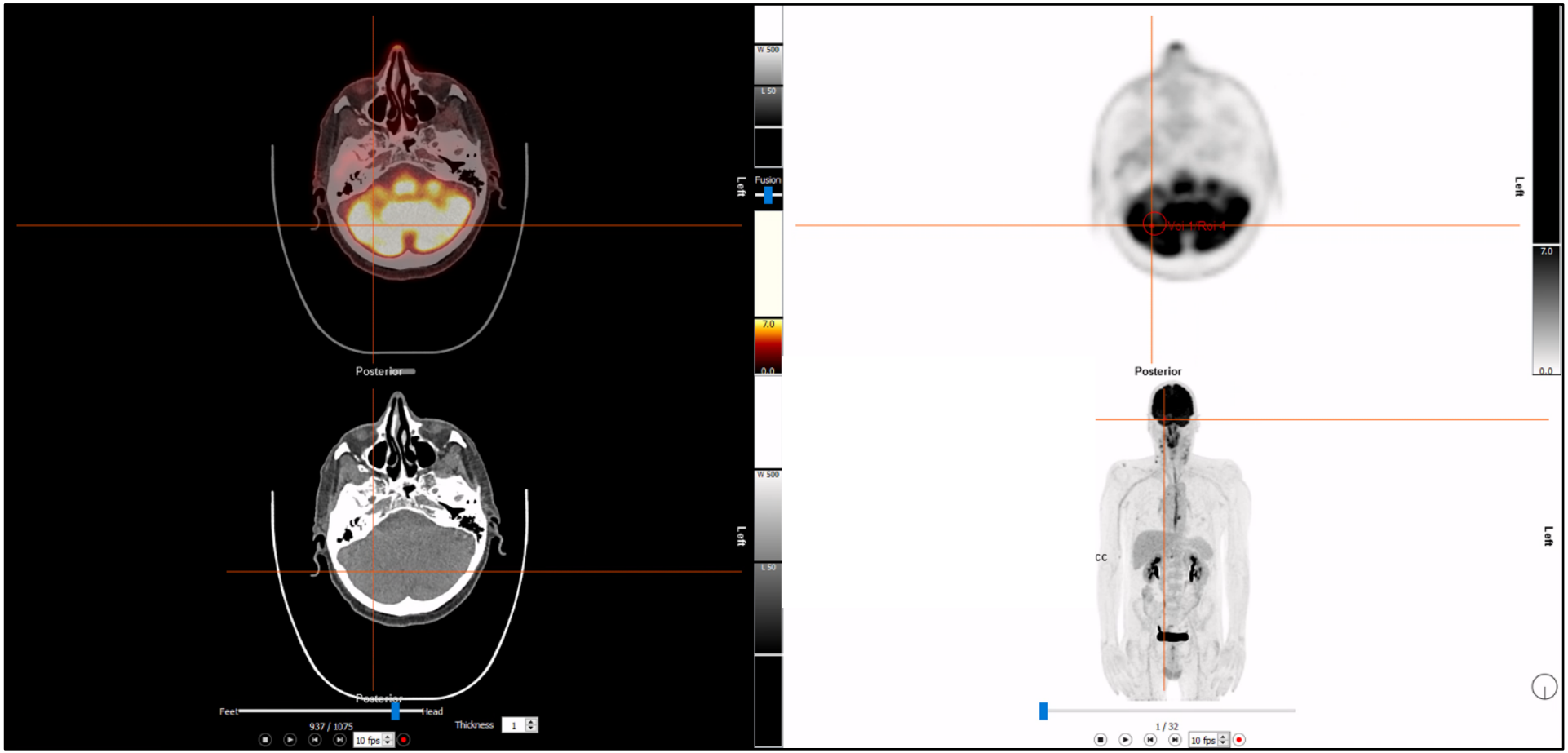


Figure 3.6: SUV_{mean} and SUV_{max} derived from a fixed-size VOI drawn over the cerebellum.

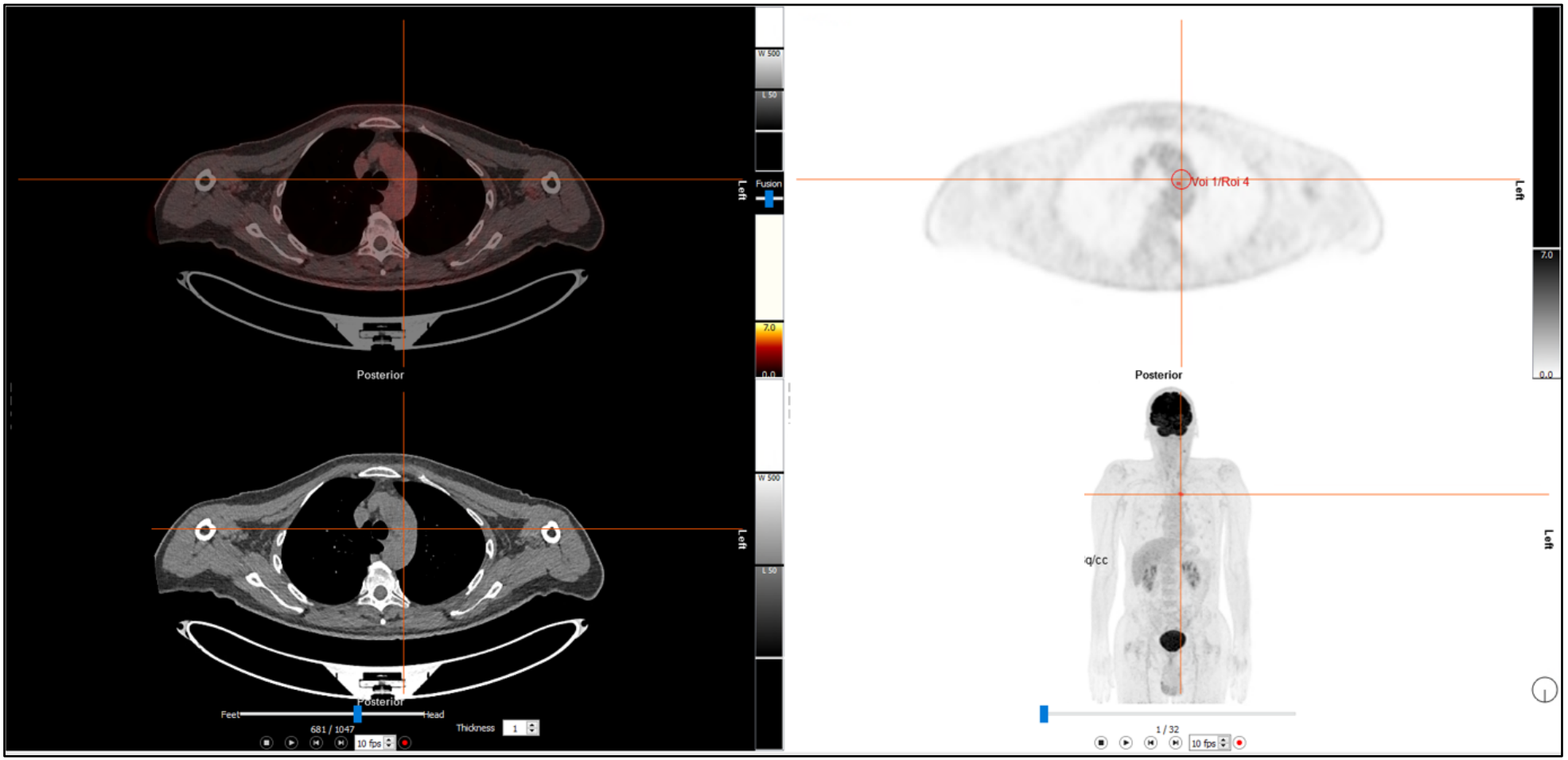


Figure 3.7: The blood pool SUV_{mean} and SUV_{max} derived from a VOI drawn over the aortic arch.

3.8 General Statistical Analysis

- The normality of the data distribution was evaluated by applying the Shapiro–Wilk test and histogram plots. Since the data were not normally distributed, nonparametric tests were applied throughout this thesis.
- For descriptive statistics, continuous variables were described by median and interquartile range (IQR), whereas binary variables were presented in numbers and percentages.
- Data analysis was conducted using Stata/SE 16.1 and IBM SPSS Statistics version 28.0.
- We considered a P value of less than 0.05 significant.

3.9 Statistical analysis and methods

3.9.1 Analysis of the correlation of PET/CT measurements with body metrics

Only patients with visible lesions with positive and equivocal PET/CT scans were included. In patients with multiple lesions, all measurable primary tumours and lymph node lesions were analysed.

In Chapter 4, each lesion SUV_{max}, SUV_{peak}, MTV and TLG normalised to total body weight, LBM, and BSA were determined as previously described. The liver SUV_{mean} and SUV_{max} were also determined.

Spearman correlation analysis was used to assess the correlation between PET/CT metrics and various patient body size measurements, such as height (cm), weight (kg), LBM (kg), BSA (m^2), and BMI (kg/m^2) (Sanghera et al., 2009, Sarikaya et al., 2020). This was conducted to determine the normalisation body size metric that has the least effect on PET/CT metrics calculations and so could be viewed as producing data with the least bias.

The Spearman correlation coefficient was denoted by r. The magnitude of the correlation was described according to the following rules: *r*=0.00 to 0.10 indicates negligible correlation; *r*=0.10 to 0.39 suggests weak correlation; *r*=0.40 to 0.69 shows moderate correlation; *r*=0.70 to 0.89 indicates strong correlation; and *r*=0.90 to 1.00 indicates very strong correlation (Schober et al., 2018).

The same methodology was also used to analyse the relationship between liver SUV_{mean} and SUV_{max} and various body size parameters. The purpose of this was to see if the effect of employing different body size measurements differed when applied to normal background regions, such as the liver and malignant lesions, such as HNSCC.

Scatter plots were also generated to visualise the relationship between each normalised parameter and various body size factors (height, weight, LBM, BSA, and BMI).

Since our data showed that the type of normalisation factor did not greatly influence how MTV was calculated, only the correlation between MTV normalised to weight and patient factors was examined and presented.

All included patients were scanned using the Siemens Biograph 64-mCT.

3.9.2 Analysis of the diagnostic performance of post-treatment quantitative metrics

3.9.2.1 The diagnostic performance study: A cohort study

The results of the study were reported using the Standards for the Reporting of Diagnostic Accuracy Studies (STARD) guidelines (Cohen et al., 2016) (Appendix 4).

This was a retrospective study. Three-month post-treatment PET/CT scans of patients with HNSCC were retrospectively analysed.

The diagnostic performance of several post-treatment PET/CT metrics was evaluated, including the following:

- Primary tumours and lymph nodes SUV_{max} and SUV_{peak} normalised by weight, LBM, and BSA.
- Lesions-to-background SUV relative metrics derived from the liver, cerebellum, and BP (aortic arch).

- Patients' data with positive, equivocal, and negative PET/CT scans were included. All primary tumour lesions, as well as the lymph node lesions with the highest FDG uptake, were analysed. Also, lesion-to-background ratios (relative metrics) obtained from three different backgrounds regions were also derived.
- The index tests were post-treatment PET/CT quantitative metrics. The methods used for determining these metrics were previously described in Sections 3.6.1, 3.6.2, 3.6.2.1, 3.6.3, 3.6.4, and 3.7.
- The reference standard was the histopathology results of a biopsy, tumour resection, or neck dissection. In the absence of a histopathology report, a negative clinical and/or radiological follow-up was used to determine the positivity of the lesion.
- PET/CT imaging reports were available when PET/CT metrics were analysed. In some challenging cases, a consultant radiologist assisted in ensuring accuracy in identifying the target lesions.
- The reader was blinded to the histology results during the image analysis process.
- The imaging procedure is described in Section 3.4.

The primary endpoint was disease status at three months post-treatment completion.

Disease status was defined as follows:

- Local and regional controls were defined as the absence of residual disease at the primary tumour or lymph node sites on neck dissection, biopsy, or clinical one-year follow-up.
- The visual disappearance of the intense uptake of target lesions on post-treatment PET/CT images was defined as a complete metabolic response to treatment (negative PET/CT) (Sagardoy et al., 2016).

- The visual appearance of focal intense uptake or uptake that was greater than uptake in the surrounding background regions on post-treatment PET/CT images was defined as an incomplete metabolic response to treatment (positive and equivocal PET/CT).
- **Reference standard:** The presence of histologically confirmed disease on a biopsy or neck dissection was considered a residual disease. If not available, a one-year negative clinical/radiological follow-up was used to confirm the positivity of the lesions.

The flowcharts in Figures 3.8 and 3.9 illustrate the analysis plan used to evaluate the diagnostic performance of post-treatment HNSCC PET/CT metrics.

- The first analysis investigated the diagnostic performance of 12 different absolute imaging parameters to detect residual disease at three months post-treatment PET/CT (Figure 3.8).
- The second analysis investigated the diagnostic performance of an additional 12 different relative imaging parameters (lesions-to-background ratios) to detect residual disease at three months post-treatment PET/CT (Figure 3.9).

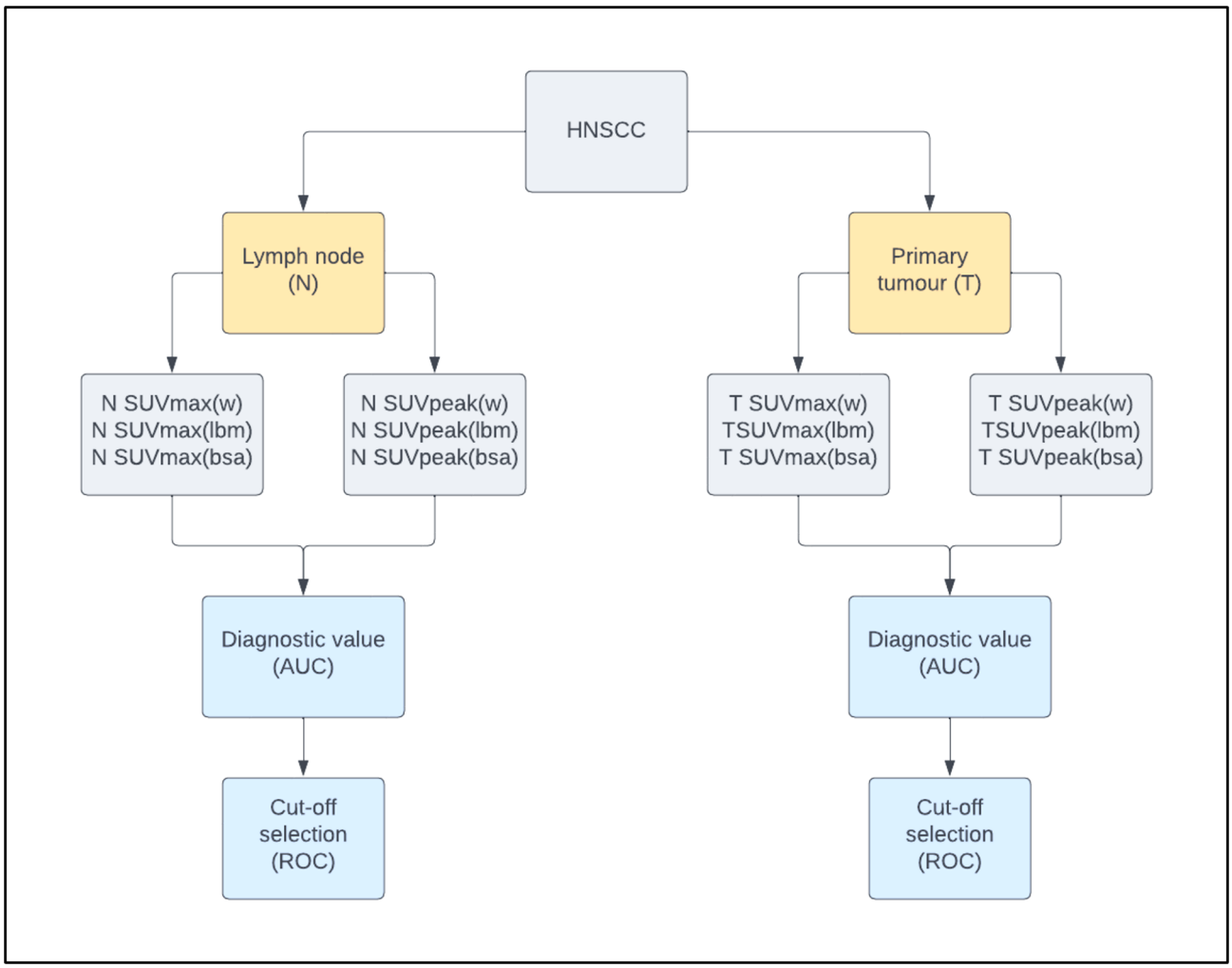


Figure 3.8: Analysis Plan Flowchart of the absolute metrics.

The initial analysis investigated the diagnostic performance of post-treatment primary and lymph node lesion SUV absolute metrics for discriminating residual disease. The area under the receiver operator characteristic (AUC) curves was used to assess the diagnostic performance of the imaging metrics.

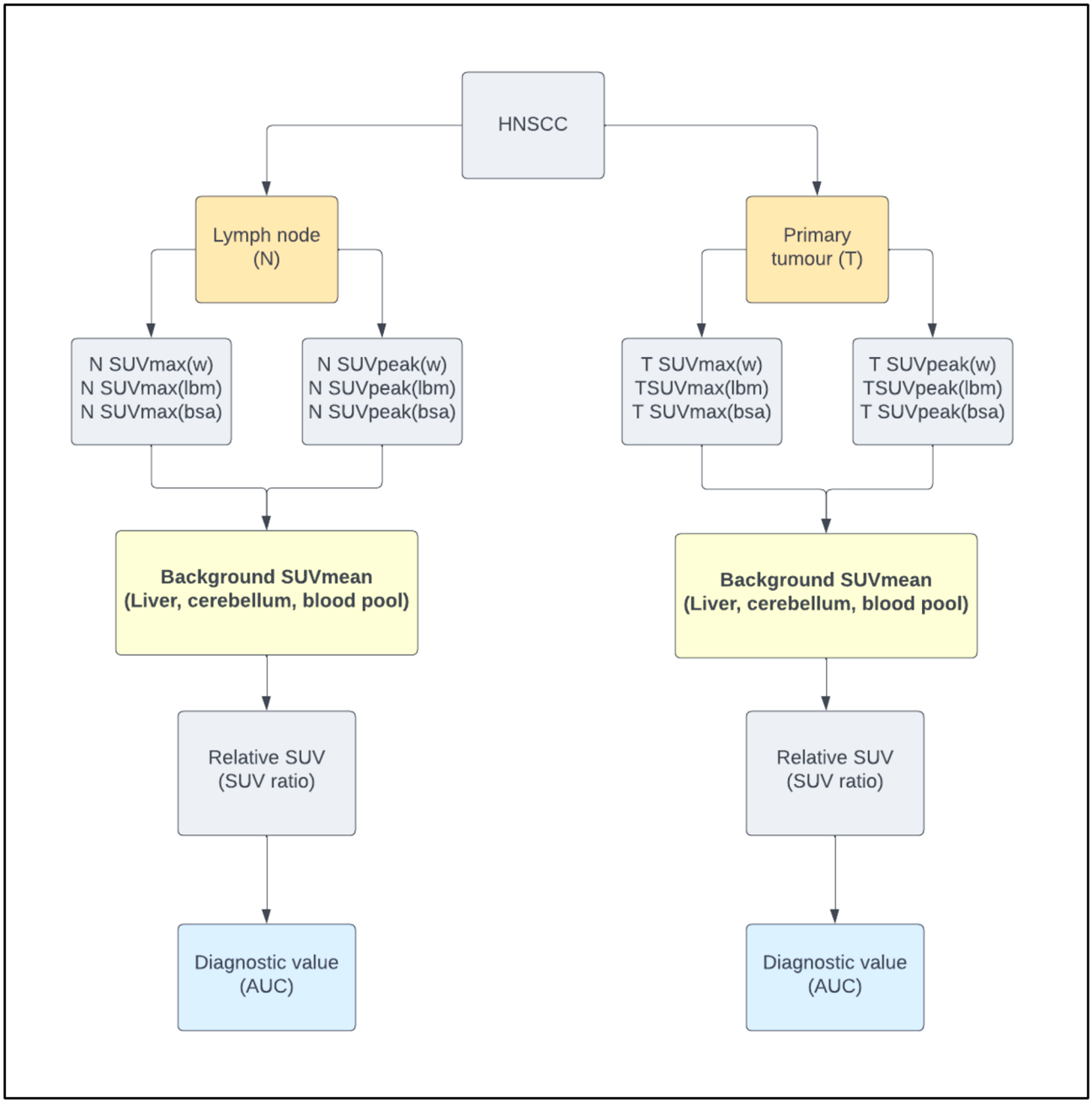


Figure 3.9: Analysis Plan Flowchart of the relative metrics.

The second analysis investigated the diagnostic performance of post-treatment primary and lymph node lesion SUV **relative metrics** for discriminating residual disease. These metrics were calculated by dividing the lesion SUV_{max} or SUV_{peak} by the SUV_{mean} obtained from the background region (SUV ratios). The diagnostic performance of the relative metrics was then assessed using the area under the receiver operator characteristic (AUC) curve.

The ROC curves were plotted for the post-treatment HNSCC PET/CT absolute and relative metrics and the final findings of the histology results. To compare their diagnostic performance, AUCs were computed and compared (Habibzadeh et al., 2016).

SUV lesions-to-background ratios were calculated by dividing the target lesion SUV_{max} and SUV_{peak} by the SUV_{mean} obtained from the background regions. SUV_{mean} from three background regions was calculated, including the liver SUV_{mean}, the cerebellum SUV_{mean}, and the BP SUV_{mean}.

For threshold selection, ROC curves were used to examine threshold sensitivity and specificity. The ROC curve illustrates the sensitivity of various post-treatment SUV metrics versus (1-specificity). Sensitivity, also known as the 'true positive rate', indicates the ability of a test to yield a positive result when a subject has a disease determined by a reference standard test. Specificity, on the other hand, is the 'true negative rate', indicating the ability to find a negative result if there is no disease. If either test is perfect, it will have a sensitivity or specificity of 100%. However, just because a test has 100% sensitivity or specificity does not imply that it is flawless. A test with 100% sensitivity or specificity, but no utility, would always return a positive or negative result. For example, a test that returns a positive result for every single person in the population has a sensitivity of 100%, but is completely ineffective (Wang et al., 2021).

Consequently, we have also analysed the associated NPV and PPV, which consider disease prevalence and are important clinical decision-making factors (Dejaco et al., 2020). NPV is the likelihood that a randomly selected person with a negative test will not have a true disease, whereas

PPV is the likelihood that a randomly selected person with a positive test will actually have the disease.

Previous studies have shown that PET/CT has a high NPV and a low PPV (Sagardoy et al., 2016, Nelissen et al., 2017). We regard a high NPV as preferable to a high PPV from a clinical standpoint, and the cutoff values were chosen based on this preference (Sjovall et al., 2016). This was carried out by selecting thresholds that produced the greatest possible sensitivity while maintaining a specificity >80% (Helsen et al., 2020a). The aim was to improve a cutoff's NPV so that we could exclude patients with no residual lesions, possibly minimising the number of unnecessary invasive procedures, such as neck dissection. Where NPV values were very similar for different SUVs, we selected the optimal one that yielded higher specificity.

There is a trade-off between sensitivity and specificity (Table 3.2); as sensitivity increases, specificity and PPV also decrease, while NPV increases. Increasing specificity increases the PPV while decreasing the NPV. Decreasing PPV results in an increase in NPV (Wang et al., 2021).

Table 3.2: Associations between sensitivity, specificity, PPV, and NPV (Wang et al., 2021).

	Specificity	PPV	NPV
Sensitivity	-	-	+
Specificity		+	-
PPV			-

Abbreviations: PPV, positive predictive value; NPV, negative predictive value; +, directly proportional; -, inversely proportional.

Therefore, to provide a more comprehensive interpretation of the cutoffs, we examined sensitivity, specificity, PPV, and NPP and discussed the trade-off effect between these parameters derived from three months of post-treatment HNSCC lesions.

- Patients with missing histopathology results or insufficient follow-up information were excluded.
- Possible confounding factors, such as the timing of post-treatment imaging and the influence of HPV status, are discussed in Chapter 7.

3.9.2.2 The reproducibility of SUV metrics

In this sub-analysis, 20% of the total cohort was randomly selected using an online randomization tool (Research Randomizer). The total number of lesions examined was 42, and the total number of SUVs in the background region was 25.

To examine the intraobserver reproducibility, the same rater acquired the SUV metrics twice within at least five months using the same image processing methodology previously described.

These methods were repeated after five months, and the two readings for each parameter were compared.

The rater was blinded to the measurements obtained the first time. Clinical reports, on the other hand, were used to confirm the correct location of the target lesions on each occasion.

The intraclass correlation coefficient (ICC) was used to assess the agreement between measurements (Koo and Li, 2016), or, in other words, the variation in the SUV measurements. The goal was to examine whether the image processing method was consistent when all SUV metrics were extracted.

The ICC model used was a two-way mixed model in which the rater was fixed (the same rater). Therefore, the results reflect the repeatability of a specific rater, which cannot be generalised to other raters. Furthermore, to evaluate how close the SUV measurements were when acquired twice at different time points, an absolute agreement analysis was applied (Koo and Li, 2016).

The analysed SUV measurements included lesion SUV_{max} and SUV_{peak} all normalised by total weight (W), LBM and BSA. Furthermore, the SUV_{mean} and SUV_{max} of the liver, BP, and cerebellum were all normalised by W, LBM and BSA. Variabilities among 24 different PET/CT parameters were investigated.

The following general guidelines were used to assess the ICC: values less than 0.50 represented poor reliability, values between 0.50 and 0.75 represented moderate reliability, values between 0.75 and 0.90 represented good reliability, and values greater than 0.90 represented excellent reliability (Koo and Li, 2016).

For a graphical presentation of global agreement between the two measurements, Bland–Altman plots with 95% limits of agreement were constructed. These plots were used to visualise the differences in the two measures against their means. Constructing Bland–Altman plots involves calculating the mean of the two values, the difference between the two values, the average of the difference, and the upper and lower 95% CIs of the average of the difference value (Bland and Altman, 1999).

In addition to the Bland–Altman plot, a simple linear regression was utilised to determine whether the bias, if any, between two measures was constant across their range or whether it varied with the magnitude of the measurements (proportional bias) (Arakawa et al., 2020). If the *P* value was >0.05, this suggested the nonoccurrence of proportional bias. If the *P* value was <0.05, this indicated the occurrence of proportional bias.

3.9.3 Analysis of the prognostic value of post-treatment quantitative parameters in predicting survival outcomes in patients with HNSCC

The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies (Appendix 5) were used to report the study's results (von Elm et al., 2007).

This was a retrospective study. Three-month post-treatment PET/CT scans of patients with HNSCC were retrospectively analysed. The prognostic value of several post-treatment PET/CT metrics was evaluated, including the following:

- Primary tumours SUV_{max} and SUV_{peak} normalised by weight, LBM, and BSA.
- Lymph nodes SUV_{max} and SUV_{peak} normalised by weight, LBM, and BSA.
- After CRT completion, patients were retrospectively followed for a minimum of three years and a maximum of five years. Time of death and data on last follow-up within 5 years of treatment completion were collected to analyse the association between PET/CT metrics and the average 5-year OS.
- Patients' data with positive, equivocal, and negative PET/CT scans were included. All primary tumour lesions, as well as the lymph node lesions with the highest FDG uptake, were analysed.

Survival outcome variables

The primary endpoints of Chapter 6 are as follows:

- Three-year PFS
- Five-year OS

PFS was defined as the time interval between the final fraction of CRT and the time to detection of persistent disease, distant metastatic disease, or the time to locoregional recurrence or death, whichever occurred first, or the time of the last follow-up for patients who did not have an event. Therefore, this analysis allowed the evaluation of post-treatment metrics to predict treatment response at three months, as well as disease relapse and death from any cause.

OS was defined as the time interval between the final fraction of CRT and the time to death from any cause, or the time to the last follow-up.

Exposure variables (predictors)

- Primary tumours and lymph nodes SUV_{max} and SUV_{peak} normalised by weight, LBM, and BSA obtained from overall HNSCC lesions.
- Primary tumours and lymph nodes SUV_{max} and SUV_{peak} normalised by weight, LBM, and BSA obtained from HPV^{+ve} oropharyngeal cancer.
- Primary tumours and lymph nodes SUV_{max} and SUV_{peak} normalised by weight, LBM, and BSA obtained from HPV^{-ve} HNSCC.
- The source of data was previously described in Section 3.1.

In Chapter 6, two analyses were conducted, as illustrated in Figure 3.10.

- The initial analysis investigated the prognostic value of post-treatment PET/CT SUV metrics derived from both primary tumour and cervical lymph node lesions, taken from a cohort comprising multiple anatomical HNSCC subsites, including oropharyngeal, laryngeal, and unknown HNSCC.

- The second analysis compared the prognostic value of PET/CT SUV metrics derived from both primary tumour and cervical lymph node lesions in a cohort comprising both HPV^{+ve} oropharyngeal cancer and HPV^{-ve} HNSCC. This was due to the distinct disease clinical characteristics and a unique molecular profile and outcome of HPV-positive-associated OPCC (Lechner et al., 2022) in comparison to other subtypes.

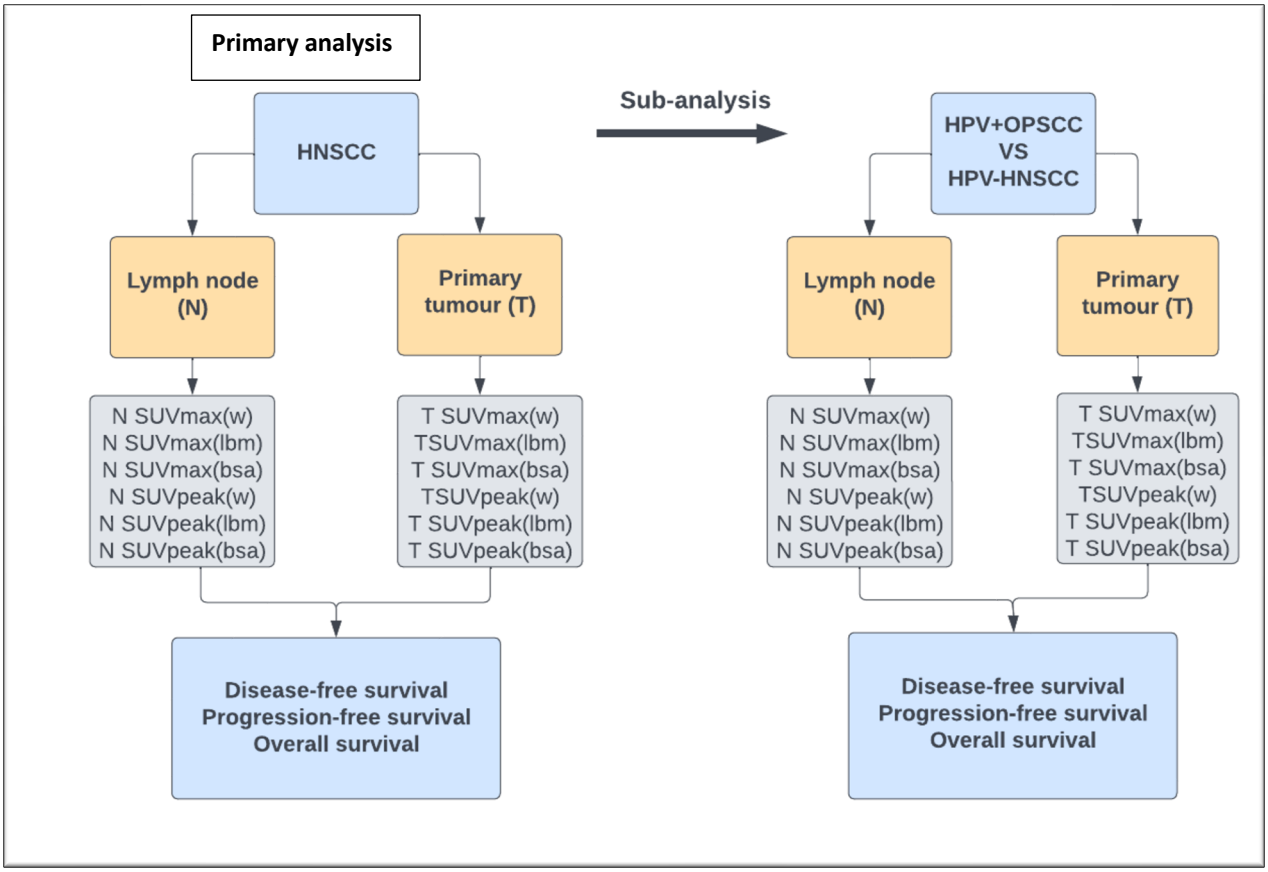


Figure 3.10: Flowcharts of the two analysis plans of Chapter 6.

The initial analysis investigated the prognostic value of post-treatment primary and lymph node lesions SUV derived from the whole cohort (different HNSCC subtypes) for predicting three-year PFS and five-year OS. A second analysis was performed to evaluate whether the prognostic value of these quantitative metrics differed when lesions were divided based on HPV status (HPV^{+ve} oropharyngeal cancer and HPV^{-ve} HNSCC).

Abbreviations: SUV_{max}, maximum standardised uptake value; SUVpeak, peak standardised uptake value; w, weight in kg; LBM, lean body mass; BSA, body surface area; HPV, human papilloma virus; OPSCC, oropharyngeal squamous cell carcinoma; HNSCC, head and neck squamous cell carcinoma.

The flowchart in Figure 3.11 summarises the statistical plan used in Chapter 6.

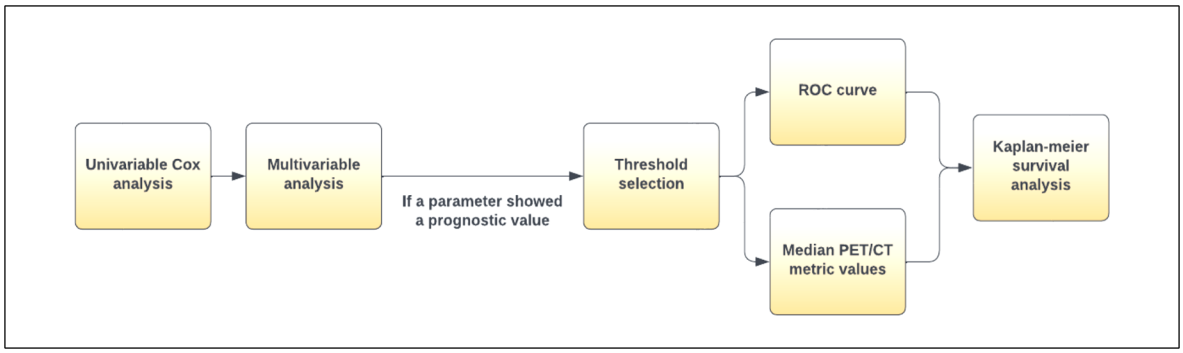


Figure 3.11: Statistical analysis plan to investigate the prognostic value of post-treatment PET/CT metrics.

In Chapter 6, we carried out univariable and multivariable Cox analysis tests, the ROC curve analysis, and the Kaplan-Meir survival analysis. First, Cox's proportional hazards model was used to conduct a univariable analysis on all post-treatment PET/CT metrics independently to identify their prognostic significance as potential risk factors for three-year PFS and five-year OS. In other words, the univariable Cox analysis was used to investigate the effect of a single variable (independent predictor), such as SUV_{max}, on the risk of an event (dependent outcome) over time. This event could be the risk of disease progression in the case of analysing the outcome of PFS or the risk of death in the case of analysing the outcome of OS.

Next, PET/CT parameters were also integrated into a multivariable Cox's proportional hazards model. This multivariable analysis was used to investigate the influence of multiple combined predictors on outcomes. It considered some potential confounding factors and provided an adjusted risk analysis. These factors included older age, tumour stage, HPV status, and locoregional response

to treatment (LRT). LRT indicated a complete or incomplete tumour response to CRT at three months post-treatment PET/CT.

Also, as stated in Chapter 1, a prevalent risk factor for HNSCC is exposure to HPV. Studies have shown that more than 70% of oropharyngeal cancers are associated with this virus (Johnson et al., 2020). However, little data exist on how HPV status affects the predictive power of SUVs (Helsen et al., 2020a, Connor et al., 2021). Therefore, a subgroup analysis was conducted to evaluate the potential predictive power of SUV metrics when patients were analysed based on HPV positive and negative status. In OPSCC patients with unknown HPV status, those patients were excluded from the sub-analysis (15 patients).

In the Cox proportional hazards model, SUV metrics and age variables were analysed as continuous variables. T stage and N stage were analysed as ordinal variables. LRT was analysed as a binary variable.

Also, LRT was excluded from the analysis of PFS because of the similarity in the definitions of the predictor (LRT) and the outcome PFS. As previously mentioned, disease progression included residual disease at three months, which represented a complete tumour response at three months post-treatment PET/CT imaging (LRT).

In this study, Cox proportional hazards model models served as the foundation upon which all subsequent analyses were based. This was carried out to enhance the accuracy of the prognostic analysis by preserving the PET/CT metrics as continuous variables in the Cox model.

If a parameter showed significance in the multivariable analysis, Kaplan–Meier curves with the log-rank test were used to compare survival outcomes between patients with high and low SUV_{max} and SUV_{peak} thresholds.

A critical point to note was that even though converting a continuous variable, such as SUV_{max} to a binary variable seemed inappropriate, keeping it as a continuous variable would render its value in clinical practice obsolete. Therefore, to maximise the utility of PET/CT metrics in the clinical setting as potential predictors of survival, and as no consensus could be identified from the literature on the best approach for selecting specific thresholds for prognostication purposes, SUV thresholds were determined based on two different methods:

1) The first approach was to use the ROC curve coordinate table and select a threshold based on sensitivity and specificity. This method has been utilised repeatedly in numerous publications (Sagardoy et al., 2016, Oyama et al., 2020, Wang et al., 2020). Our selected threshold was based on a minimal specificity of 80% (Helsen et al. 2020).
2) The second approach was to use the median values of PET/CT parameters to establish thresholds for the survival analyses. This method has been used in similar studies evaluating quantitative MRI factors in post-treatment HNSCC (Connor et al., 2021) and non-small-cell lung cancer PET/CT imaging factors (Shin et al., 2017). In this study, patient survival was compared using both methods.

- The reverse Kaplan-Meier approach was used to estimate the median follow-up
- Patients with insufficient follow-up information were excluded. Therefore, Four out of 124 patients were excluded

3.9.4 Sensitivity analysis

A few patients (8/120) were scanned using a different scanner. Different factors, including slice thickness, attenuation correction method, and reconstruction algorithm, may be implemented when utilising two different scanners, which may influence SUV readings. However, a sample sensitivity analysis was performed. Our analysis confirmed the insignificant difference in SUV readings when these patients (8 patients) were removed, suggesting the minimal effect of using the two different scanners in our analysis. The results of the sensitivity analysis are shown in Appendix 6. Notably, the eight patients were not excluded from the analysis in order to prevent the possibility of selection bias.

3.10 SUV quality control/assurance

To ensure the accuracy of quantitative analysis for clinical use, the EANM Guidelines (2015) defined an expected normal range for FDG PET/CT metrics, as follows:

- For the liver: SUV_{mean} (1.3 to 3.0) and SUV_{lbm} (1.0 to 2.2)
- For the BP SUV_{mean} around 1.6 and 1.2 for SUV_{lbm}

Values above or below this range could suggest inaccuracies in FDG dose delivery or technical difficulties (Westerterp et al., 2008).

The mean values of the liver $SUV_{mean(w)}$, liver $SUV_{mean(lbm)}$, BP $SUV_{mean(w)}$, and BP $SUV_{mean(lbm)}$ investigated in this thesis were all within the EANM-defined range (Table 3.3). Thus, this conformity may ensure that the correct protocol was followed.

Table 3.3: Mean values of the liver and BP $SUV_{mean(w)}$ and $SUV_{mean(lbm)}$

Background region	$SUV_{mean(w)}$	$SUV_{mean(lbm)}$
Liver	2.1	1.7
Blood pool	1.7	1.3

CHAPTER 4

CORRELATION OF POST CHEMORADIOTHERAPY FDG PET/CT PARAMETERS WITH BODY SIZE MEASUREMENTS IN PATIENTS WITH HEAD AND NECK SQUAMOUS CELL CARCINOMA

4. Chapter 4: Correlation of FDG PET/CT Measurements with Body Metrics

4.1 Introduction

As described in Chapter 1, quantitative PET/CT parameters are sensitive to several factors, including the type of body size measures utilized for normalisation (Ziai et al., 2016). In Chapter 1, we found that several studies have recommended the use of LBM or BSA for SUV calculations derived from normal background regions (Kim et al., 1994, Sanghera et al., 2009, Keramida and Peters, 2019, Sarikaya et al., 2020); however, a limited number of studies have evaluated the effect of different normalisation methods obtained from malignant lesions (Hallett et al., 2001). Little is also known about the association between different body size metrics and volumetric parameters, especially in the context of post-treatment settings in patients diagnosed with HNSCC.

Thus, the purpose of this chapter was:

1) To investigate the association of HNSCC lesions SUV_{max}, $SUV_{peak,}$ MTV, and TLG normalised by different body measures with patient body size measurements to identify the normalisation method that is least sensitive to these body factors when reporting post-treatment PET/CT parameters in patients with HNSCC.
2) To explore whether this association differed when SUV metrics were derived from the normal reference organ of the liver (liver SUVs) versus HNSCC SUVs acquired from the same patients: sub-analysis.

4.2 Materials and methods

Patients, scanning protocols, image analysis, and quantitative metrics were previously described in Chapter 3.

The statistical methods used were described in Chapter 3, Section 3.9.1.

4.3 Results

4.3.1 Patient Characteristics:

A total of 69 patients, including 52 (75.3%) males and 17 (24.6%) females with histologically proven HNSCC with a median age of 60.1 years (IQR: 54.7–66.2 years), were included. Patients had a median height of 171.0 cm (166.5-177.9 cm), weight of 67.0 kg (55.0-80.0 kg), LBM of 53.1 kg (44.7-59.6 kg), BSA of 1.8 m^2 (1.6-1.9 m^2), and BMI of 27.0 kg/m^2 (18.0-37.0 kg/m^2). Patient and tumour characteristics are tabulated in Table 4.1.

In total, 108 measurable HNSCC lesions were analysed (45 primary tumours and 63 involved lymph nodes). Specifically, SUV_{max} of 108 lesions and SUV_{peak} of 107 lesions were studied (SUV_{peak} of one lesion was not possible to derive because of the small size of the lesion). MTV and TLG of 73 lesions with clear boundaries were also assessed.

The HPV status was determined by examining p16. It is a cheap, reliable test and is the standard for clinical evaluation of HPV positivity during the time we conducted our research.

Table 4.1: Patient and tumour characteristics

Characteristics	**n=69**
Age, y, median (IQR)	60.1(54.7–66.2)
Gender, n (%)	
Male	52(75.3)
Female	17(24.6)
Body size measurements, median (IQR)	
Height	171.0(166.5-177.9)
Weight	67.0(55.0-80.0)
LBM	53.1(44.7-59.6)
BSA	1.8(1.6-1.9)
BMI	27.0(18.0-37.0)
Tumour location, n (%)	
Oropharynx	46(66.6)
Larynx	6(8.7)
Oropharynx + Larynx	2(2.8)
Hypopharynx	6(8.7)
Occult primary	9(13.0)
HPV status*, n (%)	
Positive	36(63.2)
Negative	15(26.3)
Not recorded	6(10.5)
T stage, n (%)	
T1	9(13.0)
T2	21(30.4)
T3	12(17.3)
T4	25(36.2)
Not recorded	2(2.9)
N stage, n (%)	
N0	3(4.3)
N1	4(5.8)
N2	60(86.9)
Not recorded	2(2.9)
Treatment, n (%)	
CRT	61(88.5)
RT only	8(11.6)

* HPV status of oropharynx, oropharynx+ larynx, and occult primary cancers.

Abbreviations: IQR, interquartile range; LBM, lean body mass; BSA, body surface area; BMI, body mass index; HPV, human papillomavirus; T, tumour, N, lymph node; CRT, chemoradiotherapy; RT, radiotherapy.

4.3.2 Lesion SUV_{max}

The median (IQR) values of post-treatment HNSCC lesions $SUV_{max(w)}$, $SUV_{max(lbm)}$, $SUV_{max(bsa)}$ were 4.6 (3.2-6.4), 3.5 (2.4-5.2), and 1.2 (0.8-1.7), respectively (Table 4.2).

SUV_{max} normalised by weight did not statistically significantly correlate with patient height (*r*=-0.010, *P*=0.920), weight (*r*=0.067, *P*=0.489), LBM (*r*=0.063, *P*=0.515), BSA (*r*=0.055, *P*=0.569), and BMI (*r*=0.060, *P*=0.539) (Table 4.3 and Figure 4.1).

Similarly, this was true between SUV_{max} normalised by LBM with patient height (*r*=0.111, *P*=0.254), weight (*r*=-0.102, *P*=0.294), LBM (*r*=0.075, *P*=0.442), BSA (*r*=-0.067, *P*=0.488), and BMI (*r*=-0.120, *P*=0.215) (Figure 4.2).

Data analysis also showed that there was a weak relationship between SUV_{max} normalised by BSA with weight (*r*=-0.204, *P*=0.034), BSA (*r*=-0.191, *P*=0.048), and BMI (*r*=-0.214, *P*=0.026). However, no correlation was seen between SUV_{max} normalised by BSA with height (*r*=-0.035, *P* =0.721) and LBM (*r*=-0.103, *P*=0.287) (Table 4.3 and Figure 4.3).

Overall, all post-treatment HNSCC lesions SUV_{max} showed no or weak correlation with different body size measurements.

4.3.3 Lesion SUV_{peak}

The median (IQR) values of post-treatment HNSCC $SUV_{peak(w)}$, $SUV_{peak(lbm)}$, and $SUV_{peak(bsa)}$ were 3.4 (2.4-4.5), 2.6 (1.9-3.5), and 0.94 (0.7-1.2), respectively (Table 4.2).

When SUV_{peak} indices were correlated with body size metrics, similar findings to SUV_{max} were observed. For instance, no statistically significant correlation was observed between $SUV_{peak(w)}$ with patient height (*r*=-0.016, *P*=0.869), weight (*r*=0.052, *P*=0.596), LBM (*r*=0.044, *P*=0.650), BSA (*r*= 0.042, *P*=0.666), and BMI (*r*=0.046, *P*=0.635) (Table 4.3 and Figure 4.4). Similarly, a lack of significant

correlation was found between $SUV_{peak(lbm)}$ with patient height (r=0.105, P =0.280), weight (r=-0.141, P=0.148), LBM (r=0.046, P=0.641), BSA (r=-0.100, P=0.305), and BMI (r=-0.159, P=0.102) (Figure 4.5). In addition, data analysis revealed weak correlations between $SUV_{peak(bsa)}$ with weight (r=-0.266, P=0.006), BSA (r=-0.246, P=0.011), and BMI (r=-0.276, P=0.004). No correlation was observed between $SUV_{peak(bsa)}$ with height (r=-0.050, P=0.607) and LBM (r=-0.152, P=0.118) (Table 4.3 and Figure 4.6).

Overall, all post-treatment lesions SUV_{peak} exhibited either no or weak correlation with different body size metrics.

4.3.4 Lesion TLG

The median (IQR) values for post-treatment HNSCC $TLG_{(w)}$, $TLG_{(lbm)}$, $TLG_{(bsa)}$ were 2.9 (1.4-5.1), 2.3 (1.1-4.2), and 0.8 (0.4-1.4), respectively (Table 4.2).

The correlation analyses revealed that TLG normalised to weight exhibited no statistically significant correlation with height (r=0.069, P=0.562), weight (r=0.032, P=0.790), LBM (r=0.063, P=0.597), BSA (r=0.051, P=0.665), and BMI (r=0.018, P=0.880) (Table 4.3 and Figure 4.7).

Furthermore, no correlation was also observed when TLG normalised to LBM was correlated with all the variables, including height (r=0.101, P=0.397), weight (r=-0.030, P=0.800), LBM (r=0.054, P=0.649), BSA (r=-0.009, P=0.940), and BMI (r=-0.050, P=0.674) (Table 4.3 and Figure 4.8).

Similarly, TLG normalised to BSA was not correlated with height (r=0.018, P=0.880), weight (r=-0.102, P=0.389), LBM (r=-.039, P=0.741), BSA (r=-0.074, P=0.534), and BMI (r=-0.117, P=0.322) (Table 4.3 and Figure 4.9).

Overall, all post-treatment lesions TLG showed no correlation with different body size metrics.

4.3.5 Lesion MTV

The median (IQR) value of post-treatment HNSCC lesions MTV was 0.7 (0.5-1.1) (Table 4.2).

Similar to post-treatment TLG measurements, lesions MTV was not correlated with any body size metrics, including height (*r*=0.071, *P*=0.550), weight (*r*=0.045, *P*=0.704), LBM (*r*=0.064, *P*=0.592), BSA (*r*=0.067, *P*=0.574), and BMI (*r*=0.036, *P*=0.763) (Table 4.3 and Figure 4.10).

4.3.6 Background liver SUV_{mean} and SUV_{max}

The median (IQR) values of liver SUV_{max} and SUV_{mean} were 3.0 (2.6-3.3) and 2.2 (2.0-2.4), respectively (Table 4.2).

Liver SUV_{mean} was moderately and statistically significantly correlated with weight (*r*=0.544, *P*=<0.001), BSA (*r*=-0.462, *P*=<0.001), and BMI (*r*=0.563, *P*=<0.001), while weakly correlated with LBM (*r*=0.267, *P*=0.005) (Table 4.3 and Figure 4.11).

Similarly, liver SUV_{max} showed moderate statistically significant correlations with weight (*r*=-0.585, *P*=<0.001), BSA (*r*=0.510, *P*=<0.001), and BMI (*r*=0.598, *P*=<0.001), while only a weak correlation was observed between liver SUV_{max} with LBM (*r*=0.325, *P*=<0.001) (Table 4.3 and Figure 4.12).

Therefore, the weak correlation between liver SUV metrics and LBM indicates that the use of LBM as a normalisation factor for analysing liver SUV metrics appears to be the most suitable body size measurement compared to total body weight and BSA.

Some of the scatter plots might show outliers, which may be attributable to lesions with extremely high or low FDG uptake relative to the rest, or to the location of the ROI/VOI being different on the negative PET/CT scans (scans lacking visible FDG uptake).

Table 4.2: The median values of post-treatment metrics

Parameter	**Median (IQR)**
Lesion SUVmax(w)	4.6 (3.2-6.4)
Lesion SUVmax(lbm)	3.5 (2.4-5.2)
Lesion SUVmax(bsa)	1.2 (0.8-1.7)
Lesion SUVpeak(w)	3.4 (2.4-4.5)
Lesion SUVpeak(lbm)	2.6 (1.9-3.5)
Lesion SUVpeak(bsa)	0.94 (0.7-1.2)
Lesion TLG(w)	2.9 (1.4-5.1)
Lesion TLG(lbm)	2.3 (1.1-4.2)
Lesion TLG(bsa)	0.8 (0.4-1.4)
Lesion MTV	0.7 (0.5-1.1)
Liver SUVmean(w)	2.2 (2.0-2.4)
Liver SUVmax(w)	3.0 (2.6-3.3)

Abbreviations: IQR, interquartile range; SUVmax, maximum standardised uptake value; SUVpeak, peak standardised uptake value, TLG, total lesion glycolysis, MTV, metabolic tumour volume; w, weight in kg; LBM, lean body mass; BSA, body surface area.

Table 4.3: Spearman's correlation analysis results of the relationships between head and neck cancer lesion maximum and peak SUV, TLG, and MTV, as well as the liver SUVmean and SUVmax with several body size measurements.

Spearman's correlation

			Height (cm)	Weight (kg)	LBM (kg)	BSA (m^2)	BMI
Spearman's rho	Height (cm)	Correlation Coefficient	1.000	.377**	.738**	.568**	.305**
		Sig. (2-tailed)	.	<.001	<.001	<.001	.001
		N	108	108	108	108	108
	Weight (kg)	Correlation Coefficient	.377**	1.000	.799**	.964**	.993**
		Sig. (2-tailed)	<.001	.	<.001	<.001	<.001
		N	108	108	108	108	108
	LBM (kg)	Correlation Coefficient	.738**	.799**	1.000	.883**	.763**
		Sig. (2-tailed)	<.001	<.001	.	<.001	<.001
		N	108	108	108	108	108
	BSA (m^2)	Correlation Coefficient	.568**	.964**	.883**	1.000	.938**
		Sig. (2-tailed)	<.001	<.001	<.001	.	<.001
		N	108	108	108	108	108
	BMI	Correlation Coefficient	.305**	.993**	.763**	.938**	1.000
		Sig. (2-tailed)	.001	<.001	<.001	<.001	.
		N	108	108	108	108	108
	Lesion SUVmax (W)	Correlation Coefficient	-.010	.067	.063	.055	.060
		Sig. (2-tailed)	.920	.489	.515	.569	.539
		N	108	108	108	108	108
	Lesion SUVmax (LBM)	Correlation Coefficient	.111	-.102	.075	-.067	-.120
		Sig. (2-tailed)	.254	.294	.442	.488	.215
		N	108	108	108	108	108
	Lesion SUVmax (BSA)	Correlation Coefficient	-.035	-.204*	-.103	-.191*	-.214*
		Sig. (2-tailed)	.721	.034	.287	.048	.026
		N	108	108	108	108	108
	Lesion SUVpeak (W)	Correlation Coefficient	-.016	.052	.044	.042	.046
		Sig. (2-tailed)	.869	.596	.650	.666	.635
		N	107	107	107	107	107
	Lesion SUVpeak (LBM)	Correlation Coefficient	.105	-.141	.046	-.100	-.159
		Sig. (2-tailed)	.280	.148	.641	.305	.102
		N	107	107	107	107	107
	Lesion SUVpeak (BSA)	Correlation Coefficient	-.050	-.266**	-.152	-.246*	-.276**
		Sig. (2-tailed)	.607	.006	.118	.011	.004
		N	107	107	107	107	107
	Lesion TLG (W)	Correlation Coefficient	.069	.032	.063	.051	.018
		Sig. (2-tailed)	.562	.790	.597	.665	.880
		N	73	73	73	73	73
	Lesion TLG (LBM)	Correlation Coefficient	.101	-.030	.054	.009	-.050
		Sig. (2-tailed)	.397	.800	.649	.940	.674
		N	73	73	73	73	73
	Lesion TLG (BSA)	Correlation Coefficient	.018	-.102	-.039	-.074	-.117
		Sig. (2-tailed)	.880	.389	.741	.534	.322
		N	73	73	73	73	73
	Lesion MTV	Correlation Coefficient	.071	.045	.064	.067	.036
		Sig. (2-tailed)	.550	.704	.592	.574	.763
		N	73	73	73	73	73
	Liver SUVmean(W)	Correlation Coefficient	-.071	.544**	.267**	.462**	.563**
		Sig. (2-tailed)	.465	<.001	.005	<.001	<.001
		N	108	108	108	108	108
	Liver SUVmax(W)	Correlation Coefficient	-.025	.585**	.325**	.510**	.598**
		Sig. (2-tailed)	.798	<.001	<.001	<.001	<.001
		N	108	108	108	108	108

*. Correlation is significant at the 0.05 level (2-tailed).

The relationship between lesion SUV_{max} normalised by weight in kg ($SUV_{max(w)}$) with patient height, weight(kg), LBM, BSA, and BMI

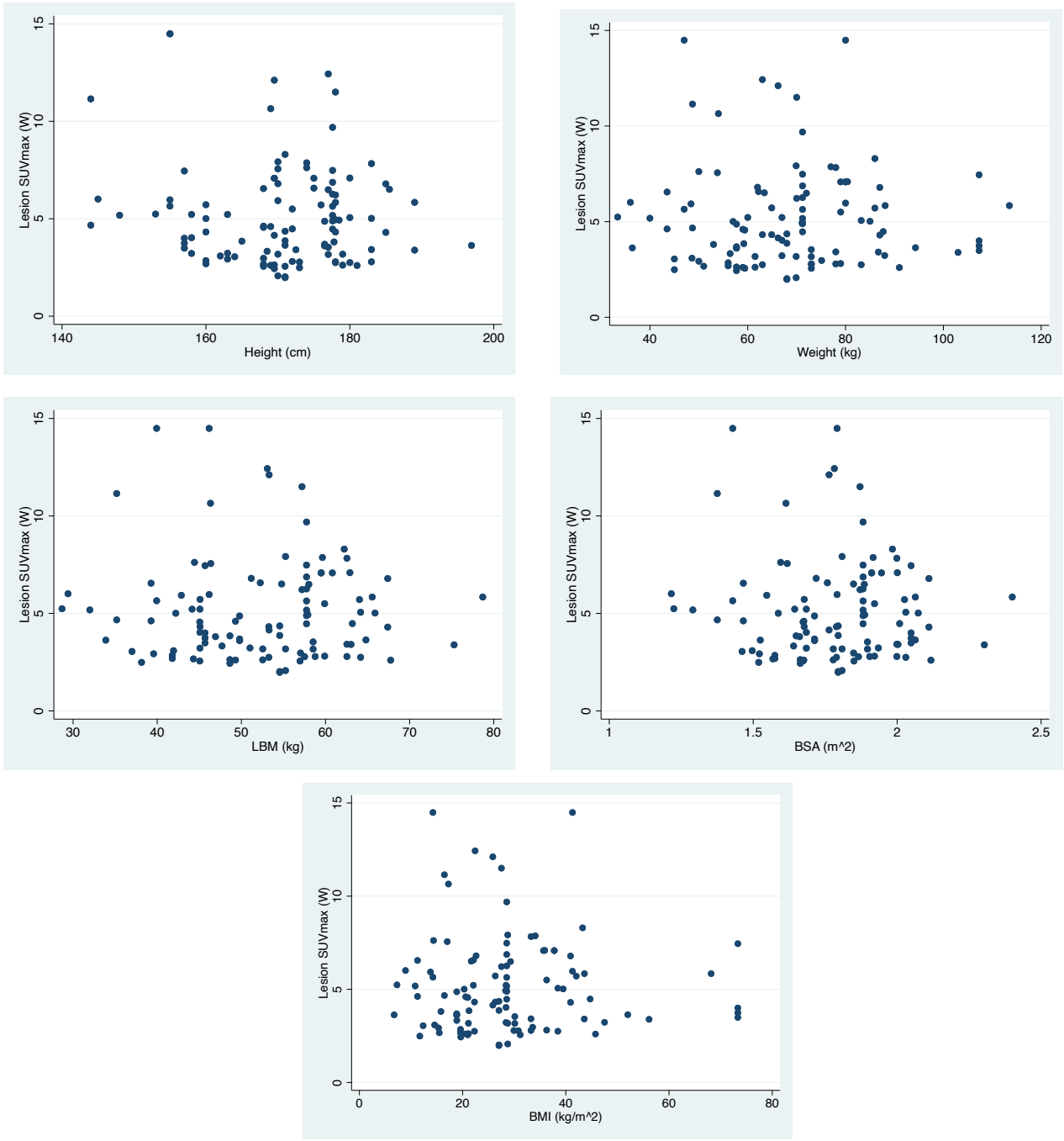


Figure 4.1: Scatter plots of lesion SUVmax(w) and patient factors.
The plots of the lesions maximum standardised uptake value normalised by weight in kg (SUVmax(w)) with patient height (cm), weight (kg), LBM (kg), BSA (m^2), and BMI show no correlation with any of the variables.

The relationship between lesion SUV_{max} normalised by LBM ($SUV_{max(lbm)}$) with patient height, weight(kg), LBM, BSA, and BMI

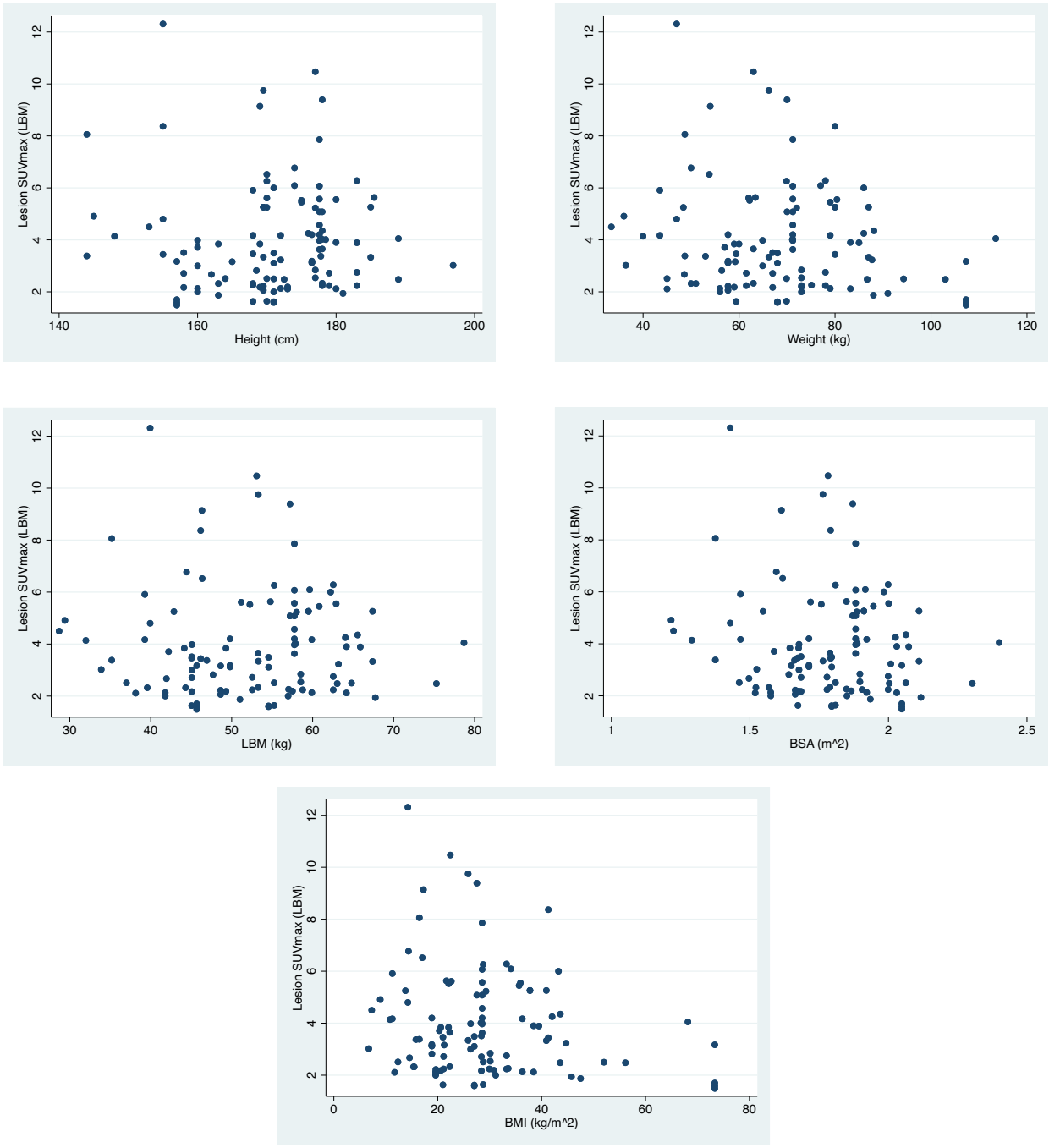


Figure 4.2: Scatter plots of lesion SUVmax(lbm) and patient factors.
The plots of lesions maximum standardised uptake value normalised by LBM (SUVmax(lbm)) with patient height (cm), weight (kg), LBM (kg), BSA (m^2), and BMI show no correlation with any of the variables.

The relationship between lesion SUV_{max} normalised by BSA ($SUV_{max(bsa)}$) with patient height, weight(kg), LBM, BSA, and BMI

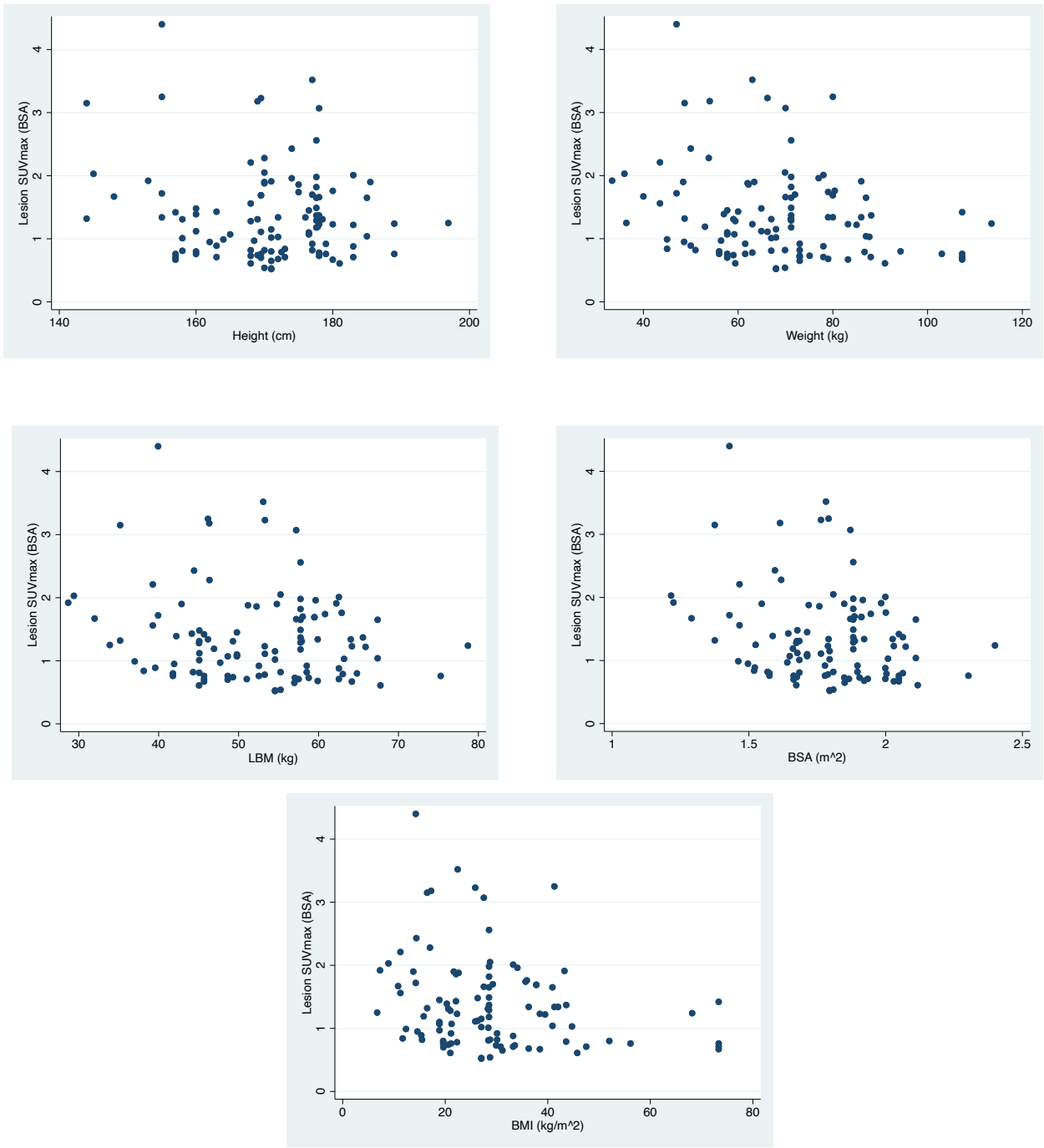


Figure 4.3: Scatter plots of lesion SUVmax(bsa) and patient factors.
The plots of the lesions maximum standardised uptake value normalised by BSA (SUVmax(bsa)) with patient height (cm), weight(kg), LBM (kg), BSA (m^2), and BMI show no correlation with any of the variables.

The relationship between lesion SUV_{peak} normalised by weight in kg ($SUV_{peak(w)}$) with patient height, weight(kg), LBM, BSA, and BMI

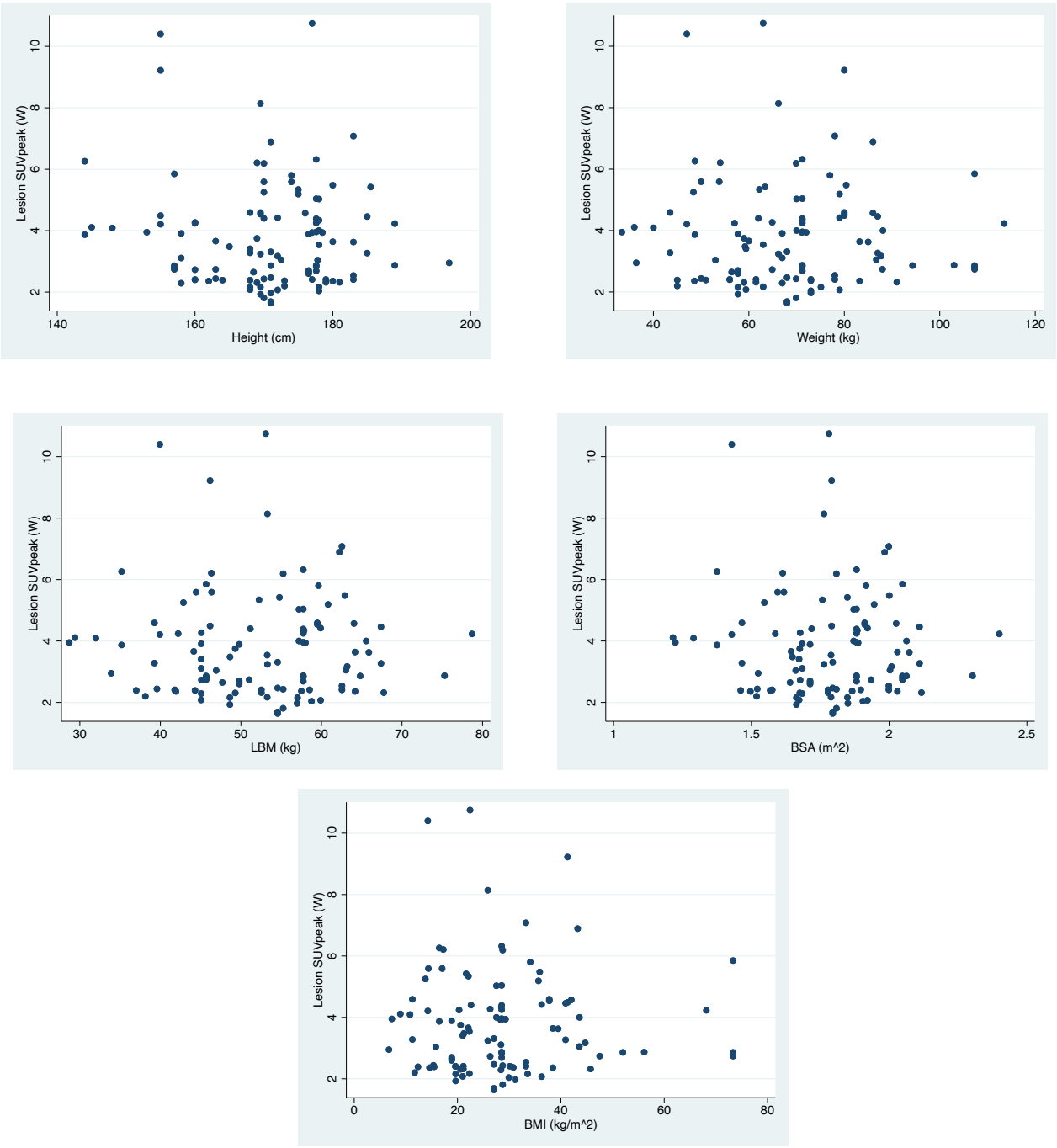


Figure 4.4: Scatter plots of lesion SUVpeak(w) and patient factors.

The plots of lesions peak standardised uptake value normalised by weight in kg (SUVpeak(w)) with patient height (cm), weight (kg), LBM (kg), BSA (m^2), and BM show no correlation with any of the variables.

The relationship between lesion SUV_{peak} normalised by LBM in kg ($SUV_{peak(lbm)}$) with patient height, weight(kg), LBM, BSA, and BMI

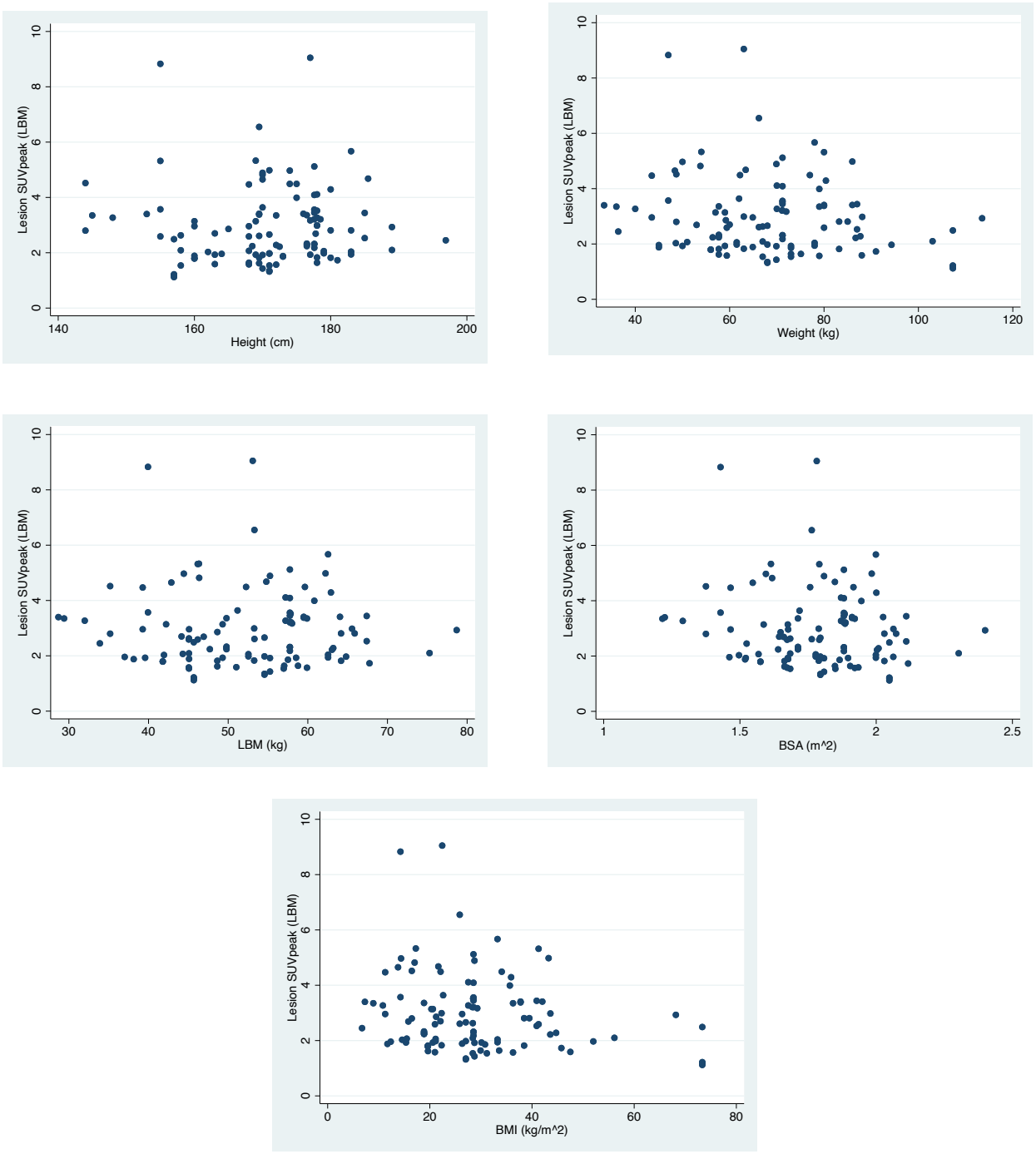


Figure 4.5: Scatter plots of lesion SUVpeak(lbm) and patient factors.

The plots of lesions peak standardised uptake value normalised by LBM (SUVpeak(lbm)) with patient height(cm), weight (kg), LBM (kg), BSA(m^2), and BMI show no correlation with any of the variables.

The relationship between lesion SUV_{peak} normalised by BSA in m^2 ($SUV_{peak(bsa)}$) with patient height, weight(kg), LBM, BSA, and BMI

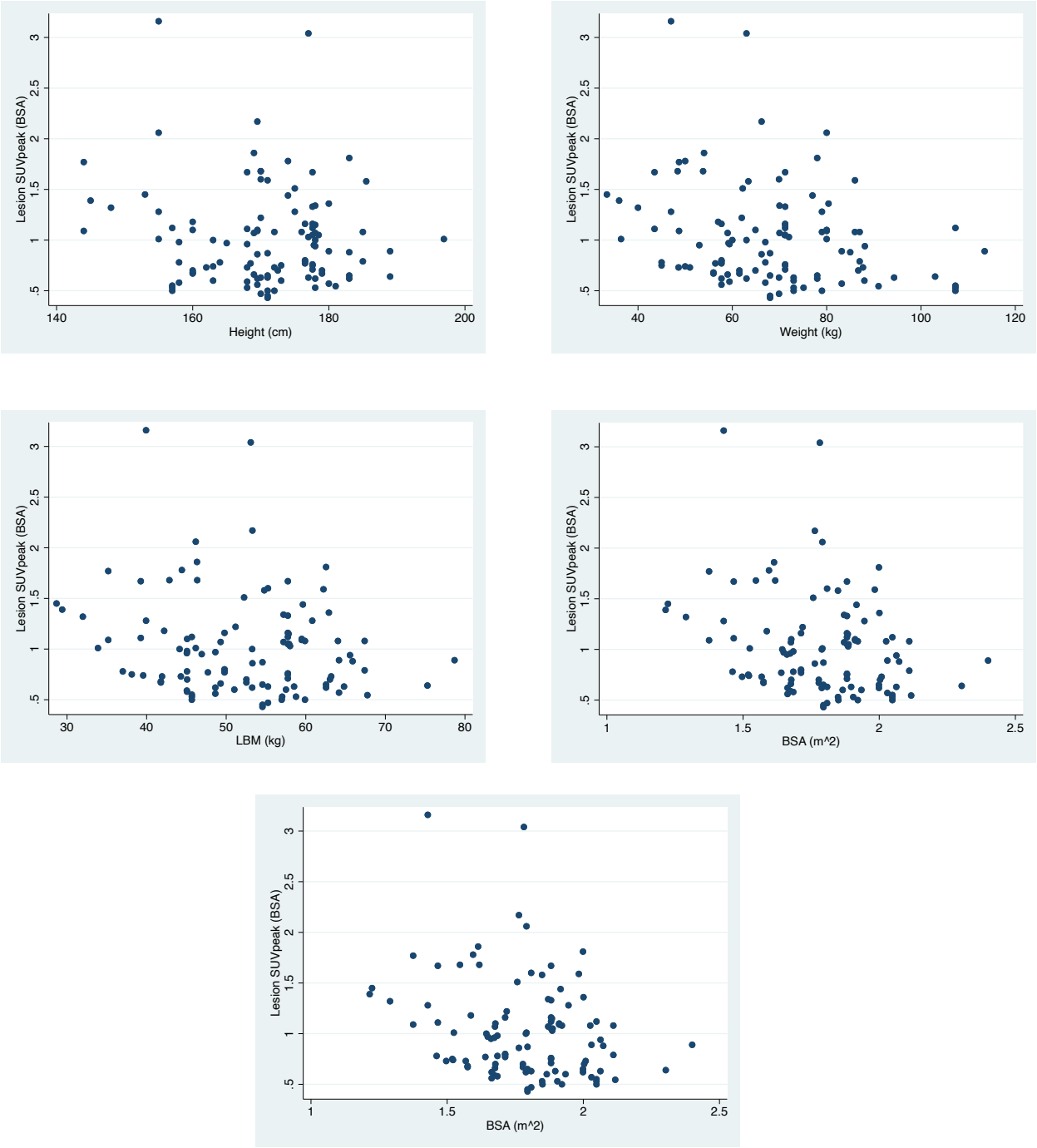


Figure 4.6: Scatter plots of lesion SUVpeak(bsa) and patient factors.
The plots for correlating the lesions peak standardised uptake value normalised by BSA (SUVpeak(bsa)) with patient height (cm), weight (kg), LBM (kg), BSA (m^2), and BMI show no correlation with any of the variables.

The relationship between lesion TLG normalised by weight in kg (TLG_w) with patient height, weight(kg), LBM, BSA, and BMI

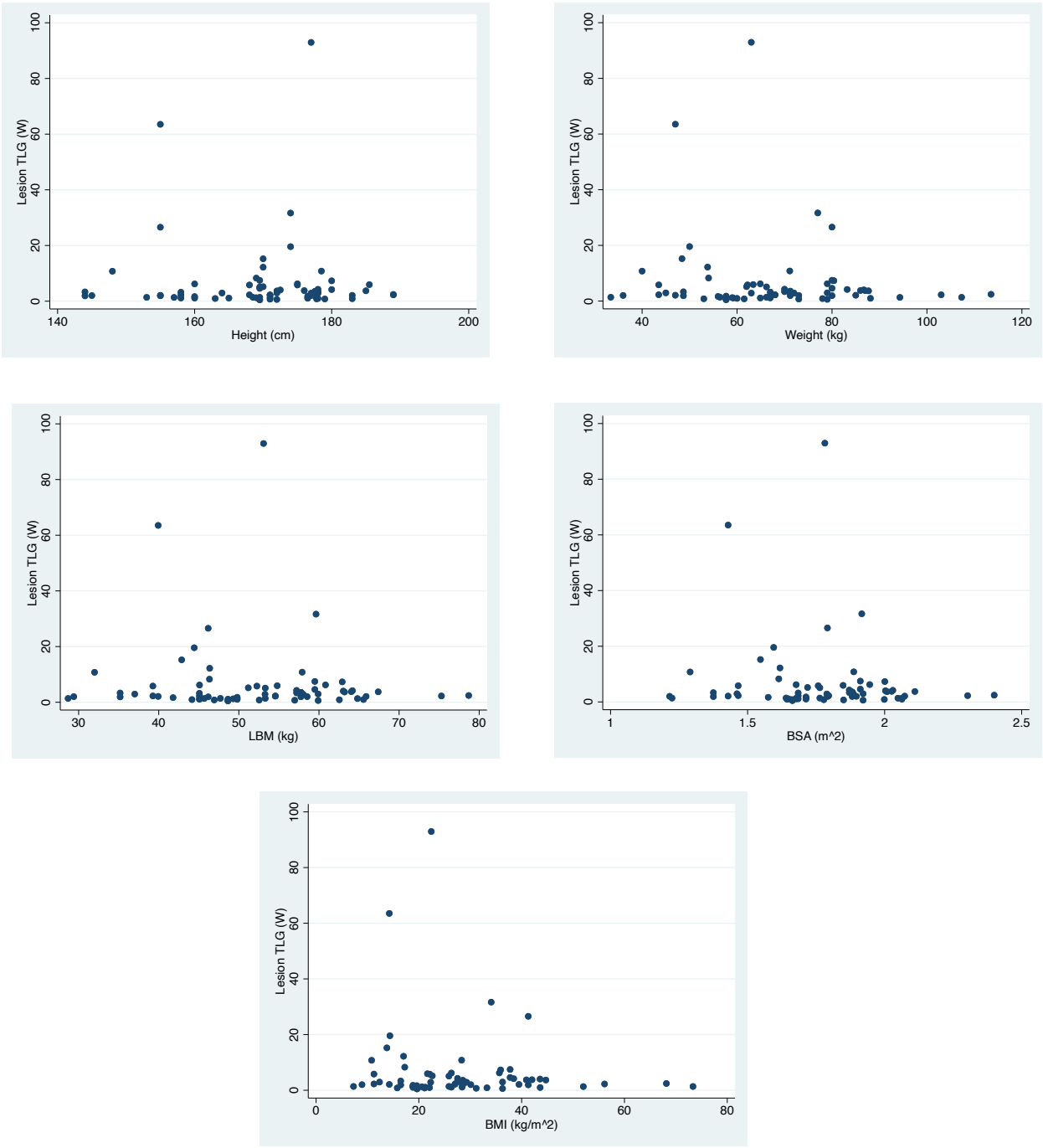


Figure 4.7: Scatter plots of lesion TLG(w) and patient factors.
The plots for correlating the lesions TLG normalised by weight in kg (TLGw) with patient height (cm), weight (kg), LBM (kg), BSA (m^2), and BMI show no correlation with any of the variables.

The relationship between lesion TLG normalised by LBM in kg (TLG_{lbm}) with patient height, weight(kg), LBM, BSA, and BMI

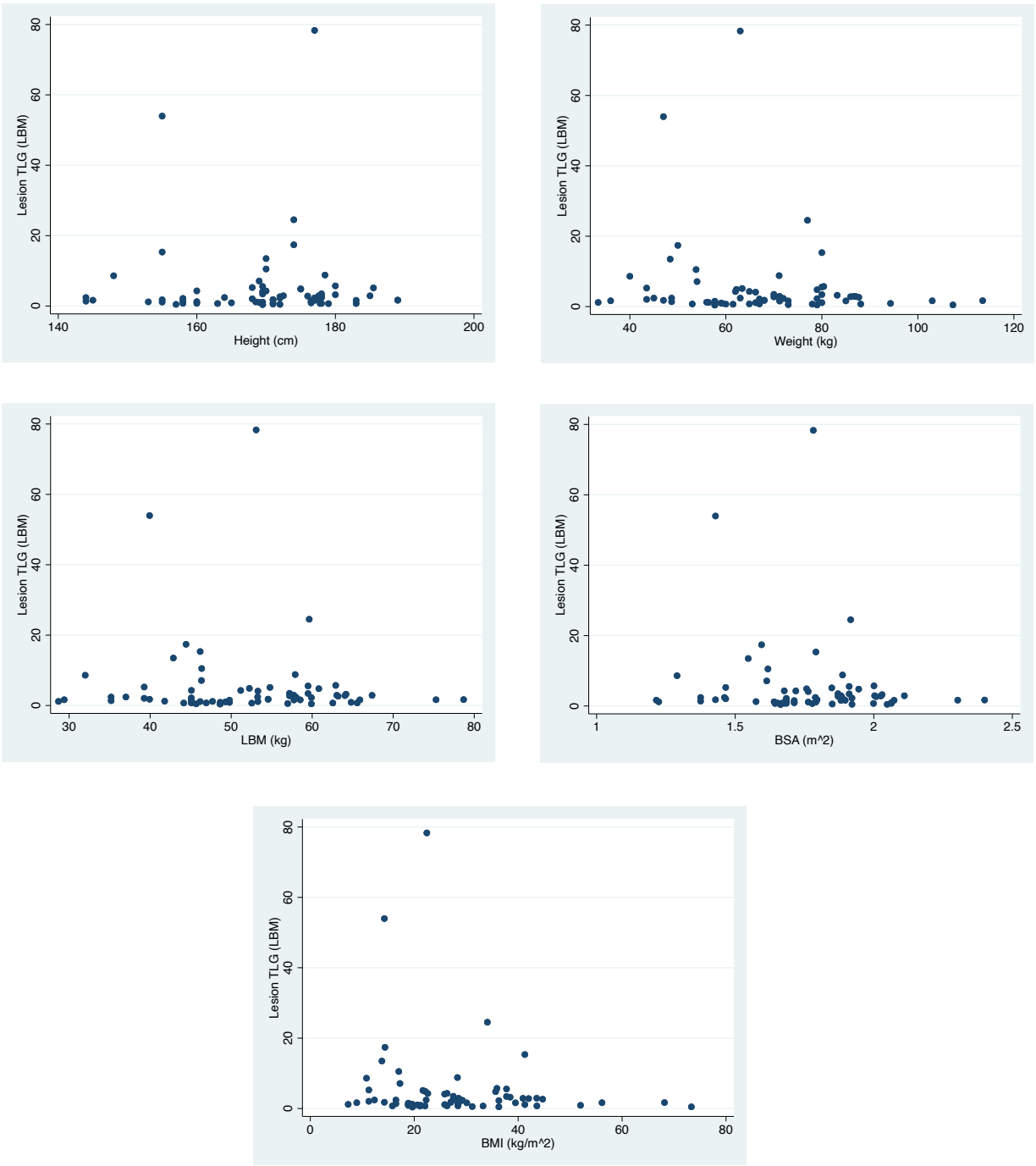


Figure 4.8: Scatter plots of lesion TLG(lbm) and patient factors.
The plots analysis for correlating the lesions TLG normalised by LBM (TLGlbm) with patient height (cm), weight (kg), LBM (kg), BSA (m^2), and BMI show no correlation with any of the variables.

The relationship between lesion TLG normalised by BSA in m^2 (TLG_{bsa}) with patient height, weight(kg), LBM, BSA, and BMI

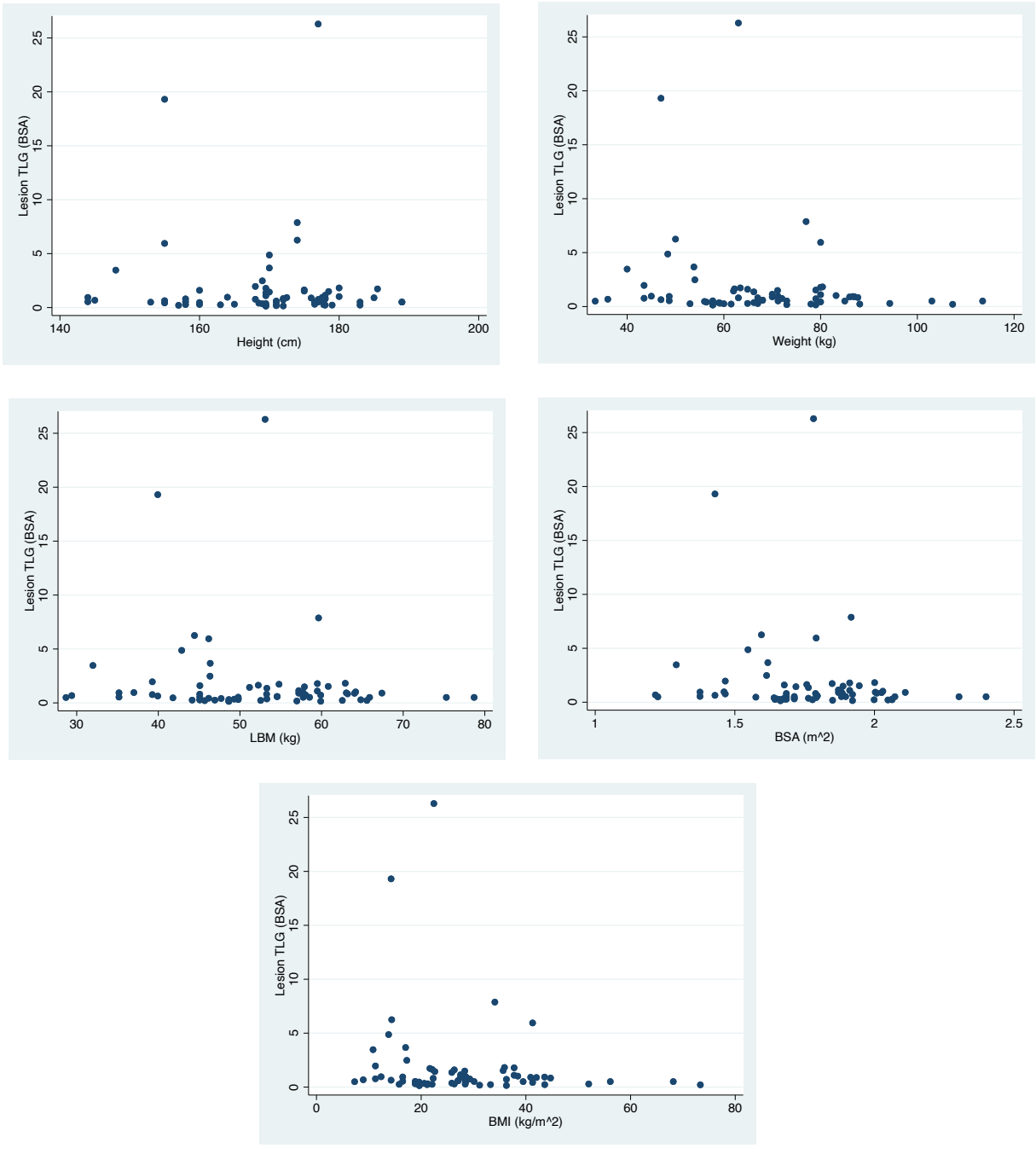


Figure 4.9: Scatter plots of lesion TLG(bsa) and patient factors.

The plots for correlating the lesions TLG normalised by BSA (TLGbsa) with patient height (cm), weight (kg), LBM (kg), BSA (m^2), and BMI show no correlation with any of the variables.

The relationship between lesion MTV with patient height, weight(kg), LBM, BSA, and BMI

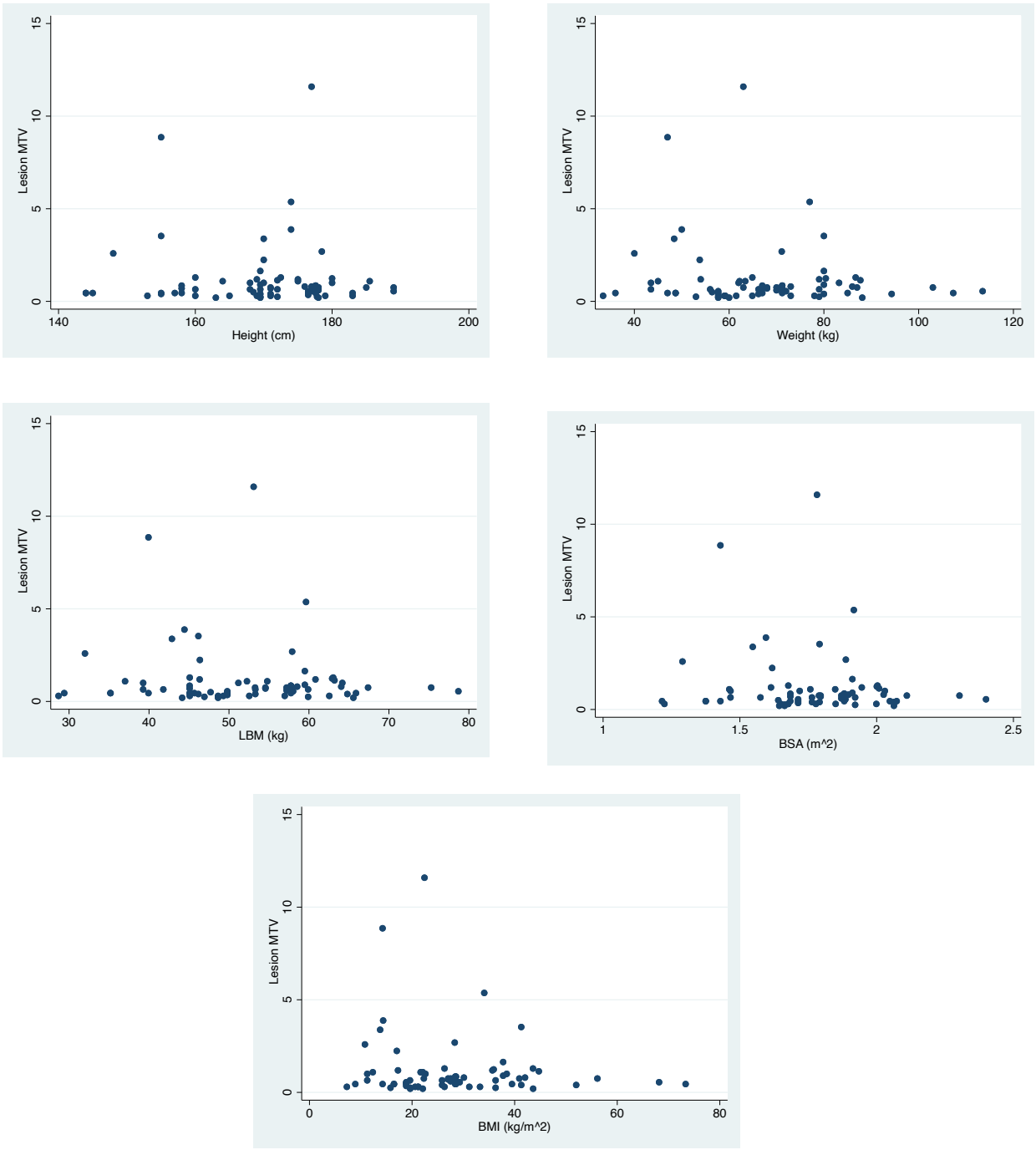


Figure 4.10: Scatter plots of lesion MTV and patient factors.
The plots for correlating the lesions MTV with patient height(cm), weight (kg), LBM (kg), BSA (m^2), and BMI show no correlation with any of the variables.

The relationship between liver SUV_{mean} normalised by weight in kg with patient height, weight(kg), LBM, BSA, and BMI

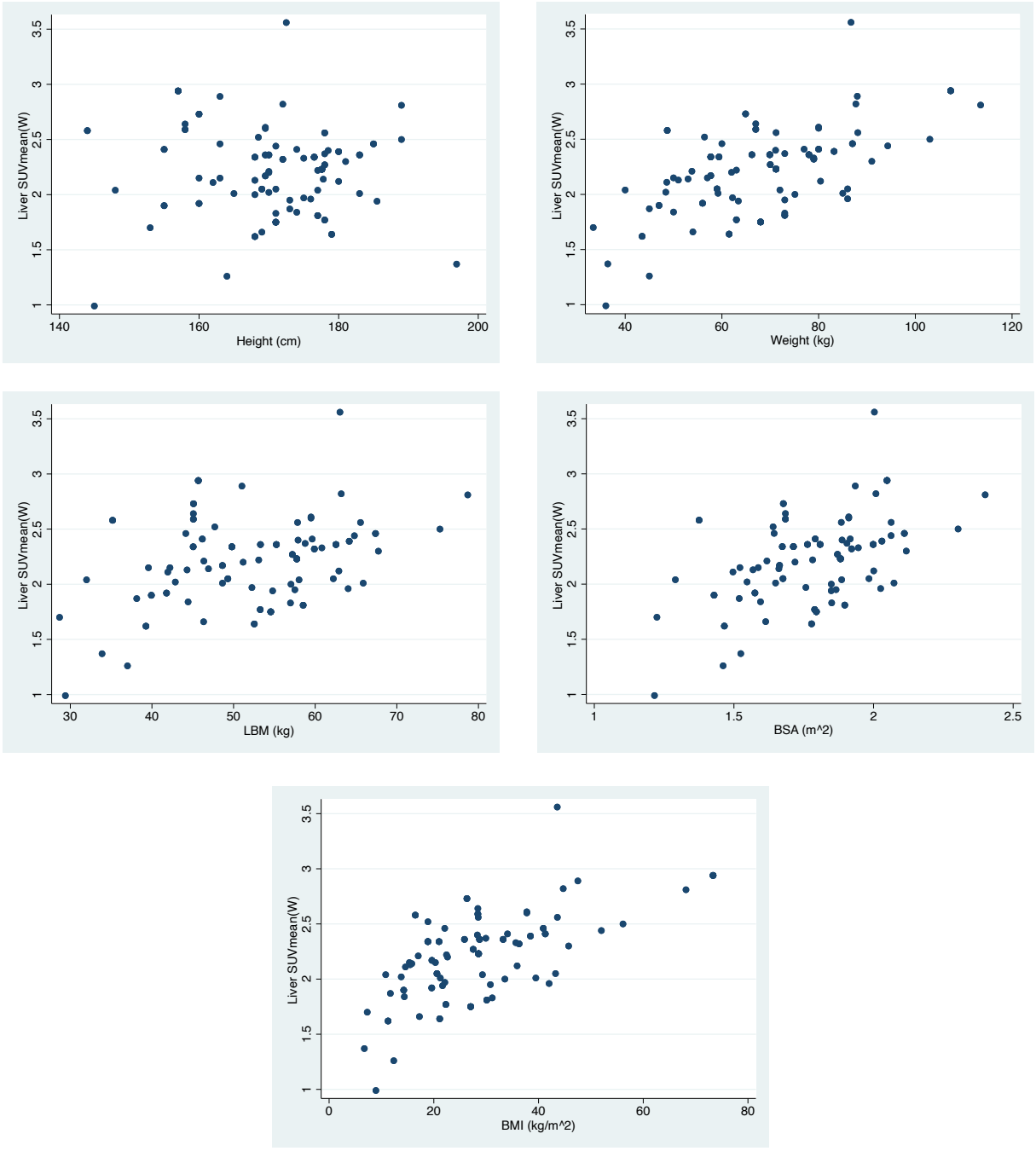


Figure 4.11: Scatter plots of the liver SUVmean and patient factors.
The plots for correlating the liver mean standardised uptake value normalised by weight in kg (liver SUVmean(w)) with patient height (cm), weight (kg), LBM (kg), BSA (m^2), and BMI show moderate correlations with weight (r=0.544, *P*=<0.001), BSA (r=-0.462, *P*=<0.001), and BMI (r=0.563, *P*=<0.001), no correlation with height (r=-0.071, *P*=0.465), and weak correlation with LBM (r=0.267, *P*=0.005).

The relationship between liver SUV_{max} normalised by weight in kg with patient height, weight(kg), LBM, BSA, and BMI

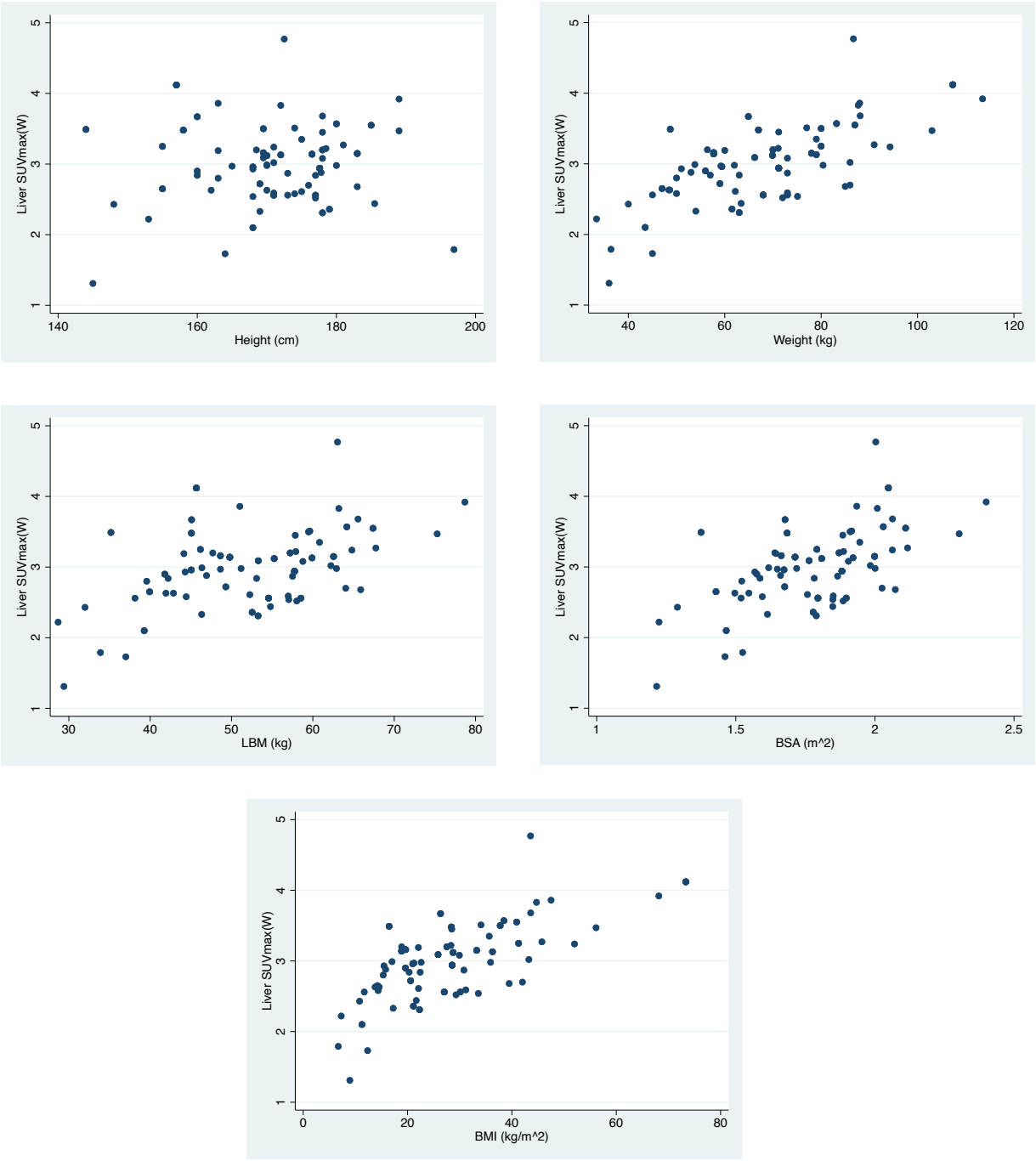


Figure 4.12: Scatter plots of the liver SUVmax and patient factors.
The plots for correlating the liver maximum standardised uptake value normalised by weight in kg (liver SUVmax(w)) with patient height (cm), weight (kg), LBM (kg), BSA (m^2), and BMI show moderate correlations with weight (r=-0.585, *P*=<0.001), BSA (r=0.510, *P*=<0.001), and BMI (r=0.598, *P*=<0.001), no correlation with height (r=-0.025, *P*=0.798), and weak correlation with LBM (r=0.325, P=<0.001).

4.4 Discussion

The results of this chapter revealed that post-treatment HNSCC SUV_{max} and SUV_{peak} normalised to weight, LBM, and BSA exhibited no or only minimal correlation with all the patient body size measurements, including height, weight, LBM, BSA, and BMI. In addition, none of the post-treatment volumetric parameters, including MTV and TLG, showed correlations with patient body size measurements. Therefore, variations in body size might have only a limited impact on HNSCC post-treatment PET/CT metrics.

Furthermore, the sub-analysis of the liver metrics revealed that the liver SUV_{mean} and SUV_{max} correlated less with LBM than with total body weight and BSA. Therefore, LBM appears to be the preferable method for calculating SUV metrics and is potentially less likely to introduce bias when reporting liver SUV readings. This indicates that the effect of applying alternative body size measures may be more noticeable when analysing SUV metrics for normal tissues as opposed to SUV metrics for lesions obtained post-treatment. Overall, we found that the use of any body size factors appears to be acceptable when analysing post-treatment FGD PET/CT quantitative metrics in HNSCC.

The next chapter will examine the diagnostic performance of post-treatment PET/CT quantitative metrics to discriminate residual HNSCC disease. It will also evaluate the reproducibility of the post-treatment SUV metrics analysed in this thesis.

CHAPTER 5

THE DIAGNOSTIC PERFORMANCE OF POST-CRT QUANTITATIVE METRICS

5. Chapter 5: The diagnostic performance of post-CRT quantitative metrics

5.1 Introduction:

In the last chapter, we found that post-treatment quantitative PET/CT metrics are not very sensitive to patient body size measures. In this chapter, we aimed to investigate whether these quantitative metrics can discriminate post-treatment HNSCC residual lesions.

Although several studies have examined the diagnostic performance of SUV_{max} normalised by total body weight (Vainshtein et al., 2014, Shimomura et al., 2014, Katahira-Suzuki et al., 2015, Sagardoy et al., 2016, Fatehi et al., 2019a, Dejaco et al., 2020, Helsen et al., 2020a), and less commonly SUV_{peak}, no consensus has yet been reached on defining their thresholds. In addition, little is known about the effect of normalisation by LBM and BSA and lesion-to-background ratios on the quantitative metrics' diagnostic performance.

Therefore, the aims of this chapter were to:

- Evaluate the diagnostic performance of SUV_{max} and SUV_{peak} normalised by total body weight obtained separately from the primary tumour and involved lymph nodes to distinguish residual disease at three months post-treatment PET/CT imaging.
- Assess whether using LBM and BSA in SUV calculation improved the diagnostic performance of these parameters.
- Assess whether lesion-to-background ratios (relative metrics) improved the diagnostic performance of the parameters.
- Identify optimal thresholds for primary tumours and involved lymph node SUV_{max} and SUV_{peak} to predict residual lesions, while considering the effect of sensitivity, specificity, PPV, and NPV on threshold selection.

- Analyse reproducibility of the post-treatment SUV measurements to ensure that these metrics were reproducible when using the same image analysis method.

5.2 Materials and methods:

Patients, scanning protocols, image analysis, and quantitative metrics were previously described in Chapter 3.

The statistical methods used were described in Chapter 3, Section 3.9.2.

5.3 Results:

5.3.1 Patient Characteristics:

Based on the inclusion and exclusion criteria, 124 patients were included in the study (92 males and 32 females). The median age of both T+ and N+ groups was 59.3 years (IQR: 53.1–65.7 years). Of those, 72/124 (58.1%) patients showed a partial or incomplete metabolic response (positive or equivocal PET/CT scans) at either the primary 49 (50.5%) or lymph node sites 40 (43.5%), while a complete metabolic response to CRT was noted in 52 (41.9%) patients (negative PET/CT scans) (Table 5.1).

Oropharyngeal cancer was the most common subtype with 88 patients (70.96%), followed by laryngeal cancer with 11 patients (8.87%). The majority of patients from both groups 111 (89.9%) were treated with CRT, while only 13 patients (10.48%) received RT. Post-CRT PET/CT imaging was performed at a median of 13.4 weeks (IQR: 12.6-13.9 weeks).

Comparing PET/CT results to HPR reports or follow-ups at one year, 96 patients (77.4%) from both groups showed locoregional control, while 28 patients (22.6%) showed local or regional failure or both (positive on histopathology or follow-up).

5.3.1.1 T+ group (primary site assessment)

97 patients (70 males and 27 females) were included in this analysis, with a median age of 60.1 years (IQR: 54.2-60.1 years). The median follow-up time was 47.5 months (IQR: 44.1-49.9 months). On post-treatment PET/CT scans, 49/97 (50.5%) patients showed a partial or incomplete metabolic response at the primary tumour site (equivocal or positive PET/CT scans), while a complete metabolic response to CRT was noted in 48/97 (49.5%) patients (negative PET/CT scans). Among these patients with PET/CT imaging showing an incomplete response to treatment, 14/49 (28.6%) patients had evidence of residual disease at the primary site (local failure) determined by histology (3 lesions) or one-year follow-up (11 lesions), while no patient showed evidence of residual disease (local failure) in the group of patients with a complete response to treatment (negative PET/CT) (Table 5.1 and Figure 5.1).

5.3.1.2 N+ group (nodal site assessment)

92 patients (66 males and 26 females) were included in this analysis, with a median age of 58.7 years (IQR: 52.2-65.4 years). The median follow-up time was 46.1 months (IQR: 42.5-47.5 months). On post-treatment PET/CT scans, 40/92 (43.5%) patients showed a partial or incomplete metabolic response at nodal sites (equivocal or positive PET/CT scans), while a complete metabolic response to CRT was noted in 52/92 (56.5%) patients (negative PET/CT scans). Among these patients with an incomplete response to treatment, 14/40 (35.0%) patients had evidence of residual disease at the lymph node site (regional failure) confirmed by histology (6 lesions), or one-year follow-up (8

lesions). In the group of patients with complete metabolic responses, only 1/52 (1.92%) patients had residual disease at the lymph node site (regional failure) (Table 5.1 and Figure 5.1).

Patient and tumour characteristics are tabulated in Table 5.1.

The study diagram illustrating the flow of participants through the study is shown in Figure 5.1.

Table 5.1: Patients and tumour characteristics

Characteristics	**Global population**	**T+ group**	**N+ group**
No. of patients	124	97	92
Age, y, median (IQR)	59.3 (53.1– 65.7)	60.1 (54.2- 60.1)	58.7 (52.2-65.4)
Gender, n (%)			
Male	92 (74.2)	70 (72.2)	66 (71.7)
Female	32 (25.8)	27 (27.8)	26 (28.3)
Tumour location			
Oropharynx	88 (70.96)	70 (88.6)	67 (72.8)
Larynx	11 (8.87)	11 (11.3)	5 (5.4)
Oropharynx + Larynx	4 (3.2)	4 (4.1)	3 (3.3)
Hypopharynx	9 (7.25)	5 (5.2)	7 (7.6)
Occult primary	12 (9.7)	7 (7.2)	10 (10.9)
HPV status, n (%)			
Positive	67 (63.2)	49 (59.6)	55 (67.1)
Negative	24 (22.6)	18 (21.9)	14 (17.1)
Not recorded	15 (14.2)	15 (18.3)	13 (15.9)
T stage, n (%)			
T1	19 (15.3)	16 (16.5)	16 (17.4)
T2	39 (31.5)	28 (28.9)	31 (33.7)
T3	21 (17)	18 (18.6)	14 (15.2)
T4	43 (34.7)	34 (35.1)	30 (32.6)
Not recorded	2 (1.6)	1 (1.0)	1 (1.1)
N stage, n (%)			
N0	5 (4.0)	5 (5.6)	–
N1	7 (5.64)	7 (7.2)	5 (5.4)
N2	109 (88)	83 (85.6)	85 (92.4)
N3	1 (0.8)	1 (1.0)	1 (1.1)
Not recorded	2 (1.6)	1 (1.0)	1 (1.1)
Treatment, n (%)			
CRT	111 (89.5)	86 (88.7)	84 (91.3)
RT only	13 (10.48)	11 (11.3)	8 (8.7)
Time between CRT and PET/CT, weeks, median (QR)	13.4 (12.6-13.9)	13.4 (12.7-14.1)	13.4 (12.6-13.9)

Post-treatment PET/CT image			
Complete response	52 (41.9)	48 (49.5)	52 (56.5)
Incomplete response	72 (58.1)	49 (50.5)	40 (43.5)
Overall LRC, n (%)			
Yes	96 (77.4)	–	–
No	28 (22.6)	–	–
Local control, n (%)			
Yes	–	**83 (85.6)	–
No	–	**14 (14.4)	–
Regional control, n (%)			
Yes	–	–	**77 (83.7)
No	–	–	**15 (16.3)

*Oropharyngeal cancers Human papillomavirus (HPV) status

** This includes overall cases with complete and incomplete responses to treatment as determined by post-treatment PET/CT. Detailed data on local control, local failure, regional control, and regional failure are provided in Figure 5.1.

Abbreviations: IQR, interquartile range; T, primary tumour; N, lymph node, LRC, locoregional control; CRT, chemoradiotherapy; RT; radiotherapy; HPV, human papilloma virus.

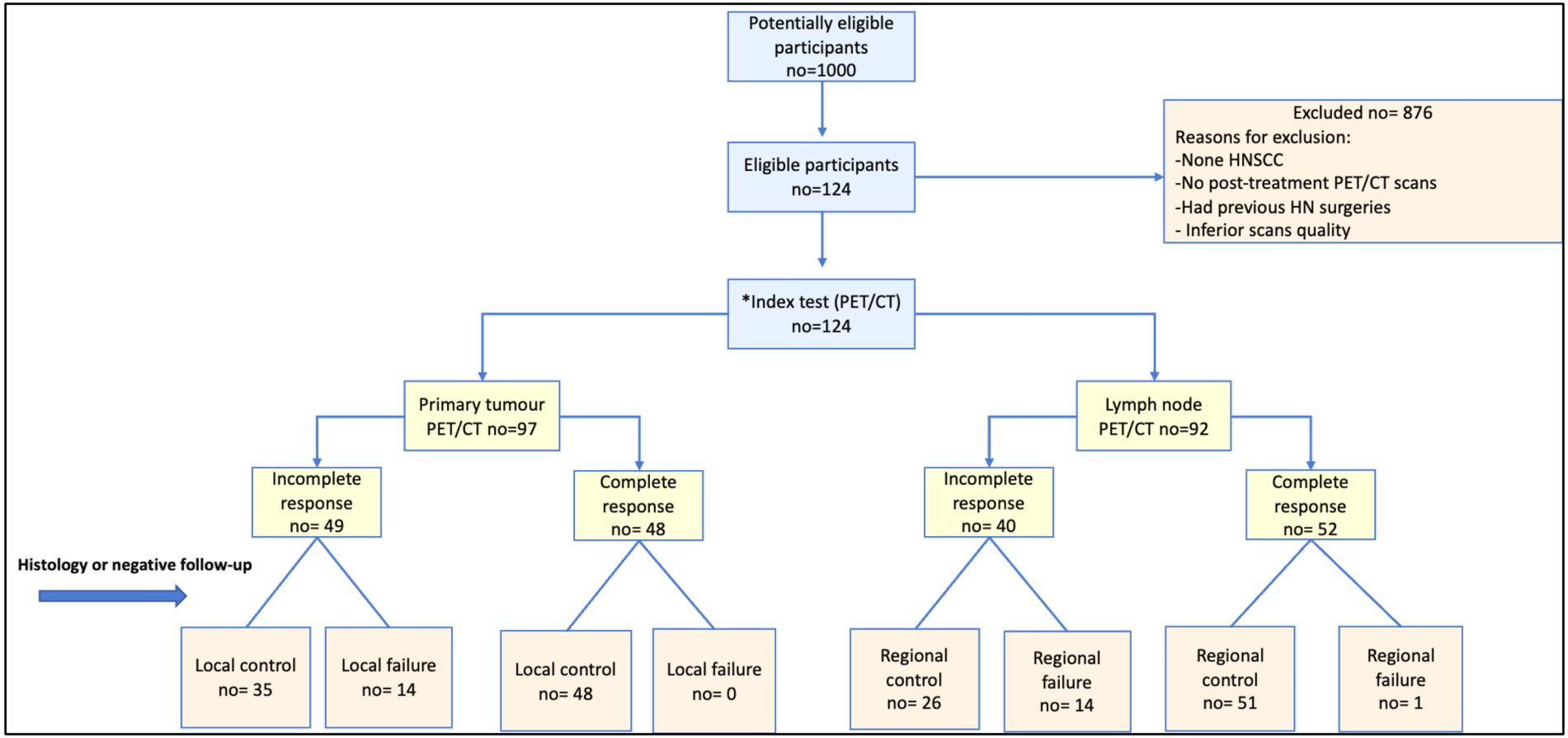


Figure 5.1: Study diagram illustrating the flow of participants through the study.

5.3.2 The diagnostic performance of primary tumour (T) and lymph node (N) SUV_{max} and SUV_{peak}

5.3.2.1 T SUV_{max}

Median post-treatment HNSCC primary tumour SUV_{max} normalised by total body weight (T $SUVmax_{w}$) was 3.85 (IQR: 3.01-6.19). Primary tumour SUV_{max} normalised by LBM (T $SUVmax_{lbm}$) showed a median of 3.17 (IQR: 2.49-5.23). Median primary tumour SUV_{max} normalised by BSA (T $SUVmax_{bsa}$) was 1.11 (IQR: 0.83-1.69) (Table 5.2).

In Table 5.3, the results show that when HNSCC post-treatment primary tumour SUV_{max} was normalised by total body weight or LBM, there was no difference observed in the values of the AUCs (0.88, 95% CI: 0.78-0.97) and (0.88, 95% CI: 0.79-0.98), respectively. However, when BSA was used for post-treatment primary tumour SUV_{max} normalisation, the AUC value increased slightly to 0.91 (95% CI: 0.83-0.99). Thus, while HNSCC primary lesion SUV_{max} normalised by BSA appeared to have the highest discriminative ability at the time of post-treatment PET/CT, all body size normalisation metrics generally yielded comparable high AUC values (Figure 5.2.a and Table 5.3).

5.3.2.2 T SUV_{peak}

Median post-treatment HNSCC primary tumour SUV_{peak} normalised by body weight (T $SUVpeak_{w}$) was 3.31 (IQR: 2.36-4.46). Similarly, median primary tumour SUV_{peak} normalised by LBM (T $SUVpeak_{lbm}$) was 2.59 (IQR: 1.86-3.41). Primary tumour SUV_{peak} normalised by BSA (T $SUVpeak_{bsa}$) had a median value of 0.87 (IQR: 0.65-1.22) (Table 5.2).

Although we can see that normalising primary tumour SUV_{peak} by BSA resulted in a slightly higher AUC of 0.89 (95% CI: 0.81- 0.98), comparable AUC values were produced by all body size normalisation methods (Figure 5.2.b and Table 5.3).

5.3.2.3 N SUV_{max}

Median post-treatment HNSCC lymph node SUV_{max} normalised by body weight (N $SUVmax_w$) was 2.58 (IQR: 2.08–3.40). Median lymph node SUV_{max} normalised by LBM (N $SUVmax_{lbm}$) was 1.97 (IQR: 1.65–2.54). Median lymph node SUV_{max} normalised by BSA (N $SUVmax_{bsa}$) was 0.68 (IQR: 0.56-0.86) (Table 5.2).

While both nodal SUV_{max} normalised by total body weight (0.94, 95% CI: 0.88-0.99) and nodal SUV_{max} normalised by LBM (0.93, 95% CI: 0.86-0.99) appeared to yield slightly greater discriminative ability compared to nodal SUV_{max} normalised by BSA (0.90, 95% CI: 0.83-0.98), all body size normalisation metrics generally yielded comparably high AUC values to detect head and neck nodal lesions (Figure 5.3.a and Table 5.3).

5.3.2.4 N SUV_{peak}

Median post-treatment HNSCC lymph node SUV_{peak} normalised by body weight (N $SUVpeak_w$) was 2.08 (IQR: 1.67-2.65). Median lymph node SUV_{peak} normalised by LBM (N $SUVpeak_{lbm}$) was 1.66 (IQR: 1.33-1.99). Median lymph node SUV_{peak} normalised by BSA (N $SUVpeak_{bsa}$) was 0.55 (IQR: 0.47-0.70) (Table 5.2).

While both nodal SUV_{peak} normalised by weight and nodal SUV_{peak} normalised by LBM appeared to yield slightly greater discriminative ability with AUCs of 0.93 (95% CI: 0.87-1.00) and 0.93 (95% CI:

0.85-1.00), respectively, in comparison to nodal SUV_{peak} normalised by BSA (0.89, 95% CI: 0.80-0.98), all body size normalisation metrics generally yielded comparably high AUC values to detect head and neck nodal lesions (Figure 5.3.b and Table 5.3).

Table 5.2: The medians and interquartile range (IQR) of post-treatment HNSCC primary tumour and lymph node SUV_{max} and SUV_{peak} normalised by total body weight, LBM, and BSA.

Parameters	**Median**	**IQR**	
		Lower	**Upper**
T SUVmax(w)	3.85	3.01	6.19
T SUVmax(lbm)	3.17	2.49	5.23
T SUVmax(bsa)	1.11	0.83	1.69
T SUVpeak(w)	3.31	2.36	4.46
T SUVpeak(lbm)	2.59	1.86	3.41
T SUVpeak(bsa)	0.87	0.65	1.22
N SUVmax(w)	2.58	2.08	3.40
N SUVmax(lbm)	1.97	1.65	2.54
N SUVmax(bsa)	0.68	0.56	0.86
N SUVpeak(w)	2.08	1.67	2.65
N SUVpeak(lbm)	1.66	1.33	1.99
N SUVpeak(bsa)	0.55	0.47	0.70

Abbreviations: T, primary tumour; N, lymph node; IQR, interquartile range; W, total body weight; LBM, lean body mass; BSA, body surface area.

Table 5.3: The diagnostic performance of each primary tumour and lymph node SUV_{max} and SUV_{peak} metrics was evaluated and compared using the AUC of the ROC curve.

Parameters	**AUC**	**SE**	**[95% CI]**	
T SUVmax(w)	0.88	0.05	0.78	0.97
T SUVmax(lbm)	0.88	0.05	0.79	0.98
T SUVmax(bsa)	0.91	0.04	0.83	0.99
T SUVpeak(w)	0.86	0.05	0.76	0.95
T SUVpeak(lbm)	0.88	0.05	0.79	0.97
T SUVpeak(bsa)	0.89	0.04	0.81	0.98
N SUVmax(w)	0.94	0.03	0.88	0.99
N SUVmax(lbm)	0.93	0.03	0.86	0.99
N SUVmax(bsa)	0.90	0.04	0.83	0.98
N SUVpeak(w)	0.93	0.03	0.87	1.00
N SUVpeak(lbm)	0.93	0.04	0.85	1.00
N SUVpeak(bsa)	0.89	0.05	0.80	0.98

The AUC of primary tumour and lymph node SUV_{max} and SUVpeak normalised by weight, LBM, and BSA with 95% confidence intervals for detection of residual disease at the primary and lymph node sites.

Abbreviations: T, primary tumour; N, lymph node; W, total body weight; LBM, lean body mass; BSA, body surface area; AUC, area under the ROC curve; SE, standard error; CI, confidence interval.

Primary tumour metrics ROC area

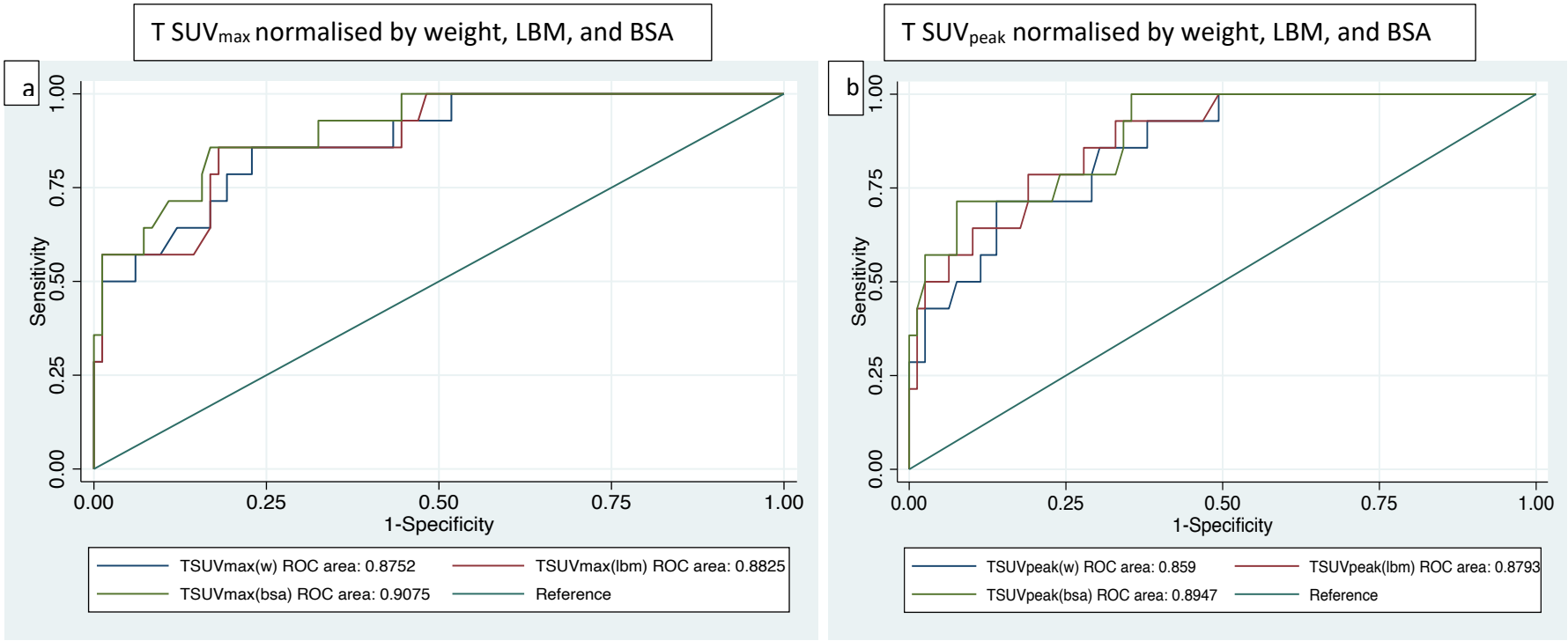


Figure 5.2: The ROC curves of T SUVmax and T SUVpeak

The ROC curves illustrate the discriminative performance of **a)** T SUVmax and **b)** T SUVpeak normalised by weight, LBM, and BSA in detecting persistent HNSCC primary lesion (T) at three-month post-CRT PET/CT imaging.

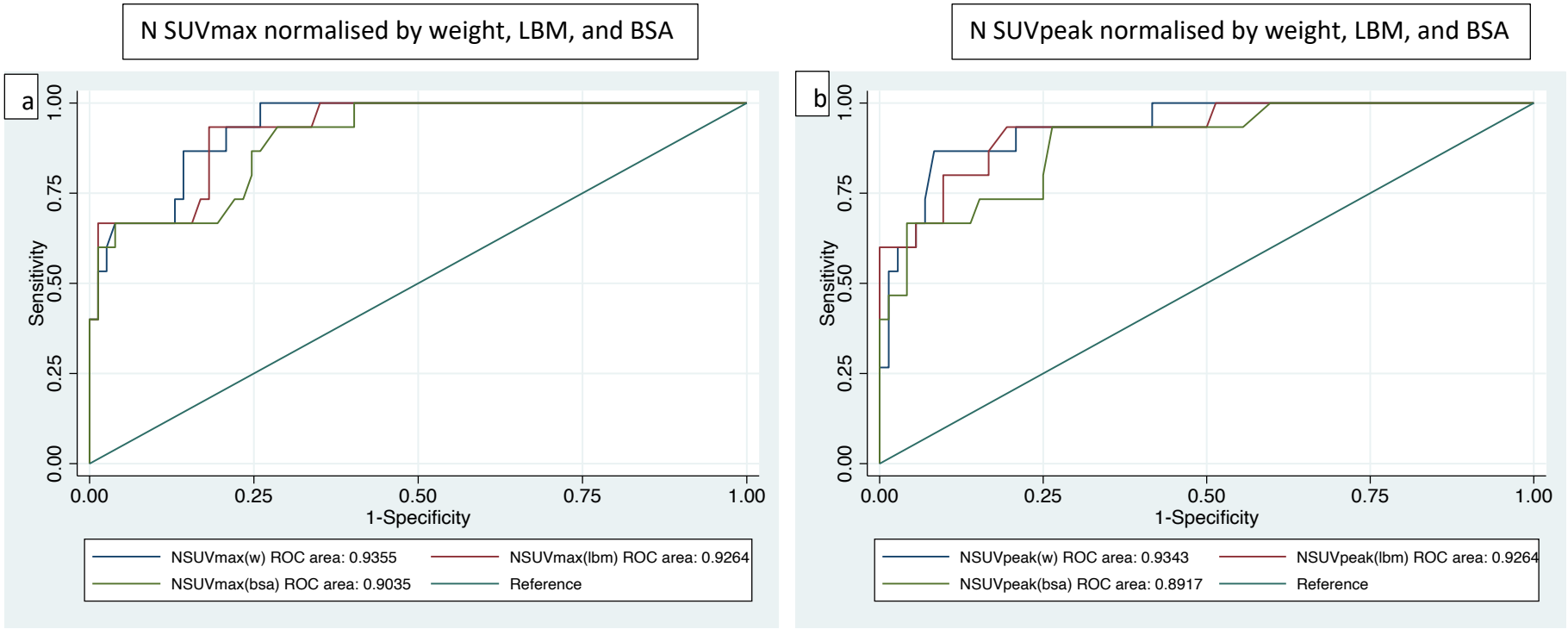


Figure 5.3: The ROC curves of N SUVmax and N SUVpeak

The curves illustrate the discriminative performance of **a)** N SUVmax and **b)** N SUVpeak normalised by weight, LBM (lean body mass), and BSA (body surface area) in detecting persistent HNSCC nodal lesion (N) at three-month post-CRT PET/CT imaging.

5.3.3 Thresholds selection- the ROC curve

The ROC curve coordinates were then used to select the optimal primary tumour and lymph node SUV_{max} and SUV_{peak} normalised by weight, LBM, and BSA thresholds. Two thresholds for each parameter were determined (Table 5.4). The first cutoff was selected with a minimal specificity of 80% and choosing the cutoff with the highest corresponding sensitivity (Helsen et al., 2020b) and NPV. A second cutoff was also chosen to demonstrate the relations between sensitivity, specificity, PPV, and NPV in HNSCC lesions. Choosing a lower cutoff value improves its sensitivity and its corresponding NPV. This was further discussed in the preceding sections.

The cutoffs selected for each primary tumour and lymph node SUV_{max} and SUV_{peak} metric normalised by weight, LBM, and BSA and their associated sensitivities and specificities are shown in Figures 5.4, 5.5, 5.6, and 5.7.

Table 5.4: The diagnostic performance of various post-treatment metrics

Parameter	Cutoff	Sensitivity	Specificity	PPV	NPV
T SUVmax(w)	**5.93**	**78.57%**	**80.72%**	**40.74%**	**95.71%**
T SUVmax(w)	5.24	85.71%	77.11%	38.71%	96.97%
T SUVmax(lbm)	**4.50**	**85.71%**	**81.93%**	**44.44%**	**97.14%**
T SUVmax(lbm)	3.16	92.86%	55.42%	26.00%	97.87%
T SUVmax(bsa)	**1.69**	**85.71%**	**84.34%**	**48.00%**	**97.22%**
T SUVmax(bsa)	1.25	92.86%	67.47%	32.50%	98.25%
T SUVpeak(w)	**4.59**	**71.43%**	**86.08%**	**47.62%**	**94.44%**
T SUVpeak(w)	3.94	85.71%	69.62%	33.33%	96.49%
T SUVpeak(lbm)	**3.40**	**78.57%**	**81.01%**	**42.31%**	**95.52%**
T SUVpeak(lbm)	4.21	64.29%	89.87%	52.94%	93.42%
T SUVpeak(bsa)	**1.45**	**71.43%**	**92.41%**	**62.50%**	**94.81%**
T SUVpeak(bsa)	1.10	78.57%	75.95%	36.67%	95.24%
N SUVmax(w)	**3.39**	**86.67%**	**85.71%**	**54.17%**	**97.06%**
N SUVmax(w)	4.00	66.67%	90.91%	58.80%	93.33%
N SUVmax(w)	5.22	66.67%	96.10%	76.92%	93.67%
N SUVmax(lbm)	**2.48**	**93.33%**	**81.82%**	**50.00%**	**98.44%**
N SUVmax(lbm)	3.84	66.67%	98.70%	90.91%	93.83%
N SUVmax(bsa)	0.79	86.67%	75.32%	40.62%	96.67%
N SUVmax(bsa)	**1.24**	**66.67%**	**96.10%**	**76.92%**	**93.67%**
N SUVpeak(w)	2.41	93.33%	79.17%	48.28%	98.28%
N SUVpeak(w)	**2.86**	**86.67%**	**91.67%**	**68.42%**	**97.06%**
N SUVpeak(lbm)	1.93	93.33%	80.56%	50.00%	98.31%
N SUVpeak(lbm)	**2.10**	**80.00%**	**90.28%**	**63.16%**	**95.59%**
N SUVpeak(bsa)	0.63	93.33%	73.61%	42.42%	98.15%
N SUVpeak(bsa)	**0.70**	**73.33%**	**84.72%**	**50.00%**	**93.85%**

The sensitivity, specificity, PPV, and NPV of HNSCC primary tumour and lymph node SUV_{max} and SUV_{peak} normalised by weight, LBM, an BSA for detecting residual disease at three-month post-treatment PET/CT imaging in comparison to the results of HPR or negative follow-up. Two thresholds for each parameter were determined.

Abbreviations: PPV, positive predictive value; NPV, negative predictive value; LBM, lean body mass; BSA, body surface area. ***Suggested optimal thresholds are in bold.***

5.3.3.1 T SUVmax(w)

Two thresholds were presented (Figure 5.4.a). T SUVmax(w) threshold of 5.93 resulted in 78.57% sensitivity, 80.72% specificity, 40.74% PPV, and 95.71% NPV, respectively.

The lower primary tumour SUVmax normalised by weight cutoff at 5.24 showed sensitivity, specificity, PPV, and NPV of 85.71%, 77%, 38.71%, and 96.97%, respectively (Table 5.4).

As mentioned before, sensitivity and specificity do not consider disease prevalence, which is an important aspect in clinical research to allow us to find out a patient's likelihood of having cancer. Determining the optimal SUV metric threshold, the number of false positives/negatives that clinicians are willing to accept in exchange for one additional true positive/negative is a crucial factor.

Looking at the NPV values (Table 5.4), the first cutoff had a NPV of 95.71%. This means that around 95.71% of all the patients with a primary tumour SUVmax(w) cutoff of <5.93 on three months post-CRT PET/CT were likely to have true negatives confirmed by histology or negative follow-up. The second SUV cutoff of 5.24 had an almost similar NPV (96.97%). However, the cutoff at 5.93 showed slightly improved specificity than the other cutoff (80.72% vs. 77.11%). Therefore, this suggests that T SUVmax(w) cutoff of 5.93 is proposed to be the optimal cutoff with a high NPV of 95.71% to discriminate true HNSCC residual disease at the primary site on three-month post-treatment FDG PET/CT imaging.

5.3.3.2 T SUVmax(lbm)

Two thresholds were presented (Figure 5.4.b). T SUVmax(lbm) threshold of 4.50, which showed sensitivity, specificity, PPV, and NPV of 85.71%, 81.93%, 44.44%, and 97.14%, respectively. The second cutoff of 3.16 had sensitivity, specificity, PPV, and NPV of 92.86%, 55.42%, 26.00%, and 97.87%, respectively. Both cutoffs had almost similar NPVs (97.14% and 97.87%), with improved specificity observed with the cutoff at 4.50 (81.93% vs. 55.42%) (Table 5.4). As a result, the T SUVmax (lbm) cutoff of 4.50 is proposed to be the optimal cutoff for distinguishing true HNSCC residual disease at the primary tumour site on three-month post-treatment PET/CT imaging, with a high NPV of 97.14%.

5.3.3.3 T SUVmax(bsa)

T SUVmax(bsa) threshold of 1.69 (Figure 5.4.c) showed sensitivity, specificity, PPV, and NPV of 85.71%, 84.34%, 48.00%, and 97.22%, respectively. The second cutoff at 1.25 had sensitivity, specificity, PPV, and NPV of 92.86%, 67.47%, 32.5%, and 98.25%, respectively. Both cutoffs had almost similar NPVs (97.22% and 98.25%), with improved specificity observed with the cutoff at 1.69 (84.34% vs. 67.47%) (Table 5.4). As a result, on three-month post-treatment PET/CT imaging, the T SUVmax (BSA) cutoff of 1.69 is proposed to be the optimal cutoff with a high NPV of 97.22% to discriminate true HNSCC residual disease at the primary tumour site.

5.3.3.4 T SUVpeak(w)

T SUVpeak(w) threshold of 4.59 (Figure 5.5.a) resulted in 71.43% sensitivity, 86.08% specificity, 47.62% PPV, and 94.44% NPV. The second cutoff of 3.94 had sensitivity, specificity, PPV, and NPV of

85.71%, 69.62%, 33.33%, and 96.49%, respectively. Both cutoffs had almost similar NPVs (94.44% and 96.49%), with improved specificity observed with the cutoff at 4.59 (86.08% vs. 69.62%). As a result, the T SUVpeak(bsa) cutoff of 4.59 is proposed to be the optimal cutoff for distinguishing true HNSCC residual disease at the primary tumour site on three-month post-treatment PET/CT imaging, with a high NPV of 94.44%.

5.3.3.5 T SUVpeak(lbm)

T SUVpeak(lbm) threshold of 3.40 (Figure 5.5.b) showed sensitivity, specificity, PPV, and NPV of 78.57%, 81.01%, 42.31%, and 95.52%, respectively. The second cutoff of 4.21 had sensitivity, specificity, PPV, and NPV of 64.29%, 89.87%, 52.94%, and 93.42%, respectively. Looking at Figure 5.5.b, we can see that the cutoff at 3.40 seems closer to the left upper corner. Even though it shows less specificity, it has a higher NPV compared to the other cutoff. As a result, we propose that the T SUVpeak(lbm) cutoff of 3.40, with a high NPV of 95.52%, is the best cutoff for distinguishing true HNSCC residual disease at the primary tumour site on three-month post-treatment PET/CT imaging.

5.3.3.6 T SUVpeak(bsa)

T SUVpeak(bsa) threshold of 1.45 (Figure 5.5.c) showed sensitivity, specificity, PPV, and NPV of 71.43%, 92.41%, 62.5%, and 94.81%, respectively. The second cutoff of 1.10 had sensitivity, specificity, PPV, and NPV of 78.57%, 75.95%, 36.67%, and 95.24%, respectively. Both cutoffs had nearly identical NPVs (94.81% and 95.24%), with improved specificity observed with the cutoff at 1.45 (92.41% vs. 75.95%) (Table 5.4). As a result, the T SUVpeak (BSA) cutoff of 1.45 is proposed to

be the optimal cutoff for distinguishing true HNSCC residual disease at the primary tumour site on three-month post-treatment PET/CT imaging, with a high NPV of 94.81%.

5.3.3.7 N SUVmax (w)

The sensitivity, specificity, PPV, and NPV of an N SUVmax(w) threshold of 3.39 (Figure 5.6.a) were 86.67%, 85.71%, 54.17%, and 97.06%, respectively. The second cutoff of 5.22 had sensitivity, specificity, PPV, and NPV of 66.67%, 96.1%, 76.92%, and 93.67%, respectively. The cutoff of 3.39 had a greater NPV than the other cutoff (97.06% vs. 93.67%). As a result, the N SUVmax(w) cutoff of 3.39 is proposed to be the best cutoff for distinguishing true HNSCC residual disease at the lymph node site on three-month post-treatment PET/CT imaging. The third cutoff of 4.0 was discussed later in the discussion section. It was presented because it was suggested in two previous publications (Sagardoy et al., 2016, Nelissen et al., 2017).

5.3.3.8 N SUVmax (lbm)

N SUVmax(lbm) threshold of 2.48 (Figure 5.6.b) showed sensitivity, specificity, PPV, and NPV of 93.33%, 81.82%, 50.00%, and 98.44%, respectively. The second cutoff of 3.84 had sensitivity, specificity, PPV, and NPV of 66.67%, 98.7%, 90.91%, and 93.83%, respectively. The cutoff of 2.48 had a greater NPV than the other cutoff (98.44% and 93.83%) (Table 5.4). As a result, the N SUVmax(lbm) cutoff of 2.48 is proposed to be the best cutoff for distinguishing true HNSCC residual disease at the lymph node site on three-month post-treatment PET/CT imaging.

5.3.3.9 N SUVmax(bsa)

N SUVmax(bsa) threshold of 0.79 (Figure 5.6.c) showed sensitivity, specificity, PPV, and NPV of 86.67%, 75.32%, 40.62%, and 96.67%, respectively. The second cutoff of 1.24 had sensitivity, specificity, PPV, and NPV of 66.67%, 96.10%, 76.92%, and 93.67%, respectively. The cutoff of 1.24 had greater specificity than the other cutoff (96.10% and 75.32%) (Table 5.4). As a result, the N SUVmax(lbm) cutoff of 1.24 is proposed to be the best cutoff for distinguishing true HNSCC residual disease at the lymph node site on three-month post-treatment PET/CT imaging.

5.3.3.10 N SUVpeak(w)

The sensitivity, specificity, PPV, and NPV of an N SUVpeak(w) threshold of 2.86 (Figure 5.7.a) were 86.67%, 91.67%, 68.42%, and 97.06%, respectively. The second cutoff of 2.41 had sensitivity, specificity, PPV, and NPV of 93.33%, 79.17%, 48.28%, and 98.28%, respectively. Both cutoffs had almost similar NPVs (98.28% and 97.06%), with improved specificity observed with the cutoff at 2.86 (91.67% vs. 79.17%) (Table 5.4). As a result, the N SUVpeak(w) cutoff of 2.86 is proposed to be the optimal cutoff for distinguishing true HNSCC residual disease at the lymph node site on three-month post-treatment PET/CT imaging, with a high NPV of 97.06%.

5.3.3.11 N SUVpeak(lbm)

N SUVpeak (lbm) threshold of 2.10 (Figure 5.7.b) showed sensitivity, specificity, PPV, and NPV of 80%, 90.28%, 63.16%, and 95.59%, respectively. The second cutoff of 1.93 had sensitivity, specificity, PPV, and NPV of 93.33%, 80.56%, 50.00%, and 98.31%, respectively. Both cutoffs had very close NPVs (95.59% and 98.31%), with improved specificity observed with the cutoff at 2.10

(90.28% vs. 80.56%) (Table 5.4). As a result, on three-month post-treatment PET/CT imaging, the N SUVpeak(lbm) cutoff of 2.10 is proposed to be the optimal cutoff with a high NPV of 95.59% to discriminate true HNSCC residual disease at the lymph node site.

5.3.3.12 N SUVpeak(bsa)

N SUVpeak(bsa) threshold of 0.70 (Figure 5.7.c) had sensitivity, specificity, PPV, and NPV of 73.33%, 84.72%, 50.00%, and 93.85%, respectively. The second cutoff was 0.63, which showed sensitivity, specificity, PPV, and NPV of 93.33%, 73.61%, 42.42%, and 98.15%, respectively. We can see that the cutoff of 0.70 had greater specificity than the other cutoff (84.72% vs. 73.61%) and a reasonable NPV (Table 5.4). Therefore, this suggests that N SUVpeak(bsa) cutoff of 0.70 is proposed to be the optimal cutoff with a high NPV of 93.85% to discriminate true HNSCC residual disease at the lymph node site on three-month post-treatment PET/CT imaging.

Primary tumour SUV_{max} thresholds

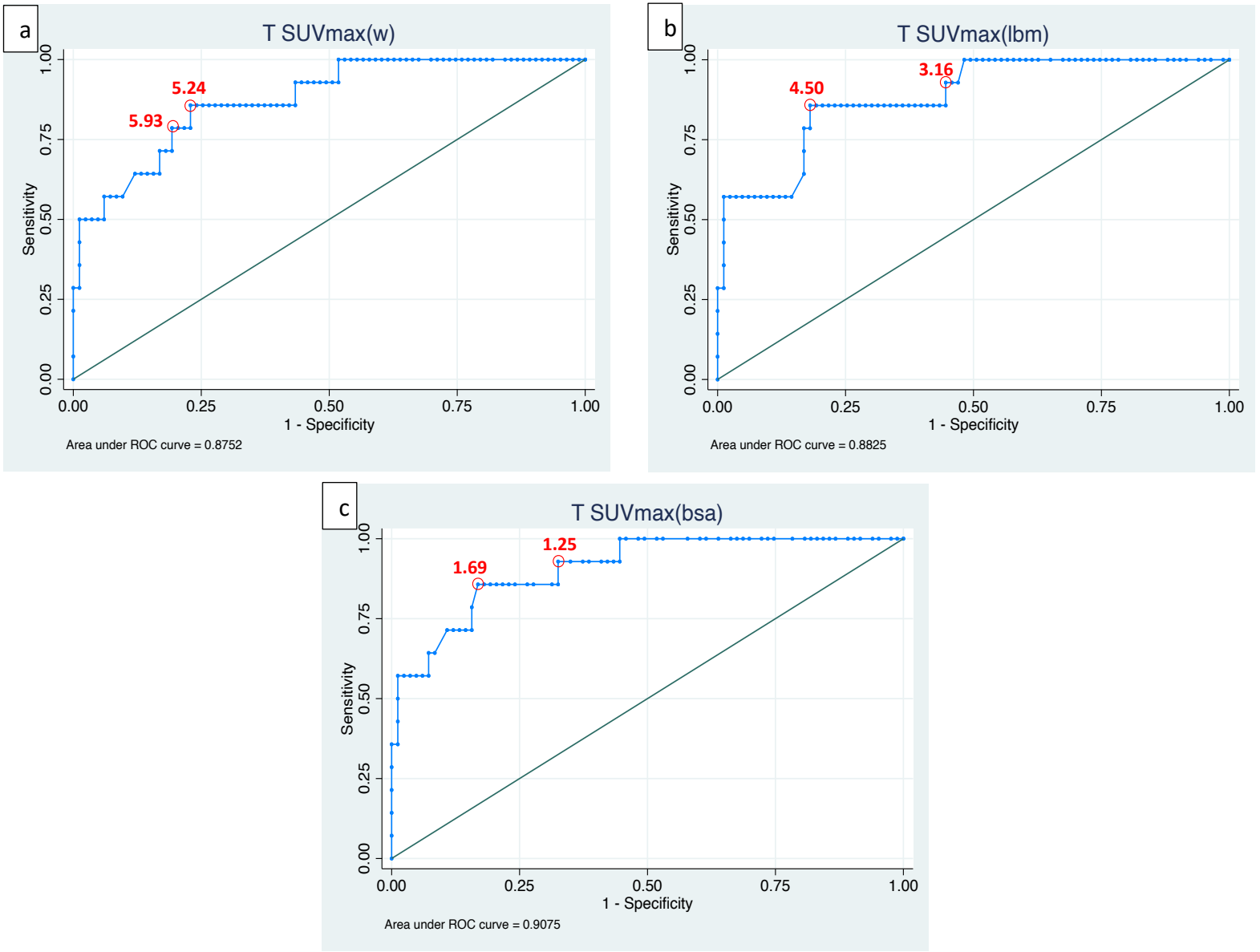


Figure 5.4: The ROC curve of the post-treatment primary tumour SUVmax with two different potential optimal thresholds for a) T SUVmax(w), b) T SUVmax(lbm), and c) T SUVmax(bsa)

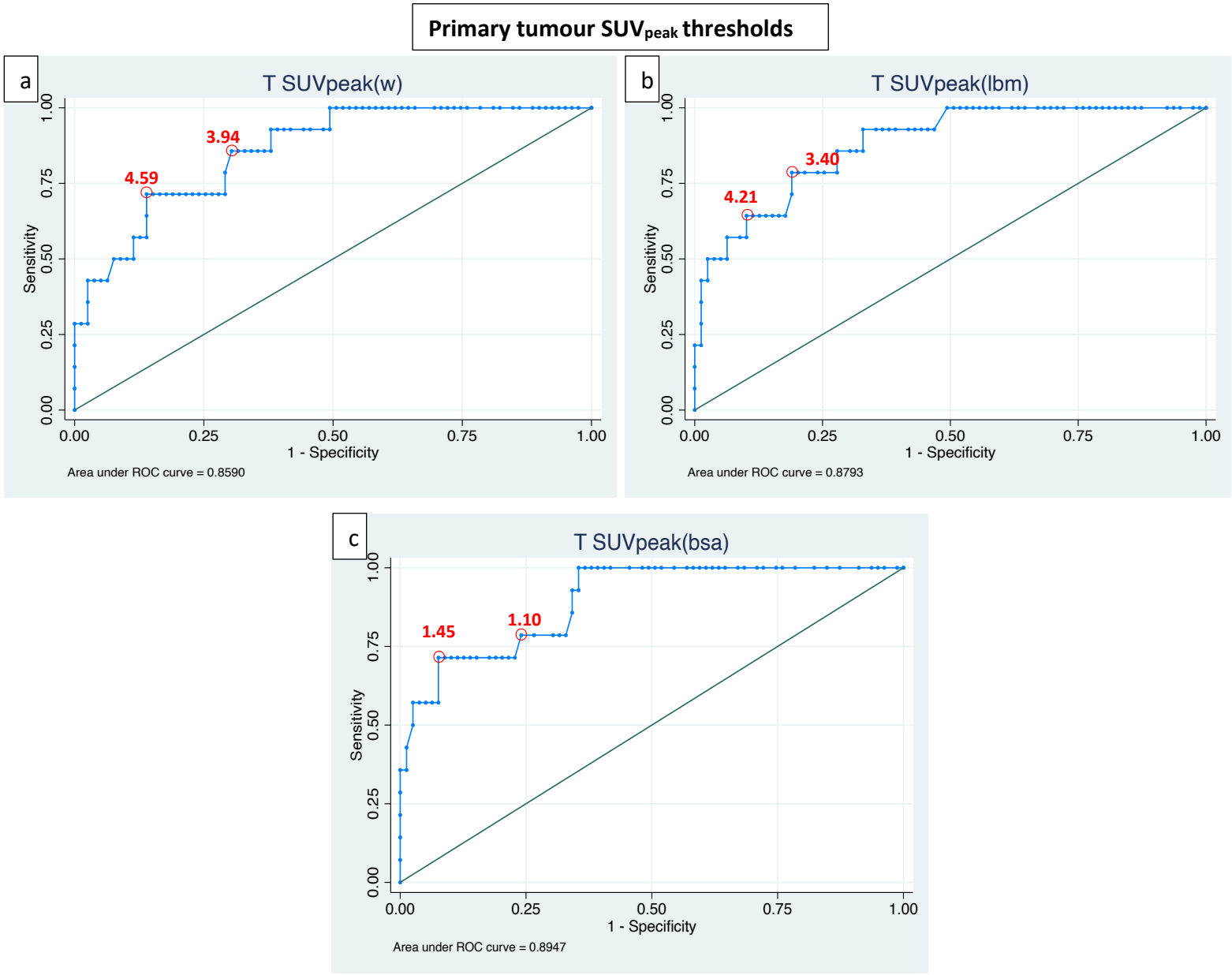


Figure 5.5: The ROC curve of the post-treatment primary tumour SUVpeak with two different potential optimal thresholds for a) T SUVpeak(w), b) T SUVpeak(lbm), and c) T SUVpeak(bsa)

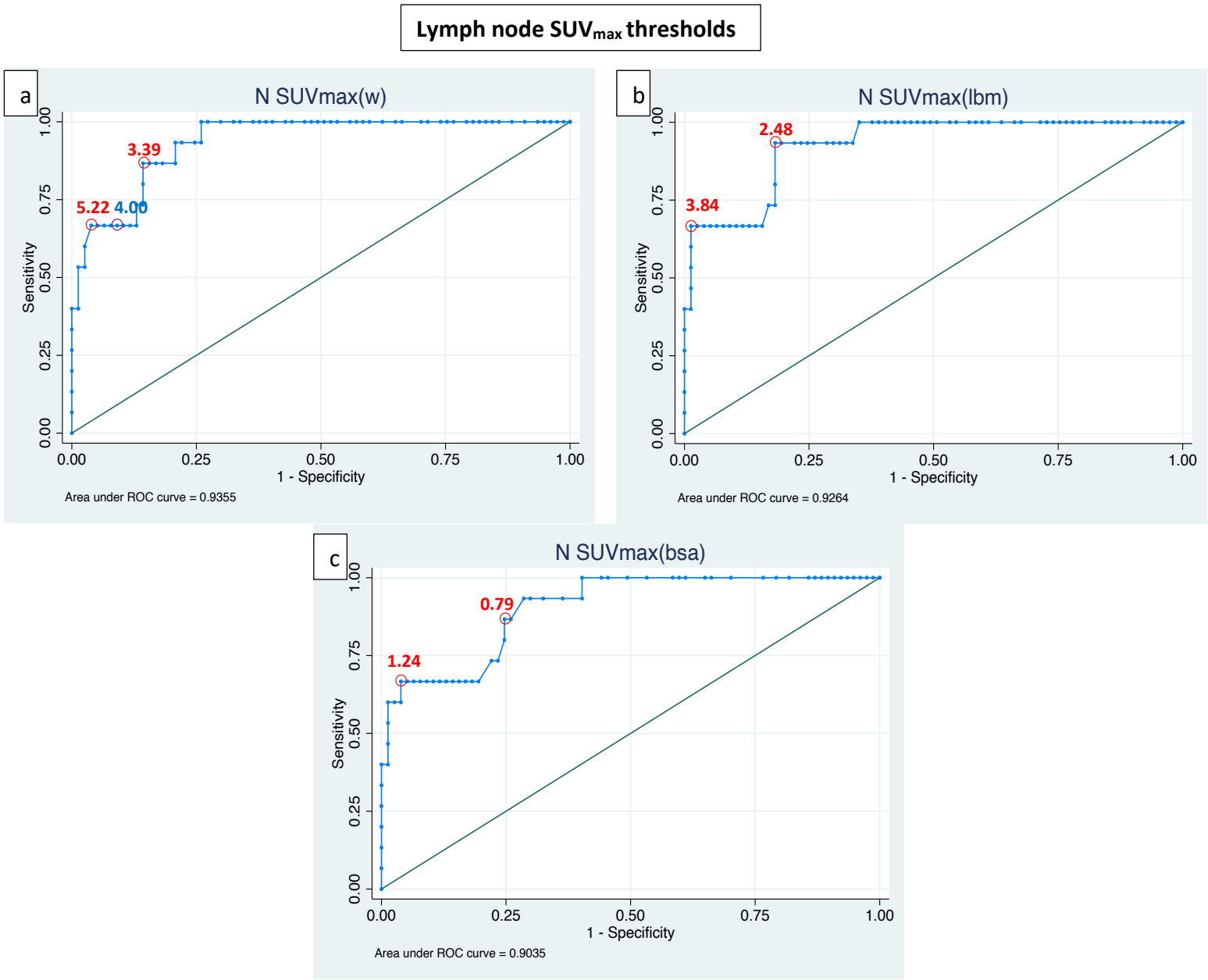


Figure 5.6: The ROC curve of the post-treatment lymph node SUVmax with two different potential optimal thresholds for a) N SUVmax(w), b) N SUVmax(lbm), and c) N SUVmax(bsa)

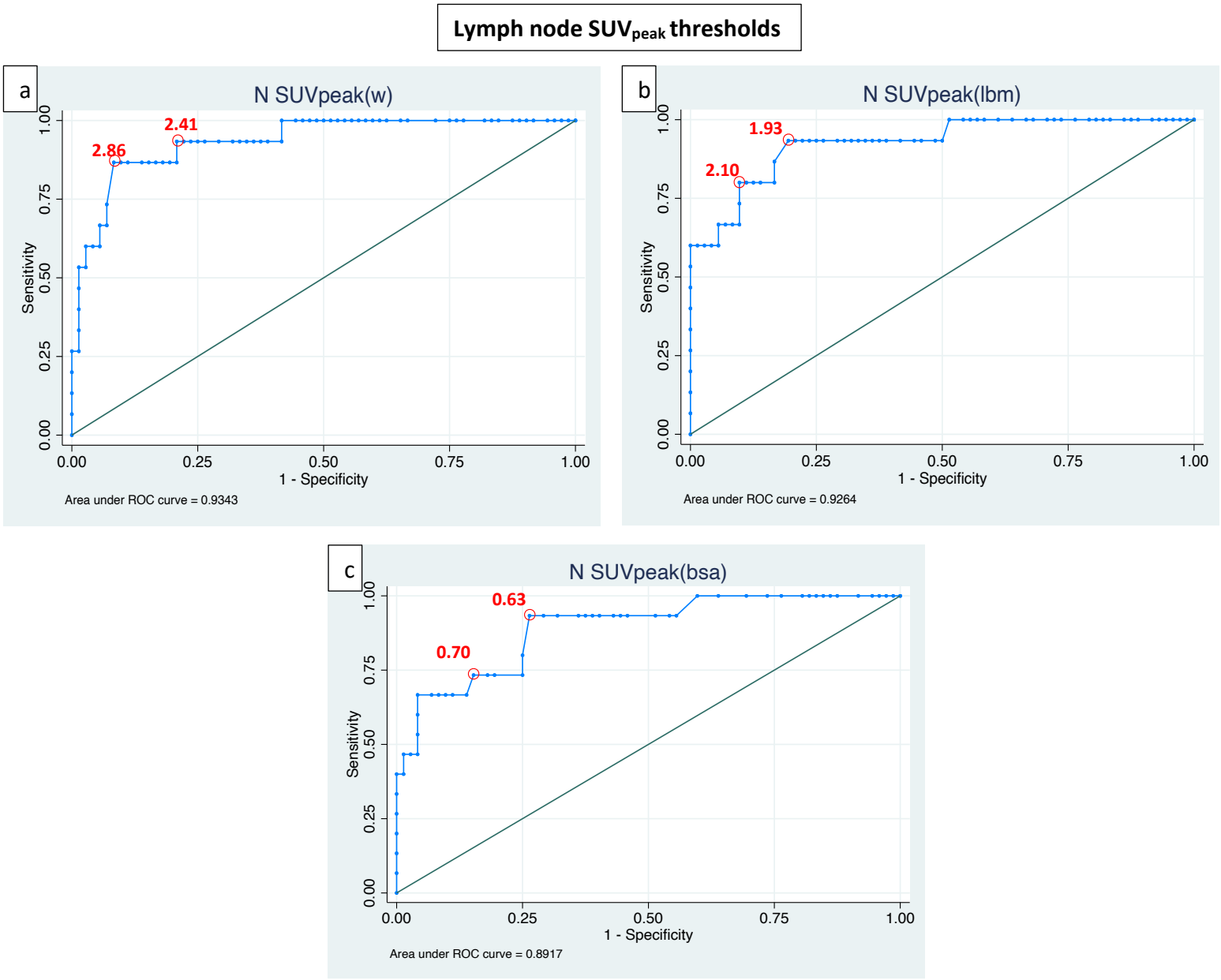


Figure 5.7: The ROC curve of the post-treatment lymph node SUVpeak with two different potential optimal thresholds for a) N SUVpeak(w), b) N SUVpeak(lbm), and c) N SUVpeak(bsa)

5.3.4 The diagnostic performance of T and N SUV_{max} in HPV^{+ve} OPSCC and HPV^{-ve} HNSCC.

We undertook a sub-analysis comparing the SUV median values and proposed thresholds for HPV^{+ve} OPSCC and HPV^{-ve} HNSCC patient groups (Table 5.5). This reveals that lesions with HPV^{+ve} disease had SUV median values that were lower than those lesions with HPV^{-ve} disease. This was most likely because HPV^{+ve} lesions are more sensitive to therapy than HPV^{-ve} lesions (Perri et al., 2020). They responded better; hence, their FDG uptake was on average lower, resulting in lower SUV readings. Proposed thresholds for each group are illustrated in Table 5.5.

Table 5.5: Post-treatment PET/CT metrics in HPV^{+ve} OPSCC and HPV^{-ve} HNSCC.

		Primary tumour		Lymph node	
		HPV+ve	HPV-ve	HPV+ve	HPV-ve
SUVmax(w)	Median (IQR)	3.76 (2.82- 5.72)	4.77 (3.09-6.19)	2.61 (2.09- 3.18)	2.74 (2.195- 5.68)
	Threshold	6.51	5.93	3.17	3.39
SUVmax(lbm)	Median (IQR)	3.15 (2.37- 4.02)	3.55 (2.52- 5.25)	1.94 (1.65- 2.33)	2.23 (1.74- 4.58)
	Threshold	4.25	4.5	2.32	3.18

This table illustrates the comparison of two imaging metrics, including SUV_{max} normalised to total body weight and SUV_{max} normalised to LBM obtained from primary tumours and nodal sites in HPV^{+ve} OPSCC and HPV^{-ve} HNSCC. Primary tumour and lymph node metrics showed lower SUV median values in HPV^{+ve} disease compared to HPV^{-ve} disease. The thresholds proposed were based on 80.0% minimal specificity.
Abbreviations: IQR, interquartile range.

5.3.5 Lesion-to-background ratios

Each of the primary tumour and lymph node SUV_{max} and SUV_{peak} normalised to total body weight, LBM, and BSA were further normalised to the average FDG uptake (SUV_{mean}) in three different background regions, including the liver, cerebellum, and BP. This was carried out to investigate whether lesion-to-background normalisation improved the diagnostic performance for detecting persistent disease at the site of the primary HNSCC tumour (T) and the lymph node (N) at three-month post-CRT PET/CT imaging.

Table 5.6 illustrates the AUC values of primary tumour and lymph node HNSCC SUV_{max} and SUV_{peak} normalised by weight, LBM, and BSA and standardised further by the average FDG uptake (SUV_{mean}) in the liver, cerebellum, and BP.

5.3.5.1 T SUVmax(w)- to-background ratios

From Table 5.6, we can see that when the T SUVmax(w) was further normalised by the average uptake (SUVmean) in the liver, cerebellum, and BP, the ratio of T SUVmax(w) normalised by the cerebellum exhibited the highest AUC value (0.92, 95% CI: 0.86-0.99), whereas standardisation by the BP yielded the lowest AUC value (0.88, 95% CI: 0.79-0.96). This may suggest that, while generally comparable AUC values were found, normalisation to the cerebellum (SUV_{mean}) produced the greatest AUC value.

Overall, when comparing the absolute metric to the relative metric, the absolute metric of T SUVmax(w) showed an AUC of 0.88 (95% CI: 0.78-0.97) (Table 5.4). When T SUVmax(w) was further normalised to the liver and cerebellum SUV_{mean}, the AUC increased to 0.90 (95% CI: 0.83-0.98) and 0.92 (95% CI: 0.86-0.99), respectively, whereas it remained almost the same when T SUVmax(w) was further normalised to the BP SUV_{mean}, with an AUC of 0.88 (95% CI: 0.79-0.96) (Table 5.6).

5.3.5.2 T SUVmax(lbm)-to-background ratios

Similarly, T SUVmax(lbm) normalised by the cerebellum showed the greatest AUC value (0.92, 95% CI: 0.86-0.99), while normalisation by BP had the lowest AUC value (0.88, 95% CI: 0.79-0.96). This may suggest that, while generally comparable AUC values were found, the normalisation of T SUVmax(lbm) to the cerebellum (SUV_{mean}) produces the highest AUC value (Table 5.6).

Overall, T SUVmax(lbm) showed an AUC of 0.88 (95% CI: 0.79-0.98) (Table 5.4). When T SUVmax(lbm) was further normalised to the liver and cerebellum SUV_{mean}, the AUC increased to 0.90 (95% CI: 0.82-0.98) and 0.92 (95% CI: 0.86-0.99), respectively, whereas it remained almost the same when it was further normalised to the BP SUV_{mean} with an AUC of 0.88 (95% CI: 0.79-0.96) (Table 5.6).

5.3.5.3 T SUVmax(bsa)- to-background ratios

T SUVmax(bsa) also showed similar results. T SUVmax(bsa) normalised by the cerebellum showed the greatest AUC value (0.92, 95% CI: 0.85-0.98), while normalisation by the BP had the lowest AUC value (0.88, 95% CI: 0.79-0.96). This may suggest that, while generally comparable AUC values were found, the normalisation of T SUVmax(bsa) to the cerebellum (SUV_{mean}) produces the largest AUC value (Table 5.6).

Overall, the absolute metric T SUVmax(bsa) showed an AUC of 0.91 (95% CI: 0.83-0.99) (Table 5.4). When T SUVmax(bsa) was further normalised to the cerebellum SUV_{mean}, the AUC increased to 0.92 (95% CI: 0.85- 0.98), whereas it decreased when it was further normalised to the liver and the BP SUV_{mean}, with AUCs of 0.90 (95% CI: 0.82-0.98) and 0.88 (95% CI: 0.79-0.96), respectively (Table 5.6).

5.3.5.4 All T SUV_{peak}- to-background ratios

Similar to the T SUV_{max} parameters, when T SUV_{peak} normalised by weight, LBM, and BSA were further normalised to the average uptake in the liver, cerebellum, and BP, all T SUV_{peak} metrics showed the greatest AUC values when further normalised by the cerebellum and the lowest AUC values when further normalised by the BP (Table 5.6). This may suggest that, while generally comparable AUC values were found, the normalisation of T SUV_{peak} metrics to the cerebellum (SUV_{mean}) produces the largest AUC value.

Overall, when the absolute metrics of T SUV_{peak} were compared to the relative-to-background ratios, the AUC values increased when the absolute SUV_{peak} metrics were further normalised to the cerebellum SUV_{mean}, while less or no improvement was found when further normalised by the BP SUV_{mean} (Table 5.6).

5.3.5.5 N SUVmax(w)-to-background ratios

N SUVmax(w) normalised by the cerebellum showed the greatest AUC value (0.92, 95% CI: 0.85- 0.98), while normalisation by the liver had the lowest AUC value (0.89, 95% CI: 0.78- 0.99) (Table 5.6). This may suggest that, while generally comparable AUC values were found, the normalisation of N SUVmax(w) to the cerebellum (SUV_{mean}) yielded the greatest AUC value.

Overall, when comparing the absolute metric to the relative metric, N SUVmax(w) showed an AUC of 0.94 (95% CI: 0.88- 0.99) (Table 5.4). When N SUVmax(w) was further normalised to the liver, cerebellum, and BP, the AUC values decreased to 0.89 (95% CI: 0.78- 0.99), 0.92 (95% CI: 0.85- 0.98), and 0.91(95% CI: 0.83- 0.98), respectively (Table 5.6), suggesting that no improvement was achieved when the nodal SUVmax(w) metric was further normalised by the average FDG uptake in the background regions.

5.3.5.6 N SUVmax(lbm)- to-background ratios

Similarly, N SUVmax(lbm) normalised to the cerebellum showed the greatest AUC value (0.92, 95% CI: 0.85-0.98), while normalisation by the liver had the lowest AUC value (0.89, 95% CI: 0.78-0.99) (Table 5.6). This may suggest that, while generally comparable AUC values were found, the normalisation of N SUVmax(lbm) to the cerebellum (SUV_{mean}) produced the greatest AUC value.

Overall, when comparing the absolute metric to the relative metric, the absolute N SUVmax(lbm) showed an AUC of 0.93 (95% CI: 0.86- 0.99) (Table 5.4). When N SUVmax(lbm) was further normalised to the liver, cerebellum, and BP, the AUC decreased to 0.89 (95% CI: 0.78- 0.99), 0.92 (95% CI: 0.85- 0.98), and 0.91(95% CI: 0.83- 0.98), respectively (Table 5.6), suggesting that no improvement was achieved when the nodal SUVmax(lbm) metric was further normalised by the average FDG uptake in background regions.

5.3.5.7 N SUVmax(bsa)- to-background ratios

Both N SUVmax(bsa) normalised to the cerebellum and the BP showed the greatest AUC values (0.91, 95% CI: 0.84–0.98) and (0.91, 95% CI: 0.84-0.98), respectively, while normalisation by the BP yielded the lowest AUC value (0.89, 95% CI: 0.79-0.99). This may suggest that, while generally comparable AUC values were found, the normalisation of N SUVmax(bsa) to the cerebellum and the liver (SUV_{mean}) provided the largest AUC value.

Overall, when comparing the absolute metric to the relative metric, the absolute N SUVmax(bsa) showed an AUC of 0.90 (95% CI: 0.83-0.98) (Table 5.4). When N SUVmax(bsa) was further normalised to the cerebellum and BP, the AUC increased to 0.91 (95% CI: 0.84-0.98) and 0.91 (95%

CI: 0.84-0.98), respectively, while decreased when further normalised to the liver SUV_{mean}, with an AUC of 0.89 (95% CI: 0.78-0.99) (Table 5.6).

5.3.5.8 All N SUVpeak- to-background ratios

Similar findings to the N SUVmax metrics were found.

No improvement was achieved when the nodal SUVpeak(w) and nodal SUVpeak(lbm) were further normalised to the average FDG uptake in the liver, cerebellum, and BP. However, when N SUVpeak(bsa) was further normalised to the average FDG uptake in the cerebellum and blood pool, the diagnostic performance improved slightly (Tables 5.4 and 5.6).

Table 5.6: AUC values of lesion-to-background ratios

	SUV relative metrics	AUC	SE	[95% CI]	
T SUVmax(w)	T SUVmax(w)/ liver	0.90	0.04	0.83	0.98
	T SUVmax(w)/ cerebellum	0.92	0.03	0.86	0.99
	T SUVmax(w)/ blood pool	0.88	0.04	0.79	0.96
T SUVmax(lbm)	T SUVmax(lbm)/ liver	0.90	0.04	0.82	0.98
	T SUVmax(lbm)/ cerebellum	0.92	0.03	0.86	0.99
	T SUVmax(lbm)/ blood pool	0.88	0.04	0.79	0.96
T SUVmax(bsa)	T SUVmax(bsa)/ liver	0.90	0.04	0.82	0.98
	T SUVmax(bsa)/ cerebellum	0.92	0.03	0.85	0.98
	T SUVmax(bsa)/ blood pool	0.88	0.04	0.79	0.96
T SUVpeak(w)	T SUVpeak(w)/ liver	0.89	0.04	0.81	0.96
	T SUVpeak(w)/ cerebellum	0.90	0.03	0.84	0.97
	T SUVpeak(w)/ blood pool	0.86	0.04	0.77	0.94
T SUVpeak(lbm)	T SUVpeak(lbm)/ liver	0.89	0.04	0.82	0.97
	T SUVpeak(lbm)/ cerebellum	0.90	0.03	0.84	0.97
	T SUVpeak(lbm)/ blood pool	0.85	0.04	0.77	0.94
T SUVpeak(bsa)	T SUVpeak(bsa)/ liver	0.89	0.04	0.81	0.96
	T SUVpeak(bsa)/ cerebellum	0.90	0.03	0.83	0.97
	T SUVpeak(bsa)/ blood pool	0.85	0.05	0.76	0.94
N SUVmax(w)	N SUVmax(w)/ liver	0.89	0.05	0.78	0.99
	N SUVmax(w)/ cerebellum	0.92	0.03	0.85	0.98
	N SUVmax(w)/ blood pool	0.91	0.04	0.83	0.98
N SUVmax(lbm)	N SUVmax(lbm)/ liver	0.89	0.05	0.78	0.99
	N SUVmax(lbm)/ cerebellum	0.92	0.03	0.85	0.98
	N SUVmax(lbm)/ blood pool	0.91	0.04	0.83	0.98
N SUVmax(bsa)	N SUVmax(bsa)/ liver	0.89	0.05	0.79	0.99
	N SUVmax(bsa)/ cerebellum	0.91	0.04	0.84	0.98
	N SUVmax(bsa)/ blood pool	0.91	0.04	0.84	0.98
N SUVpeak(w)	N SUVpeak(w)/ liver	0.88	0.05	0.77	0.99
	N SUVpeak(w)/ cerebellum	0.92	0.03	0.85	0.98
	N SUVpeak(w)/ blood pool	0.91	0.04	0.84	0.97
N SUVpeak(lbm)	N SUVpeak(lbm)/ liver	0.88	0.06	0.77	0.99
	N SUVpeak(lbm)/ cerebellum	0.91	0.04	0.84	0.98
	N SUVpeak(lbm)/ blood pool	0.90	0.04	0.83	0.97
N SUVpeak(bsa)	N SUVpeak(bsa)/ liver	0.88	0.06	0.77	0.99
	N SUVpeak(bsa)/ cerebellum	0.91	0.04	0.84	0.98
	N SUVpeak(bsa)/ blood pool	0.91	0.04	0.83	0.98

Abbreviations: T, primary tumour; N, lymph node; W, total body weight; LBM, lean body mass; BSA, body surface area; AUC, area under the ROC curve; SE, standard error; CI, confidence interval.

5.3.6 Intraobserver variability among PET/CT metrics of patients with HNSCC

5.3.6.1 Descriptive statistics

The means and SDs of HNSCC lesion SUV_{max} and SUV_{peak}, as well as SUV_{mean} and SUV_{max} for the liver, cerebellum, and BP acquired at two separate time points, are shown in Table 5.7. All parameters were normalised to W, LBM, and BSA and collected twice by a single rater at least five months apart.

Table 5.7: HNSCC lesions SUVmax and SUVpeak, and the liver, cerebellum, BP SUVmean and SUVmax all normalised to weight(w), lean body mass (LBM), and body surface area (BSA) acquired at two different time point.

Parameter	First Reading		Second Reading	
	Mean	SD	Mean	SD
Lesion SUVmax(w)	4.85	3.19	4.84	3.20
Lesion SUVmax(lbm)	3.74	2.39	3.70	2.42
Lesion SUVmax(bsa)	1.30	0.87	1.29	0.88
Lesion SUVpeak(w)	3.51	2.11	3.50	2.05
Lesion SUVpeak(lbm)	2.70	1.62	2.67	1.57
Lesion SUVpeak(bsa)	0.93	0.58	0.94	0.56
Liver SUVmean(w)	2.33	0.38	2.31	0.40
Liver SUVmean(lbm)	1.80	0.28	1.78	0.30
Liver SUVmean(bsa)	0.62	0.09	0.62	0.10
Liver SUVmax(w)	3.13	0.49	3.09	0.48
Liver SUVmax(lbm)	2.41	0.33	2.38	0.33
Liver SUVmax(bsa)	0.83	0.09	0.82	0.11
Cerebellum SUVmean(w)	7.01	1.06	7.16	1.22
Cerebellum SUVmean(lbm)	5.42	0.87	5.53	0.95
Cerebellum SUVmean(bsa)	1.88	0.32	1.92	0.35
Cerebellum SUVmax(w)	9.03	1.50	9.19	1.66
Cerebellum SUVmax(lbm)	6.99	1.20	7.11	1.35
Cerebellum SUVmax(bsa)	2.42	0.44	2.46	0.48
Blood pool SUVmean(w)	1.82	0.27	1.82	0.23
Blood pool SUVmean(lbm)	1.41	0.19	1.41	0.18
Blood pool SUVmean(bsa)	0.48	0.07	0.49	0.06
Blood pool SUVmax(w)	2.37	0.41	2.26	0.33
Blood pool SUVmax(lbm)	1.82	0.27	1.75	0.25
Blood pool SUVmax(bsa)	0.63	0.09	0.60	0.09

Abbreviations: W, total body weight; LBM, lean body mass; BSA, body surface area; SD, standard deviation.

5.3.6.2 Intraclass correlation coefficients (ICC)

5.3.6.2.1 Lesion SUV_{max} and SUV_{peak}

The results of the ICC are illustrated in Table 5.8. It showed an excellent ICC between two reading times for HNSCC lesions SUV_{max} and lesion SUV_{peak} all normalised to weight, LBM, and BSA. The ICC results for lesion SUV_{max} when normalised to W, LBM, and BSA were 0.998 (95% CI: 0.996-0.999), 0.996 (95% CI: 0.994-0.998), and 0.997 (95% CI: 0.995-0.998), respectively.

The ICC results for lesion SUV_{peak} when normalised to W, LBM, and BSA were 0.992 (95% CI: 0.985-0.996), 0.990 (95% CI: 0.981-0.994), and 0.989 (95% CI: 0.980-0.994), respectively (Table 5.8).

Looking at the Bland-Altman plots in Figures 5.8 and 5.9, the data points of lesions SUV_{max} and SUV_{peak} normalised to W, LBM, and BSA are very close to the middle lines (representing the mean difference), which may indicate a good level of agreement between the two readings.

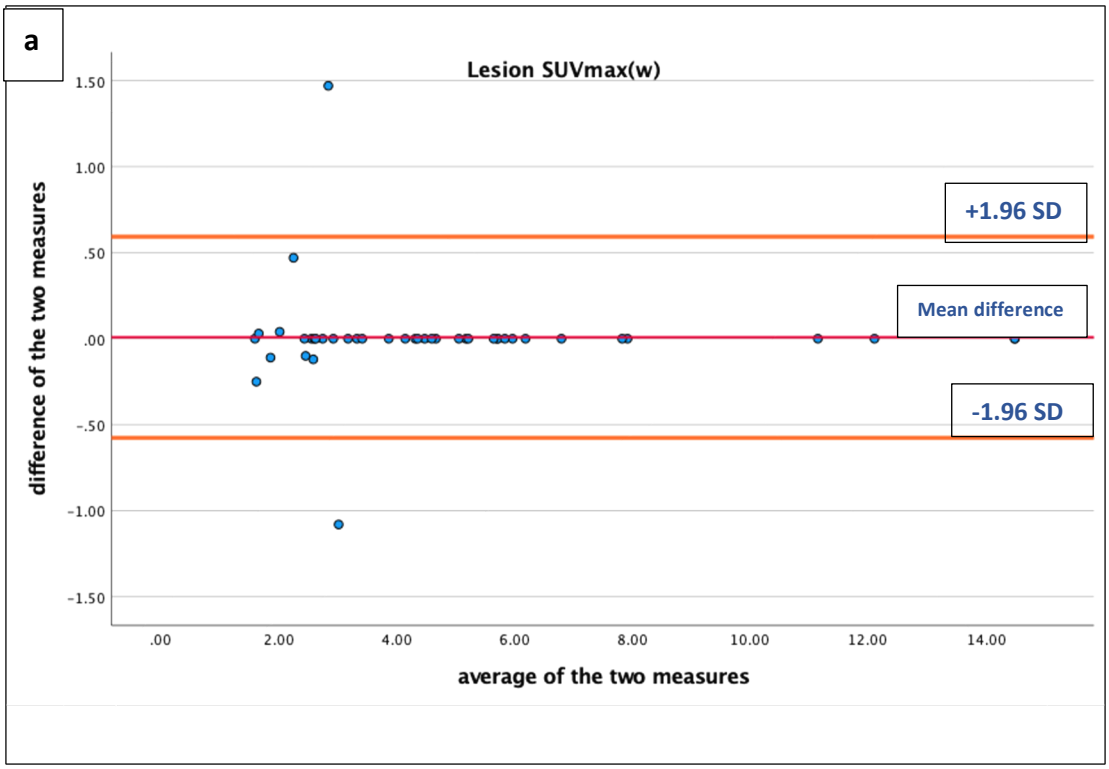
The linear regression analysis of the lesions SUV_{max} and lesions SUV_{peak} normalised to W, LBM, and BSA revealed that there was no proportional bias with all *P* values greater than 0.05 (Table 5.9).

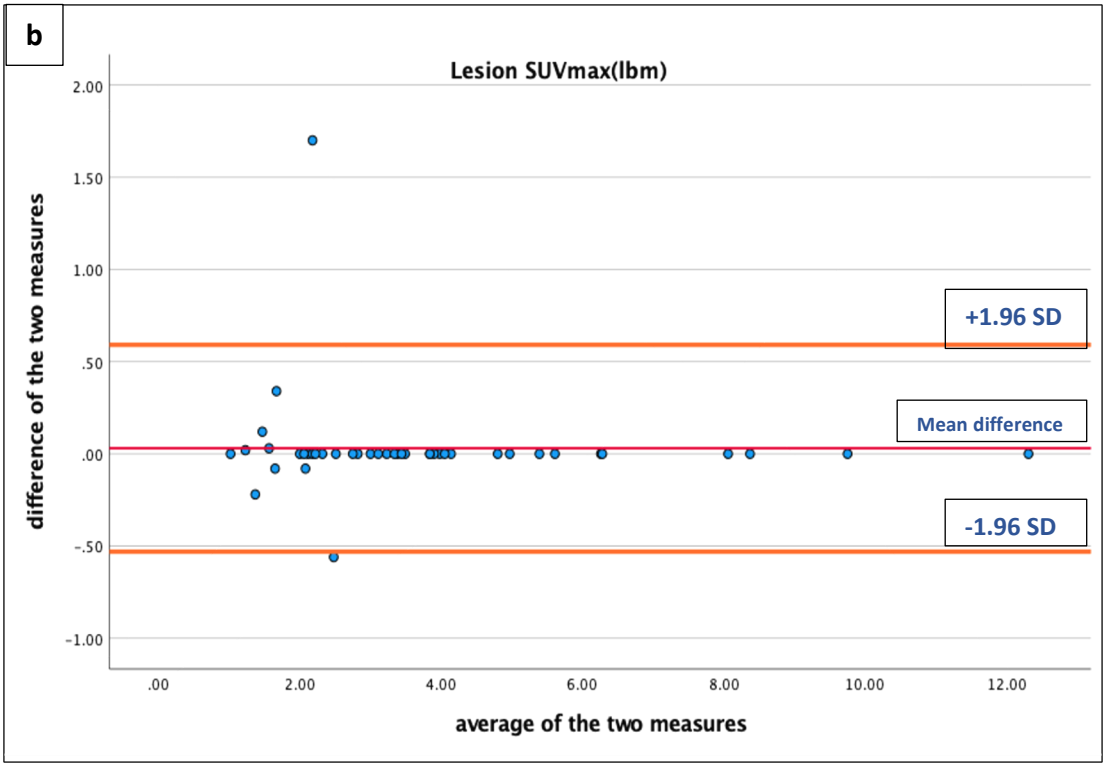
This excellent agreement suggests that when HNSCC target lesions are correctly identified, lesion SUV_{max} and SUV_{peak} are highly reproducible metrics. Also, this excellent agreement shows that very low variability was observed, suggesting that consistent image analysis when HNSCC lesions SUV_{max} and lesions SUV_{peak} were taken at different time points.

Table 5.8: Intraobserver agreement of the lesions SUV_{max} and SUV_{peak}, as well as background reference region SUV_{mean} and SUV_{max} all normalised to total body weight (w), lean body mass (LBM), and body surface area (BSA) by intraclass correlation coefficient (ICC) analysis.

PET/CT Parameters	No.	Intraclass Correlation	[95% CI]	
Lesion SUVmax(w)	42	0.998	0.996	0.999
Lesion SUVmax(lbm)	42	0.996	0.994	0.998
Lesion SUVmax(bsa)	42	0.997	0.995	0.998
Lesion SUVpeak(w)	42	0.992	0.985	0.996
Lesion SUVpeak(lbm)	42	0.990	0.981	0.994
Lesion SUVpeak(bsa)	42	0.989	0.980	0.994
Liver SUVmean(w)	25	0.966	0.924	0.985
Liver SUVmean(lbm)	25	0.963	0.916	0.984
Liver SUVmean(bsa)	25	0.961	0.911	0.983
Liver SUVmax(w)	25	0.945	0.877	0.976
Liver SUVmax(lbm)	25	0.926	0.833	0.967
Liver SUVmax(bsa)	25	0.906	0.788	0.958
Cerebellum SUVmean(w)	25	0.955	0.898	0.980
Cerebellum SUVmean(lbm)	25	0.960	0.908	0.982
Cerebellum SUVmean(bsa)	25	0.963	0.916	0.984
Cerebellum SUVmax(w)	25	0.947	0.881	0.977
Cerebellum SUVmax(lbm)	25	0.950	0.888	0.978
Cerebellum SUVmax(bsa)	25	0.957	0.903	0.981
Blood pool SUVmean(w)	25	0.950	0.887	0.978
Blood pool SUVmean(lbm)	25	0.947	0.879	0.977
Blood pool SUVmean(bsa)	25	0.938	0.861	0.973
Blood pool SUVmax(w)	25	0.765	0.473	0.896
Blood pool SUVmax(lbm)	25	0.738	0.420	0.883
Blood pool SUVmax(bsa)	25	0.743	0.430	0.886

- *3 decimal places were used to show the slight difference in the results.*


a
Lesion SUVmax(w)
difference of the two measures
1.50
1.00
.50
.00
-.50
-1.00
-1.50
+1.96 SD
Mean difference
-1.96 SD
.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
average of the two measures


b
Lesion SUVmax(lbm)
difference of the two measures
2.00
1.50
1.00
.50
.00
-.50
-1.00
+1.96 SD
Mean difference
-1.96 SD
.00
2.00
4.00
6.00
8.00
10.00
12.00
average of the two measures

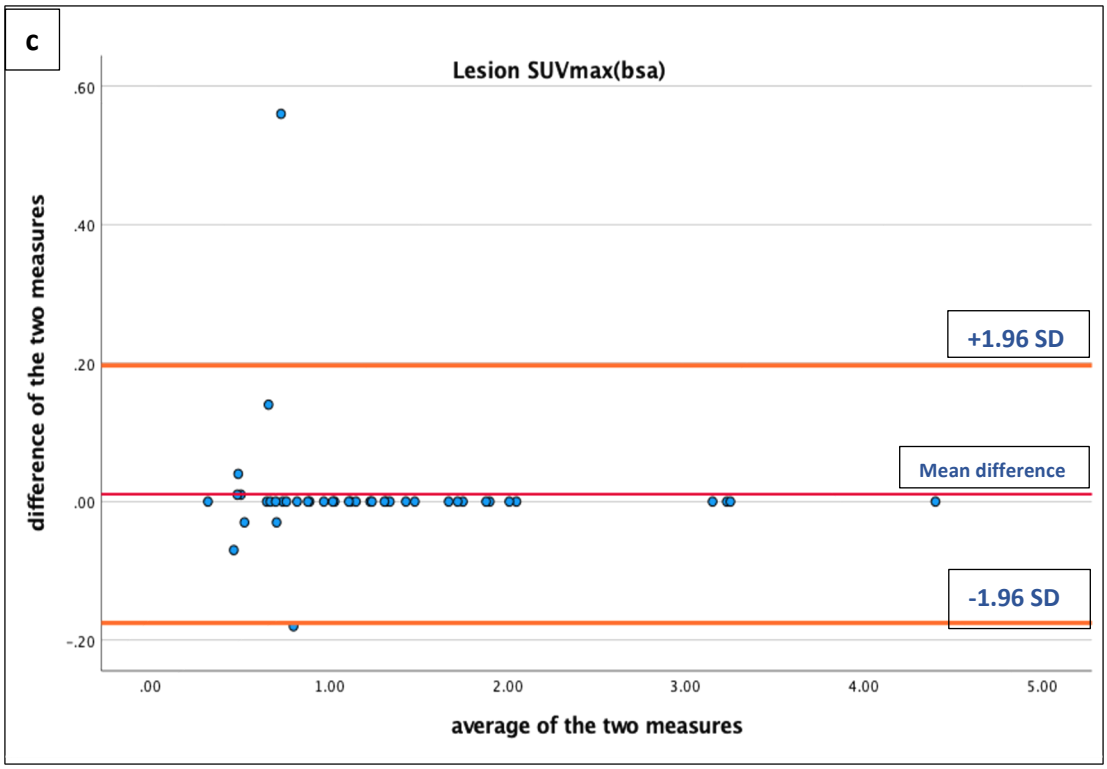
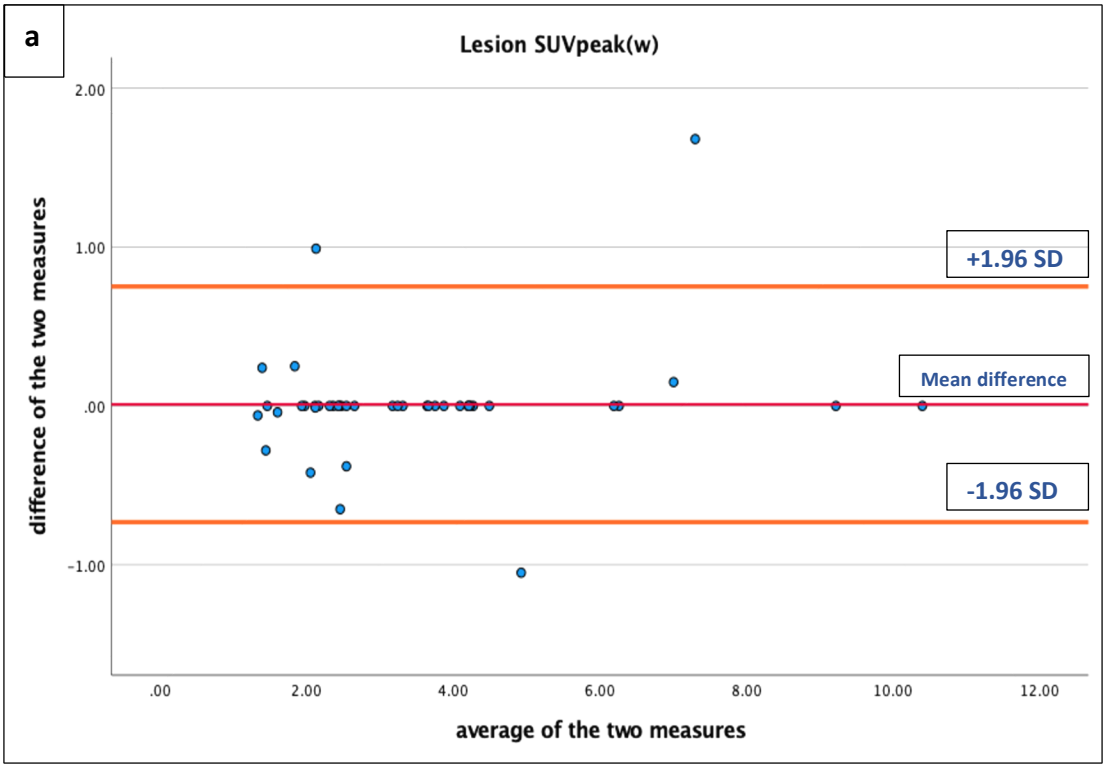
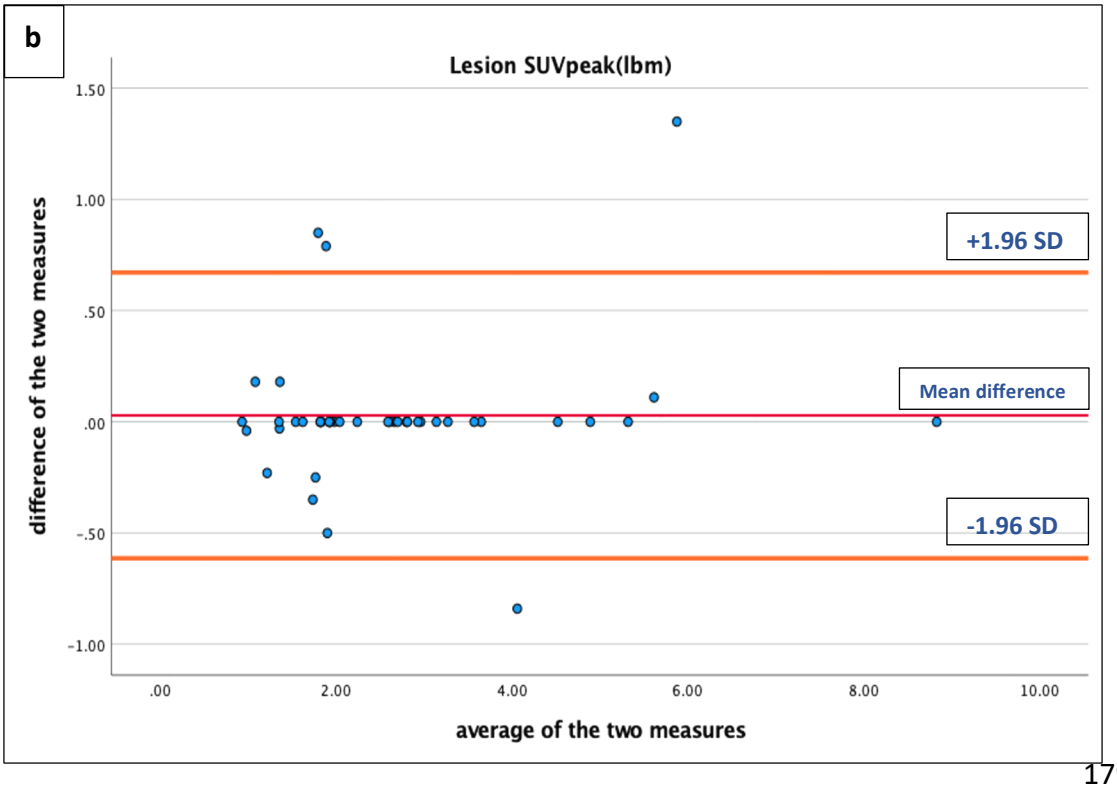


Figure 5.8: Bland-Altman plots display scatter diagrams of the differences in the two reading times of the a) lesion SUVmax(w), b) lesion SUVmax(lbm) c) lesion SUVmax(bsa) plotted against the averages of the two measurement lines.

Horizontal lines are drawn at the mean difference (red lines) and at the limits of agreement (orange lines (+1.96 SD for the upper lines and -1.96 SD for the lower lines)). The data points are very close to the middle lines (mean difference), which indicates a good level of agreement between the two readings. In addition, the presence of outliers may be attributable to the variation in VOI location across negative PET/CT scans.


a
Lesion SUVpeak(w)
2.00
1.00
.00
-1.00
difference of the two measures
+1.96 SD
Mean difference
-1.96 SD
.00
2.00
4.00
6.00
8.00
10.00
12.00
average of the two measures


b
Lesion SUVpeak(lbm)
1.50
1.00
.50
.00
-.50
-1.00
difference of the two measures
+1.96 SD
Mean difference
-1.96 SD
.00
2.00
4.00
6.00
8.00
10.00
average of the two measures

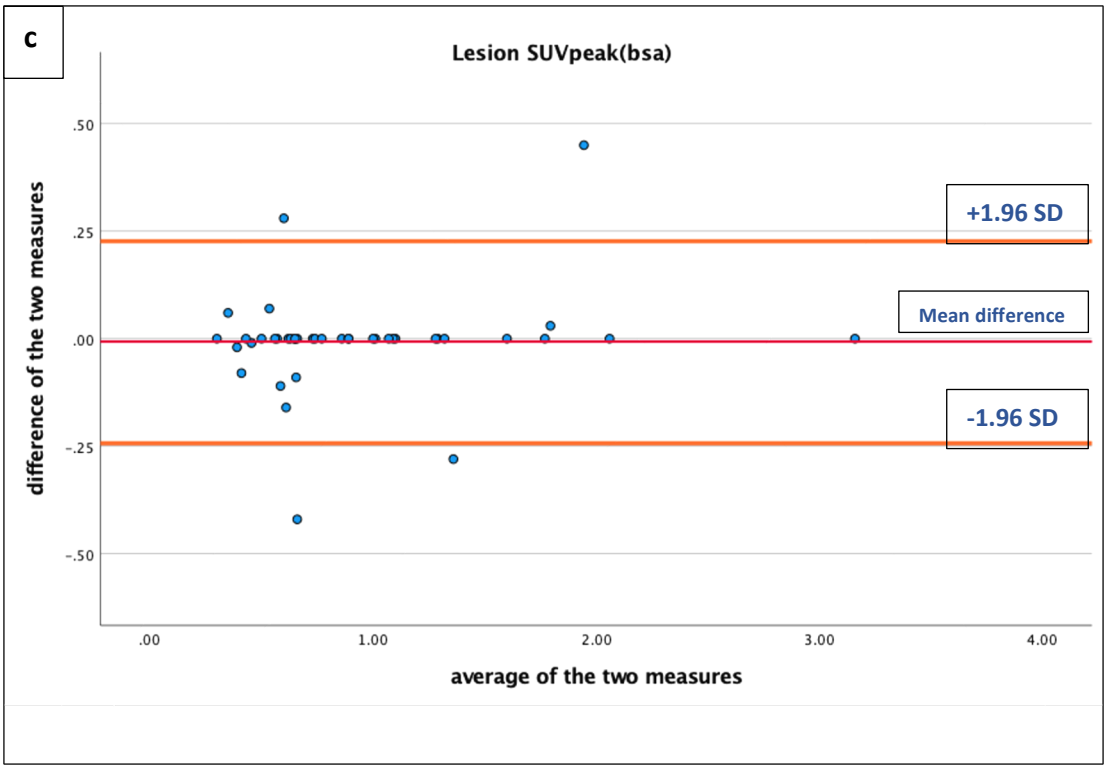


Figure 5.9: Bland-Altman plots display scatter diagrams of the differences in the two reading times of the a) lesion SUVpeak(w), b) lesion SUVpeak(lbm), and c) lesion SUVpeak(bsa) plotted against the averages of the two measurement lines.

Horizontal lines are drawn at the mean difference (red lines) and at the limits of agreement (orange lines (+1.96 SD for the upper lines and -1.96 SD for the lower lines)). The data points are very close to the middle lines (mean difference), which indicates a good level of agreement between the two readings. In addition, the presence of outliers may be attributable to the variation in VOI location across negative PET/CT scans.

Table 5.9: The results of univariate linear regression analysis for the reliability among SUV metrics acquired at two different time points.

Parameter	B	Upper CI	lower CI	P value
Lesion SUVmax(W)	-0.002	-0.032	0.028	0.903
Lesion SUVmax(lbm)	-0.009	-0.047	0.029	0.625
Lesion SUVmax(bsa)	-0.009	-0.043	0.026	0.620
Lesion SUVpeak(W)	0.032	-0.026	0.089	0.272
Lesion SUVpeak(lbm)	0.029	-0.037	0.094	0.381
Lesion SUVpeak(bsa)	0.037	-0.030	0.104	0.268
Liver SUVmean(W)	-0.053	-0.211	0.106	0.498
Liver SUVmean(lbm)	-0.055	-0.223	0.112	0.500
Liver SUVmean(bsa)	-0.070	-0.242	0.101	0.406
Liver SUVmax(W)	0.024	-0.178	0.226	0.807
Liver SUVmax(lbm)	-0.008	-0.243	0.228	0.947
Liver SUVmax(bsa)	-0.125	-0.385	0.136	0.332
Cerebellum SUVmean(W)	-0.144	-0.310	0.022	0.087
Cerebellum SUVmean(lbm)	-0.091	-0.254	0.072	0.259
Cerebellum SUVmean(bsa)	-0.111	-0.264	0.042	0.146
Cerebellum SUVmax(W)	-0.105	-0.297	0.087	0.269
Cerebellum SUVmax(lbm)	-0.121	-0.305	0.063	0.188
Cerebellum SUVmax(bsa)	-0.078	-0.253	0.097	0.365
Blood pool SUVmean(W)	0.188	0.009	0.367	0.040
Blood pool SUVmean(lbm)	0.055	-0.147	0.256	0.580
Blood pool SUVmean(bsa)	0.096	-0.117	0.308	0.361
Blood pool SUVmax(W)	0.248	-0.140	0.637	0.199
Blood pool SUVmax(lbm)	0.062	-0.369	0.493	0.768
Blood pool SUVmax(bsa)	-0.018	-0.447	0.411	0.931

Abbreviations: SUVmax, maximum standardised uptake value; SUVpeak, peak standardised uptake value; W, total body weight; LBM, lean body mass; BSA, body surface area; B, beta coefficient; CI, confidence interval.

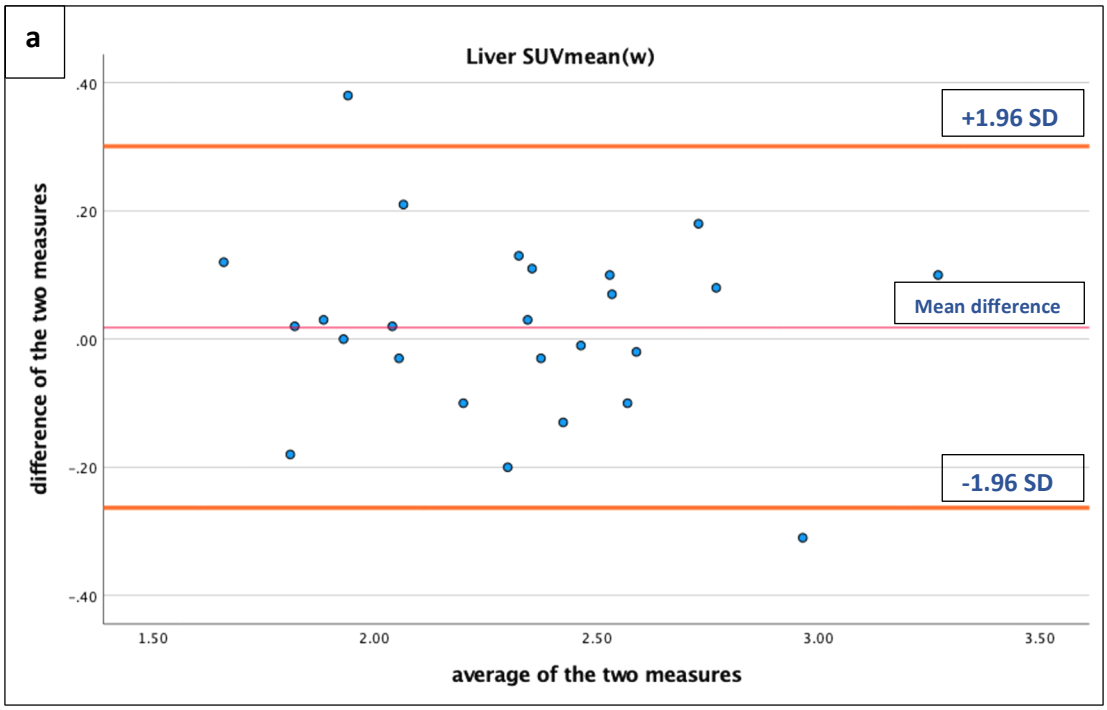
5.3.6.2.2 Background regions

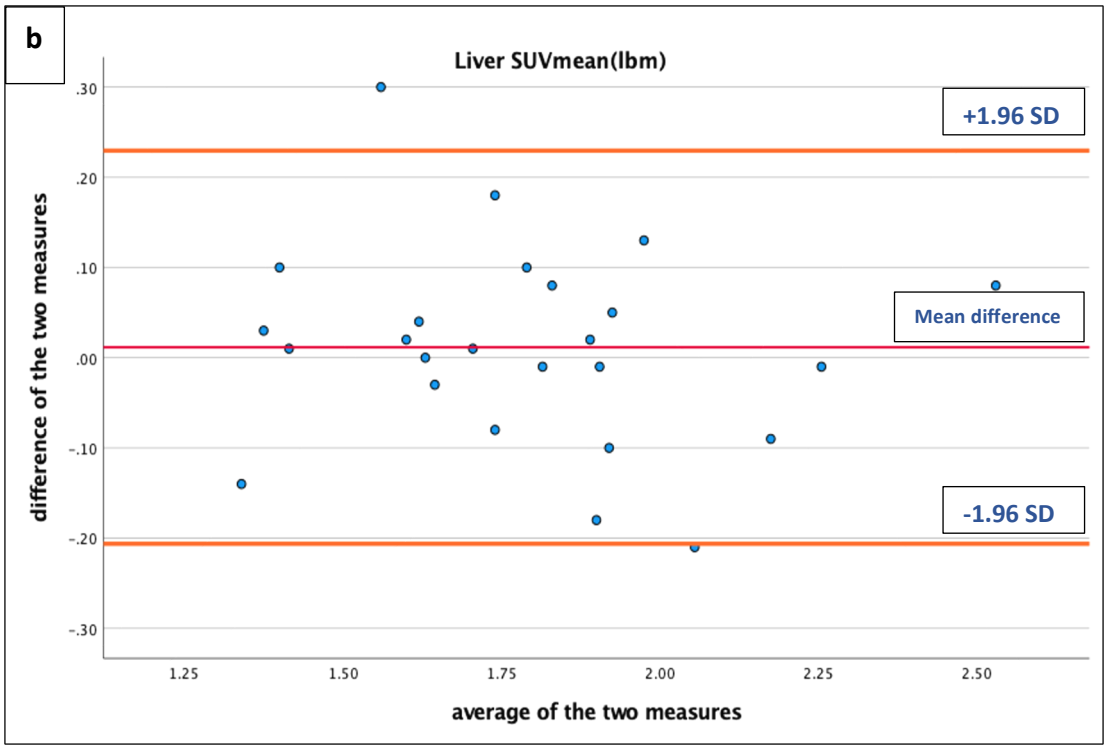
Liver

When it comes to liver SUV metrics, an excellent ICC was also found between two reading times for the liver SUV_{mean} normalised by total body weight 0.966 (95% CI: 0.924-0.985), LBM 0.963 (95% CI: 0.916-0.984), and BSA 0.961 (95% CI: 0.911-0.983). The liver SUV_{max} normalised to weight, LBM, and BSA also showed an excellent ICC (Table 5.8). This means that low variability was observed when the liver SUV metrics were taken twice.

Looking at the Bland-Altman plots of the liver SUV metrics in Figure 5.10, we can see that some data points are far from the middle lines (mean difference), which could be an indication of a weak level of agreement.

When the linear regression analysis was used to further check evidence of proportional bias, the analysis revealed that there was no proportional bias between the two reading times with P values >0.05 (Table 5.9). This may indicate that even though there was some variability when the liver SUV metrics were determined, a generally consistent image analysis process was followed when the liver SUV measurements were taken at different time points.


a
Liver SUVmean(w)
+1.96 SD
Mean difference
-1.96 SD
difference of the two measures
.40
.20
.00
-.20
-.40
1.50
2.00
2.50
3.00
3.50
average of the two measures


b
Liver SUVmean(lbm)
+1.96 SD
Mean difference
-1.96 SD
difference of the two measures
.30
.20
.10
.00
-.10
-.20
-.30
1.25
1.50
1.75
2.00
2.25
2.50
average of the two measures

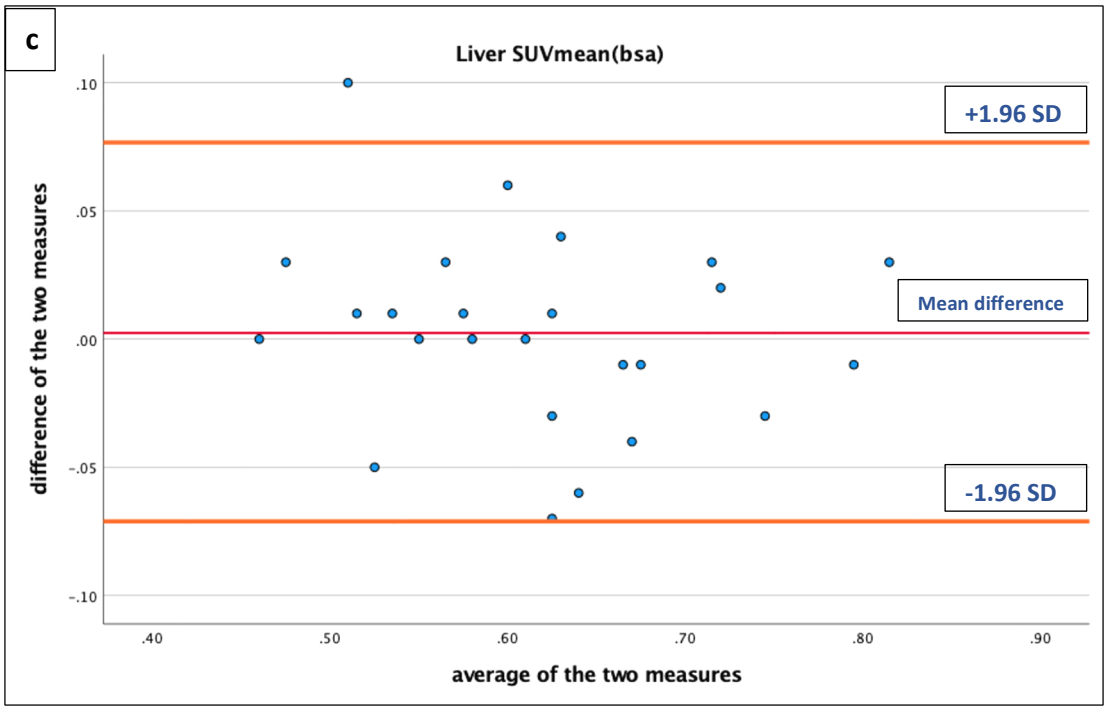


Figure 5.10: Bland-Altman plots display scatter diagrams of the differences in the two reading times of the a) liver SUVmean(w), b) liver SUVmean(lbm), and c) liver SUVmean(bsa)plotted against the averages of the two measurement lines.

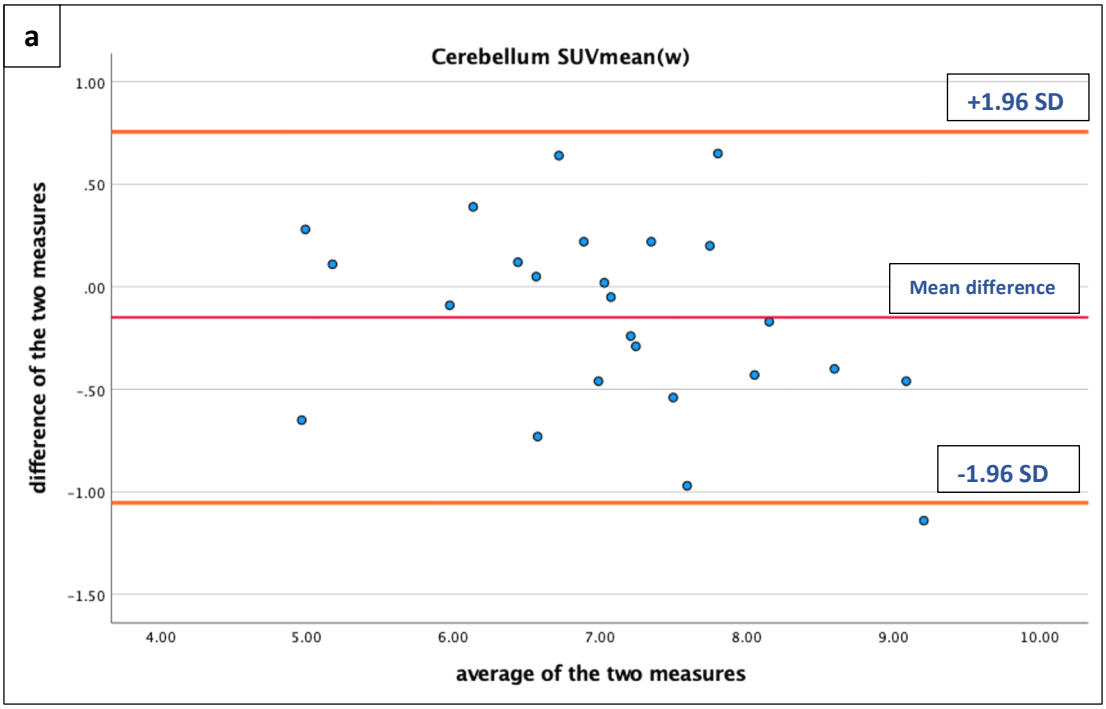
Horizontal lines are drawn at the mean difference (red lines) and at the limits of agreement (orange lines (+1.96 SD for the upper lines and -1.96 SD for the lower lines)). Some data points are far from the middle lines (mean difference), which might indicate a weak level of agreement between the two readings. However, we can see that almost all the dots are within the lines of the agreement limits.

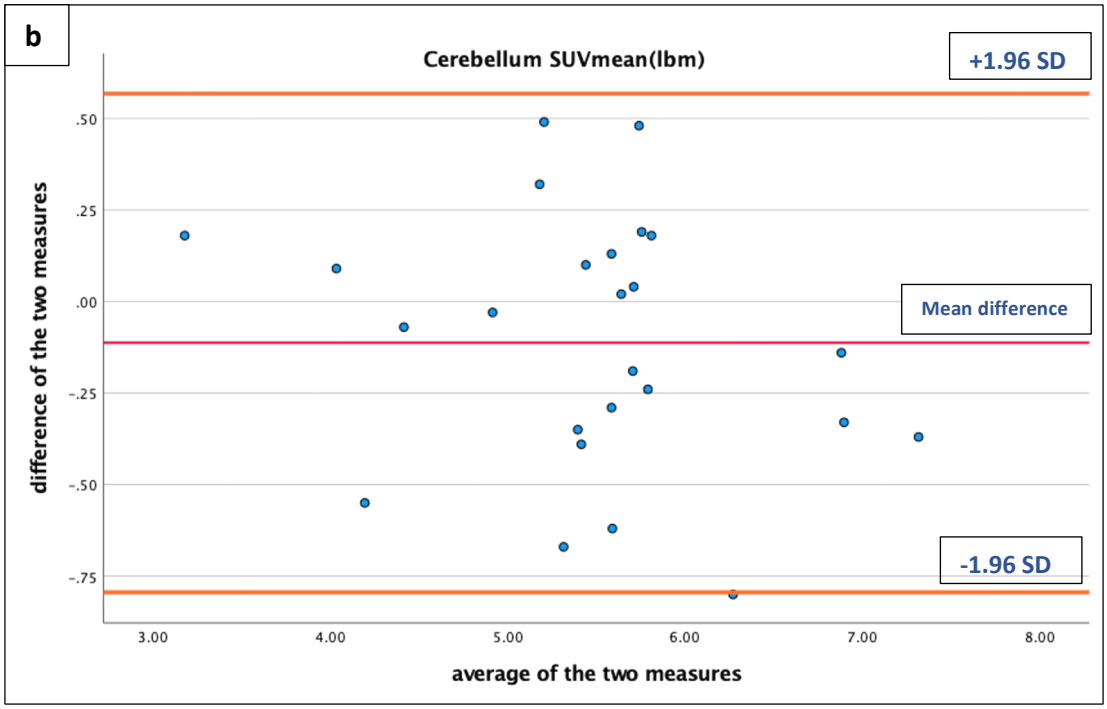
Cerebellum

Similarly, an excellent ICC was also found between two reading times for the cerebellum SUV_{mean} normalised by weight 0.955 (95% CI: 0.898-0.980), LBM 0.960 (95% CI: 0.908-0.982), and BSA 0.963 (95% CI: 0.916-0.984) (Table 5.8). Similar results were also found when the cerebellum SUV_{max} was obtained twice and compared (Table 5.8).

Looking at the Bland-Altman plots of the cerebellum SUV_{mean} normalised by W, LBM, and BSA metrics in Figure 5.11, we can see that some data points are also far from the middle lines (mean difference), which might indicate a weak level of agreement.

However, when the linear regression analysis was used to further check evidence of promotional bias, the analysis revealed no proportional bias between the two readings with P values >0.05 (Table 5.9). This may indicate that even though there was some variability when the cerebellum SUV metrics were analysed, a generally consistent image analysis method was used each time.


a
Cerebellum SUVmean(w)
+1.96 SD
Mean difference
-1.96 SD
difference of the two measures
1.00
.50
.00
-.50
-1.00
-1.50
4.00
5.00
6.00
7.00
8.00
9.00
10.00
average of the two measures


b
Cerebellum SUVmean(lbm)
+1.96 SD
Mean difference
-1.96 SD
difference of the two measures
.50
.25
.00
-.25
-.50
-.75
3.00
4.00
5.00
6.00
7.00
8.00
average of the two measures

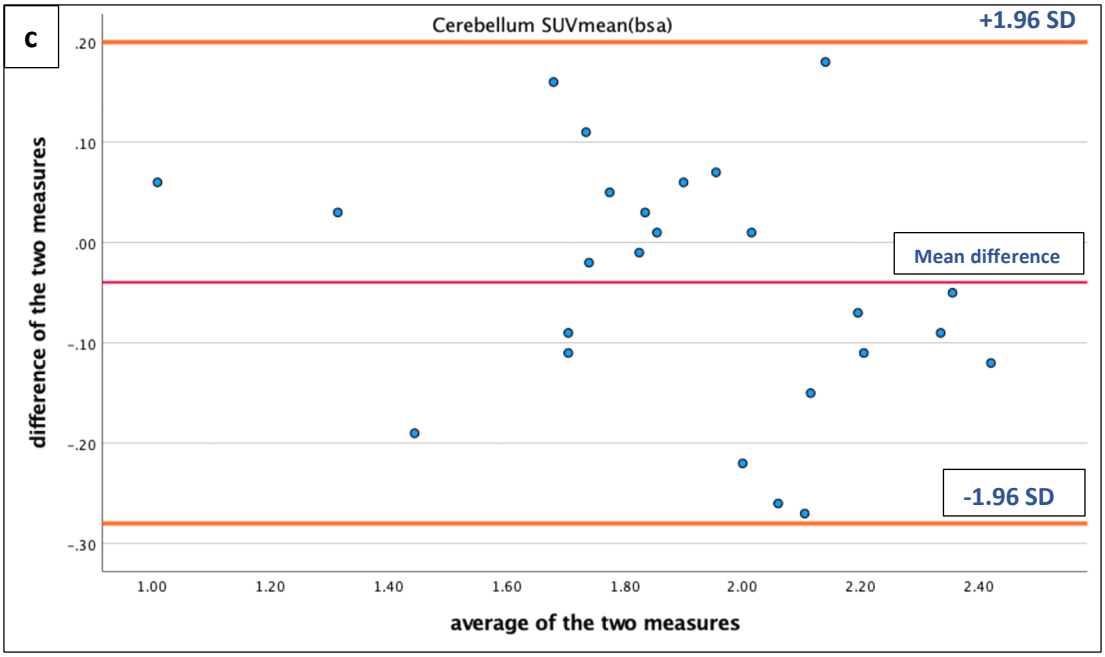


Figure 5.11: Bland-Altman plots display scatter diagrams of the differences in the two reading times of the a) cerebellum SUVmean(W), b) cerebellum SUVmean(lbm) and c) cerebellum SUVmean(bsa) plotted against the averages of the two measurement lines.

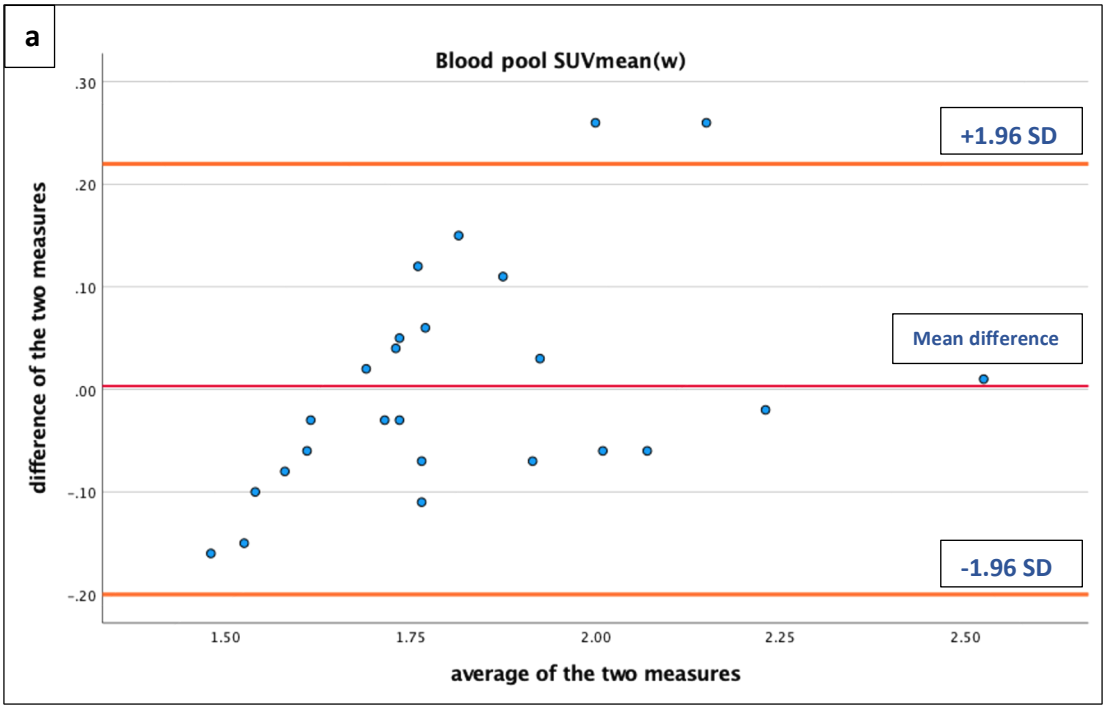
Horizontal lines are drawn at the mean difference (red lines) and at the limits of agreement (orange lines (+1.96 SD for the upper lines and -1.96 SD for the lower lines)). Some data points are far from the middle lines (mean difference), which might indicate a weak level of agreement between the two readings. However, we can see that almost all the dots are within the lines of the agreement limits.

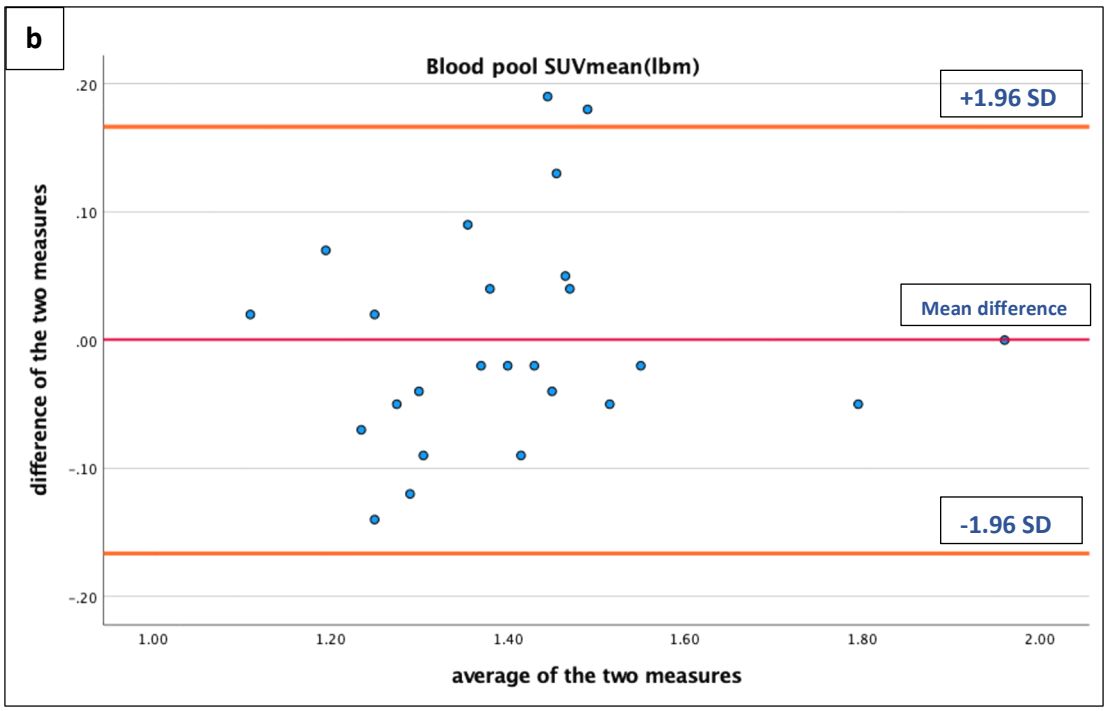
Blood pool (BP)

Regarding the blood pool metrics, a good agreement was also found between the two reading times for the blood pool SUV_{mean} normalised by weight with an ICC of 0.950 (95% CI: 0.887–0.978), LBM 0.947 (95% CI: 0.879–0.977), and BSA 0.938 (95% CI: 0.861–0.973). However, a moderate ICC was found between the reading times for blood pool SUV_{max} normalised by weight, with an ICC of 0.765 (95% CI: 0.473-0.896), LBM 0.738 (95% CI: 0.420-0.883), and BSA 0.743 (95% CI: 0.430-0.886) (Table 5.8), suggesting that some variability could be present when the blood pool measurements were taken at two different time points.

Looking at the Bland-Altman plots in Figure 5.12, we can see that some data points are also far from the middle line (mean difference), which might indicate a weak level of agreement.

The linear regression analysis revealed the non-occurrence of proportional bias between the two reading times of blood pool SUV mean metrics normalised to LBM and BSA (P>0.05). However, it revealed the occurrence of a proportional bias only between the two reading times for the blood pool SUV_{mean} when normalised by total body weight (P =0.040) (Table 5.9). This may indicate that the image analysis process was consistent when the blood pool SUV means of LBM and BSA were analysed, except for the blood pool $SUV_{mean(W)}$.


a
Blood pool SUVmean(w)
+1.96 SD
Mean difference
-1.96 SD
difference of the two measures
.30
.20
.10
.00
-.10
-.20
1.50
1.75
2.00
2.25
2.50
average of the two measures


b
Blood pool SUVmean(lbm)
+1.96 SD
Mean difference
-1.96 SD
difference of the two measures
.20
.10
.00
-.10
-.20
1.00
1.20
1.40
1.60
1.80
2.00
average of the two measures

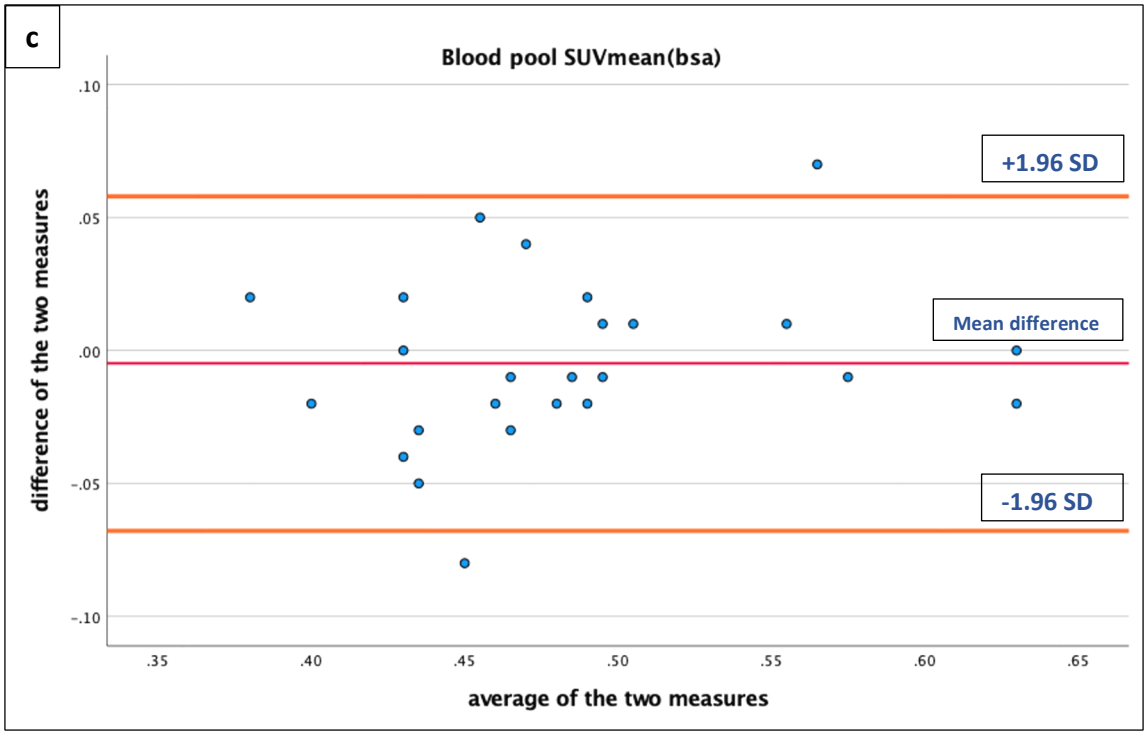


Figure 5.12: Bland-Altman plots display scatter diagrams of the differences in the two reading times of the a) blood SUVmean(w), b) blood SUVmean(lbm), and c) blood SUVmean(bsa) plotted against the averages of the two measurement lines.

Horizontal lines are drawn at the mean difference (red lines) and at the limits of agreement (orange lines (+1.96 SD for the upper lines and -1.96 SD for the lower lines)). Some data points are far from the middle lines (mean difference), which might indicate a weak level of agreement between the two readings. However, we can see that most of the dots are within the lines of the agreement limits.

5.4 Discussion

The main purpose of this chapter was to assess the diagnostic performance of three-month post-treatment primary tumour and lymph node SUV_{max} and SUV_{peak} normalised by the commonly applied normalisation factor, total body weight, to distinguish HNSCC residual disease at three-month post-treatment PET/CT imaging. We also explored if SUV normalisation using LBM, BSA, and lesion-to-background ratio metrics enhanced their diagnostic performance. We then proposed optimal thresholds for primary tumour and nodal SUV_{max} and SUV_{peak}. We have also examined the repeatability of the SUV metrics to check whether the image analysis method used was consistent throughout the process of calculating these quantitative metrics in HNSCC.

Our findings showed that all SUV metrics, including SUV_{max} and SUV_{peak} normalised by weight, LBM, and BSA, were able to differentiate residual HNSCC when assessed at primary and cervical lymph node lesions. We also found that there was no significant performance difference as a result of using different body size metrics. This implies that the use of any size metric for SUV calculation could be acceptable in the interpretation of HNSCC PET/CT imaging following treatment.

No clear evidence was found for diagnostic superiority when lesion SUV metrics were further normalised by the average uptake in the background regions (liver, cerebellum, and BP) to produce reliable measures. We also found low variations in the SUV measurements when

acquired at two times five months apart, indicating that a consistent reproducible image analysis method was used in analysing SUV metrics in this thesis.

The next chapter will evaluate the prognostic value of three-month post-treatment FDG PET/CT metrics to predict survival outcomes.

CHAPTER 6

THE PROGNOSTIC VALUE OF POST CRT FDG PET/CT QUANTITATIVE PARAMETERS IN PREDICTING SURVIVAL OUTCOMES IN PATIENTS WITH HNSCC

6. Chapter 6: The prognostic value of post-CRT FDG PET/CT quantitative parameters in predicting survival outcomes in patients with HNSCC

6.1 Introduction

In the last chapter, we found that post-treatment primary tumour and involved lymph node quantitative metrics could have the potential of discriminating residual disease; however, their prognostic value is still a matter of debate.

Previous studies have investigated the role of SUV_{max} in pre-treatment setting (Kim et al., 2015, Akagunduz et al., 2015, Cacicedo et al., 2017a, Wang et al., 2019a, Creff et al., 2021). While conflicting results have been reported, other studies in post-treatment setting have demonstrated that SUV_{max} could offer more prognostic value for predicting treatment failure and survival outcomes (Moeller et al., 2010, Katahira-Suzuki et al., 2015, Oyama et al., 2020). However, relatively few studies have investigated the ability of other SUV metrics derived at three months post-treatment to predict survival outcomes in HNSCC (Ito et al., 2014, Kim et al., 2016a, Mayo et al., 2019, Connor et al., 2021). Furthermore, little is known about the influence of HPV status on SUV prognostic ability (Helsen et al., 2020a, Connor et al., 2021).

Therefore, the aims of this chapter were:

1) To determine the association of three-month post-treatment SUV_{max} and SUV_{peak} obtained from primary tumour (T) and lymph node (N) sites with:
    - 3-year PFS
    - 5-year OS. Time of death and data on last follow-up within 5 years of treatment completion were collected to analyse the association between PET/CT metrics and the average 5-year OS.

2) To determine whether the prognostic value of PET/CT parameters for survival outcomes is enhanced by employing other SUV normalisation methods such as LBM and BSA.
3) To investigate the difference in prognostic significance based on HPV status.

6.2 Methodology

Patients, scanning protocols, image analysis, and quantitative metrics were previously described in Chapter 3.

The statistical methods used were described in Chapter 3, Section 3.9.3.

6.3 Results

6.3.1 Patient characteristics

Table 6.1 provides a summary of patient and tumour characteristics. 120 patients were included in the analysis. The cohort comprised 90 males (75%) and 30 females (25%), with a median age of 59.2 years (IQR: 52.7-67.7 years).

Based on histology findings, 28/120 (23%) patients experienced residual disease at the time of post-treatment PET/CT (13 primary, 12 neck, and 3 at both sites). 41/120 (34%) patients had overall disease progression within three years post treatment completion.

42/120 (35%) patients died within a median duration of 16 months (IQR: 10.0-30.0 months). The median follow-up duration after treatment completion was 47.0 months (IQR: 40.0-54.0 months).

After completion of CRT, patients underwent PET/CT imaging at a median of 13.4 weeks (IQR: 12.6-13.9 weeks). A median FDG dose of 225.3MBq (IQR: 202.0-265.0 MBq) was administered, followed by a median uptake period of 60.5 min (IQR: 57.0-70.0 min).

Table 6.1: Patient and tumour characteristics

Characteristics	n=120	no (%)
Timing of PET/CT (weeks)	Median	13.4
	IQR	12.6-13.9
Injected dose (MBq)	Median	225.3
	IQR	202.0-265.0
Uptake time (min)	Median	60.5
	IQR	57.0-70.0
Patient age (years)	Median	59.2
	IQR	52.7-65.7
Gender	Male	90 (75.0)
	Female	30 (25.0)
Tumour location	Oropharynx	85 (70.8)
	Larynx	11 (9.2)
	Oropharynx + Larynx	4 (3.3)
	Hypopharynx	8 (6.7)
	Occult primary	12 (10.0)
HPV status*	Positive	66 (55.0)
	Negative	39 (32.5)
	Not recorded	15 (12.5)
T stage	T1	19 (15.8)
	T2	39 (32.5)
	T3	19 (15.8)
	T4	41 (34.2)
	Not recorded	2 (1.7)
N stage	N0	5 (4.2)
	N1	7 (5.8)
	N2	105 (87.5)
	N3	1 (0.8)
	Not recorded	2 (1.7)
Treatment	CRT	107 (89.2)
	RT only	13 (10.8)
Overall LRT at three-month	Yes	92 (76.7)
Local control	Yes	104 (86.7)
Regional control	Yes	105 (87.5)
3-year overall disease progression	Yes	41 (34.2)
3-year primary tumour progression	Yes	21 (17.5)
3-year lymph node progression	Yes	18 (15.0)

5-year overall survival	Alive	78 (65.0)
	Dead	42 (35.0)

*Human papillomavirus (HPV) status of oropharyngeal tumours.
Abbreviations: IQR, interquartile range; T, primary tumour; N, lymph node; LRC, locoregional control, which describes tumour status three months after CRT.

- LRT represents a complete or incomplete response to treatment combining both the primary tumour and the involved lymph node sites.
- Local control represents a complete response to treatment at the primary tumour site.
- Regional control represents a complete response to treatment at the involved lymph node site.
- Overall disease progression represents persistent locoregional disease, locoregional recurrence, or the development of distant metastatic disease.

6.3.2 Prognostic value of imaging variables in HNSCC: initial analysis

6.3.2.1 3-year progression-free survival (PFS): HNSCC

Univariable analysis: PFS

In the analysis of PFS, post-treatment T SUV_{max}, T SUV_{peak}, N SUV_{max}, and N SUV_{peak} all normalised to weight, LBM, and BSA were independently evaluated. This was done to determine whether these variables significantly predicted disease progression within three years after the completion of treatment.

Post-treatment T SUV_{max} demonstrated a significant association in the univariable analysis with PFS, whether it was normalised to weight (HR=1.21; 95% CI: 1.10-1.33, *P*=<0.0001), LBM (HR=1.28; 95% CI: 1.14-1.43, *P*=<0.0001), or BSA (HR=2.12; 95% CI: 1.54-2.91, *P*=<0.0001) (Table 6.2). This suggests that an increase in the primary tumour SUV signifies an increase in the risk of disease progression by a factor of 1.21, 1.28, and 2.12, respectively.

In addition to T SUV_{max}, post-treatment T SUV_{peak} metrics were also shown to be significantly associated with PFS. Post-treatment T SUV_{peak} normalised to weight (HR=1.31; 95% CI: 1.13-1.53, *P*=0.001), LBM (HR=1.42; 95% CI: 1.18-1.70, *P*=<0.0001), and BSA (HR=2.98; 95% CI: 1.78-4.97, *P*=<0.0001) were all prognostic factors for PFS (Table 6.2). This indicates that an increase in these parameters would increase the risk for disease progression within three years post-treatment completion.

Similarly, post-treatment N SUV_{max} normalised to weight (HR=1.42; 95% CI: 1.26-1.60, *P*=<0.0001), LBM (HR=1.58; 95% CI: 1.35-1.86, *P*=<0.0001), and BSA (HR=3.37; 95% CI: 2.18-5.19, *P*=<0.0001) were also prognostic factors for PFS. Post-treatment N SUV_{peak} normalised

by weight (HR=1.74; 95% CI: 1.38-2.19, *P*=<0.0001), LBM (HR=2.06; 95% CI: 1.55-2.74, *P*=<0.0001), and BSA (HR=6.87; 95% CI: 3.04-15.56, *P*=<0.0001) were also significantly associated with PFS (Table 6.2).

Multivariable analysis: PFS

In the multivariable analysis, each imaging parameter was evaluated separately (not all together) in a model incorporating multiple clinical variables (Table 6.3). This was done to prevent the model from being overfit (consisting of too many variables with a limited sample size). Therefore, separate evaluations were conducted for each quantitative metric.

In the multivariable analysis of PFS, HPV status was the only clinical factor that was significantly associated with 3-year PFS (Table 6.3).

All post-treatment primary tumour (T SUV) imaging parameters, including T SUV_{max} normalised to weight (HR=1.24; 95% CI: 1.08-1.42, *P*=0.003), T SUV_{max} normalised to LBM (HR=1.28; 95% CI: 1.09-1.50, *P*=0.003), T SUV_{max} normalised to BSA (HR=2.16; 95% CI: 1.34-3.46, *P*=0.001), T SUV_{peak} normalised to weight (HR=1.46; 95% CI: 1.14-1.88, *P*=0.003), T SUV_{peak} normalised to LBM (HR=1.53; 95% CI: 1.15-2.04, *P*=0.004), and T SUV_{peak} normalised to BSA (HR=3.71; 95% CI: 1.58-8.68, *P*=0.003) showed significant associations with PFS (Table 6.3).

Similarly, all post-treatment N SUV_{max} normalised to weight (HR=1.37; 95% CI: 1.13-1.67, *P*= 0.001), N SUV_{max} normalised to LBM (HR=1.44; 95% CI: 1.13-1.83, *P*=0.003), N SUV_{max}

normalised to BSA (HR=2.89; 95% CI: 1.37-6.09, *P*=0.005), N SUV_{peak} normalised to weight (HR=1.86; 95% CI: 1.31-2.66, *P*=0.01), N SUV_{peak} normalised to LBM (HR=2.11; 95% CI: 1.36-3.27, *P*=0.001), and N SUV_{peak} normalised to BSA (HR=8.18; 95% CI: 2.12-31.57, *P*=0.002) were significantly associated with PFS (Table 6.3). This means that an increase in these SUV metrics suggests an increase in the risk of disease progression, which could suggest the presence of residual disease after treatment, the likelihood of locoregional disease recurrence, or distant metastases within three years of treatment.

6.3.2.2 5-year overall survival (OS): HNSCC

Univariable analysis: OS

In the univariable analysis of OS, the prognostic value of post-treatment T SUV_{max}, T SUV_{peak}, N SUV_{max}, and N SUV_{peak} , all normalised to weight, LBM, and BSA, was independently evaluated for predicting mortality (Table 6.2).

On univariable analysis, post-treatment T SUV_{max} normalised to weight, LBM, and BSA showed significant associations for OS (HR=1.18; 95% CI: 1.08-1.29, *P*=<0.0001), (HR=1.25; 95% CI: 1.12-1.39, *P*=<0.0001), and (HR=1.96; 95% CI: 1.45-2.63, *P*=<0.0001), respectively. Post-treatment T SUV_{peak} normalised to weight, LBM, and BSA all showed significant associations for OS (HR=1.40; 95% CI: 1.18-1.68, *P*=<0.0001), (HR=1.69; 95% CI: 1.32-2.15, *P*=<0.0001), and (HR=5.06; 95% CI: 2.53-10.11, *P*=<0.0001), respectively (Table 6.2).

Similarly, post-treatment N SUV_{max} normalised to weight, LBM, and BSA all had significant associations with OS (HR=1.37; 95% CI: 1.21-1.55, *P*=<0.0001), (HR=1.49; 95% CI: 1.29-1.73, *P* =<0.0001), and (HR=3.36; 95% CI: 2.17-5.22, *P*=<0.0001), respectively. Post-treatment N SUV_{peak} normalised to weight, LBM, and BSA also showed significant associations for OS (HR=1.62; 95% CI: 1.31-1.99, *P*=<0.0001), (HR=1.87; 95% CI: 1.46-2.41, *P*=<0.0001), and (HR=6.87; 95% CI: 3.23-14.61, *P*=<0.0001), respectively (Table 6.2).

Multivariable analysis: OS

Of the clinical factors examined, tumour response to treatment (LRT) and HPV status showed significant associations with 5-year OS (Table 6.3). This suggests that patients with residual

disease at three-month post-treatment PET/CT imaging experienced poorer overall survival. Patient age also showed a trend for association when some of the primary metrics were analysed. However, T stage did not show any association with 5-year OS (Table 6.3).

The only primary tumour imaging metric that showed a significant association with 5-year OS was post-treatment T SUVpeak normalised to BSA (HR=3.73; 95% CI: 1.00-13.89, *P*=0.049) (Table 6.3). On the other hand, post-treatment T SUV_{max} normalised to weight (HR=1.10; 95% CI: 0.95-1.28, *P*=0.219), LBM (HR=1.10; 95% CI: 0.93-1.32, *P*=0.266), and BSA (HR =1.41; 95% CI: 0.83-2.37, *P*=0.201) showed no significant associations with 5-year OS. Similarly, post-treatment T SUV_{peak} normalised to weight (HR=1.45; 95% CI: 1.0-2.10, *P*=0.052) and LBM (HR=1.48; 95% CI: 0.96-2.28, *P*=0.074) also showed no significant associations with 5-year OS.

All post-treatment N SUV_{peak} metrics, including N SUVpeak normalised to weight (HR=1.78; 95% CI: 1.15-2.76, *P*=0.010), LBM (HR=1.95; 95% CI: 1.16-3.26, *P*=0.011), and BSA (HR=7.24; 95% CI: 1.54-34.11, *P*=0.012) showed significant associations with 5-year OS. On the other hand, N SUV_{max} normalised by weight (HR=1.21; 95% CI: 0.97-1.51, *P*=0.091), N SUV_{max} normalised to LBM (HR=1.25; 95% CI: 0.96-1.63, *P*=0.100), and N SUV_{max} normalised to BSA (HR=1.94; 95% CI: 0.88-4.28, *P*=0.100) did not show any association with 5-year OS (Table 6.3).

Table 6.2: Univariable analysis of imaging variables with Cox proportional hazard model for 3-year progression-free survival and 5-year overall survival in combined HNSCC

Variables	3-year Progression-free-survival				5-year Overall-survival			
	HR	**[95% CI]**		***P* value**	**HR**	**[95% CI]**		***P* value**
T SUVmax(w)	1.21	1.10	1.33	**<0.0001**	1.18	1.08	1.29	**<0.0001**
T SUVmax(lbm)	1.28	1.14	1.43	**<0.0001**	1.25	1.12	1.39	**<0.0001**
T SUVmax(bsa)	2.12	1.54	2.91	**<0.0001**	1.96	1.45	2.63	**<0.0001**
T SUVpeak(w)	1.31	1.13	1.53	**0.001**	1.40	1.18	1.68	**<0.0001**
T SUVpeak(lbm)	1.42	1.18	1.70	**<0.0001**	1.69	1.32	2.15	**<0.0001**
T SUVpeak(bsa)	2.98	1.78	4.97	**<0.0001**	5.06	2.53	10.11	**<0.0001**
N SUVmax(w)	1.42	1.26	1.60	**<0.0001**	1.37	1.21	1.55	**<0.0001**
N SUVmax(lbm)	1.58	1.35	1.86	**<0.0001**	1.49	1.29	1.73	**<0.0001**
N SUVmax(bsa)	3.37	2.18	5.19	**<0.0001**	3.36	2.17	5.22	**<0.0001**
N SUVpeak(w)	1.74	1.38	2.19	**<0.0001**	1.62	1.31	1.99	**<0.0001**
N SUVpeak(lbm)	2.06	1.55	2.74	**<0.0001**	1.87	1.46	2.41	**<0.0001**
N SUVpeak(bsa)	6.87	3.04	15.56	**<0.0001**	6.87	3.23	14.61	**<0.0001**

Abbreviations: T, primary tumour; N, lymph node; HR, hazard ratio; CI, confidence interval; SUV_{max}, maximum standardised uptake value; SUV_{peak}, peak standardised uptake value.

Table 6.3: Multivariable analysis of clinical variables with Cox proportional hazard model for 3-year progression-free survival and 5-year overall survival in combined HNSCC

Variables	3-year Progression-free-survival				5-year Overall-survival			
	HR	[95% CI]		*P* value	HR	[95% CI]		*P* value
T SUVmax(w)	1.24	1.08	1.42	**0.003**	1.10	0.95	1.28	0.219
HPV status	0.23	0.09	0.61	**0.003**	0.22	0.06	0.73	**0.014**
Age	1.02	0.96	1.08	0.549	1.06	0.99	1.13	0.100
T Stage (1)	1.73	0.32	9.31	0.521	3.59	0.40	32.42	0.255
T Stage (2)	2.86	0.54	14.98	0.215	1.67	0.16	17.63	0.670
T Stage (3)	1.46	0.30	7.06	0.640	1.33	0.15	12.14	0.801
LRT	–	–	–	–	11.46	2.92	45.03	**<0.0001**
T SUVmax(lbm)	1.28	1.09	1.50	**0.003**	1.10	0.93	1.32	0.266
HPV status	0.26	0.10	0.68	**0.006**	0.23	0.07	0.78	**0.019**
Age	1.01	0.95	1.08	0.644	1.05	0.99	1.13	0.122
T Stage (1)	1.79	0.33	9.58	0.496	3.76	0.42	33.88	0.238
T Stage (2)	2.89	0.55	15.17	0.210	1.71	0.16	18.24	0.655
T Stage (3)	1.53	0.32	7.39	0.599	1.39	0.15	12.65	0.770
LRT	–	–	–	–	11.86	3.04	46.19	**<0.0001**
T SUVmax(bsa)	2.16	1.34	3.46	**0.001**	1.41	0.83	2.37	0.201
HPV status	0.28	0.10	0.73	**0.010**	0.23	0.07	0.82	**0.024**
Age	1.02	0.96	1.08	0.575	1.06	0.99	1.13	0.100
T Stage (1)	1.71	0.32	9.14	0.529	3.64	0.40	32.76	0.250
T Stage (2)	2.87	0.55	15.06	0.213	1.76	0.17	18.71	0.640
T Stage (3)	1.43	0.29	6.96	0.658	1.31	0.14	12.11	0.810
LRT	–	–	–	–	11.47	2.93	44.96	**<0.0001**
T SUVpeak(w)	1.46	1.14	1.88	**0.003**	1.45	1.00	2.10	0.052
HPV status	0.19	0.06	0.58	**0.004**	0.19	0.05	0.76	**0.019**
Age	1.04	0.97	1.11	0.315	1.10	1.02	1.19	**0.010**
T Stage (1)	2.28	0.22	23.41	0.487	1.67	0.16	17.21	0.668
T Stage (2)	4.06	0.45	36.57	0.211	0.65	0.05	8.50	0.746
T Stage (3)	2.59	0.32	20.90	0.372	0.73	0.07	7.11	0.786
LRT	–	–	–	–	18.36	3.71	90.79	**<0.0001**
T SUVpeak(lbm)	1.53	1.15	2.04	**0.004**	1.48	0.96	2.28	0.074
HPV status	0.21	0.07	0.64	**0.006**	0.22	0.05	0.95	**0.043**
Age	1.03	0.96	1.11	0.402	1.09	1.02	1.18	**0.017**
T Stage (1)	2.48	0.25	25.00	0.441	1.83	0.18	18.42	0.606
T Stage (2)	4.17	0.46	37.58	0.203	0.71	0.05	9.52	0.796
T Stage (3)	2.78	0.34	22.43	0.338	0.76	0.08	7.57	0.819
LRT	–	–	–	–	19.18	3.92	93.82	**<0.0001**
T SUVpeak(bsa)	3.71	1.58	8.68	**0.003**	3.73	1.00	13.89	**0.049**
HPV status	0.23	0.07	0.72	**0.011**	0.24	0.05	1.08	0.063
Age	1.03	0.96	1.10	0.413	1.10	1.02	1.18	**0.012**
T Stage (1)	2.39	0.24	24.30	0.461	1.80	0.18	18.29	0.619
T Stage (2)	4.17	0.46	37.66	0.203	0.79	0.06	10.56	0.856
T Stage (3)	2.61	0.32	21.20	0.369	0.72	0.07	7.12	0.776
LRT	–	–	–	–	18.75	3.80	92.50	**<0.0001**
N SUVmax(w)	1.37	1.13	1.67	**0.001**	1.21	0.97	1.51	0.091
HPV status	0.31	0.12	0.79	**0.015**	0.09	0.03	0.32	**<0.0001**
Age	1.02	0.97	1.07	0.482	0.98	0.93	1.04	0.527
T Stage (1)	1.15	0.22	5.86	0.870	0.86	0.08	8.71	0.895
T Stage (2)	1.41	0.24	8.40	0.707	1.35	0.12	15.80	0.809
T Stage (3)	1.17	0.23	5.92	0.845	1.13	0.12	10.86	0.913
LRT	–	–	–	–	6.13	1.78	21.10	**0.004**
N SUVmax(lbm)	1.44	1.13	1.83	**0.003**	1.25	0.96	1.63	0.100
HPV status	0.33	0.13	0.89	**0.028**	0.10	0.03	0.34	**<0.0001**
Age	1.02	0.97	1.07	0.509	0.98	0.93	1.03	0.476
T Stage (1)	1.22	0.24	6.16	0.814	0.91	0.09	9.16	0.938
T Stage (2)	1.45	0.24	8.71	0.686	1.42	0.12	16.41	0.779
T Stage (3)	1.23	0.24	6.19	0.806	1.21	0.13	11.58	0.867
LRT	–	–	–	–	6.62	1.95	22.43	**0.002**
N SUVmax(bsa)	2.89	1.37	6.09	**0.005**	1.94	0.88	4.28	0.100
HPV status	0.34	0.13	0.92	**0.033**	0.10	0.03	0.34	**<0.0001**
Age	1.02	0.97	1.07	0.505	0.98	0.93	1.03	0.467
T Stage (1)	1.21	0.24	6.12	0.818	0.89	0.09	8.97	0.923
T Stage (2)	1.43	0.23	8.76	0.702	1.36	0.12	15.95	0.809
T Stage (3)	1.19	0.24	5.99	0.836	1.16	0.12	11.06	0.899
LRT	–	–	–	–	6.85	2.04	23.06	**0.002**
N SUVpeak(w)	1.86	1.31	2.66	**0.001**	1.78	1.15	2.76	**0.010**
HPV status	0.17	0.06	0.51	**0.001**	0.03	0.01	0.17	**<0.0001**
Age	1.02	0.98	1.07	0.322	0.99	0.94	1.04	0.679
T Stage (1)	1.17	0.22	6.13	0.853	0.35	0.03	4.84	0.434
T Stage (2)	1.01	0.16	6.53	0.988	0.57	0.04	8.02	0.678
T Stage (3)	0.86	0.16	4.63	0.860	0.67	0.06	7.62	0.748
LRT	–	–	–	–	9.65	2.10	44.37	**0.004**
N SUVpeak(lbm)	2.11	1.36	3.27	**0.001**	1.95	1.16	3.26	**0.011**
HPV status	0.20	0.07	0.57	**0.003**	0.04	0.01	0.19	**<0.0001**
Age	1.02	0.98	1.08	0.325	0.98	0.93	1.04	0.572
T Stage (1)	1.24	0.24	6.45	0.802	0.41	0.03	5.38	0.497
T Stage (2)	1.04	0.16	6.69	0.971	0.66	0.05	8.83	0.752
T Stage (3)	0.90	0.17	4.83	0.898	0.80	0.07	8.86	0.855
LRT	–	–	–	–	11.36	2.45	52.68	**0.002**
N SUVpeak(bsa)	8.18	2.12	31.57	**0.002**	7.24	1.54	34.11	**0.012**
HPV status	0.21	0.07	0.62	**0.005**	0.04	0.01	0.19	**<0.0001**
Age	1.03	0.98	1.08	0.296	0.98	0.93	1.04	0.545
T Stage (1)	1.24	0.24	6.47	0.795	0.39	0.03	5.18	0.477
T Stage (2)	1.10	0.17	7.20	0.918	0.62	0.05	8.54	0.723
T Stage (3)	0.87	0.16	4.74	0.876	0.73	0.07	8.11	0.795
LRT	–	–	–	–	12.41	2.69	57.38	**0.001**

Abbreviations: T, primary tumour; N, lymph node; HR, hazard ratio; CI, confidence interval; SUV_{max}, maximum standardised uptake value; SUV_{peak}, peak standardised uptake value; LRT, locoregional control after treatment.

6.3.3 Thresholds' selection

Thresholds' selection methods were previously described in Chapter 3, Section 3.9.3. Tables 6.4, 6.5, and 6.6 illustrate the thresholds used for the Kaplan-Meier analyses.

For PFS, for example, based on a minimal specificity of 80%, the most effective threshold for post-treatment T SUVmax (w) using the ROC curve method was 5.93, showing a sensitivity of 44.12% and a specificity of 81.67%. For N SUVmax (w), it was 3.39 (sensitivity of 47.06% and specificity of 87.50%) (Table 6.4).

For OS, based on a minimal specificity of 80%, the most effective threshold for post-treatment T SUVpeak (bsa) using the ROC curve method was 1.18, showing a sensitivity of 44.83% and a specificity of 81.97%. For N SUVpeak (w), it was 2.43 (sensitivity of 55.56% and specificity of 80.00%) (Table 6.5).

If using the median values, the median values for the imaging metric are detailed in Table 6.6.

Table 6.4: ROC curve threshold values of PET parameters for predicting disease progression.

Parameters	Thresholds	Sensitivity	Specificity
T SUVmax(w)	≥5.93	44.12%	81.67%
T SUVmax(lbm)	≥4.17	50.00%	80.00%
T SUVmax(bsa)	≥1.69	47.06%	85.00%
T SUVpeak(w)	≥4.42	41.94%	81.36%
T SUVpeak(lbm)	≥3.65	35.48%	88.14%
T SUVpeak(bsa)	≥1.29	41.94%	88.14%
N SUVmax(w)	≥3.39	47.06%	87.50%
N SUVmax(lbm)	≥2.48	55.88%	85.71%
N SUVmax(bsa)	≥0.82	50.00%	82.14%
N SUVpeak(w)	≥2.37	62.50%	80.00%
N SUVpeak(lbm)	≥1.86	62.50%	80.00%
N SUVpeak(bsa)	≥0.70	46.88%	89.09%

Abbreviations: T, primary tumour; N, lymph node; SUV_{max}, maximum standardised uptake value; SUV_{peak}, peak standardised uptake value; W, weight in kg; LBM, lean body mass; BSA, body surface area.

Table 6.5: ROC curve threshold values of PET parameters for predicting overall survival.

Parameters	Thresholds	Sensitivity	Specificity
T SUVpeak(bsa)	≥1.18	44.83%	81.97%
N SUVpeak(w)	≥2.43	55.56%	80.00%
N SUVpeak(lbm)	≥2.10	51.85%	93.33%
N SUVpeak(bsa)	≥0.70	51.85%	88.33%

Abbreviations: T, primary tumour; N, lymph node; SUV_{max}, maximum standardised uptake value; SUV_{peak}, peak standardised uptake value; W, weight in kg; LBM, lean body mass; BSA, body surface area.

Table 6.6: Post-treatment PET/CT metrics median values

Parameters	**Median value**
T SUVmax(w)	≥3.90
T SUVmax(lbm)	≥3.22
T SUVmax(bsa)	≥1.12
T SUVpeak(w)	≥3.32
T SUVpeak(lbm)	≥2.62
T SUVpeak(bsa)	≥0.88
N SUVmax(w)	≥2.58
N SUVmax(lbm)	≥1.97
N SUVmax(bsa)	≥0.67
N SUVpeak(w)	≥2.07
N SUVpeak(lbm)	≥1.64
N SUVpeak(bsa)	≥0.55

Abbreviations: T, primary tumour; N, lymph node; SUV_{max}, maximum standardised uptake value; SUV_{peak}, peak standardised uptake value; W, weight in kg; LBM, lean body mass; BSA, body surface area.

6.3.4 Survival analysis

Only parameters that showed a significant association (P<0.05) in the multivariable Cox proportional hazards model (Table 6.3) were incorporated into the Kaplan-Meier analysis (Tables 6.7 and 6.8). This analysis was used to compare patient survival. Due to the lack of clinically pre-defined thresholds, two threshold-selection approaches that were outlined earlier in the chapter were presented. The first one was by using the ROC curve (Sagardoy et al., 2016, Oyama et al., 2020, Wang et al., 2020), and the second one was by using the median values of PET/CT metrics (Shin et al., 2017, Connor et al., 2021). The mean survival values in months were presented instead of the median value. This was because the survival median values were undefined in the analysis. This indicates that more than 50% of the patients in this cohort have not had an event within the follow-up period, as illustrated in Figure 6.1.

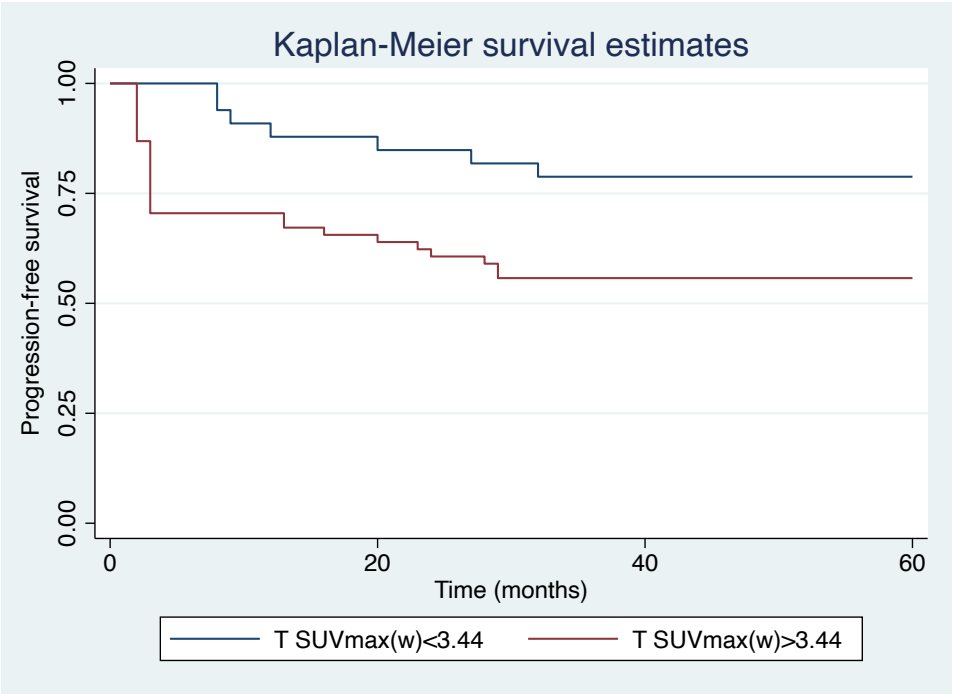


Figure 6.1: The Kaplan–Meier curve depicting the progression-free survival according to the primary tumour (T) SUVmax(w) threshold of 3.44.

One can see that both survival curves in patients with SUVmax <or ≥3.44 did not cross 50% survival. This means that more than half of the patients in both groups had PFS within the follow-up time.

6.3.4.1 The Kaplan-Meir survival analysis of 3-year PFS in HNSCC

From Table 6.7, patient survival differed significantly when groups were divided based on the median PET/CT metrics or ROC curve threshold values. For example, patients with a T SUVmax(lbm) value ≥4.17 had a mean survival time of 19 months, while those with a T SUVmax(lbm) <4.17 had a significantly longer survival time (mean of 30 months); *P*<0.001. When the median value of T SUVmax(lbm) was used as the threshold value (3.22), patients with a threshold ≥3.22 had a mean survival time of 24 months compared to 30 months for the other group (<3.22) (*P*=0.128).

Significant differences in the time to PFS were observed when all post-treatment primary tumour and lymph node metrics were divided according to the ROC curve threshold values (Table 6.7).

Significant differences in the time to PFS were also observed when the thresholds of post-treatment primary tumour SUV_{peak} normalised to weight, LBM, and BSA were set according to the thresholds' median values. However, no significant survival difference was seen when post-primary tumour SUV_{max} normalised to weight, LBM, and BSA thresholds were set according to the threshold' median values (Table 6.7).

Regarding lymph node (N) metrics, all lymph node metrics showed a significant difference in the time to PFS when thresholds were set according to the median value (Table 6.7).
Kaplan-Meier survival curves for some of the parameters were presented in Figures 6.2 and 6.3.

Table 6.7: Comparisons of three-year progression-free survival durations according to median and ROC threshold values of various post-treatment HNSCC primary tumour and nodal SUV metrics normalised to weight, LBM, and BSA

Parameters	Methods	Thresholds	Total N	N of events	Mean Survival (m)	[95% CI]		P value
T SUVmax(w)	ROC	< 5.93	68	19	30	27	32	**0.002**
		≥ 5.93	26	15	19	13	25	
	Median	< 3.9	47	15	29	26	32	0.266
		≥ 3.9	47	19	24	20	29	
T SUVmax(lbm)	ROC	< 4.17	65	17	30	28	33	**<0.001**
		≥ 4.17	29	17	19	13	24	
	Median	< 3.22	47	14	30	27	33	0.128
		≥ 3.22	47	20	24	19	28	
T SUVmax(bsa)	ROC	< 1.69	69	18	30	28	33	**<0.001**
		≥ 1.69	25	16	17	11	23	
	Median	< 1.12	47	14	30	27	33	0.111
		≥ 1.12	47	20	24	19	28	
T SUVpeak(w)	ROC	< 4.42	66	18	30	28	33	**0.004**
		≥ 4.42	24	13	18	12	25	
	Median	< 3.32	45	11	31	28	34	**0.027**
		≥ 3.32	45	20	23	18	28	
T SUVpeak(lbm)	ROC	< 3.65	72	20	29	27	32	**0.002**
		≥ 3.65	18	11	18	10	25	
	Median	< 2.62	45	11	31	28	34	**0.027**
		≥ 2.62	45	20	23	18	28	
T SUVpeak(bsa)	ROC	< 1.29	70	18	30	27	33	**<0.001**
		≥ 1.29	20	13	17	11	24	
	Median	< 0.88	45	10	32	30	34	**0.007**
		≥ 0.88	45	21	22	18	27	
N SUVmax(w)	ROC	< 3.39	67	18	30	28	33	**<0.001**
		≥ 3.39	23	16	14	8	20	
	Median	< 2.58	45	12	31	28	34	**0.011**
		≥ 2.58	45	22	21	17	26	
N SUVmax(lbm)	ROC	< 2.48	63	15	31	29	34	**<0.001**
		≥ 2.48	27	19	14	8	20	
	Median	< 1.97	45	11	32	29	34	**0.003**
		≥ 1.97	45	23	20	16	25	
N SUVmax(bsa)	ROC	< 0.82	63	17	30	27	33	**<0.001**
		≥ 0.82	27	17	16	12	23	
	Median	< 0.675	45	12	31	28	34	**0.011**
		≥ 0.675	45	22	21	17	26	
N SUVpeak(w)	ROC	< 2.37	56	12	32	30	34	**<0.001**
		≥ 2.37	31	20	16	10	21	
	Median	< 2.07	43	10	32	29	34	**0.004**
		≥ 2.07	44	22	21	16	26	
N SUVpeak(lbm)	ROC	< 1.86	53	11	32	29	34	**<0.001**
		≥ 1.86	34	21	17	12	23	
	Median	< 1.64	42	10	31	28	34	**0.009**
		≥ 1.64	45	22	22	17	26	
N SUVpeak(bsa)	ROC	< 0.70	66	17	30	27	33	**<0.001**
		≥ 0.70	21	15	14	7	20	
	Median	< 0.55	43	11	31	28	34	**0.016**
		≥0.55	44	21	22	17	26	

Abbreviations: T, primary tumour; N, lymph node; Total N, number of patients in both groups; m: month; CI, confidence interval, *P* value: Long-rank test P value.

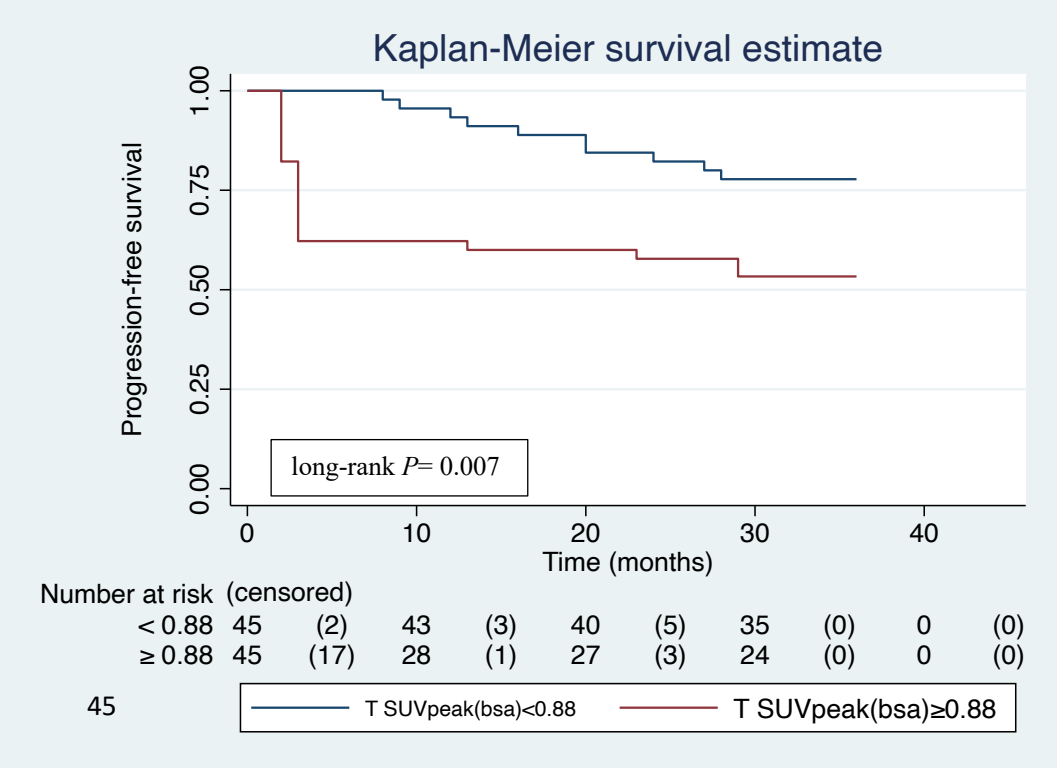
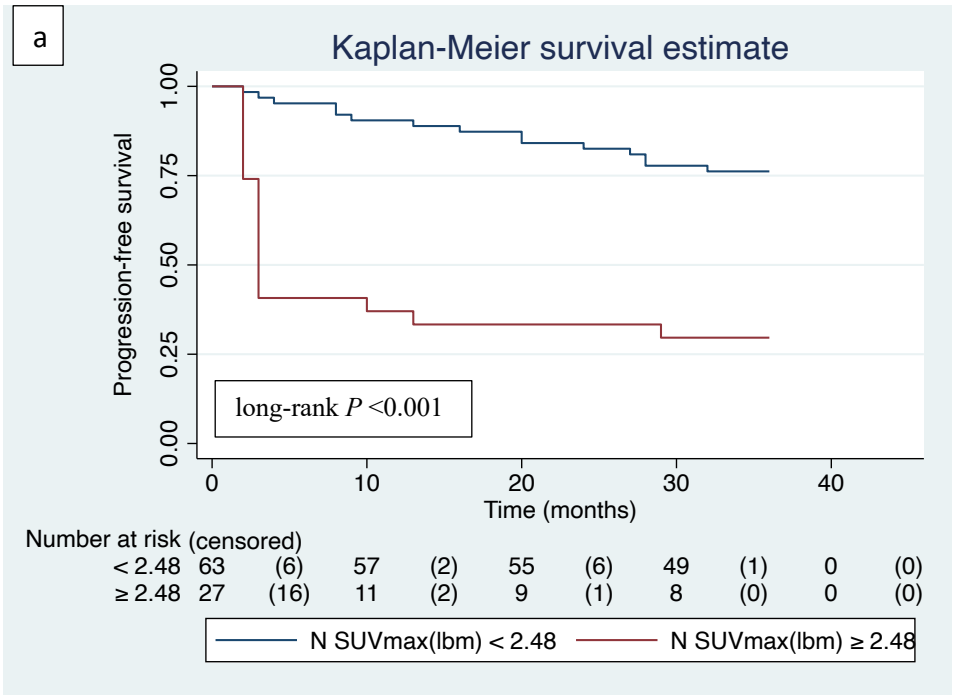


Figure 6.2: The Kaplan–Meier curve depicting 3-year progression-free survival according to the primary tumour (T) SUVpeak(bsa) threshold of 0.88.

At around three months post-treatment, 45 patients had an SUVpeak(bsa) <0.88. Similarly, 45 patients had an SUVpeak(bsa) ≥0.88 at the same time. Of these patients, 10 and 21 patients were censored, respectively, on average of 30 months after treatment. Censored patients included those who experienced disease progression or death. Overall, patients with T SUVpeak(bsa) ≥0.88 appeared to have a shorter mean 3-year PFS (22 vs. 32 months).



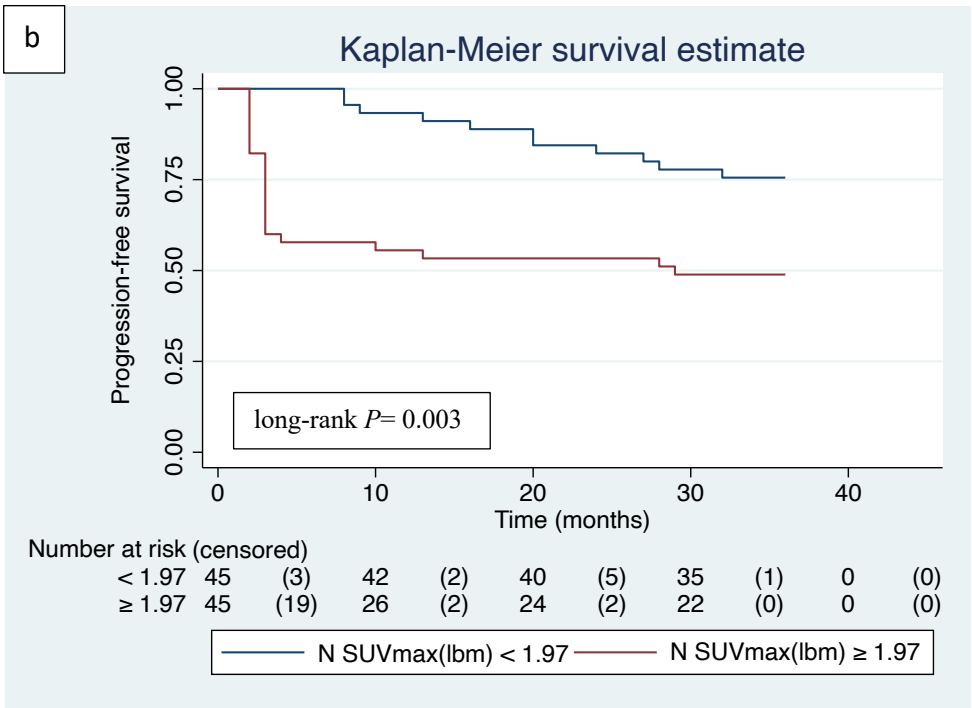


Figure 6.3: The Kaplan–Meier curve depicting 3-year progression-free survival according to the a) lymph node SUVmax(lbm) threshold of 2.48 and b) lymph node SUVpeak(lbm) threshold of 1.97.

Patients with greater thresholds appeared to have shorter mean 3-y PFS (14 vs. 31 months) and (20 vs. 32 months), respectively.

6.3.4.2 The Kaplan-Meir survival analysis of 5-year OS in HNSCC

Table 6.8: Comparisons of five-year overall survival durations according to median and ROC threshold values of various post-treatment HNSCC primary tumour SUVpeak(bsa), and lymph node SUVpeak normalised by weight, lean body mass (LBM), and body surface area (BSA).

Parameters	Methods	Thresholds	Total N	N of events	Mean Survival (m)	[95% CI]		P value
T SUVpeak(bsa)	ROC	< 1.18	66	16	50	46	54	**<0.001**
		≥ 1.18	24	13	38	29	47	
	Median	< 0.88	45	9	52	48	57	0.134
		≥ 0.88	45	20	41	34	48	
N SUVpeak(w)	ROC	< 2.43	60	12	54	50	57	**<0.001**
		≥ 2.43	27	15	36	28	44	
	Median	< 2.07	43	9	53	48	58	**0.025**
		≥ 2.07	44	18	43	37	49	
N SUVpeak(lbm)	ROC	< 2.10	69	13	54	51	58	**<0.001**
		≥ 2.10	18	14	25	18	32	
	Median	< 1.66	42	9	52	48	57	**0.041**
		≥ 1.66	43	17	44	38	51	
N SUVpeak(bsa)	ROC	< 0.70	64	12	54	50	58	**<0.001**
		≥ 0.70	21	14	31	23	40	
	Median	< 0.55	42	10	52	47	57	**0.079**
		≥ 0.55	44	17	43	37	49	

Abbreviations: T, primary tumour; N, lymph node; Total N, number of patients in both groups; m: month; CI, confidence interval, *P* value: Long-rank test P value.

6.3.5 Sub-group analysis for HPV^{+ve} OPSCC versus HPV^{-ve} HNSCC

In this sub-analysis, two groups were compared. The first group involved patients with HPV^{+ve} OPSCC. The second group included all patients with HPV^{-ve} HNSCC. We excluded patients with OPSCC whose HPV status was unknown (15 patients). Patient age and imaging factors were evaluated in the multivariable analysis. However, due to small frequencies in some sub-categories, the variable T stage was removed to prevent unstable risk estimates.

From Table 6.9, we can see those patients with HPV^{-ve} cancer had poor treatment outcomes and survival compared to patients with HPV^{+ve} OPSCC. This is to be expected, as the impact of HPV infection on prognosis in HNSCC is well documented (Fakhry et al., 2008, Lechner et al., 2022). In the group of patients with HPV^{+ve} disease, only 10 (15.2%) had residual disease at the time of post-treatment PET/CT, 14 (21.2%) experienced disease progression, and 10 (15.2%) died within five-year post-treatment. On the other hand, in the group of patients with HPV^{-ve} HNSCC, 16 (41.0%) patients had residual disease at the time of post-treatment PET/CT, 21 (53.9%) experienced disease progression, and 25 (64.1%) died within five years post-treatment (Table 6.9).

Table 6.9: Comparison of the overall outcome rates of residual disease, progressed disease, and death in patients with HPV^{+ve} OPSCC and in patients with HPV^{-ve} HNSCC.

	HPV^{+ve} OPSCC (66 patients)		HPV^{-ve} HNSCC (39 patients)	
Outcome	Yes	No	Yes	No
LRT	56 (84.8%)	10 (15.2%)	23 (59.0%)	16 (41.0%)
Progression	14 (21.2%)	52 (78.8%)	21 (53.9%)	18 (46.2%)
Death	10 (15.2%)	56 (84.8%)	25 (64.1%)	14 (35.9%)

Abbreviations: LRT, locoregional control at 3-month after treatment completion; HPV^{+ve}, positive HPV status; HPV^{-ve}, negative HPV status.

6.3.5.1 3-year progression-free survival (PFS) in HPV^{+ve} OPSCC versus HPV^{-ve} HNSCC

Patient age was not a significant prognostic factor for PFS.

All post-treatment primary tumour and lymph node metrics in both groups appeared to be significantly associated with PFS (Table 6.10).

Primary tumour - PFS

In the sub-group of HPV^{+ve} OPSCC, T SUV_{max} normalised to weight (HR=1.26; 95% CI: 1.03-1.54, *P*=0.026), LBM (HR=1.32; 95% CI: 1.01-1.72, *P*=0.044), and BSA (HR=2.42; 95% CI: 1.10-5.33, *P*=0.028), as well as T SUV_{peak} normalised to weight (HR =1.44; 95% CI: 1.10-1.90, *P*=0.009), LBM (HR=1.45; 95% CI: 1.04-2.02, *P*=0.027), and BSA (HR=3.15; 95% CI: 1.19-8.35, *P*=0.021) were associated with disease progression.

Nodal disease - PFS

N SUV_{max} normalised to weight (HR=1.54; 95% CI: 1.21-1.95, *P*=<0.0001), LBM (HR=1.68; 95% CI: 1.24-2.29, *P*=0.001), and BSA (HR = 4.64; 95% CI: 1.85-11.61, *P*=0.001), as well as N SUV_{peak} normalised to weight (HR=1.90; 95% CI: 1.31-2.76, *P*=0.001), N SUV_{peak} LBM (HR=2.20; 95% CI: 1.36-3.55, *P*=0.001), and BSA (HR=9.95; 95% CI: 2.39-41.45, *P*=0.002) were associated with disease progression.

Furthermore, similar associations were also observed in the subgroup of patients with HPV^{-ve} HNSCC (Table 6.10). Therefore, this suggests that an increase in all these metrics is predictive of the risk of disease progression within three years post-treatment.

Table 6.10: Multivariable analysis of clinical variables with Cox proportional hazard model for 3-year progression-free survival in HPV^{+ve} OPSCC and HPV^{-ve} HNSCC.

Variables	3-year Progression-free-survival							
	HPV positive OPSCC				HPV negative HNSCC			
	HR	[95% CI]		*P* value	HR	[95% CI]		*P* value
T SUVmax(w)	1.26	1.03	1.54	**0.026**	1.20	1.04	1.38	**0.011**
Age	1.03	0.96	1.09	0.450	1.01	0.95	1.07	0.736
T SUVmax(lbm)	1.32	1.01	1.72	**0.044**	1.24	1.05	1.46	**0.009**
Age	1.02	0.95	1.09	0.633	1.01	0.95	1.07	0.704
T SUVmax(bsa)	2.42	1.10	5.33	**0.028**	1.96	1.22	3.15	**0.005**
Age	1.02	0.96	1.09	0.546	1.02	0.96	1.08	0.615
T SUVpeak(w)	1.44	1.10	1.90	**0.009**	1.58	1.09	2.29	**0.015**
Age	1.06	0.98	1.15	0.125	1.01	0.95	1.07	0.772
T SUVpeak(lbm)	1.45	1.04	2.02	**0.027**	1.71	1.13	2.58	**0.011**
Age	1.06	0.98	1.15	0.154	1.01	0.95	1.08	0.727
T SUVpeak(bsa)	3.15	1.19	8.35	**0.021**	5.74	1.67	19.78	**0.006**
Age	1.06	0.98	1.15	0.143	1.02	0.96	1.08	0.597
N SUVmax(w)	1.54	1.21	1.95	**<0.0001**	1.34	1.12	1.60	**0.001**
Age	1.00	0.94	1.07	0.880	0.97	0.91	1.03	0.306
N SUVmax(lbm)	1.68	1.24	2.29	**0.001**	1.47	1.16	1.86	**0.001**
Age	1.00	0.94	1.07	0.995	0.97	0.91	1.04	0.416
N SUVmax(bsa)	4.64	1.85	11.61	**0.001**	2.74	1.44	5.22	**0.002**
Age	1.00	0.94	1.07	0.927	0.96	0.90	1.03	0.290
N SUVpeak(w)	1.90	1.31	2.76	**0.001**	1.82	1.20	2.76	**0.005**
Age	1.03	0.96	1.10	0.427	0.96	0.90	1.03	0.286
N SUVpeak(lbm)	2.20	1.36	3.55	**0.001**	2.11	1.25	3.57	**0.005**
Age	1.02	0.95	1.10	0.535	0.97	0.91	1.04	0.415
N SUVpeak(bsa)	9.95	2.39	41.45	**0.002**	5.45	1.45	20.48	**0.012**
Age	1.03	0.96	1.10	0.454	0.97	0.90	1.03	0.319

Abbreviations: T, primary tumour; N, lymph node; HR, hazard ratio; CI, confidence interval; SUV_{max}, maximum standardised uptake value; SUV_{peak}, peak standardised uptake value; OPSCC, oropharyngeal squamous cell carcinoma, HNSCC, head and neck squamous cell carcinoma.

6.3.5.2 5-year overall survival (OS) in HPV^{+ve} OPSCC versus HPV^{-ve} HNSCC

The only clinical factor that showed an association in most sub-groups with OS was LRT (Table 6.11). This suggested that patients with incomplete treatment response at three-month PET/CT experienced unfavourable survival. In addition, patient age was associated with 5-year OS in some subgroups.

Primary tumour - OS

In the analysis of patients with HPV^{+ve} OPSCC, T SUV_{max} normalised to LBM (HR=1.43; 95% CI: 1.01-2.03-1.54, *P*=0.044) and BSA (HR=3.14; 95% CI: 1.06-9.29, *P*=0.039), in addition to T SUV_{peak} normalised to weight (HR=1.92; 95% CI: 1.02-3.61, *P*=0.044), were associated with the time to death.

Nodal disease - OS

In the analysis of patients with HPV^{+ve} OPSCC , N SUV_{max} normalised to weight (HR=1.47; 95% CI: 1.02-2.11, *P*=0.039), LBM (HR = 1.62; 95% CI: 1.03-2.54, *P* =0.036), and BSA (HR=3.97; 95% CI: 1.05-15.01, *P*=0.043), as well as N SUV_{peak} normalised to weight (HR=2.88; 95% CI: 1.07-7.74, *P*=0.036) and BSA (HR=24.01; 95% CI: 1.02-566.34, *P*=0.049), were associated with the time to death.

Furthermore, in the analysis of patients with HPV^{-ve} HNSCC, only two imaging parameters showed an association with OS. These are N SUV_{peak} normalised to LBM (HR= 2.09; 95% CI: 1.04-4.19, *P*=0.038) and N SUV_{peak} normalised to BSA (HR=6.21; 95% CI: 1.21-31.96, *P*=0.029) (Table 6.11).

Table 6.11: Multivariable analysis of clinical variables with Cox proportional hazard model for 5-year overall survival in HPV^{+ve} OPSCC and HPV^{-ve} HNSCC.

	5-year Overall survival							
	HPV positive OPSCC				HPV negative HNSCC			
Variables	HR	[95% CI]		*P* value	HR	[95% CI]		*P* value
T SUVmax(w)	1.29	0.99	1.68	0.055	1.01	0.85	1.19	0.924
Age	1.07	0.99	1.15	0.073	1.01	0.96	1.07	0.726
LRT	8.24	1.25	54.47	**0.029**	4.89	1.60	14.96	**0.005**
T SUVmax(lbm)	1.43	1.01	2.03	**0.044**	1.01	0.83	1.23	0.899
Age	1.05	0.98	1.13	0.129	1.01	0.96	1.07	0.719
LRT	10.12	1.68	60.84	**0.011**	4.84	1.57	14.93	**0.006**
T SUVmax(bsa)	3.14	1.06	9.29	**0.039**	1.02	0.56	1.85	0.948
Age	1.07	0.99	1.14	0.079	1.01	0.95	1.07	0.736
LRT	8.79	1.34	57.55	**0.023**	4.91	1.54	15.64	**0.007**
T SUVpeak(w)	1.92	1.02	3.61	**0.044**	1.05	0.69	1.61	0.808
Age	1.12	1.01	1.23	**0.025**	1.01	0.96	1.07	0.675
LRT	11.90	1.04	136.09	**0.046**	4.98	1.53	16.19	**0.008**
T SUVpeak(lbm)	2.03	0.96	4.31	0.066	1.07	0.66	1.73	0.784
Age	1.08	1.00	1.18	0.065	1.01	0.96	1.07	0.664
LRT	18.63	2.00	173.19	**0.010**	4.91	1.49	16.25	**0.009**
T SUVpeak(bsa)	10.15	0.84	122.03	0.068	1.11	0.24	5.08	0.889
Age	1.10	1.00	1.20	**0.042**	1.01	0.95	1.08	0.704
LRT	16.03	1.62	158.96	**0.018**	5.07	1.43	17.90	0.012
N SUVmax(w)	1.47	1.02	2.11	**0.039**	1.18	0.95	1.47	0.144
Age	0.96	0.90	1.02	0.215	1.00	0.95	1.05	0.972
LRT	4.63	0.75	28.66	0.099	3.79	1.00	14.28	**0.049**
N SUVmax(lbm)	1.62	1.03	2.54	**0.036**	1.28	0.97	1.69	0.081
Age	0.95	0.89	1.02	0.153	1.00	0.95	1.06	0.916
LRT	5.58	0.97	31.97	0.054	3.75	1.04	13.58	**0.044**
N SUVmax(bsa)	3.97	1.05	15.01	**0.043**	2.06	0.97	4.34	0.058
Age	0.95	0.89	1.02	0.159	1.00	0.95	1.05	0.947
LRT	5.68	0.99	32.52	**0.051**	3.93	1.11	13.89	**0.034**
N SUVpeak(w)	2.88	1.07	7.74	**0.036**	1.44	0.87	2.37	0.155
Age	0.98	0.92	1.05	0.546	1.00	0.95	1.05	0.918
LRT	3.98	0.44	35.57	0.217	3.69	0.85	16.14	0.082
N SUVpeak(lbm)	3.85	0.83	17.78	0.084	2.09	1.04	4.19	**0.038**
Age	0.96	0.90	1.04	0.313	1.00	0.95	1.06	0.910
LRT	6.03	0.76	47.59	0.088	3.02	0.70	13.04	0.140
N SUVpeak(bsa)	24.01	1.02	566.34	**0.049**	6.21	1.21	31.96	**0.029**
Age	0.97	0.91	1.04	0.349	0.99	0.94	1.05	0.803
LRT	6.89	0.93	50.86	0.058	3.64	0.93	14.35	0.064

Abbreviations: T, primary tumour; N, lymph node; HR, hazard ratio; CI, confidence interval; SUV_{max}, maximum standardised uptake value; SUV_{peak}, peak standardised uptake value; OPSCC, oropharyngeal squamous cell carcinoma, HNSCC, head and neck squamous cell carcinoma

6.3.6 Discussion

This chapter investigated the significance of three-month post-CRT SUV indices determined independently from primary tumour and involved lymph node lesions in predicting 3-year PFS and 5-year OS. It also examined the influence of HPV status as well as the use of different body size metrics in SUV calculation on their prognostic ability.

We found that post-treatment primary tumour and nodal SUV_{max} and SUV_{peak} normalised by weight in all subgroups (combined HNSCC, HPV^{+e} OPSCC, and HPV^{-ve} HNSCC) appeared to be significant prognostic factors for predicting three-year PFS. An increase in the values of post-treatment SUV_{max} and SUV_{peak} could be indicative of disease progression within the first three years post-CRT. For the analysis of 5-year OS, some metrics demonstrated an association with OS, while others did not. Therefore, we were unable to reach a definitive conclusion regarding that outcome.

Importantly, we found that the use of different body size normalisations had no clear effect on the predictive value of the imaging parameters.

We discussed this in more detail in Chapter 7.

CHAPTER 7

GENERAL DISCUSSION

7. Chapter 7: General discussion

The research aim of this thesis is to investigate the utility of post-treatment FDG PET/CT quantitative metrics to reliably identify residual disease and predict survival outcomes in patients with HNSCC. This aim is met with the findings presented in Chapters 4 to 6. First, Chapter 4 presents the impact of body size factors on quantitative metrics and identifies the best method for normalising these imaging metrics. Chapter 5 investigates the diagnostic performance of both absolute and relative SUV metrics, e.g. lesion-to-background ratios. Finally, Chapter 6 examines the prognostic value of these parameters in predicting survival outcomes, as well as the impact of HPV status on their prognostic value.

7.1 Discussion of findings

7.1.1 Correlation of PET/CT measurements with body metrics

It is widely claimed that normalisation of SUV to LBM, rather than body weight alone, is the recommended standard to achieve greater accuracy when analysing FDG uptake in normal tissues, especially in obese patients (Keramida and Peters, 2019, Sarikaya et al., 2020). However, few studies have assessed this in patients with cancer lesions (Hallett et al., 2001).

The study presented in this thesis demonstrated that post-treatment HNSCC SUV_{max} and SUV_{peak} normalised to weight, LBM, or, BSA had no or only minimal correlation with all the patient body size measurements, including height, weight, LBM, BSA, and BMI. Similarly, MTV and TLG did not correlate with these factors.

The results of our study are consistent with a previous publication in the field of cancer (Hallett et al., 2001). In that study of 154 lung cancer patients, it was found that lesions SUV_w and lesions SUV_{lbm} based on height and weight from a nomogram chart, and lesions SUV_{bsa} did not correlate with body size. However, a significant correlation was found between lesions SUV_{lbm} and height when LBM was calculated as a function of height alone (Hallett et al., 2001). This supports our findings that HNSCC SUV_{max} normalised to total body weight, LBM, and BSA do not appear to be correlated with different body size measurements, suggesting that the use of any body measurements for calculating SUV_{max} and SUV_{peak} can be used when analysing post-treatment scans of HNSCC patients.

We speculated on a possible reason for the absence of a significant correlation. The median value of the patients' weight in this cohort was close to a population average (BMI of 24.4;

IQR: 18.0–37.0 kg/m^2). According to Sarikaya et al. (2020), the impact of using different body size metrics for SUV calculation is more prominent in obese patients than in patients within the recommended weight range (Sarikaya et al., 2020). Also, the absence of a correlation may be the result of a statistical error, specifically Type II errors, which can occur if using a too small sample size, resulting in missing significant associations. To answer this question, therefore, it would be necessary to identify a larger cohort of post-treatment patients who have a higher-than-average BMI to determine the true effect of LBM and BSA on SUV readings.

As mentioned previously in Chapter 1, the effect of body size normalisation was more commonly studied in normal tissues than in malignant lesions (Zasadny and Wahl, 1993, Kim et al., 1994, Sanghera et al., 2009, Keramida and Peters, 2019, Sarikaya et al., 2020). Hence, we performed a sub-analysis of the liver to compare the effects of using different body size measures in both normal and abnormal tissues acquired from the same patient. We found that the normal liver SUV metrics correlated more with total body weight and less with LBM and BSA. Our findings on liver metrics are consistent with those of previous studies (Keramida and Peters, 2019, Sarikaya et al., 2020). Sarikaya et al. (2020) found that although SUV overestimation was prevalent in heavy patients, LBM was found to be a better body size metric in calculating SUVs derived from the liver in patients with both normal and higher BMI (≥30) (Sarikaya et al., 2020). Another study found that the liver SUV_{mean} and SUV_{max} normalised by total body weight correlated significantly with total body weight (Keramida and Peters, 2019). Other studies of normal tissues have also favoured the use of LBM (Zasadny and Wahl, 1993, Sanghera et al., 2009, Keramida and Peters, 2019). The increased correlation observed with total body weight could be attributed to the relatively limited uptake of FDG

into adipose tissue (Christen et al., 2010), resulting in an increase in SUV values in other tissues. In contrast, LBM has an advantage over total body weight because it ignores the percentage of body fat. This may explain why SUV normalised to LBM is less sensitive to patients' weight. SUV normalised to BSA, on the other hand, may be underestimated in obese patients because it increases with a rising fat percentage, with no change in LBM (Keramida and Peters, 2019). Therefore, LBM appears to be the preferable method for calculating SUV metrics and is potentially less likely to introduce bias when reporting liver SUV readings.

Overall, our findings show that the effect of applying alternative body size measures may be greater when analysing SUV metrics for normal tissues as opposed to SUV metrics for lesions obtained post-treatment. Furthermore, additional research with a larger sample size (including a greater number of patients with a raised BMI) examining SUVs from tumour lesions and normal background regions from the same patients is recommended to validate our conclusions.

Notably, finding some correlation between the liver SUV metrics and the body size factors, while not seeing this with lesion SUV metrics and patients' factors obtained from the same cohort, was unexpected. This could be related to the difference in the rate of FDG uptake and excretion between normal and pathological tissues, or it may be due to the intratumour heterogeneity of head and neck tumours. The current consensus classifies tumours of HNSCC according to the molecular classification of HNSCC, which divides tumours into classical (CL), basal (BA), mesenchymal (MS) and atypical (AT) subtypes, each carrying a unique gene expression profile and biological characteristic (Canning et al., 2019). These subtypes may

have variations in tumour metabolism compared to each other, and also to normal tissues, with consequent differential FDG uptake and discrepancies in lesion SUV readings compared to those determined from normal tissues. In our study, subgroup analysis was not performed due to the insufficient data regarding the type of tumour that was included in this cohort. Therefore, further research is necessary to examine the influence of normal tissues and subtypes of head and neck cancer tumours on SUV readings.

Volumetric parameters constitute the second type of PET/CT measurements. These consist of MTV and TLG. A major difference between SUV metrics and volumetric parameters is that SUV metrics, such as SUV_{max}, represent the activity concentration derived from a single voxel, whereas MTV and TLG reflect radioactivity from the entire tumour volume (Im et al., 2018). Therefore, the latter are potentially considered better indicators of tumour burden. Similar to PET/CT semiquantitative parameters, numerous studies have investigated the utility of MTV and TLG in predicting treatment response and patient outcome (Pak et al., 2014, Wang et al., 2019b, Rijo-Cedeno et al., 2021). This study found that neither MTV nor TLG normalised to weight, LBM, and BSA correlated with any of the body measurement variables. Therefore, tumour MTV and TLG are statistically independent of patient body measurements. To the best of our knowledge, no previous studies have examined the correlation between volumetric parameters and patient body size measurements in HNSCC. The limited number of studies in this area could be attributed to the difficulty of segmenting HNSCC tumours in a post-treatment setting. Post-treatment tumours often have lower FDG uptake compared to pre-treatment tumours, requiring additional time when segmenting their borders, particularly in cases with equivocal lesions. Hence, further work analysing quantitative and

volumetric measurements using an advanced image analysis algorithm, such as radiomic and artificial intelligence approaches, could facilitate and provide a more accurate method for image quantification. This is discussed further in Section 7.4.

In summary, we found that all post-treatment quantitative metrics were independent of patient body size. This has important clinical implications, as any normalisation method can be used without concern that body weight will bias imaging metrics.

The next questions we aimed to investigate were whether post-treatment primary tumour and lymph node metrics can reliably identify residual disease and whether using LBM, BSA, and lesion-to-background ratios improved their diagnostic performance. A secondary objective was to investigate whether the method of image analysis applied in this thesis was consistent to ensure the reproducibility of these parameters.

7.1.2 The diagnostic performance of the post-treatment metrics

Few studies have investigated the diagnostic performance of PET/CT quantitative parameters acquired three months post-CRT in HNSCC (Sagardoy et al., 2016, Nelissen et al., 2017, Fatehi et al., 2019b, Helsen et al., 2020a). As a result, there is no consensus regarding optimal thresholds derived from primary tumour and involved lymph nodes to predict residual disease at three months post-treatment.

To establish an optimal threshold, SUV harmonisation is necessary. This process could include using a standardised imaging protocol and image quality and employing harmonisation phantoms to acquire reproducible quantitative metrics (Akamatsu et al., 2023). Despite the

inability of using harmonisation phantom in our study, the sensitivity analysis (Appendix 6) confirmed that the use of two different scanners had no significant impact on SUV readings.

Primary tumour

Our study found that post-treatment primary tumour SUV_{max} normalised to total body weight yielded a high AUC value (0.88, 95% CI: 0.78-0.97) (Chapter 5, Table 5.3), at an optimal threshold of 5.93.

Our suggested threshold differs from a previous publication (Sagardoy et al., 2016), which in 47 patients (with post-treatment imaging at a median of 3.3 months) found that a primary tumour $SUV_{max(w)}$ threshold of 3.7 (sensitivity: 73.3%; specificity: 87.1%) provided the best discriminative performance. This large difference in the threshold values (3.7 and 5.93) could be attributable to the limited sample size in Sagardoy's study, which evaluated only 39 patients (primary tumour analysis) compared to our study (124 patients). Another possible explanation could be the difference in the subsites of the lesions studied. Most of the patients in our study had oropharyngeal cancer. Sagradoy, on the other hand, looked at all HNSCC, although he did not provide data on the number of each head and neck cancer subtype. Compared to other subsites, the oropharynx exhibits a higher rate of inflammation due to treatment (Sroussi et al., 2017). The presence of inflammation could therefore interfere with SUV readings and their corresponding thresholds. Importantly, the influence of HPV status on treatment response and survival should not be ignored. Lesions that are HPV^{+ve} are more sensitive to treatment and respond better than lesions that are HPV^{-ve} (Perri et al., 2020). We hypothesised that SUV metrics would be lower in HPV^{+ve} lesions than in HPV^{-ve} lesions. This was confirmed by our sub-analysis in Chapter 5 (Table 5.5), which compared the median

values of post-treatment primary tumour SUV in HPV^{+ve} OPSCC and HPV^{-ve} HNSCC. As expected, we found that all SUVs in HPV^{+ve} disease have lower values than in HPV^{-ve} disease, resulting in differences in the optimal thresholds. This was similar to the findings of Tahari et al. (2014), who also found that SUV values in lesions with HPV^{+ve} were lower than in lesions with HPV^{-ve} disease.

Consequently, based on a minimal specificity of 80%, we have suggested optimal thresholds of 6.51 and 5.93 for primary tumour SUV_{max} in HPV^{+ve} and HPV^{-ve} lesions, respectively. Our proposed threshold of 6.51 for primary tumour $SUV_{max(w)}$ in HPV^{+ve} OPSCC was similarly reported in two prior studies. The first study included only patients with HPV-related squamous cell carcinoma (Vainshtein et al., 2014), and the second study comprised mostly patients with oropharyngeal cancer (Moeller et al., 2009).

Our findings imply that when investigating post-treatment primary tumour SUV_{max} originating from oropharyngeal carcinoma, it may be advisable to use a higher threshold value, such as a threshold of 6.51 in HPV^{+ve} disease. However, this finding needs to be externally validated prior to its clinical implications.

Lymph node metastases

Regarding the lymph node metrics, our study found that post-treatment nodal SUV_{max} normalised to total body weight yielded a high AUC value (0.94, 95% CI: 0.88-0.99) (Chapter 5, Table 5.3), at an optimal threshold of 3.39. Several previous studies found lymph node SUV_{max} with lower threshold values ranging from 2.0 to 2.8 (Moeller et al., 2009, Chan et al.,

2012, Vainshtein et al., 2014, Sjovall et al., 2016), while others suggested higher thresholds ranging from 3.75 to 4.62 (Sagardoy et al., 2016, Fatehi et al., 2019b, Dejaco et al., 2020).

Specifically, a nodal SUV_{max} normalised by total body weight threshold of 4.0 was found to be the best cutoff obtained from HNSCC nodal lesions (sensitivity: 85.70%; specificity: 100.00%) (Sagardoy et al., 2016). A higher nodal SUV_{max} threshold of 4.62 (sensitivity: 73.50%; specificity: 92.30%) was found to be the optimal cutoff in another study (Fatehi et al., 2019a).

Yet another study by Dejaco et al. (2020) of patients with advanced HNSCC who subsequently underwent PET/CT at a median of 10 weeks post-treatment found that post-treatment nodal SUVs normalised by total body weight demonstrated high discriminative power, with an AUC value of 0.90 (±0.05). The maximum Youden Index used indicated that a nodal $SUV_{max(w)}$ value of 3.75 was the optimal threshold between positive and negative nodal lesions (sensitivity: 69.00%; specificity: 100.00%) (Dejaco et al., 2020). The reported ROC curve value for nodal $SUV_{max(w)}$ in that study (AUC of 0.90) was slightly lower than our finding of 0.94. This could be attributed to the earlier timing of post-treatment imaging (10.0 vs. 13.4 weeks in our study). False-positive and false-negative rates are known to be greater when post-treatment imaging is performed prior to three months after completing CRT (Nelissen et al., 2017). For this reason, research indicates that the duration between therapy and post-treatment imaging should not be less than 12 weeks to minimise treatment effects (Baxi et al., 2015, Mehanna et al., 2016, Mayo et al., 2019). Furthermore, the authors of that study suggested an optimal $SUV_{max(w)}$ of 3.75, which was slightly greater than the threshold suggested in our study at 3.39. This might also be due to the timing of post-treatment imaging or the small sample size of

their analysis of 33 patients. This might also be due to the method used to select thresholds. They used the Youden Index, which is applied by giving both sensitivity and specificity equal weight (Ruopp et al., 2008), resulting in a higher cutoff value. On the other hand, we selected optimal thresholds by maximising the NPV, thus choosing a lower threshold value. Furthermore, most patients in their study had HPV^{+ve} HNSCC (54.5%), but in our cohort, this was the case in only 32.5%. In our sub-analysis on HPV^{+ve} and HPV^{-ve} disease (Chapter 5, Table 5.5), we found that nodal SUVs in HPV^{+ve} showed slightly lower values than in HPV^{-ve} disease (2.61 vs. 2.74). This was similar to the findings of Tahari et al. (2014), who found that SUV values in HPV^{+ve} and HPV^{-ve} nodal lesions were nearly identical. Based on a minimum specificity of 80%, we suggested optimal thresholds of 3.17 and 3.39 for nodal SUV_{max} in HPV^{+ve} and HPV^{-ve} lesions, respectively.

Additional factors that could potentially contribute to the variation in thresholds across studies include different uptake times, scan acquisition parameters (Boellaard, 2009), and image reconstruction techniques (Hirji et al., 2019).

Several methodologies for calculating SUV thresholds have been proposed in the literature (Chapter 2, Table 2.3). Some studies used the mean (Vainshtein et al., 2014) or median SUV values (Nelissen et al., 2017) for setting thresholds, while the majority of studies used the ROC curve (Sagardoy et al., 2016, Fatehi et al., 2019b, Dejaco et al., 2020, Helsen et al., 2020b). Using the mean or median SUV can be beneficial for providing an overview of SUV statistics. However, using means or medians to create threshold values could be misleading, as both have disadvantages. For example, the median values do not account for the individual

value of each SUV (Manikandan, 2011a), resulting in a loss of analysis efficacy, whereas the main problem with the mean is its sensitivity to outliers and extreme values, particularly when the sample size is small (Manikandan, 2011b). On the other hand, utilising the ROC curve for threshold selection appears favourable. However, it has a drawback in that threshold selection is dependent on sensitivity and specificity; therefore, there is no universal consensus on how these parameters should be adjusted.

Compared to other published research, our cutoffs, overall, were slightly lower. This may be because we opted for a threshold with improved NPV, whereas some other investigations focused only on sensitivity and specificity. We believe that selecting thresholds based on NPV might be a better indicator for assessing a threshold's ability. This is because choosing thresholds with a high NPV improves the rate of true negative findings; hence, it increases the likelihood of identifying patients with no actual disease who could be spared with more invasive diagnostic tests and treatment. As mentioned previously, sensitivity and specificity do not account for disease prevalence, while predictive values do (Wang et al., 2021). However, predictive values have the disadvantage of being heavily dependent on disease prevalence. When disease prevalence is high, the PPV is high but the NPV is low, and vice versa (Ranganathan and Aggarwal, 2018). Therefore, predictive values should be employed only when disease prevalence is within the average range. In our analysis, approximately 22.6% of the lesions were proven to have positive residual disease by histology and/or negative follow-up. This was comparable to the rate of residual disease reported by Sagardoy et al. (2015), which was 21.3%. Therefore, selecting thresholds based on NPV seemed appropriate in our study.

A further aspect of PET/CT quantification is the implication of using different body size metrics for SUV normalisation, such as LBM and BSA versus total body weight. Similar to the findings in the correlation study (Chapter 4), our study found that all post-treatment primary tumour and lymph node SUV_{max} normalised to weight, LBM, and BSA yielded generally high ROC values. Therefore, all of these measures may be equally capable of identifying residual primary or nodal lesions after the completion of CRT. We believe this is the first study to compare the effect of normalisation with the use of different body metrics on the diagnostic performance of post-treatment HNSCC PET/CT metrics. Consequently, direct comparisons of our results with other published studies were not performed. Potential reasons that might contribute to potentially falsely concluding a lack of a significant difference in different normalisation procedures to discriminate lesions were previously discussed in Section 7.1.1.

An approach that has been investigated repeatedly to reduce variation and thus improve the standardisation of different SUV metrics is by further normalising them to SUVs in unaffected regions, such as the liver and BP, resulting in an SUV ratio (Blautzik et al., 2019). In this study, we found conflicting results when SUV metrics were further normalised by the average FDG uptake in the liver. When further normalised by the liver SUV_{mean}, the diagnostic performance of some primary tumour absolute metrics improved, while the lymph node metrics did not (Chapter 5, Tables 5.3 and 5.5). The inconsistency found among our primary tumour and lymph node metrics could be related to lesion size and rate of metabolic activity. Primary tumours are often larger and therefore more hypermetabolic than smaller lymph node lesions. This can be demonstrated by the median values of these metrics. For example,

primary tumour $SUV_{max(w)}$ showed a median value of 3.85, while lymph node $SUV_{(w)}$ showed a median of 2.58 (Chapter 5, Table 5.2). Therefore, the effect of further normalising lesions on the average uptake in background organs could be more apparent in larger lesions with high FDG uptake, whereas in small lesions this could be indistinguishable. In addition, some lymph nodes often have a significant necrotic (photopenic) component (Baik et al., 2019), which may serve as a contributing factor possibly affecting SUV readings.

Similar to our primary tumour findings, in a study of rectal cancer patients who underwent neoadjuvant chemoradiotherapy (NCRT), it was found that SUV_{max} normalised to liver uptake was a better predictor of pathologic complete response rate in comparison to SUV_{max} and SUV_{max} normalised to BP uptake (Park et al., 2014). In HNSCC, a study analysed nodal lesion-to-background ratios and determined that liver normalisation provided the highest performance relative metric (Helsen et al., 2020a). However, that study evaluated the overall mean AUC values for several metrics (SUV_{70}, SUV_{50}, SUV_{90}, SUV_{max}, SUV_{mean}, and SUV_{peak}), whereas our study investigated metrics independently, which could have contributed to the difference in the results.

We also found no or even less improvement in diagnostic performance when the primary tumour and lymph node metrics were further normalised by the average uptake in the BP. This contradicts an early study by van den Hoff et al. (2013) that found that the tumour-to-blood SUV ratio was superior to tumour SUV in patients with colorectal cancer as a surrogate imaging metric of tumour FDG metabolism (van den Hoff et al., 2013). This difference might

be related to the presence of some variability when the BP measurements were measured, as concluded in Chapter 5, Section 5.3.6.2.2.

Furthermore, when we compared the overall diagnostic performance of the HNSCC absolute and relative metrics, we found that the diagnostic performance for detecting residual lesions appeared to improve when SUV metrics were further normalised by the average FDG uptake in the cerebellum. The improved results observed when normalising to the cerebellum's SUV_{mean} could be due to the cerebellum's proximity to the neck region as opposed to the liver or aortic arch (BP). This may also be because the cerebellum is more homogenous than the liver or BP.

The difference in diagnostic performance among the liver, cerebellum and BP relative metrics may be related to a variety of factors. For example, there is a difference in the levels of FDG uptake and clearance in these background regions (Chin et al., 2009). Liver disease and the potential effect of treatment on the liver such as drug-induced injury may potentially impact liver SUV values. Reduced renal function can increase BP activity, hence diminishing the accuracy of the tumour-to-blood SUV ratio (Sarikaya and Sarikaya, 2020).

Overall, in our study, when the primary tumours and lymph node HNSCC SUV metrics were further normalised by the average FDG uptake in the background regions, no clear evidence for diagnostic superiority was found, implying that normalising to the background regions has no additional benefit. Further studies with a larger patient population may be required to better characterise the diagnostic performance of the relative metrics in post-treatment

HNSCC and to clarify the effect of different biological and physiological factors on FDG uptake and clearance in these background regions.

Regarding the repeatability of SUV metrics, overall, apart from BP SUV_{mean} normalised to total body weight, good to excellent ICCs were found between two reading times for all PET/CT metrics, suggesting that the two readings obtained at different times were very similar and hence these metrics are reproducible. This also suggests that a consistent protocol for image acquisition and analysis was followed throughout this thesis when HNSCC and background regions' SUV metrics were analysed.

In summary, we found that all post-treatment SUV indices had comparable AUC values. In addition, using LBM and BSA in SUV normalisation and lesion-to-background ratios did not result in a statistically significant difference in the diagnostic performance of discriminating residual HNSCC disease compared to SUVs normalised to total body weight.

We then sought to investigate whether these indices have prognostic value for predicting survival outcomes and whether HPV status influences their predictive ability.

7.1.3 Prognostic value of post CRT PET/CT quantitative parameters

Few studies have investigated the role of PET/CT semiquantitative metrics as potential prognostic imaging predictors. The most studied imaging parameter was SUV_{max} (Ito et al., 2014, Kim et al., 2016a, Mayo et al., 2019, Connor et al., 2021), while none investigated post-treatment SUV_{peak}. The following sections discuss findings on the correlation of post-

treatment metrics with survival endpoints (Chapter 6). The discussion will be based on the initial analysis involving different HNSCC subtypes and the second analysis based on HPV status (HPV^{+ve} OPSCC and HPV^{-ve} HNSCC).

Progression-free survival (PFS)

In the initial results of the combined HNSCC cohort, we found that all primary tumour and involved lymph node HNSCC SUV_{max} and SUV_{peak} metrics were associated with three-year PFS (Chapter 6, Table 6.3). This finding was replicated when patients were classified based on HPV status (Chapter 6, Table 6.10).

There is a similarity between the association we found between SUV_{max} and three-year PFS to that described by Kim et al. (2016). They examined 78 patients with locally advanced HNSCC and found that SUV_{max} measured from the highest uptake from either primary or nodal combined HNSCC SUV_{max} threshold of 4.4 has prognostic value for predicting PFS. Patients with post-treatment SUV_{max} ≥4.4 had a shorter PFS compared to patients with SUV_{max} <4.4 (three-year PFS: 42.9 vs. 81.1%; P<0.001). However, they did not examine SUV_{peak}, and they analysed SUV_{max}, which was determined from the highest lesions' uptake, as opposed to examining primary tumour and lymph node SUVs separately. In general, therefore, it seems that post-treatment primary tumour and lymph node SUV_{max} are useful aids in predicting disease progression in all groups.

An implication of this is the possibility that an increase in post-treatment SUV_{max} values with thresholds established in Chapter 6, Table 6.7, may indicate an increased risk of disease

progression, including the risks of residual disease, locoregional or systematic recurrence, or even death within three years of CRT completion. Therefore, patients presenting with lesions with SUV_{max} greater than those thresholds should be investigated further. However, further research should be conducted in larger prospective studies to investigate the prognostic value of SUV_{peak} and validate the thresholds established in this work.

Overall survival (OS)

In the initial results of the combined HNSCC cohort, we found that only T $SUV_{peak(bsa)}$, N $SUV_{peak(w)}$, N $SUV_{peak(lbm)}$, and N $SUV_{peak(bsa)}$ were associated with overall survival (Chapter 6, Table 6.3). When analysing by HPV status, in patients with HPV^{+ve} OPSCC, most of the primary and nodal metrics exhibited an association with five-year OS, whereas in the group of patients with HPV^{-ve} HNSCC, only nodal SUV_{peak} normalised to LBM and BSA were associated with five-year OS (Chapter 6, Table 6.11).

Compared to the literature, our findings for SUV_{max} normalised to total body weight differ from those of some published studies (Ito et al., 2014, Kim et al., 2016b). However, we did not find other studies that investigated SUV_{max} normalised to LBM and BSA, as well as SUV_{peak}. In 36 patients with advanced HNSCC who received combined intra-arterial chemotherapy and radiotherapy (IACR), the authors found that patients with a primary tumour SUV_{max} value greater than 6.1 experienced a poorer OS in comparison to patients with a lower SUV_{max} (<6.1). The mean OS was estimated to be 12.1 months (95% CI, 6.3–18.0 months) and 44.6 months (95% CI, 39.9–49.3 months), respectively. This suggested that post-treatment primary tumour $SUV_{max(w)}$ had a significant prognostic value to predict survival (Ito et al., 2014). Kim

et al. (2016) found that post-CRT SUV_{max} measured from the highest uptake from either primary or nodal lesions has prognostic value for predicting OS. Patients with post-treatment SUV_{max} ≥4.4 had shorter survival compared to patients with SUV_{max} <4.4 (three-year OS: 56.9% vs. 87.7%; *P*=0.005). In contrast to our findings, we did not observe an association between both post-CRT primary tumour and lymph node $SUV_{max(w)}$ and OS (Chapter 6, Table 6.3). However, our study differed from theirs in some ways. For example, Ito et al. (2014) used the long-rank test to compare survival rates, while we used risk estimates (Cox proportional model). Therefore, comparing survival rates against the risk of an event outcome (death) is incomparable.

Also, both studies evaluated shorter endpoint durations (two- and three-year OS), while we looked at a five-year OS. We believe that SUV metrics might lose their prognostic significance over time, explaining the inconsistent findings in our analysis of five-year OS in comparison to the significant association reported in the other two investigations.

Furthermore, in the subgroup analysis of HPV^{+ve} OPSCC and HPV^{-ve} HNSCC, we found that patients with HPV^{-ve} HNSCC experienced a higher mortality rate compared to patients with HPV^{+ve} disease (Chapter 6, Table 6.9). Despite this, it was interesting to find more associations between post-treatment metrics in the group of patients with HPV^{+ve} disease compared to patients with HPV^{-ve} disease. This could be due to the small number of patients with HPV^{-ve} HNSCC resulting in type II error. It could also be related to the lesion's metabolism and heterogeneity. Based on our sub-analysis in Chapter 5, Table 5.5, primary tumour SUV_{max} was higher in lesions with HPV^{-ve} disease than in HPV^{+ve}, while nodal SUV_{max} was nearly the same in both groups. Also, according to Tahari and colleagues (2014), primary tumours of HPV^{-ve} disease are slightly more heterogeneous compared to those lesions with HPV^{+ve} disease, while

nodal lesions with HPV^{+ve} disease tend to be more heterogeneous than nodes with HPV^{-ve} disease. All of these reasons could account for our inconsistent findings, and the heterogeneity in nodal lesions could explain the presence and absence of associations between some of the metrics and 5-year OS. Another potential reason that might have contributed to the inconsistence findings was that fifteen different imaging metrics and four clinical factors (age, HPV status, T stage, and LRT) were included in the multivariable analysis. The inclusion of multiple factors could lead to incorrect findings.

Furthermore, the presence of a significant association between primary tumour $SUV_{peak(bsa)}$ and nodal SUV_{peak} normalised by weight, LBM and BSA in the combined HNSCC group (Chapter 6, Table 6.3) may be due to the fact that SUV_{peak} is less sensitive to image noise than SUV_{max} (Vanderhoek et al., 2012), thereby enhancing their prognostic value. However, not all SUV_{peak} metrics were superior predictors of five-year OS than SUV_{max} in all subgroups, including combined HNSCC, HPV^{+ve} OPSCC, and HPV^{-ve} HNSCC (Chapter 6, Tables 6.3 and 6.11). Again, these findings may be due to the heterogeneity of the lesions or a type II error.

Overall, due to the reasons discussed above, we were unable to draw clear conclusions about the analysis of 5-year OS, especially in patients with HPV^{+ve} OPSCC and HPV^{-ve} HNSCC. Therefore, evaluating the prognostic value of post-treatment metrics across different OS durations is necessary to determine whether these indices truly lose significance over time. Further studies with larger sample sizes need to be undertaken before the associations between post-treatment SUV metrics, including SUV_{max} and SUV_{peak} in patients with HPV positive and negative disease and OS, are more clearly understood.

While the focus of the analysis was on exploring the prognostic value of various imaging parameters, we also examined several clinical factors in the multivariable analysis. This was carried out to assess the impact these variables have on the metrics' prognostic value. Extensive research has been conducted on prognostic clinical factors for outcome prediction in head and neck cancer. For example, advanced tumour stage has been shown to be a prognostic factor in several studies (Kim et al., 2015, Cadoni et al., 2017). However, in this study, T stage was not a prognostic factor for PFS and OS (Chapter 6, Table 6.3). The difference in results could be attributed to the highly selective nature of this cohort with predominant OPSCC patients, the majority of whom had HPV^{+ve} disease (55%) and 12.5% had unknown HPV status. According to Budach and Tinhofer (2019), the use of the TNM staging system's seventh edition has been shown to be a poor predictor of five-year OS in cases of HPV^{+ve} OPCC (Budach and Tinhofer, 2019). This is because TNM could lose its discriminatory strength between disease stages—that is, when employing TNM 7 for HPV^{+ve} disease, early and late-stage disease survival curves end up being comparable. A reason for this could be that the primary tumour disease in HPV^{+ve} disease is frequently reported with a lower T stage.

Another imperative prognostic factor in head and neck cancer is carcinogenic HPV infection (Lechner et al., 2022). Studies indicate that patients with HPV^{+ve} disease have better treatment outcomes and prognoses than those with HPV^{-ve} disease. For example, a study on oropharyngeal cancer showed that eight-year OS differed in patients with positive and negative p16 statuses. Patients with p16 negative tumours had inferior survival compared to patients with p16 positive tumours (eight-year survival, 30.2% vs. 70.9%) (Nguyen-Tan et al., 2014). Similarly, a prospective clinical trial found that HPV status was significantly associated

with both PFS and OS (Fakhry et al., 2008). In line with these findings, we found that patients with HPV^{+ve} OPSCC experienced favourable treatment outcomes and survival compared to patients with HPV^{-ve} HNSCC (Chapter 6, Table 6.9). Despite this, in our study, we could not find a major difference in the association between post-treatment PET/CT metrics and PFS when patients were divided according to HPV status, except for the slight difference observed in the analysis of OS (Chapter 6, Table 6.11). This might be due to the small sample size in the subgroup of patients with HPV^{-ve} HNSCC (type II error) or the low incidence of events such as disease recurrence encountered in subgroups of patients with HPV^{-ve} cancers.

Similar to the findings in the correlation and diagnostic performance studies (Chapters 4 and 5), we found that there was no major difference in the prognostic value when SUV metrics were corrected for total body weight, LBM, or BSA. Further discussion is provided in Section 7.1.1.

In summary, it was found that post-treatment primary tumour and nodal SUV_{max} and SUV_{peak} normalised to total body weight in all subgroups appear to be significant prognostic factors for predicting three-year PFS. An increase in the value of these imaging indices could be indicative of disease progression within the first three years post-CRT. Therefore, when SUV readings are significantly higher than the suggested thresholds, more monitoring is needed to detect recurrence sooner and to potentially improve survival rates. The study also found inconclusive results regarding the analysis of 5-year OS, as some of the parameters showed an association with 5-year OS while others did not across the different subgroups (combined

HNSCC, HPV^{+} OPSCC, and HPV^{-}HNSCC). Our findings should be validated in prospective, larger cohort studies.

7.2 Limitations

This study has several limitations:

1) The overall study cohort consisted of 124 patients, and a limited sample size may bias conclusions by being underpowered statistically, such as in the analysis of HPV^{+ve} and HPV^{-ve} disease. Nonetheless, the sample size was comparable to some previously published studies and larger than many.
2) Due to the retrospective design of the study, it was challenging to avoid missing data, such as data on HPV status for all included OPSCC patients and data on other potential prognostic factors, such as tumour size, lymph node involvement and extracapsular spread (Budach and Tinhofer, 2019).
3) We analysed three-year PFS rather than five-year PFS. This was done because we did not have five-year post-treatment recurrence data for all patients, but we did have five-year post-treatment data on overall survival. Therefore, three-year PFS and five-year OS analyses were conducted.
4) The data covered a cohort of patients over three years. Two PET/CT scanners were used during that period. Despite the use of two different scanners for PET/CT acquisition and the inability to perform SUV harmonisation, both scanners underwent daily quality assurance/regular service maintenance to ensure proper operation. Also, the effect of non-harmonisation is likely to be minimal, as only eight patients were

scanned using a different scanner. A sample sensitivity analysis confirmed that the impact of using two scanners on SUV readings was minimal (Appendix 6).

5) Another limitation is that the LBM in the present study was calculated based on an equation that incorporates height and weight. While this method is acceptable, other suggested LBM calculation methods might yield different estimations (Tahari et al., 2014) or the use of direct determination by CT (Aide et al., 2017).
6) Several studies have used different methods to classify equivocal post-treatment PET/CT scans, such as considering PET/CT scans with a near-complete response (CR) as CR or as <CR (Vainshtein et al., 2014). Other methods used involved using a nodal SUV_{max} cutoff of 2 (Sjovall et al., 2016, Nelissen et al., 2017) or visually comparing lesion FDG intensity to uptake in the surrounding areas (Sagardoy et al., 2016). In our study, we considered equivocal PET/CT scans with an uptake higher than the surrounding regions to be positive. This yielded a high number of PET/CT scans with incomplete treatment responses in both groups (T+ and N+). Even though using a different approach could affect the number of incomplete and complete PET/CT scans, it would not have an impact on the SUV results and interpretation, as we used thresholds to determine false-positive and -negative findings that were later applied to calculate predictive values.
7) To test the repeatability of SUV measurements, a small sample size was used. However, even though the acquisition process of SUV metrics was not fully automated, the image analysis software (Hermes) allowed for a semi-automated process, reducing the chance of errors introduced during image analysis. Therefore, our sample size was considered adequate for the purposes of this study.

8) Some studies have shown that volumetric parameters and changes in pre-treatment and post-treatment metrics could be better predictors of survival (Matoba et al., 2017, Rijo-Cedeno et al., 2020). However, due to the inavailability of pre-treatment scans and the low uptake of FDG in some of the target lesions in the post-treatment scans, which made tumour segmentation difficult, we were unable to evaluate the significance of post-treatment PET/CT volumetric parameters, such as MTV and TLG, as well as the percentage change in SUV metrics. Other methods that were unavailable to us showed promising results, including the use of advanced radiomics in lesion segmentation (Beichel et al., 2019, Bruixola et al., 2021).

7.3 Conclusion and clinical implications

This thesis primarily aimed to investigate the correlation between post-treatment HNSCC metrics and several patient body size factors to identify a normalisation metric that is less sensitive to patients' body size measurements (Chapter 4). Next, it aimed to investigate the diagnostic performance (Chapter 5) and prognostic value (Chapter 6) of SUV metrics acquired separately from primary and nodal sites, as well as the effect of HPV status on their prognostic ability.

Based on the findings of this thesis, we have reached the following conclusions with relevance for clinical implications:

1) Post-treatment metrics normalised to total body weight, LBM, and BSA derived from HNSCC lesions did not significantly correlate with different body size measurements.

This implies that the use of any method in SUV normalisation is suitable for analysing post-treatment scans of HNSCC. This includes evaluating quantitative metrics for the purposes of discriminative ability, as well as for prognostication.

2) In combined HNSCC from different subtypes, a three months post-treatment primary tumour lesion with an $SUV_{max(w)}$ threshold ≥5.59 is more likely to harbour residual disease at the primary sites. A three months post-treatment lymph node lesion with an $SUV_{max(w)}$ threshold ≥3.39 is more likely to harbour residual disease at the nodal sites.
3) In HPV^{+ve} OPSCC, three months post-treatment primary tumour and involved lymph node lesions with $SUV_{max(w)}$ thresholds of ≥6.51 and ≥3.17 are more likely to harbour residual disease at the primary and nodal sites, respectively.
4) In HPV^{-ve} HNSCC, three months post-treatment primary tumour and involved lymph node lesions with $SUV_{max(w)}$ thresholds of ≥5.93 and ≥3.39 are more likely to harbour residual disease at the primary and nodal sites, respectively.
5) The use of lesion-to-background ratios yielded inconsistent findings. Hence, normalisation of the background regions should be avoided until strong evidence of their diagnostic superiority is established.
6) Post-treatment primary tumour and nodal SUV_{max} and SUV_{peak} normalised by weight in all subgroups appeared to be significant prognostic predictors for three-year PFS. Therefore, an increase in post-treatment SUV_{max} and SUV_{peak} values higher than the thresholds established in Chapter 6, Table 6.7, may indicate an increased risk of disease progression, including the risks of residual disease, locoregional or systematic

recurrence, or even death within three years of CRT completion. Patients with lesions SUV_{max} and SUV_{peak} greater than those thresholds should be closely followed.

7.4 Future work

1) External validation study

Our study was the first to evaluate the correlation between quantitative metrics and body size factors and was one of the few studies that assessed the diagnostic and prognostic performance of post-treatment SUV_{max} acquired three months after the completion of the treatment. It was also the first study to evaluate SUV_{peak} in a post-treatment setting and the second study to evaluate lesions-to-background relative metrics. Therefore, further prospective research with a larger sample size is required to confirm the true effect of body size measures on SUV readings in normal-weight and obese patients. Further studies with larger sample sizes are also needed to better characterise the diagnostic performance of both absolute and relative metrics and to determine their prognostic significance. There is also a need for additional research focusing on particular subtypes of head and neck cancer and clarifying the role of HPV status in SUV readings. Importantly, our established thresholds should be externally validated in a larger prospective study.

2) Advanced radiomics and machine learning methods for image analysis

Tumour heterogeneity in terms of molecular and biological heterogeneity is a characteristic of HNSCC. As previously described, SUV metrics could be highly affected by tumour

heterogeneity. Radiomics, on the other hand, applies advanced computational analysis of medical images, which can assist in overcoming some of the obstacles. Machine learning is an advanced technology of artificial intelligence (AI) that uses experience (data) to improve the performance of measurements or predictors (Bruixola et al., 2021). The heterogeneity of image voxel intensities can be quantified by different image processing and analysis methods, including texture analysis (TA). TA is a procedure by which a significant number of features are extracted from images to quantify the characteristics of organs and tissues beyond the capability of visual interpretation or simple quantitative metrics (Hatt et al., 2017). In head and neck cancer imaging, AI could improve a variety of clinical tasks. These tasks may include assisting with tumour segmentation, characterization, prognostication, treatment response assessment, and prediction of metastatic lymph node disease, in addition to improving image quality by reducing motion and noise artefacts (Pham et al., 2022).These techniques could have the potential to provide a more accurate assessment of imaging metrics. Hence, future studies are suggested to explore the predictive and prognostic role of imaging metrics in HNSCC, analysed using advanced radiomics and machine learning techniques.

List of references

Risk of bias assessment tool (QUIPS).

ABD EL-HAFEZ, Y. G., MOUSTAFA, H. M., KHALIL, H. F., LIAO, C. T. & YEN, T. C. 2013. Total lesion glycolysis: a possible new prognostic parameter in oral cavity squamous cell carcinoma. *Oral Oncol,* 49**,** 261-8.

ABOUZIED, M. M., CRAWFORD, E. S. & NABI, H. A. 2005. 18F-FDG imaging: pitfalls and artifacts. *J Nucl Med Technol,* 33**,** 145-55; quiz 162-3.

ADAMS, M. C., TURKINGTON, T. G., WILSON, J. M. & WONG, T. Z. 2010. A systematic review of the factors affecting accuracy of SUV measurements. *AJR Am J Roentgenol,* 195**,** 310-20.

AIDE, N., LASNON, C., VEIT-HAIBACH, P., SERA, T., SATTLER, B. & BOELLAARD, R. 2017. EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. *Eur J Nucl Med Mol Imaging,* 44**,** 17-31.

AKAGUNDUZ, O. O., SAVAS, R., YALMAN, D., KOCACELEBI, K. & ESASSOLAK, M. 2015. Can adaptive threshold-based metabolic tumor volume (MTV) and lean body mass corrected standard uptake value (SUL) predict prognosis in head and neck cancer patients treated with definitive radiotherapy/chemoradiotherapy? *Nuclear Medicine & Biology,* 42**,** 899-904.

AKAMATSU, G., TSUTSUI, Y., DAISAKI, H., MITSUMOTO, K., BABA, S. & SASAKI, M. 2023. A review of harmonization strategies for quantitative PET. *Ann Nucl Med,* 37**,** 71-88.

ARAKAWA, I., ABOU-AYASH, S., GENTON, L., TSUGA, K., LELES, C. R. & SCHIMMEL, M. 2020. Reliability and comparability of methods for assessing oral function: Chewing, tongue pressure and lip force. *J Oral Rehabil,* 47**,** 862-871.

ASLAN, H., CEKIN, G., PINAR, E., YAZIR, M., IMRE, A., SONGU, M., ISLEK, A., ALADAG, I. & ONCEL, I. S. 2019. Prognostic value of 18F-FDG PET/CT parameters and histopathologic variables in head and neck cancer. *Brazilian Journal of Otorhinolaryngology*.

AWAN, M. J., LAVERTU, P., ZENDER, C., REZAEE, R., FOWLER, N., KARAPETYAN, L., GIBSON, M., WASMAN, J., FAULHABER, P., MACHTAY, M. & YAO, M. 2017. Post-treatment PET/CT and p16 status for predicting treatment outcomes in locally advanced head and neck cancer after definitive radiation. *Eur J Nucl Med Mol Imaging,* 44**,** 988-997.

BAIK, S. H., SEO, J. W., KIM, J. H., LEE, S. K., CHOI, E. C. & KIM, J. 2019. Prognostic Value of Cervical Nodal Necrosis Observed in Preoperative CT and MRI of Patients With Tongue Squamous Cell Carcinoma and Cervical Node Metastases: A Retrospective Study. *AJR Am J Roentgenol,* 213**,** 437-443.

BAXI, S. S., DUNN, L. & PFISTER, D. G. 2015. Evaluating the potential role of PET/CT in the posttreatment surveillance of head and neck cancer. *J Natl Compr Canc Netw,* 13**,** 252-4.

BEICHEL, R. R., ULRICH, E. J., SMITH, B. J., BAUER, C., BROWN, B., CASAVANT, T., SUNDERLAND, J. J., GRAHAM, M. M. & BUATTI, J. M. 2019. FDG PET based prediction of response in head and neck cancer treatment: Assessment of new quantitative imaging features. *PLoS ONE,* 14.

BLAND, J. M. & ALTMAN, D. G. 1999. Measuring agreement in method comparison studies. *Stat Methods Med Res,* 8**,** 135-60.

BLAUTZIK, J., GRELICH, L., SCHRAMM, N., HENKEL, R., BARTENSTEIN, P. & PFLUGER, T. 2019. What and how should we measure in paediatric oncology FDG-PET/CT? Comparison of commonly used SUV metrics for differentiation between paediatric tumours. *EJNMMI Res,* 9**,** 115.

BOELLAARD, R. 2009. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med,* 50 Suppl 1**,** 11S-20S.

BOELLAARD, R., DELGADO-BOLTON, R., OYEN, W. J., GIAMMARILE, F., TATSCH, K., ESCHNER, W., VERZIJLBERGEN, F. J., BARRINGTON, S. F., PIKE, L. C., WEBER, W. A., STROOBANTS, S., DELBEKE, D., DONOHOE, K. J., HOLBROOK, S., GRAHAM, M. M., TESTANERA, G., HOEKSTRA, O. S., ZIJLSTRA, J., VISSER, E., HOEKSTRA, C. J., PRUIM, J., WILLEMSEN, A., ARENDS, B., KOTZERKE, J., BOCKISCH, A., BEYER, T., CHITI, A., KRAUSE, B. J. & EUROPEAN ASSOCIATION OF NUCLEAR, M. 2015. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging,* 42**,** 328-54.

BONOMO, P., MERLOTTI, A., OLMETTO, E., BIANCHI, A., DESIDERI, I., BACIGALUPO, A., FRANCO, P., FRANZESE, C., ORLANDI, E., LIVI, L. & CAINI, S. 2018. What is the prognostic impact of FDG PET in locally advanced head and neck squamous cell carcinoma treated with concomitant chemo-radiotherapy? A systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging,* 45**,** 2122-2138.

BRUIXOLA, G., REMACHA, E., JIMENEZ-PASTOR, A., DUALDE, D., VIALA, A., MONTON, J. V., IBARROLA-VILLAVA, M., ALBERICH-BAYARRI, A. & CERVANTES, A. 2021. Radiomics and radiogenomics in head and neck squamous cell carcinoma: Potential contribution to patient management and challenges. *Cancer Treat Rev,* 99**,** 102263.

BUDACH, V. & TINHOFER, I. 2019. Novel prognostic clinical factors and biomarkers for outcome prediction in head and neck cancer: a systematic review. *Lancet Oncol,* 20**,** e313-e326.

CACICEDO, J., FERNANDEZ, I., DEL HOYO, O., NAVARRO, A., GOMEZ-ITURRIAGA, A., PIJOAN, J. I., MARTINEZ-INDART, L., ESCUDERO, J., GOMEZ-SUAREZ, J., DE ZARATE, R. O., PEREZ, J. F., BILBAO, P. & RADES, D. 2017a. Prognostic value of maximum standardized uptake value measured by pretreatment 18F-FDG PET/CT in locally advanced head and neck squamous cell carcinoma. *Clinical & Translational Oncology: Official Publication of the Federation of Spanish Oncology Societes & of the National Cancer Institute of Mexico,* 19**,** 1337-1349.

CACICEDO, J., FERNANDEZ, I., DEL HOYO, O., NAVARRO, A., GOMEZ-ITURRIAGA, A., PIJOAN, J. I., MARTINEZ-INDART, L., ESCUDERO, J., GOMEZ-SUAREZ, J., DE ZARATE, R. O., PEREZ, J. F., BILBAO, P. & RADES, D. 2017b. Prognostic value of maximum standardized uptake value measured by pretreatment 18F-FDG PET/CT in locally advanced head and neck squamous cell carcinoma. *Clin Transl Oncol,* 19**,** 1337-1349.

CACICEDO, J., NAVARRO, A., DEL HOYO, O., GOMEZ-ITURRIAGA, A., ALONGI, F., MEDINA, J. A., ELICIN, O., SKANJETI, A., GIAMMARILE, F., BILBAO, P., CASQUERO, F., DE BARI, B. & DAL PRA, A. 2016. Role of fluorine-18 fluorodeoxyglucose PET/CT in head and neck oncology: the point of view of the radiation oncologist. *Br J Radiol,* 89**,** 20160217.

CADONI, G., GIRALDI, L., PETRELLI, L., PANDOLFINI, M., GIULIANI, M., PALUDETTI, G., PASTORINO, R., LEONCINI, E., ARZANI, D., ALMADORI, G. & BOCCIA, S. 2017. Prognostic factors in head and neck cancer: a 10-year retrospective analysis in a single-institution in Italy. *Acta Otorhinolaryngologica Italica,* 37**,** 458-466.

CANNING, M., GUO, G., YU, M., MYINT, C., GROVES, M. W., BYRD, J. K. & CUI, Y. 2019. Heterogeneity of the Head and Neck Squamous Cell Carcinoma Immune Landscape and Its Impact on Immunotherapy. *Front Cell Dev Biol,* 7**,** 52.

CASTELLI, J., DE BARI, B., DEPEURSINGE, A., SIMON, A., DEVILLERS, A., ROMAN JIMENEZ, G., PRIOR, J., OZSAHIN, M., DE CREVOISIER, R. & BOURHIS, J. 2016. Overview of the predictive value of quantitative 18 FDG PET in head and neck cancer treated with chemoradiotherapy. *Crit Rev Oncol Hematol,* 108**,** 40-51.

CASTELLI, J., SIMON, A., LAFOND, C., PERICHON, N., RIGAUD, B., CHAJON, E., DE BARI, B., OZSAHIN, M., BOURHIS, J. & DE CREVOISIER, R. 2018. Adaptive radiotherapy for head and neck cancer. *Acta Oncol,* 57**,** 1284-1292.

CHAN, J. Y., SANGUINETI, G., RICHMON, J. D., MARUR, S., GOURIN, C. G., KOCH, W., CHUNG, C. H., QUON, H., BISHOP, J. A., AYGUN, N. & AGRAWAL, N. 2012. Retrospective review of positron emission tomography with contrast-enhanced computed tomography in the posttreatment setting in human papillomavirus-associated oropharyngeal carcinoma. *Archives of Otolaryngology -- Head & Neck Surgery,* 138**,** 1040-6.

CHEUNG, M. K., ONG, S. Y., GOYAL, U., WERTHEIM, B. C., HSU, C. C. & YI, S. K. 2017. False Positive Positron Emission Tomography / Computed Tomography Scans in Treated Head and Neck Cancers. *Cureus,* 9**,** e1146.

CHIN, B. B., GREEN, E. D., TURKINGTON, T. G., HAWK, T. C. & COLEMAN, R. E. 2009. Increasing uptake time in FDG-PET: standardized uptake values in normal tissues at 1 versus 3 h. *Mol Imaging Biol,* 11**,** 118-22.

CHOW, L. Q. M. 2020. Head and Neck Cancer. *N Engl J Med,* 382**,** 60-72.

CHRISTEN, T., SHEIKINE, Y., ROCHA, V. Z., HURWITZ, S., GOLDFINE, A. B., DI CARLI, M. & LIBBY, P. 2010. Increased glucose uptake in visceral versus subcutaneous adipose tissue revealed by PET imaging. *JACC Cardiovasc Imaging,* 3**,** 843-51.

COHEN, J. F., KOREVAAR, D. A., ALTMAN, D. G., BRUNS, D. E., GATSONIS, C. A., HOOFT, L., IRWIG, L., LEVINE, D., REITSMA, J. B., DE VET, H. C. & BOSSUYT, P. M. 2016. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open,* 6**,** e012799.

CONNOR, S., SIT, C., ANJARI, M., LEI, M., GUERRERO-URBANO, T., SZYSZKO, T., COOK, G., BASSETT, P. & GOH, V. 2021. The ability of post-chemoradiotherapy DWI ADCmean and (18)F-FDG SUVmax to predict treatment outcomes in head and neck cancer: impact of human papilloma virus oropharyngeal cancer status. *J Cancer Res Clin Oncol,* 147**,** 2323-2336.

CREFF, G., JEGOUX, F., PALARD, X., DEPEURSINGE, A., ABGRAL, R., MARIANOWSKI, R., LECLERE, J. C., EUGENE, T., MALARD, O., DE CREVOISIER, R., DEVILLERS, A. & CASTELLI, J. 2021. FDG-PET/CT-based prognostic survival model after surgery for head and neck cancer. *Oral Oncology,* 118.

DAISAKI, H., KITAJIMA, K., NAKAJO, M., WATABE, T., ITO, K., SAKAMOTO, F., NAKAHARA, T., ISHIBASHI, M. & TORIIHARA, A. 2021. Usefulness of semi-automatic harmonization strategy of standardized uptake values for multicenter PET studies. *Sci Rep,* 11**,** 8517.

DE BREE, R., VAN DER PUTTEN, L., BROUWER, J., CASTELIJNS, J. A., HOEKSTRA, O. S. & LEEMANS, C. R. 2009. Detection of locoregional recurrent head and neck cancer after (chemo)radiotherapy using modern imaging. *Oral Oncol,* 45**,** 386-93.

DEJACO, D., UPRIMNY, C., WIDMANN, G., RIEDL, D., MOSER, P., ARNOLD, C., STEINBICHLER, T. B., KOFLER, B., SCHARTINGER, V. H., VIRGOLINI, I. & RIECHELMANN, H. 2020. Response evaluation of cervical lymph nodes after chemoradiation in patients with head and neck cancer - does additional [18F]FDG-PET-CT help? *Cancer Imaging,* 20**,** 69.

DESCHLER, D. G., MOORE, M. G. & SMITH, R. V. 2014. Quick Reference Guide to TNM Staging of Head and Neck Cancer and Neck Dissection Classification. *In:* FOUNDATION, T. A. A. O. O. H. A. N. S. (ed.) 4th ed.: The American Academy of Otolaryngology— Head and Neck Surgery Foundation.

DIRIX, P., VANDECAVEYE, V., DE KEYZER, F., STROOBANTS, S., HERMANS, R. & NUYTS, S. 2009. Dose painting in radiotherapy for head and neck squamous cell carcinoma: value of repeated functional imaging with (18)F-FDG PET, (18)F-fluoromisonidazole PET, diffusion-weighted MRI, and dynamic contrast-enhanced MRI. *J Nucl Med,* 50**,** 1020-7.

ECONOMOPOULOU, P., DE BREE, R., KOTSANTIS, I. & PSYRRI, A. 2019. Diagnostic Tumor Markers in Head and Neck Squamous Cell Carcinoma (HNSCC) in the Clinical Setting. *Front Oncol,* 9**,** 827.

FAKHRY, C., WESTRA, W. H., LI, S., CMELAK, A., RIDGE, J. A., PINTO, H., FORASTIERE, A. & GILLISON, M. L. 2008. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst,* 100**,** 261-9.

FARWELL, M. D., PRYMA, D. A. & MANKOFF, D. A. 2014. PET/CT imaging in cancer: current applications and future directions. *Cancer,* 120**,** 3433-45.

FATEHI, K. S., THIAGARAJAN, S., DHAR, H., PURANDARE, N., A.K, D. C., CHAUKAR, D., LASKAR, S. G., PRABHASH, K. & RANGARAJAN, V. 2019a. Utility of response assessment PET-CT to predict residual disease in neck nodes: A comparison with the Histopathology. *Auris Nasus Larynx,* 46**,** 599-604.

FATEHI, K. S., THIAGARAJAN, S., DHAR, H., PURANDARE, N., AK, D. C., CHAUKAR, D., LASKAR, S. G., PRABHASH, K. & RANGARAJAN, V. 2019b. Utility of response assessment PET-CT to predict residual disease in neck nodes: A comparison with the Histopathology. *Auris Nasus Larynx,* 46**,** 599-604.

GUPTA, T., MASTER, Z., KANNAN, S., AGARWAL, J. P., GHSOH-LASKAR, S., RANGARAJAN, V., MURTHY, V. & BUDRUKKAR, A. 2011. Diagnostic performance of post-treatment FDG PET or FDG PET/CT imaging in head and neck cancer: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging,* 38**,** 2083-95.

HABIBZADEH, F., HABIBZADEH, P. & YADOLLAHIE, M. 2016. On determining the most appropriate test cut-off value: the case of tests with continuous results. *Biochem Med (Zagreb),* 26**,** 297-307.

HALLETT, W. A., MARSDEN, P. K., CRONIN, B. F. & O'DOHERTY, M. J. 2001. Effect of corrections for blood glucose and body size on [18F]FDG PET standardised uptake values in lung cancer. *Eur J Nucl Med,* 28**,** 919-22.

HATT, M., TIXIER, F., PIERCE, L., KINAHAN, P. E., LE REST, C. C. & VISVIKIS, D. 2017. Characterization of PET/CT images using texture analysis: the past, the present... any future? *Eur J Nucl Med Mol Imaging,* 44**,** 151-165.

HELSEN, N., VAN DEN WYNGAERT, T., CARP, L., DE BREE, R., VANDERVEKEN, O. M., DE GEETER, F., CAMBIER, J. P., SPAEPEN, K., MARTENS, M., HAKIM, S., BEELS, L.,

HOEKSTRA, O. S., VAN DEN WEYNGAERT, D., STROOBANTS, S., VAN LAER, C., SPECENIER, P., MAES, A., DEBRUYNE, P., HUTSEBAUT, I., VAN DINTER, J., HOMANS, F., GOETHALS, L., LENSSEN, O. & DEBEN, K. 2020a. Quantification of 18F-fluorodeoxyglucose uptake to detect residual nodal disease in locally advanced head and neck squamous cell carcinoma after chemoradiotherapy: results from the ECLYPS study. *European Journal of Nuclear Medicine and Molecular Imaging,* 47**,** 1075-1082.

HELSEN, N., VAN DEN WYNGAERT, T., CARP, L., DE BREE, R., VANDERVEKEN, O. M., DE GEETER, F., MAES, A., CAMBIER, J. P., SPAEPEN, K., MARTENS, M., HAKIM, S., BEELS, L., HOEKSTRA, O. S., VAN DEN WEYNGAERT, D., STROOBANTS, S., CONSORTIUM, E., VAN LAER, C., SPECENIER, P., MAES, A., DEBRUYNE, P., HUTSEBAUT, I., VAN DINTER, J., HOMANS, F., GOETHALS, L., LENSSEN, O. & DEBEN, K. 2020b. Quantification of 18F-fluorodeoxyglucose uptake to detect residual nodal disease in locally advanced head and neck squamous cell carcinoma after chemoradiotherapy: results from the ECLYPS study. *Eur J Nucl Med Mol Imaging,* 47**,** 1075-1082.

HIRATA, K. & TAMAKI, N. 2021. Quantitative FDG PET Assessment for Oncology Therapy. *Cancers (Basel),* 13.

HIRJI, H., SULLIVAN, K., LASKER, I., SHARIF, M. S., NUNES, A., SHEPHERD, C., WONG, W. L. & SANGHERA, B. 2019. Effect of PET Image Reconstruction Techniques on Unexpected Aorta Uptake. *Mol Imaging Radionucl Ther,* 28**,** 1-7.

HOGG, P. & TESTANERA, G. Principles and Practice of PET/CT. *Part 1 A Technologist's Guide.* EANM.

HSIEH, R. W., BORSON, S., TSAGIANNI, A. & ZANDBERG, D. P. 2021. Immunotherapy in Recurrent/Metastatic Squamous Cell Carcinoma of the Head and Neck. *Front Oncol,* 11**,** 705614.

HUANG, Y., FENG, M., HE, Q., YIN, J., XU, P., JIANG, Q. & LANG, J. 2017. Prognostic value of pretreatment 18F-FDG PET-CT for nasopharyngeal carcinoma patients. *Medicine,* 96**,** e6721.

IM, H. J., BRADSHAW, T., SOLAIYAPPAN, M. & CHO, S. Y. 2018. Current Methods to Define Metabolic Tumor Volume in Positron Emission Tomography: Which One is Better? *Nucl Med Mol Imaging,* 52**,** 5-15.

ITO, K., SHIMOJI, K., MIYATA, Y., KAMIYA, K., MINAMIMOTO, R., KUBOTA, K., OKASAKI, M., MOROOKA, M. & YOKOYAMA, J. 2014. Prognostic value of post-treatment (18)F-FDG PET/CT for advanced head and neck cancer after combined intra-arterial chemotherapy and radiotherapy. *Chin J Cancer Res,* 26**,** 30-7.

JOHNSON, D. E., BURTNESS, B., LEEMANS, C. R., LUI, V. W. Y., BAUMAN, J. E. & GRANDIS, J. R. 2020. Head and neck squamous cell carcinoma. *Nat Rev Dis Primers,* 6**,** 92.

KAMARUDIN, A. N., COX, T. & KOLAMUNNAGE-DONA, R. 2017. Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Med Res Methodol,* 17**,** 53.

KAPOOR, V., MCCOOK, B. M. & TOROK, F. S. 2004. An introduction to PET-CT imaging. *Radiographics,* 24**,** 523-43.

KARAM, M., DOROUDINIA, A., GOODARZI, S., KAGHAZCHI, F., KOMA, A., MEHRIAN, P. & ALIZADEH, N. 2018. Prognostic value of 18F-fluorodeoxyglucose-positron emission tomography/computed tomography scan volumetric parameters in head-and-neck

cancer patients after treatment. *Biomedical and Biotechnology Research Journal (BBRJ),* 2**,** 196-202.
KATAHIRA-SUZUKI, R., HATA, M., TATEISHI, U., TAGUCHI, T., TAKANO, S., OMURA-MINAMISAWA, M. & INOUE, T. 2015. Definitive chemo-radiotherapy for squamous cell carcinoma of the pharynx: impact of baseline low hemoglobin level (<12 g/dL) and post-radiation therapy F-18 FDG-PET/CT. *Ann Nucl Med,* 29**,** 37-45.
KERAMIDA, G. & PETERS, A. M. 2019. The appropriate whole body metric for calculating standardised uptake value and the influence of sex. *Nucl Med Commun,* 40**,** 3-7.
KIM, C. K., GUPTA, N. C., CHANDRAMOULI, B. & ALAVI, A. 1994. Standardized uptake values of FDG: body surface area correction is preferable to body weight correction. *J Nucl Med,* 35**,** 164-7.
KIM, J. W., OH, J. S., ROH, J. L., KIM, J. S., CHOI, S. H., NAM, S. Y. & KIM, S. Y. 2015. Prognostic significance of standardized uptake value and metabolic tumour volume on 18F-FDG PET/CT in oropharyngeal squamous cell carcinoma. *European Journal of Nuclear Medicine & Molecular Imaging,* 42**,** 1353-61.
KIM, R., OCK, C. Y., KEAM, B., KIM, T. M., KIM, J. H., PAENG, J. C., KWON, S. K., HAH, J. H., KWON, T. K., KIM, D. W., WU, H. G., SUNG, M. W. & HEO, D. S. 2016a. Predictive and prognostic value of PET/CT imaging post-chemoradiotherapy and clinical decision-making consequences in locally advanced head & neck squamous cell carcinoma: a retrospective study. *BMC Cancer,* 16**,** 116.
KIM, R., OCK, C. Y., KEAM, B., KIM, T. M., KIM, J. H., PAENG, J. C., KWON, S. K., HAH, J. H., KWON, T. K., KIM, D. W., WU, H. G., SUNG, M. W. & HEO, D. S. 2016b. Predictive and prognostic value of PET/CT imaging post-chemoradiotherapy and clinical decision-making consequences in locally advanced head & neck squamous cell carcinoma: A retrospective study. *BMC Cancer,* 16, 116.
KIM, S. Y., BEER, M. & TSHERING VOGEL, D. W. 2021. Imaging in head and neck cancers: Update for non-radiologist. *Oral Oncol,* 120**,** 105434.
KOO, T. K. & LI, M. Y. 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine,* 15**,** 155-163.
LECHNER, M., LIU, J., MASTERSON, L. & FENTON, T. R. 2022. HPV-associated oropharyngeal cancer: epidemiology, molecular biology and clinical management. *Nat Rev Clin Oncol,* 19**,** 306-327.
LEE, S. W., NAM, S. Y., IM, K. C., KIM, J. S., CHOI, E. K., AHN, S. D., PARK, S. H., KIM, S. Y., LEE, B. J. & KIM, J. H. 2008. Prediction of prognosis using standardized uptake value of 2-[(18)F] fluoro-2-deoxy-d-glucose positron emission tomography for nasopharyngeal carcinomas. *Radiother Oncol,* 87**,** 211-6.
LELL, M., BAUM, U., GREESS, H., NOMAYR, A., NKENKE, E., KOESTER, M., LENZ, M. & BAUTZ, W. 2000. Head and neck tumors: imaging recurrent tumor and post-therapeutic changes with CT and MRI. *Eur J Radiol,* 33**,** 239-47.
LIBERATI, A., ALTMAN, D. G., TETZLAFF, J., MULROW, C., GOTZSCHE, P. C., IOANNIDIS, J. P., 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 Med,* 6**,** e1000100.
MANCA, G., VANZI, E., RUBELLO, D., GIAMMARILE, F., GRASSETTO, G., WONG, K. K., PERKINS, A. C., COLLETTI, P. M. & VOLTERRANI, D. 2016. (18)F-FDG PET/CT

quantification in head and neck squamous cell cancer: principles, technical issues and clinical applications. *European Journal of Nuclear Medicine & Molecular Imaging,* 43**,** 1360-75.

MANIKANDAN, S. 2011a. Measures of central tendency: Median and mode. *J Pharmacol Pharmacother,* 2**,** 214-5.

MANIKANDAN, S. 2011b. Measures of central tendency: The mean. *J Pharmacol Pharmacother,* 2**,** 140-2.

MARTENS, R. M., NOIJ, D. P., ALI, M., KOOPMAN, T., MARCUS, J. T., VERGEER, M. R., DE VET, H., DE JONG, M. C., LEEMANS, C. R., HOEKSTRA, O. S., DE BREE, R., DE GRAAF, P., BOELLAARD, R. & CASTELIJNS, J. A. 2019. Functional imaging early during (chemo)radiotherapy for response prediction in head and neck squamous cell carcinoma; a systematic review. *Oral Oncology,* 88**,** 75-83.

MATOBA, M., TUJI, H., SHIMODE, Y., KONDO, T., OOTA, K. & TONAMI, H. 2017. The role of changes in maximum standardized uptake value of FDG PET-CT for post-treatment surveillance in patients with head and neck squamous cell carcinoma treated with chemoradiotherapy: preliminary findings. *British Journal of Radiology,* 90**,** 20150404.

MAYO, Z., SEYEDIN, S. N., MALLAK, N., MOTT, S. L., MENDA, Y., GRAHAM, M. & ANDERSON, C. 2019. Clinical Utility of Pretreatment and 3-Month ¹⁸F-Fluorodeoxyglucose Positron Emission Tomography/Computed Tomography Standardized Uptake Value in Predicting and Assessing Recurrence in T3-T4 Laryngeal Carcinoma Treated With Definitive Radiation. *Annals of Otology, Rhinology & Laryngology,* 128**,** 595-600.

MCGUINNESS, L. A. & HIGGINS, J. P. T. 2021. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Res Synth Methods,* 12**,** 55-61.

MEHANNA, H., PALERI, V., WEST, C. M. & NUTTING, C. 2010. Head and neck cancer--Part 1: Epidemiology, presentation, and prevention. *Bmj,* 341**,** c4684.

MEHANNA, H., WONG, W. L., MCCONKEY, C. C., RAHMAN, J. K., ROBINSON, M., HARTLEY, A. G., NUTTING, C., POWELL, N., AL-BOOZ, H., ROBINSON, M., JUNOR, E., RIZWANULLAH, M., VON ZEIDLER, S. V., WIESHMANN, H., HULME, C., SMITH, A. F., HALL, P., DUNN, J. & GROUP, P.-N. T. M. 2016. PET-CT Surveillance versus Neck Dissection in Advanced Head and Neck Cancer. *N Engl J Med,* 374**,** 1444-54.

MODY, M. D., ROCCO, J. W., YOM, S. S., HADDAD, R. I. & SABA, N. F. 2021. Head and neck cancer. *Lancet,* 398**,** 2289-2299.

MOELLER, B. J., RANA, V., CANNON, B. A., WILLIAMS, M. D., STURGIS, E. M., GINSBERG, L. E., MACAPINLAC, H. A., LEE, J. J., ANG, K. K., CHAO, K. S., CHRONOWSKI, G. M., FRANK, S. J., MORRISON, W. H., ROSENTHAL, D. I., WEBER, R. S., GARDEN, A. S., LIPPMAN, S. M. & SCHWARTZ, D. L. 2009. Prospective risk-adjusted [18F]Fluorodeoxyglucose positron emission tomography and computed tomography assessment of radiation response in head and neck cancer. *J Clin Oncol,* 27**,** 2509-15.

MOELLER, B. J., RANA, V., CANNON, B. A., WILLIAMS, M. D., STURGIS, E. M., GINSBERG, L. E., MACAPINLAC, H. A., LEE, J. J., ANG, K. K., CHAO, K. S., CHRONOWSKI, G. M., FRANK, S. J., MORRISON, W. H., ROSENTHAL, D. I., WEBER, R. S., GARDEN, A. S., LIPPMAN, S. M. & SCHWARTZ, D. L. 2010. Prospective imaging assessment of mortality risk after head-and-neck radiotherapy. *Int J Radiat Oncol Biol Phys,* 78**,** 667-74.

MURPHY, J. D., LA, T. H., CHU, K., QUON, A., FISCHBEIN, N. J., MAXIM, P. G., GRAVES, E. E., LOO, B. W., JR. & LE, Q. T. 2011. Postradiation metabolic tumor volume predicts outcome in head-and-neck cancer. *Int J Radiat Oncol Biol Phys,* 80**,** 514-21.

NELISSEN, C., SHERRIFF, J., JONES, T., GUEST, P., COLLEY, S., SANGHERA, P. & HARTLEY, A. 2017. The Role of Positron Emission Tomography/Computed Tomography Imaging in Head and Neck Cancer after Radical Chemoradiotherapy: a Single Institution Experience. *Clin Oncol (R Coll Radiol),* 29**,** 753-759.

NGUYEN-TAN, P. F., ZHANG, Q., ANG, K. K., WEBER, R. S., ROSENTHAL, D. I., SOULIERES, D., KIM, H., SILVERMAN, C., RABEN, A., GALLOWAY, T. J., FORTIN, A., GORE, E., WESTRA, W. H., CHUNG, C. H., JORDAN, R. C., GILLISON, M. L., LIST, M. & LE, Q. T. 2014. Randomized phase III trial to test accelerated versus standard fractionation in combination with concurrent cisplatin for head and neck carcinomas in the Radiation Therapy Oncology Group 0129 trial: long-term report of efficacy and toxicity. *J Clin Oncol,* 32**,** 3858-66.

NIEDERKOHR, R. D., GREENSPAN, B. S., PRIOR, J. O., SCHODER, H., SELTZER, M. A., ZUKOTYNSKI, K. A. & ROHREN, E. M. 2013. Reporting guidance for oncologic 18F-FDG PET/CT imaging. *J Nucl Med,* 54**,** 756-61.

NISHIMURA, G., YABUKI, K., HATA, M., KOMATSU, M., TAGUCHI, T., TAKAHASHI, M., SHIONO, O., SANO, D., ARAI, Y., TAKAHASHI, H., CHIBA, Y. & ORIDATE, N. 2016. Imaging strategy for response evaluation to chemoradiotherapy of the nodal disease in patients with head and neck squamous cell carcinoma. *Int J Clin Oncol,* 21**,** 658-667.

OYAMA, T., HOSOKAWA, Y., ABE, K., HASEGAWA, K., FUKUI, R., AOKI, M. & KOBAYASHI, W. 2020. Prognostic value of quantitative FDG-PET in the prediction of survival and local recurrence for patients with advanced oral cancer treated with superselective intra-arterial chemoradiotherapy. *Oncol Lett,* 19**,** 3775-3780.

PAK, K., CHEON, G. J., NAM, H. Y., KIM, S. J., KANG, K. W., CHUNG, J. K., KIM, E. E. & LEE, D. S. 2014. Prognostic value of metabolic tumor volume and total lesion glycolysis in head and neck cancer: a systematic review and meta-analysis. *J Nucl Med,* 55**,** 884-90.

PARK, J., CHANG, K. J., SEO, Y. S., BYUN, B. H., CHOI, J. H., MOON, H., LIM, I., KIM, B. I., CHOI, C. W. & LIM, S. M. 2014. Tumor SUVmax Normalized to Liver Uptake on (18)F-FDG PET/CT Predicts the Pathologic Complete Response After Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Nucl Med Mol Imaging,* 48**,** 295-302.

PERRI, F., LONGO, F., CAPONIGRO, F., SANDOMENICO, F., GUIDA, A., DELLA VITTORIA SCARPATI, G., OTTAIANO, A., MUTO, P. & IONNA, F. 2020. Management of HPV-Related Squamous Cell Carcinoma of the Head and Neck: Pitfalls and Caveat. *Cancers (Basel),* 12.

PHAM, N., JU, C., KONG, T. & MUKHERJI, S. K. 2022. Artificial Intelligence in Head and Neck Imaging. *Semin Ultrasound CT MR,* 43**,** 170-175.

PLAXTON, N. A., BRANDON, D. C., COREY, A. S., HARRISON, C. E., KARAGULLE KENDI, A. T., HALKAR, R. K. & BARRON, B. J. 2015. Characteristics and Limitations of FDG PET/CT for Imaging of Squamous Cell Carcinoma of the Head and Neck: A Comprehensive Review of Anatomy, Metastatic Pathways, and Image Findings. *AJR Am J Roentgenol,* 205**,** W519-31.

RANGANATHAN, P. & AGGARWAL, R. 2018. Common pitfalls in statistical analysis: Understanding the properties of diagnostic tests - Part 1. *Perspect Clin Res,* 9**,** 40-43.
RIAZ, S., BASHIR, H., IQBAL, H., JAMSHED, A., MURTAZA, A. & HUSSAIN, R. 2017. Impact and prognostic value of (18)F-fluorodeoxyglucose positron emission tomography-computed tomography scan in the evaluation of residual head and neck cancer: Single-center experience from Pakistan. *South Asian J Cancer,* 6**,** 81-83.
RIJO-CEDENO, J., MUCIENTES, J., ALVAREZ, O., ROYUELA, A., SEIJAS MARCOS, S., ROMERO, J. & GARCIA-BERROCAL, J. R. 2020. Metabolic tumor volume and total lesion glycolysis as prognostic factors in head and neck cancer: Systematic review and meta-analysis. *Head Neck,* 42**,** 3744-3754.
RIJO-CEDENO, J., MUCIENTES, J., SEIJAS MARCOS, S., ROMERO, J., ROYUELA, A., CARBONELL, S., BENLLOCH, R. & GARCIA-BERROCAL, J. R. 2021. Adding value to tumor staging in head and neck cancer: The role of metabolic parameters as prognostic factors. *Head Neck,* 43**,** 2477-2487.
ROBBINS, K. T., CLAYMAN, G., LEVINE, P. A., MEDINA, J., SESSIONS, R., SHAHA, A., SOM, P., WOLF, G. T., AMERICAN, H., NECK, S., AMERICAN ACADEMY OF, O.-H. & NECK, S. 2002. Neck dissection classification update: revisions proposed by the American Head and Neck Society and the American Academy of Otolaryngology-Head and Neck Surgery. *Arch Otolaryngol Head Neck Surg,* 128**,** 751-8.
RUOPP, M. D., PERKINS, N. J., WHITCOMB, B. W. & SCHISTERMAN, E. F. 2008. Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biom J,* 50**,** 419-30.
SAGARDOY, T., FERNANDEZ, P., GHAFOURI, A., DIGUE, L., HAASER, T., DE CLERMONT-GALLERAN, H., CASTETBON, V. & DE MONES, E. 2016. Accuracy of (18) FDG PET-CT for treatment evaluation 3 months after completion of chemoradiotherapy for head and neck squamous cell carcinoma: 2-year minimum follow-up. *Head Neck,* 38 Suppl 1**,** E1271-6.
SANDERSON, R. J. & IRONSIDE, J. A. 2002. Squamous cell carcinomas of the head and neck. *BMJ,* 325**,** 822-7.
SANGHERA, B., EMMOTT, J., WELLSTED, D., CHAMBERS, J. & WONG, W. L. 2009. Influence of N-butylscopolamine on SUV in FDG PET of the bowel. *Ann Nucl Med,* 23**,** 471-8.
SARIKAYA, I., ALBATINEH, A. N. & SARIKAYA, A. 2020. Revisiting Weight-Normalized SUV and Lean-Body-Mass-Normalized SUV in PET Studies. *J Nucl Med Technol,* 48**,** 163-167.
SARIKAYA, I. & SARIKAYA, A. 2020. Assessing PET Parameters in Oncologic (18)F-FDG Studies. *J Nucl Med Technol,* 48**,** 278-282.
SCHOBER, P., BOER, C. & SCHWARTE, L. A. 2018. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth Analg,* 126**,** 1763-1768.
SCHUTTRUMPF, L., MARSCHNER, S., SCHEU, K., HESS, J., RIETZLER, S., WALCH, A., BAUMEISTER, P., KIRCHNER, T., GANSWINDT, U., ZITZELSBERGER, H., BELKA, C. & MAIHOEFER, C. 2020. Definitive chemoradiotherapy in patients with squamous cell cancers of the head and neck - results from an unselected cohort of the clinical cooperation group "Personalized Radiotherapy in Head and Neck Cancer". *Radiat Oncol,* 15**,** 7.
SHEIKHBAHAEI, S., TAGHIPOUR, M., AHMAD, R., FAKHRY, C., KIESS, A. P., CHUNG, C. H. & SUBRAMANIAM, R. M. 2015. Diagnostic Accuracy of Follow-Up FDG PET or PET/CT in

Patients With Head and Neck Cancer After Definitive Treatment: A Systematic Review and Meta-Analysis. *AJR Am J Roentgenol,* 205**,** 629-39.
SHIMOMURA, H., SASAHIRA, T., YAMANAKA, Y., KURIHARA, M., IMAI, Y., TAMAKI, S., YAMAKAWA, N., SHIRONE, N., HASEGAWA, M., KUNIYASU, H. & KIRITA, T. 2014. [¹⁸F]fluoro-2-deoxyglucose-positron emission tomography for the assessment of histopathological response after preoperative chemoradiotherapy in advanced oral squamous cell carcinoma. *International Journal of Clinical Oncology.,* 19.
SHIN, S., PAK, K., KIM, I. J., KIM, B. S. & KIM, S. J. 2017. Prognostic Value of Tumor-to-Blood Standardized Uptake Ratio in Patients with Resectable Non-Small-Cell Lung Cancer. *Nucl Med Mol Imaging,* 51**,** 233-239.
SJOVALL, J., BITZEN, U., KJELLEN, E., NILSSON, P., WAHLBERG, P. & BRUN, E. 2016. Qualitative interpretation of PET scans using a Likert scale to assess neck node response to radiotherapy in head and neck cancer. *Eur J Nucl Med Mol Imaging,* 43**,** 609-16.
SROUSSI, H. Y., EPSTEIN, J. B., BENSADOUN, R. J., SAUNDERS, D. P., LALLA, R. V., MIGLIORATI, C. A., HEAIVILIN, N. & ZUMSTEG, Z. S. 2017. Common oral complications of head and neck cancer radiation therapy: mucositis, infections, saliva change, fibrosis, sensory dysfunctions, dental caries, periodontal disease, and osteoradionecrosis. *Cancer Med,* 6**,** 2918-2931.
STEPNICK, D. & GILPIN, D. 2010. Head and neck cancer: an overview. *Semin Plast Surg,* 24**,** 107-16.
TAHARI, A. K., CHIEN, D., AZADI, J. R. & WAHL, R. L. 2014. Optimum lean body formulation for correction of standardized uptake value in PET imaging. *J Nucl Med,* 55**,** 1481-4.
TANG, W. H., SUN, W. & LONG, G. X. 2020. Concurrent cisplatin or cetuximab with radiotherapy in patients with locally advanced head and neck squamous cell carcinoma: A meta-analysis. *Medicine (Baltimore),* 99**,** e21785.
UK, C. R. 2020. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers#heading-One [Accessed April 2020].
ULANER, G. A. 2018. Fundamentals of Oncologic PET/CT Elsevier Health Sciences.
VAINSHTEIN, J. M., SPECTOR, M. E., STENMARK, M. H., BRADFORD, C. R., WOLF, G. T., WORDEN, F. P., CHEPEHA, D. B., MCHUGH, J. B., CAREY, T., WONG, K. K. & EISBRUCH, A. 2014. Reliability of post-chemoradiotherapy F-18-FDG PET/CT for prediction of locoregional failure in human papillomavirus-associated oropharyngeal cancer. *Oral Oncol,* 50**,** 234-9.
VAN DEN HOFF, J., OEHME, L., SCHRAMM, G., MAUS, J., LOUGOVSKI, A., PETR, J., BEUTHIEN-BAUMANN, B. & HOFHEINZ, F. 2013. The PET-derived tumor-to-blood standard uptake ratio (SUR) is superior to tumor SUV as a surrogate parameter of the metabolic rate of FDG. *EJNMMI Res,* 3**,** 77.
VAN DER VELDT, A., SMIT, E. & LAMMERTSMA, A. 2013. Positron Emission Tomography as a Method for Measuring Drug Delivery to Tumors in vivo: The Example of [11C]docetaxel. *Frontiers in Oncology,* 3.
VANDERHOEK, M., PERLMAN, S. B. & JERAJ, R. 2012. Impact of the definition of peak standardized uptake value on quantification of treatment response. *J Nucl Med,* 53**,** 4-11.

VON ELM, E., ALTMAN, D. G., EGGER, M., POCOCK, S. J., GOTZSCHE, P. C., VANDENBROUCKE, J. P. & INITIATIVE, S. 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet,* 370**,** 1453-7.

WAHL, R. L., JACENE, H., KASAMON, Y. & LODGE, M. A. 2009. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med,* 50 Suppl 1**,** 122S-50S.

WANG, C., ZHAO, K., HU, S., HUANG, Y., MA, L., LI, M. & SONG, Y. 2020. The PET-Derived Tumor-to-Liver Standard Uptake Ratio (SUV TLR) Is Superior to Tumor SUVmax in Predicting Tumor Response and Survival After Chemoradiotherapy in Patients With Locally Advanced Esophageal Cancer. *Front Oncol,* 10**,** 1630.

WANG, H., WANG, B., ZHANG, X. & FENG, C. 2021. Relations among sensitivity, specificity and predictive values of medical tests based on biomarkers. *Gen Psychiatr,* 34**,** e100453.

WANG, L., BAI, J. & DUAN, P. 2019a. Prognostic value of 18F-FDG PET/CT functional parameters in patients with head and neck cancer: a meta-analysis. *Nucl Med Commun,* 40**,** 361-369.

WANG, L., BAI, J. & DUAN, P. 2019b. Prognostic value of ¹⁸F-FDG PET/CT functional parameters in patients with head and neck cancer: A meta-analysis. *Nuclear Medicine Communications,* 40**,** 361-369.

WESTERTERP, M., SLOOF, G. W., HOEKSTRA, O. S., TEN KATE, F. J., MEIJER, G. A., REITSMA, J. B., BOELLAARD, R., VAN LANSCHOT, J. J. & MOLTHOFF, C. F. 2008. 18FDG uptake in oesophageal adenocarcinoma: linking biology and outcome. *Journal of Cancer Research & Clinical Oncology,* 134**,** 227-36.

WONG, W. L. (2018). Manegment of head and neck cancer. *PET/CT In Head and Neck Cancer*, Springer, pp. 11-14.

XIE, P., YUE, J. B., FU, Z., FENG, R. & YU, J. M. 2010. Prognostic value of 18F-FDG PET/CT before and after radiotherapy for locally advanced nasopharyngeal carcinoma. *Ann Oncol,* 21**,** 1078-82.

ZASADNY, K. R. & WAHL, R. L. 1993. Standardized uptake values of normal tissues at PET with 2-[fluorine-18]-fluoro-2-deoxy-D-glucose: variations with body weight and a method for correction. *Radiology,* 189**,** 847-50.

ZHONG, J., SUNDERSINGH, M., DYKER, K., CURRIE, S., VAIDYANATHAN, S., PRESTWICH, R. & SCARSBROOK, A. 2020. Post-treatment FDG PET-CT in head and neck carcinoma: comparative analysis of 4 qualitative interpretative criteria in a large patient cohort. *Sci Rep,* 10**,** 4086.

ZIAI, P., HAYERI, M. R., SALEI, A., SALAVATI, A., HOUSHMAND, S., ALAVI, A. & TEYTELBOYM, O. M. 2016. Role of Optimal Quantification of FDG PET Imaging in the Clinical Practice of Radiology. *Radiographics,* 36**,** 481-96.

Appendices

Appendix 1

Search strategy:
Date of search: 16/05/2020
I: MEDLINE database: No limits
(1) (((Fluorine OR deoxyglucose) AND ("Positron Emission Tomography Computed Tomography" OR "Positron-Emission Tomography")).MESH) **OR** (((deoxyglucose OR desoxyglucose OR deoxy-glucose OR desoxy-glucose OR deoxy-d-glucose OR desoxy-d-glucose OR 2deoxyglucose OR 2deoxy-d-glucose OR fluorodeoxyglucose OR fluorodesoxyglucose OR fludeoxyglucose OR fluordeoxyglucose OR fluordesoxyglucose OR 18fluorodeoxyglucose OR 18fluorodesoxyglucose OR 18fluordeoxyglucose OR fdg* OR 18fdg* OR 18f-dg*OR 18f-fdg*) AND ((CAT OR CT OR ''computed tomograph*'' OR functional) AND (scan* OR X-ray* OR Imaging))).ti,ab) **OR** ((((2fluor* OR fluor* OR fludeoxy OR 18f* OR ''18 F*'') AND glucose) AND ((pet* OR positr*) AND (CAT OR CT OR ''computed tomograph*''))) OR (PET/CT OR PET-CT))
(2) (("Head and Neck Neoplasms" OR "Mouth Neoplasms" OR "Otorhinolaryngologic Neoplasms" OR "Tracheal Neoplasms" OR "Neoplasms, Squamous Cell" OR "carcinoma, squamous cell" OR "Palatine Tonsil" OR Palate OR "Vocal Cords" OR Mouth OR Tongue OR Larynx OR Pharynx OR Oropharyngeal OR Neck OR Head OR Lip).MESH) **OR** ((((Neck AND head) OR pharynx OR laryngopharyn* OR pharyn* OR oropharyn* OR hypopharyn* OR laryn* OR nasopharyn* OR Tongue OR Palat* OR tonsil* OR vocal cord* OR Lip* OR mouth OR oral OR Otorhinolaryngol* OR Trache*) AND (HNSCC* OR cancer* OR malignan* OR tumor* OR tumour* OR neoplas*OR oncolog* OR metasta* OR SCC OR squam*)).ti,ab)

(3) (("predictive value of tests" OR "Disease-Free Survival" OR "Survival Analysis" OR "Survival Rate" OR "Neoplasm Recurrence, Local" OR "Neoplasm, Residual" OR "Recurrence").MESH) **OR** (((sensitiv* AND specifici*) OR Chemother* OR radiother* OR posttest OR prognos* OR "disease free survival" OR recurren* OR relapse* OR residual* OR posttherap* OR ''Post-therap*'' OR SUV OR MTV OR TLG OR "standard* uptake value" OR "metabolic tumo* volume" OR "Total lesion glycolysis").ti,ab)
(4) (1 **AND** 2 **AND** 3)
Total: 2340

II. EMBASE database: No limits
(1) (((Fluorine OR deoxyglucose) AND ("Positron Emission Tomography Computed Tomography" OR "Positron-Emission Tomography")).Emtree) **OR** (((deoxyglucose OR desoxyglucose OR deoxy-glucose OR desoxy-glucose OR deoxy-d-glucose OR desoxy-d-glucose OR 2deoxyglucose OR 2deoxy-d-glucose OR fluorodeoxyglucose OR fluorodesoxyglucose OR fludeoxyglucose OR fluordeoxyglucose OR fluordesoxyglucose OR 18fluorodeoxyglucose OR 18fluorodesoxyglucose OR 18fluordeoxyglucose OR fdg* OR 18fdg* OR 18f-dg*OR 18f-fdg*) AND ((CAT OR CT OR ''computed tomograph*'' OR functional) AND (scan* OR X-ray* OR Imaging))).ti,ab) **OR (**((2fluor* OR fluor* OR fludeoxy

OR 18f*) AND glucose) AND ((pet* OR positr*) AND (CAT OR CT OR ''computed tomograph*''))) OR (PET/CT OR PET-CT))
(2) (("Head and Neck Neoplasms" OR "Mouth Neoplasms" OR "Otorhinolaryngologic Neoplasms" OR "Tracheal Neoplasms" OR "Neoplasms, Squamous Cell" OR "carcinoma, squamous cell" OR "Palatine Tonsil" OR Palate OR "Vocal Cords" OR Mouth OR Tongue OR Larynx OR Pharynx OR Oropharyngeal OR Neck OR Head OR Lip).Emtree) **OR** (((Neck AND head) OR pharynx OR laryngopharyn* OR pharyn* OR oropharyn* OR hypopharyn* OR laryn* OR nasopharyn* OR Tongue OR Palat* OR tonsil* OR vocal cord* OR Lip* OR mouth OR oral OR Otorhinolaryngol* OR Trache*) AND (HNSCC* OR cancer* OR malignan* OR tumor* OR tumour* OR neoplas*OR oncolog* OR metasta* OR SCC OR squam*)).ti,ab)
(3) ("predictive value of tests" OR "Disease-Free Survival" OR "Survival Analysis" OR "Survival Rate" OR "Neoplasm Recurrence, Local" OR "Neoplasm, Residual" OR "Recurrence").Emtree) **OR** ((sensitiv* AND specifici*) OR Chemother* OR radiother* OR posttest OR prognos* OR "disease free survival" OR recurren* OR relapse* OR residual* OR posttherap* OR ''Post-therap*'' OR SUV OR MTV OR TLG OR "standard* uptake value" OR "metabolic tumo* volume" OR "Total lesion glycolysis").ti,ab)
(4) (1 **AND** 2 **AND** 3)
Total: 5288

III: Cochrane CENTRAL database: No limits
(1) (((Fluorine OR deoxyglucose) AND ("Positron Emission Tomography Computed Tomography" OR "Positron-Emission Tomography")).MESH) **OR** (((deoxyglucose OR desoxyglucose OR deoxy-glucose OR desoxy-glucose OR deoxy-d-glucose OR desoxy-d-glucose OR 2deoxyglucose OR 2deoxy-d-glucose OR fluorodeoxyglucose OR fluorodesoxyglucose OR fludeoxyglucose OR fluordeoxyglucose OR fluordesoxyglucose OR 18fluorodeoxyglucose OR 18fluorodesoxyglucose OR 18fluordeoxyglucose OR fdg* OR 18fdg* OR 18f-dg*OR 18f-fdg*) AND ((CAT OR CT OR ''computed tomograph*'' OR functional) AND (scan* OR X-ray* OR Imaging))).ti,ab) **OR** ((((2fluor* OR fluor* OR fludeoxy OR 18f* OR ''18 F*'') AND glucose) AND ((pet* OR positr*) AND (CAT OR CT OR ''computed tomograph*''))) OR (PET/CT OR PET-CT))
(2) (("Head and Neck Neoplasms" OR "Mouth Neoplasms" OR "Otorhinolaryngologic Neoplasms" OR "Tracheal Neoplasms" OR "Neoplasms, Squamous Cell" OR "carcinoma, squamous cell" OR "Palatine Tonsil" OR Palate OR "Vocal Cords" OR Mouth OR Tongue OR Larynx OR Pharynx OR Oropharyngeal OR Neck OR Head OR Lip).MESH) **OR** ((((Neck AND head) OR pharynx OR laryngopharyn* OR pharyn* OR oropharyn* OR hypopharyn* OR laryn* OR nasopharyn* OR Tongue OR Palat* OR tonsil* OR vocal cord* OR Lip* OR mouth OR oral OR Otorhinolaryngol* OR Trache*) AND (HNSCC* OR cancer* OR malignan* OR tumor* OR tumour* OR neoplas*OR oncolog* OR metasta* OR SCC OR squam*)).ti,ab)

(3) (("predictive value of tests" OR "Disease-Free Survival" OR "Survival Analysis" OR "Survival Rate" OR "Neoplasm Recurrence, Local" OR "Neoplasm, Residual" OR "Recurrence").MESH) **OR** (((sensitiv* AND specifici*) OR Chemother* OR radiother* OR posttest OR prognos* OR "disease free survival" OR recurren* OR relapse* OR residual* OR posttherap* OR ''Post-therap*'' OR SUV OR MTV OR TLG OR "standard* uptake value" OR "metabolic tumo* volume" OR "Total lesion glycolysis").ti,ab)
(4) (1 **AND** 2 **AND** 3)

Total: 3411

Total from all three databases after deduplication and applying [English only] limit: 3150

Appendix 2

Appendix 1. Quality in Prognostic Studies (QUIPS) tool

Domains	Prompting items for Consideration	Ratings
Study Participation	a. Adequate participation in the study by eligible persons b. Description of the source population or population of interest c. Description of the baseline study sample d. Adequate description of the sampling frame and recruitment e. Adequate description of the period and place of recruitment f. Adequate description of inclusion and exclusion criteria	**High bias:** The relationship between the PF and outcome is very likely to be different for participants and eligible nonparticipants **Moderate bias**: The relationship between the PF and outcome may be different for participants and eligible nonparticipants **Low bias**: The relationship between the PF and outcome is unlikely to be different for participants and eligible nonparticipants
Study Attrition	a. Adequate response rate for study participants b. Description of attempts to collect information on participants who dropped out c. Reasons for loss to follow-up are provided d. Adequate description of participants lost to follow-up e. There are no important differences between participants who completed the study and those who did not	**High bias**: The relationship between the PF and outcome is very likely to be different for completing and non-completing participants **Moderate bias**: The relationship between the PF and outcome may be different for completing and non-completing participants **Low bias**: The relationship between the PF and outcome is unlikely to be different for completing and non-completing participants
Prognostic Factor Measurement	a. A clear definition or description of the PF is provided b. Method of PF measurement is adequately valid and reliable c. Continuous variables are reported or appropriate cut points are used d. The method and setting of measurement of PF is the same for all study participants e. Adequate proportion of the study sample has complete data for the PF f. Appropriate methods of imputation are used for missing PF data	**High bias**: The measurement of the PF is very likely to be different for different levels of the outcome of interest **Moderate bias**: The measurement of the PF may be different for different levels of the outcome of interest **Low bias**: The measurement of the PF is unlikely to be different for different levels of the outcome of interest
Outcome Measurement	a. A clear definition of the outcome is provided b. Method of outcome measurement used is adequately valid and reliable c. The method and setting of outcome measurement is the same for all study participants	**High bias**: The measurement of the outcome is very likely to be different related to the baseline level of the PF **Moderate bias**: The measurement of the outcome may be different related to the baseline level of the PF **Low bias**: The measurement of the outcome is unlikely to be different related to the baseline level of the PF
Study Confounding	a. All important confounders are measured b. Clear definitions of the important confounders measured are provided c. Measurement of all important confounders is adequately valid and reliable d. The method and setting of confounding measurement are the same for all study participants e. Appropriate methods are used if imputation is used for missing confounder data f. Important potential confounders are accounted for in the study design g. Important potential confounders are accounted for in the analysis	**High bias:** The observed effect of the PF on the outcome is very likely to be distorted by another factor related to PF and outcome **Moderate bias:** The observed effect of the PF on outcome may be distorted by another factor related to PF and outcome **Low bias:** The observed effect of the PF on outcome is unlikely to be distorted by another factor related to PF and outcome
Statistical Analysis and Reporting	a. Sufficient presentation of data to assess the adequacy of the analytic strategy b. Strategy for model building is appropriate and is based on a conceptual framework or model c. The selected statistical model is adequate for the design of the study d. There is no selective reporting of results	**High bias**: The reported results are very likely to be spurious or biased related to analysis or reporting **Moderate bias**: The reported results may be spurious or biased related to analysis or reporting **Low bias**: The reported results are unlikely to be spurious or biased related to analysis or

Appendix 3

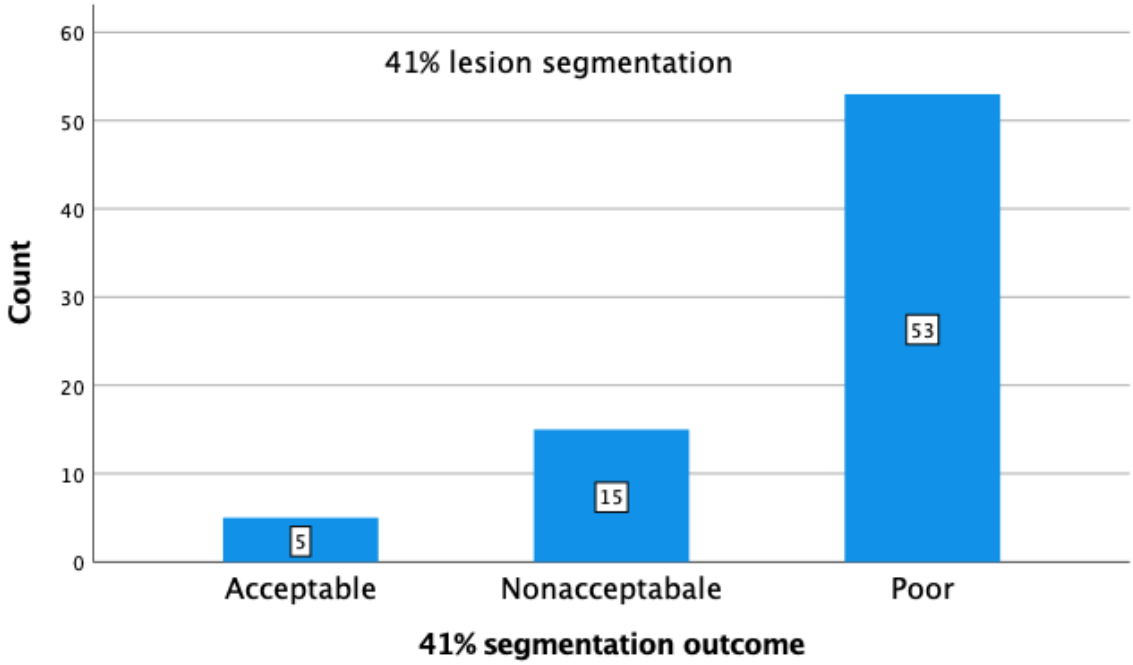


Figure 1**.** The visual accuracy of the 41% of the SUV_{max} segmentation in delineating post-treatment HNSCC MTV lesions. The Bar graph shows the non-accuracy of using the 41% segmentation in the majority of examined HNSCC lesions.

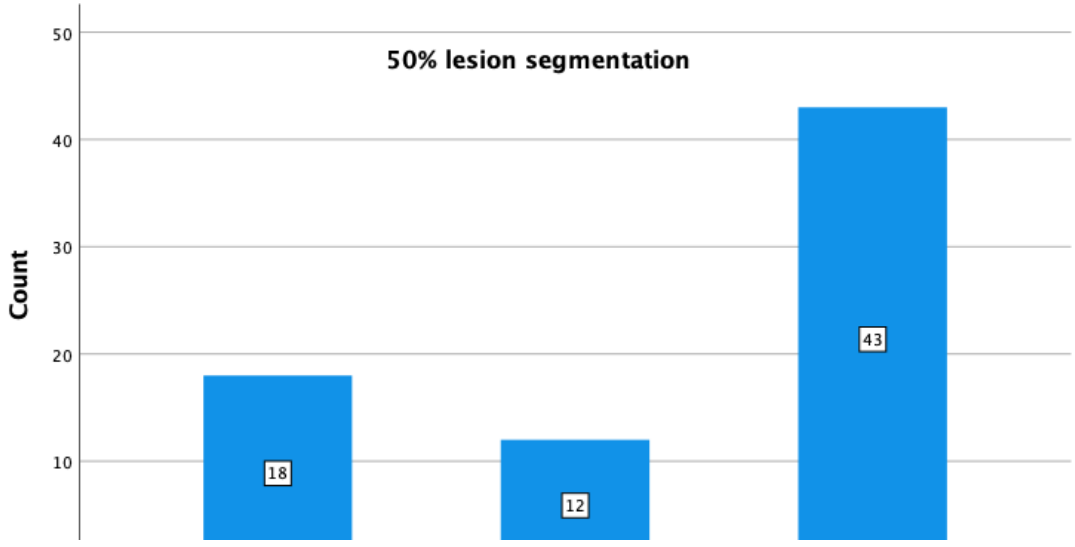


Figure 2**.** The visual accuracy of the 50% of the SUV_{max} segmentation in delineating post-treatment HNSCC MTV lesions. The Bar graph shows the non-accuracy of using the 41% segmentation in the majority of examined HNSCC lesions.

The results in this cohort showed that the 41% tumour segmentation was poor in 53 lesions (73%), nonacceptable in 15 lesions (26%), and acceptable in only five lesions (7%) (Figure 1). In addition, the 50% tumour segmentation was poor in 43 lesions (59%), nonacceptable in 12 lesions (16%), and acceptable in only 18 lesions (25%) (Figure 2).

Appendix 4

Section & Topic	No	Item
TITLE OR ABSTRACT		
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
ABSTRACT		
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)
INTRODUCTION		
	3	Scientific and clinical background, including the intended use and clinical role of the index test
	4	Study objectives and hypotheses
METHODS		
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)
Participants	**6**	Eligibility criteria
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)
	8	Where and when potentially eligible participants were identified (setting, location and dates)
	9	Whether participants formed a consecutive, random or convenience series
Test methods	**10a**	Index test, in sufficient detail to allow replication
	10b	Reference standard, in sufficient detail to allow replication
	11	Rationale for choosing the reference standard (if alternatives exist)
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test
	13b	Whether clinical information and index test results were available to the assessors of the reference standard
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy
	15	How indeterminate index test or reference standard results were handled
	16	How missing data on the index test and reference standard were handled
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory
	18	Intended sample size and how it was determined
RESULTS		
Participants	**19**	Flow of participants, using a diagram
	20	Baseline demographic and clinical characteristics of participants
	21a	Distribution of severity of disease in those with the target condition
	21b	Distribution of alternative diagnoses in those without the target condition
	22	Time interval and any clinical interventions between index test and reference standard
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
	25	Any adverse events from performing the index test or the reference standard
DISCUSSION		
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability
	27	Implications for practice, including the intended use and clinical role of the index test
OTHER INFORMATION		
	28	Registration number and name of registry
	29	Where the full study protocol can be accessed
	30	Sources of funding and other support; role of funders

Appendix 5

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) If applicable, explain how loss to follow-up was addressed
		(*e*) Describe any sensitivity analyses
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
		(c) Summarise follow-up time (eg, average and total amount)
Outcome data	15*	Report numbers of outcome events or summary measures over time
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period

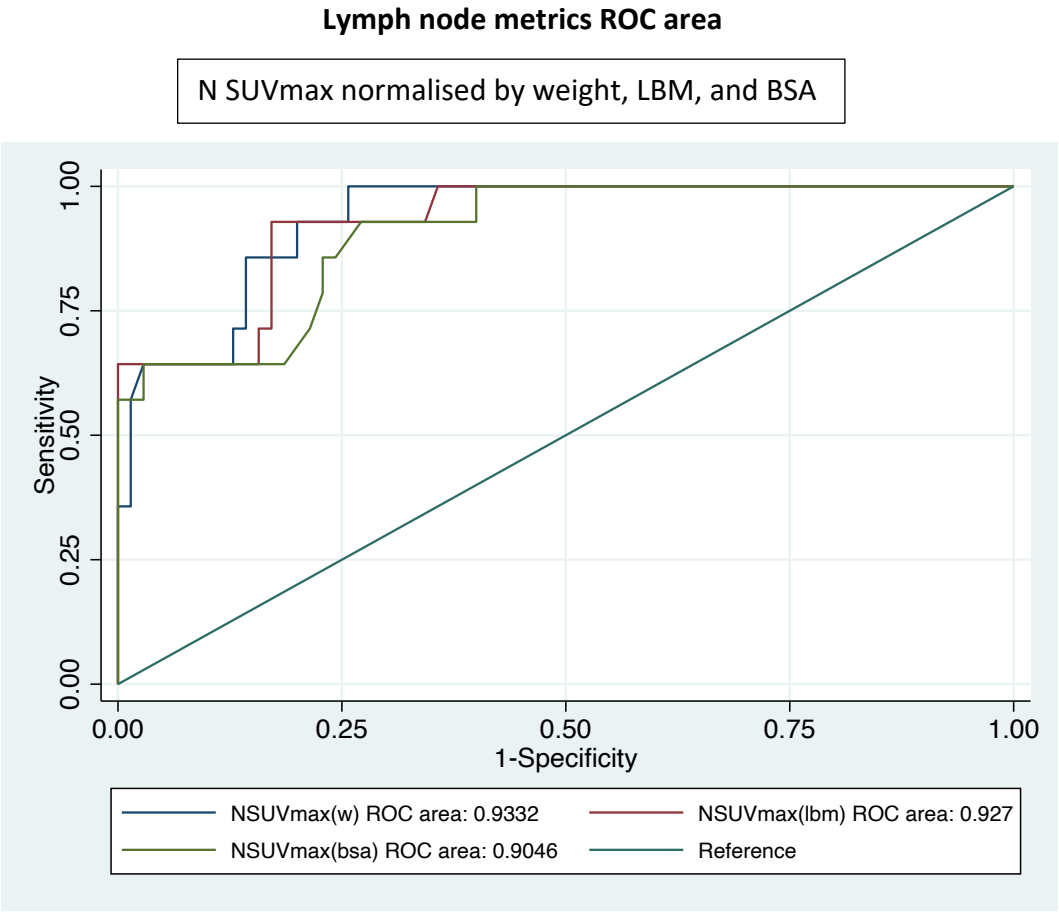
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

Appendix 6

Chapter 5: Sensitivity analysis



This figure illustrates the discriminative performance of N SUVmax normalised by weight, LBM (lean body mass), and BSA (body surface area) in detecting persistent HNSCC nodal lesions (N) at three months post-(chemo)radiotherapy. Patients who were scanned by the second scanner (about 8 patients) were excluded. The figure shows that there is a minimal difference in the ROC area values in comparison to the values in Figure 5.3 (Chapter 5), confirming the minimal effect of using two scanners.

Chapter 6: Sensitivity analysis

	Progression-free-survival			
Variables	**HR**	**[95% CI]**		***P* value**
N SUVmax(w)	1.41	1.24	1.60	**<0.0001**
N SUVmax(lbm)	1.57	1.33	1.86	**<0.0001**
N SUVmax(bsa)	3.24	2.06	5.10	**<0.0001**
N SUVpeak(w)	1.76	1.39	2.23	**<0.0001**
N SUVpeak(lbm)	2.08	1.55	2.80	**<0.0001**
N SUVpeak(bsa)	6.96	3.00	16.13	**<0.0001**

Univariable analysis of nodal SUVmax normalised by weight, LBM (lean body mass), and BSA (body surface area) with the Cox proportional hazard model for progression-free survival in combined HNSCC Patients who were scanned by the second scanner were excluded. The table shows that there is a minimal difference in the prognostic value of the nodal SUV metrics in comparison to the values in Table 6.2 (Chapter 6), confirming the minimal effect of using two scanners.

The End