Exploring the impact of metabolic imaging in head and neck cancer treatment

Diana Raquel Dias Domingues MSc, Michelle M. Leech PhD

1Cambridge University Hospitals NHS Trust, Cambridge, UK
2Trinity College Dublin, Dublin, Ireland

Correspondence
Diana Raquel Dias Domingues, Cambridge University Hospitals NHS Trust, 16 Hartree Lane, Trumpington, Cambridge CB2 9EQ, UK.
Email: diasdomd@tcd.ie or ddiasdomingues@gmail.com

Abstract
Background: Target volume delineation is performed with anatomical imaging for head and neck cancer. Molecular imaging allows the recognition of specific tumor regions. Its inclusion in the pathway could lead to changes in delineation and resultant treatment plans.

Methods: PRISMA methodology was adhered to when selecting the articles for analysis and only full articles were quality assessed.

Results: Seventeen articles were included. Gross tumor volume (GTV) primary, GTV nodal, and other target volumes were evaluated. Positron emission tomography/computerized tomography (PET/CT) produced smaller primary GTVs, although not with diffusion-weighted imaging-magnetic resonance imaging (DWI-MRI) or PET/MRI. The impact of these image modalities on GTV nodal did not display any consistency. Additionally, there was considerable heterogeneity in metrics comparing delineations. Four studies included appraised the dosimetric impact of the changes in target volume delineation.

Conclusion: Quantifying the impact of molecular imaging is difficult, due to heterogeneity in reporting metrics in molecular imaging modalities and a paucity of detail regarding delineation method and guideline adherence.

Keywords
delineation, head and neck, imaging, metabolic, radiotherapy

1 INTRODUCTION

Site and stage of the disease in addition to pathologic findings are used to establish treatment options in head and neck cancer. Surgery or radiation therapy (RT) as the sole treatment modality is recommended for approximately 30%–40% of patients with early stage disease. Both treatment approaches result in similar survival for various head and neck cancers. Additionally, around 60% of patients with locally or regionally advanced disease at diagnosis will require combined therapy.1

Therefore, it is important to evaluate methods to improve treatment efficacy in the RT treatment chain. This review evaluates the influence of metabolic imaging in target volume delineation and its impact in RT planning.

One of the primary stages in the RT pathway, regardless of treatment site, is the delineation of target volumes
and organs at risk (OARs). In head and neck cancer primary and nodal target volumes are routinely delineated. The gross tumor volume (GTV) is defined as “the gross palpable or visible/demonstrable extent and location of the malignant growth.” This volume can consist of the primary tumor (GTV primary), metastatic lymphadenopathy (GTV nodal), or another metastasis. In this review, both GTV primary and GTV nodal will be evaluated due to their importance in RT planning in this cohort of patients. Other target volumes will also be appraised where described. Dosimetrically, the clinical target volume (CTV) and the planning target volume (PTV) are of extreme importance. The CTV can be defined as the GTV with a margin including surrounding volumes of local subclinical involvement, edited for bone and air cavities. After the definition of the CTV planning margins are added to this volume to achieve the PTV. The PTV accounts for disparities in tumor position, size, and shape, as well as for discrepancies in patient position and beam geometries, both during and between fractions. This volume is used for the reporting of treatment doses and is considered to be representative of its respective CTV. In current practice, the delineation of volumes for head and neck cancer cases is completed using images from the planning computed tomography (CT) scan, and/or in combination with other imaging modalities.

With the development of complex RT techniques, the CTV to PTV margins of target volumes and planning organ at risk volumes for OARs have become smaller and therefore, it is crucial that GTVs and CTVs are outlined as accurately as possible. Inaccurate delineation can result in inadequate dose delivery to the target volume and consequently influence tumor control and potentially treatment related toxicities. Consequently, it is important to understand the factors that impact target volume delineation. Multiple factors have been cited as minimizing uncertainties in target volume delineation including the establishment of delineation protocols, education and training, a multidisciplinary approach, peer-review of delineated volumes, and the use of multimodality imaging.

A high number of articles have been published on interobserver variability (IOV) but few have considered its dosimetric impact. Additionally, IOV has been reduced using PET/CT in some cancer sites. Each imaging modality used in RT planning has its own set of characteristics, for example, Hounsfield units on CT scan, better soft tissue contrast on MRI than on CT and biological information on molecular imaging. The literature has shown that the use of multimodality imaging to assist with the contouring of structures for RT planning reduces IOV. FDG-PET/CT and other metabolic imaging techniques, such as diffusion-weighted magnetic resonance imaging (DWI-MRI) and PET/MRI, provide useful molecular information in terms of areas of higher metabolic activity within the tumor. Consequently, molecular imaging could influence target volume definition as it accounts for metabolic activity and potentially facilitates the visualization of subclinical and microscopic invasions. Molecular imaging techniques allow the visualization of numerous pathophysiological features of tumor tissue, such as hypoxia, proliferation, metabolism, and perfusion. Therefore, integrating these imaging modalities into the delineation process could assist in overcoming radiation resistance in tumor biology by allowing the recognition and targeting of those areas. In addition, the use of molecular imaging can assist in the escalation of dose to boost areas within the target volume with higher radiation resistance. This introduces the concept of “biological target volume,” which refers to those areas of the PTV that could receive a higher dose due to their lower radiosensitivity. A more accurate delineation of target volumes, in addition to the use of metabolic information has the potential to alter planning margins and provide more individualized treatment approaches based on tumor biology. The routine use of metabolic imaging for target volume delineation for RT planning has expanded with PET/CT being the most frequently used metabolic imaging modality offering potential definition of areas within the tumor with higher metabolic activity. Metabolic changes within the tumor micro-environment can affect tumor radiation resistance and consequently influence tumor control. The incorporation of this information into RT treatment planning might allow different doses to be delivered to different subvolumes, creating a more personalized treatment plan. The role of PET/CT has been established for tumors such as lung cancer, however in head and neck cancer at present there is no consensus about the role of metabolic imaging in delineation. In an area of treatment where organs at risk (OAR) with high seriality such as the spinal cord and optic apparatus are in very close proximity to the target volumes, accurate target volume delineation is of extreme importance. Therefore, it is important to evaluate if the incorporation of metabolic information into the process of target volume delineation impacts the definition of these volumes and the resultant dosimetry. Although the use of PET/CT for this purpose has increased there are other image modalities that can also provide metabolic information, such as diffusion weighted magnetic resonance imaging (DWI-MRI) and PET/MRI hybrids. Consequently, this review includes not only PET/CT but also other metabolic image modalities that are becoming more prevalent.
## METHODOLOGY

### 2.1 Study selection

Four hundred and seventy-three articles were identified through the initial electronic search using PRISMA methodology (Figure 1). After the application of the database filters, 247 remained. These studies were screened based on inclusion/exclusion criteria by title and abstract, which resulted in a total of 35 articles. From those 35 articles, 12 were excluded due to duplication and language other than English. This resulted in a total of 23 full-text articles to be assessed for eligibility. At this step, six records were excluded since three of them measured set up shifts and not volumetric or dosimetric outcomes and another three did not perform a quantitative comparison between anatomical and metabolic imaging. This process resulted in 17 studies for this review.\(^1\)\(^,\)\(^2\)

The metrics used to assess target volume delineation are volume, dice similarity coefficient (DSC), conformity/concordance index (CI), and modified Hausdorff distance (mHD). Metrics in relation to interobserver variability include DSC, CI, and mean difference. Dose to 95% of the volume (D95%) and dose to OAR are the two mostly commonly reported measures for dosimetric comparison. In addition, the percentage of change between anatomical and metabolic imaging target volume sizes was calculated in this review.

A quality assessment of each individual article was performed using the methodological index for nonrandomized studies (MINORS) (Table A1).

### 2.2 Evaluation metrics

In this review, many of the included studies did not report the same evaluation metric, making it challenging to establish recommendations for the optimal imaging modality for delineation of the GTV primary and GTV nodal in head and neck cancer. Mean GTV primary volume was the only metric common to all included articles. With respect to GTV nodal, mean volume was reported in eight of 17 included studies. In addition, three other metrics were reported in at least two of the 17 studies, which were the DSC, CI, and mHD. These metrics were also used in the evaluation of IOV between image modalities.

The DSC measures the similarity between two elements. It is calculated as the ratio between the interception of the two compared volumes and their average volume. This metric ranges from 0 to 1, the closer the DSC is to 1 the more similar the studied objects.\(^1\)\(^8\) In this

---

**PRISMA Flow Diagram**

![PRISMA Flow Diagram](wileyonlinelibrary.com)

**FIGURE 1** PRISMA flow diagram [Color figure can be viewed at wileyonlinelibrary.com]
review, the DSC was used to report the similarity in volumes both between observers using the same imaging modality and the difference between volumes contoured by the same person using both anatomical and metabolic imaging modalities.

CI is the ratio of intersection of two volumes as compared with the union of the two volumes under comparison. Slightly different values are usually achieved between CI and DSC when using the same sample, with the results from the DSC in general, higher. The DSC uses the average of the two compared volumes to create the ratio instead of their interception which can explain the higher values.

The mHD uses the mean distance between two-point sets as the measurement of comparison between them. This metric allows the comparison between similar and dissimilar points between the two sets with the purpose of ascertaining how similar they are. Comparison of these metrics for the contours delineated on anatomical and metabolic image sets permits an evaluation of the impact of the image modality on target volume delineation. Despite their limitations, metrics such as DSC and shift measures, such as mHD are helpful in evaluation of target volume variability by image modality.

Additionally, each metric has its own limitations. When calculating the DSC between two objects the resultant value will be the average of all volumes which introduces some bias. In addition, this metric does not take into account the extent of the experts agreements and disagreements which is very important in the context of target volume delineation. The mHD is a metric of positional displacement; therefore, it does not provide simple volumetric evaluation. Volumetric data are advantageous when comparing delineations on two different imaging modalities since it allows the calculation of the percentage change between the two.

3 | RESULTS AND DISCUSSION

3.1 | Gross tumor volume primary

3.1.1 | Volumetric evaluations

All the included studies reported either mean or median primary GTV volumes for both anatomical and metabolic imaging (Table 1). Fourteen of the studies used PET or PET/CT as the metabolic imaging modality, with two of the studies using DWI-MRI and two studies using a PET/MRI hybrid. All articles reported CT as the anatomical imaging modality, with eight using both CT and MRI. In four of the studies, the mean GTV primary volume was increased on the metabolic imaging modality compared to CT, while a similar result was observed in two of the studies comparing MRI and PET/CT. The overall relative difference in mean GTV volume when contoured on both anatomical and metabolic imaging was −61.8% to +29% (Figure 2).

Five of the seventeen papers evaluated IOV in the delineation of GTV primary. The study by Felice et al. assessed IOV and presented mean volumes for two different observers. The use of DWI-MRI coregistered with CT resulted in a decrease in target volume dimension for each of the observers. The volume delineated by observer 1 decreased by 45.3% with the use of metabolic imaging, while the volume delineated by observer two noticed a decrease of 48.3%. The same study also reported that the mean GTV difference between observers was smaller on DWI-MRI, with a mean difference of 0.17 cm³, while the mean difference between CT images was reported as 0.37 cm³.

To establish the difference in target volume the percentage of change in volume between delineations on anatomical and metabolic GTV primary was calculated for each included study. Eleven of the included studies reported a decrease in GTV primary mean volume when metabolic imaging was used in the delineation process. An increase in GTV primary volume was only reported in six of the included studies with most of them using CT as the comparator anatomical imaging modality. Across other tumor sites, similar conflicting results have been reported. In a study by Zheng et al., the use of PET/CT for the delineation of target volumes in patients with lung cancer found that PET/CT resulted in smaller GTVs for 10 of 20 patients. Another study by Vees et al. demonstrated that coregistered PET/CT decreased target volume size in patient with recurrent or residual gynecological tumors compared to CT alone.

Chauhan et al. reported an increase in target volume with the use of PET/CT for delineation in their study. Their hypothesis was that some areas deemed metabolically active on PET/CT might have been considered as edema or as inflammation in other image modalities or PET avid areas such as tonsil, base of tongue, among others, present in the delineation area. Another explanation is the disease site and stage of these patients. From the 21 patients included in this study 18 were classified as having either stage III or IV disease. In addition, 14 of these patients had oropharynx carcinoma. Locally advanced oropharyngeal carcinoma includes spread to structures such as the epiglottis, esophageal muscle, hyoid bone, and/or nasopharynx, among others. On anatomical imaging alone, unless these structures show a displacement in their position, it is difficult to assess if the tumor has spread to those areas and the clinician is reliant on the examination under anesthesia report.
| Study               | Sample size | Anatomical imaging used | Metabolic imaging used | Volume measurements for anatomical imaging\(^a\) | Volume measurements for metabolic imaging\(^a\) | % of change | DSC\(^b\) | CT\(^c\) | mHD\(^d\) | Measured IOV\(^e\) |
|--------------------|-------------|-------------------------|------------------------|-----------------------------------------------|-----------------------------------------------|------------|----------|---------|---------|-----------------|
| Igdem et al.\(^4\) | 26          | CT                      | PET/CT                 | Mean GTV-CT: 26.5 cm\(^3\)                     | Mean GTV-PET/CT: 35.5 cm\(^3\)               | CT vs. PET/CT: +20% | N/R     | N/R     | N/R     | N/R              |
| Guden et al.\(^2\) | 14          | CT-MRI (coregistered images) | PET-CT (coregistered images) | Mean GTVCT-MR: 49.25 cm\(^3\) (41–122.9 cm\(^3\)) | Mean GTVPET-CT: 18.8 cm\(^3\) (2.2–110.1 cm\(^3\)) | CT-MRI vs. PET-CT: −61.8% | N/R     | N/R     | N/R     | N/R              |
| Delouya et al.\(^1\) | 29          | CT                      | PET/CT                 | Mean GTV-CT: 24 cm\(^3\)                      | Mean GTV-PET/CT: 18 cm\(^3\)                 | CT vs. PET/CT: −25% | N/R     | N/R     | N/R     | N/R              |
| Thiagarajan et al.\(^1\) | 40        | CT and MRI              | PET/CT                 | Mean GTV-CT: 50.1 ml                           | Mean GTV-PET/CT: 33.9 ml                    | CT vs. PET: −32.3% CTMR vs. PET/CT: −29% | CTPET vs. CT: 0.54 CTMR vs. CT: 0.55 CTPETMRI vs. CT: 0.84 | N/R     | N/R     | N/R     |
| Chatterjee et al.\(^2\) | 20        | CECT\(^f\)              | PET/CT                 | Mean GTV-CECT: 36.56 cm\(^3\) (3.38–184 cm\(^3\); SD: 44.14) | Mean GTVPET/CT: 25.16 cm\(^3\) (1.62–166 cm\(^3\); SD: 35.81) | CETCT vs. PET/CT: −31.2% CECT vs. PET/CT: 0.68 (range: 0.2–1) | N/R     | N/R     | N/R     |
| Venkada et al.\(^4\) | 26          | CECT                    | PET/CT                 | Mean CT-GTV: 54.58 cm\(^3\) (SD: 64.47 cm\(^3\)) | Mean GTVPET/CT: 48.43 cm\(^3\) (SD: 53.21 cm\(^3\)) | CECT vs. PET/CT: −15.8% (SD: 41.49%) | N/R     | N/R     | N/R     |
| Arslan et al.\(^4\) | 37          | CT                      | PET/CT                 | Median GTV-CT: 55.77 (7.16–390.13)             | Median GTV-PET/CT: 32.71 (3.14–311.23)       | CT vs. PET/CT: −36.8% | N/R     | N/R     | N/R     |
| Anderson et al.\(^3\) | 14         | CT and MRI              | PET/CT                 | Mean GTV-CT: 45 cm\(^3\)                       | Mean GTV-PET/CT: 35 cm\(^3\)               | CT vs. PET/CT: −22.2% MRI vs. PET/CT: −28.6% | N/R     | N/R     | N/R     | Between observers: CT: 35% MRI: 27% PET/CT: 28% |
| Chauhan et al.\(^2\) | 21         | CECT and MRI and CECT-MRI (coregistered images) | PET/CT                 | Mean GTV-CT: 29.65 cm\(^3\) (1.58–115.33 cm\(^3\); SD: 31.27) | Mean GTV-PET/CT: 32.05 cm\(^3\) (1.48–108.20 cm\(^3\); SD: 33.75) | CT vs. MRI: −16.2% CT vs. PET/CT: +8.1% MRI vs. PET/CT: +29% | CT-MRI: 0.63 (SD: 0.23) PET-CT: 0.61 (SD: 0.23) PET-MRI: 0.57 (SD: 0.16) | N/R     | N/R     | N/R     |
| Bird et al.\(^2\) | 11          | CT and MRI              | PET/CT                 | Mean GTV-CT: 11.9 cm\(^3\) (1.6–34.5 cm\(^3\); SD: 4.5) | Mean GTVPET/CT: 9.5 m\(^3\) (1.5–24.6 cm\(^3\)) | CT vs. PET/CT: −20.2% CTMRI vs. PET/CT: −32.6% MRI vs. PET/CT: −25.2% | CT-PET/CT: 0.55 (SD: 0.11) MR-PET/CT: 0.61 (SD: 0.06) CTMRI-PET/CT: 0.08 (SD: 0.08) CT-MRI: 0.57 (SD: 0.09) | N/R     | N/R     | N/R     | CT: 0.37 (SD: 0.12) CT-MRI: 0.44 (SD: 0.09) MRI: 0.47 (SD: 0.09) DSC: |
| Study                      | Sample size | Anatomical imaging used | Metabolic imaging used | Volume measurements for anatomical imaginga | Volume measurements for metabolic imaginga | % of change | DSCb | CIc | mHDd | Measured IOVc |
|---------------------------|-------------|-------------------------|------------------------|---------------------------------------------|---------------------------------------------|-------------|------|-----|------|---------------|
| Leclerc et al. 40         | 41          | CT                      | PET/CT                | GTV-CT: 40.4 cm³                             | GTV-PET/CT: 28.8 cm³                        | CT vs. PET/CT: −28.7% | CT-CTMRI: 0.62 (SD: 0.09) | MRI-CTMRI: 0.87 (SD: 0.1) | MRI-CTMRI: 0.74 (SD: 0.17) | CT: 0.57 (SD: 0.15) |
| De Felice et al. 22       | 8           | CECT                    | DW-MRI                | Mean GTV-CT investigator 1(12a): 11.28 cm³ (2.21–27.62 cm³) | Mean GTV-DW/MRI investigator 1(12a): 6.17 cm³ (1.17–16.35 cm³) | Investigator 1 CT vs. DW/MRI: −45.3% | N/R | N/R | N/R | Mean difference GTV-CT: 0.37 cm³ (−1.68–0.93 cm³) |
| Wu et al. 28              | 20          | Ct and MRI              | PET/CT                | Mean GTVCT-MRI: 12.9 cm³ (SD: 7.7 cm³)       | Mean GTVCT-MRIPET: 13.5 cm³ (SD: 6.5 cm³)   | CTMRI vs. CTMRIPET: +4.4% | N/R | N/R | N/R | Mean difference GTV-DWI/ MRI: 0.17 cm³ (−0.26–0.61 cm³) |
| Gudi et al. 36            | 10          | CECT                    | PET/CT                | Mean GTV-CT: 25.9 cm³ (SD: 14.2 cm³)         | Mean GTV-PET/CT: 21.4 cm³ (SD: 16.7 cm³)    | CT vs. PET/CT: −17% | N/R | N/R | N/R | DSC: GTV-CT: 0.57 (SD: 0.12) |
| Wang et al. 29            | 11          | CT                      | PET/MRI               | GTV-CT: 13.2 cm³                              | GTV-PET/MRI: 14.3 cm³                        | CT vs. PET/MRI: +8.3% | CT vs. PET/MRI: 0.63 (SD: 0.11) | N/R | CT vs. PET/MRI: 1.6 mm (SD: 0.7 mm) | N/R |
| Samolyk-Kogaczewska et al.13 | 10            | CT and MRI              | PET and PET/MRI       | Mean GTV-CT: 21.1 cm³ (4.08–82.25 cm³)       | Mean GTV-PET/CT: 23.2 cm³ (1.73–100.6 cm³)  | CT vs. PET: +10% MRI vs. PET: −8.3% | CT and MRI: 0.74 (SD: 0.66–0.85) | CT and PET: 0.72 (0.57–0.79) | CT and PET/MRI: 0.55 (0–0.82) | CT and MRI: 13.2 mm (4–19 mm) |

(Continues)
| Study                          | Sample size | Anatomical imaging used | Metabolic imaging used | Volume measurements for anatomical imaging^a^ | Volume measurements for metabolic imaging^a^ | % of change | DSC^b^ | CI^c^ | mHD^d^ | Measured IOV^e^ |
|-------------------------------|-------------|-------------------------|------------------------|-----------------------------------------------|-----------------------------------------------|-------------|--------|-------|--------|-----------------|
| Cardoso et al. ^1^            | 10          | CT and MRI              | DWI-MRI and PET/CT     | GTV-CT: 10.92 cm³ (8.32–13.52 cm³)            | GTV-CTPET: 10.52 cm³ (8.25–12.78 cm³)         | CT vs. CTPET/CT: −3.7% | N/R    | N/R   | N/R    | N/R             |
|                              |             |                         |                        | GTV-CTPETMRI: 13.38 cm³ (10.84–15.92 cm³)    | GTV-CTPETMRI/DWI-MRI: 13.37 cm³ (10.35–16.39 cm³) | CT vs. CTPETMRI: +22.5% |        |       |       |                  |
|                              |             |                         |                        |                                               |                                               | CT vs. CTPETMRI/DWI-MRI: +22.4%               |        |       |       |                  |

^a^Mean, standard deviation, median, range.

^b^Dice similarity coefficient.

^c^Concordance index/conformity index.

^d^Modified Hausdorff distance.

^e^Interobserver variation in target volume delineated.

^f^Contrast enhancement CT.
However, the presence of microscopic spread will be observed on PET images as highly active, which may cause the increase in GTV primary volume. Hence, the ratio of patients with locally advanced head and neck cancer is considerably higher in this study which may influence the results.

Bird et al. also assessed patients with oropharyngeal carcinoma, although in contrast to Chauhan et al., the majority of patients had early-stage disease. In this study, metabolic images resulted in smaller target volumes relative to anatomical images. In the study by Wu et al. the use of PET/CT for the delineation of GTV primary illustrated an increase in volume of +4.4%, relative to anatomical imaging. The authors hypothesize that this small increase in volume could be related to the high metabolic rate in areas where spread is still at the microscopic level. Such changes would not be observed on anatomical images, though one could speculate also difficult to fully appreciate on metabolic imaging.

Wang et al. studied the impact of a PET/MRI hybrid in target volume delineation, showing an increase in volume with the use of this modality relative to anatomical imaging alone. However, the difference was not statistically significant. The authors theorized that there was an improvement in gross tumor visualization due to the PET/MRI in the individual cases where there was a large volumetric difference. It is well established that MRI is useful in head and neck cancer delineation since it offers better soft tissue contrast compared with CT images. Therefore, it is plausible that the use of the PET/MRI hybrid brings the same advantage to this field. In the studies of Anderson et al., Bird et al., and Samolyk-Kogaczewska et al., the GTV-MRI was slightly larger than the GTV-CT. In the study by Samolyk-Kogaczewska et al., in 80% of patients the GTV-MRI volume was larger than the GTV-CT. Again, this is associated with a more accurate definition of the tumor infiltration boundary on MRI than on CT. The authors also theorized that the increase in volume between GTV-CT and GTV-PET could be a result of the artifacts caused by dental fillings.

Cardoso et al. evaluated the use of DWI-MRI in target volume delineation. Their study reported an increase in GTV primary volume when combined anatomical and metabolic MRI images were used for delineation, compared to anatomical MRI alone. In the same study, when comparing CT and PET/CT images without MRI image sets, the target volumes delineated were smaller. This result again supports the theory that the enhanced soft tissue contrast provided by MRI results in increased target volumes in head and neck delineation. However, another study by Felice et al. showed that GTVs delineated using DWI-MRI images decreased in volume compared with those delineated on CT images. This discrepancy may be due to the susceptibility of DWI-MRI to various artifacts and only shows areas of high cellularity. The study by Cardoso et al. also mentioned that 50% of their observers found DWI images more useful in the delineation of patients with oropharynx-based tumors when compared with nasopharynx. This could explain the disparity of the results between those studies, as all the patient population in the study by Felice et al. had oropharynx-based primary tumors.

It is important to note that while mean volume of the GTV primary is reported in most of the included studies, a large range of volumes is reported for both anatomical and metabolic imaging. Patients included in these studies represent a wide population with tumors of any site and stage, rather than specific focus on, for example, locally...
advanced tumors only, which explains the disparity in results. Additionally, it is common in clinical practice to use contrast-enhanced CT in planning CTs for this cohort of patients. However, only 5 of the 17 studies mention its use. Contrast-enhanced CT allows static and dynamic vascular assessments of a tumor, improving target volume delineation.\(^{33}\)

Based on the results of the included studies in this review, it can be concluded that there is a trend for a decrease in the GTV primary volume when PET/CT is used compared to CT alone. It is not possible to draw conclusions from the other modalities but should be noted that Jager et al.\(^{34}\) indicated that MRI volumes were significantly larger than pathological volumes in locally advanced hypopharyngeal and laryngeal cancers, up to twice the volume in some instances. A decrease in target volume size would allow the dose to this volume to possibly be escalated, which will increase the probability of local disease control can while the toxicity of OAR remains acceptable.

### 3.1.2 Dice similarity coefficient

The DSC was reported in six of the 17 articles. The smallest reported DSC value was 0.55 between CT and PET/MRI in the study by Samolyk-Kogaczewska et al.\(^{15}\) and the study by Bird et al.\(^{27}\) in the comparison between CT and PET/CT image sets. The highest reported DSC value was 0.79 between anatomical and metabolic images. It was reported by Wu et al.\(^{28}\) in the comparison between CT and MRI coregistered images and coregistered CT, MRI, and PET/CT. Concerning IOV, the highest reported DSC between functional and metabolic image sets was 0.73, in the comparison between CT and PET in the study by Cardoso et al.\(^{32}\) However, the smallest interobserver DSC value reported was also between CT and PET image sets in the study by Anderson et al.\(^{31}\) The DSC metric in the included articles was used to evaluate the similarity between the delineated anatomical and metabolic GTVs. Bird et al.\(^{27}\) reported a DSC of 0.55 between CT and PET/CT images, which increased to 0.6 between CT-MRI and PET-CT. This is likely related to the better soft tissue contrast provided by the MRI compared to CT only. One study reported a DSC value of 0.79, which compared coregistered CT-MRI images with coregistered CT-MRI-PET,\(^{28}\) again indicating that MRI allows the enhanced appreciation of soft tissue infiltration. However, in two studies that used PET/MRI the DSC between these images and CT image sets were 0.55 and 0.63. This questions if the PET/MRI hybrid retains all the soft tissue contrast advantages of anatomical MRI or if in the presence of a metabolic image modality the observer is heavily influenced by the biological information provided.

The DSC was also evaluated for the GTV primary in the context of IOV. In this review, PET/CT images improved IOV when compared with CT alone. A similar result has been observed in other cancer sites. Vesprini et al.\(^{6}\) appraised observer variability in target volume delineation for gastro-esophageal cancer between PET/CT and CT alone. Interobserver and intraobserver both improved using PET/CT. Eight out of ten patients had an increase in PTV overlap between observers when PET-CT was used. In the study by Cardoso et al.,\(^{32}\) the same results are demonstrated. Though, when MRI and DWI-MRI images sets are added to the comparison, the DSC value decreased. The influence of this image modality in this coefficient might be related to multiple factors, such as the lack of familiarization of the clinicians with this type of images and the fact that the scans were acquired in different positions.\(^{30,32}\)

### 3.1.3 Conformity index

Three of the seventeen articles evaluated conformity index (CI) with the highest value of 0.84, reported by Thiagarajan et al.\(^{35}\) between CT and coregistered CT, PET, and MRI. The smallest reported CI was 0.33 and was reported by Bird et al.\(^{27}\) comparing CT and PET. The study by Cardoso et al.\(^{32}\) reported the smallest interobserver CI in the comparison between CT, PET and MRI coregistered image sets and DWI-MRI. Two of the studies showed very limited similarity between target volumes based on different image modalities. CI ranges from 0 to 1, with the score of 1 representing the perfect match. Chauhan et al.\(^{25}\) reported a CI of 0.47, while Bird et al.\(^{27}\) reported a value of 0.33 between CT and PET/CT images. However, in the study by Bird et al.\(^{27}\) this value increased slightly to 0.36 with the use of coregistered CT-MRI as a comparator to PET/CT. Routinely, the MRI and PET/CT images that are used in delineation are performed in different positions than the planning CT scan which makes their use in the delineation process challenging. Both studies utilized rigid registration, which in the head and neck area creates a challenge due to the mobility of structures. It would be beneficial in the future to perform studies where the patient position is identical for all scans.

In addition, the use of PET/CT improved the CI between observers.\(^{32,36}\) However, if MRI and DWI-MRI image sets are added this appears to decrease.\(^{32}\) This may again be related to coregistration issues and lack of familiarity of clinicians with the DWI-MRI images, with limited research regarding standardized apparent diffusion
3.1.4 | Modified Hausdorff distance

Only 2 of the 17 articles measured the modified Hausdorff distance (mHD) with a difference of 12.4 mm between CT and PET being reported by Samolyk-Kogaczewska et al. The same study also reported a difference of 16.2 mm in the comparison between CT and PET/MRI. The other study that evaluated mHD was performed by Wu et al., reporting a difference between CT and PET/MRI of just 1.6 mm. The range of mHD values varied between 1.6 and 16.2 mm, with both values reported between CT and PET/MRI. A smaller mHD value suggests greater similarity between two objects. Samolyk-Kogaczewska et al. mentioned that the high mHD reported on their study is suboptimal and that this might be caused by mHD being responsive to shape changes of the measured contours, since it matches two objects based on their edge points.

3.2 | Gross tumor volume nodal

3.2.1 | Volumetric evaluations

Eight of the seventeen included articles evaluated GTV nodal volumes for both anatomical and metabolic image modalities (Table 2). Four of these articles reported PET/CT as the metabolic imaging modality and the other two PET/MRI hybrids. All of the studies reported CT as the anatomical image modality of comparison, with three using both CT and MRI. From the eight articles, six reported mean GTV nodal measurements, with the other two only reporting range values. Two of the studies reported a decrease in the mean GTV nodal when comparing CT and PET/CT image sets. Thiagarajan et al. reported a decrease of 14.7% in GTV nodal volume when comparing the two previously mentioned image sets. However, in the same study when comparing coregistered CT and MRI images and PET/CT there was an increase in volume of 1.45%. The other decrease in the mean GTV was reported by Chatterjee et al. and was 31.2% between the CT and PET/CT datasets. All other articles reported an increase in nodal target volumes, ranging between 15.2% and 21.2%. In addition, Delouya et al. reported that the percentage of change between anatomical and metabolic imaging was statistically insignificant (Figure 3). Eight of the seventeen included articles reported volumetric data for GTV nodal. Three reported an increase in GTV nodal with the use of metabolic imaging relative to anatomical imaging. As observed in Table 2, it is not possible to fully establish if the use of metabolic imaging increases or decreases target volumes for RT planning and even a trend cannot be recognized. Additionally, one of the studies deemed the difference between GTV-CT and GTV-PET/CT as statistically insignificant, while two other report an increase in GTV nodal volume.

Thiagarajan et al. reported a decrease in GTV nodal when comparing CT and PET/CT images, but an increase in volume with CT/MRI images as the anatomical comparator. Additionally, Chatterjee et al. also reported a decrease in GTV nodal volume when comparing CT and PET/CT images. These studies included very similar populations, both in tumor site and TNM staging. The tumor sites included in the studies were tonsil, base of tongue, pharyngeal wall and soft palate. This proximity between included populations might explain the decrease in GTV nodal volume in both comparisons between CT and PET/CT images. Additionally, the literature has already showed that MRI images provide better soft tissue definition and has the potential to allow a more accurate delineation of target volumes, which might explain the small decrease in target volume size between CT only and CT/MRI, although this image modality has its own limitations such as geometrical distortion. The use of coregistered images is common practice in the world of head and neck delineation, although since the patient is, in most cases, scanned in different positions, some problems can be encountered. Thiagarajan et al. acknowledge this and confirm that when registration was deemed suboptimal for planning purposes the case was excluded from the study. The same authors also mention that, in their study, nodal volumes were identified with the use of metabolic imaging that were missed on both CT and MRI. This might explain the slight increase in GTV nodal volume noted in this study between MRI and fused CT and PET image sets. Though Delouya et al. found differences in nodal target volumes between CT and PET/CT statistically insignificant. The authors theorized that PET/CT might overstage patients identifying benign inflammatory nodules as cancer. This hypothesis is controversial since it is already established that PET/CT is accurate in the detection of lymph node involvement.

3.2.2 | Dice similarity coefficient

The DSC for GTV nodal volumes was reported in two of the eight articles. Chatterjee et al. reported a value of 0.91 comparing CT and PET/CT. Wang et al. evaluated the DSC between CT and PET/MRI image sets with a resultant value of 0.69. This metric was used to evaluate...
| Study                  | Sample size | Anatomical imaging used                                      | Metabolic imaging used | Volume measurements for anatomical imaging\(^a\) | Volume measurements for metabolic imaging\(^a\) | % of change | DSC\(^b\) | CT\(^c\) | mHD\(^d\) |
|-----------------------|-------------|--------------------------------------------------------------|------------------------|--------------------------------------------------|---------------------------------------------|-------------|-----------|---------|----------|
| Guden et al.\(^{42}\) | 14          | CT-MR (coregistered images)                                  | PET-CT (coregistered images) | CT: 2–95.3 cm\(^3\)                              | PET-CT: 1.3–123.1 cm\(^3\)                | N/R         | N/R      | N/R     | N/R      |
| Delouya et al.\(^{14}\) | 29          | CT                                                           | PET/CT                 | N/R                                              | N/R                                         | No statistically significant volume change | N/R      | N/R      | N/R      |
| Thiagarajan et al.\(^{38}\) | 40          | CT and MRI                                                   | PET/CT                 | Mean GTV-CT: 40.9 ml                             | Mean GTV-PET/CT: 34.9 ml                  | N/R         | CTPET vs. CT: 0.76 | N/R      | N/R      |
| Chatterjee et al.\(^{37}\) | 20          | CECT                                                         | PET/CT                 | Mean GTV-CECT: 32.48 cm\(^3\) (1.06–111.05 cm\(^3\); SD: 36.63) | Mean GTV-PET/CT: 32.21 cm\(^3\) (1.06–74.61 cm\(^3\); SD: 37.09) | 0.83%       | 0.91 (range: 0–1) | N/R      | N/R      |
| Venkada et al.\(^{45}\) | 26          | CECT                                                         | PET/CT                 | Mean GTV-CT: 11.04 cm\(^3\) (SD: 14.87 cm\(^3\)) | Mean GTV-PET: 12.72 cm\(^3\) (SD: 15.46 cm\(^3\)) | +15.2%      | N/R      | N/R     | N/R      |
| Arslan et al.\(^{46}\) | 37          | CT                                                           | PET/CT                 | Median GTV-CT: 5.29 (0–126) | Median GTV-PET: 7.49 (0–114.46) | +41.6%      | N/R      | N/R     | N/R      |
| Wang et al.\(^{29}\)  | 11          | CT                                                           | PET/MRI                | GTV-CT: 19.0 cm\(^3\)                            | GTV-PET/MRI: 23.0 cm\(^3\)                | +21.2%      | 0.69 (SD: 0.1) | N/R      | CT vs. PET/MRIGTV: 2.3 mm (SD: 1.5 mm) |
| Samolyk-Kogaczewska et al.\(^{35}\) | 10          | CT and MRI                                                   | PET and PET/MRI        | GTV-CT: 0–2.58 cm\(^3\) | GTV-PET: 0–5.07 cm\(^3\)                | N/R         | N/R      | N/R     | N/R      |

\(^a\)Mean, standard deviation, median, range.

\(^b\)Dice similarity coefficient.

\(^c\)Concordance index/conformity index.

\(^d\)Modified Hausdorff distance.
the similarity between the delineated anatomical and metabolic GTV nodal, and it was only reported in two studies. Chatterjee et al.\textsuperscript{37} reported a DSC of 0.91 between contrast-enhanced CT and PET/CT images. This value represents the highest similarity coefficient from all the reported values. The high degree of similarity could be explained by the consistency in tumor site and T and N staging of the included population. Relative to the nodal staging PET/CT only upstaged one of 20 patients from N0 to N2b. However, the fact that the same radiation oncologist outlined all target volumes may have created an intraobserver bias, leading to this DSC of 0.91. Wang et al.\textsuperscript{29} reported a DSC of 0.69. This value is more in line with the values reported for GTV primary. The authors mention that there were numerous patients in the study that had the GTV substantially altered using PET/MRI, whether increasing or decreasing in volume, which explains the 0.69 DSC value. The paucity of DSC values for GTV nodal reported does not allow any conclusions to be drawn.

3.2.3 Conformity index

Only one of the eight studies that evaluated GTV nodal volumes reported CI results. Thiagarajan et al.\textsuperscript{35} found the highest CI between functional and metabolic image sets to be between CT and PET/MRI with a value of 0.84 compared to a CI of 0.76 between CT and PET/CT. Thiagarajan et al.\textsuperscript{35} demonstrated an increase in CI in GTV nodal when using CT/PET/MRI images compared to CT only. This suggests that the use of multimodality imaging may improve conformity in target volume delineation of GTV nodal volumes. However, this is a single report and is based on oropharyngeal patients only.

3.2.4 Modified Hausdorff distance

Wang et al.\textsuperscript{29} was the only publication reporting mHD regarding GTV nodal. The mHD was reported as 2.3 mm between PET/MRI and CT images, which is greater for GTV nodal than for GTV primary. As previously discussed, mHD is highly influenced by the edge point of the compared volumes instead of the volumes as a whole. GTV nodal volumes are more complex to delineate than GTV primary volumes.\textsuperscript{39} This may explain the slight increase in mHD for GTV nodal.

3.3 Other target volumes evaluated

3.3.1 Volumetric assessments on clinical target volume and planning target volume

CTV and PTV volumetric assessments were performed in only three out of the 17 included papers. Three of the 17 articles reported data related to CTV delineation, and one reported on the PTV (Table 3). For both articles that compared CT image sets with PET/CT there was a decrease in mean volumes when metabolic imaging was used. The most noticeable variation was reported by Leclerc et al.\textsuperscript{40} in the CTV delineation, with a percentage of change of $-26.8\%$. In the same study, the reported percentage change for PTV volumes delineated with anatomical versus metabolic imaging decreased to $-24.1\%$. Cardoso et al.\textsuperscript{32} compared CT and MRI with PET/CT and DWI-MRI which resulted in an increase in CTV volume in all CTVs between anatomical and metabolic imaging. The highest percentage change was calculated between CTV-CT and CTV-CT/PET/MRI with a value of $+23.5\%$. The percentage change between CTV-CT and CTV-CT/PET/MRI/DWI-MRI was slightly smaller with a value of
| Study          | Sample size | Target volume evaluated | Anatomical imaging used          | Metabolic imaging used | Volume measurements for anatomical imaging<sup>a</sup> | Volume measurements for metabolic imaging<sup>a</sup> | % of change | Measured IOV<sup>b</sup> |
|---------------|-------------|-------------------------|---------------------------------|------------------------|------------------------------------------------------|------------------------------------------------------|-------------|------------------------|
| Guden et al.  | 14          | CT-MR (coregistered images) | PET-CT (coregistered images)    | CTVprimary             | 169.1 cm<sup>3</sup> (86–287.5 cm<sup>3</sup>) | 132.3 cm<sup>3</sup> (64.3–246.1 cm<sup>3</sup>) | −21.7%      | N/R                    |
| Leclerc et al. | 41          | CT                      | PET/CT                          | CTVprimary and PTVprimary | CTV–CT: 73.1 cm<sup>3</sup> (53.5 cm<sup>3</sup>) | PTV–CT: 124.7 cm<sup>3</sup> | CTV: −26.8% | PTV: −24.1% |
| Cardoso et al. | 10          | CT and MRI              | DWI-MRI and PET/CT              | CTVprimary             | CTV–CT: 37.87 cm<sup>3</sup> (30.91–44.84 cm<sup>3</sup>) | CTV–CTPET: 38.48 cm<sup>3</sup> (32.86–44.1 cm<sup>3</sup>) | CTV–CTPETMRI: 46.62 cm<sup>3</sup> (39.22–54.02 cm<sup>3</sup>) | CTV–CTPETMRI/DWI-MRI: 45.49 cm<sup>3</sup> (38.59–52.38 cm<sup>3</sup>) | CT vs. CTPET: +1.6% | CT vs. CTPETMRI: +23.5% | CT vs. CRPETMRI/DWI-MRI: +20.1% |

<sup>a</sup>Mean, standard deviation, median, range.  
<sup>b</sup>Metrics in interobserver variability.  
<sup>c</sup>Dice similarity coefficient.  
<sup>d</sup>Concordance index/conformity index.
+20.1%. All the volumes followed their original trend of decreasing or increasing with the use of metabolic imaging with only one exception. In the study by Cardoso et al. there was an increase in volume of 1.6% when comparing CTV-CT with CTV-PET/CT, while there was a decrease in volume of 3.7% when comparing GTV-CT with GTV-PET/CT. All the articles used a margin of 5 mm when growing the GTV to account for the microscopic tumor infiltration, yielding the CTV. As a comparison, Ligtenberg et al. reported reductions of 45%–52% in CTV delineations when modality-specific CTV margins were used based on “ground truth” pathology in locally advanced laryngohypopharyngeal cases. They noted that PET-based CTVs were significantly smaller compared to CT- and MRI-based CTVs. However, their standard CTV margin was 10 mm unlike the 5 mm reported above. Leclerc et al. used a margin of 4–5 mm to establish the final PTV. These margins are consistent with current guidelines used in clinical practice.

The percentage of change calculated in the study by Guden et al. decreased considerably after delineation margins were added. The original difference was calculated as a decrease in volume of 61%, between GTV-CT-MRI and GTV-PET-CT, while the decrease in volume between CTV-CT-MRI and CTV-PET-CT was only 21%. Grégoire et al. indicate that GTV to CTV expansions cannot be solely a geometric expansion and that modifications due to patients’ anatomy must be made. The same guidelines divide this modification into four categories: modification for air cavities, adaptation for the complex head and neck anatomy, adaptation for strong anatomic barriers preventing tumor cells from free diffusion, and adaptation to consider the experience gained from surgical series. Therefore, some anatomical structures might not be completely included in the CTV, which explains why the percentage of change is smaller when these margins are grown. In the study by Leclerc et al. the percentage of change between CTVs is also reduced although to a much smaller degree. The percentage of change between GTVs was 28.7% and was 26.8% between CTVs. A likely explanation for this is the locally advanced stage of disease of the included population. The cancer types presented in this study were oropharynx, oral cavity, hypopharynx and larynx, all of which incorporate large air cavities that are adapted when growing the GTV to CTV. However, since these patients have advanced disease, it is possible that the exclusion of nonaffected air cavities was reduced.

### 3.3.3 Conformity index

Cardoso et al. reported CI in the context of IOV with the highest conformity between observers reported when CT and PET coregistered images were used. The use of PET/CT increased IOV, with a CI reported value of 0.68. However, when adding MRI and DWI-MRI the conformity decreased marginally to 0.65.

### 3.4 Dosimetric evaluation

#### 3.4.1 Dose to 95% of the volume (D95%)

Four of the seventeen included articles incorporated some level of dosimetric evaluation of their reported changes in target volume delineation (Table 4). In two of these articles, CT and MRI coregistered image sets were used as the anatomical imaging techniques. In addition, two of these articles had PET/CT as the metabolic imaging used and the other a PET/MRI hybrid. As previously mentioned, there was also inconsistency observed in the reported metrics for these assessments, with the most used ones being dose to OAR and D95%. However, even the OARs and the D95% were not reported consistently. Some of the studies only mention dose to some OARs
| Study          | Sample size | Anatomical imaging used | Metabolic imaging used | Target volume evaluated                           | Dose to OARs<sup>a</sup>                                                                 |
|---------------|-------------|-------------------------|------------------------|----------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| Guden et al.  | 14          | CT-MR (coregistered images) | PET-CT                | GTVprimary and CTVprimary and GTVnodal             | Mean dose to parotids higher on CT-MR for 8 of the 14 patients                                                   |
| Leclerc et al.| 41          | CT                      | PET-CT                | GTVprimary and CTVprimary and PTVprimary           | Sequential planning:                                                                                               |
|               |             |                         |                       |                                                    | CT based:                                                                                                         |
|               |             |                         |                       |                                                    | • Ipsilateral parotid mean dose: 28.4 Gy                                                                         |
|               |             |                         |                       |                                                    | • Contralateral parotid mean dose: 23.2 Gy                                                                       |
|               |             |                         |                       |                                                    | • Oral cavity mean dose: 34.6 Gy                                                                               |
|               |             |                         |                       |                                                    | • Larynx D5%: 38.1 Gy                                                                                            |
|               |             |                         |                       |                                                    | • PRV spinal cord D2%: 27.4 Gy                                                                                  |
|               |             |                         |                       |                                                    | PET-based:                                                                                                        |
|               |             |                         |                       |                                                    | • Ipsilateral parotid mean dose: 27.8 Gy                                                                         |
|               |             |                         |                       |                                                    | • Contralateral parotid mean dose: 22.6 Gy                                                                       |
|               |             |                         |                       |                                                    | • Oral cavity mean dose: 32.7 Gy                                                                               |
|               |             |                         |                       |                                                    | • Larynx D5%: 37.9 Gy                                                                                            |
|               |             |                         |                       |                                                    | • PRV spinal cord D2%: 27.2 Gy                                                                                  |
|               |             |                         |                       |                                                    | SIB planning:                                                                                                    |
|               |             |                         |                       |                                                    | CT based:                                                                                                        |
|               |             |                         |                       |                                                    | • Ipsilateral parotid mean dose: 29.4 Gy                                                                         |
|               |             |                         |                       |                                                    | • Contralateral parotid mean dose: 23.6 Gy                                                                       |
|               |             |                         |                       |                                                    | • Oral cavity mean dose: 34.4 Gy                                                                               |
|               |             |                         |                       |                                                    | • Larynx D5%: 39.2 Gy                                                                                            |
|               |             |                         |                       |                                                    | • PRV spinal cord D2%: 28.7 Gy                                                                                  |
|               |             |                         |                       |                                                    | PET-based:                                                                                                        |
|               |             |                         |                       |                                                    | • Ipsilateral parotid mean dose: 28.4 Gy                                                                         |
|               |             |                         |                       |                                                    | • Contralateral parotid mean dose: 22.8 Gy                                                                       |
|               |             |                         |                       |                                                    | • Oral cavity mean dose: 32.6 Gy                                                                               |
|               |             |                         |                       |                                                    | • Larynx D5%: 38.9 Gy                                                                                            |
|               |             |                         |                       |                                                    | • PRV spinal cord D2%: 28.2 Gy                                                                                  |
| Wu et al.     | 20          | CT and MR               | PET                   | GTVprimary and PTVprimary                          | Dose constraints were met for all plans                                                                          |
|               |             |                         |                       |                                                    | PTV-CTMR: 71.1 Gy (SD: 2.1 Gy)                                                                                   |
|               |             |                         |                       |                                                    | PTV-CTMRPET: 68.6 Gy (SD: 2.6 Gy)                                                                                |
| Wang et al.   | 11          | CT                      | PET/MRI               | GTVprimary and GTVnodal                            | N/R                                                                                                              |
|               |             |                         |                       |                                                    | Mean GTVprim CT: 65.3 Gy                                                                                         |
|               |             |                         |                       |                                                    | Mean GTVprim-PET/MRI: 65.2 Gy                                                                                  |
|               |             |                         |                       |                                                    | Mean GTVnodal CT: 62.3 Gy                                                                                       |
|               |             |                         |                       |                                                    | Mean GTVnodal-PET/MRI: 62.3 Gy                                                                                  |

<sup>a</sup>Dose to organs at risk.

<sup>b</sup>Dose to 95% of the volume.
such as parotids or mention OARs as a whole. Some studies reported D95\% to PTV and others to GTV.

Two studies evaluated the impact of metabolic imaging on dose to target volumes. Wu et al.\textsuperscript{28} reported the D95\% as 71.1 Gy to PTV-CT-MRI and 68.6 Gy to PTV-CT-MRI-PET. Although their volumetric analysis had shown a limited change in GTV geometry, the dosimetric influence was more significant and most noticeable when evaluating PTV volumes. Additionally, when the same RT plan was applied to both PTVs the PTV resulting from the delineation with PET-CT was not fully covered by the 95\% isodose line. The authors theorize that the differences in PTV coverage might be explained by the exacerbation of the differences in GTV delineation between anatomical and metabolic image sets.\textsuperscript{28} This explains the results of Wang et al.,\textsuperscript{29} where the difference in D95\% was measured between GTVs and showed little variation. However, it should be noted that this study compares PET/MRI images with CT alone instead of PET/CT and CT images, and that for both GTV primary and GTV nodal PET/MRI images have led to an increase in target volumes. In addition, this study did not evaluate the impact of routine expansions used in treatment planning between GTV, CTV and PTV which could possibly have changed the outcome.

In summary, it is difficult to quantify the impact of metabolic imaging on target volume delineation for head and neck cancer. Multiple systems are used in the reporting of target volume metrics, with mostly volumetric assessments used. With respect to GTV primary, the use of PET/CT has been shown to decrease target volumes delineated for many of the included studies. However, the same trend was not definitively established when other forms of metabolic images were used.

The use of PET/CT in addition to CT images improved IOV. PET/CT is superior to CT alone in the recognition of lymph node involvement, which would lead to the hypotheses that GTV nodal volumes would be larger when these type of images are used in the delineation process.\textsuperscript{30} However, no trend was identified in this review that supports that hypothesis, leading to the conclusion that as PET/CT discriminates between involved and noninvolved nodal regions, this results in personalized nodal targets for each individual patient, not a particular trend towards larger nodal volumes than on anatomical images.

For all other target volumes, the percentage of change was less pronounced, possibly due to the added setup margins. The use of PET/CT also demonstrated an improvement in IOV in these cases. Using metabolic imaging in the delineation process may lead to a reduction in dose in OAR, but the evidence for head and neck cancer as reported in this review is limited at present.

### 3.4.2 Dose to organs at risk

Three of the four articles explored dosimetric impact of target volume delineation variability between anatomical and metabolic image on OARs. Guden et al.\textsuperscript{42} reported

| Study                      | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total |
|----------------------------|---|---|---|---|---|---|---|---|------|
| Igdem et al.\textsuperscript{2} | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 8    |
| Guden et al.\textsuperscript{42} | 2 | 2 | 2 | 2 | 0 | 1 | 2 | 0 | 11   |
| Delouya et al.\textsuperscript{34} | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 10   |
| Thiagarajan et al.\textsuperscript{35} | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 10   |
| Chatterjee et al.\textsuperscript{37} | 2 | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 9    |
| Venkata et al.\textsuperscript{45} | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 10   |
| Arslan et al.\textsuperscript{46} | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 8    |
| Anderson et al.\textsuperscript{31} | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 10   |
| Chauhan et al.\textsuperscript{25} | 2 | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 9    |
| Bird et al.\textsuperscript{27} | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 10   |
| Leclerc et al.\textsuperscript{40} | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 0 | 12   |
| Felice et al.\textsuperscript{22} | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 8    |
| Wu et al.\textsuperscript{28} | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 10   |
| Gudi et al.\textsuperscript{36} | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 10   |
| Wang et al.\textsuperscript{29} | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 9    |
| Samolyk-Kogaczewska et al.\textsuperscript{13} | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 8    |
| Cardoso et al.\textsuperscript{32} | 2 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 10   |

Note: The matrix for these scorings can be found in Table A1.
an increase in dose to the parotid glands in eight of 14 patients when anatomical imaging was used. Leclerc et al.\(^4\) compared sequential and simultaneous integrated boost (SIB) planning, reporting the doses to OARs with both planning approaches and comparing CT and PET/CT volumes. In both planning methods, the dose for all reported OARs was higher in the CT-based volume, with minimal changes being reported for the larynx and spinal cord planning organ at risk volume (PRV). In addition, in the article by Wu et al.\(^8\) the dose constraints were met for all OARs for all developed plans.

3.5 | Quality assessment of included studies

All included articles scored between 8 and 12 out of a maximum of 16, meaning that their quality is considered as medium. This is because in the matrix of MINORS there are two items regarding patients follow-up that scored 0 for 15 of the 17 studies. All included studies had a score that corresponded to at least 50% of the maximum possible score. The individual scores for each publication are given in Table 5.

3.6 | Limitations of this review

All included papers had a small number of patients (maximum \( n = 41 \)), as well as heterogeneous populations. Delineation guidelines were not stated in the papers and metrics reported were also heterogeneous. This limits the potential to evaluate the data. In addition, many included articles used PET/CT as the metabolic imaging option which does not allow meaningful comparisons with other types of metabolic imaging, such as PET/MRI or DWI-MRI.

4 | CONCLUSION

It can be concluded that molecular imaging impacts on target volume delineation, but inconsistently. The lack of homogeneity in the reporting metrics used in the studies included hinders a definitive conclusion on its impact as does a lack of detail on target delineation guidance adhered to. PET/CT showed a trend in decreased target volumes for GTV primary delineations as well as reducing IOV.

No apparent trend in the resultant changes in GTV nodal target volumes using metabolic imaging can be concluded.

A trend to decreased dose to OARs with the use of molecular imaging when delineating target volumes was found; however, the evidence was very limited.

The impact of molecular imaging on CTV and PTV volumes was not as pronounced as for GTV. This was due to the minimization of the imaging effect when GTV to CTV and CTV to PTV margins were added.

In summary, more homogeneous reporting practices, details on delineation method together with additional focus on other metabolic imaging modalities along with PET/CT would better evaluate its impact on target volume delineation.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Diana Raquel Dias Domingues \( \text{@https://orcid.org/0000-0001-7799-8428} \)

REFERENCES

1. Pfister DG, Spencer S, Adelstein D, et al. Head and neck cancers, version 2.2020, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2020;18(7):873-898.
2. Berthelsen AK, Dobbs J, Kjellen E, et al. What’s new in target volume definition for radiologists in ICRU Report 71? How can the ICRU volume definitions be integrated in clinical practice? Cancer Imaging. 2007;7:104-116.
3. Jalil ur Rehman, Zahra N, Khalid M, et al. Intensity modulated radiation therapy: a review of current practice and future outlooks. J Radiat Res Appl Sci. 2018;11:361-367.
4. Segedin B, Petric P. Uncertainties in target volume delineation in radiotherapy—Are they relevant and what can we do about them? Radiol Oncol. 2016;50:254-262.
5. Lin R. Target volume delineation and margins in the management of lung cancers in the era of image guided radiation therapy. J Med Radiat Sci. 2014;61(1):1-3.
6. Vesprini D, Ung Y, Dinniwell R, et al. Improving observer variability in target delineation for gastro-oesophageal cancer the role of 18F-fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography. Radiother Oncol. 2008;20:631-638.
7. Pereira GC, Traughber M, Muzic RF. The role of imaging in radiation therapy planning: past, present, and future. Biomed Res Int. 2014;2014:231090.
8. Bentzen SM, Gregoire V. Molecular imaging-based dose painting: a novel paradigm for radiation therapy prescription. Semin Radiat Oncol. 2011;21(2):101-110.
9. Berwouts D, Madani I, Duprez F, et al. Long-term outcome of 18 F-fluorodeoxyglucose-positron emission tomography-guided dose painting for head and neck cancer: matched case-control study. Head Neck. 2017;39(11):2264-2275.
10. Wang J, Zheng J, Tanig T, et al. A randomized pilot trial comparing position emission tomography (PET)-guided dose escalation radiotherapy to conventional radiotherapy in chemoradiotherapy treatment of locally advanced nasopharyngeal carcinoma. PLoS One. 2015;10(4):e0124018.
11. Pigorsch S U, Wilkens J, Kampfer S, et al. Do selective radiation dose escalation and tumour hypoxia status impact the
loco-regional tumour control after radio-chemotherapy of head & neck tumours? The ESCALOX protocol. Radiat Oncol. 2017;12(1):45.

12. Tang L, Wei F, Wu Y, et al. Role of metabolism in cancer cell radioresistance and radiosensitization methods. J Exp Clin Cancer Res. 2018;37(1):87.

13. Samołyk-Kogaczewska N, Sierko E, Zuzda K, et al. PET/MRI-guided GTV delineation during radiotherapy planning in patients with squamous cell carcinoma of the tongue. Strahlenther Onkol. 2019;195(9):780-791.

14. Delouya G, Igidbashian L, Houle A, et al. 18F-FDG-PET imaging in radiotherapy tumor volume delineation in treatment of head and neck cancer. Radiother Oncol. 2011;101(3):362-368.

15. Tol JP, Dahele M, Doornaert P, et al. Different treatment planning protocols can lead to large differences in organ at risk sparing. Radiother Oncol. 2014;113(2):267-271.

16. Xiao Y, Watson M. Guidance on conducting a systematic literature review. J Plan Educ Res. 2017;39:93-112.

17. Liberati A, Altman D, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.

18. Rawan Aliozzi XL, White J, Apte A, et al. Tools for consensus analysis of experts’ contours for radiotherapy structure definitions. Radiother Oncol. 2010;97:10.

19. Hanna GG, Hounsell AR, O’Sullivan JM. Geometrical analysis of radiotherapy target volume delineation: a systematic review of reported comparison methods. Clin Oncol. 2010;22:10.

20. Zhang J, Pang J, Yu J, Wang P. An efficient assembly retrieval method based on Hausdorff distance. Robot Comput Integr Manufact. 2018;51:103-111.

21. Juneja P, Evans PM, Harris EJ. The validation index: a new metric for validation of segmentation algorithms using two or more expert outlines with application to radiotherapy planning. IEEE Trans Med Imaging. 2013;32(8):1481-1489.

22. De Felice F, Musio D, Bulzonetti N, et al. Target volume delineation based on diffusion-weighted magnetic resonance imaging for locally advanced head and neck cancer. Anticancer Res. 2016;36(9):4181-4185.

23. Zheng Y, Sun X, Wang J, et al. FDG-PET/CT imaging for tumor staging and definition of tumor volumes in radiation treatment planning in non-small cell lung cancer. Oncol Lett. 2014;7:1015-1020.

24. Vees H, Casanova N, Zilli T, et al. Impact of 18F-FDG PET/CT on target volume delineation in recurrent or residual gynaecologic carcinoma. Radiat Oncol. 2012;7:176.

25. Chauhan D, Rawat S, Sharma M, et al. Improving the accuracy of target volume delineation by combined use of computed tomography, magnetic resonance imaging and positron emission tomography in head and neck carcinomas. J Cancer Res Ther. 2015;11(4):746-751.

26. Lydiatt WM, Patel S, O’Sullivan B, et al. Head and neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual. CA Cancer J Clin. 2017;67(2):122-137.

27. Bird D, Scarsbrook A, Sykes J, et al. Multimodality imaging with CT, MR and FDG-PET for radiotherapy target volume delineation in oropharyngeal squamous cell carcinoma. BMC Cancer. 2015;15(1):844.

28. Wu VVC, Leug W, Wong K, et al. The impact of positron emission tomography on primary tumour delineation and dosimetric outcome in intensity modulated radiotherapy of early T-stage nasopharyngeal carcinoma. Radiother Oncol. 2016;111(1):109.

29. Wang K, Leug W, Wong K, et al. Evaluation of PET/MRI for tumor volume delineation for head and neck cancer. Front Oncol. 2017;7:8.

30. Prestwich RJ, Sykes J, Carey B, Sen M, Dyker K, Scarsbrook A. Improving target definition for head and neck radiotherapy: a place for magnetic resonance imaging and 18-fluorodeoxyglucose positron emission tomography? Clin Oncol. 2012;24(8):577-589.

31. Anderson C.M., Sun W, Buatti J, et al. Interobserver and intermodality variability in GTV delineation on simulation CT, FDG-PET, and MR images of head and neck cancer. Jacobs J Radiat Oncol. 2014;1(1):6.

32. Cardoso M, Min M, Jameson M, et al. Evaluating diffusion-weighted magnetic resonance imaging for target volume delineation in head and neck radiotherapy. J Med Imaging Radiat Oncol. 2019;63(3):399-407.

33. Lewis-Jones H, Colley S, Gibson D. Imaging in head and neck cancer: United Kingdom National Multidisciplinary Guidelines. J Laryngol Otol. 2016;130(S2):S28-S31.

34. Jager EA, Litenberg H, Caldas-Magalhaes J, et al. Validated guidelines for tumor delineation on magnetic resonance imaging for laryngeal and hypopharyngeal cancer. Acta Oncol. 2016;55(11):1305-1312.

35. Thiagarajan A, Caria N, Schoder H, et al. Target volume delineation in oropharyngeal cancer: impact of PET, MRI, and physical examination. Int J Radiat Oncol Biol Phys. 2012;83(1):220-227.

36. Gudi S, Ghosh-Laskar S, Agarwal J, et al. Interobserver variability in the delineation of gross tumour volume and specified organs-at-risk during IMRT for head and neck cancers and the impact of FDG-PET/CT on such variability at the primary site. J Med Imaging Radiat Sci. 2017;48(2):184-192.

37. Chatterjee S, Frew J, Mott J, et al. Variation in radiotherapy target volume definition, dose to organs at risk and clinical target volumes using anatomic (computed tomography) versus combined anatomic and molecular imaging (positron emission tomography/computed tomography): intensity-modulated radiotherapy delivered using a tomotherapy Hi Art machine: final results of the VortigERN study. Clin Oncol (R Coll Radiol). 2012;24(10):e173-e179.

38. Mak D, Corry J, Lau E, Rischin D, Hicks R. Role of FDG-PET/CT in staging and follow-up of head and neck squamous cell carcinoma. Q J Nucl Med Mol Imaging. 2011;55(5):487-499.

39. Grégoire V, Evan M, Le Q, et al. Delineation of the primary tumour Clinical Target Volumes (CTV-P) in laryngeal, hypopharyngeal, oropharyngeal and oral cavity squamous cell carcinoma: AIRO, CACA, DAHANCA, EORTC, GEORCC, GORTEC, HKNPCSG, HNCIG, IAG-KHT, LPRHHT, NCIC CTG, NCRI, NRG Oncology, PHNS, SBRT, SOMERA, SRO, SSHNO, TROG consensus guidelines. Radiother Oncol. 2018;126(1):3-24.

40. Leclerc M., Lartiguerie T, Daisne J, Kramar A, Grégoire V. Primary tumor delineation based on 18FDG PET for locally advanced head and neck cancer treated by chemoradiotherapy. Radiother Oncol. 2015;116(1):87-93.
41. Ligtenberg H, Jager E, Caldas-Magalhaes J, et al. Modality-specific target definition for laryngeal and hypopharyngeal cancer on FDG-PET, CT and MRI. *Radiother Oncol*. 2017;123(1): 63-70.

42. Guden M, Ceylan C, Berberoğlu K, et al. Contribution of PET-CT to staging, gross tumour volume definition, planning and response assessment in IMRT for nasopharyngeal carcinoma. *J Radiother Pract*. 2010;10(4):272-282.

43. Rosa C, Pizzi A, Augurio A, et al. Volume delineation in cervical cancer with T2 and diffusion-weighted MRI: agreement on volumes between observers. *In Vivo*. 2020;34(4):1981-1986.

44. Igdem, S, Alco G, Recan T, et al. The application of positron emission tomography/computed tomography in radiation treatment planning: effect on gross target volume definition and treatment management. *Clin Oncol (R Coll Radiol)*. 2010;22(3): 173-178.

45. Venkada, M.G., Rawat S, Choudhury P, et al. A quantitative comparison of gross tumour volumes delineated on [18F]-FDG PET-CT scan and CECT scan in head and neck cancers. *Indian J Nucl Med*. 2012;27(2):95-100.

46. Arslan, S., Abakay C, Sen F, et al. Role of PET/CT in treatment planning for head and neck cancer patients undergoing definitive radiotherapy. *Asian Pac J Cancer Prev*. 2014;15(24):10899-10903.

**How to cite this article:** Dias Domingues DR, Leech MM. Exploring the impact of metabolic imaging in head and neck cancer treatment. *Head & Neck*. 2022;44(10):2228-2247. doi:10.1002/hed.27131
APPENDIX A

TABLE A1 The revised and validated version of MINORS.
The items are scored 0 (not reported), 1 (reported but inadequate), or 2 (reported and adequate). The global ideal score being 16 for noncomparative studies and 24 for comparative studies.

| Methodological items for nonrandomized studies | Score |
|------------------------------------------------|-------|
| 1. A clearly stated aim: The question addressed should be precise and relevant in the light of available literature | |
| 2. Inclusion of consecutive patients: All patients potentially fit for inclusion (satisfying the criteria for inclusion) have been included in the study during the study period (no exclusion or details about the reasons for exclusion) | |
| 3. Prospective collection of data: Data were collected according to a protocol established before the beginning of the study | |
| 4. Endpoints appropriate to the aim of the study: Unambiguous explanation of the criteria used to evaluate the main outcome which should be in accordance with the question addressed by the study. Also, the endpoints should be assessed on an intention-to-treat basis | |
| 5. Unbiased assessment of the study endpoint: Blind evaluation of objective endpoints and double-blind evaluation of subjective endpoints. Otherwise the reasons for not blinding should be stated | |
| 6. Follow-up period appropriate to the aim of the study: The follow-up should be sufficiently long to allow the assessment of the main endpoint and possible adverse events | |
| 7. Loss to follow-up less than 5%: All patients should be included in the follow up. Otherwise, the proportion lost to follow up should not exceed the proportion experiencing the major endpoint | |
| 8. Prospective calculation of the study size: Information of the size of detectable difference of interest with a calculation of 95% confidence interval, according to the expected incidence of the outcome event, and information about the level for statistical significance and estimates of power when comparing the outcomes | |

Additional criteria in the case of comparative study

| Methodological items for comparative studies | Score |
|---------------------------------------------|-------|
| 9. An adequate control group: Having a gold standard diagnostic test or therapeutic intervention recognized as the optimal intervention according to the available published data | |
| 10. Contemporary groups: Control and studied group should be managed during the same time period (no historical comparison) | |
| 11. Baseline equivalence of groups: The groups should be similar regarding the criteria other than the studied endpoints. Absence of confounding factors that could bias the interpretation of the results | |
| 12. Adequate statistical analyses: Whether the statistics were in accordance with the type of study with calculation of confidence intervals or relative risk | |