Introduction

Head and neck cancer (HNC) encompasses multitudinous malignancies, including laryngocarcinoma, oral carcinoma and nasopharyngeal carcinoma (1). Tobacco and alcohol abuse and human papillomavirus (HPV) infection are two primary causes of HNC. HNC was commonly prevalent in several countries around the world, with an estimated prevalence of HNC in 4% of malignancies and more frequently in China in approximately 4.8–6.1% of malignancies (2,3). The clinical management of HNC requires precise diagnosis, accurate staging and a standardized multidisciplinary treatment approach. The HNC stage at diagnosis predicts prognosis and guides treatment decisions; American Joint Committee on Cancer TNM staging classification was used as a basis for clinical oncology practice. However, due to various tumor sites and specific anatomic regions of HNC, tumor staging is not...
sufficient and usually generates error (4). Therefore, many concerns have been raised regarding the weakness of T staging for several HNC sites.

Recently, several studies implied that anatomic primary tumor volumes on CT or MRI and primary gross tumor volume (GTVp) play a meaningful predictive role for HNC long time prognosis. Strongin et al. conducted a study including 78 participants with stage III–IV oropharyngeal, laryngeal or hypopharyngeal cancer; patients with GTVp <35 mL had significantly improved local control (71% vs. 41%), as well as the long-term survival, like progression free survival (61% vs. 33%), and overall survival (OS) (84% vs. 41%) rates (5). Those small retrospective studies implied that anatomic tumor volumes on imaging tests have prognostic value for HNC patients (6). With advancements in imaging technologies, positron emission tomography-computed tomography (PET-CT) has been used in HNC diagnosis and clinical staging, although routine PET-CT scan for staging of cancer remains controversial. However, PET-CT is helpful in accurately evaluating distant metastasis and therapy response by tumor metabolic and morphological changes (7,8). Currently, in many studies using PET-CT to calculate primary tumor volume, named primary tumor metabolic tumor volume (MTVp), MTVp was defined as tumor volume with elevated fluorodeoxyglucose (FDG) uptake (7,9). Several studies also suggested that MTVp as a novel parameter in PET-CT has potential to become a standardized prognostic indicator (10,11).

From the literature, it appears that tumor volume (GTVp and MTVp) is a significant factor associated with outcomes in HNC. However, the prognosis role of tumor volume (GTVp and MTVp) in HNC is controversial and limited according to the previous investigations. This comprehensive meta-analysis was conducted to accurate evaluate of the relationship between tumor volume (GTVp and MTVp) and the prognosis of HNC.

**Methods**

**Literature search**

Initially, the systematic literature search was performed by using the electronic databases PubMed, ISI Web of Science, and EBSCO up to January 2016. This search strategy utilized the following selected common tags: “HEAD AND NECK CANCER” and “TUMOR VOLUMES” and “PROGNOSIS”. The language was limited to English, and unpublished results and abstracts were not included. Hand search was also conducted according to reference lists of relevant articles and reviews for additional relevant researches.

**Eligibility criteria**

Eligible studies include randomized controlled studies or observational studies that reported estimates of the associations between primary gross tumor volume (GTVp or MTVp) and HNC prognosis. Included studies in this analysis should meet all of the following criteria: (I) articles evaluating the association between tumor volume (GTVp or MTVp) and HNC parameters such as clinicopathological characteristic and long-term prognosis (OS or DFS) of HNC; (II) articles containing sufficient information to estimate of hazard ratios (HRs) and 95% confidence intervals (CIs); and (III) English papers and full text available. Studies were excluded based on the following: (I) duplicate publications; (II) reviews, editorials, clinical guidelines, case reports, comments or letters and meeting abstract; (III) lacking relevant data.

**Data extraction and quality assessment**

Relevant data were extracted from each eligible study in a standardized manner: first author’s name; publication year and country; average age of participants; duration of follow-up; sample size; tumor volume (GTVp or MTVp) detection methods and cutoff scores; tumor stage; tumor location; and local control rate, five years OS and disease-free survival (DFS) rates. Data extraction conflicts were resolved through discussion. The cutoff value for the tumor volume (GTVp or MTVp) varied among the included articles; we a defined large tumor volume (GTVp or MTVp) value according to receiver operating characteristic (ROC) curve analysis from original articles. Newcastle-Ottawa Quality Assessment Scale (NOS) was adopted to assess included studies quality, when scores greater than six were judged as high-quality studies.

**Statistical analysis**

Statistical analysis was conducted to calculate the pooled HRs with 95% CI to evaluate tumor volume (GTVp or MTVp) predict prognosis in HNC. Analysis endpoints for survival outcomes including local control rate, OS and DFS. Q test and I² test were used to assess the heterogeneity of
included studies. The DerSimonian-Laird method (random-effects model) was applied when statistical heterogeneity was found ($I^2 > 50\%$ or $P<0.10$). While, Mantel-Haenszel method (fixed-effects model) was considered when $P>0.1$ and $I^2 < 50\%$. Funnel plots, Begg’s and Egger’s tests were performed to assess the publication bias on the reported outcomes. All statistical analyses were following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline. And performed with STATA software (Version 12.0, Stata Corporation, College Station, TX, http://www stata.com).

Results

Eligible studies characteristics

Initially, 517 potential articles were retrieved utilizing electronic database searches above. Those potentially relevant articles were screened for eligibility by duplication and language, and 266 studies were excluded. Two hundred and twenty-four records were excluded after titles and abstracts screened. Finally, 27 articles were reviewed. Due to insufficient of survival data, we removed 12 papers. Finally, 15 observational retrospective studies were included.

Figure 1 Flow diagram of study identification. 517 potential articles were acquired from PubMed, ISI Web of Science and EBSCO. After screened for eligibility by duplication and language, 266 studies were excluded. 224 records were excluded after titles and abstracts screened. Finally, 27 articles were reviewed. Due to insufficient of survival data, we removed 12 papers. Finally, 15 observational retrospective studies were included.

Evaluation of the GTVp

The detection and definition method of tumor volume...
### Table 1 Characteristics of the included studies

<table>
<thead>
<tr>
<th>Source</th>
<th>Year</th>
<th>Country</th>
<th>Cancer location</th>
<th>Study</th>
<th>Number (M/F)</th>
<th>Age (y)</th>
<th>Follow-up (m)</th>
<th>Median volume</th>
<th>Survival data</th>
<th>NOS scorer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timmermans (12)</td>
<td>2016</td>
<td>Netherlands</td>
<td>LC (T3–T4)</td>
<td>Retrospective</td>
<td>166 (124/42)</td>
<td>61.9±11.3</td>
<td>37 (13.8–74)</td>
<td>11.6 (5.7–21.3)</td>
<td>Local control, OS, DFS</td>
<td>7</td>
</tr>
<tr>
<td>Wu (16)</td>
<td>2013</td>
<td>China</td>
<td>NPC</td>
<td>Retrospective</td>
<td>305 (241/64)</td>
<td>43 [13–78]</td>
<td>78 [6–128]</td>
<td>16.47 (1.04–146.86)</td>
<td>Local control, OS, DFS</td>
<td>8</td>
</tr>
<tr>
<td>Noij (13)</td>
<td>2015</td>
<td>Netherlands</td>
<td>HNSCC</td>
<td>Retrospective</td>
<td>78 (45/33)</td>
<td>–</td>
<td>18 [9–25]</td>
<td>7.3 (2.7–14.6)</td>
<td>DFS</td>
<td>6</td>
</tr>
<tr>
<td>Guo (18)</td>
<td>2012</td>
<td>China</td>
<td>NPC</td>
<td>Retrospective</td>
<td>694 (534/160)</td>
<td>43 [13–78]</td>
<td>59.1 (3.1–104.9)</td>
<td>29.1 (4.6–92.3)</td>
<td>Local control, OS, DFS</td>
<td>8</td>
</tr>
<tr>
<td>Feng (17)</td>
<td>2013</td>
<td>China</td>
<td>NPC</td>
<td>Retrospective</td>
<td>363 (285/78)</td>
<td>48 [16–78]</td>
<td>63 [9–82]</td>
<td>31.5 (1.8–112.4)</td>
<td>Local control, OS, DFS</td>
<td>7</td>
</tr>
<tr>
<td>Strongin (5)</td>
<td>2012</td>
<td>USA</td>
<td>HNSCC (III–IV)</td>
<td>Retrospective</td>
<td>78 (57/11)</td>
<td>57.5 [38–82]</td>
<td>22.6 (2.4–65.9)</td>
<td>38.7 (3.1–123.2)</td>
<td>OS, DFS</td>
<td>6</td>
</tr>
<tr>
<td>Hou (14)</td>
<td>2015</td>
<td>USA</td>
<td>HNSCC (LA)</td>
<td>Retrospective</td>
<td>30 (23/7)</td>
<td>59.8 [48–79]</td>
<td>–</td>
<td>34.7±24.2</td>
<td>Local control, OS</td>
<td>6</td>
</tr>
<tr>
<td>Janssens (15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>2014</td>
<td>Netherlands</td>
<td>LC (cT2–4)</td>
<td>Retrospective</td>
<td>137 (109/28)</td>
<td>44 [2–84]</td>
<td>3.6 (0.4–57.7)</td>
<td></td>
<td>Local control</td>
<td>7</td>
</tr>
<tr>
<td>ARCON</td>
<td>2014</td>
<td>Netherlands</td>
<td>LC (cT2–4)</td>
<td>Retrospective</td>
<td>133 (108/25)</td>
<td>59 [18–84]</td>
<td>3.5 (91.0–65.2)</td>
<td></td>
<td>Local control</td>
<td>7</td>
</tr>
<tr>
<td>Romesser (10)</td>
<td>2012</td>
<td>USA</td>
<td>HNSCC</td>
<td>Retrospective</td>
<td>41 (31/10)</td>
<td>58.9±10.3</td>
<td>24.2 (2.7–56.3)</td>
<td>GTVp: 22.2 (1.5–162.5)</td>
<td>Local control, OS, DFS</td>
<td>6</td>
</tr>
<tr>
<td>Seol (11)</td>
<td>2010</td>
<td>Korea</td>
<td>HNSCC</td>
<td>Retrospective</td>
<td>59 (56/3)</td>
<td>65 [47–81]</td>
<td>–</td>
<td>MTV: 7.2 (0.40–43.5)</td>
<td>Local control, OS, DFS</td>
<td>6</td>
</tr>
<tr>
<td>Chen (20)</td>
<td>2009</td>
<td>Tai Wan</td>
<td>NPC</td>
<td>Retrospective</td>
<td>76 (74/2)</td>
<td>57 [36–79]</td>
<td>37 [13–95]</td>
<td>33.4 (3.8–152.4)</td>
<td>Local control, DFS</td>
<td>6</td>
</tr>
<tr>
<td>Chang (23)</td>
<td>2003</td>
<td>Tai Wan</td>
<td>HNSCC</td>
<td>Retrospective</td>
<td>52 (32/20)</td>
<td>46 [16–75]</td>
<td>62.2 (2.8–84.4)</td>
<td>19.15 (3.23–131.82)</td>
<td>DFS</td>
<td>6</td>
</tr>
<tr>
<td>Kurek (22)</td>
<td>2003</td>
<td>Germany</td>
<td>NPC</td>
<td>Retrospective</td>
<td>107 (100/7)</td>
<td>61 [34–87]</td>
<td>43</td>
<td>32.5 (2.1–220.1)</td>
<td>OS</td>
<td>7</td>
</tr>
<tr>
<td>Chung (19)</td>
<td>2009</td>
<td>Korea</td>
<td>NPC</td>
<td>Retrospective</td>
<td>82 (69/13)</td>
<td>53.8 [11–70]</td>
<td>–</td>
<td>39 (1.9–231.9)</td>
<td>Local control, OS, DFS</td>
<td>7</td>
</tr>
</tbody>
</table>
(GTVp or MTVp) used in the included studies were imaging techniques. Among the 15 studies, two articles used PET-CT, whereas computed tomography (CT) and/or magnetic resonance imaging (MRI) was used in 12 studies, and one study used both CT and PET. The cut-off value for larger tumor volume depended on ROC curve analysis, the median cutoff value of GTVp was 24.48 mL (range: 8.38 to 48 mL).

**GTVp predicts local control**

Nine studies provided the association between GTVp and HNC local control, including two studies on laryngeal cancer (LC), three studies on nasopharyngeal cancer (NPC), and four articles on squamous cell carcinoma. The combined analysis from the published data showed that larger GTVp was significantly associated with worse local control rate (pooled HR =1.34, 95% CI =1.22–1.47). Next, we found that heterogeneity existed (I²=65%, P=0.002) (Figure 2A). Furthermore, the subgroup analysis was conducted according to location of HNC, two studies including three datasets were about LC demonstrated that larger GTVp was correlated with poor local control (pooled HR =1.21, 95% CI=1.03–1.42), NPC (pooled HR =1.53,
95% CI =1.22–1.93) and HNSCC (pooled HR =1.37, 95% CI =1.21–1.55) (Figure 2B).

GTVp implies OS

There were ten studies assessing GTVp and OS of HNC. The pooled analysis found that larger GTVp could predict a worse OS rate (pooled HR =1.27, 95% CI =1.17–1.37). Next, we found that heterogeneity existed ($I^2=75\%$, $P=0.000$) (Figure 3A). Subgroup meta-analysis was performed by the location, four studies on NPC demonstrated that larger GTVp was correlated with unfavorable OS (pooled HR =1.34, 95% CI =1.13–1.58), LC (pooled HR =1.19, 95% CI =1.00–1.42) and HNSCC (pooled HR =1.77, 95% CI =1.14–1.41) (Figure 3B).

GTVp and DFS

Ten included articles assessed the relationship between GTVp and DFS of HNC. The pooled analysis found that a larger GTVp was significantly associated with a poorer DFS rate (pooled HR =1.34, 95% CI =1.23–1.46). $I^2=78.8\%$ indicted heterogeneity was existed ($P=0.000$) (Figure 4A). Subgroup meta-analysis was conducted according to HNC.
location, three studies on NPC demonstrated that a larger GTVp was correlated with a poorer DFS (pooled HR =1.28, 95% CI =1.08–1.52), LC (pooled HR =1.27, 95% CI =1.01–1.60) and HNSCC (pooled HR =1.39, 95% CI =1.24–1.55) (Figure 4B).

**MTVp in HNC**

Three studies assessed the MTVp and HNC. The combined analysis found that a larger MTVp was significantly associated with a poorer local control (pooled HR =3.51, 95% CI =1.87–6.59). For longtime survival, a larger MTVp also correlated with a poorer OS (pooled HR =4.13, 95% CI =2.63–6.47), and an unfavorable DFS (pooled HR =5.70, 95% CI =4.30–7.5) (Figure 5).

**Publication bias**

Begg’s rank correlation and Egger’s weighted regression methods were used to investigate the publication bias of included studies. Analysis results indicted no significant publication bias was found in this meta-analysis (data not

---

**Figure 4** Forest plots showing results of studies on the impact of primary tumor volume (GTVp) of disease free survival (DFS). (A) Impact of GTVp on DFS of patients with head and neck cancer; (B) subgroup analysis of GTVp on DFS of patients with different location of head and neck cancer.
shown). However, publication bias might exist and difficult to confirm because of the limited studies included.

**Discussion**

The present systematic literature review and meta-analysis studied the published data to examine the tumor volumes (GTVp or MTVp) on predict prognosis of HNC and included fifty studies with 2,447 patients. This comprehensive analysis incorporated the most recent data, used both random and fixed effects models, and yielded the evidence that GTVp on CT/MRI (GTVp) is a marker of poor prognosis. These results not only observed on local control, but also OS and DFS. Similarly, tumor volume also could be assessed by PET-CT, named MTVp. The pooled analysis implied that a larger MTVp also indicted a worse prognosis.

According to National Comprehensive Cancer Network (NCCN 2016) guideline, the T classification plays a major role in HNC treatment decisions (1). However, T classification was dependent on clinical examination, laryngoscopy, CT or MRI, which was not sufficient to predict survival in clinical practice. The clinical T stage of HNC is not representative of tumor burden; many T stages have a wide range of tumor burden, but guidelines lack of further nonsurgical sub-grouping for clinical stages, helping oncologists to select therapy methods and to predict survival outcomes. In addition, basic studies indicated that HNC with large volumes usually have more clonogenic tumor cells, tumor volume could be an ideal surrogate for tumor
burden. A larger volume indicates a greater tumor burden and requires larger doses of radiation or chemoradiotherapy to acquire radical cure (16). Previous studies demonstrated that volumetric staging system was superior to TNM stage system in correlating long-term survival in any site of HNC undergoing definitive intensity modulation radiated therapy (IMRT) (24). Furthermore, Studer’s retrospective research adopted the volumetric stratification to divide clinical T4 HNC into four volume subgroups (1 to 15 vs. 16 to 70 vs. 71 to 130 vs. >130 mL); they found that the OS was 90%/72%/58%/18% for V1/2/3/4, respectively. Those results implied that volumetric staging could potentially predict outcomes for different volume sizes in HNC (25).

Recently, evidence regarding the prognostic impact of tumor volume and as a supplement for T stage in HNC was increasing. However, the exact relationship was not well established, and tumor volume was not being used for clinical practice. There are many reasons why it is hard to draw a final conclusion. First, due to malignancy, tumors might be more aggressive and finally grow into irregularly shapes. Therefore, the accurate measurement of the tumor volume is difficult under conventional imaging examinations (e.g., CT and MRI). Second, the cut-off value of tumor volume was difficult to define; among the eligible studies in this meta-analysis cut-off values were different (median cutoff value was 24.48 mL, range from 7.3 to 48 mL). Various assessment software and different treatment methods might account for cutoff value biases (26,27). Furthermore, all the studies were retrospective studies; no prospective randomized trials performed to verify the prognostic impact of tumor volume on HNC. The therapy methods adopted in the included studies were radiotherapy or chemoradiotherapy, but the tumor stages and location varied. The unavoidable differences in baseline characteristics in each study might also explain the controversial results in association between tumor volume and HNC survival.

Furthermore, studies indicated that total HNC volume is a major prognostic factor, it mainly impacted by hypoxic volume, while not by non-hypoxic volume (8). Advancements in imaging technologies have significant implications for the use of measured tumor volumes, especially for molecular imaging (e.g., PET-CT). MTVp appears to be a trend to instead of GTVp. PET-CT could accurately assess the metabolic tumor volume in HNC not influenced by irregularly shaped tumors, and MTVp measurements do not need special software and was acquired easier than GTVp. Furthermore, in our analysis, the results revealed that a small advantage of MTVp (OS: HR =4.13, 95% CI =2.63–6.47) was that it is better than GTVp (OS: HR =1.27, 95% CI =1.17–1.37) as a prognosis factor. Therefore, MTVp is an exciting parameter as metabolic information provided by the PET-CT scan, which could help more accurately measure tumor volume.

Limitations might remain in this study. Firstly, the method and cut-off values for assessing GTVp and MTVp are inconsistent. Secondly, some heterogeneity observed across studies could not be completely accounted for despite the use of appropriate meta-analytic techniques with random-effect models. Finally, because this is a literature-based analysis, inevitable could cause publication bias due to several negative results not been published.

Conclusions

Results from this analysis demonstrated that tumor volume (GTVp or MTVp) was a useful prognostic marker, and HNC patients with larger tumor volumes are associated with unfavorable local control and long-term survival. It is recommended that tumor volume should be included in a new staging system combined with TNM classification to provide oncologists predict long-term survival. Without doubt, further studies to assess the standard measurement and improve new imaging technologies are urgently needed.

Acknowledgements

Funding: This work was supported by the Nature science foundation of Zhejiang Province (Grant No. LY12H13002), project of Zhejiang Traditional Chinese Medicine Administration (Grant No. 2012ZZ012), project of Zhejiang medical health science plan (Grant No. 2015RCB026), project of Zhejiang traditional Chinese medicine research (Grant No. 2013ZB115, 2015ZB107) and Nature science foundation of Ningbo City (Grant No. 2012A610205, 2012A610206), Science and Technology foundation of Cixi (CN2013013).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

2013;189:867-73.
