NASOPHARYNGEAL CANCERS: WHICH METHOD SHOULD BE USED TO MEASURE THESE IRREGULARLY SHAPED TUMORS ON CROSS-SECTIONAL IMAGING?


Departments of *Diagnostic Radiology and Organ Imaging and †Clinical Oncology, and ‡Centre for Clinical Trials, Faculty of Medicine, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong S.A.R., China

Purpose: To determine whether the standard techniques of measuring tumor size and change in size after treatment could be applied to the measurement of nasopharyngeal cancers, which are often irregular in shape.

Methods and Materials: The standard measurements of bidimensional (BDM) (World Health Organization criteria) and unidimensional (UDM) (Response Evaluation Criteria in Solid Tumors [RECIST] criteria), together with the maximum depth of the tumor perpendicular to the pharyngeal wall (DM), were acquired from axial magnetic resonance images of primary nasopharyngeal carcinoma in 44 patients at diagnosis and in 29 of these patients after treatment. Tumor volume measurements (VM), acquired from the summation of areas from the axial magnetic resonance images, were used as the reference standard.

Results: There was a significant association between VM and BDM with respect to tumor size at diagnosis (p = 0.002), absolute change in tumor size after treatment (p < 0.001), and percentage change in tumor size after treatment (p = 0.044), but not between VM and UDM. There was also a significant association between VM and DM with respect to percentage change in tumor size after treatment (p = 0.0001) but not absolute change (p = 0.222).

Conclusion: When using simple measurements to assess irregularly shaped nasopharyngeal cancers, the BDM should be used to measure size at diagnosis and the BDM and percentage change in size with treatment. Unidimensional measurement does not reflect size or change in size, and therefore the RECIST criteria may not be applicable to all tumor shapes. The use of DM requires further evaluation. © 2007 Elsevier Inc.

Magnetic resonance imaging, Tumor measurement, Head and neck, Pharyngeal carcinoma.

INTRODUCTION

Cancers of the head and neck arising from the upper aerodigestive tract, such as those in the nasopharynx, are often irregular in shape. They spread along the walls of a hollow organ and frequently infiltrate extensively into the adjacent soft tissues. Their irregular shape (Fig. 1) causes difficulty for the oncologist and radiologist when it comes to determining the best method to assess tumor size and response to therapy.

Although accurate cancer measurement may not always be required in the routine clinical setting, it is important for clinical trial researchers. There is a pressing need to establish a method for measurement of irregularly shaped tumors in the head and neck for use in the increasing number of clinical trials of patients with advanced-stage or recurrent cancer. Where possible, measurement of the primary tumor should be included in trial assessments. In many patients this may be the only site of disease. At present, because of the confusion in methods of measurement, some studies exclude the primary cancer, whereas others use bidimensional measurements (BDM) from the World Health Organization (WHO) guidelines (1, 2), but even the ability of this latter measurement to reflect the size of irregular tumors has not been examined. Furthermore, the increasing use of unidimensional measurements (UDM) from the Response Evaluation Criteria in Solid Tumors (RECIST) guidelines (3) to measure regular-shaped neck nodes, could increase the confusion over which technique to use when the primary tumor is measured on the same scan. This is a difficult and controversial area, but there is an important need to explore...
methods of tumor measurement on which the results of clinical trials are based.

The hypothesis for this study was that standard techniques for measuring tumor size could not be applied to the irregularly shaped nasopharyngeal cancer. The aim of this study was to use magnetic resonance (MR) imaging to determine whether BDM used in the WHO guidelines and UDM used in the RECIST guidelines would reflect tumor size at diagnosis or the change in size after treatment. A new measurement, the maximum depth of the tumor projecting into the pharynx taken perpendicular to the pharyngeal wall (DM), was evaluated also with respect to change in tumor size with treatment. Tumor volume measurements (VM) obtained from the summation of areas from cross-sectional MR images were used as the reference standard. The study aimed only to address the problem of how to measure this irregularly shaped cancer; it did not set out to determine the clinical outcome of the group of patients selected in this study.

METHODS AND MATERIALS

Patient selection

Patients undergoing conventional MR imaging of the head and neck for the staging of previously untreated nasopharyngeal carcinoma (NPC) were included in this retrospective study. The patients in this study were either consecutive patients with advanced-stage disease (Stage III or IV) who were enrolled in a neoadjuvant chemotherapy study (Group A) with informed consent, or consecutive patients with early T stage disease (T1 or T2) (Group B) who underwent MR scanning as part of the routine staging assessment. One patient was excluded because MR was unable to delineate a primary tumor from a metastatic lateral retropharyngeal node.

Fig. 1. T1-weighted postcontrast magnetic resonance image in the coronal plane of a 39-year-old woman, demonstrating the irregular shape of the nasopharyngeal carcinoma, which is invading the nasopharynx (small arrows) with invasion into the left parapharyngeal region (arrowhead), skull base and sphenoid sinus (long arrows), and cranium (curved arrow).

MR technique and measurements

The examinations were performed on a 1.5-T whole-body MR imaging system (Gyroscan ACS-NT; Philips, Best, The Netherlands) with a 23 mT/m maximum gradient capability. A standard volume head-and-neck coil was used to obtain T1-weighted spin-echo images before and after contrast and T2-weighted turbo spin-echo images with and without fat saturation. Images were obtained in at least two planes. Measurements were made on the axial T1-weighted postcontrast images (repetition time of 500 ms, echo time of 20 ms, 22-cm field of view, 4-mm slice thickness with no interslice gap) using a 512 × 512 matrix after a bolus injection of 0.1 mmol/kg gadolinium dimeglumine (Schering, Berlin, Germany).

No single sequence or plane provides all the information required to accurately depict the extent of the tumor; therefore, all sequences were examined by the radiologist to determine the full extent of the tumor. Measurements then were made electronically on the axial T1-weighted postcontrast images displayed on a monitor using the Philips Viewforum 3.2. The whole tumor, including any portion invading the bony skull base, was included in the measurements. The VM was obtained by manually outlining the area of the tumor on each axial image and adding these areas together and multiplying by the slice thickness (Fig. 2a). The UDM was obtained by measuring the maximum diameter of the tumor, and BDM was obtained by multiplying the UDM by the greatest measurement perpendicular to the UDM (Fig. 2b). The aforementioned measurements were made by a head-and-neck radiologist with more than 9 years experience of MR imaging of NPC. Measurements from the same set of scans were repeated after 2 months to assess intraexaminer reliability.

Assessment of tumor response

Patients in Group A with advanced-stage disease (overall Stage III or IV) were included in this part of the study because these patients represent the group that is recruited for clinical trials. In this study patients in Group A were randomized into two further groups, the first group received neoadjuvant chemotherapy followed by chemoradiotherapy (Group A1), and the second group received chemoradiotherapy alone (Group A2). The assessment of tumor response was performed after neoadjuvant chemotherapy in the Group A1 and after chemoradiotherapy in Group A2.

For the purpose of assessment of tumor response, the tumor size at diagnosis was recalculated after excluding any areas of tumor invasion into the skull base. For VM areas of tumor invasion into the main bony structures on the axial images (i.e., into the clivus, petrous apices, or pterygoid processes) were excluded; these areas were also excluded from measurement on the posttreatment scans. For UDM and BDM the measurements were placed to avoid traversing bony structures. In addition, one other measurement technique was used, whereby the maximum depth of the tumor projecting into the pharynx taken perpendicular to the pharyngeal...
Fig. 2. Magnetic resonance (MR) images of the 39-year-old woman with a primary nasopharyngeal carcinoma shown in Fig. 1. Measurements were made electronically on the axial T1-weighted postcontrast images, taking into account the extent of tumor on all other sequences and planes performed in the MR examination. (a) T1-weighted postcontrast MR sequential images in the axial plane. The volume measurement was obtained by manually outlining the area of the tumor on each axial image and adding these areas together and multiplying by the slice thickness. (b) T1-weighted postcontrast
wall (DM) (Fig. 2c) was measured before and after treatment. This technique was chosen because for many years, as well as in some routine clinical practices today, assessment of the primary tumor response has been based on the response of the nasopharyngeal intraluminal component as judged by endoscopy with or without biopsy rather than imaging.

**Statistical methods**

A general linear model with stepwise model building procedure was used to assess whether UDM and BDM were associated with VM. A \( p \) value of \(<0.05\) was considered significant. Intrarater reliability was measured using the intraclass correlation coefficient. To have an 80% power of detecting an intraclass correlation coefficient of 0.4 using a two-sided 5% level test with two raters, we required at least 25 subjects in the study (4).

Response of each group was calculated using the absolute change and the percentage change in the measurements. Response rate, for the established techniques using VM, UDM, and BDM, was categorized into complete response, partial response, stable disease, and progressive disease according to the criteria set out in Table 1. The comparisons among response as measured by the different approaches were made using the \( \kappa \) statistic and McNemar’s test to assess the degree of agreement and discordance in response classification, respectively.

**RESULTS**

**Patient characteristics**

The study group comprised 44 patients (33 men, 11 women; mean age, 48.3 years) with the following T stages: T1 \( (n = 14)\), T2 \( (n = 8)\), T3 \( (n = 16)\), T4 \( (n = 6)\). Twenty-nine of these patients with advanced-stage disease were enrolled in a study to assess neoadjuvant chemotherapy (Group A) (mean age, 49 years; 22 male, 7 female). Seventeen of these patients underwent MR imaging after neoadjuvant chemotherapy (Group A1) and 12 after chemoradiotherapy (Group A2).

**Measurement of tumor size at diagnosis**

Tumor volumes ranged from 3590 to 116,813 mm\(^3\), with a mean of 27,371 mm\(^3\), standard deviation of 20,701 mm\(^3\), and median of 21,566 mm\(^3\). The association of VM with BDM and UDM is shown in Table 2. Bidimensional measurement was significantly associated with VM at diagnosis and significantly associated with absolute change and percentage change in VM. For UDM there was no significant association. The intrarater reliability for VM, BDM, and UDM is shown in Table 3. The intraclass correlation coefficients for VM were very high, and the intraclass correlation coefficients for BDM were in general higher than those of UDM, especially for the early-stage NPC patients.

**FIG. 2.** (Cont’d)

MR image in the axial plane. The unidimensional measurement (UDM) was obtained by measuring the maximum diameter of the tumor (arrows) that includes tumor invasion into the left pterygoid process and pterygopalatine fossa of the skull base, and the bidimensional measurement was obtained by multiplying the UDM by the greatest measurement perpendicular (arrowhead) to the UDM. (c) T1-weighted postcontrast MR image in the axial plane. The depth measurement was obtained by measuring the maximum depth of the tumor perpendicular to the pharyngeal wall (arrowhead).
**Assessment of tumor response**

The association of VM with BDM, UDM, and DM for absolute change in size after treatment and percentage change in size after treatment is shown in Table 2. Bidimensional measurement was significantly associated with both absolute and percentage change in size, whereas DM was associated with the percentage change in size.

In terms of response according to VM, there were 4 complete responses (CR), 16 partial responses (PR), and 8 stable diseases (SD). According to UDM (RECIST criteria) there were 4 CR, 12 PR, and 12 SD, with a κ of 0.42 (95% confidence interval [CI], 0.11–0.72). Taking VM results as the reference standard, UDM misclassified 7 PR cases as SD and misclassified 3 SD as PR (p = 0.1714, McNemar’s test). For BDM (WHO criteria) there were 4 CR, 13 PR, and 11 SD, with a κ of 0.47 (95% CI, 0.16–0.77). Taking VM results as the reference standard, BDM misclassified 6 PR as SD and misclassified 3 SD as PR (p = 0.2525, McNemar’s test). Using VM as the reference standard, the use of BDM and UDM misclassified the response in 9 of 28 cases (32%) and 10 of 28 cases (36%), respectively.

**DISCUSSION**

Volume methods using the summation of areas from cross-sectional images of the tumor outlined on a computer workstation provide the most accurate representation of tumor size. They exclude normal structures and do not make presumptions on the shape of the tumor (5). However, this technique is time consuming, and the technology, expertise, and manpower are not available for routine clinical practice or the assessment of multicenter drug trials. New semiautomated techniques for measuring tumors are being assessed in an attempt to replace the manual tracing method (6, 7), but they remain relatively labor intensive and are not widely available. Therefore, at present the simple methods of measurement prevail. The most widely adopted of these methods uses the WHO guidelines (1, 2) and is based on measuring the tumor in two dimensions (BDM). More recently the emphasis has started to shift toward UDM, which are used in the RECIST guidelines (3). Unidimensional measurements have been shown to be sufficient for assessing both tumor size (8, 9) and change in size of solid tumors (10, 11). However, many of the studies on tumor measurement are based on the measurement of metastatic cancer, which frequently produces spherical or oval lesions rather than irregularly shaped lesions. Although extremely irregularly shaped tumors, such as lymphangitis carcinomatosis, are considered immeasurable (3), little attention has been given to the problem of assessing measurable irregularly shaped tumors, such as primary NPCs that spread along the walls of a hollow viscus.

The first part of the present study evaluated UDM and BDM for measuring tumor size at diagnosis. The results confirmed our hypothesis that UDM would not reflect the volume of the tumor, but surprisingly BDM showed a significant association with tumor volume. This association remained even when the measurements were repeated to exclude any areas of bony tumor invasion. In the case of NPC, the size of the primary tumor has been shown to be an independent prognostic factor of local control (12), and these results suggest that if size is incorporated into the T staging for NPC, BDM can be used in the staging process.

The second part of the study evaluated techniques to measure change in tumor size and to categorize tumor response. Change in tumor volume is important not only for clinical management but also for assessing the efficacy of new drugs, whereby tumor response on imaging is used to expedite the study and reduce the number of patients. However, studies with smaller numbers of patients require even more accurate and reproducible methods of measuring size. One of the problems in measuring change in the size of a pharyngeal cancer is related to the difficulty of measuring the component of the cancer that spreads along the pharyngeal walls. As the tumor shrinks, the depth of the tumor projecting into the pharyngeal cavity may decrease, whereas the extent along the length of the wall may remain the same. Therefore UDM and BDM may be expected to underestimate the true tumor response. It was therefore of interest to determine whether the use of DM would overcome this problem and reflect the total tumor response.

A further problem in the measurement of tumor response arises when there is tumor invasion into bone. Nasopharyngeal carcinoma has a propensity to invade the skull base. The posttreatment MR scan of these patients usually displays persistent abnormalities in the bone marrow, although ultimately most do not prove to have residual disease at this site. Therefore, including the bony skull base would have produced an underestimate of the clinical response. It was for this reason that areas of bony tumor invasion were excluded from the pretreatment (by re-measuring the cancer at diagnosis) and posttreatment measurements in this study. In addition, NPC often invades the paranasal sinuses and cranium to produce significant volumes of tumor both below and above the complex-shaped bones that form the skull base (Fig. 1). In such cases, UDM or BDM measurements could be expected to show even less association with VM.

Despite all the difficulties discussed above, a significant association between the BDM and VM remained for both absolute and percentage change in tumor size. This result suggests that BDM, which is a quicker and more widely applicable method than VM, is sufficient for assessment of tumor response.

Once again there was no significant association between UDM and VM, bringing into question the application of UDM and the RECIST criteria to all tumors’ shapes. This finding has the potential to influence the measurement not only of nasopharyngeal tumors but also other irregularly shaped tumors, including other pharyngeal cancers and primaries at other sites, such as cerebral gliomas.

There was no association between the absolute change in...
Disease progression results highlight the impact that the different methods of dance in response classification when compared with dance. In terms of the degree of agreement and discordance, although there was no significant difference in discordance, the agreement showed a slightly better agreement than UDM, with BDM showing a better agreement than VM. Bidimensional measurements. Discrepancies arose in assignment to the parapharyngeal region, rarely leaves a mass of scar tissue. However, there was an association between VM and DM for percentage change in tumor size. This finding is difficult to explain, although it is interesting that for many years it was only this intraluminal component that could be assessed clinically. Therefore, further evaluation of the role of DM in tumor response is warranted, especially with respect to patient outcome.

The response category, using either UDM or BDM compared with VM, was altered in approximately one third of patients. A similar observation has been made by Hopper et al. (13) and Prasad et al. (14) when comparing volumetric techniques with uni- and bidimensional measurements. Discrepancies arose in assignment to the partial response and static disease categories, influencing the number of responders and nonresponders to treatment. Both UDM and BDM decreased the number of responders when compared with VM. Bidimensional measurement showed a slightly better agreement than UDM, although there was no significant difference in discordance. In terms of the degree of agreement and discordance in response classification when compared with VM, there was no difference in these two methods. These results highlight the impact that the different methods of measurement have on treatment response, and clearly this factor has to be taken into account when comparing the results of drug trials.

The main limitation of this study was that the reference standard for tumor size was volume measurement on MR rather than volume of the resected specimen, because surgery is not the primary treatment for NPC. The use of volume measurements as the reference standard may be criticized because of the reported poor reproducibility of this technique when measuring poorly defined and irregularly shaped tumors (15, 16). It was for this reason that we evaluated the intraexaminer intraclass correlation coefficient in this study. Despite the irregular shape of NPC, the reproducibility of VM was actually very high across all T stages. Reproducibility also is considered to be one of the advantages of using the simple methods of measurement, such as UDM and BDM, because they entail less observer variation than tumor volume estimates. However, in this study the reproducibility of BDM and UDM was lower than that for VM, with BDM showing an advantage over UDM for all T stages. These results probably reflect the difficulty in correctly identifying the largest measurements for UDM and BDM because of the greater variation in possible sites for measurement in irregularly shaped tumors.

A second limitation of the study was that we only assessed the measurements for NPC, and these measurements were made in the axial plane. Although NPC is an irregularly shaped tumor, it is possible that there are tumors that are even more irregular. In such tumors, even the use of BDM may not be sufficient to reflect true tumor size. Furthermore, with the more widespread use of multidetector CT and MR, the images are no longer restricted to the axial plane. The effect of using an alternative plane to measure UDM or BDM is, to the authors’ knowledge, unknown.

It should be stressed that the focus of this study was “how to measure” in terms of measurement methods. For the assessment of treatment response, the next step is “what to measure.” This is a controversial topic and one that is usually overlooked. Exploration of this topic is beyond the scope of this study, but we have tried to partly address the problems by defining the structures that we measured on the pre- and posttreatment scans.

### Table 1. Tumor measurements and treatment response criteria

<table>
<thead>
<tr>
<th>Category</th>
<th>Unidimensional criteria</th>
<th>Bidimensional criteria</th>
<th>Volumetric criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete response</td>
<td>Tumor disappearance</td>
<td>Tumor disappearance</td>
<td>Tumor disappearance</td>
</tr>
<tr>
<td>Partial response</td>
<td>&gt;30% reduction in size</td>
<td>&gt;50% reduction in size</td>
<td>&gt;65% reduction in size</td>
</tr>
<tr>
<td>Stable disease</td>
<td>Size between that for partial response and progressive disease</td>
<td>Size between that for partial response and progressive disease</td>
<td>Size between that for partial response and progressive disease</td>
</tr>
<tr>
<td>Disease progression</td>
<td>&gt;20% increase in size</td>
<td>&gt;25% increase in size</td>
<td>&gt;40% increase in size</td>
</tr>
</tbody>
</table>

* According to Response Evaluation Criteria in Solid Tumors (RECIST) guidelines (3).
† According to World Health Organization (WHO) guidelines (1, 2).
‡ Based on relationship of change in area to change in volume (using the criteria from WHO guidelines).

VM and DM. The latter result is not surprising given that the general impression of clinicians and researchers is that the nasopharyngeal intraluminal component responds more quickly to therapy and, unlike tumor deep in the soft tissues of the parapharyngeal region, rarely leaves a mass of scar tissue. However, there was an association between VM and DM for percentage change in tumor size. This finding is difficult to explain, although it is interesting that for many years it was only this intraluminal component that could be assessed clinically. Therefore, further evaluation of the role of DM in tumor response is warranted, especially with respect to patient outcome.

The response category, using either UDM or BDM compared with VM, was altered in approximately one third of patients. A similar observation has been made by Hopper et al. (13) and Prasad et al. (14) when comparing volumetric techniques with uni- and bidimensional measurements. Discrepancies arose in assignment to the partial response and static disease categories, influencing the number of responders and nonresponders to treatment. Both UDM and BDM decreased the number of responders when compared with VM. Bidimensional measurement showed a slightly better agreement than UDM, although there was no significant difference in discordance. In terms of the degree of agreement and discordance in response classification when compared with VM, there was no difference in these two methods. These results highlight the impact that the different methods of measurement have on treatment response, and clearly this factor has to be taken into account when comparing the results of drug trials.

The main limitation of this study was that the reference standard for tumor size was volume measurement on MR rather than volume of the resected specimen, because surgery is not the primary treatment for NPC. The use of volume measurements as the reference standard may be criticized because of the reported poor reproducibility of this technique when measuring poorly defined and irregularly shaped tumors (15, 16). It was for this reason that we evaluated the intraexaminer intraclass correlation coefficient in this study. Despite the irregular shape of NPC, the reproducibility of VM was actually very high across all T stages. Reproducibility also is considered to be one of the advantages of using the simple methods of measurement, such as UDM and BDM, because they entail less observer variation than tumor volume estimates. However, in this study the reproducibility of BDM and UDM was lower than that for VM, with BDM showing an advantage over UDM for all T stages. These results probably reflect the difficulty in correctly identifying the largest measurements for UDM and BDM because of the greater variation in possible sites for measurement in irregularly shaped tumors.

A second limitation of the study was that we only assessed the measurements for NPC, and these measurements were made in the axial plane. Although NPC is an irregularly shaped tumor, it is possible that there are tumors that are even more irregular. In such tumors, even the use of BDM may not be sufficient to reflect true tumor size. Furthermore, with the more widespread use of multidetector CT and MR, the images are no longer restricted to the axial plane. The effect of using an alternative plane to measure UDM or BDM is, to the authors’ knowledge, unknown.

It should be stressed that the focus of this study was “how to measure” in terms of measurement methods. For the assessment of treatment response, the next step is “what to measure.” This is a controversial topic and one that is usually overlooked. Exploration of this topic is beyond the scope of this study, but we have tried to partly address the problems by defining the structures that we measured on the pre- and posttreatment scans.

### Table 2. Probability (p) values for association of UDM, BDM, and DM with VM

<table>
<thead>
<tr>
<th></th>
<th>UDM</th>
<th>BDM</th>
<th>DM</th>
</tr>
</thead>
<tbody>
<tr>
<td>VM at diagnosis (including areas of bone invasion)</td>
<td>0.273</td>
<td>0.0002</td>
<td>NA</td>
</tr>
<tr>
<td>VM at diagnosis (excluding areas of bone invasion)</td>
<td>0.5856</td>
<td>0.0005</td>
<td>NA</td>
</tr>
<tr>
<td>Absolute change in VM after treatment</td>
<td>0.2034</td>
<td>&lt;0.0001</td>
<td>0.2221</td>
</tr>
<tr>
<td>Percentage change in VM after treatment</td>
<td>0.8620</td>
<td>0.0438</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

* Abbreviations: UDM = unidimensional measurement; BDM = bidimensional measurement; DM = depth measurement; VM = volume measurement; NA = not applicable.
In conclusion, when using simple measurements to assess irregularly shaped nasopharyngeal tumors, the BDM should be used to measure size at diagnosis and response to treatment. Unidimensional measurement does not reflect tumor size or change in tumor size, and therefore the RECIST criteria may not be applicable to all tumor shapes. This finding could have implications for the measurement of other irregularly shaped tumors. Finally, DM is a quick and easy measurement, but its use requires further evaluation.

**REFERENCES**