The following publication Lam, J., Ying, M., Cheung, S. Y., Yeung, K. H., Yu, P. H., Cheng, H. C., & Ahuja, A. T. (2016). A comparison of the diagnostic accuracy and reliability of subjective grading and computer-aided assessment of intranodal vascularity in differentiating metastatic and reactive cervical lymphadenopathy. Ultraschall in der Medizin-European Journal of Ultrasound, 37(01), 63-67 is available at https://doi.org/10.1055/s-0034-1384939.

A comparison of the diagnostic accuracy and reliability of subjective grading and computer-aided assessment of intranodal vascularity in differentiating metastatic and reactive cervical lymphadenopathy.

Introduction

Evaluation of cervical lymph nodes is important for head and neck cancer patients because it assesses patient prognosis and aids treatment planning. In a patient with proven head and neck cancer, the presence of a metastatic cervical lymph node decreases the 5 year survival rate by 50% and the survival rate reduces to 25% if bilateral metastatic lymph nodes are found in the neck [1].

Clinical examination of cervical lymph nodes by neck palpation is not a reliable method in differentiating benign and malignant cervical lymphadenopathy. It has been previously reported that ultrasound is more sensitive than neck palpation in the assessment of cervical lymphadenopathy [2-4]. Grey scale ultrasound assesses the morphology of cervical nodes, whereas colour or power Doppler ultrasound evaluates intranodal vascularity. Power Doppler ultrasound assessment of intranodal vascular pattern has been found to be useful in distinguishing benign and malignant nodes. The presence of peripheral vascularity, regardless of the presence or absence of hilar vascularity, is highly suggestive of malignancy [4,5]. In addition, due to tumour angiogensis and the associated proliferation of blood vessels, malignant lymph nodes tend to be more vascular than benign nodes [6,7]. Therefore, intranodal vascular abundance, which is commonly expressed as vascularity index, could be a useful parameter to differentiate benign and malignant nodes [6].

The intranodal vascular abundance can be assessed qualitatively or quantitatively. For qualitative assessment, the intranodal vascularity was subjectively graded according to its degree of vascular abundance within the lymph node [8]. For quantitative assessment, the intranodal vascularity was quantified with the use of an image processing program [9]. However, literature is devoid of any comparison of the diagnostic accuracy of the two techniques in differentiating benign and malignant lymph nodes. Therefore, this study was undertaken to evaluate and compare the diagnostic accuracy of the subjective grading and computer-aided approach in differentiating benign and malignant lymph nodes, to determine optimal cut-off value of the two assessment approaches in the differential diagnosis, and to compare the reliability of the two approaches in the assessment of intranodal vascularity of cervical lymph nodes.

Materials and Methods

A retrospective review of power Doppler sonograms of 99 patients with palpable cervical lymph nodes was conducted. The patients were selected consecutively from our patient archive. Patients who did not have fine-needle aspiration cytology (FNAC) of the lymph nodes were excluded. The study was approved by the Human Subject Ethics Subcommittee of the Department of Health Technology and Informatics, the Hong Kong Polytechnic University. Among these 99 patients, 64 patients had known carcinoma of the head and neck or other regions (nasopharyngeal carcinoma, 28; papillary carcinoma of the thyroid, 12; oral cavity carcinomas, 11; pharyngeal and laryngeal carcinomas, 8; lung carcinoma, 5), and FNAC-proven metastatic cervical lymph nodes. The remaining 35 patients had no known carcinoma in the head and neck or any other regions, and FNAC-proven reactive cervical lymph nodes. These 35 patients had clinical follow-ups in the out-patient department and remained well. The cervical lymph nodes of these 99 patients were examined by the same operator using the same ultrasound scanning protocol. In each patient, the cervical lymph nodes were evaluated with grey scale and power Doppler ultrasound, and the lymph node that showed the most abundant vascularity was included in the study. In the circumstance that the lymph node included in the study was not the node aspirated for FNAC, lymph node in the same level that showed similar grey scale and Doppler ultrasound features (such as round in shape, absence of echogenic hilus and presence of peripheral vascularity which are common in metastatic nodes) with the aspirated node was selected. Grey scale and power Doppler ultrasound examinations were performed on Philips HDI 5000 ultrasound unit using 5-12MHz linear transducer (Bothell, WA, USA). Settings of the power Doppler ultrasound were standardized and set for high sensitivity with a low wall filter for detecting blood vessels with low blood flow velocity. Pulsed repetition frequency (PRF) was set at 700Hz and medium persistence was adopted. The colour gain was standardized and initially increased to show colour noise and then decreased until the noise disappeared [8].

For each lymph node, multiple power Dopper sonograms at different scan planes were acquired during ultrasound scanning. Since the degree of intranodal vascularity varies with the scan plane, it was assessed in the scan plane that showed the most abundant vascularity. The degree of intranodal vascularity was evaluated by two approaches: 1. a qualitative subjective grading (QSG) method; 2. a quantitative computer-aided (QCA) method.

In the QSG method, the degree of intranodal vascularity was subjective graded and classified into 5 categories: 1. Grade 0 – no vascular signal was found within the lymph node; 2. Grade 1 – the area of the vascular signals was less than one fourth of the total cross-sectional area of the node (Fig. 1); 3. Grade 2 – the area of the vascular signals was between one fourth and half of the total cross-sectional area of the node (Fig. 2); 4. Grade 3 – the area of the vascular signals was between half and three fourths of the total cross-sectional area of the node (Fig. 3); 5. Grade 4 – the area of the vascular signals was greater than three fourths of the total cross-sectional area of the node (Fig. 4).

In the QCA method, the degree of intranodal vascularity was evaluated using the software program Matlab® (version 7.3.0.267 R2006b; The MathWorks, Natick, MA, USA) and a customized algorithm for colour signal quantification of Doppler images

[10]. In the evaluation of the intranodal vascularity, the power Doppler ultrasound images retrieved from the ultrasound unit were converted into tagged image file format. The images were then processed on a computer workstation with the Matlab and Microsoft Paint (version 5.1; Microsoft Corporation, Redmond, WA, USA). With the use of the Microsoft Paint, the boundaries of the lymph node (i.e. region of interest, ROI) were manually outlined on the ultrasound image. The image with the ROI outlined was then analyzed with the Matlab. Using the customized algorithm, the ROI was initially extracted from the ultrasound image and the total number of pixels of the ROI was evaluated by the algorithm. Afterwards, the colour pixels coded by power Doppler ultrasound were extracted from the ROI, and the number of colour pixels was counted (Fig. 5). The vascularity index (VI) of the lymph node was then calculated by the following equation:

$$\text{VI (\%)} = \frac{\text{Number of colour pixels within the ROI}}{\text{Total number of pixels within the ROI}} \times 100$$

To evaluate the inter-rater variability, the vascularity of the lymph nodes was evaluated by 5 raters independently using both the qualitative and quantitative methods. All raters were blinded to the result of other raters. To ensure the image viewing condition was standardized for the 5 raters, curtains were used to cover the viewing cubicle and the ambient lighting was consistent by standardizing the adjustment of the light dimmer. The setting of the brightness and contrast of the computer monitor was kept constant for the image review and analysis. To evaluate the intra-rater variability, the 5 raters reviewed and analysed the images two weeks after the first review. The same image review and analysis protocols, and viewing conditions were used.

Assuming lower intranodal vascularity indicated reactive lymph nodes and higher intranodal vascularity represented metastatic lymph nodes, the receiver-operating characteristic (ROC) curves of the QSG and QCA methods were plotted, and the performance of the two methods in distinguishing reactive and metastatic nodes was evaluated and compared using the area under the curve (AUC). The AUC of the ROC curves was calculated using MedCalc (version12.5.0, MedCalc Software, Ostend, Belgium). The ROC curves were also used to determine the optimum cut-off of the vascularity grading and VI. The point of the curve nearest to the top left-hand corner corresponded to the cut-off that had high sensitivity and specificity in distinguishing reactive and metastatic lymph nodes. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and overall accuracy of the two methods in the differential diagnosis were also evaluated and compared.

The inter-rater and intra-rater reliability of the QSG method were evaluated by Kappa test, whilst those of the QCA method were calculated by intraclass correlation coefficient (ICC). Kappa and ICC tests were performed using the software PASW Statistic 18 (SPSS Inc, Chicago, IL, USA). For the Kappa test, the inter-rater and intra-rater reliability were interpreted according to the previously described criteria: poor to fair agreement (Kappa value < 0.4), moderate agreement (Kappa value = 0.4-0.6), substantial agreement (Kappa value = 0.61-0.8), excellent agreement (Kappa value > 0.8) [11]. For the ICC, there are no agreed interpretation values to interpret the reliability. However, it is generally agreed that ICC values greater than 0.75 indicate good reliability, whereas those smaller than 0.75 indicate poor to moderate reliability [12]. Unpaired t-test and chi square test were used to calculate the significance of difference in the VI and vascularity grading between metastatic and reactive lymph nodes respectively.

Results

From the result of the 5 raters, the mean VI of metastatic nodes (43.1 ± 19.4%) was significantly higher than that of reactive nodes (26.1 ± 10.2%) ($p < 0.05$). There was a significant difference in the grading of the intranodal vascularity between metastatic and reactive nodes (grade 1, 28% and 47%; grade 2, 37% and 39%; grade 3, 23% and 13%; grade 4, 12% and 1% respectively) ($p < 0.05$). None of the metastatic or reactive nodes was classified as grade 0. To compare the diagnostic accuracy of the two methods, the AUC of the ROC curves of the QCA method (mean ± 1 standard deviation, SD: 0.77 ± 0.03; range: 0.77-0.81) was significantly greater than that of the QSG method (mean ± 1 SD: 0.63 ± 0.03; range: 0.58-0.67) ($p < 0.05$). For the QCA method, the mean optimum cut off of VI in distinguishing metastatic and reactive nodes was 32%, ranging from 28.7-33.1% for the 5 raters. For the QSG method, the optimum cut off grading was grade 2 for three raters and grade 3 for two raters.

Using the optimum cut off values, the sensitivity, specificity, PPV, NPV and overall accuracy of the two methods were calculated. From the result of the 5 raters, the mean sensitivity, specificity, PPV, NPV and overall accuracy of QCA (67.8%, 73.3%, 73%, 54% and 72.8% respectively) were higher than those of QSG (61.9%, 57.3%, 69%, 41.6% and 60.6% respectively).

To evaluate and compare the assessment reliability of the 2 methods, the inter-rater and intra-rater reliability of QCA and QSG were calculated. Kappa statistics was used to test for the inter- and intra-rater reliability of subjective grading approach. For the inter-rater reliability test of QSG, the Kappa values of various combinations of paired comparison of the raters ranged from 0.19-0.55, which indicates a poor to moderate agreement between the results of raters [11]. For the intra-rater reliability test of QSG, the Kappa values of the 5 raters ranged from 0.53-0.69, which indicates moderate to substantial level of agreement [11].

For the inter-rater reliability test of QCA, the ICC values of various combinations of paired comparison of the raters ranged from 0.83-0.96 indicating good inter-rater reliability [12]. For the intra-rater reliability test of QCA, the ICC value of the 5 raters ranged from 0.97-0.99 which indicates good intra-rater reliability [12].

Discussion

It has reported that more than 90% of head and neck tumours are squamous cell carcinomas (SCC) in the western population [13]. However, nasopharyngeal carcinoma (NPC) is common in the Chinese population, particularly for people in the southern parts of China, whilst NPC is less common in the western population. As the patients recruited in the present study were Chinese in Hong Kong, it accounts for the larger number of NPC patients (28/64) and relatively fewer head and neck SCC patients (19/64). Result of the present study found that metastatic nodes showed higher intranodal vascularity than reactive nodes, which is consistent with the finding of a previous report [6]. The higher intranodal vascularity of metastatic nodes is probably due to the tumour angiogenesis within the lymph node. In the present study, the mean AUC of the QCA was significantly greater than that of the QSG. Moreover, the sensitivity, specificity, PPV, NPV and overall accuracy of the QCA were higher than those of QSG. These findings showed that using the QCA approach in assessing the intranodal vascularity was more accurate than using the QSG approach in distinguishing benign and metastatic cervical lymphadenopathy. In addition, results showed that the intra-rater and inter-rater reliability of QCA method was higher than that of the QSG method, and this suggested that the QCA method was more reliable and reproducible than QSG in assessing the degree of intranodal vascularity. The higher diagnostic accuracy and reliability of QCA in the assessment of intranodal vascularity was due to the fact that QCA approach could provide an objective and quantitative assessment of the intranodal vascularity, whereas QSG approach evaluated the lymph nodes subjectively. Compared with human subjective perception, computer programs accurately depict subtle changes in vascularity that cannot be distinguished by human perception. Moreover, after prolonged reading of images, reviewer fatigue sets in and his ability in assessing the degree of intranodal vascularity may decrease. Therefore, the QCA method has a higher diagnostic accuracy and reliability than the QSG method in the assessment of intranodal vascularity.

The result of the present study was consistent with that of previous studies which investigated the use of quantitative approach in assessing the vascularity of different organs. QCA approach has been used in the assessment of thyroid vascularity, and was suggested to be an objective method in evaluating intrathyroidal vascularity [10]. Moreover, Bankier et al. [14] compared subjective grading and objective quantification approaches in the diagnoses of pulmonary emphysema in CT. They found that the objective quantification approach was more reliable, and suggested that subjective grading of the images should be supplemented with the information provided from the objective quantification. In addition, it has been reported that in the evaluation of the vascularity of plantar fascia in patients with plantar fasciitis QCA method has high reliability and strong correlation with the QSG approach [15].

In the present study, the optimum cut off of VI in distinguishing metastatic and reactive lymph nodes was 32% which yielded an overall accuracy of 72.8%. A previous study reported an accuracy of 78% when using the cut off of 9% [6]. The lower optimum cut off in Wu et al. [6] is probably because of the use of high wall filter in the study which results in low levels of Doppler signals within the lymph nodes. Conversely, a low wall filter was used in the present study and thus a higher optimum cut off was found. Another study reported a sensitivity and specificity of

70% and 70% respectively for detecting metastatic nodes when using a cut off of 30.6% [7]. Similar result was found in the present study in which a sensitivity of 67.8% and specificity of 73.3% were obtained when using a cut off of 32%.

The QCA method applied in the present study used an algorithm that required manual outline of the ROI for calculating the VI of the cervical lymph nodes. In our experience, the time for loading the image from the ultrasound unit to the remote workstation as well as for the image processing of manual outlining of a lymph node and of quantifying its intranodal vascularity takes about 15-20 minutes. Further development of an automated system may reduce the image processing time and be easily accepted for routine clinical practice.

Other than the degree of intranodal vascularity, grey scale ultrasound appearance and vascular pattern of lymph nodes are also useful sonographic features in the assessment of cervical lymphadenopathy. Since the grey scale ultrasound features and vascular pattern of metastatic and reactive lymph nodes were already well documented [2,4,16], and their assessments have to be performed subjectively, we focused on the evaluation of the degree of intranodal vascularity in the present study. In routine clinical practice, assessment of the grey scale ultrasound features and vascular pattern of lymph nodes is essential. The evaluation of the degree of intranodal vascularity provides additional information and is useful in patients where the ultrasound finding is equivocal.

In the present study, the presence of intranodal necrosis (including cystic and coagulation necrosis) was not an exclusion criterion. However, in each patient the lymph node demonstrated the most abundant vascularity was chosen, and thus there was a potential exclusion of necrotic metastatic nodes which tend to have absent or lower intranodal vascularity. In the study, if more necrotic metastatic nodes were included in the analyses, it may yield a lower cut off of VI in distinguishing metastatic and reactive lymph nodes. Nevertheless, one must note that the degree of intranodal vascularity should not be used as the sole criterion in distinguishing metastatic and reactive nodes, and it should be supplemented with the information from grey scale ultrasound. In patients with a known primary cancer, lymph nodes with intranodal necrosis regardless of the degree of intranodal vascularity are highly suspicious of metastases.

In this study, if the lymph node included in the study was not the aspirated node for FNAC, lymph node in the same level that showed similar grey scale and Doppler ultrasound features with the aspirated node was selected. In our experience, lymph nodes showed similar grey scale and Doppler ultrasound appearance usually have the same cytological characteristics.

In conclusion, QCA assessment of intranodal vascularity is possible and has higher reliability and accuracy than QSG method. The optimum cut off of the VI in differentiating metastatic and reactive lymph nodes was 32%. However, QCA assessment requires post-processing of images and thus the examination time is longer. The development of an automated system would enhance the feasibility of using QCA method in routine clinical practice.

References

1. Som PM. Detection of metastasis in cervical lymph nodes: CT and MR criteria and differential diagnosis. Am J Roentgenol 1992;158:961-969
2. Adibelli ZH, Unal G, Gul E, et al. Differentiation of benign and malignant cervical lymph nodes: value of B-mode and color Doppler sonography. Eur J Radiol 1998;28:230-234
3. Haberal I, Celik H, Gocmen H, et al. Which is important in the evaluation of metastatic lymph nodes in head and neck cancer: palpation, ultrasonography, or computed tomography? Otolaryngol Head Neck Surg 2004;130:197-201
4. Giacomini CP, Jeffrey RB, Shin LK. Ultrasonographic evaluation of malignant and normal cervical lymph nodes. Semin Ultrasound CT MR 2013;34:236-247
5. Ahuja AT, Ying M, Ho SS, Metreweli C. Distribution of intranodal vessels in differentiating benign from metastatic neck nodes. Clin Radiol 2001;56:197-201.
6. Wu CH, Chang YL, Hsu WC, et al. Usefulness of Doppler spectral analysis and power Doppler sonography in the differentiation of cervical lymphadenopathies. Am J Roentgenol 1998;171:503-509
7. Kagawa T, Yuasa K, Fukunari F, Shiraishi T, Miwa K. Quantitative evaluation of

vascularity within cervical lymph nodes using Doppler ultrasound in patients with oral cancer: relation to lymph node size. Dentomaxillofac Radiol 2011;40:415-421

8. Ying M, Ahuja A, Brook F, Metreweli C. Power Doppler sonography of normal cervical lymph nodes. J Ultrasound Med 2000;19:511-517

9. Wu CH, Hsu MM, Chang YL, Hsieh FJ. Vascular pathology of malignant cervical lymphadenopathy: qualitative and quantitative assessment with power Doppler ultrasound. Cancer 1998;83:1189-1196

10. Ying M, Ng DK, Yung DM, Lee ES. A semi-quantitative approach to compare high sensitivity power Doppler sonography and conventional power Doppler sonography in the assessment of thyroid vascularity. Thyroid 2009;19:1265-1269

11. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174

12. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. Am J Phys Med Rehabil 2005;84:719-723

13. Sanderson RJ, Ironside JA. Squamous cell carcinomas of the head and neck. BMJ 2002;325:822-827

14. Bankier AA, De Maertelaer V, Keyzer C, Gevenois PA. Pulmonary emphysema:

subjective visual grading versus objective quantification with macroscopic morphometry and thin-section CT densitometry. Radiology 1999;211:851-858

15. Chen H, Ho HM, Ying M, Fu SN. Correlation between computerised findings and Newman's scaling on vascularity using power Doppler ultrasonography imaging and its predictive value in patients with plantar fasciitis. Br J Radiol 2012;85:925-929

16. Ahuja AT, Ying M, Ho SY, et al. Ultrasound of malignant cervical lymph nodes. Cancer Imaging 2008;8:48-56