Interobserver Reliability of Computed Tomography-Derived Primary Tumor Volume Measurement in Patients with Supraglottic Carcinoma

Suresh K. Mukherji, M.D.[1,2]
Alicia Y. Toledano, Sc.D.[3]
Clifford Beldon, M.D.[4]
Ilona M. Schmalfuss, M.D.[5]
Jay S. Cooper, M.D.[6]
JoRean D. Sicks, M.S.[3]
Robert Amdur, M.D.[7]
Scott Sailer, M.D.[8]
Laurie A. Loevner, M.D.[9]
Phil Kousouboris, M.D.[10]
Kian Ang, M.D.[11]

[1] Department of Radiology, University of Michigan Health System, Ann Arbor, Michigan.

[2] Department of Otolaryngology/Head Neck Surgery, University of Michigan Health System, Ann Arbor, Michigan.

[3] Center for Statistical Sciences, Brown University, Providence, Rhode Island.

[4] Department of Radiology, State University of New York, Albany Medical School, Albany, New York.

[5] Department of Radiology, University of Florida College of Medicine, Gainesville, Florida.

[6] Department of Radiation Oncology, New York University School of Medicine, New York, New York.

[7] Department of Radiation Oncology, University of Florida College of Medicine, Gainesville, Florida.

[8] Department of Radiation Oncology, University of North Carolina School of Medicine, Chapel Hill, North Carolina.

[9] Department of Radiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania.

[10] Department of Radiology, Bryn Mawr Hospital, Bryn Mawr, Pennsylvania.

[11] Department of Radiation Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas.

BACKGROUND. Prior studies have determined that macroscopic ("gross") tumor volume (GTV), as calculated from pretreatment computer tomography (CT), was capable of predicting local control in squamous cell carcinoma arising in different subsites in the head and neck in patients who were treated with nonsurgical organ-preservation therapy. The majority of these studies were single-institution, retrospective investigations. Consequently, there has been concern that GTV measurements may not be reproducible by different readers at different institutions. The objective of the current study was to measure the interobserver reliability for GTV measurements for squamous cell carcinoma of the supraglottic larynx (SG-SCCA) performed by different readers at different institutions.

METHODS. Eight experienced readers (4 neuroradiologists and 4 radiation oncologists) from different institutions independently measured the pretreatment GTV of 20 patients with SGSCCA. The CT scans were obtained from patients entered into the definitive radiation therapy arm of Radiation Therapy Oncology Group protocol 91-11, who had supraglottic carcinoma and underwent pretreatment CT scans of the neck. Statistical analysis focused on interobserver reliability as measured by the intraclass correlation coefficient.

RESULTS. The intraclass correlation coefficient was 0.81 (95% lower confidence bound, 0.71). This value was interpreted as "excellent."

CONCLUSIONS. GTV measurements were reliable and reproducible when performed by neuroradiologists and radiation oncologists who were experienced in the interpretation of CT scans of the extracranial head and neck in patients with SGSCCA. The result implied that the correlation between GTV and local control should be reproducible across institutions. *Cancer* 2005;103:2616–22.
© 2005 American Cancer Society.

KEYWORDS: cancer, tumor volume, CT, squamous cell cancer, head and neck, larynx.

Nonsurgical organ-preservation therapy is a viable treatment option for patients with head and neck squamous cell carcinoma (HNSCCA).

The final decision for each individual frequently is based on an overall assessment of the risk of local recurrence, 5-year survival, treatment-associated morbidity, and institutional and patient preference.

In many disease sites, it has been shown that cross-sectional imaging is more accurate than physical examination for assessing primary tumor size and extent.[1–3] Consequently, pretreatment imaging has become an accepted part of staging and, as stated in the American Joint Committee on Cancer Staging Manual, any diagnostic information that may contribute to the overall accuracy of a pretreatment assessment should be taken into consideration in clinical staging and treatment planning.[4]

Numerous studies have shown that the macroscopic ("gross") tumor volume (GTV) at the primary site, as calculated from pretreatment computed tomography (CT) scans, can predict local control in squamous cell carcinoma (SCC) arising in different subsites of the head and neck in patients who are treated with definitive radiation therapy (RT).[5–17] Larger volume tumors have a higher likelihood of local recurrence than smaller volume lesions arising in the same anatomic subsite.[5–17] Threshold volumes have been identified in nasopharyngeal, oropharyngeal, supraglottic, pyriform sinus, and T3 glottic carcinomas.[5–17] Several investigators also have shown that there is a stronger association between GTV and local control than the association between T classification and local control.[6,7,15–18] This additional information helps physicians to counsel patients regarding the relative likelihood of local tumor control with primary surgery or RT with or without chemotherapy. In regard to laryngeal carcinoma, information regarding the relative risk of local failure based on an objective, quantitative measure of GTV may assist patients in determining the risk they are willing to assume to preserve native laryngeal function. Quantitative volumetric analysis also may help select patients with larger volume tumors who either desire organ-preservation therapy or who are poor surgical candidates for neoadjuvant or concurrent chemotherapy.

One drawback of integrating GTV measurements into treatment decisions is concern regarding the reproducibility of GTV measurements performed by different observers at different institutions. There is some doubt that GTV measurements can be reproduced reliably when they are performed by different subspecialists at different institutions. This concern has reduced the acceptance of integrating this concept into treatment and management decisions. An inaccurate GTV measurement has the potential to misrepresent the likelihood of nonsurgical local control and adversely affect treatment decisions. In view of this,

the objective of the current investigation was to calculate the interobserver variability of GTV measurements in patients with primary site SCC of the supraglottic larynx (SGSCCA) performed by different readers at different institutions.

MATERIALS AND METHODS
Patient Accrual
This trial was supported by the American College of Radiology Imaging Network (ACRIN) (protocol 6658). ACRIN protocol 6658 was approved by the Institutional Review Board (IRB) of the American College of Radiology and by the National Cancer Institute Cancer Therapy Evaluation Program. The patient population consisted of a subset of patients entered into Radiation Therapy Oncology Group (RTOG) protocol 91-11 between 1992 and 2000. All patients eligible for ACRIN protocol 6658 had SGSCCA and had been randomized to the "definitive RT only" arm of RTOG 91-11. Informed consent for RTOG 91-11 included consent for subsequent research as long as patient confidentiality was maintained.

RTOG member institutions and affiliates that had accrued at least one potential patient were asked to submit the protocol to their local IRB. Upon notice of local IRB approval, ACRIN requested materials for appropriate patients. Institutions then submitted available case material, including pretreatment CT studies.

ACRIN obtained from the RTOG a list of 93 potentially eligible patients enrolled at 41 institutions. Eight institutions (20%) with a total of 36 eligible patients (39%) obtained local IRB approval. Images were obtained for 28 eligible patients (78% of the 36 requested; 3 images were lost, 3 images were unavailable and presumed destroyed, 1 institution chose not to reconsent a Veterans Administration patient, and 1 institution with 1 eligible patient did not return the request form). The principal investigator reviewed the image quality for inclusion into the study. Images from 4 patients (14%) did not pass this quality review (for 3 patients the image was of the wrong primary tumor location, and for 1 patient only a poor-quality image was available, with the institution unable to obtain a better quality copy). The principal investigator then reviewed the 24 sets of images that passed initial quality review, selecting 4 as training cases (2 with suboptimal copy quality, 1 with labels on films, and 1 with nonremovable wax crayon on films), and 20 patients for the study sample. The four training cases were used to familiarize the reader with the presentation of the imaging format, software capabilities, and use of the mouse to contour the outer margin of the tumor.

Of the 20 patients in the study sample who were eligible; 55% were male and 75% were white, with a

median age of 56 years (range, 40–72 years). The lowest Karnofsky performance status was 70 (15%), and 1 patient (5%) had a Karnofsky performance status of 100. Fifty-five percent of the study patients presented with weight loss (maximum, 12 kg). There were no statistically significant differences noted between the study sample and the remainder of the original 93 potential patients with respect to age, ethnicity, gender, Karnofsky performance status, proportion who presented with weight loss, or amount of weight lost among those with this symptom.

CT Studies

The CT studies consisted of contrast-enhanced studies that were performed in patients who were entered prospectively into RTOG 91-11 between 1992 and 2000. The CT imaging protocol consisted of contiguous, 5-mm-thick sections from the skull base to the thoracic inlet. It was believed that this slice thickness was adequate at the time the protocol was initiated. Because our request for the CT studies was made after the studies already had been performed, we were unable to modify the CT imaging protocol to incorporate recent technical advances.

All original CT studies were hard-copy films. These studies were digitized using a Kodak laser scanner (Eastman Kodak, Rochester, NY) and were postprocessed for display and analysis using a proprietary software package. Each image on each sheet of film was digitized and stacked sequentially. The software automatically accounted for any potential differences in slice thickness, table spacing, and field of view to ensure accurate measurements.

Readers

Eight readers independently interpreted the CT image from each patient in the study sample. The readers came from educational institutions and had a range of 2–25 years of experience in interpreting imaging studies or managing patients with HNSCCA. Four readers were neuroradiologists who had dedicated training in head and neck radiology by working with an experienced head and neck radiologist during their neuroradiology fellowship. Three of the four neuroradiologists had a certificate of added qualification. The other four readers were radiation oncologists with experience in the treatment of HNSCC. Readers provided a rating of the quality of each CT study by selecting from excellent diagnostic quality, very good diagnostic quality, average diagnostic quality, below average diagnostic quality, or poor diagnostic quality. Readers reported tumor location by selecting from epiglottis, aryepiglottic folds, arytenoids, ventricular bands, or unable to determine. The site of origin was selected from right, left, midline, or unable to determine.

Volume Analysis

Volumetric analysis was performed on all tumors in the following manner: The interpretation and contouring was performed using a Microsoft Windows-based personal computer (Microsoft Inc., Redmond, WA) with dual $2 K \times 2 K$, high-resolution monitors and running BITImage software for viewing CT images and making the volumetric measurements. For every CT study, each reader manually used a "mouse" to contour the outer margin of any abnormal laryngeal mass on each image that they believed was suspicious of demonstrating a primary SGSCCA site. (Fig. 1) The patient entry criteria into RTOG 91-11 excluded early-stage T1 and advanced-stage T4 tumors that extended through the cartilage and into the soft tissues of the neck. Consequently, all the tumors measured in our study were confined to the larynx, and no tumor extended into the soft tissues of the neck. After completion of the contouring, the tumor volume was calculated automatically from our software package by incorporating the area contoured on each image. The software automatically accounted for differences in slice thickness, table spacing, and field of view for each study to ensure accurate tumor volume measurements, as discussed above.

Statistical Analysis

The primary statistical endpoint was interreader reliability for the evaluation of GTV, as measured by the intraclass correlation coefficient (ICC). The ICC and a 95% lower confidence bound for the ICC were estimated using a two-way analysis of variance (ANOVA) with $\log_{10}$ (GTV) as the dependent variable and with random effects for patients and readers.[19] We use $\log_{10}$ (GTV) instead of GTV, because the distribution of $\log_{10}$ (GTV) values is more symmetric than the distribution of GTV values, and the spread of observed $\log_{10}$ (GTV) values for each tumor is more stable over the range of actual tumor sizes. In part of the secondary analyses for this study, generalized κ statistics were computed to evaluate interreader reliability in impressions of image quality and tumor location.[20] Reliability statistics often are interpreted as "poor" if < 0.00, "slight" if 0.00–0.20, "fair" if 0.21–0.40, "moderate" if 0.41–0.60, "substantial" if 0.61–0.80, and "almost perfect" if 0.81–1.00.[21] In addition, we used a mixed-model ANOVA to explore differences in $\log_{10}$ (GTV) measurements between the 2 groups of subspecialists that interpreted the studies.

FIGURE 1. This axial, contrast-enhanced computed tomography scan, which was obtained at the level of the thyroid cartilage in a patient with a right aryepiglottic fold squamous cell carcinoma, demonstrates a mass involving the right aryepiglottic fold. The tumor contour drawn by Reader 1 is shown in this illustration.

RESULTS

In total, 160 CT interpretations (20 patients and 8 readers) were performed at a central reading site from May 29, 2002 through September 7, 2002. Study readers predominantly rated image quality as "average" or "very good," with 116 ratings (72.5%) that were at least average. The proportion of observed agreement among 8 readers for image quality on a 5-point scale was 0.36, increasing to 0.84 when adjacent categories were considered to be in agreement; reliability beyond chance was slight ($\kappa = 0.11$; standard error [SE] = 0.03).

One reader (Reader 5) could not identify the tumor for 1 patient (Patient 6) and therefore did not provide information on primary site, site of origin, or tumor volume. This tumor was identified by the other seven readers and had the smallest median volume among the study sample. The primary site was noted as epiglottis in 77 ratings (48%), aryepiglottic fold in 45 ratings (28%), false vocal cord in 25 ratings (16%), unable to determine in 12 ratings (16%), and arytenoid in 1 reading. The site of origin was noted as midline in 60 ratings (38%), left side in 57 ratings (36%), right side in 37 ratings (23%), and unable to determine in 6 ratings (4%). The proportion of observed agreement for location (epiglottis, aryepiglottic fold, false vocal
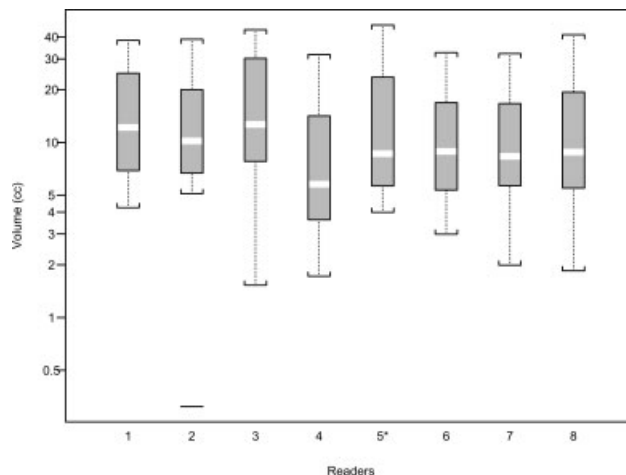

FIGURE 2. Measurements of macroscopic tumor volume were provided by study readers for patients. Each box contains the center 50% of the volumes for each reader; the white bar within the box indicates the median. The asterisk indicates the radiation oncologist who could not identify one of the tumors. The solid lines extending above and below each box indicates the range of volumes except for Reader 2, who provided a very small volume for one patient, as indicated by the small line near the bottom of the figure.

cord, unable to determine) was 0.58; reliability beyond chance was fair ($\kappa = 0.36$; SE = 0.03). The proportion of observed agreement for site of origin (left, right, midline, unable to determine) was 0.73; reliability beyond chance was moderate ($\kappa = 0.59$; SE = 0.03).

The median tumor volume among the 159 CT interpretations obtained was 9.77 cm^3 (range, 0.31–46.84 cm^3). The ICC for log^{10} GTV, which measures interreader reliability, was 0.81, with a 95% lower confidence bound of 0.71. These values are interpreted as "almost perfect" (0.81) and "substantial" (0.71). Figures 2 and 3 display the distributions of GTV for each reader and for each patient, respectively. The high degree of overlap across readers and the narrow distributions within each patient reflect the high interreader reliability for measuring GTV.

The overall median GTV measurements differed between neuroradiologists and radiation oncologists. The median GTV was 8.3 cc when measured by diagnostic neuroradiologists and 11.5 cc when measured by radiation oncologists. This difference was significant ($P = 0.02$). The ICC for log_{10} GTV was measured for the subgroups of radiologists and radiation oncologists. The ICC for neuroradiologists was 0.84 (95% lower confidence bound of 0.74), whereas the ICC for radiation oncologists was 0.88 (95% lower confidence bound, 0.79); the difference between the ICC for neuroradiologists and radiation oncologists was not significant. Figure 4 displays the distribution of GTV for each reader grouped by subspecialty. The horizontal
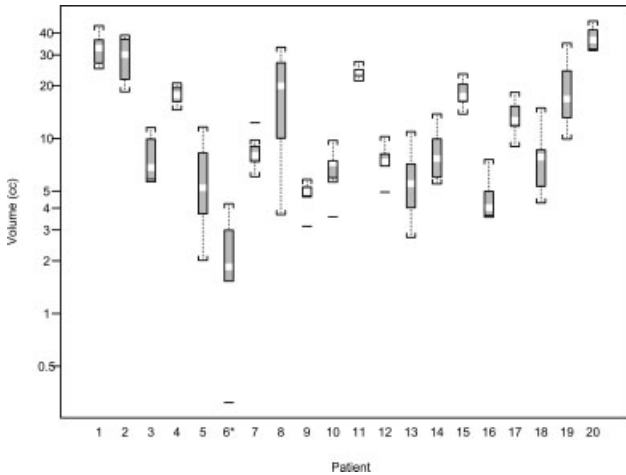
FIGURE 3. Measurements of macroscopic tumor volume for each patient were provided by the study readers. Each box contains the center 50% of the volumes for each patient; the white bar within the box indicates the median. The asterisk indicates a patient for whom one radiation oncologist could not identify a tumor. The solid lines extending above and below each box indicate the range of volumes for each patient. For Patients 6, 9, 10, and 12, 1 reader provided a volume that was notably smaller than the volume reported by other readers, as indicated by the small lines that are not connected to the boxes.
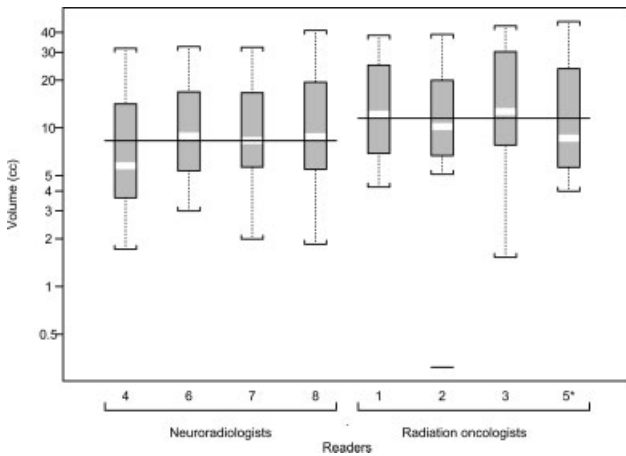

FIGURE 4. Measurements of macroscopic tumor volume provided by study readers for patients are organized by reader type. Horizontal lines indicate the median volume for each group. Each box contains the center 50% of volumes for each reader; the white bar within the box indicates the median. The dashed lines extending above and below each box indicate the range of volumes for each reader, except for Reader 2, who provided a very small volume for one patient, as indicated by the small line near the bottom of the figure. The asterisk indicates the radiation oncologist who could not identify one of the tumors.

lines indicate the median GTV for each group. The similarity in the sizes of the boxes reflects the consistency in variability over all of the readers. Figure 5 displays the range of GTV measurements for each case
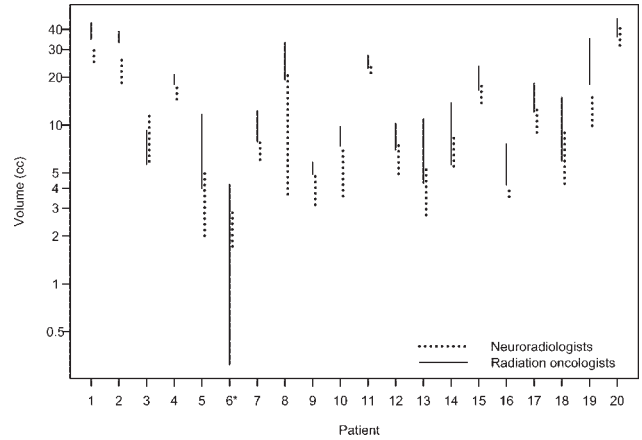

FIGURE 5. Measurements of macroscopic tumor volume are organized by patient and by the subspecialty of the reader. Line segments indicate the range of the four measurements per patient and subspecialty. The asterisk indicates a patient for whom one radiation oncologist could not identify a tumor.

by subspecialty. The similarity in the lengths of the lines indicated that the variability is similar between the two groups. The fact that the solid lines generally are higher reflects the fact that the radiation oncologists' tumor volume measurements tended to be higher.

DISCUSSION

The results of our investigation demonstrate that GTV derived from CT studies can be measured reliably across readers. The estimated ICC of 0.81 commonly is considered excellent. These findings indicate that reproducible GTV measurements can be obtained across different institutions and different subspecialties.

In 1998, Hermans et al.[22] reported on intraobserver and interobserver reliability of GTV measurements using 13 laryngeal tumors (including 5 supraglottic carcinomas) that were evaluated 4 times by each of 5 readers. In the current study, 8 readers evaluated 20 supraglottic carcinomas once. The tumors in our series tended to be larger in volume. Due to study design, the statistical analyses performed for the two studies differed. Hermans et al. generously supplied us their data, which we used to obtain interobserver ICCs for each of the 4 readings, which ranged from 0.75 to 0.82. Therefore, our results are similar to the ICCs obtained from the data of Hermans et al.[22]

Our readers consisted of both neuroradiologists and radiation oncologists who were experienced in the treatment of HNSCCA. We believed that it was important to include both groups to emphasize the interdisciplinary approach for the management and treatment of HNSCCA. GTV measurements may be

performed by either radiologists or radiation oncologists, depending on the institution. GTV may be calculated for initial evaluation by neuroradiologists. Tumor contouring for RT planning is performed by radiation oncologists for either definitive or adjuvant treatment. Because both groups are involved in tumor contouring and volume measurements, we believed that it was important to determine interobserver variability not only between institutions but also between specialists involved in the management of these patients. The overall ICC for our study was 0.81 ("excellent"). The estimated ICCs were similar for each subspecialty and were higher than the overall ICC, indicating that neuroradiologists were more likely to measure like other neuroradiologists than like radiation oncologists, and vice versa.

Exploratory data analysis showed that the median GTV measurement was slightly larger for radiation oncologists than for the neuroradiologists. This unusual finding needs to be confirmed in a prospective study, because it does have potential implications in clinical practice: Most published data were based on work done by radiologists. Unfortunately, there is no reliable method for identifying the true GTV of an in vivo tumor accurately, because the size of the tumor will change once it is devascularized and removed from the patient. Thus, it is not possible to determine whether the neuroradiologists or radiation oncologists had the more accurate measurements. The potential effect on management, based on measurements obtained by different subspecialists, would need to be evaluated in a large-scale, prospective investigation. Radiation oncologists may estimate larger sizes based on their common use of endoscopy, which reveals the tumor's surface anatomy, which is not seen on CT. Subconsciously, radiation oncologists may add a conceptual volume to the volume estimated solely from CT to compensate. Radiation oncologists may tend to be more inclusive and "round off" irregular edges of a tumor, as assessed by CT, to conform with their ideas of tumor growth, whereas diagnostic radiologists may contour exactly what they see. Unfortunately, in the current study, we could not differentiate between these possibilities and could not resolve the issue; however, it does raise questions that need to be addressed in future trials. Currently, in the absence of a true gold standard, it appears that patients are served ideally when their tumors are evaluated by a multidisciplinary group, potentially diluting any individual group's tendencies.

Specific imaging parameters appear to be strong predictors of local and locoregional outcome in patients with supraglottic carcinoma who are treated by definitive RT. GTV appears to be the strongest inde-

pendent predictor of local failure after RT.[9] Pretreatment CT measurements of GTV permit stratification of local control in patients with SGSCCA who receive treatment with RT alone. Mancuso et al. reported local control rates of 89% in tumors < 6 cc and 52% when tumor volumes were ≥ 6 cc. Some authors suggest that SGSCCA with a GTV > 16 cc should be treated with surgical resection only.[14,15] The specific threshold volume may vary; however, what is of greater importance is the linear relation between tumor volume and local control, as reported by Mancuso et al. Based on this relation, a more accurate estimate of cure can be obtained for each individual patient rather than relying on general threshold measurements.[9] Therefore, consistent and reproducible measurements are important when attempting to quantify GTV. Large variability in GTV measurements may result in an erroneous classification of a tumor as too large to be treated successfully with definitive RT.

A limitation of the current study was that the CT protocol was not optimized for slice thickness, gantry angle, or helical acquisition. Current recommendations for imaging studies in patients with laryngeal carcinoma include a slice thickness ≤ 3 mm, a gantry angle that is parallel to the laryngeal ventricle, and the image acquisition should be performed with a multidetector scanner. The CT studies were performed on patients that prospectively entered RTOG study 91-11 between 1992 and 2000. Because our request for these CT studies was made after the CT already had been performed, we were unable to modify the CT imaging protocol by incorporating current techniques and concepts. Despite the fact that an optimized protocol was not used, we feel comfortable that the available imaging studies can be used to measure interobserver variability for GTV measurements.

The results of the current study indicate that reproducible GTV measurements can be obtained at different institutions by both neuroradiologists and radiation oncologists. All readers should be knowledgeable with the appearance of normal anatomy, common tumor behavior, and pathology of the head and neck to obtain accurate and reproducible measurements. Multidisciplinary evaluation teams appear to provide the greatest level of comfort, so that an accurate measurement of tumor volume is obtained. However, for tumors that are substantially larger or substantially smaller than the quasithreshold of suitability, for nonsurgical management, our results suggest that experienced diagnostic radiologists and radiation oncologists agree that assessment by either subspecialty should lead to the correct assessment. Given these findings, we believe that primary site GTV information can be taken into account reliably when

considering treatment options in patients with SG-SCCA.

REFERENCES

1. Isaacs JH, Manucso AA, Mendenhall WM. CT scanning as an aid to selection of therapy in T2–T4 laryngeal cancer. *Head Neck Surg*. 1988;99:455–464.
2. Mancuso AA, Hanafee WN. Larynx and hypopharynx. In: Curtin HD, editor. Computed tomography and magnetic resonance imaging of the head and neck, 2nd edition. Baltimore: Williams and Wilkens, 1985:241–357.
3. Curtin HD. Larynx. In: Som PM, Curtin HD, editors. Head and neck imaging, 3rd edition. St. Louis: Mosby Year Book, 1995:612–707.
4. Fleming ID, Cooper JS, Henson DE, et al., editors. AJCC cancer staging manual, 5th edition. Philadelphia: Lippincott-Raven, 1998.
5. Mancuso AA, Mukherji SK, Kotzur I, et al. Preradiotherapy-computed tomography as a predictor of local control in supraglottic carcinoma. *J Clin Oncol*. 1999;17:631–636.
6. Hermans R, Van den Bogaert, Runders A, Baert A. Value of computed tomography as outcome predictor of supraglottic carcinoma treated by definitive radiation therapy. *Int J Radiat Oncol Biol Phys*. 1999;44:755–765.
7. Chua DTT, Sham JST, Kwong DLW, et al. Volumetric analysis of tumor extent in nasopharyngeal carcinoma and correlation with treatment outcome. *Int J Radiat Oncol Biol Phys*. 1997;39:711–719.
8. Johnson CR, Thames HD, Huang DT, Schmidt-Ullrich RK. The tumor volume and clonogen number relationship: tumor control predictions based upon tumor estimates derived from computed tomography. *Int J Radiat Oncol Biol Phys*. 1995;33:281–287.
9. Mancuso AA, Mukherji SK, Kotzur, et al. Preradiotherapy-computed tomography as a predictor of local control in supraglottic carcinoma. *J Clin Oncol*. 1999;17:631–663.
10. Freeman DE, Mancuso AA, Parsons JT, et al. Irradiation alone for supraglottic larynx carcinoma: can CT findings predict treatment results? *Int J Radiat Oncol Biol Phys*. 1990;19:485–490.
11. Pameijer FA, Mancuso AA, Mendenhall WM, Parsons JT, Kubilis PS. Can pretreatment computed tomography predict local control in T3 squamous cell carcinoma of the glottic larynx treated with definitive radiotherapy? *Int J Radiat Oncol Biol Phys*. 1997;37:1011–1021.
12. Lee WR, Mancuso AA, Saleh EM, et al. Can pretreatment computed tomography findings predict local control in T3 squamous cell carcinomas of the glottic larynx treated with radiotherapy alone? *Int J Radiat Oncol Biol Phys*. 1993;25:683–687.
13. Gilbert RW, Birt D, Shulman H, et al. Correlation of tumor volume with local control in laryngeal carcinoma treated by radiotherapy. *Ann Otol Rhinol Laryngol*. 1987;96:514–518.
14. Mukherji SK, O'Brien SM, Gerstle RJ, Castillo M. Tumor volume: an independent predictor of outcome for laryngeal cancer. *J Comput Assist Tomogr*. 1999;23:50–54.
15. Mukherji SK, Gerstle RJ, O'Brien SM, et al. Tumor volume as a predictor of local control in patients with laryngeal cancer treated surgically. *J Hong Kong Coll Radiol*. 1999;2:104–111.
16. Hermans R, Van den Bogaert W, Rijnders A, Doornaert P, Baert A. Predicting the local control of glottic squamous cell carcinoma after definitive radiation therapy: value of computed tomography-determined tumour parameters. *Radiother Oncol*. 1999;50:39–46.
17. Mukherji SK, Gerstle RJ, O'Brien SM, et al. The ability of tumor volume to predict local control in surgically treated squamous cell carcinoma of the supraglottic larynx. *Head Neck*. 2000;22:282–287.
18. Mukherji SK, Mancuso AA, Mendenhall M, O'Brien S, Weissler M, Pillsbury HC. The ability of tumor volume to predict relative risks of local failure associated with surgical and non-surgical treatment of supraglottic carcinoma. *Asian Oceanic J Radiol*. 2001;6:13–20.
19. Fleiss JL. Reliability of measurement. In: Fleming TR, Harrington DP, editors. The design and analysis of clinical experiments. New York: John Wiley & Sons, Inc., 1986:1–28.
20. Fleiss JL. The measurement of interrated agreement. In: Teutsch SM, Churchill RE, editors. Statistical methods for rates and proportions, 2nd edition. New York: John Wiley & Sons, Inc., 1981:225–232.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
22. Hermans R, Feron M, Bellon E, Dupont P, Van den Bogaert W, Baert AL. Laryngeal tumor volume measurements determined with CT: a study on intra- and interobserver variability. *Int J Radiat Oncol Biol Phys*. 1998;40:553–557.
23. Forastiere AA, Goepfert H, Maor M, et al. Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. *N Engl J Med*. 2003;349:2091–2098.