



Published in final edited form as:

Int J Radiat Oncol Biol Phys. 2007 March 15; 67(4): 972–975.

An Evaluation of the Variability of Tumor-Shape Definition Derived by Experienced Observers from CT-Images of Supraglottic Carcinomas (ACRIN protocol 6658)

Jay S. Cooper, MD¹, Suresh K. Mukherji, MD^{2,3}, Alicia Y. Toledano, ScD⁴, Clifford Beldon, MD⁵, Ilona M. Schmalfluss, MD⁶, Robert Amdur, MD⁷, Scott Sailer, MD⁸, Laurie A. Loevner, MD⁹, Phil Kousouboris, MD¹⁰, K. Kian Ang, MDPH¹¹, Jean Cormack, PhD⁴, and JoRean Sicks, MS⁴

1 Department of Radiation Oncology, Maimonides Medical Center (telephone 1-718-765-2744; fax 1-718-765-2709; email jcooper@maimonidesmed.org)

2 Department of Radiology, University of Michigan Health System

3 Department of Otolaryngology Head Neck Surgery, University of Michigan Health System

4 Center for Statistical Sciences, Brown University

5 Department of Radiology, State University of New York, Albany Medical School

6 Department of Radiology, University of Florida School of Medicine

7 Department of Radiation Oncology, University of Florida School of Medicine

8 Department of Radiation Oncology, University of North Carolina School of Medicine

9 Department of Radiology, University of Pennsylvania School of Medicine

10 Department of Radiology, Bryn Mawr Hospital

11 Department of Radiation Oncology, M. D. Anderson Cancer Center

Abstract

Purpose—Accurate target definition is considered essential for sophisticated, image-guided radiation therapy; however, relatively little information has been reported that measures our ability to identify the precise shape of targets accurately. We decided to assess the manner in which eight “experts” interpreted the size and shape of tumors based on “real life” contrast-enhanced CT scans.

Methods and Materials—Four neuroradiologists and four radiation oncologists (the authors) with considerable experience and presumed expertise in treating head and neck tumors independently contoured, slice-by-slice, his/her interpretation of the precise gross tumor volume (GTV) on each of twenty sets of CT scans taken from twenty patients who previously were enrolled in Radiation Therapy Oncology Group protocol 9111.

Results—The average proportion of overlap (i.e., the degree of agreement) was 0.532 (95% confidence interval 0.457 to 0.606). There was a slight tendency for the proportion of overlap to increase with increasing average GTV.

This work represents a group effort of the American College of Radiology Imaging Network (ACRIN) and was supported by the National Cancer Institute through the grants U01 CA079778 and U01 CA080098.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusions—Our work suggests that estimation of tumor shape currently is imprecise, even for experienced physicians. In consequence, there appears to be a practical limit to the current trend of smaller fields and tighter margins.

Keywords

imaging; target definition; GTV estimation

Introduction

Reduced to its simplest elements, the goal of radiation therapy is to irradiate all malignant cells while avoiding all normal ones. This suggests that all malignant cells can be identified and that technology can deliver radiation exclusively to the targeted cells. In recent years the technology to deliver radiation therapy has advanced substantially and the patterns of dose distribution produced by intensity modulated radiation therapy (IMRT) can now shape dose clouds to conform to targets far more closely than ever before, while relatively sparing normal tissues. Computerized tomography, magnetic resonance imaging and positron emission tomography are being used increasingly to help define targets for more sophisticated forms of radiation therapy. But, how well do we, as clinicians, interpret these modalities and how well do our targets represent reality?

The American College of Radiology Imaging Network (ACRIN) protocol 6658 was designed to calculate the interobserver variability of gross tumor volume (GTV) determined by different physicians for patients who had squamous cell carcinoma of the supraglottic larynx (SGSCCA).⁽¹⁾ Eight experienced physicians from different institutions (four neuroradiologists with considerable experience and presumed expertise in delineating head and neck tumors and four radiation oncologists with considerable experience and presumed expertise in treating head and neck tumors) independently contoured, slice-by-slice, his/her interpretation of the precise GTV on each of twenty sets of CT scans taken from twenty patients who previously were enrolled in Radiation Therapy Oncology Group (RTOG) protocol 91-11.⁽²⁾ We have already reported the outcome (1) of these investigations with respect to primary tumor volumes and essentially demonstrated that GTV measurements are reliably and reproducibly measured by neuroradiologists and radiation oncologists who were experienced in the interpretation of CT scans of the extracranial head and neck in patients with SGSCCA. Consequently, the likelihood of control of disease by radiation therapy alone, based on the volume of disease, is consistent from experienced physician to experienced physician.⁽³⁾ However, as part of this process, we noticed that the precise shape of the cross-sectional areas deemed to represent tumor on individual slices of each set of scans was not uniformly congruent among the physicians. As the sum of these slices in a given set represents the perceived shape of the tumor, we decided to investigate precisely how similar the shapes of the tumors that each of the eight physicians contoured were to each other. The assessment of shape is critical for treatment planning and therapy in Radiation Oncology, particularly when IMRT is used. This report details our findings with respect to these “real life” examples of three-dimensional tumor shape assessment by experienced physicians.

Material and Methods

ACRIN protocol 6658 was based on data from a subset of eligible patients with squamous cell carcinoma of the supraglottic larynx entered into Radiation Therapy Oncology Group (RTOG) protocol 91-11 who had a CT scan and who had been randomized to the “definitive RT only” arm. Informed consent for RTOG 91-11 included consent for subsequent research as long as patient confidentiality was maintained and the use of the images from RTOG 91-11 occurred only after the ACRIN 6658 was approved by the National Cancer Institute Cancer Therapy

Evaluation Program, by the Institutional Review Board (IRB) of the American College of Radiology, and by each local IRB of participating institutions.

Details of the procedures for delineation of slice-by-slice estimates of tumor involvement have previously been published.⁽¹⁾ Subsequently, the first author (JSC) outlined the areas of concordance (and therefore non-concordance) on every CT slice, for every pair of readers, for every case. Following the contouring of areas of concordance on each CT slice, the volumes of concordance (overlap) and non-concordance for each reader pair for each case were automatically calculated by the proprietary software package (BIT Display 07x00) which automatically accounted for differences in slice thickness, table spacing, and field of view to ensure accurate measurements.

In the absence of surgical resection and histologic measurement, the precise “true” tumor shape and volume is unknown. Typical, simple descriptive statistical analyses comparing our findings to a “true standard” therefore could not be applied. In addition, because the tumor itself typically occupies only a small portion of the area of each CT slice (or in the aggregate volumes), the area (or volumes) not included in the contours drawn by readers would dominate any statistic that simply measured spatial agreement. We therefore chose to quantify the degree of concordance by calculating the proportion of overlap between pairs of readers, i.e., the overlapping volume divided by the sum of the overlapping and non-overlapping volumes. To estimate any potential influence of the specialty training of the readers, we not only estimated average proportions of overlap for all readers, we also specifically compared the outcome within types of reader pairs (i.e., 2 diagnostic radiologists, 2 radiation oncologists, one of each), along with standard errors of those estimates.

Statistical inference is further complicated because the proportion of overlap for a pair of readers in any case is correlated with all other proportions of overlap for that case, and all other proportions of overlap involving either or both of those readers. For example, if two readers have different contouring tendencies (such as one reader contouring with relatively tight margins routinely and the other contouring with relatively large margins routinely) their estimates of shape and volume would be consistently affected. To address this, we used a regression model for proportion of overlap that incorporated a fixed mean proportion of overlap and random effects for cases, readers, and their interactions. Because we wanted to learn if the estimation of shape was influenced by the size of the tumor, we compared the observed correlations of overlap by primary tumor volumes. Since the “true” primary tumor volumes could not be measured, we estimated the “true” volume for each case as the average of all volumes estimated by all readers, mindful that this method, by itself, creates a relationship between the proportion of overlap and the estimated primary tumor volume (To compensate for this, we used a permutation distribution of the correlation under the null hypothesis of no relationship).

Results

The proportion of overlap was computed for 546 reader pairs (it could not be computed for the remaining 14 reader pairs because contours from 2 cases were missing for 1 reader due to a computer error resulting in accidental overwriting of the files). Figures 1 and 2 demonstrate examples of relatively high and low concordance of the shapes outlined on one slice of one set of scans by two readers. Overall, the average proportion of overlap was 0.532 (range: 0.000 to 0.818), with 95% confidence interval for average proportion of overlap from 0.457 to 0.606. Figure 3 shows the relationship of proportion of overlap to the average of the tumor volumes across the eight study readers. There is a slight tendency for proportion of overlap to increase with increased average GTV. The correlation between values of proportion of overlap and average tumor volume is 0.490 (p-value < 0.0002).

On average, proportion of overlap was slightly lower when one or both members of the reader pair were radiation oncologists (averages of 0.516 and 0.526, respectively, with ranges from 0.000 to 0.816 and 0.812, respectively) than when both members of the reader pair were radiologists (average = 0.580, range: 0.158 to 0.818). However, the standard error of the average proportion of overlap among reader pairs when both readers were radiologists was 0.048, compared with 0.041 when one reader was a radiologist and the other a radiation oncologist and 0.037 when both readers were radiation oncologists. Corresponding 95% confidence intervals for average proportion of overlap were: 0.454 to 0.598 when both readers were radiation oncologists, 0.485 to 0.675 when both readers were radiologists, and 0.436 to 0.597 when the reader pair was mixed.

Discussion

In a 2003 manuscript (6) titled “Image Guidance For Precise Conformal Radiotherapy”, Mackie et al. wrote, “Image guidance is beginning to be the fundamental basis for radiotherapy planning, delivery, and verification.” A guidance document on delivery, treatment planning, and clinical implementation of IMRT from the American Association of Physicists in Medicine (7) states, “With inverse planning, the physician designates targets instead of designing fields, so careful and accurate contouring is essential.” Both statements imply that currently obtainable images represent accurate depictions of the extent of the tumor and that physicians are capable of accurately translating these images into three-dimensional targets. This study was designed to test the latter hypothesis.

Two overlapping issues need to be discussed in this context. The first is the reproducibility of tumor volume measurements. We previously have reported high levels of agreement among the authors in the determination of the volumes of tumor as part of this ACRIN study (1). This is reassuring for decisions concerning the suitability of radiation therapy as a treatment modality for a given tumor in light of the data suggesting quasi-thresholds of volume above which radiation therapy alone is not likely to eradicate disease.(3⁻⁵)

The second issue is shape, because volume and shape are only loosely related. For example, a soccer ball and a football could be exactly the same volume and contain exactly the same amount of air. Yet, the soccer ball and football will have substantially different shapes. Similarly, two tumors can have identical volume, but substantially different shape. Volume will influence the likelihood of control by radiation therapy at a given dose, but shape will determine the placement of multileaf collimators/blocks, number of fields, adjacent normal tissue that must be included in the treatment portals, etc.

Identification of targets, both those to be treated and those to be avoided has been the subject of several recent reports.(8⁻¹³) However, these tend to focus on the ability of newer modalities to detect tumors with greater sensitivity and specificity, presumably leading to better recognition by physicians of targets. Hong et al.(14) have assessed the variability in IMRT targets drawn from an unambiguous “virtual” tumor and reported what they termed “remarkable variation” and “substantial variation.” Our work suggests that a relatively overlooked, but important, parallel task in the process is to improve the agreement of “experts” as to precisely what represents “tumor” when looking at exactly the same images drawn from real life cases.

Our work examines the precise three-dimensional shapes that eight experienced physicians derived from contrast enhanced CT-images of 20 real-life supraglottic carcinomas. Although the speed and resolution of CT scanners appears to be getting better and better over time and our group of physicians might have performed differently if all of the images were obtained from a 2006-vintage top-of-the-line scanner and were technically perfect, our work probably

represents the real world experience far better. It suggests that on average each pair of reviewers agreed on approximately one-half (53.17% \pm 3.80% standard dev.) of each tumor's exact shape. Some differences in the average degree of concordance are apparent between physicians grouped by their specialty (Radiologists vs. Radiation Oncologists vs. mixed), but no group came close to perfect concordance.

Furthermore, the presence of overlapping volumes does not ensure correctness. As there is no gold standard measure of a tumor's shape while it remains *in situ*, we recognize that two physicians can draw GTVs that have perfect overlap in the wrong location, or no overlap with each including only part of the tumor, and we could not measure the difference from "absolute truth".

The implications of our work suggest that estimation of tumor shape currently is less than a precise science. When the limitations of available technology dictated large parallel opposed or 3-field techniques for the management of head and neck tumors, the precise estimation of a tumor's shape was not critical. But, today's image-guided, intensity modulated technology imposes an obligation upon the profession; we must be critically aware of our limitations lest we confuse the precision of image guidance and the conformality of IMRT with accuracy of targeting tumors. How we train ourselves to practice and how we accredit facilities for IMRT protocol participation at present still appear to require some volume to be added to GTVs to compensate for our inability to know the precise shape of the intended target. Although some tumors may be easier to describe (our maximum observed concordance between two observers was 81.8%), in our worst case there was no agreement between two observers. Consequently, there appears to be a practical limit to the current trend of smaller and smaller fields with tighter and tighter margins that may ultimately influence the success and applicability of image guided, intensity modulated radiation therapy techniques.

References

1. Mukherji SK, Toledano AY, Beldon C, Schmalfuss IM, Cooper JS, Sicks JD, Amdur R, Sailer S, Loevner LA, Kousouboris P, Ang K. Interobserver reliability of computed tomography-derived primary tumor volume measurement in patients with supraglottic carcinoma. *Cancer* 2005 Jun 15;103(12):2616–22. [PubMed: 15887218]
2. Forastiere AA, Goepfert H, Maor M, Pajak TF, Weber R, Morrison W, Glisson B, Trotti A, Ridge JA, Chao C, Peters G, Lee DJ, Leaf A, Ensley J, Cooper J. Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. *N Engl J Med* 2003 Nov 27;349(22):2091–8. [PubMed: 14645636]
3. Nathu RM, Mancuso AA, Zhu TC, Mendenhall WM. The Impact Of Primary Tumor Volume On Local Control For Oropharyngeal Squamous Cell Carcinoma Treated With Radiotherapy. *Head Neck* 2000;22:1–5. [PubMed: 10585598]
4. Keberle M, Hoppe F, Dotzel S, Hahn D. Tumor volume as determined by computed tomography predicts local control in hypopharyngeal squamous cell carcinoma treated with primary surgery. *Eur Radiol* 2004 Feb;14(2):286–91. [PubMed: 12845466]
5. Mendenhall WM, Morris CG, Amdur RJ, Hinerman RW, Mancuso AA. Parameters that predict local control after definitive radiotherapy for squamous cell carcinoma of the head and neck. *Head Neck* 2003 Jul;25(7):535–42. [PubMed: 12808656]
6. Mackie TR, Kapatoes J, Ruchala K, Lu W, Wu C, Olivera G, Forrest L, Tome W, Welsh J, Jeraj R, Harari P, Reckwerdt P, Paliwal B, Ritter M, Keller H, Fowler J, Mehta M. Image guidance for precise conformal radiotherapy. *Int J Radiat Oncol Biol Phys* 2003 May 1;56(1):89–105. [PubMed: 12694827]
7. Ezzell GA, Galvin JM, Low D, Palta JR, Rosen I, Sharpe MB, Xia P, Xiao Y, Xing L, Yu CX. IMRT subcommittee; AAPM Radiation Therapy committee. Guidance document on delivery, treatment planning, and clinical implementation of IMRT: report of the IMRT Subcommittee of the AAPM Radiation Therapy Committee. *Med Phys* 2003 Aug;30(8):2089–115. [PubMed: 12945975]

8. Kauczor HU, Zechmann C, Stieltjes B, Weber MA. Functional magnetic resonance imaging for defining the biological target volume. *Cancer Imaging* 2006 Jun 1;6:51–5. [PubMed: 16766269]
9. Everitt C, Leong T. Influence of F-fluorodeoxyglucose-positron emission tomography on computed tomography-based radiation treatment planning for oesophageal cancer. *Australas Radiol* 2006 Jun; 50(3):271–4. [PubMed: 16732830]
10. Dimopoulos JC, Schard G, Berger D, Lang S, Goldner G, Helbich T, Potter R. Systematic evaluation of MRI findings in different stages of treatment of cervical cancer: potential of MRI on delineation of target, pathoanatomic structures, and organs at risk. *Int J Radiat Oncol Biol Phys* 2006 Apr 1;64 (5):1380–8. [PubMed: 16580496]
11. van Baardwijk A, Baumert BG, Bosmans G, van Kroonenburgh M, Stroobants S, Gregoire V, Lambin P, De Ruyscher D. The current status of FDG-PET in tumour volume definition in radiotherapy treatment planning. *Cancer Treat Rev* 2006 Jun;32(4):245–60. [PubMed: 16563636]
12. Lavrenkov K, Partridge M, Cook G, Brada M. Positron emission tomography for target volume definition in the treatment of non-small cell lung cancer. *Radiother Oncol* 2005 Oct;77(1):1–4. [PubMed: 16225943]
13. Heron DE, Smith RP, Andrade RS. Advances in image-guided radiation therapy - the role of PET-CT. *Med Dosim* 2006 Spring;31(1):3–11. [PubMed: 16551524]
14. Hong TS, Tome WA, Chappell RJ, Harari PM. Variations in target delineation for head and neck IMRT: An international multi-institutional study. *Int J Radiat Oncol Biol Phys* 2004;60(1 Suppl 1):S157–158. A#47.

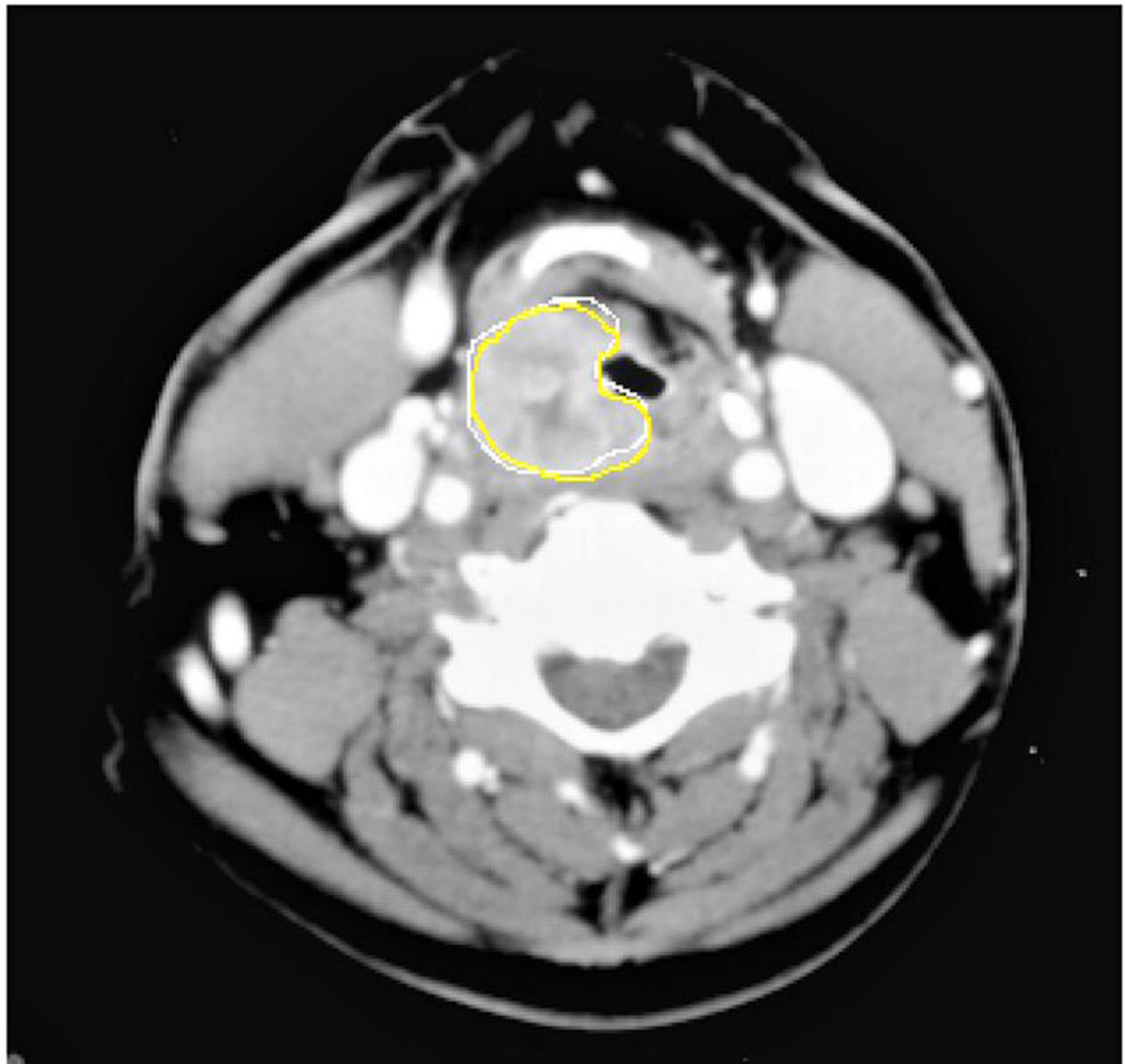


Figure 1.
Example of Relatively Good Concordance of Shape Estimates by Two Physicians.

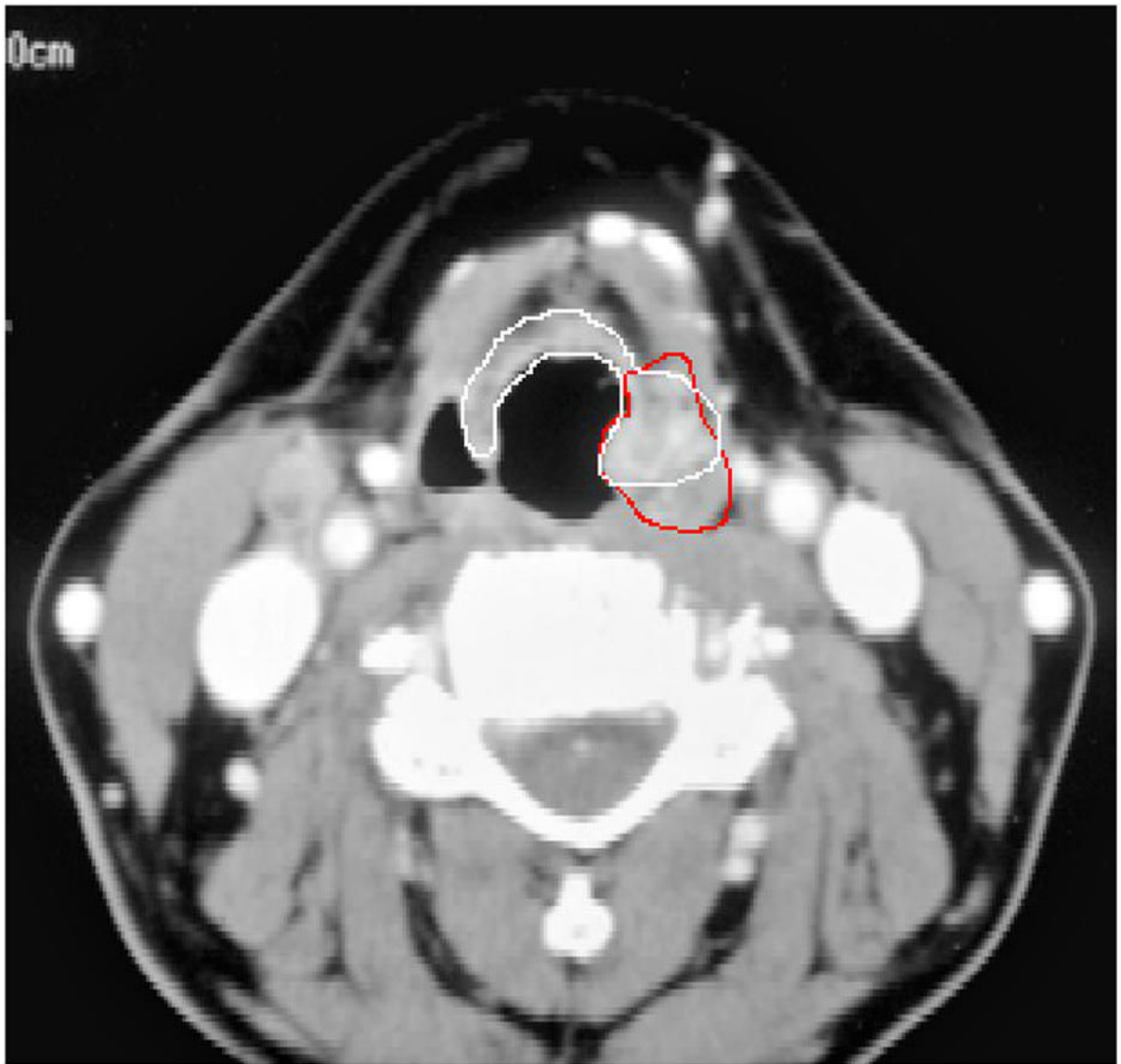


Figure 2.
Example of Relatively Poor Concordance of Shape Estimates by Two Physicians.

PR Overlap by Mean PTV

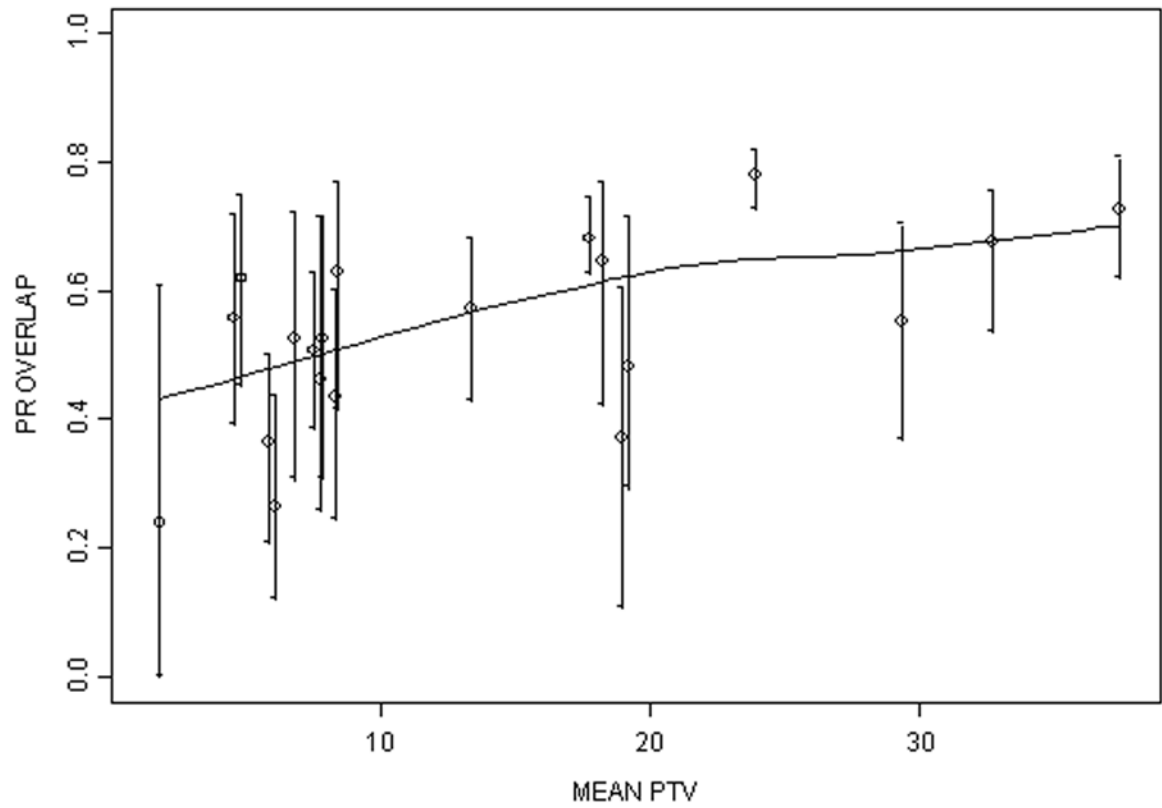


Figure 3.
Relationship of proportion of overlap to mean PTV.