

Experimental Procedure: A data set containing a variety of tumor types, including head/neck (9 cases), ovarian (5 cases), and a variety of lung, liver and head lesions (23 cases), was assembled for this study from previously acquired Phase I clinical trial data. Images for at least two time points were available for each of the 37 cases selected. Seven of these 37 cases were selected at random for inclusion in the intra-observer variability portion of this study. Each of these cases was replicated and re-identified four times, for a total of 28 additional cases. These repeated cases were then inserted into the original data set in a pseudo-random order such that, when the resulting 65 case data set was separated into five blocks of 13 cases, no block contained any of the original 37 cases more than once.

The full 65 case data set was analyzed a total of 10 times by 7 radiologists, all of whom were experienced in oncology and trained in the use of VirtualScopics' analysis software. Two radiologists made 1-D and 2-D measurements using virtual calipers. Two performed full volumetric analyses using the semi-automated method described in the box at right. The other three analyzed the full data set using both methods. Those radiologists participating in the inter-reader variability portion of the study were allowed to analyze the data sets at their own pace. The two (one manual, one semi-automated) who participated in the intra-reader variability study were required to have at least two weeks separation between analyses of 13 case blocks, and thus between any two repeated data sets.

Background: Uni-dimensional (RECIST) and bi-dimensional (WHO) axial-plane measurements are currently the standard imaging markers for structural changes in solid tumors. There is, however, currently no standardized method for obtaining these measurements.^{1,2,3} Furthermore, there is ample evidence that highly manual methods for obtaining these measurements may lead to poor measurement precision and high error rates in tumor staging and assessment of response to treatment.^{1,4} The aim of this work was to evaluate an optimized "virtual caliper" method for manual in-plane measurement, as well as a semi-automated technique that derives the WHO and RECIST measurements from complete volumetric tumor contours, and to compare these methods to film-based methods in terms of measurement variability and error rate. As a secondary goal, we also wished to explore the relative utility of full volume measurement in a variety of tumor types.⁵

Results: The results of these experiments were assessed in terms of intra- and inter-operator variability, as well as error rate in classification of cases into three categories: progressing, stable, and responding. Cut-off values for the three categories for each measurement type are given in Table 1 below.⁶ Correct classification was determined by the consensus of the 10 total reads available for each data set. The error rate is therefore a measure of the variability among readers in case categorization.

	Film-based Inter-operator	Virtual Caliper Inter-operator	Semi-automated Inter-Operator	Film-based Intra-operator	Virtual Caliper Intra-operator	Semi-automated Intra-Operator
RECIST %Error	30%	9.4%	6.9%	9.5%	1.6%	5.2%
WHO %Error	43%	8.5%	9.1%	21%	5.2%	4.6%
Volume %Error	N/A	N/A	7.9%	N/A	N/A	7.1%

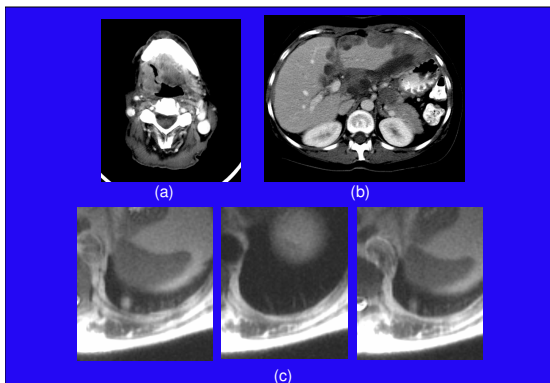
Table 2: Inter- and intra-observer error rates for film-based caliper measurements, digital virtual caliper measurements, and semi-automated tumor measurements using RECIST, WHO, and volume measures. Film-based measurements are taken from Erasmus *et al.*,¹ using criteria for progressive disease. Virtual Caliper and Semi-automated measurements include summed errors for both progression and response.

	Virtual Caliper Inter-Operator	Semi-Automated Inter-Operator	Virtual Caliper Intra-Operator	Semi-Automated Intra-Operator
RECIST CV	13.5%	9%	4.3%	5.2%
WHO CV	24.8%	16%	8.9%	9.1%

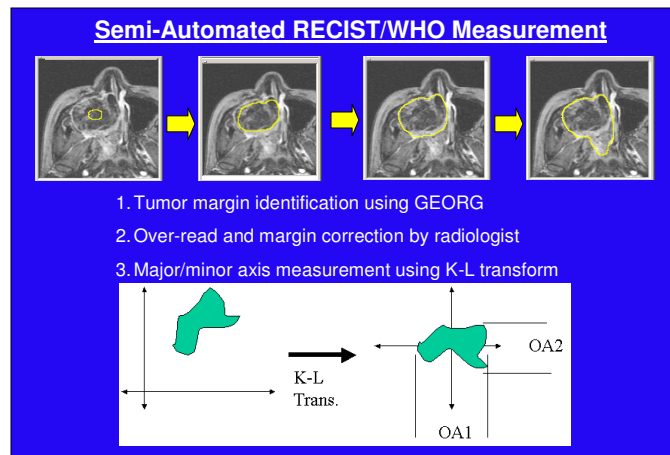
Table 3: Inter- and intra-observer coefficients of variability for RECIST and WHO measurements using Virtual Calipers and the semi-automated measurement technique described below. There is no statistical difference in intra-observer variability for either WHO or RECIST. However, semi-automated analysis shows significantly better inter-observer variability.

Table 1: Tumor measurement change criteria for disease progression and response to treatment.

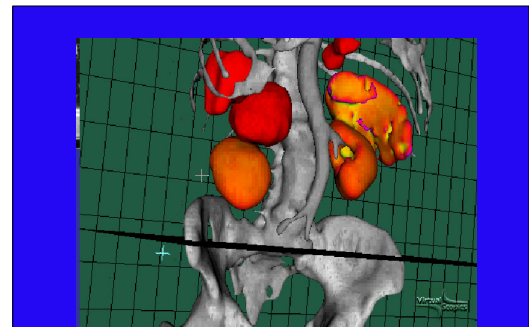
Measurement error rates for the inter- and intra-operator portions of this study are summarized in Table 2. Coefficients of variability are shown in Table 3. It should be noted that volume measurements were only possible using the semi-automated method. Two key conclusions can be drawn from these results. (1) WHO/RECIST measurements obtained using semi-automated methods show significantly better inter-operator reproducibility and lower error rates. (2) Volume measurements, which should correspond more closely to true disease progression, can be obtained with error rates similar to those of the more simplistic one and two dimensional measures.



Examples of the difficulty of the subject data used in this experiment. (a) This Head/Neck case shows a tumor with an irregular shape and infiltrating margins. (b) This Ovarian case contains many confluent lesions. (c) This series shows three consecutive slices from one lung case imaged using MR. Note that the lesion disappears on the second slice and reappears on the third. This is due to the interleaved imaging technique used to acquire this sequence.



References: [1] Erasmus J, Gladish G, *et al.*, JCO, pp. 2574 – 2582, 2003. [2] Ashton E, Takahashi C, *et al.*, JMRI, pp. 300 – 308, 2003. [3] Ashton E, Totterman S, *et al.*, IEEE-CBMS, pp. 301 – 305, 2001. [4] Caldwell C, Mah K, *et al.*, Int. J. Rad. Onco. Biol. Phys., pp. 923 – 931, 2001. [5] Werner-Wasik M, Xiao Y, *et al.*, Int. J. Rad. Onco. Biol. Phys., pp. 56 – 61, 2001. [6] Therasse P *et al.*, JNCI 92(3), 2000.



In addition to allowing a more accurate and reproducible measurement of RECIST and WHO criteria, semi-automated tumor segmentation makes it possible to measure tumor volume with high reproducibility.