A PILOT STUDY EVALUATING PULMONARY NODULE MARKING METHODS

A. M. Biancardi[a], A. P. Reeves[a], D. F. Yankelevitz[b], D. Ghiorghiu[c], M. Scott[c], H. Mann[c]

(a) School of Electrical and Computer Engineering, Cornell University, Ithaca, U.S.A.
(b) Weil Cornell Medical College, Cornell University, New York, U.S.A.
(c) AstraZeneca, Alderley Park, Macclesfield, U.K.

ABSTRACT

Assessing the precision in the estimation of lesion dimensions is a prerequisite for the determination of growth rates and response to therapy both in clinical practice and research. An initial study was designed and performed to evaluate three different marking methods: uni-dimensional (maximum diameter of nodule in-axial plane), manual volumetric and a computer assisted mark-up (CAM) method. The CAM method has a good level of agreement with the manual method. Additionally, the CAM method is more repeatable than both the manual volumetric and uni-dimensional measures (CAM 95% Limits of Agreement (LoA):[-15.8, 21.2], manual 95% LoA [-23.4, 31.8], uni-dimensional 95% LoA [-43.8, 80.2]) so this study supports the expectation that more reproducible measurements can be made by using a computer assisted method compared to standard manual methods.

Index Terms— response to therapy, diagnosis, volumetric measurements

1. INTRODUCTION

In clinical practice and research the quantification of precision in the estimation of tumor growth rates is fundamental in assessing tumor changes over the time or during the treatment. Precision, being a quantification of measurement repeatability, concerns the dependability of the estimate: a higher precision[1] assures greater stability for the result. When evaluating size changes, precision affects the minimum bounds for the time intervals between the measurements used to estimate growth: a higher precision allows shorter follow-up times.

With the continuous increase in CT scan resolution and the corresponding increase in image information, the radiologist faces the inevitable problem regarding the increased time taken to perform measurements, particularly volumetric measurements where each axial image must be evaluated for correctness of the segmentation or drawing. Current methods, based on automated or semi-automated segmentation of the nodules, are not yet able to accurately identify the actual extent of the nodule in the case of many large complex lesions and, therefore, manual lesion marking is the best

---

[1] Assuming enough sensitivity to detect the changes.

---

available option in those cases. Given this constraint, while volumetric-based estimates have the potential to be more robust than uni-dimensional (maximum diameter of nodule in-axial plane) or bi-dimensional (product of the maximum diameter in-axial plane with its maximum perpendicular diameter) estimates, they can dramatically increase user effort while still having limited success in minimizing user subjectivity. This limitation has been addressed by the use of a computer assisted mark-up (CAM) method. The addition of this new marking technique has broadened the spectrum of usable marking methods to four, in chronological order: the measurable bi-dimensional components of the method proposed by the World Health Organization (WHO) [1], the measurable uni-dimensional component proposed by the Response Evaluation Criteria In Solid Tumors (RECIST) working group [2], the completely manual volumetric method and the volumetric CAM. Current clinical practice focuses on the use of the RECIST criteria for the assessment of response using imaging, with much less emphasis on the WHO criteria, hence this study will concentrate on the comparison of the uni-dimensional measures to the volumetric measures, and the bi-dimensional measure will not be considered in this report.

The task of assessing the relative merits of each of these methods needs to address measurement reproducibility, intra- and inter-reader variability and evaluate the level of agreement between these different methods; the analysis, as the comparison of size metrics as shown [3], has to acknowledge also the presence of two quite different aspects of marking methods that impact the outcomes: the marking act itself and the approximations introduced when estimating the size. Available studies, in the context of pulmonary nodules and involving manual markings from multiple readers, have described either the differences between an automated segmentation approach and fully manual methods [4] or between the bi- and uni-dimensional methods [5]. In this paper we report data obtained from an initial study that was designed and conducted to investigate some key aspects of those marking methods and also to create an initial basis for more in-depth studies.

2. MATERIALS AND METHODS

The data set for this pilot study was comprised of twelve pulmonary nodules visible on thoracic CT scans. All of the scans, selected from the publicly accessible database of the Prevent Cancer Foundation, hosted at Cornell University, were acquired from multi-detector row CT scanners with pixel size ranging from 0.586 to 0.795 mm (average 0.694 mm, median 0.703 mm), axial resolution and slice thickness of 1.25 mm, tube current and voltage were 80 mA and 120 kVp, respectively, for all of the scans. Except for one scan acquired on a GE LightSpeed QX/i, all the other scans were acquired on a GE LightSpeed Ultra.

The study was divided into four phases, in which five experienced readers delineated, using an in-house developed system, the pulmonary nodules four times: phases one and two using the completely manual volumetric method preceded by a measurement of the maximum diameter in-axial plane and its maximum perpendicular diameter phases three and four using the computer assisted volumetric method preceded by a measurement of the maximum diameter in-axial plane and its maximum perpendicular diameter Phases two and four were essentially repeats of phases one and three, respectively, in order to be able to assess repeatability. Nodules were presented in pre-computed random orders that were different with respect to both readers and phases and a minimum time of one day was required between the end of one phase and the beginning of the following one. Readers were given detailed instructions concerning the study, the system, the marking process, and they were also expected to perform a self-evaluated training phase before performing phases one and three of the pilot study. The marking tool had magnification and windowing capabilities, although the readers were expected to keep the initial window and level settings of 1500 HU and -500 HU, respectively.

All the markings were processed to determine nodule size estimates; as the study involved size metrics belonging to two different scale spaces, the uni-dimensional measurements were also expressed as volumes. For volumetric boundaries, both manual and computer-assisted ones, the inner regions in each of the axial slices were rasterized and volumes were determined by counting the number of nodule pixels and then multiplying their sum by the voxel volume [6]; this method is frequently used in CAD/CADx tools. For one-dimensional mark-ups, the estimated volume v was assumed to be the volume of the sphere whose diameter length d was the measured value: $v = \frac{4}{3}\pi(\frac{d}{2})^3$. Agreement between each pair of marking methods (uni-dimensional vs. manual volumetric, uni-dimensional vs. CAM, CAM vs manual volumetric) was assessed using the Bland-Altman [7] 95% limits of agreement (LoA) for repeated measurements, using the logarithms of estimated volumes over the ten-mark-ups (i.e. two mark-ups from each of five readers) of each of the 12 nodules. Repeatability of mark-ups for each of the three methods was also as-

sessed using the Bland-Altman 95% LoA. All the LoA results are expressed as relative difference percentages. Lower and upper confidence bounds (LaUCB) for the lower and upper LoA were also calculated.

3. RESULTS

The median volumes from the volumetric markings ranged between 50.1 and 21347.0 mm^3 (mean 4607.6 mm^3, median 1564.1 mm^3).

The 95% LoA between uni-dimensional and the manual method were [-45.1, 217.8] (LaUCB: -52.5, 265.1), between uni-dimensional and CAM method were [-33.2, 250.2] (LaUCB: -41.4, 299.2), and between the manual and CAM methods were [-44.0, 35.0] (LaUCB: -47.8, 44.7). On average the volume derived from the uni-dimensional measure was 32% greater than the volume determined by manual delineation, and 53% higher than that of the CAM method. On average, the CAM volume was 13% smaller than the manual volume. Figure 1 is a graphical display of the Bland-Altman agreement analysis for marking method pairs.

The 95% LoA between repeated volumetric measurements for CAM were -15.8 to 21.2 (LaUCB: -19.1, 26.3) compared with 95% LoA of -23.4 to 31.8 (LaUCB: -27.9, 40.1) for the manual method and -43.8 to 80.2 (LaUCB: -48.8, 97.6) for uni-dimensional. Figure 2 is a graphical display of the Bland-Altman agreement analysis between repeated markings of the same method: (a) uni-dimensional, (b) manual volumetric, and (c) CAM volumetric.

4. DISCUSSION

The results show that, on average, estimated volumes derived from uni-dimensional measurements are substantially greater than the volumetric estimates from both the manual and CAM volumetric methods Moreover, the very wide limits of agreement underline the amount of approximation introduced by assuming that the nodule is spherical. Nodule shape influences the discrepancies between the two approaches: for example, for case AS0005, presented in Figure 3, the main axis of the nodule ellipsoidal shape is almost perpendicular to the axial direction and therefore the uni-dimensional measurement over-estimates the volume with respect to the other methods; on the other hand, when the main axis of the nodule ellipsoidal shape is parallel to the axial direction, as is the case for AS0003 displayed in Figure 4, the uni-dimensional-based estimations tend to be smaller then the volumetric ones. This behavior was already noticed previously [3] in the context of a size metric analysis; the outcomes of this study are nevertheless important because in this case the uni-dimensional measurements were performed by the readers, whereas in the other study they were derived from the readers' manual volumetric markings.
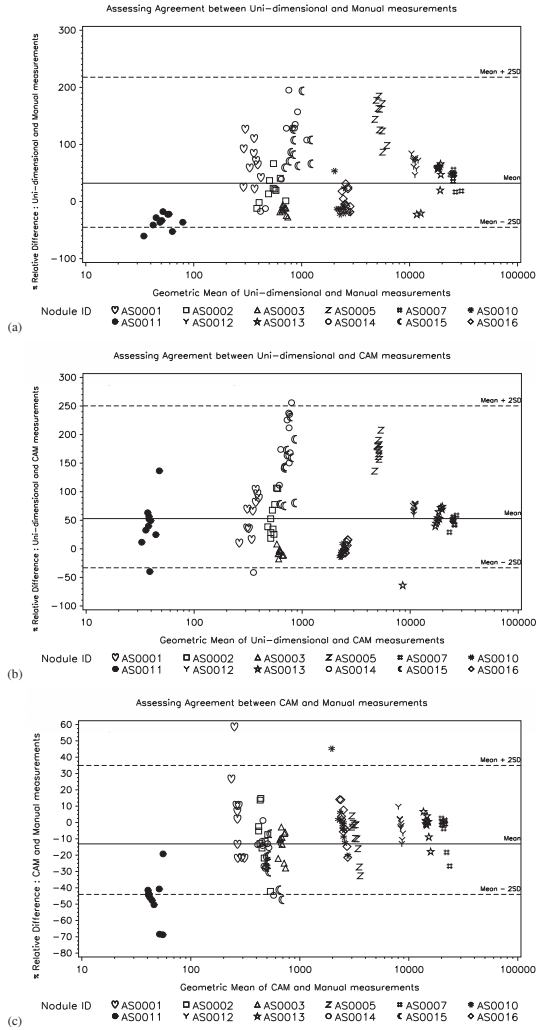
Fig. 1. Bland-Altman plots of agreements in method pairs. Notice that different intervals on the y axis had to be used due to the different distributions of the results.

Fig. 2. Bland-Altman plots of reader repeatability within the same method. Notice that different intervals on the y axis had to be used due to the different distributions of the results.

The Bland-Altman analysis of the agreement between the manual and CAM methods shows a fair agreement with a systematic tendency of CAM markings to be slightly smaller than the manual one. This was in part expected because manual markings (both uni-dimensional and volumetric) tend to suffer from over-estimation of the measurement due to readers giving the benefit of the doubt in areas of high uncertainty in the position of the nodule boundary. The computer assistance aims at minimizing those effects, by providing an algorithmic positioning of the boundary in such areas.

As regards precision, for all three methods it can be seen that average of the relative percentage difference across the repeated measures is very close to zero, indicating that there was no systematic difference in the mark-ups between the repeat phases of the experiments. Uni-dimensional measurements showed an unexpectedly high level of disagreement.
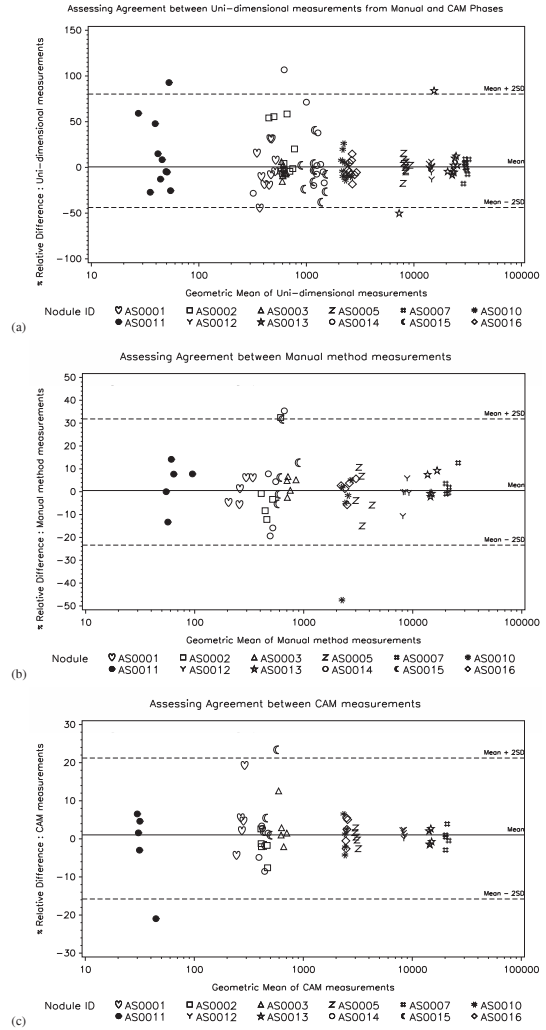
Excluding case AS0011 where the smaller size amplifies the relative differences, two factors seems to affect these measurements: the unavoidable presence of the partial-voxel effect and the higher noise level of low-dose scans, compared to standard dose ones. These factors create areas of uncertainty where the maximal-diameter extremes can be placed with unchanged soundness of the measuring act and yet with profound effects on the measure itself.

The CAM method had the tightest 95% LoAs: [-15.8, 21.2]. This means, for instance, that, if we consider a 10 mm lesion, then the 95% LoA on its size in mm would be [8.25, 12.17] for the uni-dimensional method and [9.15, 10.96] for the manual method; when using the CAM method the 95% LoA, in mm, would be reduced to [9.44, 10.66].
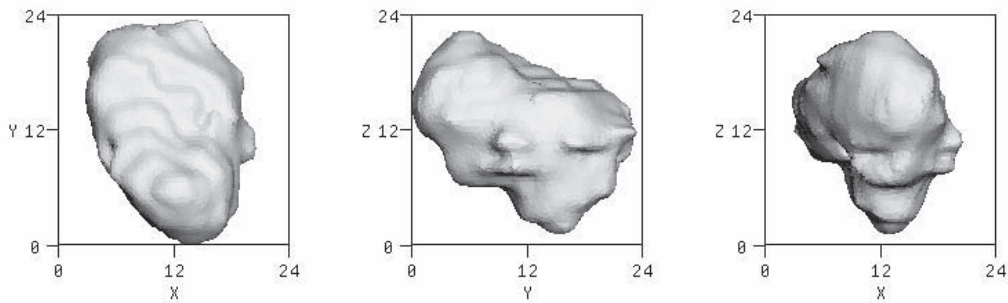
Fig. 3. Three-dimensional rendering of nodule AS0005, computed from a representative mark-up.
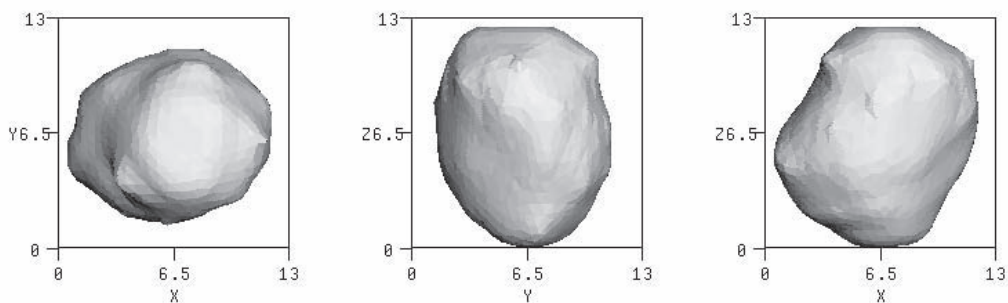

Fig. 4. Three-dimensional rendering of nodule AS0003, computed from a representative mark-up.

5. CONCLUSIONS

The viability of using a computer assisted method for deriving volumetric measurements was shown in this study by a good level of agreement with the manual method. Additionally, as superior limits of agreement were observed for the computer assisted method compared to the totally manual volumetric method and uni-dimensional method, this study supports the expectation that more reproducible measurement can be made by using a computer assisted method compared to standard manual methods.

6. REFERENCES

[1] "WHO handbook for reporting results of cancer treatment," World Health Organization Offset Publication No. 48, 1979.

[2] P Therasse, SG Arbuck, EA Eisenhauer, J Wanders, RS Kaplan, L Rubinstein, J Verweij, M Van Glabbeke, AT van Oosterom, MC Christian, and SG Gwyther, "New guidelines to evaluate the response to treatment in solid tumors," *J. Natl. Cancer Inst.*, vol. 92, no. 3, pp. 205–216, February 2000.

[3] Anthony P. Reeves and Alberto M. Biancardi et al., "The lung image database consortium (LIDC): A comparison of different size metrics for pulmonary nodule measurements," *Academic Radiology*, vol. 14, no. 12, pp. 1475–1485, Dec 2007.

[4] Katharina Marten, Florian Auer, Stefan Schmidt, Gerhard Kohl, Ernst J. Rummeny, and Christoph Engelke, "Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria," *European Radiology*, vol. 16, no. 4, pp. 781–790, April 2006.

[5] Jeremy J. Erasmus and Gregory W. Gladish et al., "Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: Implications for assessment of tumor response," *Journal of Clinical Oncology*, vol. 21, no. 13, pp. 2574–2582, 2003.

[6] R. S. Breiman, J. W. Beck, M. Korobkin, R. Glenny, O. E. Akwari, D. K. Heaston, A. V. Moore, and P. C. Ram, "Volume determinations using computed tomography," *American Journal of Roentgenology*, vol. 138, no. 2, pp. 329–333, 1982.

[7] J. M. Bland and D. G. Altman, "Measuring agreement in method comparison studies.," *Stat Methods Med Res*, vol. 8, no. 2, pp. 135–160, June 1999.