

Consistency and Efficiency of CT Analysis of Metastatic Disease: Semiautomated Lesion Management Application Within a PACS

Les R. Folio¹
 Aline Sandouk¹
 Jiaxin Huang¹
 Jeffrey M. Solomon^{1,2}
 Andrea B. Apolo³

Keywords: metastatic cancer, quantitative imaging, RECIST, tumor assessment, volumetric measurements

DOI:10.2214/AJR.12.10136

Received October 1, 2012; accepted after revision December 20, 2012.

Supported in part by the Intramural Research Program of the National Institutes of Health.

Presented in preliminary form at the 2012 annual meeting of the RSNA, Chicago, IL.

¹Department of Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, 10 Center Dr, Bldg 10, Rm 1C340 Bethesda, MD 20892. Address correspondence to L. R. Folio (les.folio@nih.gov).

²Expert Image Analysis LLC, Potomac, MD.

³Medical Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD.

AJR 2013; 201:618–625

0361–803X/13/2013–618

© American Roentgen Ray Society

OBJECTIVE. The purpose of this study was to evaluate the success, consistency, and efficiency of a semiautomated lesion management application within a PACS in the analysis of metastatic lesions in serial CT examinations of cancer patients.

MATERIALS AND METHODS. Two observers using baseline and follow-up CT data independently reviewed 93 target lesions (17 lung, five liver, 71 lymph node) in 50 patients with either metastatic bladder or prostate cancer. The observers measured the longest axis (or short axis for lymph nodes) of each lesion and made Response Evaluation Criteria in Solid Tumors (RECIST) determinations using manual and lesion management application methods. The times required for examination review, RECIST calculations, and data input were recorded. The Wilcoxon signed rank test was used to assess time differences, and Bland-Altman analysis was used to assess interobserver agreement within the manual and lesion management application methods. Percentage success rates were also reported.

RESULTS. With the lesion management application, most lung and liver lesions were semiautomatically segmented. Comparison of the lesion management application and manual methods for all lesions showed a median time saving of 45% for observer 1 ($p < 0.05$) and 28% for observer 2 ($p = 0.05$) on follow-up scans versus 28% for observer 1 ($p < 0.05$) and 9% for observer 2 ($p = 0.087$) on baseline scans. Variability of measurements showed mean percentage change differences of only 8.9% for the lesion management application versus 26.4% for manual measurements.

CONCLUSION. With the lesion management application method, most lung and liver lesions were successfully segmented semiautomatically; the results were more consistent between observers; and assessment of tumor size was faster than with the manual method.

Therapeutic response in cancer patients is determined by manual assessment of target lesions according to the international Response Evaluation Criteria in Solid Tumors (RECIST). These criteria are based on percentage change in target lesions from baseline to follow-up on images, primarily CT scans [1–3]. Obtaining timely, accurate, and consistent target measurements is a major challenge. Moreover, the process of selecting a target lesion is subject to interpretation and is limited by the number of target lesions [2, 4]. Optimally, therapeutic response should include the greatest number of metastatic lesions in a patient [5]. In current manual protocols, however, analysis of additional lesions beyond the limit set by the RECIST criteria would be impractical. Hence, manual assessment is usually limited to a few target lesions that are followed over time. As an alternative, automated localization, registration,

and recognition algorithms have been applied to organ-specific lesions. With these algorithms, neighboring regions have been successfully aligned and the same lesions have been identified in serial CT studies [6]. Lesion registration and segmentation methods have been described [7], and automated measurement methods are beginning to emerge for characterization of lesions on serial CT images [8, 9].

Our PACS contains a semiautomatic lesion management application that facilitates detection, measurement, and data capturing of target lesions. The features of this tool have potential for remedying the limitations inherent in manual tumor measurement, including subjectivity, lack of consistency, and time consumption. We are evaluating the consistency and efficiency of assessments of all lung, liver, and lymph node lesions with the lesion management application. Lymph node seg-

CT of Metastatic Disease

mentation was not available at the time of this study, but features of the application assist in measurement recording, RECIST calculation, and data input. For lung and liver lesions, the lesion management application assists with measurement recording, RECIST calculation, data input, and lesion segmentation. The purpose of this study was to assess interobserver agreement for manual and lesion management application methods and determine percentage time saved over the conventional manual method.

Materials and Methods

Patients

The CT scans of 50 patients enrolled in three institutional review board–approved studies were retrospectively evaluated. The inclusion criteria were that the patient had undergone baseline and follow-up CT examinations and the presence of soft-tissue metastatic lesions. All patients had undergone routine treatment in the studies. Nine patients with urothelial cancer and 16 patients with castration-resistant prostate cancer were enrolled in a prospective treatment protocol with the single-agent monoclonal antibody TRC105 (anti-CD105 [endoglin]) [10]. Twenty-five patients with castration-resistant prostate cancer were enrolled in a phase 2 study of bevacizumab, lenalidomide, docetaxel, and prednisone [11]. A total of 93 target lesions were analyzed: 17 lung, five liver, and 71 lymph node lesions. In primary studies all lesions were identified as target lesions according to RECIST 1.1 (lung and liver lesions ≥ 1 cm in the long axis and lymph node ≥ 1.5 cm in the short axis).

Imaging

Baseline and follow-up CT scans were obtained with any of the following scanners depending on availability: Definition (Siemens Healthcare), Brilliance 16 (Philips Healthcare), or LightSpeed (GE Healthcare). The detector collimation for primary axial acquisition was 0.6–2.5 mm, and the pitch was 1.078–1.35, depending on the vendor. Though techniques varied depending on CT vendor, they were generally limited to 120 kVp at 200–300 effective mAs with rotation times of 0.6–0.75 second. Portal venous phase images were obtained 70 seconds after initiation of IV contrast administration at 2 mL/s with a power injector. The dose of contrast material was based on body mass index and ranged from 90 to 130 mL.

Observer Measurements

Observations were made with the lesion management application in a Carestream Vue PACS 11.4 (Carestream Health). Patients were randomly distributed into two groups. Two postbaccalau-

reate students with no formal medical training served as study observers and independently reviewed target lesions selected by the patients' primary study teams and radiologist. A radiologist with more than 15 years' experience in body CT verified all measurements and segmentations at the completion of the observations (not included in the timing). Observers were timed while identifying, measuring, and recording determinations using the manual and lesion management application methods for each patient, as described by Haygood et al. [12]. Measurements were made first for the baseline examination and then for the follow-up examination over several sessions. The observers alternated groups and methods to minimize training and recall biases (Table 1).

In the manual protocol, each observer used standard electronic manual linear calipers of the PACS workstation to determine the longest axis of each lesion (short axis for lymph nodes) on a single axial CT image. An observer made a measurement by clicking first on one end of the axis and then on the other end of the perceived axis. The measurements were recorded on the National Cancer Institute RECIST form for subsequent manual calculations. Key images of target lesions, which were screen shots of pertinent axial images, were sometimes available in the patient's electronic records and helped observers identify the lesions of interest in baseline studies. CT series and image numbers were available to observers for all target lesions assessed.

In the lesion management application protocol, observers followed the same general steps but with the assistance of the semiautomated software. When measuring lesions, observers had to neither select

the single axial plane with the longest axis nor click on the two ends of the axis. Instead, they clicked on any part of the lesion, prompting the software to segment and determine the longest axis, shortest axis, and volume. When measuring liver lesions, observers clicked on the two ends of the lesion and the algorithm completed the measurement but could do so in any single axial plane. Appendix 1 summarizes the manual and lesion management application methods for baseline and follow-up examinations. The lesion management application does not support lymph node segmentation as it does segmentation of lung and liver lesions. However, the measurements are directly imported into the PACS RECIST report, saving data transfer time.

Lesion Management Application Features

The anatomic registration function entails a volumetric voxel-based algorithm similar to methods described by Hawkes [13]. Such algorithms minimize square differences between datasets [7, 14, 15]. The advanced serial coregistration capability of the application allowed overall alignment of anatomic axial images and refined alignment for specific tumors relative to the baseline examination. Registration between examinations replaced the conventional task of manually scrolling through images from the individual examinations to align them visually on the basis of interpretation of the anatomic features, a key time-saving feature [16].

The organ-specific lesion segmentation tools of the application entailed proprietary algorithms conceived for detection of abnormal tissue in specific organs. Similar segmentation tools have been reported that entail enhanced tumor-edge detection and other methods of differentiating tumors

TABLE 1: Study Design Alternating Manual and Lesion Management Application Methods of Tumor Measurement

Examination	Method	Observer 1	Observer 2
Baseline	Manual	Group 1	Group 2
Baseline	Lesion management application	Group 2	Group 1
Interval of 2–3 d			
Baseline	Manual	Group 2	Group 1
Baseline	Lesion management application	Group 1	Group 2
Interval of 10 d			
Follow-up	Manual	Group 1	Group 2
Follow-up	Lesion management application	Group 2	Group 1
Interval of 2–3 d			
Follow-up	Manual	Group 2	Group 1
Follow-up	Lesion management application	Group 1	Group 2

Note—Study design was intended to minimize training and recall bias in comparisons of manual and automated methods of tumor measurement. However, longer and shorter intervals between baseline and follow-up did occur owing to equipment and observer availability, reflecting real-world logistic difficulties and obstructing strict adherence to the stated design.

from surrounding structures [17–20]. Observers engaged the appropriate lesion tool (lung or liver), which outlined the perimeter of the selected lesion and produced axial and volumetric measurements (Fig. 1).

A bookmarking function allowed rapid recall of the slice bearing a lesion of interest along with longest axis, shortest axis, and volumetric measurements. Observers retrieved bookmarks and used them to assist in the tumor measurements on images obtained at follow-up examinations. Specifically, selecting the bookmarks on baseline images automatically prompted recall to the equivalent axial slice on follow-up images. The bookmarking feature stored identification information for selected lesions in electronic records. This feature eliminated the need to retrieve measurements of selected lesions from previous CT examinations. The lesion-pairing feature allowed observers to match the lesions on images from baseline and follow-up examinations, precluding confusion about which follow-up lesions corresponded to the original baseline lesions. This feature was particularly useful for patients who had multiple lesions of similar dimensions in the same location, which made the lesions practically indistinguishable if they shared the same slice numbers and anatomic descriptions on the RECIST form.

Lesion pairing afforded observers the option of electronically matching the baseline and follow-up measurements of the same lesion. They used the software to calculate percentage change in the longest axis (short axis for lymph nodes). The percentage change and image slice number of all lesions were exported into a Microsoft Excel 2007 spreadsheet, replacing the need for observers to manually record the information on RECIST forms, saving time and preventing transcription errors.

The lesion-tracking tool combined bookmarking and pairing features into one algorithm so that tumor measurement and analysis on the images from the follow-up examination were performed in a single step rather than multiple individual ones. This tool has been recently described by Folio et al [21]. For examinations in which the lung or liver lesion segmentation tools were used, observers would engage the lesion-tracking tool after registering images from the baseline and follow-up examinations. The lesion-tracking tool automatically located the lesions of interest on the images from the follow-up examination corresponding to the lesions targeted by observers on the baseline images and calculated the percentage change in the lesion between the two examinations. If the lesion was successfully located and segmented and the change in tumor size successfully calculated on the images from the follow-up examinations, use of the lesion-tracking tool was considered successful, and

the data were exported to the spreadsheet. If the lesion-tracking tool failed in any of these components, use of the tool was considered unsuccessful, and observers reverted to the individual manual application of segmentation or pairing steps of the lesion management application protocol.

Success Rates

The success rates with the lesion segmentation and lesion-tracking tools of the lesion management application were calculated only for lung and liver lesions. Segmentation with the application was considered successful if the software correctly outlined the longest axis, shortest axis, and volume of the lesions without oversegmenting, undersegmenting, or altogether failing to segment the lesion. The segmentation success rate with the lesion management application was defined as the number of successfully segmented lesions divided by the total number of lesions to which lesion management application segmentation was applicable in either the lung or the liver. Lesions were either segmented (successful) or not (pop-up box stated “unsuccessful”), a binary process. Resultant measurements by each observer were compared as described in Statistical Methods.

Application of the lesion-tracking tool was considered successful if the tool correctly located and segmented lesions on the follow-up images and calculated the change in tumor size between examinations. The success rate with the lesion-tracking tool was defined as the number of lesions with successful lesion-tracking tool application divided by the total number of lesions to which the lesion-tracking tool was applicable. The lesion-tracking tool was appropriate for use only in follow-up examinations with successful lung and liver segmentation on baseline images.

Statistical Methods

Bland-Altman analysis [22] was used to assess interobserver agreement by comparison of differences in percentage change in lesion measurements between two observers [23, 24] using the manual and lesion management application methods. The distribution of percentage change differences was plotted. The Wilcoxon signed rank test was used to assess the statistical significance of time saved per patient. Median, mean, and range were calculated for each observer and for time differences between the lesion management application and manual methods and for baseline and follow-up studies between observers.

Results

Time Saving

In the assessment of target lesions on CT images obtained at baseline examinations,

the median time saving was 28% (96 seconds with the manual method and 69 seconds with the lesion management application) for observer 1 ($p < 0.05$) and 9% (69 versus 63 seconds) for observer 2 ($p = 0.087$) (Table 2). On follow-up CT images, there was a 45% median time saving (median manual, 170 seconds; median automatic, 93 seconds) for observer 1 ($p < 0.05$, Wilcoxon signed rank test) and 28% (122 vs 88 seconds) for observer 2 ($p < 0.05$).

Success Rates

The success rates of lesion management application segmentation by observers 1 and 2 on CT images from baseline examinations were 88% and 65% of lung lesions ($n = 17$) and 60% and 100% of liver lesions ($n = 5$). On images from follow-up examinations, these rates were 88% and 71% of lung lesions for observers 1 and 2 and 100% of liver lesions for both observers. Figure 2 shows the success rates of the observers on baseline and follow-up images of both lung and liver lesions. Figure 3 shows the time saved with the lesion management application versus the manual method. Segmentation on follow-up images was approximately twice as fast with the lesion management application.

Interobserver Variation

We assessed the variability of measurement results between the two observers using Bland-Altman analysis, which showed mean percentage change differences of 26.4% (95% CI, -106.0% to 158.85%) for manual measurements and 8.9% (-84.9% to 102.62%) for the lesion management application measurements on images from baseline and follow-up CT examinations (Fig. 4). Notably, seven lesions had nearly zero interobserver variation with the lesion management application.

Discussion

In this study, the lesion management application method was more consistent and efficient than the traditional manual method of tumor lesion identification and measurement, RECIST calculation, and data input. Some of the features that saved time were anatomic registration, lesion pairing, automatic identification and measurement of lesions on follow-up examinations, and bookmarking.

The anatomic registration feature of the lesion management application expedited location of target lesions on the images from follow-up examinations. The organ-specific lesion

TABLE 2: Mean and Median Times for Each Observer for Baseline and Follow-Up Examinations for Both Manual and Automated Methods

Parameter	p^a	Mean Time (min:s)		Median Time (min:s)		Median Difference (min:s)
		Manual	Automated	Manual	Automated	
Observer 1						
Baseline	< 0.05	1:51	1:24	1:32	1:09	0:23
Follow-up	< 0.05	3:10	1:50	2:50	1:33	1:07
Observer 2						
Baseline	0.087	1:23	1:12	1:09	1:02	0:08
Follow-up	< 0.05	2:26	1:48	2:02	1:29	0:33

Note—The only time saving that was not statistically significant was for observer 2 in the baseline assessment ($p = 0.087$).

^aBy Wilcoxon signed rank test.

segmentation tool replaced subjective estimation of lesion axes, which can be difficult for lesions with ill-defined borders or insufficient contrast between diseased and normal tissue.

Our findings show the possibility of close-to-zero variation between observers (Fig. 4B) using the lesion management application as opposed to the manual method. Improved consistency among observers may lead to more accurate quantification of tumor burden and assessment of treatment response.

Use of the bookmarking feature eliminated the need to retrieve measurements of selected lesions from previous CT examinations, and the results show that use of the lesion-tracking tool was successful in most of the patients with lung and liver lesions. Lymph node segmentation is not available, and therefore the lesion-tracking tool cannot be applied to assessment of lymph nodes. However, other features were applicable for lymph node assessments, such as measurements on key images for which bookmarking and lesion pairing were used across serial examinations and contributed to the overall time saving.

The lesion management application is housed directly in the PACS and can be used by any clinical personnel. An advantage of having the algorithm in the PACS is that providers are freed from reliance on third-party software. Therefore, reading efficiency is enhanced, and it becomes more likely that providers will integrate the semiautomated measurements within their workflow. Because lesion measurements are saved in the PACS and exported into a spreadsheet with tumor size calculations, the lesion management application enhances the flow of information between the tumor measurement and the RECIST form with less opportunity for data transfer error. See figure 5 for a comparison of manual versus automated steps,

avoiding the need for manually filling out a paper form.

Practical Applications

The time saved shows the benefit of the lesion management application over a manual approach to lesion detection and may translate into the ability to perform faster analysis and generate a greater number of radiology reports that include results of quantitative tumor analysis. The automatic lesion segmentation, lesion pairing, and calculation of tumor size possible with the lesion management application are especially useful for radiology practices and departments in which tumor measurements are often performed by nonradiologists or physicians in training and in which multiple radiologists share the diagnostic review. Our institution and many others employ such valuable support staff to allow radiologists to focus on identifying lesions and reporting them, not measuring every lesion [25–28].

Radiology reports typically record tumor burden with language such as “improved,” “similar,” or “worse,” sometimes accompanied by lesion measurements and comparisons. In a retrospective review [29], only 26% of radiology reports were found to be sufficient for calculating quantitative response rate. A previous survey of 565 abdominal radiologists at 55 National Cancer Institute–funded cancer centers [30] showed that 86% of the respondents would include tumor size measurements in their dictation reports if the PACS could calculate tumor measurements with a click of the mouse. Tumor measurements would allow more objective evaluation of tumor analysis than would qualitative assessment.

Limitations

This study had several limitations. It was a pilot study in which a small number of ap-

plicable lung and liver lesions were evaluated. The lack of segmentation capability for lymph nodes was a major limitation. The time saving in the target lesion assessments was captured in a patient-based analysis. Patients had a combination of lung, liver, and lymph nodes metastases, limiting full interpretation of the time saving obtained from full use of the lesion management application methods in lesions that could be segmented. Because none of the lesions in this study were surgically removed, comparison with actual lesion size was not performed. However, we tested the tools using “coffee-break” serial CT of simulated pliable lesions in a phantom described by Buckler et al. [31]. In addition, even for lung and liver lesions, manual methods were used in combination with the lesion management application when segmentation failed. Some segmentation failures in the liver were likely due to the effects of chemotherapy. Folio et al. [21] identified additional failures in metastatic melanoma that were caused by beam hardening, overestimation or underestimation of lesions (for example, adjacent vessels in lungs), ill-defined lesions, and tumors abutting adjacent tissue. They also found that when lesions undergo extreme shrinkage or growth over time, the lesion-tracking tool occasionally fails; however, these issues have been markedly improved in ongoing studies (Folio LR, unpublished data).

Another limitation was that there were only two observers. However, we considered it sufficient to evaluate agreement between two persons with similar levels of training and proficiency. Although the observers were not radiologists (all measurements were verified by a radiologist), strong interobserver correlation supports the use of physician extenders such as technologists to measure tumor

with radiologist or oncologist oversight. Currently at our institution, providers and radiologic technologists (with radiologist supervision) record the examination, slice numbers, and anatomic location of the lesion of interest on RECIST forms using the image and sequence numbers in radiologist reports. This process requires frequent reference to the form and manual scroll-through of the CT images to find the lesion in baseline and follow-up examinations. It also requires extensive and meticulous bookkeeping.

Refining of the lesion management application algorithms to include lesions of all organ sites, increasing availability, and validating our findings in larger cohorts should mitigate many of the limitations of this study. The integration of these tools into standard tumor assessments in the care of patients with metastatic cancer is crucial for evaluating the accuracy and efficiency of emerging therapies and thus aiding in drug development.

Conclusion

Segmentation of most lung and liver lesions with a semiautomated lesion management application method was more consistent between observers than was use of a manual method for all lesions and was significantly faster.

Acknowledgment

The authors thank writer-editor Carol Pearce for editorial support with this manuscript.

References

- Suzuki C, Jacobsson H, Hatschek T, et al. Radiologic measurements of tumor response to treatment: practical approaches and limitations. *Radiographics* 2008; 28:329–344
- Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 2000; 92:205–216
- National Cancer Institute Cancer Imaging Program website. Clinical trials: imaging response criteria. imaging.cancer.gov/clinicaltrials/imaging. Accessed June 12, 2012
- Chalian H, Töre HG, Horowitz JM, Salem R, Miller FH, Yaghmai V. Radiologic assessment of response to therapy: comparison of RECIST versions 1.1 and 1.0. *Radiographics* 2011; 31:2093–2105
- Schwartz LH, Mazumdar M, Brown W, Smith A, Panicek DM. Variability in response assessment in solid tumors: effect of number of lesions chosen for measurement. *Clin Cancer Res* 2003; 9:4318–4323
- Sofka M, Stewart CV. Location registration and recognition (LRR) for serial analysis of nodules in lung CT scans. *Med Image Anal* 2010; 14:407–28
- Maintz JB, Viergever MA. A survey of medical image registration. *Med Image Anal* 1998; 2:1–36
- Linguraru MG, Wang S, Shah F, et al. Automated noninvasive classification of renal cancer on multiphase CT. *Med Phys* 2011; 38:5738–5746
- Xu J, Greenspan H, Napel S, Rubin DL. Automated temporal tracking and segmentation of lymphoma on serial CT examinations. *Med Phys* 2011; 38:5879–5886
- Karzai FH, Apolo AB, Adelberg D, et al. A phase I study of TRC105 (anti-CD105 monoclonal antibody) in metastatic castration resistant prostate cancer (mCRPC). (abstract) *J Clin Oncol* 2012; 30(suppl):3043
- Adesunloye B, Huang X, Ning, YM, et al. Dual antiangiogenic therapy using lenalidomide and bevacizumab with docetaxel and prednisone in patients with metastatic castration-resistant prostate cancer (mCRPC). (abstract) *J Clin Oncol* 2012; 30(suppl):4569
- Haygood TM, Wang J, Atkinson EN, et al. Timed efficiency of interpretation of digital and film-screen screening mammograms. *AJR* 2009; 192:216–220
- Hawkes DJ. Algorithms for radiological image registration and their clinical application. *J Anat* 1998; 193:347–361
- Hutton BF, Braun M, Thurjell L, Lau DY. Image registration: an essential tool for nuclear medicine. *Eur J Nucl Med Mol Imaging* 2002; 29:559–577
- Knops ZF, Maintz JB, Viergever MA, Pluim JP. Normalized mutual information based registration using k-means clustering and shading correction. *Med Image Anal* 2006; 10:432–439
- Awad J, Owringi A, Villemare L, O'Riordan E, Parraga G, Fenster A. Three-dimensional lung tumor segmentation from x-ray computed tomography using sparse field active models. *Med Phys* 2012; 39:851–865
- Kuhnigk JM, Dicken V, Bornemann L, et al. Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. *IEEE Trans Med Imaging* 2006; 25:417–434
- Foo JL, Miyano G, Lobe T, Winer E. Tumor segmentation from computed tomography image data using a probabilistic pixel selection approach. *Comput Biol Med* 2011; 41:56–65
- Bauer S, Nolte LP, Reyes M. Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. *Med Image Comput Comput Assist Interv* 2011; 14:354–361
- Xu T, Mandal M, Long R, Basu A. Gradient vector flow based active shape model for lung field segmentation in chest radiographs. *Conf Proc IEEE Eng Med Biol Soc* 2009; 2009:3561–3564
- Folio L, Choi M, Solomon J, Schaub NP. Automated registration, segmentation and measurement of metastatic melanoma tumors in serial CT Scans. *Acad Radiol* 2013; 20:604–613
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1:307–310
- Durkalski VL, Palesch YY, Lipsitz SR, Rust PF. Analysis of clustered matched-pair data. *Stat Med* 2003; 22:2417–2428
- Gajewski BJ, Thompson S, Dunton N, Becker A, Wrona M. Inter-rater reliability of nursing home surveys: a Bayesian latent class approach. *Stat Med* 2006; 25:325–344
- Grossi F, Belvedere O, Fasola G, et al. Tumor measurements on computed tomography images of non-small cell lung cancer were similar among cancer professionals from different specialties. *J Clin Epidemiol* 2004; 57:804–808
- Hopper KD, Kasales CJ, Van Slyke MA, Schwartz TA, TenHave TR, Jozefiak JA. Analysis of interobserver and intraobserver variability in CT tumor measurements. *AJR* 1996; 167:851–854
- Monsky WL, Heddens DK, Clark GM, et al. Comparison of young clinical investigators' accuracy and reproducibility when measuring pulmonary and skin surface nodules using a circumferential measurement versus a standard caliper measurement: American Association for Cancer Research/American Society of Clinical Oncology Clinical Trials Workshop. *J Clin Oncol* 2000; 18:437–444
- Nguyen TB, Wang S, Anugu V, et al. Distributed human intelligence for colonic polyp classification in computer-aided detection for CT colonography. *Radiology* 2012; 262:824–833
- Levy MA, Rubin DL. Tool support to enable evaluation of the clinical response to treatment. *AMIA Annu Symp Proc* 2008; Nov6:399–403
- Jaffe TA, Wickersham NW, Sullivan DC. Quantitative imaging in oncology patients. Part I. Radiology practice patterns at major U.S. cancer centers. *AJR* 2010; 195:101–106
- Buckler AJ, Mulshine JL, Gottlieb R, Zhao B, Mozley PD, Schwartz L. The use of volumetric CT as an imaging biomarker in lung cancer. *Acad Radiol* 2010; 17:100–106

CT of Metastatic Disease

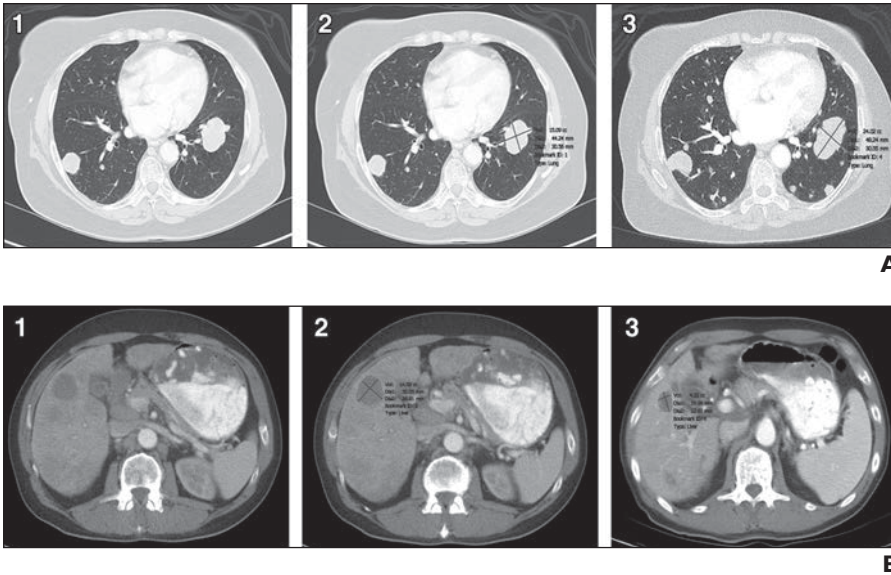


Fig. 1—Lesion segmentation in 66-year-old man with urothelial cancer. **A** and **B**, CT images show metastatic lung (**A**) and liver (**B**) lesions unmarked at baseline examinations (1), with axial diameters and segmentation marked with lesion management application (LMA) assistance at baseline examinations (2), and marked with LMA assistance at follow-up examinations (3). Examples show successful segmentation with LMA (automated identification of lesions at follow-up with resultant measurements including volumes, longest diameter, and shortest diameter). Enlargement of lung lesion and reduction in size of liver lesion are examples of size changes over time.

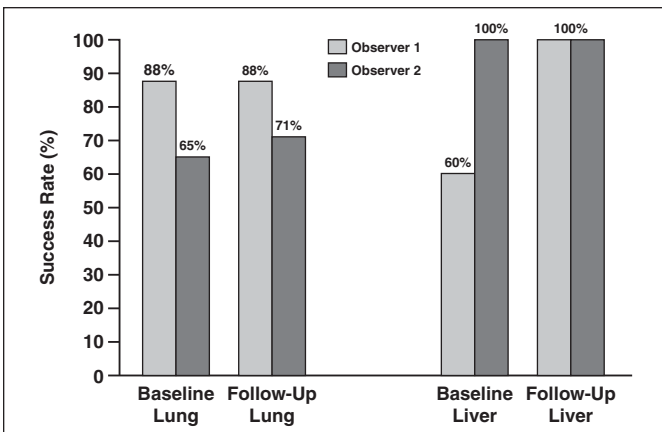


Fig. 2—Graph shows percentage segmentation with lesion management application for lung and liver lesions on CT images from baseline and follow-up examinations. For lung lesions ($n = 17$) on baseline, the success rates were 88% for observer 1 and 65% for observer 2. For follow-up lung lesions, success was 88% for observer 1 and 71% for observer 2. For baseline liver lesions ($n = 5$), 60% were successfully segmented by observer 1 and 100% for observer 2. On follow-up liver segmentations, all lesions were successfully segmented for both observers.

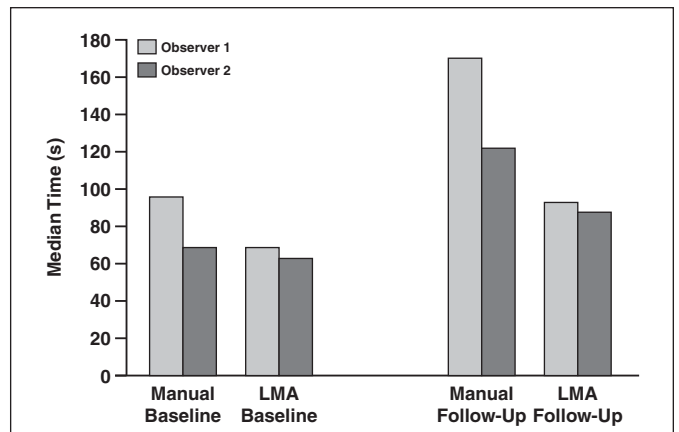


Fig. 3—Graph shows median time saving in lesion review with manual and lesion management application (LMA) methods. For follow-up images, median times were 170 seconds for manual method and 93 seconds for semiautomated method (–45%) for observer 1 and 122 versus 88 seconds for observer 2 (–28%). For baseline images, median times were 96 versus 69 seconds for observer 1 (–28%) and 69 versus 63 seconds for observer 2 (–9%). Time economy was evident for both baseline and follow-up examinations with use of LMA method but was more appreciable for follow-up.

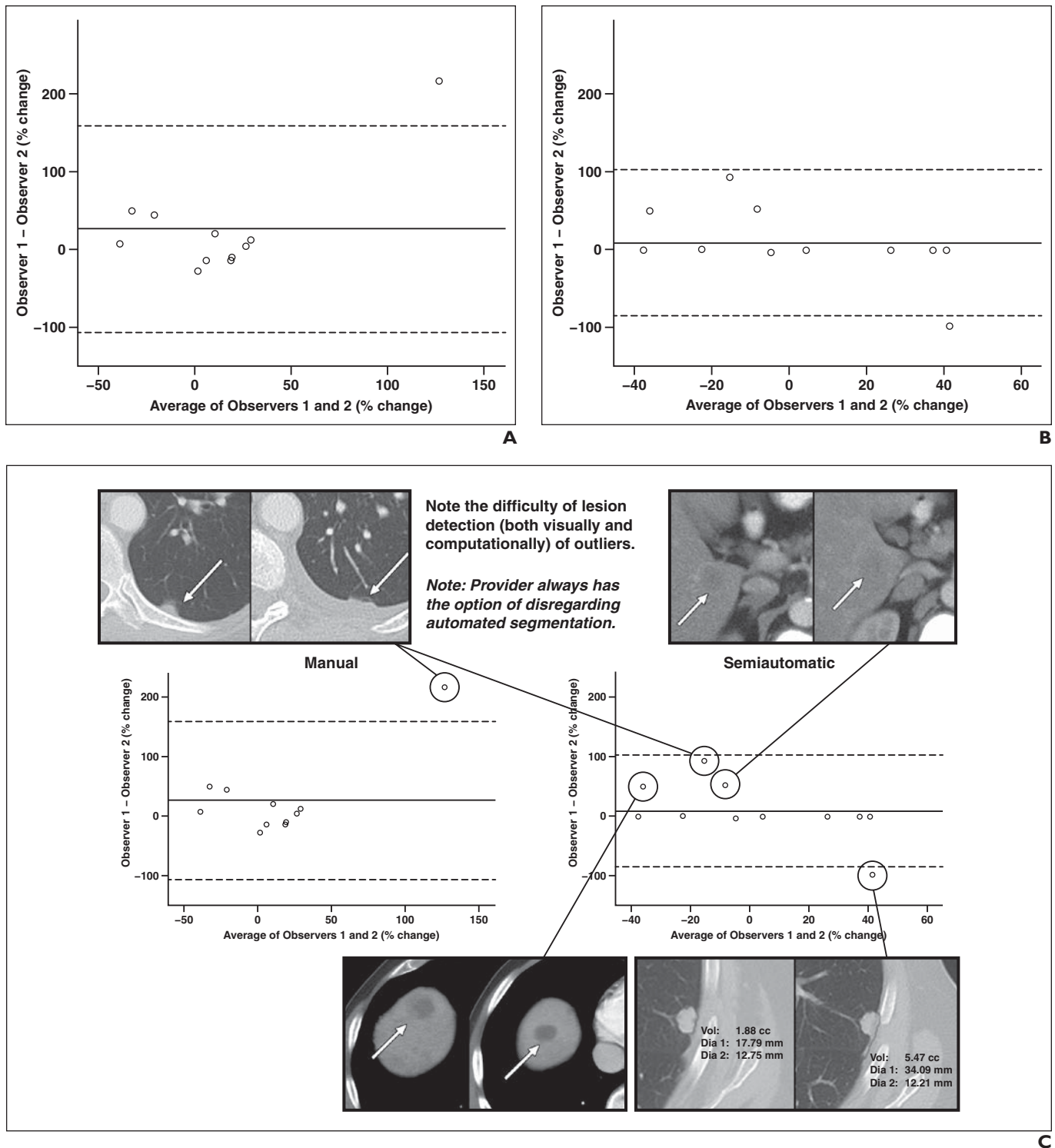


Fig. 4—Interobserver agreement.

A and **B**, Bland-Altman plots show stronger agreement with lesion management application (**B**) than manual method (**A**) with associated outlier lesions. Seven data points close to zero difference indicate strong correlation between observers using lesion management application. Data represent all liver and lung lesions that were successfully segmented ($n = 11$) on baseline and follow-up images by both observers. Mean percentage change differences were 26.4% (95% CI, -106.0% to 158.85%) for manual measurements and 8.9% (95% CI, -84.9% to 102.62%) for lesion management application.

C, Combination of Bland-Altman plots and CT images shows example segmentations that were not exactly zero with semiautomated lesion management application method (*right*). Plots show stronger agreement with the semiautomated method. Baseline (*left*) and follow-up (*right*) CT images show outlier lesions. Observers measured small juxtapleural lesion (*top left*) differently because of interval development of small pleural effusion.

CT of Metastatic Disease

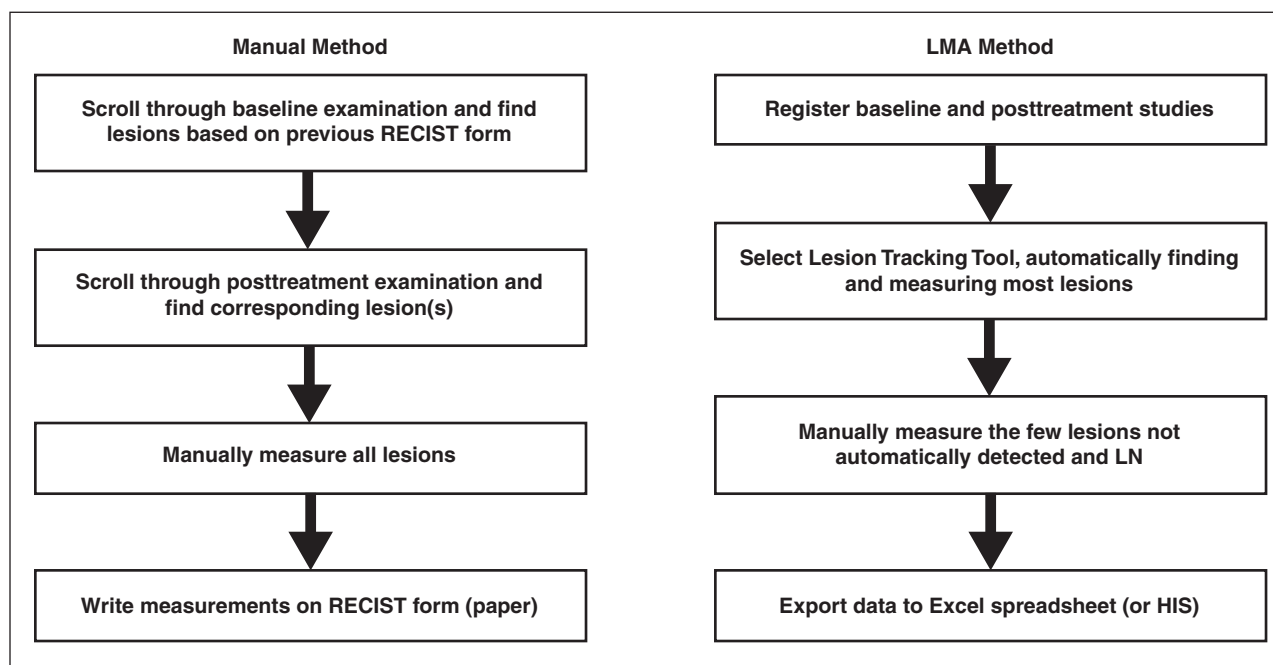


Fig. 5—Flowchart highlights differences in manual and automated steps applied in study. Manual method requires time-consuming scrolling through both baseline and follow-up examinations, whereas automated method finds most lesions previously measured and identifies and measures them as part of algorithm. Another time-saving step is exportation of data in which measurements are exported in digital format to Microsoft Excel for purposes of our study. Because data are exportable digital numbers, they can be exported directly into medical record or hospital information system (HIS). LMA = Lesion Management Application, RECIST = Response Evaluation Criteria in Solid Tumors, LN = lymph node.

APPENDIX I: Summary of Manual and Lesion Management Application Methods in Baseline and Follow-Up Examinations

Step	Manual		Lesion Management Application	
	Baseline	Follow-Up	Baseline	Follow-Up
1	Load baseline examination	Load baseline and follow-up examination	Load baseline examination	Load baseline and follow-up examination
2				Register baseline and follow-up examinations
3	Navigate to image slice with lesion of interest, as identified by oncologist ^a	Scroll through baseline examination to find lesion of interest	Navigate to image slice with lesion of interest, as identified by oncologist ^a	Open bookmark tab to locate slice with lesion of interest on baseline examination
4		Scroll through follow-up examination to find lesion of interest from baseline examination		Use software to automatically find equivalent slice on follow-up examination
5	Measure lesion of interest with caliper	Measure lesion of interest with caliper	Measure lung or liver lesion of interest with segmentation tool; measure lymph node with caliper	Apply follow-up tool if applicable. If not, measure lung or liver lesion of interest with segmentation tool, measure lymph node with caliper
6	Record measurements on RECIST form	Record measurements on RECIST form		Apply lesion pairing tool if follow-up tool was not applicable
7		Calculate percentage change in tumor size		
8		Record changes on RECIST form		Export measurement and calculation on Microsoft Excel spreadsheet

Note—Figure 5 shows flowcharts of steps taken by the observers to include CT images. RECIST = Response Evaluation Criteria in Solid Tumors.

^aOften the baseline lesions were identified by radiologists on key images, or the image slice with the lesion of interest was recorded on a sheet for the individual making the tumor measurement.