Effect of Display Magnification on Perceived Growth of Liver Lesions on Computed Tomography

Franklin N. Tessler · Michael S. Loop ·
Naomi Fineberg · Michelle M. McNamara ·
Tatum A. McArthur

Published online: 21 July 2011
© Society for Imaging Informatics in Medicine 2011

Abstract Our goal was to investigate the effect of displayed image magnification on perception of the size of hepatic lesions on abdominal computed tomography (CT) scans. Institutional review board approval and informed observer consent were obtained. Three experienced radiologists reviewed 90 CT image pairs in one session. Each image pair demonstrated a solitary, well-defined hypodense hepatic lesion measuring greater than 1 cm obtained at two points in time. The image pairs were presented three times in random order, once with the left image magnified, once with the right image magnified, and once with neither image magnified. The radiologists were asked to determine on which image the lesion was smaller or if there was no difference. The responses were analyzed statistically. The proportion of correct responses increased significantly as the difference in lesion size increased ($p < 0.001$). The percent of correct responses was higher when neither CT image was magnified. Magnification of one image decreased the accuracy of the readers' performance, especially at smaller differences, both of which were statistically significant ($p < 0.001$). Thus, accuracy of detect-

ing lesion size differences was degraded when the images were presented at differing magnification. This should be kept in mind when evaluating serial CT scans for growth or regression of tumors and other lesions.

Keywords Computed tomography · Image perception · Magnification · Measurement

Introduction

In clinical practice, physicians frequently rely on computed tomography (CT) scans to determine if tumors have changed in response to therapy [1–3]. Radiologists typically make this determination by comparing pairs of CT images from successive exams displayed side-by-side on computer monitors. Although all current display systems allow users to measure tumors electronically, radiologists may rely on their “gestalt” impression to assess growth, particularly when there are many tumors to evaluate. However, CT image pairs are often displayed at different magnification levels and we wondered if this could affect the accuracy of growth judgments.

There has been little published research regarding the effect of display size on perception of lesion growth. In a study by Melhem et al. [4], three readers were presented with simulated brain lesions. All three readers detected size differences in the brain lesions if the diameter of one of them was at least 15% greater than the other lesion. Gur et al. [5] assessed evaluation of lesions on CT scans at one and two times image magnification. Their results showed that magnification did not improve performance and increased variability. We therefore undertook this study to test the hypothesis that image magnification affects the subjective perception of size of liver lesions on CT.

F. N. Tessler (✉) · M. M. McNamara · T. A. McArthur
UAB Department of Radiology,
JTN338, 619 19th Street South,
Birmingham, AL 35249, USA
e-mail: ftessler@uabmc.edu

M. S. Loop
UAB Department of Vision Sciences,
924 18th Street South,
Birmingham, AL 35294, USA

N. Fineberg
UAB Department of Biostatistics, UAB School of Public Health,
1665 University Blvd,
Birmingham, AL 35294, USA

Fig. 1 **a** (left) and **b** (right): A representative pair of CT images of a hypodense right hepatic lobe lesion obtained at two different times. The image on the left has been magnified by a factor of 1.3 in each dimension

Materials and Methods

This study was compliant with the Health Insurance Portability and Accountability Act (HIPAA). Two of the authors (FNT, MMM) searched the departmental electronic image archive to find 90 pairs of CT images of liver lesions obtained at different points in time. Selection criteria included conspicuous hypodensity compared to the adjacent unenhanced or enhanced hepatic parenchyma, diameter greater than 1 cm, and good border definition. For lesions that were imaged following iodinated contrast administration, the images in each pair were obtained at the same phase of enhancement so that their conspicuity was similar. As well, to avoid ambiguity, none of the chosen images showed more than a single lesion.

Orthogonal bidimensional linear dimensions of each lesion were carefully measured on the picture archiving and communication system (PACS) workstation by two of the authors and recorded. In order to compensate for instances in which the two images were acquired with different fields of view, a correction factor was manually applied to one of the images to ensure that all the images were displayed at identical magnification. The images were displayed at a constant window width of 342 and a level of 56, in keeping with settings used in clinical practice. Next, the images were exported as TIFF files and anonymized. The images and measurements were imported into a customized database (FileMaker Pro 8 Advanced, FileMaker, Inc.) developed by one of the authors.

Using the same database software, the CT image pairs were presented to three observers three times in random order, once with the left image magnified by a factor of 1.3 in both dimensions (Fig. 1), once with the right image magnified to the same degree, and once with neither image magnified.

This factor was chosen following an informal prestudy review of image pairs with a range of magnification levels; we felt that it closely mimicked what we saw in our CT practice. The side with the larger lesion was determined randomly. The three reviewers were board-certified academic radiologists with a minimum of 6 years of experience in body imaging. Given this level of experience, and because the assigned task was very similar to what the radiologists routinely performed in clinical practice, we felt that training images were not needed. Institutional review board approval and informed observer consent were obtained. The radiologists were asked to indicate if the target lesion was larger on the left image, on the right image, or on neither. The software recorded the results and the time taken for the radiologist to enter a response.

Statistical analysis was performed using SPSS 16.0 software (SPSS, Inc.). Observations were divided into 20 groups based on percent difference between the two images. The Cox and Stuart test for trend was then used to assess the changes in percent correct responses as the difference between images increased. The Pearson correlation coefficient was used to assess the relationship between time to read the images and the order of observation. A *t* test was used to compare time to read the first and second half of the observations. The chi-square test was used to compare the proportion of correct responses between groups. Observations were divided into 10 groups based on percent difference between the two images to provide a larger cell size for this analysis. Three-way analysis of variance (ANOVA) was used to answer the primary question—does magnification affect the proportion of correct responses? Percent difference group (10 groups) and correct/incorrect response magnification were treated as grouping variables and reader as a random factor. This

allowed us to look at the relationship between the percent difference in size between the two images' magnification and the correctness of the response, while adjusting for the fact that there were three readers and differences in size between the two lesions. Significance levels of <0.05 were considered meaningful.

Results

Three of the 90 image pairs depicted lesions with identical areas (120, 143, and 1,295 mm²). All the readers correctly indicated that there was no size difference when images of the two smaller lesions were presented. For the third lesion, reader 1 chose the right image when it was magnified and chose the left image when there was no magnification. Reader 2 chose the right image when either image was magnified and the right image when no magnification was present. Reader 3 chose the right image when the right image was magnified and when there was no magnification. Therefore, two of the three readers chose the right-hand image when the right image was magnified and when there was no image magnification.

The remaining 87 image pairs depicted lesions that were not the same size. We divided these observations into 10 groups based on the percentage difference in lesion size between the paired images (Table 1). A response was considered correct if the larger lesion was chosen regardless of how small the actual size difference was. For lesions in which the actual size difference was less than 10%, the readers' pooled performance was worse than would be expected by chance. The proportion of correct responses increased as the difference in actual lesion size increased, and this was statistically significant with a Cox and Stuart

test for trend ($p=0.004$). The trend was also significant for images with ($p=0.035$) and without ($p=0.004$) magnification. No matter what the displayed magnification, the reviewers tended to choose the correct image if the size difference of the two compared hepatic lesions was large. Actual size difference also had a significant effect when the three readers' results were analyzed individually. However, the effect of magnification was not significant for individual readers, perhaps reflecting the smaller sample size.

A three-way ANOVA was used to look at percentage size difference, magnification and reader differences simultaneously. Overall accuracy increased as the percentage difference in lesion size increased ($p<0.001$) and was better without magnification ($p=0.025$). Only one interaction term was significant. The interaction term between readers and percent size difference ($p=0.025$) demonstrated that the difference between readers was greater when the percent difference in lesion size was small compared to when it was large.

Since reader differences were not significant in the three-way ANOVA, we repeated the analysis with a two-way ANOVA looking at the effects of difference in lesion size and magnification in the pooled data set. This analysis showed that the percentage of correct responses was higher when neither CT image was magnified ($p<0.001$). Overall accuracy also increased as percentage difference in actual lesion size increased ($p<0.001$). The image reviewers performed poorly if the size differences were small. However, the interaction term was not significant, indicating that the magnification effect was similar at all levels.

The radiologists were allowed to take as much time as they needed to enter their responses. Readers differed significantly ($p<0.001$) in the time to read the images and their percentage of correct responses ($p=0.018$) (Table 2).

Table 1 Relationship between correct responses and actual size difference for 87 lesions

Percent size difference	Correct responses (%)		
	Neither or one image magnified (%)	Neither image magnified (%)	One image magnified (%)
>0–10	46/315 (15)	14/105 (13)	32/210 (15)
>10–20	41/144 (28)	16/48 (33)	25/96 (26)
>20–30	29/81 (36)	9/27 (33)	20/54 (37)
>30–40	12/36 (33)	5/12 (42)	7/24 (29)
>40–50	26/54 (48)	12/18 (67)	14/36 (39)
>50–60	17/36 (47)	7/12 (58)	10/24 (42)
>60–70	8/9 (89)	3/3 (100)	5/6 (83)
>70–80	20/36 (56)	9/12 (75)	11/24 (46)
>80–90	27/36 (75)	9/12 (75)	18/24 (75)
>90	31/36 (86)	11/12 (92)	20/24 (83)
All observations	257/783 (33)	95/261 (36)	162/522 (31)
Cox and Stuart Test	$p=0.004$	$p=0.004$	$p=0.035$

Table 2 Time in seconds to score each of 270 observations for three readers

Reader	Mean±SD	Minimum	Maximum	% Correct
One	10.8±8.0	3.0	79.0	40%
Two	9.0±6.2	3.0	38.0	33%
Three	7.3±5.6	2.0	41.0	29%
All Observers	9.0±6.8	2.0	79.0	33%

The reader who took the longest time had the highest percentage of correct responses, while the reader who took the shortest time to respond had the lowest percentage of correct responses. There were no significant correlations between time and order of presentation for the three readers ($r=-0.06$ to -0.01). While the readers read the second half of the image pairs faster than the first half, this was statistically significant for only one reader ($p=0.013$). Overall, there was no relationship between time to score an observation and its order in the study.

Discussion

Changes in lesion size are often used as the primary criterion in assessing response to therapy in patients undergoing CT to evaluate tumors and various non-neoplastic conditions, such as abscesses or cysts. All current PACS allow radiologists to measure the size of lesions using electronic calipers. Although these measurements are subject to inter- and intra-observer variability that is related to lesion size, definition, and other factors, it is often assumed that caliper measurements are superior to a radiologist's subjective impression.

Unfortunately, while it is relatively simple to perform electronic measurements, radiologists are less likely to do so when there are multiple lesions to assess. Therefore, radiologists may rely on a “gestalt” impression when comparing serial studies. In our practice, we had observed that the comparison was sometimes performed on images that were displayed at different magnifications, in part because of changes in field of view between scans. We wondered if this might affect the subjective impression of change in size, and therefore undertook this study to assess the effect of magnification on perception of growth.

Not surprisingly, the proportion of correct responses increased significantly as the difference in lesion size increased—large size differences were readily perceived even if one image was magnified. However, the percentage of correct responses was higher when both images were displayed at the same magnification level. Magnification of

one image decreased the accuracy of reader performance, especially at smaller differences, both of which were statistically significant ($p<0.001$).

Our study suffers from several limitations, which include variable definition of our test lesions, inherent uncertainties in measuring the “true size” of the lesions, and an unequal distribution of lesion sizes and size changes between image pairs. We chose our magnification factor (1.3 times in each dimension, for an area change of 1.69) because we felt that a greater degree of magnification would have been unrealistic, while a lesser degree would have been insufficient to test an effect. Another limitation relates to our sole use of hepatic lesions; it is conceivable, though unlikely, that the effect of magnification may be different in other anatomic areas. As well, although repeated trials with interval forgetting would have been desirable, we felt that the effect of recall was minimized, since all the lesions were hypodense and because we avoided choosing scans that exhibited extraordinary characteristics.

It is also possible that the perception of size of a lesion is related to its shape, an effect that we did not test. Although we could have eliminated this variable by only choosing perfectly round lesions, that would have eliminated most noncystic masses, and we wished to consider a range of lesion morphology that is similar to what radiologists commonly encounter in clinical practice. Finally, it is also worth noting that observer errors in determining whether a lesion has grown may not be clinically significant in some situations.

Conclusions

We conclude that our study validates our hypothesis that display size affects radiologists' subjective assessment of the change in size of hepatic lesions on CT. Based on these results, we recommend that radiologists take this into account when interpreting CT studies. As well, we recommend that PACS vendors incorporate software features to automatically display images at identical magnification at the radiologist's discretion.

Acknowledgment We thank David S. Cosper for his assistance with this research.

References

1. van Persijn van Meerten EL, Gelderblom H, Bloem JL: RECIST revised: implications for the radiologist. A review article on the modified RECIST guideline. Available at: <http://www.springerlink.com/content/56316v4g6528848u/fulltext.html>. Accessed February 25, 2011

2. Erasmus JJ, Gladish GW, Broemeling L, et al: Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J Clin Oncol* 21:2574–2582, 2003
3. Zhao B, James LP, Moskowitz CS, et al: Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. *Radiology* 252:263–272, 2009
4. Melhem E, Herskovits E, Karli-Oguz K, et al: Defining thresholds for changes in size of simulated T2-hyperintense brain lesions on the basis of qualitative comparisons. *AJR* 180:65–69, 2003
5. Gur D, Klym A, King J, et al: The effect of image display size on observer performance: an assessment of variance components. *Acad Radiol* 13:409–413, 2006