Assessment of tumor size change in medical imaging is an essential aspect of the clinical evaluation of cancer treatment. Thus, precise and reliable lesion quantification is essential in response evaluation criteria in clinical oncology. The most common criteria used for tumor assessment remains Response Evaluation Criteria in Solid Tumors (RECIST). RECIST is a series of response criteria initially established in 2000 and revised in 2009 (version 1.1). Per RECIST version 1.1, up to two lymph nodes with short-axis diameter equal to or more than 1.5 cm can be selected as target lesions and be evaluated over time (1,2).

Despite substantial revisions made in RECIST 1.1, some vital questions and limitations remain unsolved, including interreader and intrareader variations of measurements, improper method of tumor assessment in the setting of image-based focal treatment (chemoembolization, radiofrequency ablation), and lack of tumor size change with new cancer therapies when necrosis occurs (3–5).

Adjacent nodes often merge as the disease progresses. Conversely, conglomerated nodes can split into smaller nodes when they shrink in response to treatment. In merged target lymph nodes, RECIST 1.1 calculates the longest short-axis diameter perpendicular to the longest axis of the resulting node (herein known as “short axis”), while in split target lymph nodes, RECIST 1.1 measures the sum of short-axis diameters of all the resulting nodal fragments (2).

Recent advancements in imaging modalities have allowed us to measure the volume of the lesions more easily. The literature indicates that volumetric measurements represent a better prediction and outcome than linear one-dimensional or two-dimensional measurements in response evaluation of cancers (6–10). Volumetric measurement is a surrogate for the number of neoplastic cells and provides a more detailed and accurate evaluation of tumor size compared with RECIST 1.1 (8,11,12).

In this study, we compared the RECIST 1.1 measurement in split and merged target lymph nodes with volumetric measurement on CT scans from clinical trials.

Materials and Methods

Patients

In this institutional review board–approved Health Insurance Portability and Accountability Act–compliant...
study, we retrospectively evaluated lymph nodes on CT scans from 166 patients (mean age, 53 years ± 14 [standard deviation] [range, 18 to 86 years]; 94 men) with various types of cancers affecting lymph nodes. Of the 166 patients, 158 were from the CT Lymph Nodes data set of The Cancer Imaging Archive Public Access (13,14). The remaining eight patients with genitourinary cancer (imaged between October 13, 2011, and June 26, 2017), including seven patients with bladder cancer and one patient with urachal cancer, were enrolled in clinical trials in our institution. The CT Lymph Nodes data have been used previously to develop artificial intelligence software to automatically detect and segment lymph nodes (14). The additional cases have been used for various clinical trials of treatments. The primary cancers of the involved nodes are as follows: lymphoma, melanoma, genitourinary tract malignancy, colorectal cancers, ovarian cancers, and others.

The inclusion criterion was a positive CT finding for lymphadenopathy in a patient with cancer. According to the RECIST 1.1 criteria, lymphadenopathy refers to any lymph node with a short-axis diameter greater than 1.5 cm. All patients were assessed for merging or splitting target lymph node lesions from baseline to follow-up. In patients with more than one group of lymph nodes that merged or split over time, we selected only one node with the best clear boundaries for segmentation.

Exclusion criteria were suboptimal image quality, absence of thin-section CT image series, no or inadequate intravenous contrast agent administration, and lack of merging or splitting nodes in serial examinations. To prevent inaccurate measurements and unrealistic results, we excluded the merged and split lesions if the boundaries were not clear for segmentation.

CT Imaging Technique

Multidetector-row CT scanners included the following: Siemens Biograph, Siemens SOMATOM Definition Flash, Siemens SOMATOM Force (Siemens Healthcare); GE Lightspeed Ultra (GE Medical Systems); Toshiba Aquilion (Toshiba Medical Systems); and Philips Brilliance 64 CT (Philips Healthcare). All scans were conducted according to our institution’s protocol for chest, abdominal, and pelvic CT. All CT examinations were performed with a single breath hold from the supraclavicular areas to symphysis pubis. Images were obtained with a collimation of 192 × 0.6 mm, 120 kV, and 2-mm section thickness with 1-mm increments, and 0.5-second gantry rotation time. Intravenous contrast agent was administered in all patients except one whom administration was withheld due to rising creatinine levels. A single dose of a nonionic intravenous contrast agent, iopamidol (Isovue-300, Bracco Diagnostics), was administered at a dosage of 1.8 mL per kilogram of body weight (maximum 130 mL) at a rate of 2 mL per second. Oral enteric contrast agent (50 mL of Omnipaque [GE Healthcare]) was combined in water with a maximum administration of 1000 mL and administered 30 minutes before the examination. We performed image analysis on the series with 2-mm section thickness, in soft-tissue kernel.

Geometric Modeling

We developed a spherical geometric model of the lymph nodes during merging (when nodes grew) or splitting (when nodes shrunk) over time. We aimed to have a simple representative method for the assessment of nodal change over time. Multiple spherical shapes were considered as a representation of the lymph nodes. Random degrees of increase in size and overlapping were given to each shape to resemble merging in one-target nodes (Fig 1, B) and two-target nodes (Fig 1, C). To illustrate splitting nodes, the overlap was eliminated, and nodes were decreased in size (Fig 1, D). The short axis per RECIST 1.1 and volume of these geometric shapes were calculated during merging or splitting.

Imaging Analysis

CT images in soft-tissue window settings (window level, 50 HU and width, 450 HU) were reviewed to determine if a target lymph node (≥1.5 cm in short axis) had merged or split. Carestream picture archiving and communication system version 12.1.5 software (Carestream Health) was used for detecting and measuring the size of these lesions at baseline and follow-up studies according to RECIST 1.1. Next, the whole thin-section soft-tissue series was exported to Vitrea Enterprise Suite V: 6.8.0 software (Vital Images) for volume measurement. The abdominal analysis module in Vitrea was used for tracing target lymph nodes that merged or split.
On the basis of RECIST 1.1, we calculated the short axis of the coalesced node in merged lymph nodes and calculated the sum of the short-axis diameters of the fragmented nodes in split nodes (1,15). Then, relative percent size changes in each group of merging and splitting nodes were calculated.

Volumetric measurement was considered as the reference standard method for tumor response evaluation (7,8,16). Volumes of the nodes before and after merging or splitting were calculated using Vitrea, and relative percent changes of the volumes were recorded.

Some studies have previously compared the percent change of the size in one dimension (RECIST), two dimensions (World Health Organization), and three dimensions (volumetric measurement) (17–19); however, mathematically comparing linear versus volumetric measurement is not easy and might not appear valid and precise (20). For a more accurate comparison of one and three dimensions, we considered each lesion as a sphere and a hypothetical diameter (HD) of the spherical lesion (volume) was calculated using this formula: \( HD = 2\sqrt[3]{\frac{3V}{4\pi}} \). In the next step, we compared changes of RECIST 1.1 with changes of...
Bland-Altman plots were made to assess the degree of difference and agreement between the percent size change in the HD measurement (variable A) and RECIST 1.1 measurement (variable B) in merging and splitting nodes. These plots represent the absolute differences between two paired measurements \((A - B)\) versus an average of these two paired measurements \([A + B]/2\) with 95% limits of agreement (21,22) (Appendix E1 [supplement]).

To confirm the accuracy of our estimates, we measured the CI and standard errors (21,22) (Appendix E2–E4 [supplement]). Because in Bland-Altman analysis clinicians determine up to what level of the random error is acceptable, we considered the acceptance cutoff discrepancies between RECIST 1.1 and HD measurements to be less than 20% in merged nodes and 30% in split nodes.

The null hypothesis in this study states that size changes measured by RECIST 1.1 are the same as HD calculated from the volume, but our alternative hypothesis states that size changes in RECIST 1.1 and HD measurements are neither compatible nor identical.

The statistical analyses were performed with statistical computing software (R version 3.5.0 [R Foundation for Statistical Computing] and Python version 3.7.2 [Python Software Foundation]).

**Results**

A total of 1698 CT scans were assessed in 166 adult patients (94 men) with lymphoma or lymph node metastases. The mean number of scans per patient was 10 to 12 (range, two to 45). The age varied between 18 and 86 years, with a median age of 57 years ± 7 (standard deviation). Unless otherwise noted, data are numbers with percentages in parentheses. Data are means with ranges in parentheses for time between baseline and follow-up.

| Characteristic | One-Target Merged | Two-Target Merged | Split Node |
|---------------|-------------------|-------------------|------------|
| No. of patients | 19                | 11                | 20         |
| No. of men     | 15 (79)           | 8 (73)            | 15 (75)    |
| Mean follow-up time (d)* | 250 (23–1269) | 130 (35–287) | 155 (30–420) |
| Cancer type    |                   |                   |            |
| Lymphoma       | 6 (32)            | 2 (18)            | 10 (50)    |
| Melanoma       | 3 (16)            | 2 (18)            | 1 (5)      |
| Bladder        | 4 (21)            | 3 (27)            | 4 (20)     |
| Urachal        | 1 (5)             | 1 (9)             | 0          |
| Colorectal     | 1 (5)             | 1 (9)             | 0          |
| Ovarian        | 2 (11)            | 1 (9)             | 0          |
| Renal cell carcinoma | 2 (11) | 1 (9) | 1 (5) |
| Adrenal cortical carcinoma | 0 | 0 | 1 (5) |
| Pheochromocytoma | 0 | 0 | 1 (5) |
| Mesothelioma   | 0                 | 0                 | 2 (10)     |

Note.—One node per patient was assessed. The mean age for all included patients \(n = 50\) was 57 years ± 7 (standard deviation). Unless otherwise noted, data are numbers with percentages in parentheses.

Bland-Altman plots were made to assess the degree of difference and agreement between the percent size change in the HD measurement (variable A) and RECIST 1.1 measurement (variable B) in merging and splitting nodes. These plots represent the absolute differences between two paired measurements \((A - B)\) versus an average of these two paired measurements \([A + B]/2\) with 95% limits of agreement (21,22) (Appendix E1 [supplement]). To confirm the accuracy of our estimates, we measured the CI and standard errors (21,22) (Appendix E2–E4 [supplement]).

Because in Bland-Altman analysis clinicians determine up to what level of the random error is acceptable, we considered the acceptance cutoff discrepancies between RECIST 1.1 and HD measurements to be less than 20% in merged nodes and 30% in split nodes.

The null hypothesis in this study states that size changes measured by RECIST 1.1 are the same as HD calculated from the volume, but our alternative hypothesis states that size changes in RECIST 1.1 and HD measurements are neither compatible nor identical.

The statistical analyses were performed with statistical computing software (R version 3.5.0 [R Foundation for Statistical Computing] and Python version 3.7.2 [Python Software Foundation]).

**Results**

A total of 1698 CT scans were assessed in 166 adult patients (94 men) with lymphoma or lymph node metastases. The mean number of scans per patient was 10 to 12 (range, two to 45). The age varied between 18 and 86 years, with a median age of 54 years. Lymphoma (23%; 38 of 166), melanoma (15%; 25 of 166), and colorectal cancer (7%; 11 of 166) were the most common cancers among the 166 patients.

### Table 1: Patient and Cancer Characteristics

| Characteristic         | One-Target Merged | Two-Target Merged | Split Node |
|------------------------|-------------------|-------------------|------------|
| No. of patients        | 19                | 11                | 20         |
| No. of men             | 15 (79)           | 8 (73)            | 15 (75)    |
| Mean follow-up time (d)* | 250 (23–1269) | 130 (35–287) | 155 (30–420) |
| Cancer type            |                   |                   |            |
| Lymphoma               | 6 (32)            | 2 (18)            | 10 (50)    |
| Melanoma               | 3 (16)            | 2 (18)            | 1 (5)      |
| Bladder                | 4 (21)            | 3 (27)            | 4 (20)     |
| Urachal                | 1 (5)             | 1 (9)             | 0          |
| Colorectal             | 1 (5)             | 1 (9)             | 0          |
| Ovarian                | 2 (11)            | 1 (9)             | 0          |
| Renal cell carcinoma   | 2 (11)            | 1 (9)             | 1 (5)      |
| Adrenal cortical carcinoma | 0 | 0 | 1 (5) |
| Pheochromocytoma       | 0                 | 0                 | 1 (5)      |
| Mesothelioma           | 0                 | 0                 | 2 (10)     |

Note.—One node per patient was assessed. The mean age for all included patients \(n = 50\) was 57 years ± 7 (standard deviation). Unless otherwise noted, data are numbers with percentages in parentheses. Data are means with ranges in parentheses for time between baseline and follow-up.
Fifty of 166 (30%) patients (76% men; 38 of 50) had at least one measurable target node that merged or split in this cohort. From these 50 nodes, 19 were one-target merged, 11 were two-target merged, and 20 were identified as split nodes. The mean age of the included patients was 57 years (range, 30 to 75 years). The mean time between baseline and follow-up for one-target merged lesions was 250 days (range, 23 to 1269 days), for two-target merged lesions was 130 days (range, 35 to 287 days), and for split lesions was 155 days (range, 30 to 420 days) (Table 1).

### Geometric Modeling

Spherical geometric modeling is not based on patient data and is only an example to show the concept of merging and splitting in target nodes. One-target merged nodes demonstrated an increase in size in both volumes (268 to 524 cm³ [+195%]) and RECIST 1.1 measurements (8 to 10 cm [+125%]) (Fig 1, A). Spherical modeling in the two-target merged nodes depicted an increase in size per volumetric measurements (302 to 580 cm³ [+192%]), while RECIST 1.1 calculation showed a decrease (12 to 10 cm [−83%]) in the ultimately merged tumor (Fig 1, C). In split nodes, our model showed a decrease in size (580 to 302 cm³ [−52%]), while RECIST 1.1 indicated an increase (10 to 12 cm [+120%]) in the resulting nodes (Fig 1, D).

#### One-Target Merged Nodes

Nineteen patients (15 men) had at least one target lymph node that merged with other nodes in the follow-up scans. Among all types of cancer, lymphoma (32%; six of 19) and bladder cancer (21%; four of 19) were more common in this group (Table 1).

**Figure 2:** Top images indicate percent change based on RECIST 1.1, tumor volume, and hypothetical diameter measurements for A, one-target merged, C, two-target merged, and E, split nodes. Bottom images illustrate the percent size change of each target lesion based on RECIST 1.1 and hypothetical diameter measurement and also demonstrate the response categories of treatment for each lesion based on RECIST 1.1 definition for B, one-target merged, D, two-target merged, and F, split nodes. RECIST 1.1 assigns four categories of response based on percent size change: complete response (the disappearance of all lesions), partial response (PR) (<30% decrease), stable disease (SD) (≤30% decrease or ≤20% increase), and progressive disease (PD) (>20% increase). RECIST = Response Evaluation Criteria in Solid Tumors.
changes between RECIST 1.1 and HD measurements might not be significantly different.

All RECIST 1.1 measurements in the one-target group fell into the PD category of the response type (19 of 19), while in HD, 18 of 19 (95%) were categorized as PD and one as SD. In just one occasion (5%), there was a discrepancy in response category between the two methods of measurements (Fig 2, B).

The Bland-Altman plot showed the mean difference in assessing lymph node size changes between RECIST 1.1 and HD was 1% (95% limits of agreement: 2 to 4; Fig 3, A and Table 3). The average discrepancy between the two methods was less than 20%.

### Two-Target Merged Nodes

In 11 patients (eight men), we found two neighboring target lymph nodes that merged with each other in the follow-up scans. The most prevalent cancer in this group was bladder cancer (27%; three of 11).

While volumetric measurements demonstrate an increase in size in all two-target merged nodes, RECIST 1.1 indicated a decrease in size in all 11 nodes (Figs 2, C and D, and 4). In two-target merged nodes, the mean size change in RECIST 1.1 was $-15\% \pm 6$ (range, $-31\%$ to $-4\%$), while mean volume change was $+121\% \pm 49$ (range, $+15\%$–$264\%$), and HD was $+29\% \pm 10$ (range, $+3\%$–$54\%$) ($P < .001$) (Table 2).

In two-target nodes, in only 18% (two of 11) of patients, both RECIST 1.1 and HD were in the SD category. In 73% (eight of 11) of patients, RECIST 1.1 showed SD; however, HD indicated PD. In one patient (9%), RECIST 1.1 showed PR but HD showed PD. Thus, on nine of 11 occasions (82%), the response category was different between RECIST 1.1 and HD (Fig 2, D).

The Bland-Altman plot demonstrated the mean difference in assessing lymph node size changes between RECIST 1.1 and HD was 44% (95% limits of agreement: 14 to 74; Figure 3, B and Table 3). The average difference between the two methods was greater than 20%.

### Split Nodes

In 20 patients (15 men) conglomerate nodes split into smaller fragments. Among all types of cancer in this group, lymphoma was the most common one (10 of 20; 50%).

While volumetric and HD measurements revealed a decrease in the size of all split nodes, RECIST 1.1 indicated an increase in size in 60% (12 of 20) of nodes (Fig 2, E and F, and 5). The mean size change of split nodes by RECIST 1.1

---

**Table 2: Mean Size Change per RECIST 1.1, Volume, and Hypothetical Diameter**

| Target Node       | RECIST 1.1 Change | Volume Change | Hypothetical Diameter Change |
|-------------------|------------------|---------------|-----------------------------|
| One-target merged | $+65\% \pm 12$   | $+376% \pm 92$| $+65\% \pm 11$              |
| Two-target merged | $-15\% \pm 6$    | $+121% \pm 49$| $+29\% \pm 10$              |
| Split node        | $+1\% \pm 12$    | $-66\% \pm 9$ | $-33\% \pm 7$               |

Note.—Values are shown as mean ± marginal error. RECIST = Response Evaluation Criteria in Solid Tumors.
was +1% ± 12 [range, −48% to +52%]. The mean volume change was −66% ± 9 (range, −98% to −13%), and the mean HD change was −33% ± 7 (range, −65% to −4%) (P < .001) (Table 2).

In a total of 60% (12 of 20) of the split nodes, RECIST 1.1 results were dissimilar and were not congruent with volumetric measurements and HD. In these 12 nodes, the mean size change in volumetric measurements was −54% ± 14 (range, −13 to −72) and HD was −23% ± 7 (range, −4 to −35) (P < .001). However, in the remaining 40% (eight of 20), RECIST 1.1 results were in the same direction as volumetric measurements and HD. The mean size change in these eight nodes was −85%
RECIST 1.1 Limitations in Evaluating Lymph Nodes That Merge or Split

± 11 (range, −67 to −98) for volumetric measurements and −49% ± 12 (range, −30 to −65) for HD (P < .001).

Response categories in the split group per RECIST 1.1 measurements demonstrated PD in five (25%), SD in 12 (60%), and PR in three (15%) nodes. However, per the HD method, 10 (50%) were SD, and 10 (50%) were in the PR category. In 11 of 20 nodes (55%), a discrepancy of response categories was seen between RECIST 1.1 and HD.

The Bland-Altman analysis showed that the mean difference in assessing lymph node size changes between RECIST 1.1 and HD was −34% (95% limits of agreement: −64 to −3; Figure 3, C and Table 3). The average difference between these two methods of measurement was greater than 30%.

Discussion

In this study, we found that there were discrepancies in the calculated size changes of lymph nodes before and after treatment between RECIST 1.1 and HD based on the three-dimensional volumetric measurement. Although the changes in RECIST 1.1 and HD method were not significantly different in one-target merged nodes (P = .95), the average percentage changes between the two methods of measurement in two-target nodes and split nodes were clearly different (P < .001). Additionally, there were differences in response categories between these different measurement methods. Taken together, RECIST 1.1 did not represent nodal changes as well as volumetric and HD methods in this study.

In most cancer trials, the treatment plan is determined on the basis of response categories of the lesions at CT. For RECIST 1.1, these categories are complete response (disappearance of all target lesions), PR (at least a 30% decrease in the sum of the diameters of target lesions), and PD (at least a 20% increase in the sum of the longest diameter of target lesions). SD does not meet the criteria for either PR or PD (1,12). Highly precise lesion measurement is crucial for tumor evaluation, as even minor errors in lesion measurement may alter treatment decisions by changing one category of response to another (4).

Volumetric measurement has been recognized as a standard reference for tumor assessment (7,8,10,16,23). Some studies have compared volume with RECIST 1.1 measurement and have shown acceptable reliability and validity between them (8,24). The literature demonstrates volumetric imaging provides higher sensitivity to measurement and greater capacity in earlier response detection and prediction of prognosis than RECIST 1.1 measurements (5,25,26). Earlier response detection and more representative quantification enable physicians to stop ineffective therapy earlier or prevent premature termination of treatment. Although volumetric measurements can be more time-consuming and expensive to undertake, some studies claimed that this method of tumor measurement is more reproducible than RECIST 1.1 and other traditional anatomic measurements, which indicates an additional benefit to the use of the tumor volume for accurate tumor response characterization (27–30).

In comparing RECIST 1.1 with volume, Therasse et al in the 2000 RECIST 1.1 guidelines stated that a 20% increase by RECIST 1.1 measurements is equal to a 73% increase by volumetric measurements during cancer progression. In PR through disease regression, a 30% decrease in size by RECIST 1.1 measurements is equal to a 65% decrease in volumetric measurements (17,25,31).

As cancer progresses, lymph nodes may enlarge and coalesce with adjacent nodes. In this study, the volumetric and
HD measurements demonstrated an increase in the size of both merged groups (one-target and two-target). Various factors can influence the accuracy of lesion measurements during merging and splitting. Some examples are lesion shape, the distance between lesions, degree of overlap when merging, growth axis or regression axis, and lesion numbers.

Even though in one-target merged nodes RECIST 1.1 may not be precise, it can be an appropriate representative of size changes compared with volume as the reference standard. However, in two-target merged nodes, RECIST 1.1 falsely indicated a reduction in the size of all the nodes, while the volume of the nodes had actually increased. The discrepancy between the results among one-target and two-target merged can be described mainly by overlapping of the merged nodes. In fact, when one-target lymph nodes merged with other adjacent nontarget nodes, there was no overlapping between target nodes (one lesion at baseline and one lesion at follow-up). As a result, an increase in size is seen in both by RECIST 1.1 and volume measurement, although with different magnitude. While in two-target merged, because of the merging of one target node with another target node, some degree of overlapping will eventually occur (two lesions at baseline and then one lesion at follow-up). Therefore, in all of the two-target merged nodal lesions, different size changes were observed (RECIST 1.1 revealed a decrease in size while volume demonstrated an increase).

In one-target merged nodes, when comparing response categories between RECIST 1.1 and HD, a difference was seen in only one of 19 cases (5%). In this case, although both RECIST 1.1 and HD reflected an increase in size, HD outcome was categorized into SD, while RECIST 1.1 result was classified as PD. This difference might be random or incidental and might not be supported in a larger group of patients with similar lesions. However, in two-target merged nodes, the discrepancy of the response categories between RECIST 1.1 and HD was demonstrated in nine of 11 (82%). RECIST 1.1 in all of the cases showed a decrease in size and indicated SD in 10 of 11 (91%) cases and PR in one case (9%). However, HD method indicated an increase in all cases and showed the trajectory of lesion growth. HD method demonstrated PD in nine of 11 (82%) cases. Although the response categories of two-target merged nodes were the same in two of 11 (18%) cases, and both RECIST 1.1 and HD showed SD, in both cases HD revealed an increase in size while RECIST 1.1 indicated a decrease. In 11 of 20 (55%) split nodes, the response categories between RECIST 1.1 and HD methods were different. Even if the split nodal lesion is decreased in size, however, RECIST 1.1 indicates PD in five of 20 (25%) and SD in 12 of 20 (60%). In these 12 SD cases, RECIST 1.1 indicated an increase in size in seven cases that were still in the SD category. In five of the SD cases, RECIST 1.1 showed a decrease in size but still less than the 30% that indicated SD. In three of 20 (15%), RECIST 1.1 showed PR. In a total of 20 splitting nodes, RECIST 1.1 indicated a decrease in size in eight (40%), in which five were categorized as SD and three as PR. This may happen when the lesion fragments get so small that even if we add the sum of all fragments, it is still smaller than the original short-axis diameter. So, if the nodal lesion is the only target lesion in the body to be selected, evaluation per RECIST 1.1 will falsely change the treatment plan in the majority of the two-target merged and in more than half of split nodal lesions. In this work, we decided to set an acceptable threshold of up to 20% and 30% average absolute difference for discrepancies between RECIST 1.1 and HD measurements in the progression (merged groups) and regression of the disease (split group), respectively. In two-target merged and split nodes, Bland-Altman plot showed a remarkable difference between RECIST 1.1 and HD measurement (mean difference was +44% in two-target merged and was −34% in split nodes). This plot indicated that these two methods are not interchangeable in these scenarios. In one-target merged method, the mean difference was 1% (near to zero), which demonstrates that there is not a significant difference between the two methods. However, the limit of agreement is very wide, which shows that RECIST 1.1 is not a good alternative method to HD. However, if the limit of agreement were small, we could say that HD and RECIST 1.1 measurements could be equivalent and can be used interchangeably.

Our study was limited by its retrospective nature and small sample size. Because some types of cancer may have a heterogeneous response to treatment, the conclusions of this study may not apply to all types of malignancies. Therefore, more testing with larger sample sizes and cancer types is required. Acquisition of scans from multiple institutions will be needed to validate these findings. Another limitation is that the HD was calculated on the basis of tumor volume measurements assuming that the tumor mass is spherical in nature, while in reality they are not spheres and do not expand nor shrink spherically. In addition, the inherent interobserver and intraobserver variabilities of manual imaging measurements for calculating lesion volume and the short-axis diameter should be considered as another limitation. Labor time may be the most limiting factor toward integrating volumetric measurements in practical evaluation of the tumor responses (11). Automatic segmentation algorithms could be a better alternative to perform such measurements.

The issue of the inaccuracy of the measurement when lesions merge or split is not specific to lymph nodes. We see splitting or merging with other soft-tissue lesions too, particularly in the lung and liver, which needs further research. This problem is also applicable to some other response evaluation criteria, such as RECIST for immunotherapy and Cheson and Lugano criteria that are being used for patients with lymphoma.

In summary, we found that when lymph nodes merge (during progression) or split (during regression), the RECIST 1.1 method of measurement does not optimally reflect the nodal size. In one-target merged lesions, even though RECIST 1.1 shows the increase in size of the lesions, it is not an accurate method of estimation for the growth changes. For two-target merged lesions, RECIST 1.1 falsely showed a decrease in size, while the size of the tumor increased. The same is true for splitting, a sign of treatment response that can be mistaken for the progression of disease by RECIST 1.1 in 60% of the patients and is inaccurate in the rest. These inaccuracies can negatively affect the overall therapeutic response estimation. The findings affirm a need for a new method or modification of the current RECIST 1.1 method of response evaluation in nodal disease.
We also recommend that target nodes should not be selected too close to other nodes, whenever possible.

**Acknowledgments:** We would like to thank Najmeh Abedzadeh for her assistance in the statistical analysis that was done using R and Python and also in the preparation of figures.

**Author contributions:** Guarantors of integrity of entire study: A.S., M.B.; study concept/design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, A.S., M.B., E.F., L.R.F., E.C.J.; clinical studies, A.S., M.B., A.B.A., N.M.B.; statistical analysis, A.S., F.F.; and manuscript editing, all authors

**Disclosures of Conflicts of Interest:** A.S. disclosed no relevant relationships. M.B. disclosed no relevant relationships. E.F. disclosed no relevant relationships. A.B.A. disclosed no relevant relationships. N.M.B. disclosed no relevant relationships. L.R.F. Activities related to the present article: author has research agreement. Activities not related to the present article: author has government-issued patents without royalties; author received royalties from Springer for “combat radiology” book; author has patent pending without royalties for multi grey trauma window and x-meter. Other relationships: disclosed no relevant relationships. R.M.S. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: author received patent royalties and license royalties from PingAn, iCAD, ScanMed, Philips, and Translation Holdings; author’s institution has cooperative research and development agreement with PingAn; author’s institution has GPU card donations from NVIDIA. Other relationships: disclosed no relevant relationships.

**References**

1. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45(2):228–247.
2. Schwartz LH, Litière S, de Vries E, et al. RECIST 1.1-Update and clarification: From the RECIST committee. Eur J Cancer 2016;62:132–137.
3. Fourrier L, Ammari S, Thiam R, Cuñod CA. Imaging criteria for assessing tumour response: RECIST, mRECIST, Cheson. Diagn Interv Imaging 2014;95(7-8):689–703.
4. Nishino M. Tumor Response Assessment for Precision Cancer Therapy: Response Evaluation Criteria in Solid Tumors and Beyond. Am Soc Clin Oncol Educ Book 2018;38(38):1019–1029.
5. 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(7):2093–2105.
6. Force J, Rajan A, Dombi E, Steinberg SM, Giaccone G. Assessment of objective responses using volumetric evaluation in advanced thymic malignancies and metastatic non-small cell lung cancer, J Thorac Oncol 2011;6(7):1267–1273.
7. Tran LN, Brown MS, Goldin JG, et al. Comparison of treatment response classifications between unidimensional, bidimensional, and volumetric measurements of metastatic lung lesions on chest computed tomography, Acad Radiol 2004;11(12):1355–1360.
8. Reza P, Pisaneschi MJ, Feg C, Yaghmai V. A radiologist’s guide to treatment response criteria in oncologic imaging: anatomic imaging biomarkers. AJR Am J Roentgenol 2013;201(2):237–245.
9. Yaghmai V, Miller FH, Reza P, Benson AR 3rd, Salem R. Response to treatment series: part 2, tumor response assessment—using new and conventional criteria. AJR Am J Roentgenol 2011;197(1):18–27.
10. Goldmacher GV, Conklin J. The use of tumour volumetrics to assess response to therapy in anticancer clinical trials. Br J Clin Pharmacol 2012;73(6):846–854.
11. Fenret A, Morello J, Bjella K, Fargoty E, Beal J, Chaudhary V. Do Diameter Measurements Provide Sufficient and Reliable Tumor Assessment? An Evaluation of Diameter, Arcometric, and Volumetric Variability of Lung Lesion Measurements on Computerized Tomography Scans. J Oncol 2015;2015:632943.
12. Lubner MG, Stabo N, Lubner SJ, Munoz Del Rio A, Song C, Pickhardt PJ. Volumetric Versus Unidimensional Measures of Metastatic Colorectal Cancer in Assessing Disease Response. Clin Colorectal Cancer 2017;16(4):324–333.
13. Clark K. CT Lymph Nodes. Cancer Imaging Archive (TClA) Public Access. https://wiki.cancerimagingarchive.net/display/Public/CT+Lymph+Nodes. Last modified June 3, 2020. Accessed March 30, 2021.
14. Roth HR, Lu L, Seff A, et al. A New 2.5D Representation for Lymph Node Detection Using Random Sets of Deep Convolutional Neural Network Observations. In: Golland P, Hata N, Barillot C, Hornegger J, Howe R, eds. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014. MICCAI 2014. Lecture Notes in Computer Science, vol 8673. Cham, Switzerland: Springer; 2014:520–527.
15. van Persijn van Meerten EL, Gelderblom H, Bloem JL. RECIST revised: implications for the radiologist. A review article on the modified RECIST guideline. Eur Radiol 2010;20(6):1456–1467.
16. Anderson O, Kidd A, Goorman K, et al. Fully Automated Volumetric Measurement of Malignant Pleural Mesothelioma from Computed Tomography Images by Deep Learning: Preliminary Results of an Internal Validation. In: Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2: BIOIMAGING, Valletta, Malta, February 24–26, 2020. Setúbal, Portugal: SciTePress, 2020:64–73.
17. James K, Eisenhauer E, Christian M, et al. Measuring response in solid tumors: unidimensional versus bidimensional measurement. J Natl Cancer Inst 1999;91(6):523–528.
18. Padhani AR, Ollivier L. The RECIST (Response Evaluation Criteria in Solid Tumors) criteria: implications for diagnostic radiologists. Br J Radiol 2001;74(887):983–986.
19. 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(3):205–216.
20. Welsh JL, Bodeker K, Fallon E, Bhuta SK, Buatti JM, Cullen JJ. Comparison of response evaluation criteria in solid tumors with volumetric measurements for estimation of tumor burden in pancreatic adenocarcinoma and hepatocellular carcinoma. Am J Surg 2012;204(5):580–585.
21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307–310.
22. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb) 2015;25(2);141–151.
23. Mozley PD, Schwartz LH, BendtSEN C, Zhao B, Petrick N, Buckler AJ. Change in lung tumor volume as a biomarker of treatment response: a critical review of the evidence. Ann Oncol 2010;21(9):1751–1755.
24. Cademartiri F, Luccichenti G, Maffe1 E, et al. Imaging for oncologic staging and follow-up: review of current methods and novel approaches. Acta Biomed 2008;79(2):85–91.
25. Hayes SA, Pietanza MC, O’Driscolld D, et al. Comparison of CT volumetric measurement with RECIST response in patients with lung cancer. Eur J Radiol 2016;85(3):524–533.
26. Shah GD, Kesari S, Xu R, et al. Comparison of linear and volumetric criteria in assessing tumor response in adult high-grade gliomas. Neuro Oncol 2008;10(1):38–46.
27. Zhao B, Oxnard GR, Moskwitz CS, et al. A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development. Clin Cancer Res 2010;16(18):4647–4653.
28. Mozley PD, Bendersen C, Zhao B, et al. Measurement of tumor volumes improves RECIST-based response assessments in advanced lung cancer. Transl Oncol 2012;5(1):19–25.
29. Nishino M, Jackman DM, DiPiro PJ, Hatahu H, Jänne PA, Johnson BE. Revisiting the relationship between tumour volume and diameter in advanced NSCLC patients: An exercise to maximize the utility of each measure to assess response to therapy. Clin Radiol 2014;69(8):841–848.
30. Fenery KE, Patrónas NJ, Heery CR, Guille JL, Folio LR. Resources required for semi-automatic volumetric measurements in metastatic chordoma: is potentially improved tumor burden assessment worth the time burden? J Digit Imaging 2016;29(3):357–364.
31. Schiavon G, Ruggiero A, Schöffski P, et al. Tumor volume as an alternative response measurement for imatinib treated GIST patients. PLoS One 2012;7(11):e48372 [Published correction appears in PLoS One 2013;8(1)].