Title

RECIST 1.1 Target Lesion Categorical Response in Metastatic Renal Cell Carcinoma: A Comparison of Conventional versus Volumetric Assessment.

Journal

Radiology: Imaging Cancer, 5(5)

Authors

Gong, Amanda
Ruchalski, Kathleen
Kim, Hyun
et al.

Publication Date

2023-09-01

DOI

10.1148/rycan.220166

Peer reviewed

Radiology: Imaging Cancer

ORIGINAL RESEARCH

RECIST 1.1 Target Lesion Categorical Response in Metastatic Renal Cell Carcinoma: A Comparison of Conventional versus Volumetric Assessment

Amanda J. Gong, BA • Kathleen Ruchalski, MD • Hyun J. Kim, PhD • Michael Douek, MD, MBA • Antonio Gutierrez, MD • Maitraya Patel, MD • Victor Sai, MD • Heidi Coy, BS • Bianca Villegas, MPH • Steven Raman, MD • Jonathan Goldin, MD, PhD

From the David Geffen School of Medicine, University of California, Los Angeles, Calif (A.J.G., K.R., H.J.K., M.D., A.G., M.P., V.S., H.C., S.R., J.G.); Department of Radiological Sciences, UCLA, Los Angeles, Calif (K.R., H.J.K., M.D., A.G., M.P., V.S., S.R., J.G.); and UCLA Center for Computer Vision and Imaging Biomarkers, 924 Westwood Blvd, Ste 615, Los Angeles, CA 90024 (A.J.G., H.J.K., H.C., B.V., J.G.). Received November 21, 2022; revision requested January 23, 2023; revision received July 5; accepted July 18. **Address correspondence to** A.J.G. (email: *ajgong@mednet.ucla.edu*).

Supported in part by a David Geffen Medical Scholarship.

Conflicts of interest are listed at the end of this article.

Radiology: Imaging Cancer 2023; 5(5):e220166 • https://doi.org/10.1148/rycan.220166 • Content codes: GU OI

Purpose: To investigate Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1) approximations of target lesion tumor burden by comparing categorical treatment response according to conventional RECIST versus actual tumor volume measurements of RECIST target lesions.

Materials and Methods: This is a retrospective cohort study of individuals with metastatic renal cell carcinoma enrolled in a clinical trial (from 2003 to 2017) and includes individuals who underwent baseline and at least one follow-up chest, abdominal, and pelvic CT study and with at least one target lesion. Target lesion volume was assessed by *(a)* V_{model}, a spherical model of conventional RECIST 1.1, which was extrapolated from RECIST diameter, and *(b)* V_{actual}, manually contoured volume. Volumetric responses were determined by the sum of target lesion volumes (V_{model}-sum TL and V_{actual}-sum TL, respectively). Categorical volumetric thresholds were extrapolated from RECIST. McNemar tests were used to compare categorical volume responses.

Results: Target lesions were assessed at baseline (638 participants), week 9 (593 participants), and week 17 (508 participants). V_{model}-sum TL classified more participants as having progressive disease (PD), compared with V_{actual}-sum TL at week 9 (52 vs 31 participants) and week 17 (57 vs 39 participants), with significant overall response discordance ($P < .001$). At week 9, 25 (48%) of 52 participants labeled with PD by V_{model}-sum TL were classified as having stable disease by V_{actual}-sum TL.

Conclusion: A model of RECIST 1.1 based on a single diameter measurement more frequently classified PD compared with response assessment by actual measured tumor volume.

ClinicalTrials.gov registration no. NCT01865747

Supplemental material is available for this article.

Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1) is the most commonly applied standardized imaging criteria to assess systemic treatment response in metastatic renal cell carcinoma (1). This evidence- and consensus-based guideline provides systematic rules for the assessment of tumor burden change during treatment. Measurable sites of disease are selected on pretreatment images and defined as target lesions to be quantitatively followed and used to inform objective categorical responses to treatment. These target lesions represent only a subset of disease and therefore serve as an estimation of total tumor burden.

While actual tumor masses are three-dimensional, RECIST 1.1 relies solely on linear two-dimensional measurements of each target lesion, usually in the axial image acquisition plane. By using a single diameter to describe lesion size, RECIST 1.1 assumes lesions have a spherical shape and undergo uniform changes in size. However, actual lesions may have complex contours that are nongeometric in shape and change in form over time, resulting in limitations on the validity of RECIST measures (2).

Recent advancements in medical imaging and the unprecedented availability of computational power allow for automated and semiautomated volumetric tumor assessment (3–6). Several prior studies have shown that measuring tumor volume is technically feasible, accurate, and reproducible (7–11). These tumor volumes better correlate with actual tumor size and form than do their two-dimensional depictions (12). In addition, changes in tumor volume are more sensitive for indicating changes in tumor burden in response to systemic treatment when compared with single-dimension assessment (7,11) and can predict a clinical benefit in certain clinical scenarios (13–16). However, much of this prior work has been in the evaluation of size changes in a single lesion or organ site and has been less commonly

Abbreviations

PD = progressive disease, PR = partial response, QIWS = quantitative imaging workstation, RECIST = Response Evaluation Criteria in Solid Tumors, SOD = sum of diameters, V_{actual} = actual target lesion volume, V_{actual}-sum TL = sum of target lesion volumes according to V_{actual}, V_{model} = spherical approximation of tumor volume, V_{model}-sum TL = sum of target lesion volumes according to V_{model}

Summary

Compared with actual tumor volume according to manual contours, the spherical model of Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1) for target lesions more frequently classified participants as having disease progression, suggesting that conventional RECIST 1.1 itself may overestimate disease progression.

Key Points

- The use of a single diameter measurement as a proxy for tumor volume overestimated actual tumor volume in individuals with metastatic renal cell carcinoma; specifically, a volumetric model of Response Evaluation Criteria in Solid Tumors version 1.1 (RECIST 1.1) overestimated manually contoured volumes by 46.1% ± 96.8 (SD) (638 lesions, $P < .001$) across all anatomic sites of disease.
- Treatment response assessment by a model of RECIST based on single diameter measurements classified more cases of disease progression than did assessment by actual tumor volume at early imaging time points ($P < .001$ at week 9 and $P < .001$ at week 17).
- A sum of diameters, as represented by a sum of spheres in the volumetric model of RECIST, may not accurately reflect changes in actual tumor volume, suggesting potential limitations of RECIST 1.1 in response assessment.

Keywords

Urinary, Kidney, Metastases, Oncology, Tumor Response, Volume Analysis, Outcomes Analysis

applied in the context of multisite metastatic disease response to systemic treatment (14,17–19).

This study investigated RECIST approximations of target lesion tumor burden by comparing categorical treatment response assessment according to conventional RECIST with actual tumor volume measurements of RECIST target lesions in the setting of metastatic renal cell carcinoma.

Materials and Methods

Participants and Study Design

Written informed consent was waived for this institutional review board–approved (approval no. 19-000495), Health Insurance Portability and Accountability Act–compliant retrospective analysis. The participants in our study cohort were identified from an anonymized imaging clinical trial database; they were previously included as part of a randomized phase III trial between 2003 and 2017 that compared the efficacy and safety of cabozantinib versus everolimus in renal cell carcinoma (ClinicalTrials.gov registration no. NCT01865747) (20), in addition to a secondary analysis that highlighted the effect of new and nontarget lesions on response assessment (21). Inclusion criteria included all participants who had at least one target lesion at baseline as defined by RECIST 1.1. Participants were excluded if there was no cross-sectional imaging performed at baseline or if baseline imaging was performed but no target lesion was identified.

Imaging Acquisition and RECIST 1.1 Assessment

Baseline and follow-up standard-of-care volumetric chest CT and either a CT or T1-weighted MRI (T1 spin-echo without breath holding) scan of the abdomen and pelvis were performed in the axial plane (with a section thickness of 5.0 mm or less). Imaging was performed at baseline and every 8 weeks for the first 12 months and every 12 weeks thereafter. Images at pretreatment baseline, week 9, and week 17 were included in this retrospective analysis. Intravenous contrast material was administered unless contraindicated. Tumor response was assessed using RECIST 1.1 (1) through a blinded independent central review by two radiologists specialized in abdominal and oncologic imaging (pool of readers: M.D., K.R., A.G., M.P., V.S., S.R.), with adjudication by a third reviewer in case of disagreement (22). At the time of interpretation, readers were blinded to all participant and clinical information. Target lesions were selected at baseline for each participant per RECIST 1.1, with a maximum of five target lesions total per participant, at most, two lesions per organ. The anatomic location or organ in which the target lesion resided was also annotated. A sum of diameters (SOD) for all target lesions was calculated (by using the longest diameter for soft-tissue lesions and short axis for nodal lesions) and reported at baseline. All other sites of disease were marked as nontarget lesions.

Response assessment for target lesions was determined by the percentage of change in the SOD from baseline. Complete response was defined by the disappearance of all soft-tissue target lesions, with all nodal lesions measuring less than 10 mm in the short axis. Partial response (PR) was defined as at least a 30% decrease in SOD. Progressive disease (PD) was defined by a relative increase in SOD of 20% or greater, with an absolute increase of at least 5 mm (Appendix S4). In RECIST 1.1, the appearance of new lesions or unequivocal worsening of nontarget lesions also resulted in a classification of PD.

Tumor Size

Tumor size was measured for each target lesion by using RECIST 1.1, requiring a minimum of a 10-mm longest diameter for soft-tissue lesions and a 15-mm short axis for lymph nodes. A spherical approximation of tumor volume (V_{model}) was calculated based on the RECIST diameter measurement for each target lesion by using the below equation, where d represents the diameter of a lesion:

$$V_{\text{model}} = \frac{\pi}{6} d^3.$$

The actual target lesion volume (V_{actual}) was obtained by manually contouring the volume of each lesion on a quantitative imaging workstation (QIWS); this contour was used as the ground truth (Fig 1). The QIWS provides infrastructure for reading, annotating, and archiving image data; the software is housed on a secure computing cluster and was developed by our group's computational team with more than 10 years of

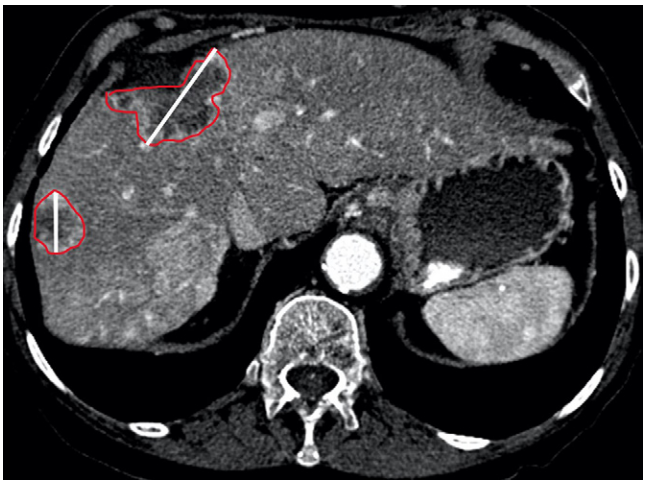

Figure 1: Illustration of manual contours and RECIST 1.1 diameter in an axial section. Axial contrast-enhanced CT image in the abdomen with liver metastases measured by manual contour (red) and RECIST 1.1 diameter (white). Manually contoured volume is calculated from a three-dimensional rendition of axial contours in adjacent sections. RECIST 1.1 = Response Evaluation Criteria in Solid Tumors version 1.1.

experience in imaging-based data management, quality control, and analysis.

Total Tumor Burden

A total tumor burden (sum of target lesion measurements per participant) was calculated for each target measurement (lesion diameter, V_{model}, and V_{actual}) at each time point. For this study, conventional RECIST is defined as response assessment based on tumor measurements obtained using RECIST 1.1 definitions (longest diameter for soft-tissue lesions and short axis for lymph nodes).

Volumetric Categorical Response

Categorical response thresholds for volumetric measurements of tumor burden were geometrically extrapolated from RECIST 1.1 response thresholds by using a well-documented simplifying assumption (Eq 4). We used the following volumetric thresholds for response assessment: *(a)* PR, requiring a 65% decrease in the sum of volumes, and *(b)* PD, requiring a 73% or greater increase in the sum of volumes (Fig S1). Volumetric thresholds for PD and PR were geometrically extrapolated from RECIST SOD thresholds by using a spherical model of tumor shape (Eq 2) and a simple calculation included by Therasse et al in the 2000 RECIST guidelines, which has been widely cited in subsequent literature (13,14,23–26). The 2009 RECIST 1.1 updated guidelines reference a 73% volumetric threshold for PD when describing "unequivocal progression" by nonmeasurable disease (1).

Despite this consensus, these sources do not comment on the implicit simplifying approximations (Eqq 1–6) made in this extrapolation. We use the volumetric thresholds described in prior literature, acknowledging that the simplification is responsible for differences in response categorization between the spherical model of RECIST and RECIST using target lesions alone.

$$T_{\text{R1.1, PD}} = 1.2 = \frac{\sum_{n'=1}^{n'} d_i'}{\sum_{n=1}^{n} d_i} \quad (1)$$

$$V_i = \frac{\pi}{6} d_i^3;\ d_i = \left(\frac{6}{\pi} V_i\right)^{\frac{1}{3}} \quad (2)$$

$$V_i = \frac{\pi}{6} d_i^3;\ d_i = \left(\frac{6}{\pi} V_i\right)^{\frac{1}{3}} \quad (3)$$

$$\text{Assume } n = n' = 1 \quad (4)$$

$$\text{PD}: 1.2^3 = 1.728 = \frac{V'}{V} \rightarrow +73\% \quad (5)$$

$$\text{PR}: 0.7^3 = 0.343 = \frac{V'}{V} \rightarrow -65\% \quad (6)$$

For each participant, the sum of target lesion volumes by V_{model} (V_{model}-sum TL) and the sum of target lesion volumes by V_{actual} (V_{actual}-sum TL) at weeks 9 and 17 were compared with respective sum baseline values to determine a percentage of change in total tumor burden and the corresponding volumetric categorical response.

Statistical Analysis

Percentage of error and paired two-tailed *t* tests were used to compare V_{model} and V_{actual} (and their respective sums over all target lesions per participant), using V_{actual} as a reference of the ground truth. To compare agreement in overall response assessment using V_{model}-sum TL and V_{actual}-sum TL, both a paired proportion test and unweighted κ statistic with 95% CIs were calculated across all response categories (PD, stable disease, PR). McNemar tests were used to compare differences in the proportion of participants with PR at week 9 and week 17 and those with PD at week 9 and week 17. Bonferroni adjustment was applied for the multiple comparisons (ie, .05/2 = .025 for all response categories and α = .05/4 = .0125 for week 9 and week 17 in PR and PD). The familywise error rate was controlled at a significance level of .05. Stata (SE, version 17.0; Stata) was used to perform statistical analyses.

Results

Participant and Lesion Characteristics

Of the original 658 participants identified, 20 were excluded because they did not undergo baseline imaging or had no target lesions (Fig 2). This resulted in a final study cohort of 638 participants who underwent baseline imaging and 587 participants who underwent baseline and at least a first follow-up (week 9) imaging session, of which 496 participants had a subsequent week 17 imaging time point (Fig 2); participants who underwent baseline imaging had a median of five total imaging time points (IQR, three).

The number of target lesions on the baseline scan selected for each participant ranged from one to five, with 129 (20.2%)

participants with one target lesion, 184 (28.8%) with two target lesions, 147 (23.0%) with three target lesions, 100 (15.7%) with four target lesions, and 78 (12.2%) with five target lesions. There were 1728 target lesions identified in total, with 1587 and 1337 measured at weeks 9 and 17, respectively (Fig 3). Anatomic sites of disease were most commonly lymph nodes (27.3%) and lung (24.7%).

Estimation of Tumor Size and Total Tumor Burden

At baseline, V_{model} overestimated the volume of the largest target lesion per participant by 46.1% ± 96.8 (SD) (638 lesions, $P < .001$) across all anatomic sites (Fig 4, Appendix S2).

Volumetric Categorical Response

The majority of participants were classified as having stable disease by RECIST target SOD, V_{model}-sum TL, and V_{actual}-sum TL at weeks 9 and 17 (Figs 5, 6). No participants were classified as having a complete response to treatment at either time point. According to conventional RECIST 1.1, PD occurred in 115 (19.6%) and 157 (31.7%) of participants at week 9 and week 17, respectively. When accounting for only those participants in which PD occurred according to target lesion SOD criteria, there were 46 (7.8%) and 67 (13.5%) participants at weeks 9 and 17, respectively, with PD. More participants were categorized as having PD by V_{model}-sum TL when compared with V_{actual}-sum TL at week 9 (52 [8.9%] vs 31 [5.3%] participants, $P < .001$) and week 17 (57 [11.5%] vs 39 [7.9%] participants, $P < .001$).

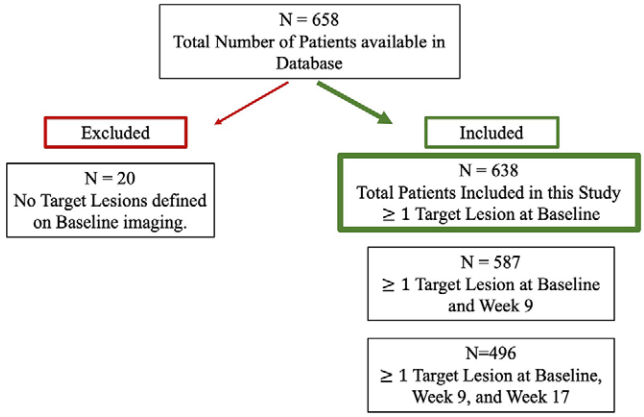

Figure 2: Selection of study participants.

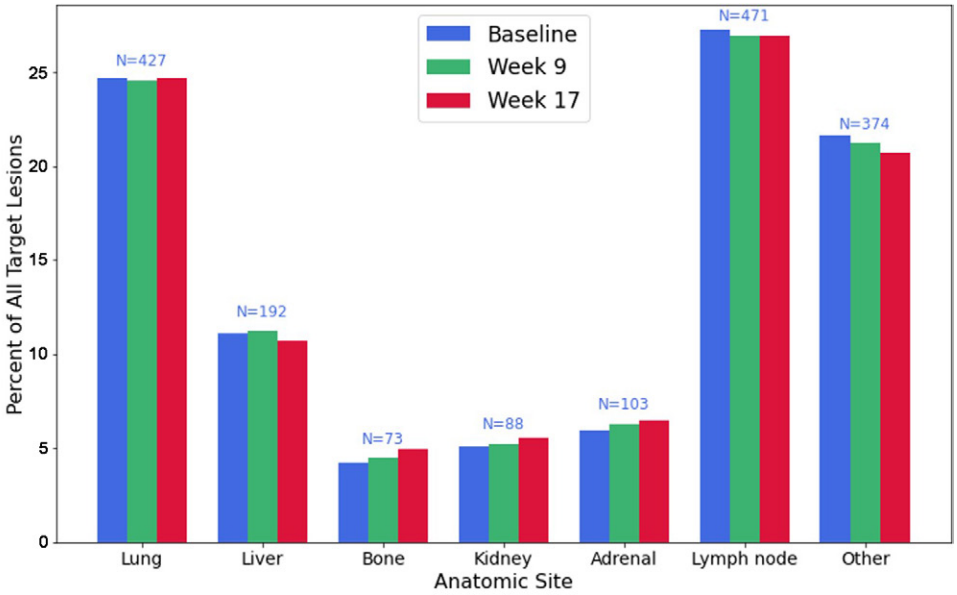

Figure 3: Anatomic sites of target lesions. Anatomic sites of disease at baseline, week 9, and week 17. N = the number of target lesions at baseline for a given site, distinct from the vertical axis, which illustrates the percentage of total number of target lesions (all anatomic sites). Anatomic site labeled as "other" includes abdominal cavity, brain, chest, gastrointestinal tract, pancreas, pelvis, peritoneum, plural cavity, retroperitoneum, and spleen.

We found no evidence of a difference in the proportion of participants with PR, stable disease, and PD between the V_{model}-sum TL and RECIST 1.1 target SOD response classification at week 9 and week 17 (Appendix S3). The spherical model of RECIST 1.1 (V_{model}-sum TL) and tumor burden by actual contoured volumes (V_{actual}-sum TL) differed in overall categorical response at week 9 ($P = .017$) and showed no evidence of a difference at week 17 ($P = .053$) (Fig 5); agreement in response assessment ranged from moderate to substantial according to κ statistic at week 9 (κ = 0.63; 95% CI: 0.54, 0.71) and week 17 (κ = 0.70; 95% CI: 0.62, 0.77) (Appendix S3).

At week 9, V_{model}-sum TL classified significantly more participants with PD than did V_{actual}-sum TL (52 [8.86%] vs 31 [5.28%], respectively; $P < .001$). Twenty-five (48%) of the 52 participants classified as having PD by V_{model}-sum TL were reclassified as having stable disease by V_{actual}-sum TL. Conversely, only four (13%) of 31 participants with progression according to V_{actual}-sum TL were reclassified with stable disease by V_{model}-sum TL (Fig 6, Appendix S1). Only 27 participants were classified as having PD at week 9 by both measures. Similarly, at week 17, V_{model}-sum TL classified more participants as having PD than did V_{actual}-sum TL ($P < .001$), with PD occurring in 57 (11.49%) participants according to V_{model}-sum TL and only 39 (7.86%) according to V_{actual}-sum TL.

Although V_{model}-sum TL classified more participants as having PR response than did V_{actual}-sum TL at both time points, the overall difference in the proportion of participants with PR was not significant ($P = .16$ at week 9 and $P = .19$ at week 17). At week 9, for example, PR occurred in 56 (9.54%) participants according to V_{model}-sum TL and 48 (8.18%) participants according to V_{actual}-sum TL (Fig 5). There was an approximately balanced reclassification, with 20 (36%) of 56 participants having

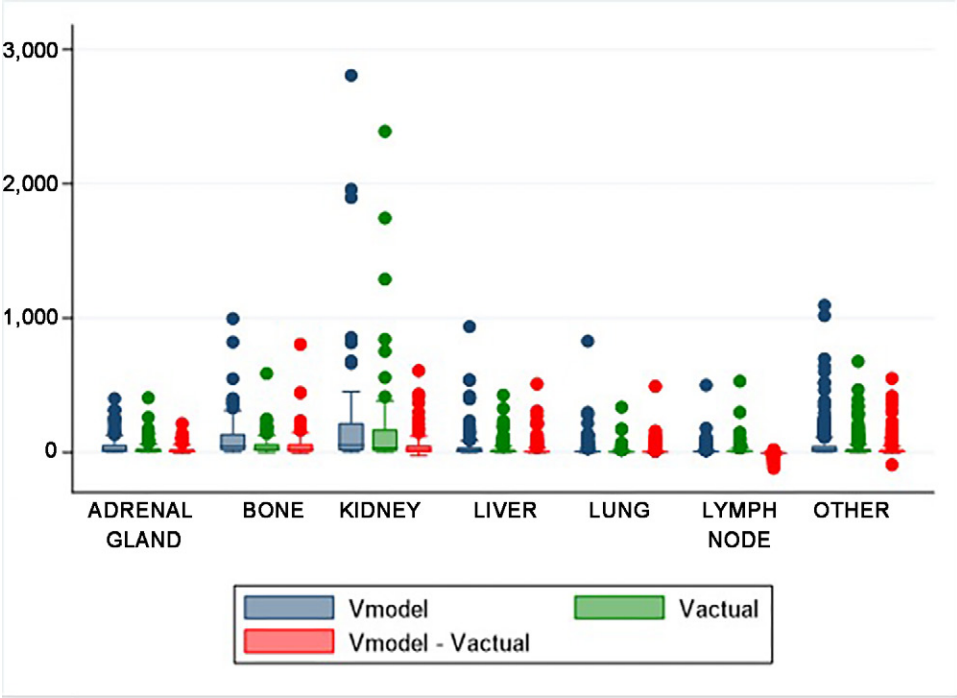

Figure 4: Target lesion size by anatomic site: a comparison of largest target lesion volume at baseline when assessed by the spherical model of RECIST 1.1 (V_{model}) and by manual contour. V_{model} was calculated from the RECIST 1.1 diameter for each lesion. V_{actual} is the lesion volume measured by manual contour. Only the largest target lesion per participant, determined by RECIST 1.1 diameter, was included in this box plot. Three kidney lesions with outlier V_{model} and V_{actual} (>1000 cm^3) values were not depicted for scaling purposes. Mean, SD, and *P* values comparing V_{model} and V_{actual} are included in Appendix S2. RECIST 1.1 = Response Evaluation Criteria in Solid Tumors version 1.1.

	Week 9 (n; %) ^^			Week 17 (n; %) ^		
	PR	**SD**	**PD**	**PR**	**SD**	**PD**
RECIST 1.1	58(9.88)	414(70.53)	115(19.59)	64(12.90)	275(55.44)	157(31.65)
RECIST 1.1 TL	59(10.05)	482(82.11)	46(7.84)	67(13.51)	362(72.98)	67(13.51)
V_{model}-sum TL	**56(9.54)**	**479(81.60)**	**52(8.86) ****	**65(13.10)**	**374(75.40)**	**57(11.49) ***
V_{actual}-sum TL	**48(8.18)**	**508(86.54)**	**31(5.28) ****	**58(11.69)**	**399(80.44)**	**39(7.86) ***

Figure 5: RECIST categorical responses by measurement: RECIST 1.1, RECIST 1.1 target sum of diameters only, V_{model}-sum TL, and V_{actual}-sum TL. ^^ indicates a *P* value of .017, and ^ indicates a *P* value of .053, comparing overall response (all categories) between V_{model}-sum TL and V_{actual}-sum TL. ** indicates a *P* value of .0001, and * indicates a *P* value of .0004, comparing differences in the proportion of classifications of progressive disease (PD vs non-PD) between V_{model}-sum TL and V_{actual}-sum TL. PD = progressive disease; PR = partial response; RECIST 1.1 = Response Evaluation Criteria in Solid Tumors version 1.1 (1) including target, nontarget, and new lesions; RECIST 1.1 TL = sum of RECIST 1.1 diameters on target lesions only; SD = stable disease; V_{actual}-sum TL = sum of target lesion volumes according to manual contours; V_{model}-sum TL = sum of target lesion volumes according to spherical model of RECIST 1.1.

	V_{actual}-sum TL					
	Week 9, (n; %)			**Week 17, (n; %)**		
V_{model}-sum TL	**PR**	**SD**	**PD**	**PR**	**SD**	**PD**
PR (n; %)	**36(6.13)**	**20(3.41)**	**0(0)**	**47(9.48)**	**18(3.63)**	**0(0)**
SD (n; %)	**12(2.04)**	**463(78.88)**	**4(0.68) ****	**11(2.22)**	**359(72.38)**	**4(0.81) ***
PD, (n; %)	**0(0)**	**25(4.26) ****	**27(4.60)**	**0(0)**	**22(4.44) ***	**35(7.06)**

Figure 6: Comparisons of volumetric categorical response according to V_{model}-sum TL and V_{actual}-sum TL. ** indicates a *P* value of .0001, and * indicates a *P* value of .0004 for PD versus non-PD at week 9 and week 17, respectively. *P* = .16 and *P* = .19 for PR versus non-PR at week 9 and week 17, respectively. Groups listed on diagonals (shaded cells) have response concordance, and those on off-diagonals have response discordance. PD = progressive disease; PR = partial response; RECIST 1.1 = Response Evaluation Criteria in Solid Tumors version 1.1 (1), including target, nontarget and new lesions; RECIST 1.1 TL = sum of RECIST 1.1 diameters on target lesions only; SD = stable disease; V_{actual}-sum TL = sum of target lesion volume according to manual contours; V_{model}-sum TL = sum of target lesion volumes according to spherical model of RECIST 1.1.

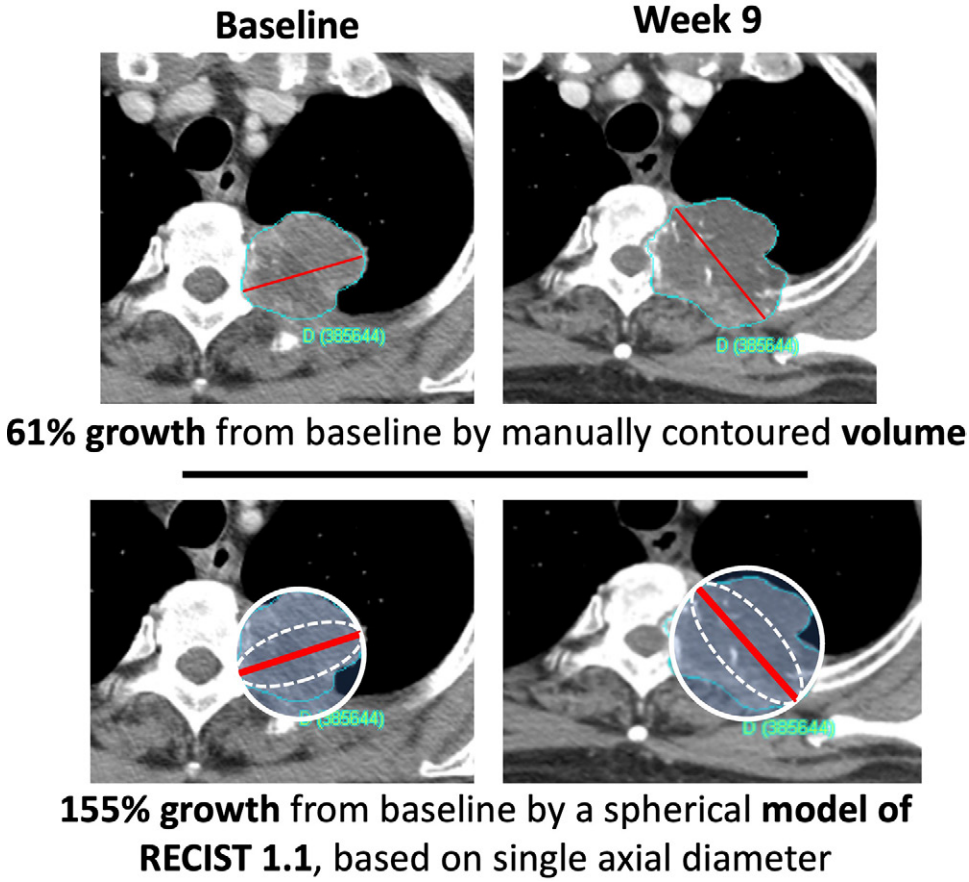

Figure 7: Target lesion example. A soft-tissue tumor illustrates anisotropic growth from baseline to week 9 follow-up. The RECIST 1.1 single diameter axial measurement (bottom row) overestimates tumor size and inflates reported change over time compared with manual contours (top row). RECIST 1.1 = Response Evaluation Criteria in Solid Tumors version 1.1.

PR according to V_{model}-sum TL being reclassified as having stable disease by V_{actual}-sum TL and 12 (25%) of 48 participants with PR according to V_{actual}-sum TL reclassified as having stable disease by V_{model}-sum TL. Despite the net agreement in the total proportion of participants with PR, only 36 participants had PR according to both measures (Fig 6).

As an example, one soft tumor illustrates how anisotropic target lesion growth can result in differences in reported change in tumor burden, when comparing a single diameter measurement with a manually contoured volume (Fig 7).

Discussion

In this study, conventional RECIST, measured by spherical approximation, overestimated true tumor burden and affected response classification. Tumor burden as defined by the RECIST sum of target lesions when measured by manual volumetric contours is a more conservative and likely more accurate measure of treatment response. While overestimation of tumor burden at a single time point (by RECIST's use of *longest* diameter) is an intuitive result, our study finds that this approximation also has an effect on response assessment because of the overestimation of change in tumor burden over multiple time points. In our cohort of participants with metastatic renal cell carcinoma, there was significant discordance in the classification of PD when using RECIST-derived spherical tumor measurements versus manually contoured tumor volumes. Of those participants classified as having PD by the spherical model, approximately 48% of participants at week 9 and 39% of participants at week 17 were reclassified as having stable disease when calculated by the sum of actual tumor volumes. In clinical trials, RECIST PD is deemed as treatment failure and commonly results in the cessation of therapy for that participant. However, according to our study, this may occur prematurely in up to one-third to one-half of participants when reassessed by target lesion tumor volume. Treatment response based on percentage of change in an SOD may result in an earlier reported time to progression and a shortened progression-free survival, with effects on clinical decisions and the evaluation of treatment efficacy.

Volumetric measures of tumor burden also offer benefits in study design. A more accurate assessment of PD should reduce the number of participants needed for clinical trial enrollment and better correlate with overall survival. Additionally, discordance in response assessment of serial imaging for a single participant requires adjudication by a third reviewer. Variability in the measured change in tumor burden is a major contributor to discordant response assessment; a dual-reader study of 179 participants with multisite small cell lung cancer required adjudication in 36.7% of participants, of whom 18.5% were due to discrepant measures of change in tumor burden (27). More

conservative, clinically stable response criteria such as volume have the potential to lower adjudication rates in clinical trials and improve reader agreement in clinical practice.

Our work supports others who have investigated the use of more complex tumor measurements, including volume, in comparison to RECIST-derived two-dimensional measurements. Notably, our findings of overclassification of progression are consistent with the study by Schiavon et al (2) in 78 participants with 139 gastrointestinal stromal tumor liver metastases. The measured change in tumor size according to geometric models (spheres and ellipsoids) better correlated with change according to manually contoured volumes for lesion size reduction than for lesion size increase, suggesting that overestimation of tumor size by RECIST's geometric approximations is more pronounced for lesions that grow. Although further investigation is needed to better relate single-lesion analysis to multisite whole tumor burden, these conclusions are consistent with our findings that RECIST overreports PD and may be a result of RECIST's inflation of the reported change in tumor burden for participants with growing lesions.

Our study supports the published literature that tumor volumetry is a preferable measure of tumor burden and treatment response (10–18,28–33). Our study is additive to prior work, much of which compares response assessment by RECIST with assessment by alternative geometric models of tumor volume, which are calculated not from actual volumetric contours but from additional orthogonal diameter measurements (14,29). To our knowledge, this is the largest study in which multisite metastases are summated to extract categorical volumetric response. In a study including 42 participants with multisite metastatic non–small cell lung cancer, Hayes et al (13) compared response assessment between semiautomated volumetric contours (using volumetric thresholds extrapolated using spheres and ellipsoids) and conventional RECIST. Although this study's strengths included an analysis of overall survival and demonstrated the benefit of lowering volumetric thresholds, the size of the study cohort precluded an analysis of PD, allowing only a comparison of PR and stable disease.

This study was limited by its retrospective nature. First, the use of anonymized clinical trial data of participants with metastatic renal cell carcinoma prohibited the evaluation of inter- and intrareader variation and its effect on tumor burden estimation and response categorization. However, methods to limit this effect were used, including formalized reader training to review RECIST criteria and a collection of sample cases to minimize reader variation. Second, given anonymized data, participant outcomes were not readily available to further compare volumetric measurement surrogacy for overall survival; further work is needed to assess the value of volumetric response assessment as an early indicator of overall survival. Third, this study's focus on methods used to quantify target lesion tumor burden represents a subset of disease included in RECIST criteria. While not quantified by RECIST 1.1, the inclusion of nontarget and new lesions may limit the differences in overall response assessment highlighted in our analysis of target lesions alone. Volumetric segmentation may be more useful when lesions are more likely to be discrete, asymmetric, and to change anisotropically (nonuniform change); therefore, these findings may not extend naturally to other tumor histologic characteristics. In addition, our use of multi-institutional data has inherently introduced variations in imaging acquisition and reconstruction parameters, with potential effects on both volumetric and conventional RECIST measures of tumor size (34). However, these variations may also improve the generalizability of our study findings. While this study's analysis does not apply RECIST 1.1 absolute increase criteria, this has a negligible impact on our results and conclusions. Last, our analysis exclusively compared manually contoured volumes with volumes calculated from a single diameter. While other tumor measurements, including bidimensional measurements, were not assessed, single diameter tumor measurements were found to be highly concordant with bidimensional assessments and have less measurement variability (35).

In conclusion, our study demonstrates that the use of change in lesion size as a measure of treatment efficacy, or lack thereof, would be strengthened by direct measures of volume rather than by the current use of lesion diameters as a proxy for tumor volume.

Acknowledgments: Thank you to the UCLA Center for Computer Vision and Imaging Biomarkers for dedicating funding and resources that make this study possible.

Author contributions: Guarantors of integrity of entire study, **A.J.G., H.C.**; study concepts/study 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.J.G., K.R.**; clinical studies, **A.J.G., K.R., H.J.K., A.G., M.P., V.S., H.C., S.R., J.G.**; statistical analysis, **A.J.G., K.R., H.J.K., B.V., J.G.**; and manuscript editing, **A.J.G., K.R., H.J.K., M.D., A.G., M.P., V.S., B.V., S.R., J.G.**

Data sharing: Data generated or analyzed during the study are available from the corresponding author by request.

Disclosures of conflicts of interest: **A.J.G.** No relevant relationships. **K.R.** GE Association of University Radiologists Radiology Research Academic Fellowship. **H.J.K.** Participation on a Data Safety Monitoring Board or advisory board for Curocell and C&R Research; research consultant for MedQIA. **M.D.** No relevant relationships. **A.G.** No relevant relationships. **M.P.** No relevant relationships. **V.S.** No relevant relationships. **H.C.** No relevant relationships. **B.V.** No relevant relationships. **S.R.** No relevant relationships. **J.G.** Founder of MedQIA, with stock or stock options.

References

1. Eisenhauer EA, Verweij J. New response evaluation criteria in solid tumors: RECIST guideline version 1.1. Eur J Cancer Suppl 2009;7(2):5.
2. Schiavon G, Ruggiero A, Bekers DJ, et al. The effect of baseline morphology and its change during treatment on the accuracy of Response Evaluation Criteria in Solid Tumours in assessment of liver metastases. Eur J Cancer 2014;50(5):972–980.
3. Li CC, Wu MY, Sun YC, et al. Ensemble classification and segmentation for intracranial metastatic tumors on MRI images based on 2D U-nets. Sci Rep 2021;11(1):20634.
4. Liu X, Han C, Cui Y, Xie T, Zhang X, Wang X. Detection and segmentation of pelvic bones metastases in MRI images for patients with prostate cancer based on deep learning. Front Oncol 2021;11:773299.
5. Kim K, Kim S, Han K, Bae H, Shin J, Lim JS. Diagnostic performance of deep learning-based lesion detection algorithm in CT for detecting hepatic metastasis from colorectal cancer. Korean J Radiol 2021;22(6):912–921.
6. Baidya Kayal E, Kandasamy D, Yadav R, Bakhshi S, Sharma R, Mehndiratta A. Automatic segmentation and RECIST score evaluation in osteosarcoma using diffusion MRI: A computer aided system process. Eur J Radiol 2020;133:109359.
7. 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.

8. Yankelevitz DF, Reeves AP, Kostis WJ, Zhao B, Henschke CI. Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. Radiology 2000;217(1):251–256.
9. Goodman LR, Gulsun M, Washington L, Nagy PG, Piacsek KL. Inherent variability of CT lung nodule measurements in vivo using semiautomated volumetric measurements. AJR Am J Roentgenol 2006;186(4):989–994.
10. Zhao B, Schwartz LH, Larson SM. Imaging surrogates of tumor response to therapy: anatomic and functional biomarkers. J Nucl Med 2009;50(2):239–249.
11. Marten K, Auer F, Schmidt S, Kohl G, Rummeny EJ, Engelke C. Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria. Eur Radiol 2006;16(4):781–790.
12. Choi SM, Choi DK, Kim TH, et al. A comparison of radiologic tumor volume and pathologic tumor volume in renal cell carcinoma (RCC). PLoS One 2015;10(3):e0122019.
13. Hayes SA, Pietanza MC, O'Driscoll D, et al. Comparison of CT volumetric measurement with RECIST response in patients with lung cancer. Eur J Radiol 2016;85(3):524–533.
14. 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).]
15. Fenerty KE, Folio LR, Patronas NJ, Marté JL, Gulley JL, Heery CR. Predicting clinical outcomes in chordoma patients receiving immunotherapy: a comparison between volumetric segmentation and RECIST. BMC Cancer 2016;16(1):672.
16. Lee JH, Lee HY, Ahn MJ, et al. Volume-based growth tumor kinetics as a prognostic biomarker for patients with EGFR mutant lung adenocarcinoma undergoing EGFR tyrosine kinase inhibitor therapy: a case control study. Cancer Imaging 2016;16(1):5.
17. 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.
18. Winter KS, Hofmann FO, Thierfelder KM, et al. Towards volumetric thresholds in RECIST 1.1: Therapeutic response assessment in hepatic metastases. Eur Radiol 2018;28(11):4839–4848.
19. Oxnard GR, Zhao B, Sima CS, et al. Variability of lung tumor measurements on repeat computed tomography scans taken within 15 minutes. J Clin Oncol 2011;29(23):3114-3119.
20. Choueiri TK, Escudier B, Powles T, et al. Cabozantinib versus everolimus in advanced renal cell carcinoma (METEOR): final results from a randomised, open label, phase 3 trial. Lancet Oncol 2016;17(7):917–927.
21. Coy HJ, Douek ML, Ruchalski K, et al. Components of radiologic progressive disease defined by RECIST 1.1 in patients with metastatic clear cell renal cell carcinoma. Radiology 2019;292(1):103–109.
22. U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research, and Center for Biologics Evaluation and Research. Clinical Trial Imaging Endpoint Process Standards Guidance for Industry. Silver Spring, Md: U.S. Food and Drug Administration, 2018.
23. 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.
24. 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.
25. Padhani AR, Ollivier L. The RECIST criteria: Implications for diagnostic radiologists. Br J Radiol 2001;74(887):983-986.
26. Zhao B, Schwartz LH, Moskowitz CS, Ginsberg MS, Rizvi NA, Kris MG. Lung cancer: Computerized quantification of tumor response—initial results. Radiology. 2006;241(3):892-898.
27. Beaumont H, Evans TL, Klifa C, et al. Discrepancies of assessments in a RECIST 1.1 phase II clinical trial - association between adjudication rate and variability in images and tumors selection. Cancer Imaging 2018;18(1):50.
28. Levine ZH, Galloway BR, Peskin AP, Heussel CP, Chen JJ. Tumor volume measurement errors of RECIST studied with ellipsoids. Med Phys 2011;38(5):2552–2557.
29. Choi H, Charnsangavej C, Faria SC, et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. J Clin Oncol 2007;25(13):1753–1759.
30. Wen PY, Macdonald DR, Reardon DA, et al. Updated response assessment criteria for high-grade gliomas: Response assessment in neuro-oncology working group. J Clin Oncol 2010;28(11):1963-1972.
31. Dunstan RW, Wharton KA, Quigley C, Lowe A. The use of immunohistochemistry for biomarker assessment—can it compete with other technologies? Toxicol Pathol 2011;39(6):988-1002.
32. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. J Clin Oncol 2007;25(5):579-586.
33. Oda Y, Tanaka K, Hirose T, et al. Standardization of evaluation method and prognostic significance of histological response to preoperative chemotherapy in high-grade non-round cell soft tissue sarcomas. BMC Cancer 2022;22(1)94.
34. Emaminejad N, Wahi-Anwar MW, Kim GHJ, Hsu W, Brown M, McNitt-Gray M. Reproducibility of lung nodule radiomic features: Multivariable and univariable investigations that account for interactions between CT acquisition and reconstruction parameters. Med Phys 2021;48(6):2906–2919.
35. Nishino M, Giobbie-Hurder A, Gargano M, Suda M, Ramaiya NH, Hodi FS. Developing a common language for tumor response to immunotherapy: immune-related response criteria using unidimensional measurements. Clin Cancer Res 2013;19(14):3936–3943.