Quantified Visual Scoring of Metastatic Melanoma Patient Treatment Response Using Computed Tomography: Improving on the Current Standard

Ronald H. Gottlieb · Elizabeth Krupinski ·
Pavani Chalasani · Lee Cranmer

Published online: 22 July 2011
© Society for Imaging Informatics in Medicine 2011

Abstract To assess whether quantitative visual scoring (QVS) is a better early predictor of progression-free survival (PFS) in patients on chemotherapy for metastatic melanoma using CT than the currently used Response Evaluation Criteria in Solid Tumors (RECIST) standard. Retrospective evaluation of 65 consecutive patients with metastatic melanoma on treatment who had a baseline and follow-up CT after two cycles of therapy. QVS was used to code imaging findings on the radiology reports considering size change, brain metastases, new lesions, mixed lesion response, and the number of organ systems involved. RECIST 1.1 criteria placed patients in the progressive disease, stable disease, or partial response groups. Multiple regression analysis was used to correlate the various independent variables with PFS. The Cox hazard proportions ratio, median survival, and Kaplan–Meier curves of the different prognostic groups were calculated. QVS of size change was found more sensitive in detecting patients deteriorating (57.1% versus 37.5%) or improving (23.8% versus 10.7%), more correlated with the median PFS for the deteriorating (1.8 versus 1.7 months), stable (5.6 versus 4.0 month), and improving (8.3 versus 5.5 months) categories and more predictive of PFS (Cox hazard proportion ratio of 3.070 versus 1.860) than RECIST 1.1 categorization. Multiple regression analysis demonstrated QVS of lesion size correlated

most closely with PFS among the variables assessed ($r=0.519$, $p<0.0001$). QVS in this study was superior to standard RECIST categorization in terms of discriminating treated metastatic melanoma patients likely to have longer PFS.

Keywords Structured reporting · Visual perception · Cancer detection · Image analysis · Image perception · Computed tomography · Body imaging · Clinical oncology · Health services research

Introduction

Little work has been undertaken to validate imaging as a surrogate endpoint for patient clinical responses to the many new cancer therapies in development. Most prior work has been directed at measuring lesion size on CT using the Response Evaluation Criteria in Solid Tumors (RECIST) [1, 2]. Unidimensional RECIST size measurements have been found to be relatively poor estimates of volume, particularly in lesions with complex shapes [3] and are restricted to a small subset of defined target lesions. In addition, even the most recent update of RECIST (version 1.1) only assesses changes in lesion size, rather than incorporating consideration of the many other findings evident in modern imaging studies [2]. This greatly restricts full exploitation of both conventional and emerging modalities.

Our objective in this project was to develop and test a quantitative visual scoring (QVS) system, exploiting the full range of imaging parameters in assessing cancer treatment response. For this, we selected a set of patients with metastatic melanoma receiving either nanoparticle albumin-encapsulated (nab-)paclitaxel, a new chemotherapy agent for this condition, or dacarbazine (DTIC), the standard agent for this condition.

R. H. Gottlieb (✉) · E. Krupinski
Department of Radiology, University of Arizona,
4291 N. Campbell Avenue,
Tucson, AZ 85719, USA
e-mail: rhgottlieb@gmail.com

P. Chalasani · L. Cranmer
Department of Medicine, University of Arizona,
1501 N. Campbell Avenue,
Tucson, AZ 85719, USA

Survival in melanoma patients with classical therapy is of relatively short duration with median overall survival times of 6–9 months and median progression-free survival (PFS) of about 2 months [4]. These outcomes have not been altered despite decades of therapeutic development.

While some recent progress has been made in metastatic melanoma management, validation of new means of imaging assessment remains a priority. RECIST defined response rates do not necessarily correlate with outcome and survival [5]. We undertook the present study to evaluate whether a new systematic, imaging interpretation paradigm, QVS, could perform better than conventional RECIST scoring in assessing patient outcomes in metastatic melanoma.

Methods

This retrospective evaluation was in accord with the principles of the Health Insurance Portability and Accountability Act and was approved by the University of Arizona's Human Subjects Protection Program. Informed consent was not required because there was no impact on patient care. Information was obtained from the radiology and clinical information systems. The records of all consecutive melanoma patients treated with either DTIC or nab-paclitaxel (Abraxane; Celgene, Summit, New Jersey, USA) at our institution from 2004 to 2009 were searched. The study group only included patients having at least a baseline and follow-up body CT (relevant anatomic regions included) after two cycles of chemotherapy (6–8 weeks). Other clinical parameters extracted from the patient records included their age, gender, serum lactic dehydrogenase (LDH) baseline, and completion of therapy levels as a comparative serologic marker of tumor response [6], date and reason of therapy discontinuation, time of last known contact with the patient, and date of death, if relevant (Table 1).

A simplified QVS scheme was used to code imaging findings on the radiology reports, if newly apparent on the follow-up CT scan as present or absent and, if apparent on the baseline and follow-up scans, as increased, no change, or decreased, considering all neoplastic lesions (Table 2). The tumor burden was quantified by recording the number of organ systems involved. Brain metastases were docu-

Table 2 Visual scoring system

Visual assessment	Score	Imaging findings
Change (if already present)		Overall tumor size change
Increased	1	
No change	0	
Decreased	−1	
Appearance on follow-up scan		New lesions
Newly apparent	1	
Absent	0	
Indeterminate	−3	All findings

Three categories of change resulted from using quantitative visual scoring: +1=overall increase, 0=no change, and −1=overall decrease in size of all neoplastic lesions corresponding to improving, stable, and deteriorating patient status on treatment. New lesion development placed patients automatically into the deteriorating category

mented and coded as present (score of 1) or absent (score of 0) on either the baseline or follow-up head CT or MRI examinations. RECIST 1.1 criteria [2] were used to calculate the percentage difference in the sum of the diameters of up to five target lesions on the baseline and follow-up imaging studies. Patients were placed in the RECIST categories of progressive disease (PD), stable disease (SD), or partial response (PR) based on these percentage changes from baseline. No patients had a complete response. One or more new lesions categorized patients automatically as PD (Table 3). If parameters could not be determined from the reports, they were defined as indeterminate. All imaging parameters, either QVS or RECIST based, were coded using the reports from the baseline and first follow-up body CT scans by a single board-certified radiologist who was blinded to all clinical information other than that the patients were treated for metastatic melanoma.

PFS was defined as beginning when the therapy under evaluation was started and ending when disease progression was documented. Overall survival (OS) was defined as beginning when the therapy under evaluation was started and ending at the time of documented death or last contact. If both nab-paclitaxel and DTIC were used at different times to treat an individual patient, then the last of the two therapies

Table 1 Patient characteristics

Males (%)	47 (72.3)
Females (%)	18 (27.7)
Median age (years (range))	65 (39–92)
Median progression-free survival (months (range))	3.5 (0.5–64.6)
Fraction of patients expiring (%)	55/65 patients (84.6)
Median overall survival (months (range))	7.2 (0.8–64.6)
Median follow-up surviving patients (months (range)) ^a	20.5 (3.6–64.6)

^a Patients were followed until they either died or, if survived, to when they were last contacted

Table 3 RECIST categorization

	RECIST 1.1 category
Difference in sum of diameters of target lesions (maximum of 5)	
≥ 20% ↑	Progressive disease
≥ 30% ↓	Partial response
< 20% ↑ or < 30% ↓	Stable disease
Complete disappearance	Complete response
New lesions	
Any	Progressive disease
None	Category defaults to diameter change

Response Evaluation Criteria in Solid Tumors Version 1.1 (RECIST 1.1) which places patients in the progressive disease, stable disease, partial response, and complete response categories based on the summed percentage change in up to five total target lesions (two per organ). The interval development of new lesions automatically places patients in the progressive disease category

was evaluated in this analysis to reduce the effect of the later salvage therapy on OS. QVS of all imaging parameters, RECIST 1.1 percentage change in target lesion diameters, and standard RECIST 1.1 categorization (size change and new lesions) were compared alone and in combination with baseline, follow-up, and percentage change in serum LDH levels in their prediction of patient PFS as our primary endpoint and OS as a secondary endpoint. Median PFS and OS were determined from the individual patient PFS and OS values resulting from using the imaging or serum LDH based categorizing schemes described above. Multiple regression analysis was used to determine the correlation of the various independent variables with PFS and OS. The Cox hazard proportions ratio and Kaplan–Meier curves of the different prognostic groups were calculated with the log-rank test used to compare the statistical significance of the survival curves for the different prognostic groups. A p value of <0.05 was considered significant.

Results

QVS of overall lesion size change was the single best correlated imaging parameter with PFS and OS, although still demonstrating only fair corresponding regression coefficients of 0.519 ($p<0.0001$) and 0.375 ($p=0.0021$), respectively. The resulting correlation coefficients improved to 0.541 ($p=0.0148$) for PFS and 0.499 ($p=0.0380$) for OS when combining QVS of overall lesion size change with other parameters including the presence or absence of new lesions, brain metastases, mixed lesion response, number of organ systems involved with metastases, and serum LDH levels. Replacing QVS of overall lesion size change with RECIST size change categorization, placing patients into traditional PD ($\geq 20\%$ summed target lesion diameter

increase), PR ($\geq 30\%$ summed target lesion diameter decrease), and SD (all other summed percent target lesion diameters), resulted in deterioration in correlation with PFS to 0.412 ($p=0.3988$) and OS to 0.394 ($p=0.4426$).

The relatively wide RECIST 1.1 SD category for the overall percentage change in the summed target lesion diameters resulted in the scoring system placing most patients in the no change or SD category. QVS assessment of overall tumor lesion size change between scans resulted in 16/65 (24.6%) versus 6/58 (10.3%) patients placed in the improving category, 12/65 (18.5%) versus 30/58 (51.7%) patients placed in the no change category, and 37/65 (56.9%) versus 22/58 (37.9%) patients placed in the deteriorating category compared with RECIST 1.1 size assessment using standard categorical boundaries of $\geq 20\%$ increase or 30% decrease in target lesion diameters between scans.

QVS assessment of overall lesion size change demonstrated superior prediction of median survival differences between the three prognostic groups for either PFS or OS compared with RECIST 1.1 size assessment alone using standard categorical boundaries for PR and PD (Tables 4 and 5). The Cox hazard ratios for predicting both PFS and OS were also superior for QVS of overall lesion size change between scans compared with standard RECIST 1.1 assessment of target lesion size change using standard categorical boundaries. Reducing the RECIST 1.1 target lesion categorical boundaries to $\geq 10\%$ decrease or increase to place patients in the PR or SD groups enhanced both sensitivity in detecting both improving and deteriorating patients on therapy and superior predictions of both PFS and OS compared with standard RECIST size change categorization (Tables 4 and 5).

Combining the presence or absence of new lesions with QVS size change assessment resulted in no change in patient categorization. New lesions were assumed to place patients in the deteriorating category regardless of the size change as defined currently by RECIST 1.1. RECIST 1.1 standard size change categorization was significantly altered when combined with the assessment of new lesions. The number of patients categorized as improving or deteriorating increased with resulting improvements in prediction of both median survival of the prognostic groups as well as the Cox hazard ratios for both PFS and OS. Despite these improvements, QVS assessment of overall lesion size change still demonstrated better sensitivity in placement of patients in the improving and deteriorating groups following treatment compared with RECIST 1.1. QVS also demonstrated enhanced prediction of PFS and equivalent prediction of OS compared with standard RECIST 1.1 categorization (size change combined with new lesion development) based on the median survival times of the resulting prognostic groups and corresponding

Table 4 CT prediction of progression-free survival

Prognostic category	QVS ^a	RECIST size (+20%, -30%)	RECIST size (+10%, -10%)	RECIST 1.1
Improvement				
1st quartile PFS (SE; months)	4.8 (0.913)	4.8 (4.382)	4.8 (1.255)	4.8 (1.212)
2nd quartile PFS (SE; months)	8.3 (3.961)	9.9 (3.889)	9.9 (5.051)	5.5 (3.123)
3rd quartile PFS (SE; months)	15.6 (23.848)	11.9 (2.549)	13.9 (2.073)	11.9 (2.602)
Number of patients (%)	15/63 (23.8)	5/53 (9.4)	14/53 (26.4)	6/56 (10.7)*
Stable				
1st quartile PFS (SE; months)	3.6 (0.675)	2.1 (0.643)	2.4 (0.606)	2.4 (0.722)
2nd quartile PFS (SE; months)	5.6 (0.779)	3.6 (0.292)	3.6 (0.500)	4.0 (1.480)
3rd quartile PFS (SE; months)	9.7 (3.750)	6.2 (2.597)	5.6 (0.520)	7.9 (4.033)
Number of patients (%)	12/63 (19.0)	34/53 (64.2)	16/53 (30.2)	29/56 (51.8)
Deterioration				
1st quartile PFS (SE; months)	1.3 (0.179)	1.1 (0.254)	1.1 (0.316)	1.2 (0.207)
2nd quartile PFS (SE; months)	1.8 (0.120)	1.6 (0.281)	1.8 (0.158)	1.7 (0.229)
3rd quartile PFS (SE; months)	2.7 (0.260)	3.8 (1.151)	3.7 (0.989)	2.6 (1.464)
Number of patients (%)	36/63 (57.1)	14/53 (26.4)	23/53 (43.4)	21/56 (37.5)

^a Overall lesion size change assessment independent or combined with new lesion assessment

QVS quantitative visual scoring, RECIST 1.1 Response Evaluation Criteria in Solid Tumors (version 1.1), RECIST (+ 20%, -30%) RECIST size assessment alone using 20% increase and 30% decrease categorical boundaries for stable disease range, RECIST (+ 10%, -10%) RECIST size assessment alone using 10% increase and decrease for stable disease range, SE standard error

* $p < 0.05$

Cox hazard ratios (Tables 4, 5, and 7). Kaplan–Meier survival curves for the three prognostic groups determined by QVS of lesions size change alone demonstrated superior separation compared with standard RECIST 1.1 size change categorization. (Fig. 1a, b).

We evaluated a standard tumor marker for patient treatment response, the serum LDH level. We recorded the baseline,

completion of therapy, and percent change in these values and found the best predictions of both PFS and OS resulted from completion of therapy LDH levels (Table 6) with the change in these levels having the worst predictive value for either PFS or OS (Table 7). No improvement resulted from combining these levels with the individual or various combinations of imaging parameters in predicting PFS or OS.

Table 5 CT prediction of overall survival

Prognostic category	QVS ^a	RECIST size +20%, -30%	RECIST size +10%, -10%	RECIST 1.1
Improvement				
1st quartile PFS (SE; months)	6.0 (2.338)	5.7 (3.779)	6.0 (3.634)	5.7 (1.155)
2nd quartile PFS (SE; months)	10.5 (0.667)	10.5 (4.491)	12.4 (2.245)	6.0 (2.939)
3rd quartile PFS (SE; months)	15.4 (2.598)	13.9 (3.354)	18.0 (9.519)	13.9 (3.423)
Fraction of patients (%)	16/65 (24.6)	5/55 (9.1)	14/55 (25.5)	6/58 (10.3)*
Stable				
1st quartile PFS (SE; months)	6.8 (1.050)	5.0 (1.034)	5.0 (1.221)	5.8 (0.335)
2nd quartile PFS (SE; months)	12.2 (3.002)	7.2 (2.365)	8.4 (3.396)	10.0 (2.602)
3rd quartile PFS (SE; months)	15.2 (11.175)	12.4 (0.869)	12.2 (1.572)	12.9 (2.432)
Fraction of patients (%)	12/65 (18.5)	35/55 (63.6)	17/55 (30.1)	30/58 (51.7)
Deterioration				
1st quartile PFS (SE; months)	4.5 (0.556)	5.1 (1.827)	4.2 (2.440)	3.6 (0.313)
2nd quartile PFS (SE; months)	5.6 (0.304)	5.7 (0.869)	5.6 (0.245)	5.4 (0.313)
3rd quartile PFS (SE; months)	9.3 (0.783)	10.3 (1.859)	7.2 (2.758)	9.3 (0.724)
Fraction of patients (%)	37/65 (56.9)	15/55 (27.3)	24/55 (43.6)	22/58 (37.9)

^a QVS overall lesion size change assessment independent or combined with new lesion assessment

QVS quantitative visual scoring, RECIST 1.1 Response Evaluation Criteria in Solid Tumors (version 1.1), RECIST (+ 20%, -30%) RECIST size assessment alone using 20% increase and 30% decrease categorical boundaries for stable disease range, RECIST (+ 10%, -10%) RECIST size assessment alone using 10% increase and decrease for stable disease range, SE standard error

* $p < 0.05$

Table 6 Completion of treatment serum lactic dehydrogenase level prediction of survival

	Serum LDH level (IU/L)	Quartile PFS in months (SE)	Fraction of patients (%)	Quartile OS in months (SE)	Fraction of patients (%)
<i>LDH completion of Rx serum lactic dehydrogenase level, IU/L international units per liter, Rx treatment, PFS progression-free survival, OS overall survival, SE standard error</i>	LDH (≤ 180)	–	28/61 (45.9%)		28/63 (44.4%)
	1st quartile	1.9 (0.687)	–	5.7 (0.956)	–
	2nd quartile	4.1 (0.706)	–	10.5 (1.323)	–
	3rd quartile	7.8 (4.353)	–	15.4 (2.406)	–
	LDH (>180 and <250)		10/61 (16.4%)		12/63 (19.1%)
	1st quartile	1.2 (0.725)	–	4.5 (3.000)	–
	2nd quartile	2.7 (1.186)	–	10.3 (1.645)	–
	3rd quartile	9.7 (6.325)	–	18.0 (10.950)	–
	LDH (≥ 250)		23/61 (37.7%)		23/63 (36.5%)
	1st quartile	1.3 (0.263)	–	3.6 (1.895)	–
	2nd quartile	1.9 (0.539)	–	5.4 (0.240)	–
	3rd quartile	3.8 (0.989)	–	6.7 (0.791)	–

Discussion

RECIST the current standard has been found to be deficient to use in clinical trials to predict patient survival by oncologists, radiologists and pharmaceutical companies [7]. RECIST is limited to anatomic evaluation of lesion size change between scans [1, 2]. Many additional potentially useful prognostic imaging findings are therefore not incorporated. These have to date been investigated in limited studies, which have prevented them from being universally adopted due to lack of a statistically adequate number of patients to do so. Changes in tumor perfusion on both CT [8] and MR [9], oxygenation and water diffusion by MRI [10], and lesion morphology on CT [11], have been compared with RECIST assessment and

found in limited evaluations to improve prediction of survival. Survival, the gold standard in these evaluations, has varied with the Cox hazard proportions ratio, median survival, and Kaplan–Meier survival curves used to compare the results from the different imaging scoring methods.

Furthermore RECIST, even in evaluating size, the one imaging parameter the scoring system has been developed to address, has severe limitations in doing so. By definition, it is restricted to a limit of only five target lesions by the current version, RECIST 1.1 [2]. This restriction prevents its quantitative application to many other readily observable changing potentially relevant lesions in the assessment of tumor response. In addition, it uses only the single longest tumor dimension (other than for lymph nodes in which the short axis is measured) which has been found inadequate in reflecting tumor 3-D volume [3]. This shortcoming is exacerbated in lesions with complex shapes, ill defined borders, or areas of non-viable tumor [12]. Another RECIST limitation is that it prohibits use of lesions below a 1.0-cm diameter size threshold, which eliminates many readily evaluable lesions such as small lung nodules or other small organ metastases which clearly change between scans but cannot be used. Lesions between 5 mm and 1.0 cm in diameter are readily observed on current CT scanners.

RECIST uses categories of target lesion summed size change, which are broad, and have been found relatively insensitive to change [13]. This issue has become increasing more important with new therapies becoming available which the patient can be transitioned to which may improve survival, or reducing the need to continue with expensive, futile toxic therapies. Wide variability exists between different observers in applying even the simplified single diameter measurements required by RECIST 1.1 between observers at the same site and at multiple sites [14].

The current project further developed a user-friendly visual scoring method (QVS) which incorporates a Likert-type scale

Table 7 Cox hazard ratios for predicting survival

	Cox hazard proportion ratio for PFS	Cox hazard proportion ratio for OS
Start of Rx LDH level	1.345	1.464
End of Rx LDH level	1.624	1.799
20% change in LDH level	1.192	0.898
Quantitative visual scoring	3.070	1.758
RECIST size (+20%, –30%)	0.645	0.829
RECIST size (+10%, –10%)	0.607	0.753
RECIST 1.1	1.860	1.785

Quantitative visual scoring results in same Cox hazard ratios when applied to overall lesion size change alone or combined with new lesions. RECIST 1.1 corresponds to current standard which includes defined target lesion summed diameter size change categorical boundaries and new lesion development. RECIST% change categories refer to summed target lesion categorical boundaries with ranges for stable disease category indicated

PFS progression-free survival, *OS* overall survival, *Rx* treatment, *LDH* serum lactic dehydrogenase

Fig. 1 **a, b** Kaplan–Meier survival curves plotting progression-free survival (*Cum. Survival*) against the number of months (horizontal axis). **a** QVS placement of patients into increasing, no change, and decreasing overall lesion size categories corresponding to patients deteriorating, stable, and improving on treatment. Differences in survival prediction between curves were statistically significant ($p < 0.0001$, log-rank test). **b** The RECIST 1.1 categorization based on target lesion size change alone without new lesions. *PD* progressive disease, *SD* stable disease, *PR* partial response. Differences in survival prediction between curves were statistically significant ($p = 0.0014$, log-rank test). QVS demonstrates improved separation in categorizing patients compared to RECIST 1.1 target lesion size categorization particularly when comparing the patients deteriorating on treatment from the stable and improving categories

[15, 16], for scoring radiologists' observations of the interval development of new or changing imaging findings. Use of QVS exploits the large body of established work [17–23] from the scientific discipline of human perception. Previous investigators have found the use of QVS to quantify the vast array of perceived imaging findings identified in daily practice by imagers thereby extending the number of evaluable parameters beyond lesion size change, to which RECIST or volumetric analysis are restricted. Prior work has found QVS of the change in lesion size between scans to be often superior to the RECIST method in the consideration of lesions with complex shapes, large interfaces with adjacent

normal structures (such as a pleural surface), or nodules too small to be considered measurable by RECIST [24]. Furthermore, QVS can more quickly and flexibly be adapted to incorporate new relevant imaging features, which may be important in predicting patient response and can more easily be introduced into the practice setting than the hardware and software upgrades required for techniques involved in computer aided diagnosis or 3-D lesion volumetric analysis [13, 24, 25].

We employed simplified Likert-type scoring scales to assess the presence or absence of an imaging finding (i.e., new lesion development) or the interval change in an imaging feature (i.e., size change) to relate quantitatively single or combined imaging findings to patient outcome (PFS or OS). Our analysis confirmed those of prior investigations coding human perception of size change of neoplastic lesions in that it was more sensitive to change in patient status on therapy and more predictive of their survival than the currently used RECIST scoring paradigm [24–26]. However it should be noted that imaging, was only slightly superior to a relatively inexpensive serum marker, the serum LDH level, in predicting patients likely to do well on therapy. Combining imaging and serologic parameters did not improve prediction of patient survival on treatment.

We envision a simplified scale, like the one we used, could be placed in structured reports and incorporated into radiology information systems if supported by vendors [27]. Radiologists could then quantitatively record the pattern based imaging features they see in their daily practices in structured reports using standardized terms, such as in the RadLex lexicon currently being developed by the RSNA [28]. The value of the imaging features recorded would be determined based on their reproducibility and accuracy in predicting patient outcomes, such as survival. A tumor registry could be developed assessing the accumulated QVS of a variety of imaging findings (size change as well as other features) recording both the scored imaging findings and the patient outcomes which they independently predict. Improvements would be necessary in linking the radiology information system with the remainder of the patient electronic medical record to allow this to occur however.

The current project was limited in that imaging interpretation data was extracted from the radiology reports by a single radiologist and not re-read to address reproducibility. However, three prior studies with two other metastatic cancers, colon and lung [24–26], have demonstrated improved agreement between radiologists and physician assistants when re-reading cases using Likert-type scales to score imaging findings. The Likert-type scale used in this project was simpler than ones utilized previously and therefore less likely to be subject to inter or intraobserver disagreements. We believe our results are sufficiently promising to warrant subsequent testing, both extracting

information from radiology reports (possibly using natural language processing programs), or from having observers re-interpret the images to measure the reproducibility of QVS in multiple reader, multisite evaluations.

A relatively short survival tumor, metastatic melanoma, was used in the current investigation because the primary and secondary survival endpoints were short and therefore readily measurable. The survival times of patients with metastatic malignancies such as melanoma, colon, and lung are still very short despite active research in these areas [29]. New drugs are being rapidly brought into clinical trials and require early evaluation to promote quick development of potentially successful therapies and termination of therapies likely to be futile [30–33].

Our study population was relatively small and additional studies are required to establish QVS as a new tumor response scoring paradigm in a variety of metastatic cancers including melanoma and other types of cancers. QVS can be used not only to code CT findings but could be potentially expanded flexibly to incorporate visual-based features from other modalities such as MRI, positron emission tomography, ultrasound, and emerging modalities which have visual-based features requiring human interpretation. Current size-based imaging parameters are only fair in predicting survival as measured in this study in metastatic melanoma patients and in other treated malignancies [25, 26]. QVS provides a methodology for scoring a variety of imaging findings which can be compared in how they predict patient outcomes with the current largely size-based anatomic parameters as a standard of reference. Visually apparent optical intensity changes coding tumor metabolic activity in response to therapy can be assessed by QVS in the same manner as was done with the size-based anatomic parameters in the current project.

In summary new therapies are currently being developed for metastatic cancer which need to be quickly evaluated for their impact on patient survival. QVS promises to enhance patient response categorization compared to RECIST 1.1, the current imaging surrogate endpoint used by the Food and Drug Administration to aid in the drug approval process and by pharmaceutical companies [33]. QVS in this project demonstrated superior discrimination compared with RECIST 1.1 in predicting patients likely to improve or deteriorate on treatment for metastatic melanoma early in the course of treatment. We believe our results justify incorporating QVS in other larger retrospective and prospective studies utilizing CT and other modalities to assess early tumor therapy response.

References

1. Therasse P, Arbuck SG, Eisenhauer EA, et al: New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst* 92(3):205–216, 2000
2. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J: New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 45(2):228–247, 2009
3. Zhao B, Schwartz LH, Moskowitz CS, Ginsberg MS, et al: Lung cancer: computerized quantification of tumor response-initial results. *Radiology* 241(3):892–898, 2006
4. Korn EL, Liu PY, Lee SJ, Chapman JA, et al: Meta-analysis of phase II cooperative group trials in metastatic stage IV melanoma to determine progression-free and overall survival benchmarks for future phase II trials. *J Clin Oncol* 26(4):527–534, 2008
5. Liu F, Zhao B, Krug L, Ishill N, et al: Assessment of therapy responses and prediction of survival in malignant pleural mesothelioma through computer-aided volumetric measurement on computed tomography scans. *J Thorac Oncol* 5(6):879–884, 2010
6. Balch CM, Gershenwald JE, Soong SJ, Thompson JF, et al: Final version of 2009 AJCC melanoma staging and classification. *J Clin Oncol* 27(36):6199–6206, 2009. Epub 2009 Nov 16
7. Husband JE, Schwartz LH, Spencer J, Olivier L, et al: Evaluation of the response to treatment of solid tumours—a consensus statement of the International Cancer Imaging Society. *Br J Cancer* 90(12):2256–2260, 2004
8. Turkbey B, Kobayash H, Ogawa M, Bernardo M, Choyke PL: Imaging of tumor angiogenesis: functional or targeted? *AJR* 193:304–313, 2009
9. Van Beers BE, Vilgrain V: Biomarkers in abdominal imaging. *Abdominal Imaging* 34(6):663–667, 2009
10. Jordan BF, Runquist M, Raghunand N, Baker A, et al: Dynamic Contrast-enhanced and diffusion mri show rapid and dramatic changes in tumor microenvironment in response to inhibition of HIF-1 α using PX-47. *Neoplasia* 7(5):475–485, 2005
11. Nishino M, Guo M, Jackman DM, DiPiro PJ, et al: CT tumor volume measurement in advanced non-small-cell lung cancer: performance characteristics of an emerging clinical tool. *Acad Radiol* 18(1):57–62, 2011
12. Choi H, Chamsangavej C, Faria SC, Macapinlac HA, Burgess MA, Patel SR, Chen LL, Podoloff DA, Benjamin RS: 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* 25:1753–1759, 2007
13. Prasad SR, Jhaveri KS, Saini S, Hahn PF, Halpern EF, Sumner JE: CT tumor measurement for therapeutic response assessment: comparison of unidimensional, bidimensional, and volumetric techniques initial observations. *Radiology* 225(2):416–419, 2002
14. Jaffe CC: Measures of response: RECIST, WHO, and new alternatives. *J Clin Oncol* 24(20):3245–3251, 2006
15. Likert R: A technique for the measurement of attitudes. *Archives of Psychology* 140:1–55, 1932
16. Rogers MP, Orav J, Black PM: The use of a simple Likert Scale to measure quality of life in brain tumor patients. *J Neurooncol* 55(2):121–131, 2001
17. Krupinski EA, Kundel HL, Judy PF, Nodine CF: The medical image perception society. Key issues for image perception research. *Radiology* 209(3):611–612, 1998
18. Krupinski EA, Nodine CF, Kundel HL: A perceptually based method for enhancing pulmonary nodule recognition. *Invest Radiol* 28(4):289–294, 1993
19. Kundel HL, Polansky M: Measurement of observer agreement. *Radiology* 228(2):303–308, 2003. Epub 2003 Jun 20
20. Beam CA, Krupinski EA, Kundel HL, Sickles EA, Wagner RF: The place of medical image perception in 21st-century health care. *J Am Coll Radiol* 3(6):409–412, 2006. Review

21. Krupinski EA, Kundel HL: Update on long-term goals for medical image perception research. *Acad Radiol* 5(9):629–633, 1998
22. Kundel HL: Medical image perception. *Acad Radiol* 2(Suppl 2): S108–S110, 1995
23. Krupinski EA, Nodine CF, Kundel HL: Perceptual enhancement of tumor targets in chest Xray images. *Percept Psychophys* 53(5):519–526, 1993
24. Gottlieb RH, Kumar P, Loud P, Klippenstein D, et al: Semiquantitative visual approach to scoring lung cancer treatment response using computed tomography: a pilot study. *J Comput Assist Tomogr* 33(5):743–747, 2009
25. Gottlieb RH, Raczky C, Hanna T, Fora A, et al: Quantitative methodology using CT for predicting survival in patients with metastatic colorectal carcinoma: a pilot study. *Clin Imaging* 34(3):196–202, 2010
26. Gottlieb RH, Litwin A, Gupta B, Taylor, et al: Qualitative radiology assessment of tumor response: does it measure up? *J Clin Imaging* 32(2):136–140, 2008
27. Langlotz CP: Automatic structuring of radiology reports: harbinger of a second information revolution in radiology. *Radiology* 222(1):5–7, 2002
28. Langlotz C: RadLex: a new method for indexing online educational materials. *Radiographics* 26:1595–1597, 2006
29. Jemal A, Siegel R, Xu J, Ward E: Cancer Statistics 2010 American Cancer Society. *CA Cancer J Clin* 60:277–300, 2010
30. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, et al: Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* 363(8):711–723, 2010. Epub 2010 Jun 5
31. Smith JJ, Sorensen AG, Thrall JH: Biomarkers in imaging: realizing radiology's future. *Radiology* 227(3):633–638, 2003
32. Pien HH, Fischman AJ, Thrall JH, Sorensen AG: Using imaging biomarkers to accelerate drug development and clinical trials. *Drug Discov Today* 10(4):259–266, 2005
33. Johnson JR, Williams G, Pazdur R: End points and United States Food and Drug Administration approval of oncology drugs. *J Clin Oncol* 21(7):1404–1411, 2003