Tool Support to Enable Evaluation of the Clinical Response to Treatment

Mia A. Levy, MD, Daniel L. Rubin, MD, MS
Stanford University, Stanford, CA

Abstract

Objective criteria for measuring response to cancer treatment are critical to clinical research and practice. The National Cancer Institute has developed the Response Evaluation Criteria in Solid Tumors (RECIST)[1] method to quantify treatment response. RECIST evaluates response by assessing a set of measurable target lesions in baseline and follow-up radiographic studies. However, applying RECIST consistently is challenging due to inter-observer variability among oncologists and radiologists in choice and measurement of target lesions. We analyzed the radiologist-oncologist workflow to determine whether the information collected is sufficient for reliably applying RECIST. We evaluated radiology reports and image markup (radiologists), and clinical flow sheets (oncologists). We found current reporting of radiology results insufficient for consistent application of RECIST, compared with flow sheets. We identified use cases and functional requirements for an informatics tool that could improve consistency and accuracy in applying methods such as RECIST.

Introduction

Objective criteria for measuring response to cancer treatment are critical to clinical research and practice. Therapeutic clinical trials for metastatic cancer often use radiographic imaging studies to visualize and measure cancer lesions at baseline and again at follow-up in order to evaluate response to treatment. Several heuristic methods have been developed to aid clinical researchers in quantifying the response to treatment including: the WHO criteria[2] which has more recently been replaced by the National Cancer Institute's Response Evaluation Criteria in Solid Tumors (RECIST)[1], and the International Harmonization Project for Response Criteria in Lymphoma[3]. These evaluation methods have complex rules for defining a set of measurable cancer lesions called target lesions, and for classifying response to treatment based on temporal changes in the size and metabolic activity of the target lesion set. However, application of these methods accurately and consistently in practice is challenging for two key reasons. First, the patient's images are evaluated by radiologists and oncologists independently, and there is a high degree of inter-observer variation in selection and measurement[4] of target lesions.

Second, the way information is recorded and reported differs between oncologists and radiologists. In the current clinical workflow, the oncologist orders an imaging study, the patient has the study performed and the radiologist is first to review the images. The radiologist summarizes their findings in a text report and records detailed measurements as image markups, using qualitative methods to report response. The report is then sent to the oncologist who independently reviews the report and images. In the context of clinical trials, oncologists often use flow sheets to record the quantitative aspects of the target lesions, and calculate a quantitative response rates using methods such as RECIST. Some institutions will share the RECIST flow sheets with the radiologists for evaluation of the follow-up study, however these flow sheets are not incorporated into the radiology report that composes the official medical record.

We hypothesize that the current radiologist-oncologist communication paradigm is not optimally coordinated with respect to tracking target lesions of interest, resulting in incomplete information for evaluating disease status being recorded in the medical record. In this paper, we study and report the deficiencies in this workflow and communication process related to evaluating quantitative criteria for measuring disease response. We also propose functional requirements for a tool to support and enable robust evaluation of the clinical response to cancer treatment.

Background

The Response Evaluation Criteria in Solid Tumors (RECIST) has a complex set of rules for defining the set of target and non-target lesions to be tracked over time. A simplified version of these rules defines a target lesion as at least 10 mm in greatest dimension by CT scan. CT scans are the most common imaging modality used when applying RECIST, and for most solid tumors, CT scans of the chest, abdomen and pelvis are used for monitoring. A maximum of ten measurable lesions are included in the target lesion set with no more than five lesions per organ. Certain types of lesions are not considered measurable such as pleural effusions and bone lesions and are thus classified as non-target lesions. Once the set of baseline target lesions has been identified, the sum of the longest diameters (SLD) is calculated as a

surrogate for tumor volume. The target and non-target lesions are tracked in follow-up studies with calculation and comparison of the SLD to evaluate response to treatment.

Response is classified into four categories as follows:
- Complete Response: disappearance of all lesions
- Partial Response: no new lesions AND at least a 30% decrease in the SLD
- Stable Disease: no new lesions AND between 30% decrease and 20% increase in SLD
- Progressive Disease: increase in number of lesions OR at least 20% increase in SLD

For example, Figure 1 shows an axial CT scan image of a patient with metastatic colon cancer with more than 10 measurable liver lesions. Image markups are shown for four of these liver lesions on a single slice. RECIST requires that a maximum of 5 lesions be selected as target lesions and the rest classified as non-target lesions. However, without coordination, there is variation between observers in selecting the set of target lesions. There is also observer variation in measuring the lesions; oncologists usually measure the longest dimension to comply with RECIST, but for some types of lesions such as lymph nodes, it is routine practice for radiologists to measure the shortest dimension.

Figure 1: CT scan showing 4 measurable liver lesions with image markup of longest diameter
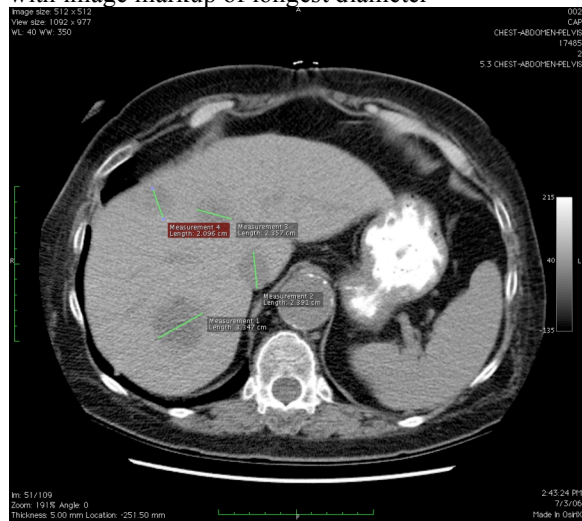
We believe there is an opportunity to develop informatics methods to improve the coordination and accuracy of physicians who wish to use quantitative response evaluation criteria such as RECIST. Several informatics methods could be developed to support image-based criteria to evaluate response to treatment including: data structures for semantic image annotation and markup, automated algorithms to apply heuristic classifiers, visualization methods, as well as a strategy for integrating these tools into operational systems supporting the clinical workflow.

In previous work, we developed a prototype tool utilizing visualization methods to summarize large amounts of imaging data by applying temporal abstraction methods to image annotations[5]. Image annotations describe the meaning in images, while markup is the visual presentation of the annotations. The Annotation and Image Markup (AIM) Project of the National Cancer Institute's cancer Biomedical Informatics Grid (caBIG)[6] has subsequently developed a minimal information model necessary to record an image annotation. The AIM project created an ontology and schema for image annotations utilizing the RadLex terminology for anatomic structures and observations, and is currently developing a tool to collect image annotations.

However, application of response evaluation criteria also requires methods that enable communication and cooperative work between the radiologist reviewing the imaging study and the oncologist making treatment decisions based on the study's findings. We evaluated this workflow process to better understand the deficiencies in this communication, and we have developed a set of functional requirements for a tool to support the quantitative evaluation of response to treatment. This tool may improve the accuracy and consistency of clinicians in applying such methods.

Methods

We obtained Institutional Review Board exemption for this study, given that all data was de-identified. Our preliminary study evaluated reports and images obtained of CT scans of the chest, abdomen and pelvis from patients enrolled in clinical trials at Stanford Comprehensive Cancer Center and Fox Chase Cancer Center where RECIST criteria were being used to evaluate response to treatment in patients with metastatic disease. Our goal was to evaluate the consistency of the radiologist reports and image markup with respect to providing sufficient information for an oncologist to apply RECIST methods to evaluate response to treatment. In order for a radiologist's report to be sufficient to calculate the response rate, the longest dimension of all target lesions would need to be reported for each baseline and follow-up study. Target lesions had been identified at baseline by the oncologists utilizing the radiology report and image markup, and recorded on RECIST flow sheets. We evaluated the radiology reports and image markup to identify if the target lesions were identified and the longest diameter measured for each baseline and follow-up study, such

that the response rate could be calculated. We have previously described our methods in detail in the Stanford subset of our data[7].

We used the results of these observations to identify areas in the workflow and reporting process where there were deficiencies in communication and application of the RECIST methods. These deficiencies present opportunities for a tool to help physicians to apply RECIST in clinical practice. From our analysis we developed functional requirements for such an informatics tool and a set of use cases to illustrate the potential value to improve communication, workflow, and enhanced accuracy in applying quantitative criteria such as RECIST.

Results: Workflow Analysis

De-identified DICOM images, image markups, radiology reports and RECIST flow sheets were acquired for thirteen patients with a total of 42 CT scans (13 baseline, 29 follow-up) interpreted by 16 different radiologists. Each patient in our study had at least three imaging studies, one baseline and two follow-ups. Fifty-five target lesions were identified at baseline by the oncologists (average 4 lesions per patient), and imaged a total of 167 times across studies.

At baseline, 71% of target lesions specified by the oncologist were identified in radiology reports and 73% by image markup, while the longest diameter was reported 55% of the time and marked up in the images 50% of the time. At follow-up 38% of target lesions were identified in reports and 70% by image markup, while the longest diameter was reported only 28% of the time and marked up 26% of the time. In only 26% of the studies were the report and image markup data sufficient to calculate the quantitative response rate as define by RECIST.

Table 1: Results of workflow analysis

	Baseline	Follow-up
Report Identified Lesion	71%	38%
Image Markup Lesion	73%	70%
Longest Diameter Reported	55%	28%
Longest Diameter Marked	50%	26%

We found that while radiologists do not often report the presence or dimensions of target lesions in follow-up studies, they do create image markups for these lesions suggesting that they are evaluating but not reporting on these findings. Yet target lesion data is precisely what the oncologist needs to calculate the response rate. These findings suggest that there is a lack of communication between the radiologist and oncologist in defining which lesions are to be tracked and reported. This deficit in communication informs our use cases which are described below.

Results: Use Cases

We describe four use cases for a Quantitative Criteria Support Tool (QCST) to help practitioners apply RECIST in the clinical trial setting. These use cases follow the sequence of steps the radiologist and oncologist encounter with patient information during the process of care, commencing with the baseline study and concluding with the follow up studies.

Use Case 1: Radiologist Baseline Study Evaluation

The *goals of the radiologist* in evaluating the *baseline study* is to review the images, identify and mark image abnormalities, create semantic image annotations, and dictate a report summarizing those findings. The radiologist first loads the DICOM images into the QCST and reviews the images looking for radiographic abnormalities. When an abnormality is identified, the radiologist selects an image markup tool in QCST and creates a visual representation of the abnormality. When the markup is created, the QCST assigns a unique number to the abnormality, recording the location of the abnormality including the study series, image and pixel numbers, and finding features derived from the markup tool including finding dimensions, density or metabolic activity. The QCST system displays a record of the abnormality in the Study Findings Table depicted in Figure 1. Using the Study Finding Table and the constraints of the RECIST rules, the radiologist further annotates the abnormality with semantic information including the finding type, anatomic location, and qualitative descriptors; and classifies the abnormality as pathologic, non-pathologic, or of uncertain pathologic significance. The radiologist may add, modify or delete a given semantic annotation or image markup for an abnormality, or may delete the finding completely. The QCST system however maintains a transaction log of all actions performed. The key functionality provided by QCST is the unambiguous identification of the abnormalities—the location of lesions and indication of how they were measured.

Use Case 2: Oncologist Baseline Study Review

The *goal of the oncologist* in reviewing the *baseline study* is to gain an understanding of the distribution of the patient's disease and to identify measurable cancer lesions that can be tracked over time to evaluate response to treatment. The oncologist's role is to identify a set of target and non-target lesions using the RECIST criteria that are used to calculate the SLD. The QCST system assists the oncologist by

categorizing the abnormalities observed by the radiologist as target and non-target lesions utilizing the RECIST rules and summarized in the Patient Findings Table shown in Figure 2. The oncologist can modify the system's choice of target lesions, and the radiologist's image markups and annotations utilizing the Patient Findings Table or the Study Findings Table. The QCST system also automatically calculates and displays the SLD.

Use Case 3: Radiologist Follow-up Study

The *goal of the radiologist* in evaluation of the *follow-up study* is to review the new imaging study in the context of the old study with careful attention to direct comparison of the abnormalities over time as well as identification and annotation of any new abnormalities. In the context of RECIST, it is especially important for the radiologist to specifically mark and annotate target and non-target lesions. The QCST system assists the radiologist in this task by identifying the target and non-target lesions that are of interest to the oncologist. The radiologist can use the Patient Finding Table to select and view abnormalities in the previous imaging study for the patient and to utilize the semantic data already

Figure 1: Example of completed Study Finding Table created by radiologist at time of baseline study

Study Type: CT scan Chest/Abdomen/Pelvis			Study Date: 6/10/2007				
Finding Number	Finding Type	Finding Location	Dimensions	Qualitative Description	Series & Image Number(s)	Pathologic Finding
1	Lung Nodule	Left Middle Lung Lobe	2.3x2.1 cm	Well circumscribed	Series 2, image 45	Yes
2	Lymph Node	Pre-tracheal Lymph Node	2.4x1.2 cm		Series 3, image 35	Yes
3	Pleural Effusion	Right pleural space	NA	Small	Series 3, image 55	Yes
4	Bone Lesion	C4 vertebral body	NA	Compression fracture	Series 5, image 23	Yes
5	Lymph Node	Left supraclavicular Lymph Node	0.9 x 0.5 cm		Series 3, image 22	No
6	Gallstones	Gallbladder	NA	Non-obstructing	Series 3, image 78	No
7	Renal cyst	Left Kidney	2.1x3.1 cm		Series 5, image 88	No
8	Lung nodule	Right upper lung lobe	0.8 cm	Irregular	Series 2, image 33	Yes
9	Lung nodule	Right upper lung lobe	0.2 cm	Calcified	Series 2, image 39	Unknown

Figure 2: Example of Patient Finding Table populated with baseline study finding annotations

		Finding Type	Finding Location	Baseline: 6/10/2007 Dimensions	Follow-up: 9/12/2007 Dimensions	Follow-up: 12/1/2007 Dimensions
Pathologic Findings	Target Lesions	Lung Nodule	Left Middle Lung Lobe	2.3x2.1 cm	1.8x1.5 cm	1.6x1.3 cm
		Lymph Node	Pre-tracheal Lymph Node	2.4x1.2 cm	2.0x0.9 cm	1.5x0.7 cm
		Total Number Target Lesions		2	2	2
		Sum Longest Diameters		4.7	3.8	3.1
		Response Rate		NA	19.1%	34%
		Response Category		NA	Stable Disease	Partial Response
	Non-Target Lesions	Pleural Effusion	Right pleural space	Moderate	Small	Resolved
		Bone Lesion	C4 vertebral body	Compression fracture	Stable	Stable
		Lung nodule	Right upper lung lobe	0.8 cm	0.5	0.5
		Total Number Non Target Lesions		3	3	3
Findings of unknown pathologic significance		Lung nodule	Right upper lung lobe	0.2 cm	Undetectable	Undetectable
Non Pathologic Findings		Lymph Node	Left supraclavicular lymph node	0.9 x 0.5 cm	0.8x0.5 cm	0.8x0.5 cm
		Gallstones	Gallbladder	NA	NA	NA
		Renal cyst	Left Kidney	2.1x3.1 cm	2.0x3.1 cm	2.0x3.0 cm

associated with the previous abnormality to create new image markups and annotations on the abnormality at the new time point. The SLD, Response Rate and Response Classification are automatically calculated and displayed in the Patient Finding Table for use by the radiologist during their dictation, eliminating the need for hand-calculation and avoiding potential error.

Use Case 4: Oncologist Follow-up Study

The *goal of the oncologist* during review of the *follow-up study* is to gain an understanding of how the patient is responding to treatment thru application of a quantitative evaluation method such as RECIST. Ideally, the radiologist has adequately annotated the target lesions in the follow-up study for the QCST system to calculate the SLD, response rate and response category. If not, the oncologist can use the same features as the radiologist in Use Case 3 to complete the image markup and annotations of lesions. Ideally, there would be nothing more for the oncologist to do other than to review the images, image annotations and calculations to confirm that measurements were made to their satisfaction. When the oncologist highlights an abnormality in the Patient Finding Table, the image annotations are displayed for both the new and the old study.

Functional Requirements

From our use cases, we have identified the following functional requirements for an informatics tool to help coordinate oncologists and radiologists in application of quantitative criteria such as RECIST during the workflow of patient encounters in the cancer care process:

- **Explicit representation of the rules comprising the quantitative response evaluation method** (RECIST)
  - o To enable real-time feedback in image markup and semantic annotation of radiographic findings required for performing the quantitative evaluation
  - o To enable semi-automated classification of findings as target and non-target lesions
  - o To automate calculation of response rates and classification of response
- **Display of quantitative lesion data in a given study and across studies for the patient**
  - o Display finding markup features
  - o Display finding semantic annotations
- **Dynamic navigation methods for reviewing and creating descriptions and measurements of lesions seen in the images**
  - o Navigation between finding tables and respective DICOM images and markup
  - o Navigation between finding tables and the annotation of new image abnormalities

We are currently developing the QCST tool, adopting these functional requirements in its design.

Conclusion

Current reporting of radiology results in the medical record focuses on summarizing imaging findings, which is insufficient for consistent application of quantitative methods to evaluate response to treatment. Our results illustrate the need for new tools to improve radiologist-oncologist workflow coordination during the cancer care process to enable more complete and consistent quantitative evaluation of response to treatment. We further describe the use cases and functional requirements for a tool to aid communication between radiologists and oncologist, and application of the complex response evaluation rules. We acknowledge that radiologists may be resistant to the additional work of creating image annotations. However, our observations suggest that radiologists are already performing the task of image markup, though their current tools do not guide or inform them as to what or how to annotate abnormalities. With careful attention to workflow considerations we hope that a tool such as we describe for QCST would make image annotation an efficient method for communicating image findings. Future work will focus on the development and testing of this tool.

References

1. Therasse P, et al. New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst. 2000 Feb 2;92(3):205-16.
2. WHO handbook for reporting results of cancer treatment. Geneva (Switzerland): World Health Organization Offset Publication No. 48; 1979.
3. Cheson BD. (2007). The International harmonization project for response criteria in lymphoma clinical trials. *Hematol Oncol Clin North Am,* **21**, 841-54.
4. Belton AL, Saini S, Liebermann K, et al. (2003). Tumour size measurement in an oncology clinical trial: comparison between off-site and on-site measurements. *Clin Radiol,* **58**, 311-4.
5. Levy MA, Garg A, Tam A, Rubin D, Garten Y. Lesion Viewer: A tool for tracking cancer lesions over time. AMIA 2007 Proceedings.
6. Rubin DL, Mongkolwat P, Kleper V, Superkar K, Channin D. Medical imaging on the semantic web: annotation and image markup. AAAI. 2007
7. Levy, MA, Rubin DL, et al. Radiologist-Oncologist Workflow Analysis in application of RECIST. ASCO 2008 Proceedings.