

Evaluation of a cloud-based local-read paradigm for imaging evaluations in oncology clinical trials for lung cancer

Naoko Sueoka-Aragane^{1,*}, Naomi Kobayashi^{1,*}, Eric Bonnard², Colette Charbonnier³, Junta Yamamichi⁴, Hideaki Mizobe⁴ and Shinya Kimura¹

Acta Radiologica Open
4(12) 1–7
© The Foundation Acta Radiologica
2015
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2058460115588103
arr.sagepub.com



Abstract

Background: Although tumor response evaluated with radiological imaging is frequently used as a primary endpoint in clinical trials, it is difficult to obtain precise results because of inter- and intra-observer differences.

Purpose: To evaluate usefulness of a cloud-based local-read paradigm implementing software solutions that standardize imaging evaluations among international investigator sites for clinical trials of lung cancer.

Material and Methods: Two studies were performed: KUMO I and KUMO I Extension. KUMO I was a pilot study aiming at demonstrating the feasibility of cloud implementation and identifying issues regarding variability of evaluations among sites. Chest CT scans at three time-points from baseline to progression, from 10 patients with lung cancer who were treated with EGFR tyrosine kinase inhibitors, were evaluated independently by two oncologists (Japan) and one radiologist (France), through a cloud-based software solution. The KUMO I Extension was performed based on the results of KUMO I.

Results: KUMO I showed discordance rates of 40% for target lesion selection, 70% for overall response at the first time-point, and 60% for overall response at the second time-point. Since the main reason for the discordance was differences in the selection of target lesions, KUMO I Extension added a cloud-based quality control service to achieve a consensus on the selection of target lesions, resulting in an improved rate of agreement of response evaluations.

Conclusion: The study shows the feasibility of imaging evaluations at investigator sites, based on cloud services for clinical studies involving multiple international sites. This system offers a step forward in standardizing evaluations of images among widely dispersed sites.

Keywords

Thorax, computed tomography (CT), lung, adults, computer applications

Date received: 28 October 2014; accepted: 29 April 2015

Introduction

Clinical trials to determine the anti-cancer effects of chemotherapeutic agents are indispensable for developing new strategies for cancer treatment. Although prolonging overall survival time is the ultimate purpose of cancer treatment, it is difficult to apply overall survival time as an endpoint, since the use of crossover therapy and therapy after progression is increasing (1–4). Since molecular targeted therapy, in particular, has shown a significant anti-cancer effect in specific populations compared to traditional chemotherapy, crossover

¹Division of Hematology, Respiratory Medicine and Oncology, Department of Internal Medicine, Faculty of Medicine, Saga University, Saga, Japan

²Radiology Department, NICE University Hospital, France

³MEDIAN Technologies, France

⁴Global Healthcare IT Project, Canon Inc., Japan

*Equal contributors.

Corresponding author:

Naoko Sueoka-Aragane, Division of Hematology, Respiratory Medicine and Oncology, Department of Internal Medicine, Faculty of Medicine, Saga University, 5-1-1 Nabeshima, Saga 849-8501, Japan.
Email: sueokan@cc.saga-u.ac.jp



treatment confounds the ability to determine the efficacy of treatment. Therefore, progression-free survival is frequently applied to clinical trials for solid tumors as a surrogate marker of overall survival (2–4). To obtain precise results of progression-free survival, accurate evaluation of tumor response is critical. To achieve objective, reproducible evaluations, an independent central review (ICR), especially blinded ICRs of radiologic images have been employed recently in many clinical trials (2,5–8). Establishing an ICR requires the highest level of site compliance and operational efficacy. In addition, rapid, real-time evaluation of progressive disease by ICR is required to eliminate non-eligible cases that have proceeded to protocol of trials without evaluation by reviewers. Since the main reasons for discordance between reviewers and investigators were differences in lesion selection, inter-reader variability, and perception of new lesions according to the previous reports (1,4), a cloud-based automated imaging system that enhances collaboration between investigators and reviewers should be useful.

Widely used standards for measuring the response of tumor are the Response Evaluation Criteria In Solid Tumor (RECIST) and the World Health Organization (WHO) criteria. The former is one-dimensional (1D) and the latter is two-dimensional (2D) (9,10). RECIST was proposed and established in 2000 and an updated version was published in 2009 (11). It is based on the rationale that maximum diameters are linearly related to cell kill compared to bi-dimensional evaluation, and on evidence showing the agreement between 1D and 2D evaluations (10). RECIST has been applied in many clinical trials, and although quite useful, it does have some problems. Considerable intra- and inter-observer variability has been noted, especially in tumors with complex shapes or located in poorly-contrasted regions (12–15). To compensate for this variability, regulatory authorities recommend an ICR with several readers to mitigate potential bias resulting from variance among investigator sites (2,5–8).

The software solution used in this study was Lesion Management Solutions (LMS) developed by MEDIAN Technologies (16,17). LMS is an image analysis software application for evaluating CT images. It allows lesion identification, quantification, and comparison of successive CT scans of the same patient. The comparison is achieved by synchronous navigation between two scans and automatic pairing of the lesions. Using these tools, we made possible a cloud-based local-read paradigm for imaging evaluations which consists of implementing a cloud software solution and a quality control at investigator sites. Readers working at distant locations were able to reliably perform radiological evaluations from the same cloud system. The objective

is to standardize the review of images at investigator sites in the frame of a clinical study. The purpose of this study was to investigate usefulness of a cloud implementation of this system in terms of evaluation of tumor response of lung cancers according to the RECIST criteria and the inter-observer agreement among sites.

Material and Methods

Study design and patient inclusion

This was a retrospective study of CT images of lung cancer patients at Saga University Hospital (Japan), Nice University Hospital (France), and other facilities. KUMO I, the first part of the study, was intended to demonstrate the feasibility of cloud implementation and to suggest technical improvements, as well as to identify issues regarding variability of evaluations among sites. The workflow of the KUMO I study is shown in Fig. 1. The Japanese investigator sites, the French independent reviewer site, the data manager (MEDIAN Technologies, Valbonne, France, www.medianttechnologies.com), and the data center (Canon IT Solutions, Tokyo, Japan, www.canon-its.co.jp) are all connected to the Canon cloud infrastructure service SOLTAGE (Canon IT Solutions, Tokyo, Japan) through a virtual private network (VPN). SOLTAGE provides storage, computing, and application services. The investigator sites provide the medical images and perform the analysis and interpretation of the medical images on Web-based MEDIAN Technologies Imaging Service. The analysis and interpretation results are centrally stored and available to the independent reviewer to perform its analysis and interpretation. This workflow is under the supervision of the data manager and data center that also have access to central information and progress status of the study.

CT images of 10 lung cancer patients were acquired at three time-points (baseline, best response, and progression) in the course of treatment. Patients who were treated with EGFR tyrosine kinase inhibitors (EGFR-TKI) were randomly selected, since obvious changes in image were observed when EGFR-TKI was administered to lung cancer patients with *EGFR* activating mutations. A well-trained radiologist was selected as a reviewer, and two medical oncologists who were experienced as specialists of more than 10 years were selected as investigators. CT scans were evaluated according to the RECIST 1.1 criteria by two oncologists from Saga University and one radiologist from Nice University Hospital, independently, through the cloud-based software. The software was hosted by the data center (Canon IT Solutions, Tokyo, Japan). Readers and data managers (Canon Inc. and MEDIAN Technologies)

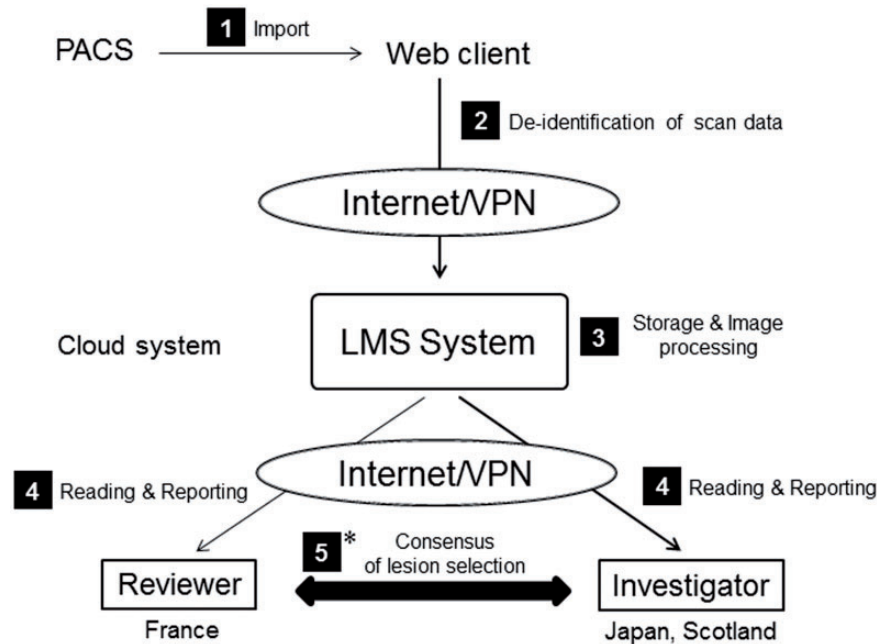


Fig. 1. KUMO I and KUMO I Extension Study work flow. Scan data were imported into the web client and anonymized by the data center, Canon IT Solutions, Japan. The data center, Canon, and the data managers (MEDIAN) stored and processed the images, and the image database was sent to each reader, in Japan and France. The study compared evaluations between readers and analyzed the reasons for discordances. Readers with different medical training and education, working at distant locations were able to reliably perform radiological evaluations from the same cloud system. The cloud quality control service detected non-conformance in applying RECIST 1.1 and had the readers change their evaluations, resolving the discrepancies. Based on KUMO I, KUMO I Extension was designed to improve the system. KUMO I Extension included additional quality controls to arrange the consensus of lesion selection (*). LMS, lesion management solutions; PACS, picture archiving and communication system, VPN; virtual private network.

were responsible for de-identification, quality control, and centralization of the images and evaluations. The study compared evaluations between the oncologists (investigators) and the radiologist (reviewer) and analyzed the reasons for discordance. The second part of the study, KUMO I Extension (Fig. 1), aimed to implement and evaluate solutions to solve issues identified by the KUMO I study and is described in the Results part of this paper. The extension study was also performed using CT scans from three timepoints, from 11 lung cancer patients. The evaluations of tumor response were performed by two oncologists (Japan and Scotland) as investigators and one radiologist (France) as a reviewer, independently. The study protocol was approved by the Clinical Research Ethics Committees of Saga University and Nice University Hospital.

Imaging technique

All images were taken with multi-detector CT scanners (LightSpeed VCT[®], GE Healthcare Japan, Tokyo, Japan; SOMATOM Definition[®], SIEMENS, Munich, Germany) at Saga University Hospital, or selected in MEDIAN images database. Slice thickness of the scans was 5 mm for KUMO I and 1–2.5 mm for KUMO I

Extension. Tube voltage was 120 kV, and field of view (FOV) was 300–500 mm. All images were properly anonymized and copied to a virtual server at the data center operated by Canon IT Solutions, Inc. The images were processed by a cloud-based prototype of Lesion Management Solutions (LMS) (MEDIAN Technologies, Valbonne, France). LMS is at the core of MEDIAN's Clinical Trial Imaging Services, which include image and workflow management, and image processing specifically designed for multi-site oncology clinical trials (Fig. 2). The image processing component of LMS offers software for detection, segmentation, and quantification of thoracic lesions (Fig. 2a). The segmentation process, which is based on a three-dimensional (3D) region-growing algorithm, begins with a simple point-and-click on the lesion of interest. Readers can make manual adjustments to the contour of the lesion as necessary. After segmentation is complete, the longest axial diameter, short axis, and volume of each lesion are extracted automatically. In the follow-up evaluation, two scans from two time-points are displayed side by side while automatic registration points to the volume of interest in the newer scan (Fig. 2b). A reader points to the corresponding lesions in the newer scan, which are then analyzed in the same manner as at the baseline

evaluation. Change in size and volume between time-points are then calculated and reported (Fig. 2c). LMS graphically displays the evolution of the tumor burden based on both diameter and volume of lesions. Finally all of the review data are used to compute the response evaluation and to categorize the response as complete response, partial response, stable disease, or progressive disease.

Statistical analysis

Kappa analysis was performed to evaluate inter-observer agreements between the reviewer and each investigator using IBM SPSS Statistics 19 (SPSS, Inc., IBM Company, Tokyo, Japan). The strength of agreement indicated with kappa values has been reported as: <0, poor; 0–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–1.00, almost perfect (18).

Results

The main reasons for inter-reader discordance were differences in the selection of target lesion at baseline and in lesion segmentation. In the KUMO I study, 10 evaluations were performed by each investigator. Therefore, the concordance between the reviewer and the two investigators include a total number of 20 evaluations.

The results showed discordance rates of 40% (8/20 evaluations) for the selection of target lesions at baseline, 70% (14/20) for response evaluation 1 (at the first time-point), and 60% (12/20) for response evaluation 2 (at the second time-point) (Table 1). The Cohen's kappa coefficient value of inter-observer agreement for response evaluations between a reviewer and investigators was 0.451 ± 0.22 (95% CI). The discordance in the RECIST overall responses was caused by differences in the selection of target lesions, differences in lesion segmentation, and by overlooked new lesions.

Table 1. Inter-reader agreement between two investigators and a reviewer (KUMO I Study).

	Agree	Reasons for disagreement
Target selection at baseline	40% (8/20)	
Response evaluation	65% (26/40)	Different target lesions, segmentation, new lesions
TP1	70% (14/20)	
TP2	60% (12/20)	

TP, time-point.

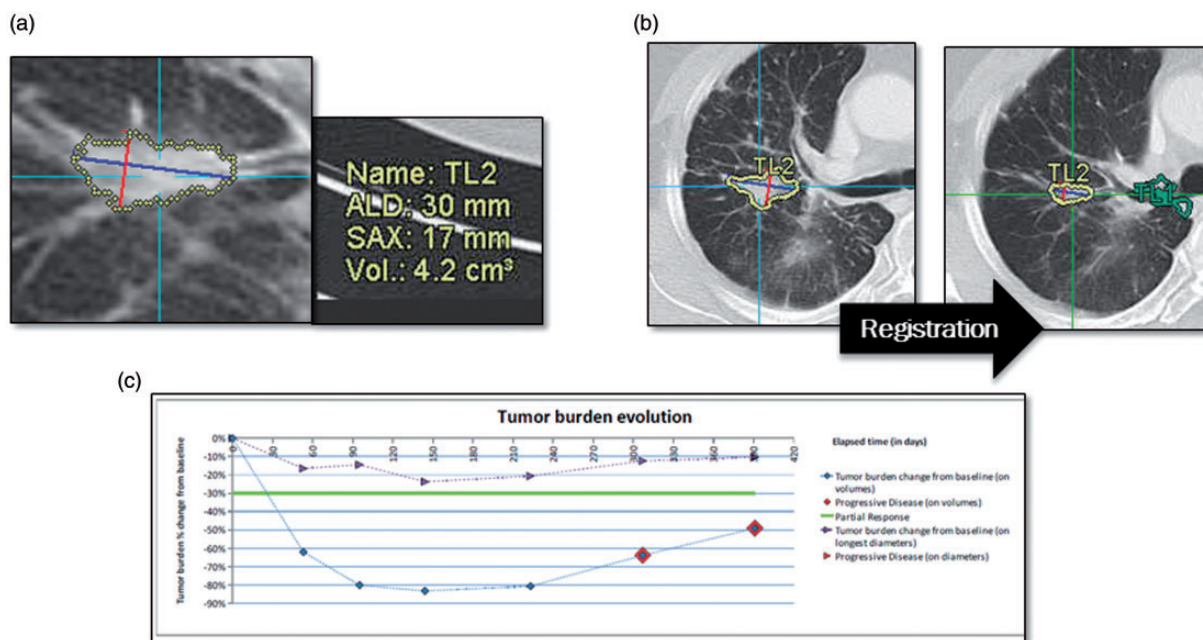


Fig. 2. The LMS system consists of three steps: (a) auto-segmentation and quantification; (b) follow-up segmentation; and (c) response evaluation. Just one click on a lesion leads to automatic segmentation, and quantification of longest diameter, short axis, and tumor volume. To compare images from two time-points, the system automatically registers the images to match the position of the lesion. Evaluation is also automatic, based on RECIST criteria and volumetry analysis.

Improvement of inter-reader agreement was observed after reaching consensus between a reviewer and investigator. Based on the results of KUMO I study, KUMO I Extension added a cloud-based quality control service to achieve a consensus on the selection of target lesions (Fig. 1). The KUMO I Extension, raised the agreement rate of response evaluations to 82% from 65% in KUMO I (Table 2). Kappa coefficient value of inter-observer agreement for response evaluations between the reviewer and the investigators was 0.724 ± 0.17 (95% CI). The cloud software solution gives the possibility for a given clinical study that all investigator sites work with the same tools, on the same database. Thanks to the data centralization provided by this cloud configuration, an ongoing quality control is made possible. A quality control performed by a clinician regarding the choice of target lesions can be implemented: in case of disagreement between the reviewer and the investigator site, the site has to explain and/or revise his choice. In spite of this system, target selection at baseline did not completely agree even after adjustment because of clinically justifiable differences between the reviewer and the investigators as shown in Fig. 3. The investigators selected a primary lesion as the target lesion even

though segmentation was difficult because of the complicated shape of the lesion and its location adjacent to the pleura (Fig. 3a). The reviewer selected a metastatic lesion, however, because evaluation of the lesion is easily reproducible (Fig. 3b). Differences in lesion segmentation between the reviewer and the investigators are shown in Table 3. Among total lesions which were evaluated in all images in KUMO I Extension Study, the frequency of indicated differences in axial diameter between a reviewer and each investigator was investigated. The lesions adjacent to the mediastinum and pleura showed bigger differences in axial diameter (Fig. 4) compared to those in the lung field. Lesions in the lung field had enough contrast on imaging so they could be segmented by automated tools, while other lesions, including lesions in the lymph nodes, could not.

Discussion

Building on the results of KUMO I, modification by the cloud quality control service could lead to improved concordance between readers in the KUMO I Extension study. Ongoing monitoring of evaluations through

Table 2. Inter-reader agreement between two investigators and a reviewer (KUMO I Extension Study).

	Agree	Reasons for disagreement
Target selection at baseline	82% (18/22)	Clinically justifiable difference
Response evaluation	82% (36/44)	Segmentation, new lesions
TP1	86% (19/22)	
TP2	77% (17/22)	

TP, time-point.

Table 3. Difference in lesion segmentation between two investigators and a reviewer (KUMO I Extension Study).

Location of lesion	Difference in axial diameter		
	≤2 mm	>2 mm, ≤4 mm	>4 mm
Lung field ($n = 42$)	83%	7%	10%
Mediastinum ($n = 18$)	44%	22%	33%
Pleural ($n = 33$)	52%	12%	36%
Lymph node ($n = 18$)	78%	11%	11%

n , total lesions which were evaluated in all images used in the KUMO I Extension Study.

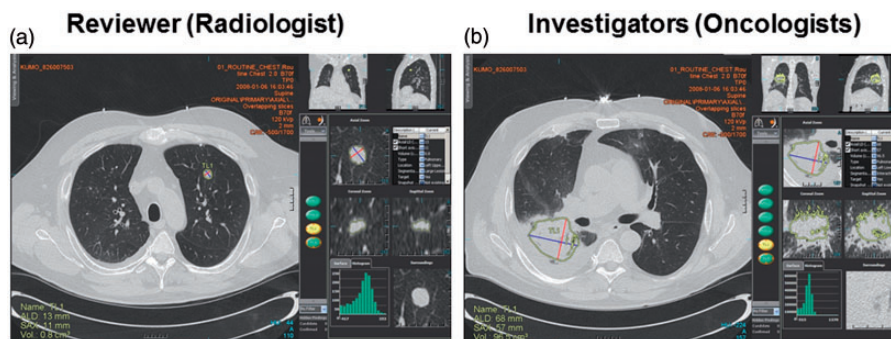


Fig. 3. Discordance of lesion selection between a reviewer (a) and investigators (b). The reviewer selected a metastatic lung lesion, and the investigator selected a primary lesion adjacent to the pleura.

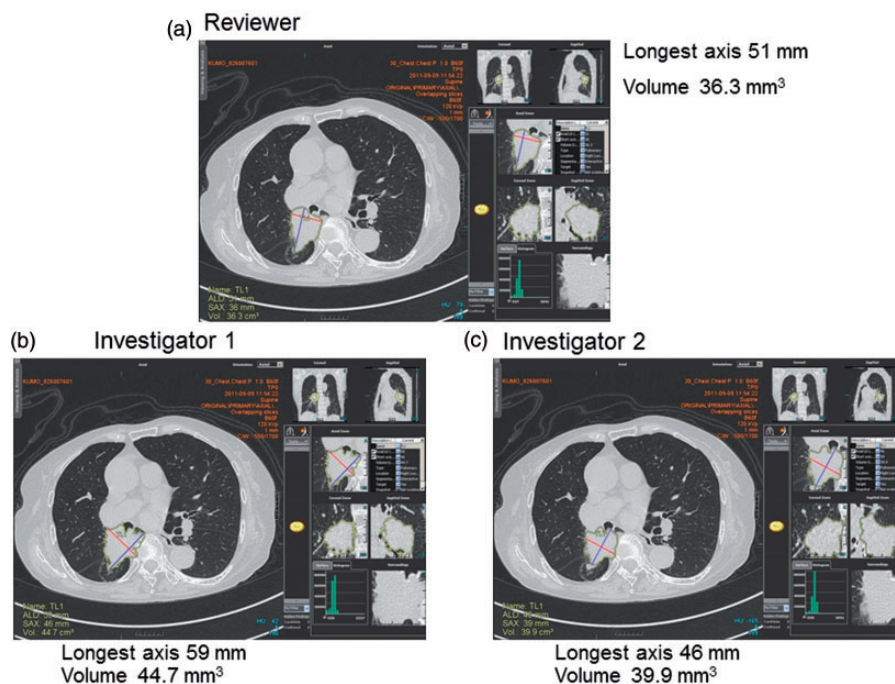


Fig. 4. Comparison of segmentations performed manually between a reviewer (a) and two investigators (b, c).

specialized services to reduce variability among sites was made possible by centralized data management.

As for cost-effectiveness, the cost required to implement solutions depends on the number of sites in usual central review systems. However, a cloud-based solution developed in this study contains centralized data processing system, and readers just need a minimum set of computers connecting to the Internet. Therefore, the total cost to set up such a cloud-based service is much lower than using locally installed software from both direct and indirect perspectives including maintenance. To secure the participants' privacy, all necessary de-identification processes are done before putting images to the LMS system. Personal information is not shared among readers. We also chose VPN connection and a highly secure data center.

Some limitations of the study are the small sample size and the lack of evaluation of lymph nodes. These studies pave the way for further investigations such as the improvement of the automated segmentation tools to better address the lesions adjacent to the pleura and mediastinum as well as the lymph nodes, since discordance in response evaluation occurred mainly when segmentation was performed manually. In addition, it would also be interesting to take advantage of this implementation to investigate the limitations of the RECIST criteria: volumetric measurement of tumor size has been reported to be reproducible and accurate compared to 1D or 2D measurements (19). LMS provides automatic evaluation of tumor volume even when the tumor contour is complicated, as with cavating lesions. With such

modification of the system, a prospective clinical study involving several hospitals around the world should be performed to confirm its feasibility.

In conclusion, the cloud-based local-read automatic imaging analysis system could become an integral component of global clinical trials for solid tumors after some modifications.

Declaration of conflicting interests

Colette Charbonnier is an employee of MEDIAN Technologies, France. Junta Yamamichi and Hideaki Mizobe are employees of Global Healthcare IT Project, Canon Inc., Japan.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

1. Johnson JR, Williams G, Pazdur R. End points and United States Food and Drug Administration approval of oncology drugs. *J Clin Oncol* 2003;21:1404–1411.
2. Dancey JE, Dodd LE, Ford R, et al. Recommendations for the assessment of progression in randomised cancer treatment trials. *Eur J Cancer* 2009;45:281–289.
3. Sherrill B, Kaye JA, Sandin R, et al. Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. *Onco Targets Ther* 2012;5:287–296.
4. Laporte S, Squifflet P, Baroux N, et al. Prediction of survival benefits from progression-free survival benefits in advanced non-small-cell lung cancer: evidence from a

- meta-analysis of 2334 patients from 5 randomised trials. *BMJ Open* 2013;3:e001802.
5. Amit O, Mannino F, Stone AM, et al. Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis. *Eur J Cancer* 2011;47:1772–1778.
 6. Tang PA, Pond GR, Chen EX. Influence of an independent review committee on assessment of response rate and progression-free survival in phase III clinical trials. *Ann Oncol* 2010;21:19–26.
 7. Ford R, Schwartz L, Dancey J, et al. Lessons learned from independent central review. *Eur J Cancer* 2009;45:268–274.
 8. Dodd LE, Korn EL, Freidlin B, et al. Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? *J Clin Oncol* 2008;6:3791–3796.
 9. James K, Eisenhauer E, Christian M, et al. Measuring response in solid tumors: unidimensional versus bidimensional measurement. *J Natl Cancer Inst* 1999;91:523–528.
 10. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst* 2000;92:205–216.
 11. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–247.
 12. Suzuki C, Torkzad MR, Jacobsson H, et al. Interobserver and intraobserver variability in the response evaluation of cancer therapy according to RECIST and WHO-criteria. *Acta Oncol* 2010;49:509–514.
 13. Sohns C, Mangelsdorf J, Sossalla S, et al. Measurement of response of pulmonary tumors in 64-slice MDCT. *Acta Radiol* 2010;51:512–521.
 14. 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;11:1355–1360.
 15. Park JO, Lee SI, Song SY, et al. Measuring response in solid tumors: comparison of RECIST and WHO response criteria. *Jpn J Clin Oncol* 2003;33:533–537.
 16. Beaumont H, Brasier-Voguet C, Butzbach A. Inter-reader agreement in response to therapy evaluation of advanced lung cancer: benefits of a volume-derived imaging biomarker. *Eur J Cancer* 2011;47(Suppl. 1): S138.
 17. Beaumont H, Oubel E, Iannesi A, et al. Reliability of imaging biomarkers for response assessment in advanced lung cancer: Influence of expertise and automation. *J Clin Oncol* 2012;30:e13547.
 18. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology* 2003;228:303–308.
 19. Sohaib SA, Turner B, Hanson JA, et al. CT assessment of tumour response to treatment: comparison of linear, cross-sectional and volumetric measures of tumour size. *Br J Radiol* 2000;73:1178–1184.