available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

# A simulation study to evaluate the impact of the number of lesions measured on response assessment

*Chaya S. Moskowitz*[a,*], *Xiaoyu Jia*[a], *Lawrence H. Schwartz*[b], *Mithat Gönen*[a]

[a]*Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, 307 East 63rd Street, 3rd Floor, NY 10065, United States*

[b]*Department of Radiology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, NY 10065, United States*

ARTICLE INFO

*Article history:*
Received 4 November 2008
Accepted 7 November 2008
Available online 16 December 2008

*Keywords:*
Cancer imaging
Ct imaging
Inter-reader variability
Random effects model
RECIST criteria
Response to therapy

ABSTRACT

The objectives of this study were to evaluate whether the number of lesions that are used to measure tumour burden affects response assessment and inter-rater variability. In order to accomplish this, a simulation study was conducted. Data were generated from a mixed-effects mixture model. Parameter values to input in the model were obtained from the analysis of real data. Response assessments based on 10, five, three, two and one lesion were evaluated. There was little difference between response assessments based on five lesions and response assessments based on 10 lesions. When fewer than five lesions were used to assess response, there were notable differences from the 10 lesion-based response assessment. Basing response assessment on a small number of lesions tends to overestimate response rates and leads to misclassification of patients' response status. Therefore, measuring five lesions per patient appears to sufficiently capture patients' response to therapy. Measuring fewer than five lesions results in the loss of information that may adversely affect clinical trial results as well as patient management.



## 1. Introduction

Response to chemotherapy is an essential part of patient care and clinical research. Responding patients are often offered prolonged treatment and non-responders are quickly switched to another treatment regimen. Phase II clinical trials using response as the primary end-point are ubiquitous and often are the primary determinants of whether a regimen should be taken to a definitive Phase III study. Hence, accurate determination of response to chemotherapy is of critical importance.

Patients who receive treatment for cancer, whether as participants in a clinical trial or simply in the course of standard therapy, usually have multiple sites of metastases, in multiple organs. It is possible that the effect of the treatment will not be identical at all sites of metastases. For example, the treatment may shrink all the lesions but by varying degrees (Fig. 1). In some instances, it is even possible for certain lesions to shrink in response to treatment whilst others grow. When assessing response to therapy, such as with RECIST guidelines,[1] it might therefore seem necessary to measure all lesions in order to best evaluate completely whether a patient is responding to a therapy. In fact, there is some empirical evidence in the literature that the variability in tumour response measurements is substantially reduced, as increasing numbers of lesions are measured.[2] Often, however, resources do not permit radiologists to evaluate every lesion, and instead a subset or selection of lesions is chosen. The original WHO criteria recommended that five lesions be measured.[3,4] In the RECIST 1.0 guidelines, recommendations were for measuring all lesions up to a total of 10. In patients with more than 10 lesions, the choice of which lesions to measure

* *Corresponding author:* Tel.: +1 (646) 735 8117; fax: +1 (646) 735 0010.
E-mail address: moskowc1@MSKCC.ORG (C.S. Moskowitz).
0959-8049/$ - see front matter 
doi:10.1016/j.ejca.2008.11.010

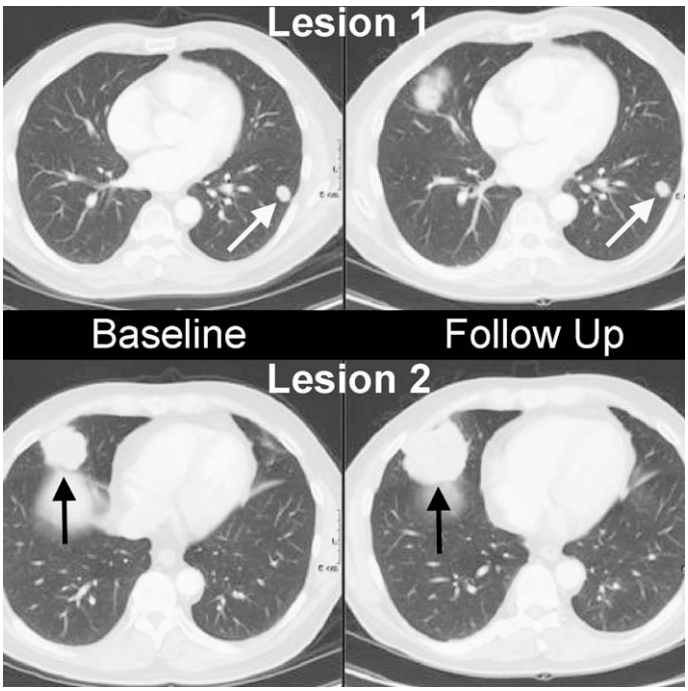


**Fig. 1 – Metastatic disease to the lungs. Note that the smaller lesion (white arrows) has not changed in size from baseline to follow-up, whilst the larger lesion (black arrows) has increased in size.**

should be based on the size of the lesion and how suitable it is for repeated measurements.

In practice, measuring up to an upper limit of 10 lesions may still be difficult and require more time and effort than many radiologists are routinely able to devote. A natural question to ask is whether fewer lesions can be measured whilst still sufficiently capturing a patient's response to therapy. If so, how many lesions must one measure?

A key difficulty in answering this question is that the truth is rarely, if ever, known. In order to know whether a radiologist's repeated assessments of tumour burden accurately reflects a patient's change in true tumour burden over the course of a therapy, after being imaged at each time point the patient would need to undergo surgery and have all of their lesions measured. Clearly this is not possible.

One potential way to address this issue is to compare the response assessment that would have been obtained had we measured fewer than 10 lesions with the response assessment obtained based on the complete 10 lesions. In some sense, this approach considers the response assessment based on the 10 lesions to be the gold standard. It must be acknowledged, however, that response assessment based on the 10 lesions is not necessarily 'the truth.' Unmeasured lesions beyond this upper limit may change the assessment. With this caveat in mind, comparing a 10 lesion-based response assessment with a response assessment based on fewer lesions would help answer the question of whether measuring less than 10 lesions would substantially alter the way tumour burden is currently evaluated under the RECIST 1.0 guidelines.

Another issue to be taken into consideration is that response assessment is radiologist-specific. That is, each radiologist selects what he or she perceives to be the 10 largest lesions and then measures these lesions to the best of their ability. Inter-rater variability in this setting, however, is not inconsequential and whether a patient is determined to have responded to treatment may in fact differ between radiologists.[5,6] We might further question, then, whether radiologists are more likely to agree in their response assessments if they measure more lesions. It seems logical to be most comfortable with response assessments that have a high level of agreement between multiple radiologists.

In this journal, the paper entitled 'Individual patient data analysis to assess modifications to the RECIST criteria' evaluates the EORTC data warehouse[7] and assesses change in response by decreasing the number of lesions. There was concern that this database collated from both industrial trials and cooperative group trials may not be truly representative of total tumour burden and number of lesions. In fact the mean number of lesions in those cooperative group trials was approximately 40% lower than the industrial independently reviewed trials. Therefore, part of the rationale of this simulation study is to more precisely approximate total tumour burden.

There are several advantages to conducting a simulation study including the ability to change the parameter settings that are used to simulate the data and the ability to explore the results in a variety of scenarios. For these reasons, we undertook this simulation study.

## 2. Methods

The primary aim of this simulation study was to evaluate whether the number of lesions measured affects response assessment. Secondarily, we were also interested in exploring whether the number of lesions measured affects inter-rater variability.