EUROPEAN JOURNAL OF CANCER 45 (2009) 248–260

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Individual patient data analysis to assess modifications to the RECIST criteria

Jan Bogaerts[a,], Robert Ford[b], Dan Sargent[c], Lawrence H. Schwartz[d], Larry Rubinstein[e], Denis Lacombe[a], Elizabeth Eisenhauer[f], Jaap Verweij[g], Patrick Therasse[h], for the RECIST Working Party*

[a]*European Organization for Research and Treatment of Cancer (EORTC) Headquarters, Avenue Mounierlaan, 83/11, Brussel 1200 Bruxelles, Belgium*
[b]*RadPharm, Princeton, USA*
[c]*Mayo Clinic, Rochester, USA*
[d]*MSKCC, New York, USA*
[e]*National Cancer Institute, Bethesda, USA*
[f]*National Cancer Institute of Canada – Clinical Trials Group, Kingston, ON, Canada*
[g]*Erasmus University Medical Center, Rotterdam, The Netherlands*
[h]*GlaxoSmithKline Biologicals, Rixensart, Belgium*

ARTICLE INFO

Article history:
Received 17 October 2008
Accepted 29 October 2008
Available online 16 December 2008

Keywords:
Tumour measurements
RECIST
Response
Progression

ABSTRACT

Background: After the initial RECIST 1.0 were published in 2000, the criteria were widely implemented in the scientific oncology community. Since then, the RECIST working group has identified several issues to examine further. Two key issues that required careful, data-based assessment were the maximum number of lesions that should be assessed at each evaluation and the added value of requiring confirmation of response.

Methods: To address these questions, data were obtained from 16 clinical trials in metastatic cancer, with patients enrolled between 1993 and 2005. A total of 6512 patients were included in the primary analysis dataset, accounting for over 18,000 potential target lesions. Nine percent of the included patients ($n = 585$) had six or more reported target lesions. The response and progression outcomes in the database were calculated using an adjusted RECIST methodology with a maximum of 5 (or 3) target lesions with/without confirmation and this was compared to the original RECIST version 1.0 which required up to 10 target lesions plus confirmation of response.

Results: Assessment of 5 lesions per patient led to a difference in best overall response assignment for an estimated 209 (3.2%) patients as compared to RECIST version 1.0. However, these changes did not affect the overall response rate. Progression-free survival was only minimally affected by measuring fewer lesions. In contrast, removing the requirement for response confirmation led to a significant increase in the numbers of patients classified as responders, resulting in a relative increase of approximately 19% in response rate. An algorithm using a maximum of three target lesions shows high concordance with the 10 lesions requirement in terms of response and TTP assignment. Concern that appropriate assessment of disease within an organ requires two lesions to be followed per organ

* *Corresponding author:* Tel.: +32 2 774 10 63; fax: +32 2 771 38 10.
E-mail address: jan.bogaerts@eortc.be (J. Bogaerts).
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2008.10.027

suggests the approach of following two target lesions per organ, up to a maximum of five target lesions overall. Both strategies seem reasonable based on the data warehouse. The requirement of response confirmation in trials where this is a primary end-point is recommended to be maintained as its removal would substantially increase reported response rates.

1. Introduction

The initial RECIST were published in 2000,[1] and the criteria have been widely implemented in the scientific oncology community. The RECIST working group has monitored the implementation of the RECIST criteria since its establishment to capture potential issues and to propose the adaptation of the methodology wherever feasible and appropriate. To gain confidence in recommending changes to RECIST, detailed measurement data on individual patients from multiple completed trials have been pooled into a data warehouse. This warehouse has allowed several analyses to explore possible refinements to RECIST. These analyses have enabled the RECIST working group to gain insights into the current use of the criteria, and to adjust the recommendations by exploring possible impact on clinical trial outcome based on alternative criteria. Specifically, the following key issues were examined: (1) What is the maximum number of lesions that should be assessed at each evaluation? and (2) What is the impact of requiring confirmation of response? This paper specifically reports on the methodology that was used, the results of the analyses, and discusses their limitations.

2. Material and methods

2.1. The data warehouse

We acquired individual patient tumour measurement data from large published trials in metastatic solid tumours from several sources (Table 2). The data were consolidated into a Data Warehouse at the EORTC Data Centre. Most, but not all of the trials in the warehouse had undergone independent radiology review. Data collected included general patient information (e.g. the start of treatment, death or last known alive date) and detailed tumour measurement data over time: measurement/evaluation date, organ of the lesion, method of measurement, longest diameter for measured lesions, information on non-target lesions and occurrence of new lesions. For analysis, a primary dataset was identified using the following criteria: at least one lesion measured at baseline, at least one disease assessment after the start of treatment, and if recorded, best response was not 'early death' or 'not assessable'. Data checks ensured that potential target lesions were followed by the same method as used at baseline, and verified that patients had a sufficient duration of (measured) follow-up.

To ensure a uniform application of standard criteria, for each patient, the best response using the established RECIST for target lesion selection, measurement and confirmation of response was calculated on the basis of the available tumour size measurements, assessments of non-measurable lesions and occurrence of new lesions, ignoring the original criteria used by each study protocol. To classify as a stable disease, a minimum six-week period was required between the baseline measurement and the (first) follow-up measurement. A sensitivity analysis was performed using a minimum stable disease duration of 3 months.

For all the analyses, additional adjustments were incorporated in RECIST version 1.1, which were not present in the original RECIST version 1.0, for lymph node assessment, progressive disease classification and for the requirement of response confirmation. The changes applied in RECIST version 1.1 are briefly summarised in Table 1.

2.2. Lymph node assessment

For malignant lymph nodes only the short axes of the lymph nodes were used in all the analyses, if available. For a few trials, only the long axis was available. In addition, only lymph nodes with a baseline short axis of at least 15 mm were considered potential target lesions, and if all lymph node short axes dropped below 10 mm (indicating that the nodes are no longer pathologic by size criteria), and all other target lesions had 0 mm measurements, the patient was classified as CR (for target lesions). To assess progressive disease and partial response, the actual short axis measurements were maintained in the sum of diameters.

2.3. Progressive disease classification

We assessed progressive disease by the growth of target lesions, using a minimum absolute increase of 5 mm from the nadir size in the sum of longest diameters in addition to the 20% relative increase in the sum of LD compared to the nadir.

2.4. Confirmation of response

To evaluate the impact of removing the requirement for response confirmation, the above analyses were conducted both with and without this requirement. Importantly, not all trials included in the warehouse had required response confirmation. Therefore, a further analysis was performed restricted to those patients obtaining a PR or CR at one time point, and for whom a further complete evaluation or a progression classification was available within 8 weeks from that time point. The latter analysis was intended to provide a less biased estimate of the percentage of additional responses that would be declared where the requirement of response confirmation to be removed.

For the evaluation of methods where a reduced number of target lesions are required to be followed, it is assumed that the lesions that were no longer used as target lesions in the

Table 1 – Overview of RECIST 1.0 and RECIST 1.1 variations analysed.

RECIST 1.0	RECIST 1.1	Variations studied
Up to 10 target lesions: longest diameter followed	Up to 'x' target lesions, longest diameter followed; lymph nodes measured short axis in nodes (⩾15 mm eligible for inclusion as target)	'x' was 1, 2, 3, 5 in a series of iterations Largest lesions selected with maximum representation of organs 2 lesions per organ, when total allowed Variation: random choice of *x* lesions
Complete response sum = 0	Same except if lymph nodes included, CR possible when nodes < 10 mm each	
Partial response 30% decrease in LD, confirmed	Same except short axis LN summed with longest diameters other target lesions	Variation: unconfirmed
PD 20% increase in sum over lowest on study; or new lesions	PD same but in addition requires minimum 5 mm absolute increase from lowest on study	

calculation would not have had measurements performed. Of note, to classify as CR, such lesions also must have decreased to 0 mm diameter measurement (except for lymph nodes, see above). Similarly, for the purpose of these analyses, the calculations did not allow such 'deselected' lesions to result in progression because of unequivocal progression. A sensitivity analysis was performed whereby a 'deselected' lesion of at least 2 cm (as measured in the original data) would cause 'unequivocal progression' in the reduced method (being considered as a non-target) if it doubled in LD.

2.5. *Number of target lesions to be followed to assess response*

The first analysis considered the requirement to follow all measurable lesions up to a maximum of 10 lesions, and whether it may be possible to relax this requirement. For this analysis, the calculated RECIST outcome with up to 10 targets was contrasted with the outcomes that would have been obtained using only up to 1, 2, 3 or 5 targets.

The presented tables focus on two of the many options that were explored, using a maximum of either 3 or 5 target lesions per patient. Both methods were selected by lesion size, but maximised the number of represented organs with a maximum of 2 target lesions per organ. The other details as described in the main article on RECIST version 1.1, and as described above regarding lymph node assessment and progression definition were applied as well. These additional criteria were applied only to the investigation of a maximum selection of 1, 2, 3 or 5 target lesions, as opposed to the 10, which utilised RECIST version 1.0 as originally published. In other words, the outcome of the 'original' RECIST (with up to 10 lesions) was compared to the performance of several alternative models that varied in the number of lesions selected, and also employed new methods for lymph node measurement and PD definition requirements. The greater the number of lesions reported at baseline per patient, the larger the potential outcome differences could be comparing the assessment of 1, 2, 3 or 5 lesions versus 10 lesions. For this reason, tabulations are provided both for the complete dataset and for the subset of patients in whom 6 or more potential target lesions were reported.

In order to further evaluate the importance of the choice of particular lesions within organs, sensitivity analyses were conducted where the rule of choosing the largest (up to two) lesions within an organ was relaxed. In these analyses, first organs to be considered were selected at random, and once an organ has been selected, the lesions within that organ were also selected at random.

2.6. *Short-term progression free survival (PFS)*

In addition to examining the response rate, we sought to investigate the time to event outcome of progression-free survival with RECIST version 1.0 compared with the proposed revised criteria. Progression-free survival was analysed considering events which are either progression of disease (on the basis of the program for RECIST(n)), or death within 60 days of the last obtained measurements. Patients without such an event were censored at last measurement date. This approach was used in order to provide an estimate of the early part of a PFS Kaplan–Meier curve, and effectively dealt with the scenario of relatively long periods of follow-up without the measurement data, followed by a later death. The censoring of such cases (for the purpose of this analysis) at last documented measurement provided a possibility to evaluate 'RECIST(n) PFS', restricting the follow-up to the period with available measurements. The reason for progression in this analysis was split into the following categories (whichever occurred first):

- New lesion, with or without other qualifying events within 3 weeks.
- Progression of non-target lesions, with or without the progression of target lesions within 3 weeks.
- Progression of target lesions (in the absence of new lesions or progression in non-target lesions within 3 weeks).
- Death within 2 months of the last measurements, in the absence of any of the above.

The reasoning for this categorisation is to identify those cases that – using a different method – would have a different outcome (progression versus non-progression), or a different timing of progression. Indeed, only progressions that are declared based on the changes in lesion measurements, in the absence of other potential causes, would alter the patient's progression date if the RECIST criteria were modified.

3. Results

Data were collected from 16 clinical trials in metastatic cancer, with patients enrolled between 1993 and 2005 (Table 2,[2–15]). In some trials patients with locally advanced disease were also eligible, but they represented a small minority. Of a total of 9147 patients for whom data were collected, 6512 could be included in the primary dataset. Reasons for exclusions were lack of recorded measurable lesions (1491, 16%), lack of valid measurement data post baseline (613, 7%), lack of any valid outcome when applying RECIST version 1.0 (mostly because not all target lesions were consistently followed, 326 cases, 4%) and a recorded best response of 'unevaluable' 205 (2%).

In the primary dataset, over 18,000 potential target lesions were available. Nine percent of patients (n = 585, Table 3) had six or more reported target lesions at baseline. 73.8% of patients had 1, 2 or 3 target lesions reported. The longest diameter at baseline was 2 cm or less in 30% of lesions, 25% were 2.1–3 cm, 25% were 3.1–5 cm and 20% were more than 5 cm in longest diameter. A large majority (88%) of potential target lesions were assessed by CT scan (Table 4). In the primary dataset, only 4% of patients had (some) lesions where the method of measurement varied over the course of follow-up. Median follow-up (including all target lesions and information on non-target lesions as applicable) was 8.0 months. The median number of visits with follow-up measurements post baseline was 4.

Table 2 – Clinical trials included in the data warehouse.

Trial	Organisation/ Company	Tumour type	Short description	N
Study of the Erasmus University Medical Centre, Rotterdam, The Netherlands	Erasmus University Medical Centre, Rotterdam, Netherlands	Breast	Phase II to III study comparing doxorubicin and docetaxel with fluorouracil, doxorubicin, and cyclophosphamide as first-line chemotherapy in patients with metastatic breast cancer	216
Intact 1	AstraZeneca	NSCLC	Gefitinib in Combination With Gemcitabine and Cisplatin in advanced non-small-cell lung cancer: a phase III trial	1093
Intact 2	AstraZeneca	NSCLC	Gefitinib in combination with paclitaxel and carboplatin in advanced non–small-cell lung cancer: a phase III trial	1037
BMS 139278	Bristol-Myers Squibb	Breast	Doxorubicin and paclitaxel versus fluorouracil, doxorubicin and cyclophosphamide as first-line therapy for women with metastatic breast cancer	267
Aventis study 303	Sanofi Aventis	Breast	Prospective randomised trial of docetaxel versus doxorubicin in patients with metastatic breast cancer	326
BCIRG study	BCIRG/Sanofi Aventis	Breast	Docetaxel and doxorubicin compared with doxorubicin and cyclophosphamide as first-line chemotherapy for metastatic breast cancer	429
Aventis study 307	Sanofi Aventis	Breast	Phase III randomised trial comparing docetaxel (T), doxorubicin (A) and cyclophosphamide (C) to FAC as first line chemotherapy (CT) for patients with metastatic breast cancer	484
Pfizer	Pfizer/RadPharm	GIST	A phase III, randomised, double blind, placebo-controlled study of SU011248 in the treatment of patients with Imatinib mesylate (Gleevec, Glivec)-resistant or intolerant malignant gastrointestinal stromal tumour	284
Pfizer	Pfizer/RadPharm	Renal Cell	A pivotal study of SU011248 in the treatment of patients with cytokine-refractory metastatic renal cell carcinoma	106
Pfizer	Pfizer/RadPharm	Renal Cell	A phase III, randomised study of SU011248 versus interferon (IFN)-a as first-line systemic therapy for patients with metastatic renal call carcinoma	695
Genentech study 2107	Genentech	Colorectal	Bevacizumab in combination with fluorouracil and leucovorin for first-line metastatic colorectal cancer	923
Tribute	Genentech	NSCLC	A phase III trial of Erlotinib hydrochloride (OSI-774) combined with carboplatin and paclitaxel chemotherapy in advanced non-small-cell lung cancer	1079
EORTC 10923	EORTC	Breast	Paclitaxel versus doxorubicin as first-line single-agent chemotherapy for metastatic breast cancer	331
EORTC 10961	EORTC	Breast	Doxorubicin and paclitaxel versus doxorubicin and cyclophosphamide as first-line chemotherapy in metastatic breast cancer	275
Roche BO16411	Roche	Lung	Phase III study of erlotinib in combination with cisplatin and gemcitabine in advanced non-small-cell lung cancer	1172
EORTC 40986	EORTC	Colorectal	Phase III study of weekly high-dose infusional fluorouracil plus folinic acid with or without irinotecan in patients with metastatic colorectal cancer	430

Table 3 – Number of target lesions per patient, primary dataset.

Number of target lesions	All patients (N = 6512) N (%)
1	1800 (27.6)
2	1784 (27.4)
3	1227 (18.8)
4	662 (10.2)
5	454 (7.0)
6	209 (3.2)
7	136 (2.1)
8	94 (1.4)
9	61 (0.9)
10	85 (1.3)

Table 4 – Method of measurement and site of target lesions.

	Potential target lesions (N = 18324) N (%)
Method	
Clinical examination	1552 (8.5)
X-ray	445 (2.4)
Conventional CT-scan	10263 (56.0)
Spiral CT-scan	5930 (32.4)
MRI	122 (0.7)
Other	11 (0.1)
Missing	1 (0.0)
Organ site	
Lymph node	3974 (21.7)
Lung	5687 (31.0)
Lliver	4430 (24.2)
Bone	156 (0.9)
Brain	6 (0.0)
Skin	235 (1.3)
Other soft tissue	583 (3.2)
Other site	3252 (17.7)
Missing	1 (0.0)

3.1. *Impact on response outcome*

Tables 5 and 6 show the differences in response outcome in the primary dataset using a maximum of 10 target lesions, as compared to a maximum of either 3 or 5 reported lesions. When the requirement for response confirmation is retained, in comparison to RECIST version 1.0 (original), RECIST version 1.1[16] using 5 *lesions* identified a different best overall response *assignment* in 209 (3.2%) patients and version 1.1 using 3 lesions identified a different best overall response *assignment* in 230 (3.5%) patients. In the subset of patients with more than 5 potential target lesions, there are 33 such cases using 5 lesions (5.6%) and 42 such cases using 3 lesions (7.5%). Compared to RECIST 1.0, a small increase in the number of cases assigned a complete response was found, related to the ability to achieve complete response in the presence of small measured lymph nodes that are considered normal size using the modifications in RECIST 1.1 (21 cases). Furthermore, the total number of responders decreased by 10 cases. Despite these changes in individual patient assignment, there was no effect on the *overall (partial and complete summed) response rate* (presumably because there was a balance in patients shifting from stable disease to partial response categories and *vice versa* for the most part) when looking at RECIST 1.1 with 5 lesions or RECIST 1.1 with 3 lesions both in comparison to RECIST 1.0.

When calculating the response rate at the trial level, the difference in the 16 trial overall response rates compared to RECIST version 1.0 ranged from −1.3% to +1.8% with a median of −0.2% using a maximum of five target lesions, and from −1.4% to +1.3% with a median of 0.0% using three target lesions.

In contrast, also shown in Table 5, when restricting the number of target lesions to be followed to 2 or 1, the discrepancies increase (for example, when only 2 targets are selected, 335 cases of the 6512 (5.1%) have a differing best response assignment).

If the requirement of response confirmation is removed, the changes in outcome compared to version RECIST 1.0 are more striking (Table 6): 1113 (17.1%) patients would have a different best overall response. In the dataset of patients with more than 5 target lesions reported at baseline, there were 75 such cases (12.8%). There would also be an increase in the number of complete responses (by 151 CR cases, almost a doubling). In addition, the overall number of responding patients would increase from 2493 to 3386, representing an increase of 13.8% in the overall response rate. When only 3 target lesions are used, the impact of eliminating response confirmation on the overall response rate is similar.

The considerable increase in the number of responses (complete responses and overall response rates) that result from dropping the requirement for response confirmation is biased due to the fact that for many such cases no further measurement was available, thereby eliminating the possibility of confirmation. To obtain a less biased estimate, we restricted analysis to initial responding patients who had a follow-up evaluation or a progression within 8 weeks. In this dataset, 1858 of 2207 initial responders (84.2%) were confirmed. Conversely, relaxing the requirement to confirm response, an extra 349 patients with response would be added, representing an increase of 19% in the overall response rate.

3.2. *Impact on progression free survival (PFS)*

In the primary dataset, the effect of restricting the number of lesions is minimal on PFS, irrespective of the requirement for response confirmation, as can be seen from Table 7. In the primary dataset, the median and the PFS point estimates at 6 months were the same regardless of whether 10, 5 or 3 lesions were considered. The difference in number of events was 29 (0.4 %) cases: 78 (1.2 %) additional progressions in version 1.1, versus 49 (0.8 %) cases that would no longer be classified as having progression.

In the 585 patients with more than 5 target lesions at baseline, the median PFS remained identical regardless of whether 10, 5 or 3 lesions were considered, and the point estimate at 6 months changed by only 0.78%. There was no overall difference in the number of events: 9 additional progressions would be classified in version 1.1, while 9 cases would no longer be classified as having progression.

Table 5 – Calculated best response: RECIST version 1.0 versus RECIST version 1.1, both with the condition of response confirmation.

	RECIST version 1.0[a] Restricted to patients with >5 reported target lesions at baseline				
RECIST version 1.1[b]	CR (N = 4) (0.68%) N (%)	PR (N = 134) (22.9%) N (%)	SD (N = 272) (46.5%) N (%)	PD (N = 175) (29.9%) N (%)	Total (N = 585) N (%)
With a maximum of 5 target lesions					
CR	4 (100.0)	1 (0.7)	0 (0.0)	0 (0.0)	5 (0.9)
PR	0 (0.0)	117 (87.3)	14 (5.1)	0 (0.0)	131 (22.4)
SD	0 (0.0)	16 (11.9)	257 (94.5)	1 (0.6)	274 (46.8)
PD	0 (0.0)	0 (0.0)	1 (0.4)	174 (99.4)	175 (29.9)
With a maximum of 3 target lesions					
CR	4 (100.0)	1 (0.7)	0 (0.0)	0 (0.0)	5 (0.9)
PR	0 (0.0)	113 (84.3)	18 (6.6)	0 (0.0)	131 (22.4)
SD	0 (0.0)	20 (14.9)	252 (92.6)	1 (0.6)	273 (46.7)
PD	0 (0.0)	0 (0.0)	2 (0.7)	174 (99.4)	176 (30.1)
With a maximum of 2 target lesions					
CR	4 (100.0)	1 (0.7)	0 (0.0)	0 (0.0)	5 (0.9)
PR	0 (0.0)	107 (79.9)	25 (9.2)	0 (0.0)	132 (22.6)
SD	0 (0.0)	26 (19.4)	244 (89.7)	2 (1.1)	272 (46.5)
PD	0 (0.0)	0 (0.0)	3 (1.1)	173 (98.9)	176 (30.1)
With 1 target lesion					
CR	4 (100.0)	1 (0.7)	0 (0.0)	0 (0.0)	5 (0.9)
PR	0 (0.0)	98 (73.1)	37 (13.6)	0 (0.0)	135 (23.1)
SD	0 (0.0)	31 (23.1)	224 (82.4)	2 (1.1)	257 (43.9)
PD	0 (0.0)	4 (3.0)	11 (4.0)	173 (98.9)	188 (32.1)
	All patients				
	CR (N = 174) (2.6%) N (%)	PR (N = 2319) (35.1%) N (%)	SD (N = 2809) (43.1%) N (%)	PD (N = 1210) (18.6%) N (%)	Total (N = 6512) N (%)
With a maximum of 5 target lesions					
CR	174 (100.0)	21 (0.9)	0 (0.0)	0 (0.0)	195 (3.0)
PR	0 (0.0)	2214 (95.5)	72 (2.6)	2 (0.2)	2288 (35.1)
SD	0 (0.0)	83 (3.6)	2723 (96.9)	10 (0.8)	2816 (43.2)
PD	0 (0.0)	1 (0.0)	14 (0.5)	1192 (98.5)	1207 (18.5)
No result	0 (0.0)	0 (0.0)	0 (0.0)	6 (0.5)	6 (0.1)
With a maximum of 3 target lesions					
CR	174 (100.0)	21 (0.9)	0 (0.0)	0 (0.0)	195 (3.0)
PR	0 (0.0)	2202 (95.0)	78 (2.8)	2 (0.2)	2282 (35.0)
SD	0 (0.0)	93 (4.0)	2716 (96.7)	11 (0.9)	2820 (43.3)
PD	0 (0.0)	3 (0.1)	15 (0.5)	1190 (98.3)	1208 (18.6)
No result	0 (0.0)	0 (0.0)	0 (0.0)	7 (0.6)	7 (0.1)
With a maximum of 2 target lesions					
CR	174 (100.0)	22 (0.9)	0 (0.0)	0 (0.0)	196 (3.0)
PR	0 (0.0)	2152 (92.8)	121 (4.3)	2 (0.2)	2275 (34.9)
SD	0 (0.0)	141 (6.1)	2668 (95.0)	16 (1.3)	2825 (43.4)
PD	0 (0.0)	4 (0.2)	20 (0.7)	1183 (97.8)	1207 (18.5)
No result	0 (0.0)	0 (0.0)	0 (0.0)	9 (0.7)	9 (0.1)
With 1 target lesion					
CR	172 (98.9)	27 (1.2)	0 (0.0)	0 (0.0)	199 (3.1)
PR	0 (0.0)	2013 (86.8)	253 (9.0)	2 (0.2)	2268 (34.8)
SD	2 (1.1)	262 (11.3)	2501 (89.0)	33 (2.7)	2798 (43.0)
PD	0 (0.0)	17 (0.7)	55 (2.0)	1159 (95.8)	1231 (18.9)
No result	0 (0.0)	0 (0.0)	0 (0.0)	16 (1.3)	16 (0.2)

a Percentages on the top bar are row percentages, showing rates per RECIST version 1.0.
b RECIST version 1.1 includes all other changes described, such as the changes for lymph nodes and a minimum increase of 5 mm for PD.

Table 8 presents the distribution of events in short-term PFS by reason: the majority of progressions occur due to new lesions being detected, either with or without growth in target lesions that qualified as progression. Some patients, while having the same reason for progression, had this progression at a different time point (being somewhat later or

Table 6 – Calculated best response: RECIST version 1.0 (including condition of response confirmation) versus RECIST version 1.1 without the condition of response confirmation.

RECIST version 1.1[b]	RECIST version 1.0[a] Restricted to patients with > 5 reported target lesions at baseline				
	CR (*N* = 4) (0.68%) *N* (%)	PR (*N* = 134) (22.9%) *N* (%)	SD (*N* = 272) (46.5%) *N* (%)	PD (*N* = 175) (29.9%) *N* (%)	Total (*N* = 585) *N* (%)
Using a maximum of 5 target lesions					
CR	4 (100.0)	5 (3.7)	0 (0.0)	0 (0.0)	9 (1.5)
PR	0 (0.0)	123 (91.8)	53 (19.5)	9 (5.1)	185 (31.6)
SD	0 (0.0)	6 (4.5)	218 (80.1)	1 (0.6)	225 (38.5)
PD	0 (0.0)	0 (0.0)	1 (0.4)	165 (94.3)	166 (28.4)
Using a maximum of 3 target lesions					
CR	4 (100.0)	5 (3.7)	0 (0.0)	0 (0.0)	9 (1.5)
PR	0 (0.0)	120 (89.6)	57 (21.0)	8 (4.6)	185 (31.6)
SD	0 (0.0)	9 (6.7)	213 (78.3)	1 (0.6)	223 (38.1)
PD	0 (0.0)	0 (0.0)	2 (0.7)	166 (94.9)	168 (28.7)
	All patients				
	CR (*N* = 174) (2.6%) *N* (%)	PR (*N* = 2319) (35.6%) *N* (%)	SD (*N* = 2809) (43.1%) *N* (%)	PD (*N* = 1210) (18.6%) *N* (%)	Total (*N* = 6512) *N* (%)
With a maximum of 5 target lesions					
CR	174 (100.0)	120 (5.2)	21 (0.7)	10 (0.8)	325 (5.0)
PR	0 (0.0)	2161 (93.2)	784 (27.9)	116 (9.6)	3061 (47.0)
SD	0 (0.0)	38 (1.6)	1993 (71.0)	8 (0.7)	2039 (31.3)
PD	0 (0.0)	0 (0.0)	11 (0.4)	1071 (88.5)	1082 (16.6)
No result	0 (0.0)	0 (0.0)	0 (0.0)	5 (0.4)	5 (0.1)
With a maximum of 3 target lesions					
CR	174 (100.0)	120 (5.2)	21 (0.7)	10 (0.8)	325 (5.0)
PR	0 (0.0)	2157 (93.0)	788 (28.1)	113 (9.3)	3058 (47.0)
SD	0 (0.0)	42 (1.8)	1988 (70.8)	9 (0.7)	2039 (31.3)
PD	0 (0.0)	0 (0.0)	12 (0.4)	1072 (88.6)	1084 (16.6)
No result	0 (0.0)	0 (0.0)	0 (0.0)	6 (0.5)	6 (0.1)

a Percentages on the top bar are row percentages, showing rates per RECIST version 1.0.
b RECIST version 1.1 includes all other changes described, such as the changes for lymph nodes and a minimum increase of 5 mm for PD.

Table 7 – PFS outcome using the three methods.

RECIST version	PFS short term			
	Patients (*N*)	Observed events (*O*)	Median (95% CI) (Months)	% at 6 months (95% CI)
Patients with more than 5 target lesions at baseline				
Version 1.1 with a maximum of 3 targets version	585	409	5.22 (4.34, 5.55)	42.89 (38.46, 47.24)
Version 1.1 with a maximum of 5 target lesions	585	409	5.22 (4.47, 5.68)	43.67 (39.24, 48.01)
Version 1.0	585	409	5.26 (4.70, 5.65)	43.02 (38.59, 47.38)
All patients				
Version 1.1 with a maximum of 3 targets version	6512	4501	6.31 (6.21, 6.44)	53.75 (52.44, 55.04)
Version 1.1 with a maximum of 5 target lesions	6512	4502	6.31 (6.21, 6.44)	53.78 (52.47, 55.07)
Version 1.0	6512	4531	6.31 (6.21, 6.44)	53.62 (52.31, 54.91)

earlier, depending on the method) but this number is small. Again, when restricting the number of targets being measured to 2 or 1, discrepancies increase, but less so than for response. The majority of patients progressing due to new lesions explain this.

3.3. *Sensitivity analyses*

As shown in Tables 5 and 8, analyses were also conducted using a maximum of 1 or 2 target lesions per patient. For these analyses, the differences in outcome between 10 and 1 or 2 lesions were larger (more than 5% discrepancy) and this level of discordance was considered unacceptable by the RECIST working group to recommend using such small numbers of reported lesions in patients where more lesions are present.

If the duration of no change in anatomical size from baseline to qualify a patient as having a stable disease were increased from 6 weeks to 3 months, then for all methods (RECIST version 1.0 as well as 3 or 5 maximum targets)

Table 8 – Reason for progression.

RECIST version 1.1	Calculated reason for progression: RECIST version 1.0 restricted to patients with more than 5 target lesions at baseline					
	No progression (*N* = 176)	New lesion +/– other (*N* = 229)	Progression of target lesions (*N* = 54)	Progression of non-target lesions (*N* = 116)	Death without progression (*N* = 10)	Total (*N* = 585)
	N (%)	*N* (%)	*N* (%)	*N* (%)	N (%)	*N* (%)
With a maximum of 5 target lesions						
No progression	170 (96.6)	0 (0.0)	6 (11.1)	0 (0.0)	0 (0.0)	176 (30.1)
New lesion +/– other	0 (0.0)	225 (98.3)	4 (7.4)	0 (0.0)	0 (0.0)	229 (39.1)
Progression of target lesions	6 (3.4)	4 (1.7)	44 (81.5)	5 (4.3)	1 (10.0)	60 (10.3)
Progression of non-target lesions	0 (0.0)	0 (0.0)	0 (0.0)	111 (95.7)	0 (0.0)	111 (19.0)
Death without progression	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	9 (90.0)	9 (1.5)
With a maximum of 3 target lesions						
No progression	167 (94.9)	0 (0.0)	9 (16.7)	0 (0.0)	0 (0.0)	176 (30.1)
New lesion +/– other	0 (0.0)	224 (97.8)	4 (7.4)	0 (0.0)	0 (0.0)	228 (39.0)
Progression of target lesions	9 (5.1)	5 (2.2)	40 (74.1)	5 (4.3)	1 (10.0)	60 (10.3)
Progression of non-target lesions	0 (0.0)	0 (0.0)	0 (0.0)	111 (95.7)	0 (0.0)	111 (19.0)
Death without progression	0 (0.0)	0 (0.0)	1 (1.9)	0 (0.0)	9 (90.0)	10 (1.7)
With a maximum of 2 target lesions						
No progression	164 (93.2)	0 (0.0)	9 (16.7)	0 (0.0)	0 (0.0)	173 (29.6)
New lesion +/– other	0 (0.0)	221 (96.5)	3 (5.6)	0 (0.0)	0 (0.0)	224 (38.3)
Progression of target lesions	12 (6.8)	8 (3.5)	41 (75.9)	6 (5.2)	1 (10.0)	68 (11.6)
Progression of non-target lesions	0 (0.0)	0 (0.0)	0 (0.0)	110 (94.8)	0 (0.0)	110 (18.8)
Death without progression	0 (0.0)	0 (0.0)	1 (1.9)	0 (0.0)	9 (90.0)	10 (1.7)
With 1 target lesion						
No progression	156 (88.6)	0 (0.0)	10 (18.5)	0 (0.0)	0 (0.0)	166 (28.4)
New lesion +/– other	0 (0.0)	214 (93.4)	7 (13.0)	0 (0.0)	0 (0.0)	221 (37.8)
Progression of target lesions	20 (11.4)	15 (6.6)	35 (64.8)	13 (11.2)	1 (10.0)	84 (14.4)
Progression of non-target lesions	0 (0.0)	0 (0.0)	1 (1.9)	103 (88.8)	0 (0.0)	104 (17.8)
Death without progression	0 (0.0)	0 (0.0)	1 (1.9)	0 (0.0)	9 (90.0)	10 (1.7)

(*continued on next page*)

Table 8 – *continued*

	Calculated reason for progression					
	RECIST version 1.0					
	All patients					
RECIST version 1.1	No progression (*N* = 1981)	New lesion +/– other (*N* = 2220)	Progression of target lesions (*N* = 1145)	Progression of non-target lesions (*N* = 946)	Death without progression (*N* = 220)	Total (*N* = 6512)
	N (%)	*N* (%)	*N* (%)	*N* (%)	*N* (%)	*N* (%)
With a maximum of 5 target lesions						
No progression	1932 (97.5)	0 (0.0)	78 (6.8)	0 (0.0)	0 (0.0)	2010 (30.9)
New lesion +/– other	0 (0.0)	2203 (99.2)	18 (1.6)	0 (0.0)	0 (0.0)	2221 (34.1)
Progression of target lesions	49 (2.5)	17 (0.8)	1037 (90.6)	11 (1.2)	3 (1.4)	1117 (17.2)
Progression of non-target lesions	0 (0.0)	0 (0.0)	9 (0.8)	935 (98.8)	0 (0.0)	944 (14.5)
Death without progression	0 (0.0)	0 (0.0)	3 (0.3)	0 (0.0)	217 (98.6)	220 (3.4)
With a maximum of 3 targets lesions						
No progression	1925 (97.2)	0 (0.0)	86 (7.5)	0 (0.0)	0 (0.0)	2011 (30.9)
New lesion +/– other	0 (0.0)	2200 (99.1)	19 (1.7)	0 (0.0)	0 (0.0)	2219 (34.1)
Progression of target lesions	56 (2.8)	20 (0.9)	1027 (89.7)	11 (1.2)	4 (1.8)	1118 (17.2)
Progression of non-target lesions	0 (0.0)	0 (0.0)	9 (0.8)	935 (98.8)	0 (0.0)	944 (14.5)
Death without progression	0 (0.0)	0 (0.0)	4 (0.3)	0 (0.0)	216 (98.2)	220 (3.4)
With a maximum of 2 target lesions						
No progression	1905 (96.2)	0 (0.0)	109 (9.5)	0 (0.0)	0 (0.0)	2014 (30.9)
New lesion +/– other	0 (0.0)	2185 (98.4)	19 (1.7)	0 (0.0)	0 (0.0)	2204 (33.8)
Progression of target lesions	76 (3.8)	35 (1.6)	1002 (87.5)	17 (1.8)	7 (3.2)	1137 (17.5)
Progression of non-target lesions	0 (0.0)	0 (0.0)	9 (0.8)	929 (98.2)	0 (0.0)	938 (14.4)
Death without progression	0 (0.0)	0 (0.0)	6 (0.5)	0 (0.0)	213 (96.8)	219 (3.4)
With 1 target lesion						
No progression	1857 (93.7)	0 (0.0)	218 (19.0)	0 (0.0)	0 (0.0)	2075 (31.9)
New lesion +/– other	0 (0.0)	2145 (96.6)	51 (4.5)	0 (0.0)	0 (0.0)	2196 (33.7)
Progression of target lesions	124 (6.3)	75 (3.4)	848 (74.1)	30 (3.2)	7 (3.2)	1084 (16.6)
Progression of non-target lesions	0 (0.0)	0 (0.0)	13 (1.1)	916 (96.8)	0 (0.0)	929 (14.3)
Death without progression	0 (0.0)	0 (0.0)	15 (1.3)	0 (0.0)	213 (96.8)	228 (3.5)

approximately 1200 (of approximately 2800) fewer cases of SD would be identified. Of these, 700 would be assigned PD, while 500 would not have an outcome for the RECIST version 1.0 calculation (their last observation is a stable disease, but within 3 months of baseline). However, the level of consistency between the measurements of lesions (10, 5 or 3) is very similar to that shown in Tables 5–8.

If a doubling of the longest diameter of lesions that were at least 2 cm at baseline and that were not selected as target lesions would be considered as an ad-hoc interpretation of unequivocal progression, an additional 5 patients would have a best response of PD when selecting a maximum of 5 targets. For the PFS end-point this would result in 18 additional events in 6512 patients, with no impact on median and point estimates.

In analyses where the selection of target lesions within organs were random, i.e. representativity of organs was preserved, but apart from that target lesions were selected at random, the overall response rates and progression-free survival did not differ from the method where the largest lesions were selected.

4. Discussion

This database served to quantitatively assess the impact of proposed modifications to RECIST version 1.0 through an evaluation using a large available set of the actual data that were prospectively collected in 16 clinical trials. It must be emphasised that the algorithmic approach used in this exercise is not equivalent to the actual conduct of performing a prospective study using these new guidelines. Previous investigations of this kind either concentrated on modelling and estimating the amount of measurement error[17,18] or followed an approach that is close to the one we have used.[19] This distinction may explain the difference in the resulting recommendations. The first class, which tries to ensure concordance with some (partially unknown) gold standard of true response status, comes to a more conservative recommendation of following six target lesions. Our analysis (and the one by Hillman et al.) observes little difference between RECIST version 1.0 and modified RECIST version 1.1[16] when maximising the number of reported targets to two or three. In a warehouse such as this, which relies on the reported measurements rather than on the detail of any individual measurement, it is impossible to determine the overall level of measurement accuracy with a hypothetical gold standard.

When determining the appropriate number of target lesions to measure one must balance collecting the data necessary to assess the true state of the tumour burden for patient management purposes and the practical limitations of conducting clinical trials. This factor leads us to prefer, all other factors being equal, to measure fewer lesions.

Our analyses suggest that it is possible to restrict the number of target lesions to report to a maximum of five, or potentially three measurable lesions, without causing meaningful changes in the reported trial outcomes.

Most of the trials in the warehouse underwent centralised review,[20] and can be assumed to have been cleaned accordingly. They thus reflect the status and quality of Phase III cancer clinical trials, as they exist today.

When compared against the best response calculated using all the lesions up to 10 per patient, inconsistencies between the calculations of RECIST version 1.0 and methods with 3 or 5 targets as proposed here, go both ways: in some cases the conclusion of version 1.0 is confirmed by the original assessment, in others the new method is closer to the original assessment.

We observed that the newer, more biologically appropriate consideration of lymph node lesions[21] as adapted in RECIST 1.1 had no major impact on the overall rates of response or progression. It does allow a small number of additional complete responses to be demonstrated. In an actual independent radiology review, these may already have been assigned as complete responses. Removing the requirement of response confirmation increases the percentage of responders by an estimated relative 19% (i.e. approximately multiplying the response rate by 1.19, e.g. an estimated response rate of 30% would increase to $.30 \times 1.19 = 36\%$).

Progression-free survival[22,23] was almost completely unaffected by any of the considered RECIST modifications. This is due in part to the fact that progressions documented solely by increases in target lesions (and not by the occurrence of new lesions or unequivocal progression of non-target lesions) constitute only 25% (1145/4531) of all PFS events in the warehouse. At the individual patient level, only 127 patients had a difference in progression status between the two versions of RECIST.

Importantly, when investigating the potential impact of individual studies on these results, no individual study stood out from the rest (thereby possibly skewing the outcome) at either the individual patient or trial level. Perhaps most importantly, measuring fewer lesions did not impact the trial level summary data of response rate and PFS.

When interpreting these results, the following (potential) sources of bias must be considered. First, the warehouse is restricted to the target lesions reported in the original trials. The data warehouse cannot guarantee that patients did not have more potential target lesions than those reported. Based on the number of reported lesions in the trials considered here, as reflected in Table 3, investigators may not be complying with the RECIST version 1.0 rule of measuring all lesions up to a maximum of 10 targets, but instead are reporting fewer lesions, likely due to the logistical constraints. While this in theory may be undesirable (see discussion on measurement error later in this paper) it does reflect the real world experience in 16 large phase III trials. Similarly, it is not straightforward to evaluate directly the importance of measurement error in the data warehouse. This analysis does not take into account reliability, reproducibility or accuracy of the actual radiological measurement other than by requiring the usual cutoffs for assigning partial response and progressive disease. While the measurement issues may not necessarily introduce significant bias in a randomised Phase III trial, clearly individual patient assignment of response or progression will be impacted.

Theoretically, the evaluation of an individual patient's overall tumour burden will improve by monitoring a greater number of target lesions. Consistent definitions of response/progression categories for patient management and trial end-point purposes is critical or trial data may be missing

because patients have been removed from study based on patient management criteria prior to attaining the corresponding trial end-point.

Measurement variability depends on tumour location. Hopper et al. demonstrated wide variability in the selection of target lesions across multiple radiologists, with a 15% inter-observer variability in CT measurements even when radiologists were presented with a relatively small set of tumours and types.[24] It is further recognised that there are specific metastatic sites where the lesions are more prone to measurement error, and therefore, the measurement of multiple lesions in these organs is a reasonable recommendation. By doing so, a greater proportion of a patient's tumour burden is assessed, and any single lesion measurement error would be minimised over time. Documented challenges in measuring, for instance, hypopharyngeal, laryngeal, renal lesions and peritoneal metastases are well known.[25]

As a specific example of these issues, the largest body of contemporary literature evaluates lung lesion measurements and change over time. Even with optimal computerised segmentation of nodules, single nodule-interscan variability range from −12% to 12.4% and in examples of lung nodules where the manual delineation of the tumour was necessary, this single lesion interscan variability increased from −27% to +30%.[26] In the worst-case scenario, for individual tumour measurements, an assessment of progression up to 27.5% tumours may be misclassified based on the measurement differences at CT.[27] The first true CT reproducibility study was performed in lung cancer. It demonstrated up to 12.5% change to be within the 95% confidence interval in a single primary lung lesion surrounded by aerated lung. Inter- and intra-observer variability was even greater when considering the repeatability of a radiologist's measurements. Even under these ideal conditions, therefore, it appears that measurement of a single lesion in an organ may not suffice to provide a reproducible or reliable metric of change.

Increasing the number of target lesions measured in an organ will reduce part of the error in the percent change of sum of diameters to the extent that it is not a systematic error (such as those due to the observer, the method, characteristics of the patient). Technological advances to reduce measurement error, as opposed to measuring more lesions, may best address many of these issues.

In some primary tumours, differential responses may be seen based upon the sites or predominant sites of metastases or those included as target measurements. For instance, inclusion of the renal primary in measurement of target disease in renal cell has been shown to impact response rate and time to progression analyses.[28] Therefore, theoretically, optimal assessment of tumour burden should encompass as many sites of target lesions with reproducible and repeatable measurements of disease. In the warehouse assembled, these theoretical concerns were not manifest, however, as the types of therapies evolve and imaging modalities and technologies change, the appropriate number of lesions to monitor for both managing patient care and evaluating treatment efficacy will be an important area of research.

Based on the analyses done on the warehouse, and based on the review of numerous individual cases where the programmed algorithm resulted in a different outcome than that documented in the originating trial database (either by the local investigator or centrally) we believe the priority in clinical practice should be to obtain consistency in the measurement of a restricted number of selected lesions. In planning the trial, imaging of selected target lesions at each intended visit should take priority over shortening the interval between assessments. In evaluating progression, in the light of these results, the importance of an adequate monitoring for the appearance of new lesions is emphasised, as the findings of new lesions are not demonstrably correlated to the evolution of the sum of lesion diameters. The last recommendation is to define carefully – at least at the study protocol level – unequivocal progression of non-target (or non-measurable) lesions, and to ensure that appropriate documentation is reported in the clinical database.

5. Conclusions

Based on our analysis of over 6000 patients from 16 completed trials, there is little difference between the results obtained at either the individual patient level or the trial level for the assessment of response or progression when measuring 10, 5 or 3 lesions. In addition, imposing criteria for lymph nodes, and a condition of a minimal 5 mm increase in sum of LD to reach progressive disease do not substantially impact conclusions, while these changes have a desirable effect of bringing assessments closer to clinical reality.

Requiring response confirmation has no impact on progressive disease findings (with the exception of an increased likelihood of finding progressive disease at the additional confirmation visit), but has a substantial impact on the response rate. We therefore retained the requirement for response confirmation in RECIST version 1.1 in trials where the objective response rate is the primary trial end-point. An analysis of which criteria for response confirmation (required or not) results in a superior prediction of overall survival would be instructive, but is beyond the scope of this exercise. Regarding the final choice of the number of lesions to be assessed for RECIST version 1.1, even though the algorithm using a maximum of three targets shows high concordance with the 10 lesion requirement in terms of response and TTP assignment, one can be concerned that the appropriate assessment of disease within an organ requires two lesions to be followed per organ. Hence, the approaches of following two target lesions per organ, up to a maximum of either three or five target lesions in total both seem reasonable strategies based on the data warehouse.

Conflict of interest statement

None declared.

Acknowledgements

This publication was supported by grant numbers 2U10 CA11488-35 and 5U10 CA11488-38 from the National Cancer Institute (Bethesda, Maryland, USA). Its contents are solely

the responsibility of the authors and do not necessarily represent the official views of the NCI. This research project was supported by Fonds Cancer/FOCA (Belgium).

The authors wish to thank the following organisations for making the data available for this analysis:

- Amgen
- AstraZeneca
- Breast Cancer International Research Group (BCIRG)
- Bristol-Myers Squibb
- European Organisation for Research and Treatment of Cancer (EORTC) Breast Cancer Group and Gastrointestinal Group
- Erasmus University Medical Centre, Rotterdam, The Netherlands
- Genentech
- Pfizer
- RadPharm
- Roche
- Sanofi Aventis

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* 2000;**92**(3).
2. Bontenbal M, Creemers G-J, Braun HJ, et al. Phase II to III study comparing doxorubicin and docetaxel with fluorouracil, doxorubicin, and cyclophosphamide as first-line chemotherapy in patients with metastatic breast cancer: results of a Dutch Community Setting Trial for the Clinical Trial Group of the Comprehensive Cancer Centre. *JCO* 2005;**23**(28):7081–8.
3. Giaccone G, Herbst RS, Manegold C, et al. Gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: a phase III trial—INTACT 1. *JCO* 2004;**22**(5):777–84.
4. Herbst RS, Giaccone G, Schiller JH, et al. Gefitinib in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: a phase III trial—INTACT 2. *JCO* 2004;**22**(5):785–94.
5. Jassem J, Pieńkowski T, Pluzanska A, et alfor the Central & Eastern Europe and Israel Paclitaxel Breast Cancer Study Group. Doxorubicin and paclitaxel versus fluorouracil, doxorubicin, and cyclophosphamide as first-line therapy for women with metastatic breast cancer: final results of a Randomized Phase III Multicenter Trial. *JCO* 2001;**19**(6):1707–15.
6. Chan S, Friedrichs K, Noel D, et alfor the 303 Study Group. Prospective randomized trial of docetaxel versus doxorubicin in patients with metastatic breast cancer. *JCO* 1999;**17**(8):2341–54.
7. Nabholtz JM, Falkson C, Campos D, et al. Docetaxel and doxorubicin compared with doxorubicin and cyclophosphamide as first-line chemotherapy for metastatic breast cancer: results of a Randomized, Multicenter, Phase III Trial. *JCO* 2003;**21**(6):968–75.
8. Mackey JR, Paterson A, Dirix L, et al. Final results of the phase III randomised trial comparing docetaxel (T), doxorubicin (A) and cyclophosphamide (C) to FAC as first line chemotherapy (CT) for patients (pts) with metastatic breast cancer (MBC). *Proc Am Soc Clin Oncol* 2002;**21**:35a [abstr 137].
9. Hurwitz HI, Fehrenbacher L, Hainsworth JD, et al. Bevacizumab in combination with fluorouracil and leucovorin: an active regimen for first-line metastatic colorectal cancer. *JCO* 2005;**23**(15):3502–8.
10. Herbst RS, Prager D, Hermann R, et al. TRIBUTE: a phase III trial of erlotinib hydrochloride (OSI-774) combined with carboplatin and paclitaxel chemotherapy in advanced non-small-cell lung cancer. *JCO* 2005;**23**(25):5892–9.
11. Paridaens R, Biganzoli L, Bruning P, et al. Paclitaxel versus doxorubicin as first-line single-agent chemotherapy for metastatic breast cancer: a European Organization for Research and Treatment of Cancer Randomized Study with cross-over. *J Clin Oncol* 2000;**18**(4):724–33.
12. Biganzoli L, Cufer T, Bruning P, et al. Doxorubicin and paclitaxel versus doxorubicin and cyclophosphamide as first-line chemotherapy in metastatic breast cancer: the European Organization for Research and Treatment of Cancer 10961 Multicenter Phase III Trial. *JCO* 2002;**20**(14):3114–21.
13. Gatzemeier U, Pluzanska A, Szczesna A, et al. Phase III Study of Erlotinib in combination with cisplatin and gemcitabine in advanced non-small-cell lung cancer: The Tarceva Lung Cancer Investigation Trial. *JCO* 2007;**25**(12):1545–52.
14. Köhne CH, van Cutsem E, Wils J, et al. Phase III study of weekly high-dose infusional fluorouracil plus folinic acid with or without irinotecan in patients with metastatic colorectal cancer: European Organisation for Research and Treatment of Cancer Gastrointestinal Group Study 40986. *JCO* 2005;**23**(22):4856–65.
15. Casali PG, Garrett CR, Blackstein ME, et al. Updated results from a phase III trial of sunitinib in GIST patients (pts) for whom imatinib (IM) therapy has failed due to resistance or intolerance, 2006 ASCO Annual Meeting Proceedings Part I. *JCO* 2006;**24**(18S (June 20 Suppl.)):9513.
16. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumors: Revised RECIST guideline (version 1.1). *Eur J Cancer*, 2009;**45**:228–47.
17. Zacharia TT, Saini S, Halpern EF, Sumner JE. *CT of colon cancer metastases to the liver using modified RECIST criteria: determining the ideal number of target lesions to measure AJR* 2006;**186**.
18. Schwartz LH, Mazumdar M, Brown W, Smith A, Panicek DM. Variability in response assessment in solid tumors: effect of number of lesions chosen for measurement. *Clin Cancer Res* 2003;**9**:4318–23.
19. Hillman SL, An M-W, O'Connell MJ, et al. An evaluation of the optimal number of lesions needed for tumor evaluation using the RECIST criteria: A North Central Cancer Treatment Group (NCCTG) Investigation, submitted for publication.
20. Ford R, Schwartz L, Dancey J, et al. Lessons learned from independent central review. *Eur J Cancer* 2009;**45**:268–74.
21. Schwartz LH, Bogaerts J, Ford R, et al. Evaluation of lymph nodes with RECIST 1.1. *Eur J Cancer* 2009;**45**:261–7.
22. Ratain MJ, Sargent DJ. Optimizing the design of phase II oncology trials: The importance of randomization. *Eur J Cancer* 2009;**45**:275–80.
23. Dancey JE, Dodd LE, Ford R, et al. Recommendations for the assessment of progression in randomized cancer treatment trials. *Eur J Cancer* 2009;**45**:281–9.
24. Hopper KD, Kasales CJ, Van Slyke MA. Analysis of interobserver and intraobserver variability in CT tumor measurements. *AJR* 1996;**167**:851–4.
25. Preda L, Lovati E, Chiesa F, et al. Measurement by multidetector CT scan of the volume of hypopharyngeal and laryngeal tumours: accuracy and reproducibility. *Eur Radiol* 2007;**17**:2096–102.

26. Gietema HA, Schaefer-Prokop CM, Mali WPTM, et al. Pulmonary nodules: Interscan variability of semiautomated volume measurements with multisection CT – influence of inspiration level, nodule size and segmentation performance. *Radiol* 2007;**245**:888–94.
27. Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and Intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *JCO* 2003;**21**(13):2574–82.
28. Punnen S, Haider MA, Lockwood G, Moulding F, O'Malley ME, Jewett MA. Variability in size measurement of renal masses smaller than 4 cm on computerized tomography. *J Urol* 2006;**176**(6 Pt 1):2386–90.