

Twenty years on: RECIST as a biomarker of response in solid tumours. An EORTC Imaging Group - ESOL joint paper

Laure Fournier^{1, 2, 3, 4, 5*}, Lioe-Fee De Geus-Oei^{6, 7, 5}, Daniele Regge^{8, 4, 9}, Daniela E. Oprea-Lager^{10, 5}, Melvin D'Anastasi^{4, 11}, Luc Bidaut^{12, 5}, Tobias Bäuerle^{4, 13}, Egesta Lopci^{14, 5}, Giovanni Cappello^{4, 8, 9}, Frederic Lecouvet^{5, 15}, Marius Mayerhoefer^{4, 16, 17}, Wolfgang G. Kunz^{18, 4, 19}, Joost J. Verhoeff^{20, 5}, Damiano Caruso^{4, 21}, Marion Smits^{22, 23, 5}, Ralf-Thorsten Hoffmann^{24, 4}, Sofia Gourtsoyianni^{25, 4}, Regina BeetsTan^{4, 26, 27}, Emanuele Neri^{4, 28}, Nandita M. DeSouza^{29, 5, 30, 31}, Christophe M. Deroose^{32, 33, 5}, Caroline Caramella^{5, 34, 35}

¹Service de radiologie cardiovasculaire, Hôpital Européen Georges-Pompidou, France, ²Université de Paris, France, ³INSERM U970 Paris-Centre de Recherche Cardiovasculaire (PARCC), France, ⁴European Society of Oncologic Imaging (ESOL), European Society of Radiology, Austria, ⁵Imaging Group, European Organisation of Research and Treatment in Cancer (EORTC), Belgium, ⁶Department of Radiology, Leiden University Medical Center, Netherlands, ⁷University of Twente, Netherlands, ⁸Department of Surgical Sciences, School of Medicine, University of Turin, Italy, ⁹Radiology Unit, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Italy, ¹⁰Department of Radiology and Nuclear Medicine, VU University Medical Center, Netherlands, ¹¹Medical Imaging Department, Mater Dei Hospital, University of Malta, Malta, ¹²College of Science, University of Lincoln, United Kingdom, ¹³Institute of Radiology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany, ¹⁴Nuclear Medicine Unit, Humanitas Clinical and Research Hospital, Italy, ¹⁵Department of Radiology, Institut de Recherche Expérimentale et Clinique (IREC), Cliniques Universitaires Saint Luc, Université Catholique de Louvain (UCLouvain), Belgium, ¹⁶Department of Radiology, Memorial Sloan Kettering Cancer Center, United States, ¹⁷Department of Biomedical Imaging and Image-Guided Therapy, Medical University of Vienna, Austria, ¹⁸European Organisation for Research and Treatment of Cancer, Belgium, ¹⁹Department of Radiology, University Hospital, LMU, Germany, ²⁰Department of Radiotherapy, University Medical Center Utrecht, Netherlands, ²¹Department of Medical Surgical Sciences and Translational Medicine, Faculty of Medicine and Psychology, Sapienza University of Rome, Italy, ²²Department of Radiology and Nuclear Medicine, Erasmus Medical Center, Netherlands, ²³Department of Neurology, Brain Tumor Center, Erasmus MC Cancer Institute, Netherlands, ²⁴Institute and Polyclinic for Diagnostic and Interventional Radiology, University Hospital Carl Gustav Carus, Germany, ²⁵Department of Radiology, School of Medicine, National and Kapodistrian University of Athens, Areteion Hospital, Greece, ²⁶Department of Radiology, The Netherlands Cancer Institute, Netherlands, ²⁷School for Oncology and Developmental Biology, Maastricht University, Netherlands, ²⁸Diagnostic and Interventional Radiology, Department of Translational Research and of New Surgical and Medical Technologies, University of Pisa, Italy, ²⁹Division of Radiotherapy and Imaging, Institute of Cancer Research (ICR), United Kingdom, ³⁰European Imaging Biomarkers Alliance (EIBALL), European Society of Radiology, Austria, ³¹Quantitative Imaging Biomarkers Alliance, Radiological Society of North America, United States, ³²Department of Nuclear Medicine, University Hospitals Leuven, Belgium, ³³Department of Imaging and Pathology, KU Leuven, Belgium, ³⁴Radiology Department, Hôpital Marie Lannelongue, Groupe Hospitalier Paris Saint Joseph, France, ³⁵Université Paris-Saclay, France

Submitted to Journal:

Frontiers in Oncology

Specialty Section:

Cancer Imaging and Image-directed Interventions

Article type:

Systematic Review Article

Manuscript ID:

800547

Received on:

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

All authors contributed to conception and design. CC and LF wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version

Keywords

tumour, biomarker, imaging, response, RECIST

Abstract

Word count: 329

Response evaluation criteria in solid tumours (RECIST) v1.1 are currently the reference standard for evaluating efficacy of therapies in patients with solid tumours who are included in clinical trials, and they are widely used and accepted by regulatory agencies. This expert statement discusses the principles underlying RECIST, as well as their reproducibility and limitations. While the RECIST framework may not be perfect, the scientific bases for the anticancer drugs that have been approved using a RECIST-based surrogate endpoint remain valid. Importantly, changes in measurement have to meet thresholds defined by RECIST for response classification within thus partly circumventing the problems of measurement variability. The RECIST framework also applies to clinical patients in individual settings even though the relationship between tumour size changes and outcome from cohort studies is not necessarily translatable to individual cases. As reproducibility of RECIST measurements is impacted by reader experience, choice of target lesions and detection/interpretation of new lesions, it can result in patients changing response categories when measurements are near threshold values or if new lesions are missed or incorrectly interpreted. There are several situations where RECIST will fail to evaluate treatment-induced changes correctly: knowledge and understanding of these is crucial for correct interpretation. Also, some patterns of response/progression cannot be correctly documented by RECIST, particularly in relation to organ-site (e.g. bone without associated soft-tissue lesion) and treatment type (e.g. focal therapies). These require specialist reader experience and communication with oncologists to determine the actual impact of the therapy and best evaluation strategy. In such situations, alternative imaging markers for tumour response may be used but the sources of variability of individual imaging techniques need to be known and accounted for. Communication between imaging experts and oncologists regarding the level of confidence in a biomarker is essential for the correct interpretation of a biomarker and its application to clinical decision-making. Though measurement automation is desirable and potentially reduces the variability of results, associated technical difficulties must be overcome, and human adjudications may be required.

Contribution to the field

This paper is a position paper written by EORTC imaging groups members and ESOI (ESR) group members. It addresses several questions and controversies, that have been debated over time about RECIST criteria. How were RECIST thresholds established? Do RECIST categories predict outcome? How reproducible is RECIST? How reproducible are other biomarkers? What are common RECIST limitations? When is RECIST response assessment misleading? Should (could) RECIST be automated? RECIST in novel drug development RECIST: only as good as its users?

Funding statement

no funding

Data availability statement

Generated Statement: The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

1 **Twenty years on: RECIST as a biomarker of response in solid tumours**

2 **An EORTC Imaging Group – ESOI joint paper**

3
4 **AUTHORS AND AFFILIATIONS**

5
6
7
8 Fournier L (1, 2, 3), de Geus-Oei L-F (1,4,5), Regge D (2, 6, 7), Oprea-Lager D.E (1, 8), Danastasi M (2, 9), Bidaut L (1, 10), Baeuerle T (2, 11), Lopci E (1, 12),
9 Cappello G (6, 7), Lecouvet F (1, 13), Mayerhoefer M E (2, 14, 15), Kunz W G (1, 2, 16), Verhoeff J (1, 17), Caruso D (2, 18), Smits M (1, 19, 20), Hoffmann R-
10 T (2, 21), Gourtsoyianni S (2, 22), Beets-Tan RGH (2, 23, 24), Neri E (2, 25), de Souza N (1, 26, 27, 28), Deroose C M (1, 9), Caramella C (1, 30)

- 11
12
13 1- Imaging Group, European Organisation of Research and Treatment in Cancer (EORTC), Brussels, Belgium.
14 2- European Society of Oncologic Imaging (ESOI), European Society of Radiology, Vienna, Austria.
15 3- Université de Paris, AP-HP, Hopital europeen Georges Pompidou, Department of Radiology, PARCC UMRS 970, INSERM, Paris, France.
16 4- Department of Radiology, Leiden University Medical Center, Leiden (NL)
17 5- Biomedical Photonic Imaging Group, University of Twente, Enschede (NL)
18 6- Department of Surgical Sciences, University of Turin, 10124 Turin, Italy.
19 7- Radiology Unit, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, 10060 Turin, Italy.
20 8- Department of Radiology & Nuclear Medicine, Cancer Centre Amsterdam, Amsterdam University Medical Centers (VU University), Amsterdam, The
21 Netherlands
22 9- Medical Imaging Department, Mater Dei Hospital, University of Malta, Msida, Malta.
23 10- College of Science, University of Lincoln, Lincoln, LN6 7TS, UK
24 11- Institute of Radiology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany.
25 12- Nuclear Medicine Unit, IRCCS – Humanitas Research Hospital, Via Manzoni 56, 20089 Rozzano (MI), Italy
26 13- Department of Radiology, Institut de Recherche Expérimentale et Clinique (IREC), Cliniques Universitaires Saint Luc, Université Catholique de Louvain
27 (UCLouvain), B-1200, Brussels, Belgium
28 14- Dept. of Radiology, Memorial Sloan Kettering Cancer Center, USA
29 15- Dept. of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Austria
30 16- Department of Radiology, University Hospital, LMU Munich, Munich, Germany.
31 17- Department of Radiotherapy, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
32 18- Department of Medical-Surgical Sciences and Translational Medicine, Sapienza University of Rome.
33 19- Dept. of Radiology & Nuclear Medicine, Erasmus MC, University Medical Centre Rotterdam, Rotterdam (NL)
34 20- Brain Tumour Centre, Erasmus MC Cancer Institute, Rotterdam (NL)
35 21- Institute and Policlinic for Diagnostic and Interventional Radiology University Hospital Carl-Gustav-Carus Technical University Dresden
36 22- Department of Radiology, School of Medicine, National and Kapodistrian University of Athens, Areteion Hospital

- 37 **23-** Department of Radiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands.
38 **24-** GROW School for Oncology and Developmental Biology, Maastricht University, Maastricht, The Netherlands.
39 **25-** Diagnostic and Interventional Radiology, Department of Translational Research and of New Surgical and Medical Technologies, University of Pisa, Pisa
40 56126, Italy.
41 **26-** Division of Radiotherapy and Imaging, The Institute of Cancer Research and Royal Marsden NHS Foundation Trust, London, UK.
42 **27-** European Imaging Biomarkers Alliance (EIBALL), European Society of Radiology, Vienna, Austria.
43 **28-** Quantitative Imaging Biomarkers Alliance, Radiological Society of North America, Oak Brook, IL, USA.
44 **29-** Nuclear Medicine, University Hospitals Leuven; Nuclear Medicine & Molecular Imaging, Department of Imaging and Pathology, KU Leuven; both in
45 Leuven, Belgium
46 **30-** Radiology Department, Hôpital Marie Lannelongue, Groupe Hospitalier Paris Saint Joseph. Centre International des Cancers Thoraciques, Université
47 Paris-Saclay, Le Plessis-Robinson, France.
48

49

Corresponding author: Laure Fournier. Laure.fournier@aphp.fr

50

51

52

53

54

55

56

Key words: tumour, response, biomarker, imaging, RECIST

57 **Abstract**

58 Response evaluation criteria in solid tumours (RECIST) v1.1 are currently the reference standard for evaluating efficacy of therapies in patients with solid
59 tumours who are included in clinical trials, and they are widely used and accepted by regulatory agencies. This expert statement discusses the principles
60 underlying RECIST, as well as their reproducibility and limitations.

61 While the RECIST framework may not be perfect, the scientific bases for the anticancer drugs that have been approved using a RECIST-based surrogate
62 endpoint remain valid. Importantly, changes in measurement have to meet thresholds defined by RECIST for response classification within thus partly
63 circumventing the problems of measurement variability. The RECIST framework also applies to clinical patients in individual settings even though the
64 relationship between tumour size changes and outcome from cohort studies is not necessarily translatable to individual cases. As reproducibility of RECIST
65 measurements is impacted by reader experience, choice of target lesions and detection/interpretation of new lesions, it can result in patients changing
66 response categories when measurements are near threshold values or if new lesions are missed or incorrectly interpreted.

67 There are several situations where RECIST will fail to evaluate treatment-induced changes correctly; knowledge and understanding of these is crucial for
68 correct interpretation. Also, some patterns of response/progression cannot be correctly documented by RECIST, particularly in relation to organ-site (e.g.
69 bone without associated soft-tissue lesion) and treatment type (e.g. focal therapies). These require specialist reader experience and communication with
70 oncologists to determine the actual impact of the therapy and best evaluation strategy. In such situations, alternative imaging markers for tumour response
71 may be used but the sources of variability of individual imaging techniques need to be known and accounted for. Communication between imaging experts
72 and oncologists regarding the level of confidence in a biomarker is essential for the correct interpretation of a biomarker and its application to clinical
73 decision-making. Though measurement automation is desirable and potentially reduces the variability of results, associated technical difficulties must be
74 overcome, and human adjudications may be required.

75

76 **Introduction**

77 Imaging plays a major role in the evaluation of tumour response to cancer treatments. It provides an objective in-vivo measurement of tumour burden,
78 and helps oncologists determine whether a treatment should be pursued, interrupted or adapted.

79 Response evaluation criteria in solid tumours (RECIST) v1.1 currently is the reference standard for evaluating efficacy of therapies in patients with solid
80 tumours who are included in clinical trials, and it is widely used and accepted by regulatory agencies(1). However, many publications question both the
81 reproducibility and the clinical relevance of RECIST. This paper is an expert statement aiming to answer some of the questions regarding the principles
82 underlying RECIST and its reproducibility compared to other biomarkers, as well as the limitations to its application and continued role in an era where
83 other biomarkers exist that are more explicitly geared towards tumour-specific properties.

84

85 **How were RECIST thresholds established?**

86 RECIST has instituted several overarching principles underpinning its approach to tumour response evaluation. Primarily, RECIST defines which lesions
87 are measurable in a reliable manner. Among these, it defines a maximal number of lesions ('target lesions') to be measured to yield a quantitative value
88 representative of tumour burden. The remainder are considered 'non-target lesions' and are evaluated qualitatively. On follow-up scans, new lesions
89 indicate progression (Table 1). The threshold for response is defined as a decrease of at least 30% of sum of diameters (SOD) of target lesions compared
90 to baseline, AND no progression of non-target lesions AND no new lesions. The threshold for progressive disease (PD) is defined as an increase of at
91 least 20% of SOD of target lesions compared to nadir AND/ OR unequivocal progression of non-target lesions AND/OR appearance of new lesions.

92

93 The first publication addressing thresholds for determining treatment efficacy was published by Moertel and Hanley in 1976(2). In this study, 16 observers
94 were asked to measure by clinical examination using a caliper the diameters of solid spheres of variable sizes arranged randomly underneath a mattress.
95 Authors suggested the product of two diameters should be used, as this would be more reliable if lesions were not spherical. For this 'clinical' estimate, a
96 50% reduction in the product of two diameters was shown to have an acceptable measurement error estimated between 7-8%. Interestingly, the authors
97 specifically stated that "the purpose was not to predict long-term efficacy but to determine what change in bidimensional size could be confidently
98 considered a change". Progression, on the other hand, was defined as an increase in the product of diameters $\geq 25\%$, but the authors could not justify
99 this cut-off, other than by specifying it "should not necessarily be regarded as influencing the management of the patient".

100 In 1979, the World Health Organization (WHO) provided recommendations for the evaluation of cancer treatments in clinical trials on imaging. Criteria
101 were based not only on the bidimensional measurement of lesions on clinical examination, but also CT or standard radiography (3), transposing results
102 of Moertel and Hanley's study and setting cut-offs for definition of response to -50% and of progression to +25%. However, many technical aspects were
103 not detailed, such as the number of lesions to be measured or what constituted a measurable lesion.

104 In 2000, a working group of European, American and Canadian cancer research organizations (EORTC, NCI, NCIC) defined the Response Evaluation
105 Criteria In Solid Tumours – RECIST(4). They used data from over 4,600 patients enrolled in 14 clinical trials to formulate criteria based on imaging.
106 RECIST used unidimensional measurement of lesions, justified by an extensive comparison of methods of measurement (1D vs. 2D)(5). Moreover, this
107 working group specified conditions of measurement, number of lesions, and detailed how to document progression. Regarding cut-off values for response
108 and progression, the -50% value for response for bidimensional measurement was altered to -30% for unidimensional measurements, and the +25%
109 value for progression for bidimensional measurement was altered to +20% for unidimensional measurements (table 2).

110 RECIST was then revised in 2009 (version 1.1)(1), introducing specific rules for measurement of small axis of lymph nodes and reducing the number of
111 target lesions to five per patient. This new version was also based on data analysis, including a literature review and a simulation using a database of
112 over 6,500 patients and 18,000 lesions. The number of target lesions for example, was chosen by determining the minimum number for which response
113 rates and time to progression were not altered from RECIST 1.0 results(6,7).

114 *Statement #1: RECIST thresholds were chosen to produce a comparable classification of patients in a given category of response when comparing trials*
115 *or even when comparing patients, taking into account tumour measurement variability.*

116 **Table 2: relationship between diameter and corresponding volume**

Diameter ("long axis")	Percentage of variation	Corresponding volume	Percentage of variation
20 mm		4.2 cm ³	
26 mm	+30%	9.2 cm ³	+120%
34 mm	+30%	20.6 cm ³	+ 120%
27 mm	-20%	10.3 cm ³	-50%

117

118 **Do RECIST categories predict outcome?**

119 RECIST criteria were originally tested and validated to provide an objective and reproducible assessment of treatment effect in cancer patients, without
120 any references to patient outcome(8). Yet it seems intuitive that when a tumour decreases in size, a patient will have a better outcome, and vice versa.
121 There is evidence to support this, including some large studies, which pool data from various trials. In over 500 patients with metastatic colorectal cancer
122 treated with combination chemotherapy, a decrease in size resulted in a decreased hazard ratio for overall survival (OS) (9). In a meta-analysis of 24
123 phase I trials, a linear relationship was shown between change in tumour size and survival (10). In a pooled analysis of over 2,700 patients with metastatic
124 renal cell carcinoma treated with anti-angiogenic agents, tumor shrinkage of $\geq 30\%$ resulted in improved OS and progression-free survival (PFS) (11). In
125 addition, the authors demonstrated that tumour shrinkage between 60% and 100% at 6-month follow-up represented an independent prognostic factor
126 for OS. Litière et al also demonstrated in an even larger pooled analysis of over 23,000 patients treated with targeted agents, chemotherapy or a
127 combination thereof(12), that a decrease in tumour size was consistently associated with a lower hazard ratio, while an increase in size was associated
128 with a higher hazard ratio.

129 Tumour response according to RECIST can only be quantified by a decrease in size or number of target lesions, as non-target lesions are not taken into
130 consideration for partial response (PR). Regarding progression however, it is important to consider non-target lesions, as unequivocal progression of non-
131 target lesions or emergence of new lesions defines tumour progression. In over 3,700 patients from 13 trials in the RECIST trial database, the presence
132 of new lesions and progression of non-target lesions were most strongly associated with worse OS (hazard ratios range 1.5–2.3) regardless of tumour
133 type, whereas percentage tumour growth in target lesions contributed less in a multivariate model of OS (13).

134 Finally, in two separate studies (14,15), An et al. compared the predictive ability of RECIST categories vs. longitudinal tumour measurement-based
135 continuous metrics and alternative categorical response metrics such as slope (absolute change in tumour size) and percent change (relative change in
136 tumour size) to predict OS. The databases consisted respectively of almost 2,100 patients from 13 trials and over 1,500 patients from 3 trials with breast
137 cancer, non-small cell lung cancer (NSCLC) or colorectal cancer. Although there seemed to be a slightly better performance for continuous variables, it
138 was not statistically significant, which led the authors to conclude there was no evidence that growth rate or a continuous evaluation of percent change
139 would improve prediction of outcome. However, it may be noted that timing of evaluations, particularly when considering non-continuous variables, may
140 have an impact on their performance and results.

141 *Statement #2: Tumour size changes correlate to outcome at a statistical (cohort) level.*

142

143 **How reproducible is RECIST?**

144 When considering whether RECIST evaluates tumour response correctly, metrology principles guide us to consider two aspects(16): is the measurement
145 “true” (when compared to a “real” value, which defines its accuracy), and is the measurement “precise” (i.e. repeatable and reproducible)?

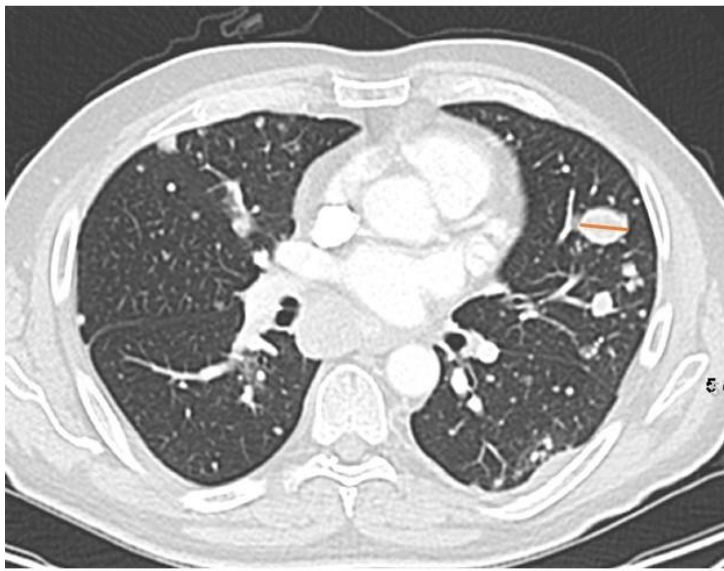
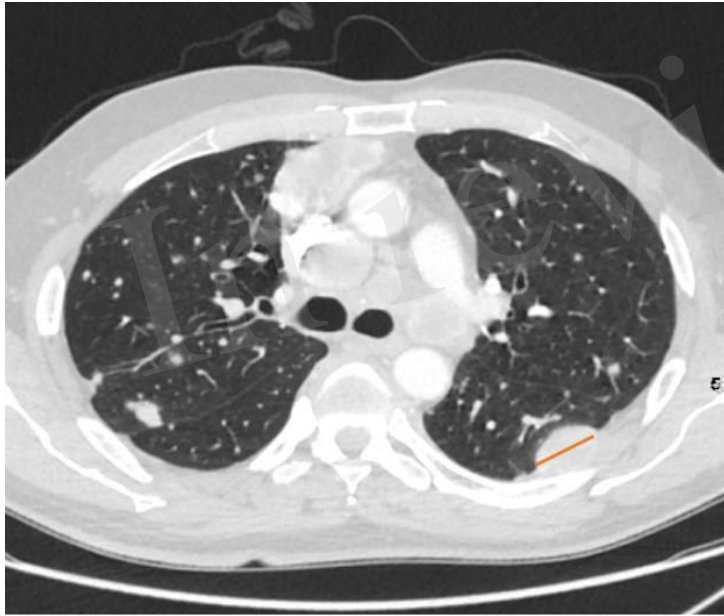
146 Assessing accuracy of change in size measurements would require obtaining “true” values of change in size. As it is not possible to surgically excise all
147 tumours for comparison with imaging, and often inaccurate to compare *ex vivo* with *in vivo* measurements, the true value of an imaging biomarker must
148 be derived from data obtained through a combination of primary tumour excision and phantom studies.

149 Precision refers to the variability of the measurement process and can be evaluated by repeatability (when measurement conditions do not change) and
150 reproducibility (when measurement conditions vary). The precision of RECIST and of response categories has been studied extensively. Table 3 lists the
151 documented reproducibility of RECIST and factors that may impact it. Overall, SOD reproducibility is in the order of +/-20% in multi-observer studies, and
152 +/-10% in single observer studies(17). Important factors associated with RECIST measurement reproducibility are the choice and number of target lesions
153 (Figure 1) and the experience of the reader(s). Where multiple target lesions are used, their selection affects variability: agreement ranges from 0.58 when
154 different targets are chosen to 0.97 when the same targets are used(18). Variability also increases with the number of target lesions selected. For this
155 reason, it has been recommended that a central review in clinical trials should include two readers and one adjudicator(19). Finally, reader experience
156 has major impact on variability, from the selection of the correct reference examination (baseline vs. previous CT) to the detection and proper interpretation
157 of new lesions(20–23). Measurements of well-demarcated lesions and bigger lesions are also more reproducible(17,24,25), which vindicates RECIST
158 recommendations for the choice of target lesions.

159 *Statement #3: RECIST reproducibility is impacted by reader experience, choice of target lesions, lesion characteristics, and detection/interpretation of*
160 *new lesions. At an individual level, this can result in patients being categorised incorrectly when values of SOD are near thresholds or when new lesions*
161 *are either missed or incorrectly interpreted.*

162

163



165 Figure 1 :Selecting target lesions in a 58 yo patient with metastatic renal cell carcinoma. Multiple lung, lymph node, pancreatic and adrenal
166 metastases are present. Lymph nodes should be sampled from different locations where possible. Selection of target lesions at baseline from
167 multiple organ sites is important for response evaluation at a patient level.

168 **How reproducible are other biomarkers?**

169 Table 4 summarises repeatability and reproducibility of some of the other biomarkers suggested or used as alternatives to RECIST for evaluating
170 response. With the abundance of suggested candidate biomarkers in the published literature, the purpose here is not to be comprehensive, but to give a
171 general overview of some of the most frequently explored options for providing a level of comparison with RECIST.

172 A first alternative to measuring a single size dimension as a response biomarker, would be to measure volume of a single or several lesions as an indicator
173 of tumour bulk. This seems particularly important when lesions are irregular in shape, or when they change orientation and are therefore not identically
174 represented on standard axial follow-up scans. Volumetric response on first follow-up CT has been shown to better predict OS than RECIST response
175 (26). Tumour volume response has been utilised in lung (27), cervical (28), and other solid malignancies (29). Despite a trend towards better intra- and
176 inter-observer reproducibility, the routine use of volume has been hampered by the need for manual segmentation, which is user-dependent and time-
177 consuming and does not improve the discrepancies linked to the choice of target lesions (30,31). Aside from tumour bulk, metabolic activity of tumours
178 through functional imaging (e.g. positron emission tomography - PET) is highly predictive of response in lymphoma(32), lung cancer (33), and metastatic
179 melanoma(34). Other radioligands are utilised for response or recurrence detection, e.g. ^{18}F -fluoroestradiol (FES) in hormone-dependent breast cancer
180 (35) and ^{18}F - or ^{68}Ga Prostate-Specific Membrane Antigen (PSMA) ligands in prostate cancer(36). Additionally, radiolabelled ligands of various metabolites
181 and biologically active molecules can assess proliferation, hypoxia, angiogenesis, apoptosis and gene transfection(37).While parameters used for the
182 quantification and measurement of tumour metabolism by PET are generally based on semi-quantitative assessments, these can be made relatively
183 reproducible and harmonised throughout the world through standardised imaging protocols and dedicated initiatives promoted by the international
184 scientific societies(38,39), such as the accreditation program developed by the EANM Research Ltd. (EARL)(40,41).

185 Other alternate imaging biomarkers include perfusion and diffusion imaging. As tumours are commonly characterised by neo-angiogenesis, perfusion and
186 permeability derived from dynamic-contrast enhanced studies (e.g. with MR or CT) have been contenders for measuring early response(42), and
187 vascularity can be quantified using most imaging techniques, such as MRI, CT, ultrasound and PET. The utility of biomarkers of vascularity has been
188 demonstrated particularly where anti-angiogenic agents such as bevacizumab have been part of the therapeutic strategy(43). However, their quantitation,
189 which depends on measuring or estimating an arterial input function, is susceptible to large potential variations(44), and the reproducibility of such data
190 is often low, thus limiting their clinical utility(45). Another biomarker reflecting tissue cellularity, the apparent diffusion coefficient (ADC) from DW-MRI, has

191 proven a robust quantitative measure with good repeatability and reproducibility across vendor platforms(46), and has the potential to detect therapeutic
192 response earlier than size measurements. It is increasingly being introduced routinely into scanning protocols, as it does not require injection of an extrinsic
193 contrast agent and is simple and fast to acquire and analyse. Increasing automation with artificial intelligence (AI) systems may aid the translation of
194 biomarkers indicative of tumour characteristics other than bulk into routine clinical workflows. Unfortunately, tightened legal rules are slowing down the
195 process of their adoption (47).

196 Although historically dependent on imaging, response assessment for malignancies may now also include liquid biopsies (quantification of circulating
197 tumour cells or DNA [CTC, ctDNA]), as well as histological sampling. ctDNA shedding is influenced by the overall tumour burden (cells) and may thus
198 inform the use of imaging in relation to likely tumour size (48), because ctDNA estimations require less workflow and infrastructure than repeated
199 monitoring with imaging. Initial clinical evaluations showed that ctDNA detected response earlier than imaging-based assessment(49). The simplest
200 clinical implementation of ctDNA may be in postoperative monitoring of disease recurrence(50) but even here reproducibility and standardisation issues
201 remain limiting. In one study, ctDNA quantities based on measurement of some target genes (e.g. TERT) were, on average, more than two-fold higher
202 than those of other assays (e.g. ERV3) (145). In another, quantities of cell-free DNA for the different isolation methods for detection of EGFR variants in
203 NSCLC varied between medians of 1.6 ng/mL and 28.1 ng/mL (146). Moreover, concordance between tissue and plasma variant detection for leading
204 platforms has been shown to range from 70% to 90% (147). Thus, ctDNA extraction /isolation methods (145,146) may need to be standardised before
205 routine clinical use.

206 Finally, histopathology may also be a method for tumour response evaluation. However, serial histological sampling is not routinely used for response
207 assessment and has thus far shown agreement with imaging-based responses only in a few studies (51). Histopathological evaluation of response is
208 performed usually after neoadjuvant therapy, when the organ is surgically resected. Qualitative or semi-quantitative histopathological evaluation also
209 presents variable reproducibility according to organs, methods and published studies (52–55). Agreement between pathologists yielded kappa values
210 ranging from 0.21 for extent in prostate cancer (53), to 0.49 for multiple well-trained observers in cervical cancer (54), 0.64 for a 5-point tumour regression
211 grade in rectal cancer (51) and 0.83 for a central review in bladder cancer (52). As with macroscopic imaging, reader experience (55), and central review
212 (53) improve reproducibility.

213 *Statement #4: Alternative biomarkers for tumour response yield reproducibility generally comparable to RECIST. Each technique has its sources of*
214 *variability, and it is important to understand inherent variability and limitations of individual biomarkers. It is critical that imaging experts communicate their*
215 *level of confidence in any chosen biomarker.*

216 **What are common RECIST limitations?**

217 *Challenging organs: bone*

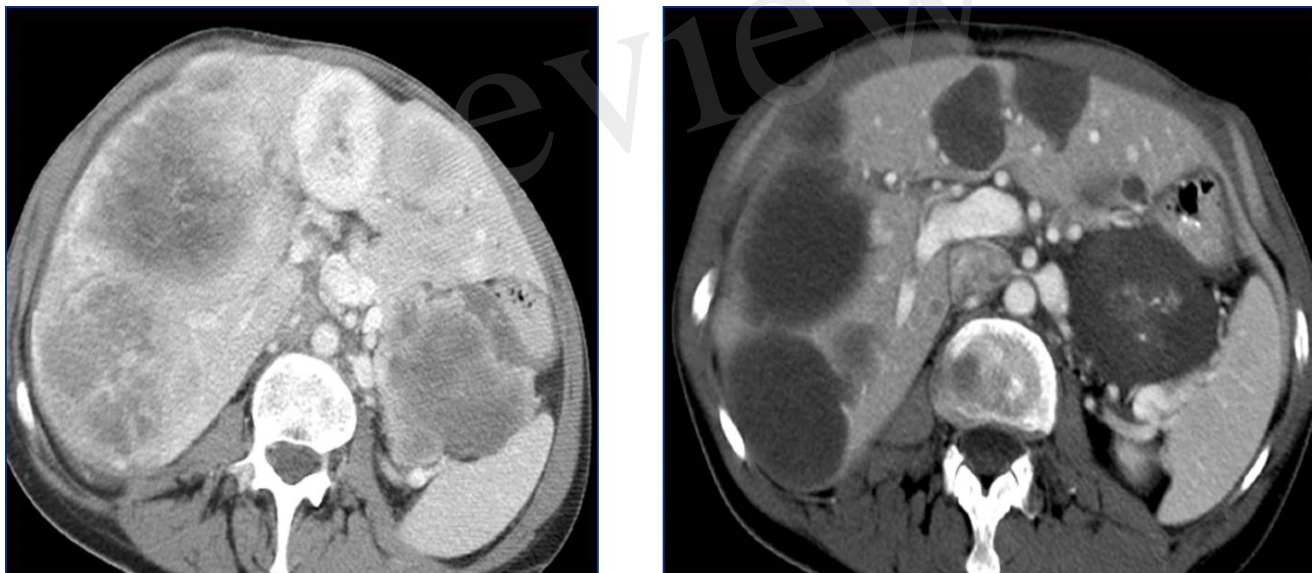
218 Bone metastases were considered unmeasurable in the initial RECIST initiative, because of the lack of sensitivity of existing techniques to bone marrow
219 infiltration(4). On CT it is the bone's osteolytic or osteosclerotic reaction to the presence of tumour, or its response to therapy (flare lesions) that is
220 visualised rather than the tumour itself (56,57). With the updated RECIST 1.1. version, bone metastases with soft tissue masses ≥ 10 mm are recognized
221 as measurable target lesions(1). Nevertheless, bone lesions without soft tissue involvement, whether lytic, mixed or sclerotic, remain unmeasurable by
222 RECIST. Since the early 1990s, bone marrow MRI has been shown to be superior to bone scintigraphy and CT for the assessment of bone metastatic
223 disease. Bone marrow replacement by neoplastic foci is detected and quantified on T1-weighted and fat-suppressed T2-weighted MRI sequences (58,59),
224 more recently complemented with diffusion-weighted imaging (DWI) sequences(60,61). However, to date, RECIST 1.1 has not validated quantitative bone
225 MRI for tumour response assessment. Positron Emission Tomography Response Criteria in Solid Tumours (PERCIST), introduced in 2009 (62,63),
226 enables response to be measured in ^{18}F -fluorodeoxyglucose (^{18}F -FDG) avid bone metastatic lesions based on their metabolic activity in the absence of
227 any obvious anatomic changes. Finally, PSMA-PET appears promising for identifying bone marrow invasion due to prostate cancer, regardless of the
228 impact on the bone mineral content (64,65) .

229 *Challenging diseases: GIST and mCRC*

230 As RECIST is not organ-specific, it might not capture the key parameters that are associated with survival outcomes in certain cancer types, and under
231 certain types of treatment. In gastrointestinal malignancies, the hepatic tumour burden and its response commonly outperform other sites of metastatic
232 disease for survival prediction. A study in metastatic colorectal cancer (mCRC) showed that the depth and uniformity of response in liver metastases
233 represented a highly useful and clinically relevant indicator for therapy monitoring(66). Organ-specific response patterns may also occur under
234 immunotherapy possibly due to varying immune microenvironments in organs or the lymphatic system(67–69). Thus, choice of target lesions would largely
235 impact the response observed according to the organ, as well as the predictive ability of RECIST. In this case also, reader experience and knowledge of
236 the disease is crucial for proper target lesion selection.

237 Response to therapy in patients with advanced GIST was drastically improved by the introduction of imatinib, a tyrosine-kinase inhibitor. Imatinib treatment
238 as been shown to induce necrosis with a marked decrease in vascularity of GIST lesions, resulting in a decrease in CT density often before any significant
239 decrease in size is seen, thus leading to underestimation of the initial tumour response (70,71) (Figure 2). A paradoxical increase in volume is occasionally
240 observed, simulating progression (72). Choi et al therefore proposed adapted criteria for GIST, combining changes in tumour density on contrast-enhanced
241 CT expressed in Hounsfield units (HU) and/or size to determine tumour response(71) : PR is defined as a decrease of $\geq 10\%$ in the SOD or a decrease of
242 $\geq 15\%$ in tumour density of target lesions, whereas PD is defined as a $\geq 10\%$ increase in size and not meeting the PR criteria by tumour density. PD may

243 also occur if new intra-tumoural nodules are present or existing intra-tumoural nodules show an increase in size, factors which are not catered for in
244 RECIST. In patients treated with imatinib, Choi criteria showed a significantly better correlation with survival rates than RECIST(73).



245
246 *Figure 2: Response unrelated to tumour size in a 66 yo patient treated with imatinib for a gastrointestinal stromal tumour (GIST). Compared to the baseline*
247 *image (left), after treatment (right) the tumour shows a dramatic decrease in density rather than in size.*

248

249 *Challenging treatments: focal therapies*

250 Treatment of tumour lesions with ablative therapies, such as radiofrequency ablation, microwave ablation or cryoablation, results in a larger defect than
251 the original lesion and such treated lesions are not considered measurable unless there is progression at this site (1), such as the development of a new
252 measurable nodule within the ablation defect. Distinguishing normal post-ablation changes from residual disease and recurrence can be challenging (74).
253 Intravascular therapies are also a challenge for the use of RECIST. Trans-arterial radioembolization (TARE) induces inflammatory changes with a
254 generally delayed morphologic response (74). A reduction of ^{18}F -FDG uptake on early PET-CT has been found to be helpful in predicting further outcome
255 of these patients(75). As a consequence, both TARE and intra-arterial therapies such as trans-arterial chemoembolization (TACE) in hepatocarcinoma

256 require modified RECIST (mRECIST) criteria derived from arterial and portal venous enhancement phases of CT or MRI(76), and which take into account
257 both lesion size and vascularity.

258 High-intensity focused ultrasound (HIFU), under the guidance of ultrasound or MRI, has also been used as a non-invasive technique for tissue ablation
259 in prostate cancer and more recently in recurrent gynecological malignancy (77). The use of HIFU for hepatic tumour lesions is still in the exploratory
260 stage. As for other ablative therapies and for similar reasons (78), RECIST 1.1 appears to be unsuitable for local response evaluation following HIFU
261 applied to liver lesions.

262 Finally, tumour lesions in a previously irradiated area (via CyberKnife, stereotactic radiotherapy or traditional fractionated radiation therapy) are not
263 considered measurable(1) and must be excluded from RECIST evaluation due to the inflammatory or fibrotic changes that may be observed, thus making
264 evaluation of size unreliable .

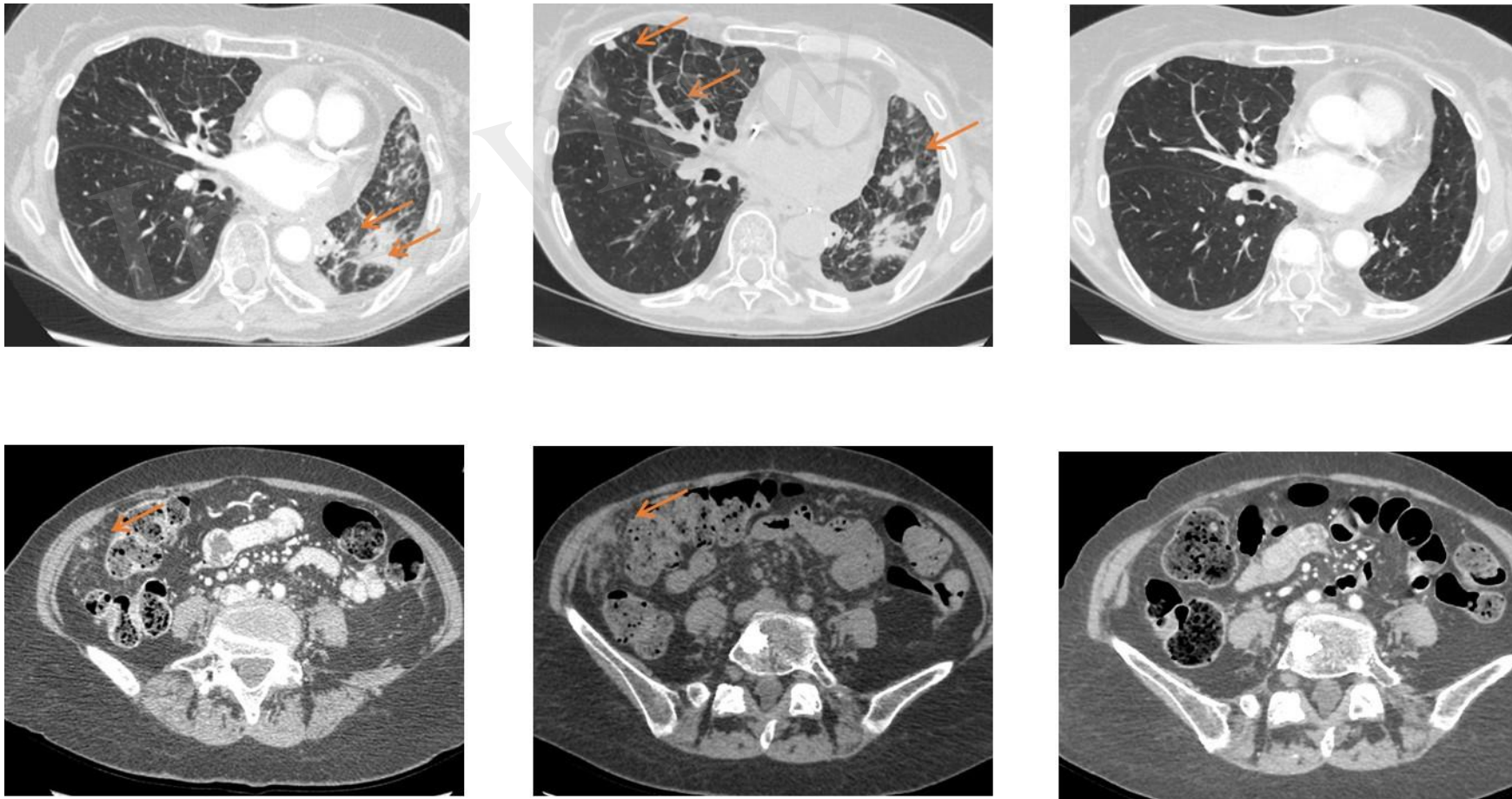
265 *Statement #5: There are several scenarios in which RECIST criteria fail to evaluate treatment-induced changes correctly. Informed appreciation that*
266 *RECIST criteria are not applicable to all tumour sites and situations is thus crucial for proper interpretation and again dependent on reader experience.*

267

268 **When is RECIST response assessment misleading?**

269 *Pseudo-progression*

270 During immunotherapy, RECIST may describe progression that can be misleading and is thus classified as “pseudo-progression”. In fact, in around 5 to
271 10% of patients with metastatic disease treated with check-point inhibitors, an initial increase of tumour burden has been observed, followed by actual
272 response or long-term stabilisation of disease (79–81). This phenomenon relates to the mechanism of action of immunotherapy, which stimulates the
273 immune response and initially induces inflammation and tumour swelling, thus delaying visible tumour shrinkage. For this reason, adaptations of RECIST
274 criteria for assessing treatment response to immunotherapy (iRECIST) have been developed. The first ascertainment of progression by iRECIST is
275 considered as “immune unconfirmed progressive disease”(or iUPD), and requires, if possible, a subsequent evaluation 4 to 8 weeks later in order to
276 confirm true progression(82) (Figure 3).



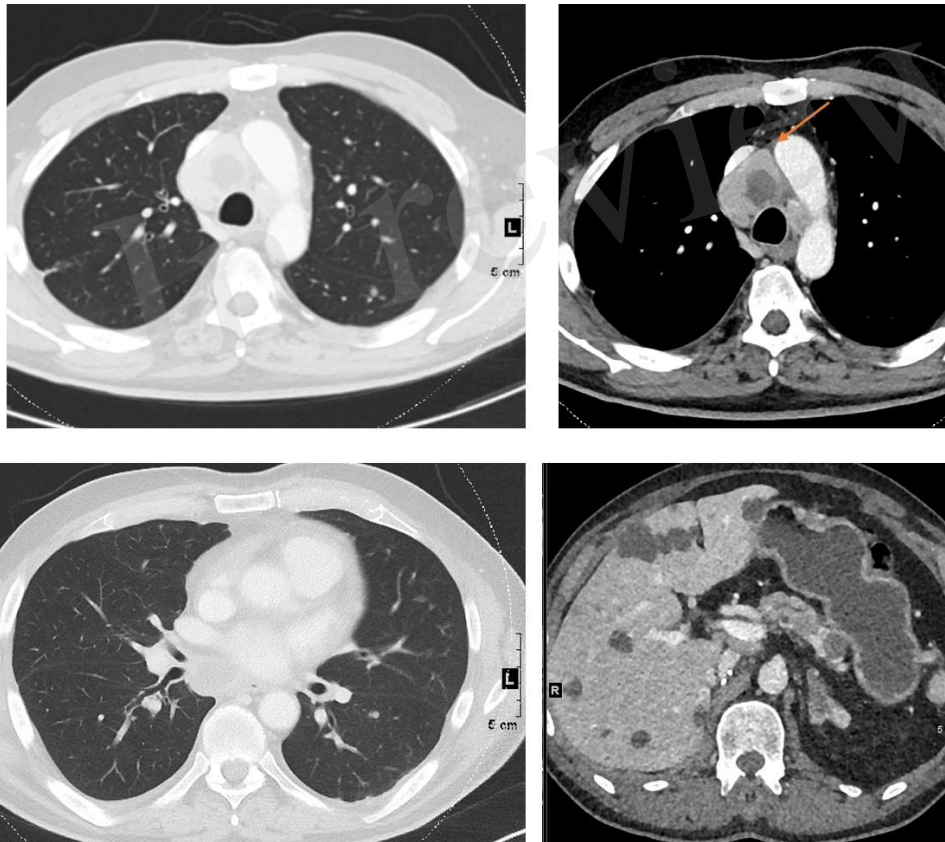
277

278 Figure 3: Pseudoprogession on immunotherapy in a 56 yo patient with metastatic non small cell lung cancer . The baseline image (left) shows lung and
 279 peritoneal nodules (arrows). After 4 wks of antiPDL1 therapy (middle), CT shows an increase in previous lesions and the appearance of new lung nodules.
 280 Disease was considered immune unconfirmed progressive disease. Siix weeks later (right) a dramatic response in all previous lesions was seen classifying
 281 the patiets as a complete responder and endorsing an ealier diagnosis of pseudoprogession.

282 *Mixed response/progression*

283 In some patients, the tumour bulk does not respond homogeneously, with some lesions increasing and others decreasing. Mixed or heterogeneous
284 response is defined as an increase in size of some tumour lesions and decrease of others in the same patient during treatment. This lesion-specific
285 response has been attributed to the emergence of drug-resistant clones and indicates that tumour heterogeneity is likely causing treatment failure(83,84).
286 Mixed response has the same incidence in patients treated with targeted cancer agents and those undergoing chemotherapy alone or even combined
287 with targeted agents(12,31).

288 Since RECIST records overall patient response rather than individual lesion response, the choice of target lesions critically affects the objective
289 assessment of overall patient response in patients with mixed response in individual lesions (Figure 4)(12). As lesions escaping treatment control will
290 weigh negatively on patient prognosis (85), their presence should be annotated in order to offer the best alternative treatment for the patient.



291

292 Figure 4: Mixed response to treatment in the same patient illustrated in Figure 1. Eight weeks after targeted therapy lung, adrenal and pancreatic
293 metastases decreased, whereas one mediastinal lymph node (top right, arrow) increased

294 Lesion cavitation, necrosis and residual non-viable masses represent other forms of response than decrease in size and may complicate RECIST
295 assessment(86). Tumour necrosis with cavitation is present in approximately 14-24% of NSCLC patients undergoing anti-angiogenic drug therapy(87-
296 89). When cavitation is present, lesion size may not change significantly and RECIST may therefore under-estimate the effect of therapy. Conversely,
297 cavitation also risks missing progression if there is tumour regrowth inside the cavity. While alternative criteria have been proposed in such cases, e.g.
298 subtracting the longest cavitation diameter from the largest lesion diameter (such as Crabb criteria)(88), these are not commonly used.

299 When residual tissue is present after therapy, evaluation with RECIST criteria is subject to pitfalls. First, an asymmetric shrinkage of the tumour may result
300 in a similar longest diameter and consequent stable disease (SD) rating not reflecting the real response to treatment (Figure 2). Second, it may be difficult
301 to distinguish between viable tumour and fibrosis. In such cases, best response assessment, an important endpoint in phase 2 studies (partial vs. complete
302 response; PR vs. CR) may be affected(88). According to RECIST guidelines, in equivocal cases, residual lesions should be evaluated by either biopsy or
303 PET(-CT) (Figure 2). This may well then allow upgrading PR to CR. However, false positive PET findings are not uncommon (90). Alternatively, other
304 advanced imaging tests, such as DWI-MRI or perfusion imaging (e.g. from MR or CT) could be used.

305 *Statement #6: Some patterns of response/progression cannot be correctly documented by RECIST. These require specialist reader experience and*
306 *communication with oncologists to determine appropriate evaluation approaches and/or therapeutic options.*

307

308 **Should (could) RECIST be automated?**

309 The core assumption of RECIST is that a single diameter on the cross-sectional imaging slice presenting the largest cross-section of a given lesion (or
310 sum thereof) is a surrogate for tumour burden. This assumes that lesions are grossly spherical and that their size represents their overall activity. To
311 streamline the determination of this single diameter and make it less subject to possible human-induced variability, semi- or fully-automated 2D or even
312 3D segmentation techniques can be applied to target lesions, which can also be semi- or fully-automatically tracked between scans acquired at different
313 time points(91–96). The 2D or 3D mask resulting from the segmentation process then readily permits the automated and accurate extraction of the largest
314 diameter from the segmented lesion. With 3D segmentation, the full volume of a target lesion can be provided alongside an automatically extracted largest
315 diameter, which may not be oriented in the 2D plane of the source images in a broader RECIST interpretation, together with any other geometric metric
316 of relevance. Using the largest 3D diameter would allow RECIST to be used beyond 2D constraints, and can account for non-orthogonal motion of target
317 lesions between scans at different time-points. While segmentation and tracking can now plausibly be fully automated, especially with newer approaches
318 using machine learning, and such capabilities are already implemented in several commercially available clinical systems, some challenges remain with
319 key RECIST operations, such as the proper selection of target lesions and dealing with new or disappeared lesions. These are currently still best addressed
320 or verified with a human (e.g. a radiologist) in the loop (21,97).

321 *Statement #7: Though automation is desirable to streamline the process and potentially reduce the variability of results within the RECIST paradigm,*
322 *remaining technical challenges must be overcome to ensure proper repeatability, and human adjudication is still required.*

323

324 **RECIST in novel drug development**

325 RECIST measurements play a pivotal role in the development of novel oncological drugs(98). In most registered randomised controlled trials (RCTs),
326 studies are powered to meet primary endpoints such as OS/PFS, which determines the number of patients recruited. A study of RCTs between 2006 and
327 2015 looking for evidence of clinical efficacy of novel oncology drugs in order to gain US Food and Drug Administration (FDA) approvals had PFS as
328 primary endpoint in 28 out of 42 RCTs (66%), and OS in 14 (33%). In 2012, 12 novel anticancer drugs were approved by the FDA; only three drugs
329 showed improvement of overall survival(99). Similarly, a study of drugs approved by the European Medicine Agency (EMA) between 2009 and 2013 also
330 showed that only 18 of 68 (26%) novel drug uses were supported by OS data, whereas PFS was used in 31 (46%)(100). In the vast majority of trials, PFS
331 is determined using the RECIST1.1 framework, or iRECIST for immune-oncology trials. It is acknowledged however, that in some disease types other
332 criteria are used: e.g. Lugano criteria for ¹⁸F-FDG PET/CT or RECIL in lymphoma (101,102)and RANO criteria for brain tumours (103,104). The fact that
333 PFS can predict OS outcome in large patients cohorts with commonly occurring cancers, reinforces the use of RECIST criteria in clinical trials(105).
334 Moreover, rapid progress in drug development will make the reliance on OS as endpoint for novel drugs in oncology increasingly challenging because
335 treatment options on progression on trial, including in-trial cross-over, are increasing.

336 *Statement #8: Although the RECIST framework might not be perfect, the scientific basis for the anticancer drugs that have been approved using a*
337 *RECIST-based surrogate endpoint remains valid.*

338

339 **RECIST: only as good as its users?**

340 RECIST criteria were developed for clinical trials and thresholds chosen to produce a comparable classification of patients, taking into account tumour
341 measurement variability. These criteria are widely used in clinical trials and accepted by regulatory agencies. Despite some limitations, the scientific basis
342 for the anticancer drugs that have been approved using a RECIST-based surrogate endpoint remains valid. Reader experience, choice of target lesions
343 and detection of new lesions impact RECIST reproducibility, which necessitates adequate training of radiologists using these criteria. Automation is not
344 currently sufficiently reliable to replace human experience. Unfortunately, some organ-, disease- or drug-specific patterns of response/progression cannot
345 be correctly documented by RECIST.

346 This expert statement concludes that RECIST remains a tool for radiologists that needs to be used with discrimination and good understanding of its purpose
347 and limitations. Training of radiologists is essential to improve its application and reproducibility. RECIST conclusions should not go against common (or
348 informed) sense. Furthermore, RECIST criteria have the advantage of simplicity, availability, cost-effectiveness, and intuitiveness. Overall, therefore, RECIST

349 provides a common language between oncologists and imaging experts (e.g. radiologists), provided there is full understanding of how measurements are
350 made, what they represent, and their inherent limitations.

In review

351 **References**

- 352 1. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised
353 RECIST guideline (version 1.1). *European Journal of Cancer*. 2009 Jan 1;45(2):228–47.
- 354 2. Moertel CG, Hanley JA. The effect of measuring error on the results of therapeutic trials in advanced cancer. *Cancer*. 1976 Jul;38(1):388–
355 94.
- 356 3. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer*. 1981;47(1):207–14.
- 357 4. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New Guidelines to Evaluate the Response to Treatment
358 in Solid Tumors. *JNCI: Journal of the National Cancer Institute*. 2000 Feb 2;92(3):205–16.
- 359 5. James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldal A, et al. Measuring Response in Solid Tumors: Unidimensional Versus
360 Bidimensional Measurement. *JNCI: Journal of the National Cancer Institute*. 1999 Mar 17;91(6):523–8.
- 361 6. Bogaerts J, Ford R, Sargent D, Schwartz LH, Rubinstein L, Lacombe D, et al. Individual patient data analysis to assess modifications to the
362 RECIST criteria. *European Journal of Cancer*. 2009 Jan 1;45(2):248–60.
- 363 7. Moskowitz CS, Jia X, Schwartz LH, Gönen M. A simulation study to evaluate the impact of the number of lesions measured on response
364 assessment. *European Journal of Cancer*. 2009 Jan 1;45(2):300–10.
- 365 8. Booth CM, Eisenhauer EA. Progression-Free Survival: Meaningful or Simply Measurable? *JCO*. 2012 Feb 27;30(10):1030–3.
- 366 9. Suzuki C, Blomqvist L, Sundin A, Jacobsson H, Byström P, Berglund Å, et al. The initial change in tumor size predicts response and
367 survival in patients with metastatic colorectal cancer treated with combination chemotherapy. *Annals of Oncology*. 2012 Apr 1;23(4):948–54.
- 368 10. Jain RK, Lee JJ, Ng C, Hong D, Gong J, Naing A, et al. Change in tumor size by RECIST correlates linearly with overall survival in phase I
369 oncology studies. *J Clin Oncol*. 2012 Jul 20;30(21):2684–90.
- 370 11. Grünwald V, McKay RR, Krajewski KM, Kalanovic D, Lin X, Perkins JJ, et al. Depth of Remission is a Prognostic Factor for Survival in
371 Patients with Metastatic Renal Cell Carcinoma. *European Urology*. 2015 May 1;67(5):952–8.
- 372 12. Litière S, Isaac G, De Vries EGE, Bogaerts J, Chen A, Dancey J, et al. RECIST 1.1 for Response Evaluation Apply Not Only to Chemotherapy-
373 Treated Patients But Also to Targeted Cancer Agents: A Pooled Database Analysis. *J Clin Oncol*. 2019 May 1;37(13):1102–10.
- 374 13. Litière S, de Vries EGE, Seymour L, Sargent D, Shankar L, Bogaerts J. The components of progression as explanatory variables for overall
375 survival in the Response Evaluation Criteria in Solid Tumours 1.1 database. *European Journal of Cancer*. 2014 Jul;50(10):1847–53.
- 376 14. An M-W, Mandrekar SJ, Branda ME, Hillman SL, Adjei AA, Pitot HC, et al. Comparison of continuous versus categorical tumor
377 measurement-based metrics to predict overall survival in cancer treatment trials. *Clin Cancer Res*. 2011 Oct 15;17(20):6592–9.
- 378 15. An M-W, Dong X, Meyers J, Han Y, Grothey A, Bogaerts J, et al. Evaluating Continuous Tumor Measurement-Based Metrics as Phase II
379 Endpoints for Predicting Overall Survival. *J Natl Cancer Inst*. 2015 Nov;107(11):djv239.

- 380 16. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, et al. Metrology Standards for Quantitative Imaging
381 Biomarkers. *Radiology*. 2015 Dec 1;277(3):813–25.
- 382 17. Yoon SH, Kim KW, Goo JM, Kim D-W, Hahn S. Observer variability in RECIST-based tumour burden measurements: a meta-analysis. *Eur J*
383 *Cancer*. 2016 Jan;53:5–15.
- 384 18. Kuhl CK, Alparslan Y, Schmoe J, Sequeira B, Keulers A, Brümmendorf TH, et al. Validity of RECIST Version 1.1 for Response Assessment
385 in Metastatic Cancer: A Prospective, Multireader Study. *Radiology*. 2019 Feb;290(2):349–56.
- 386 19. Ford R, Schwartz L, Dancey J, Dodd LE, Eisenhauer EA, Gwyther S, et al. Lessons learned from independent central review. *Eur J Cancer*.
387 2009 Jan;45(2):268–74.
- 388 20. Bellomi M, De Piano F, Ancona E, Lodigiani AF, Curigliano G, Raimondi S, et al. Evaluation of inter-observer variability according to
389 RECIST 1.1 and its influence on response classification in CT measurement of liver metastases. *Eur J Radiol*. 2017 Oct;95:96–101.
- 390 21. Beaumont H, Evans TL, Klifa C, Guerhazi A, Hong SR, Chadja M, et al. Discrepancies of assessments in a RECIST 1.1 phase II clinical trial
391 - association between adjudication rate and variability in images and tumors selection. *Cancer Imaging*. 2018 Dec 11;18(1):50.
- 392 22. Karmakar A, Kumtakar A, Sehgal H, Kumar S, Kalyanpur A. Interobserver Variation in Response Evaluation Criteria in Solid Tumors 1.1.
393 *Acad Radiol*. 2019 Apr;26(4):489–501.
- 394 23. Suzuki C, Torkzad MR, Jacobsson H, Aström G, Sundin A, Hatschek T, et al. Interobserver and intraobserver variability in the response
395 evaluation of cancer therapy according to RECIST and WHO-criteria. *Acta Oncol*. 2010 May;49(4):509–14.
- 396 24. Oxnard GR, Zhao B, Sima CS, Ginsberg MS, James LP, Lefkowitz RA, et al. Variability of lung tumor measurements on repeat computed
397 tomography scans taken within 15 minutes. *J Clin Oncol*. 2011 Aug 10;29(23):3114–9.
- 398 25. McErlean A, Panicek DM, Zabor EC, Moskowitz CS, Bitar R, Motzer RJ, et al. Intra- and interobserver variability in CT measurements in
399 oncology. *Radiology*. 2013 Nov;269(2):451–9.
- 400 26. Hayes SA, Pietanza MC, O'Driscoll D, Zheng J, Moskowitz CS, Kris MG, et al. Comparison of CT volumetric measurement with RECIST
401 response in patients with lung cancer. *Eur J Radiol*. 2016 Mar;85(3):524–33.
- 402 27. Lee JH, Lee HY, Ahn M-J, Park K, Ahn JS, Sun J-M, et al. Volume-based growth tumor kinetics as a prognostic biomarker for patients with
403 EGFR mutant lung adenocarcinoma undergoing EGFR tyrosine kinase inhibitor therapy: a case control study. *Cancer Imaging*. 2016 Mar 16;16:5.
- 404 28. deSouza NM, Soutter WP, Rustin G, Mahon MM, Jones B, Dina R, et al. Use of neoadjuvant chemotherapy prior to radical hysterectomy
405 in cervical cancer: monitoring tumour shrinkage and molecular profile on magnetic resonance and assessment of 3-year outcome. *Br J Cancer*.
406 2004 Jun 14;90(12):2326–31.
- 407 29. Fenerty KE, Folio LR, Patronas NJ, Marté JL, Gulley JL, Heery CR. Predicting clinical outcomes in chordoma patients receiving
408 immunotherapy: a comparison between volumetric segmentation and RECIST. *BMC Cancer*. 2016 Aug 23;16(1):672.

- 409 30. Keil S, Barabasch A, Dirrichs T, Bruners P, Hansen NL, Bieling HB, et al. Target lesion selection: an important factor causing variability of
410 response classification in the Response Evaluation Criteria for Solid Tumors 1.1. *Invest Radiol*. 2014 Aug;49(8):509–17.
- 411 31. Zimmermann M, Kuhl CK, Engelke H, Bettermann G, Keil S. Factors That Drive Heterogeneity of Response-to-Treatment of Different
412 Metastatic Deposits Within the Same Patients as Measured by RECIST 1.1 Analyses. *Acad Radiol*. 2020 Jun 29;
- 413 32. Schmitz C, Hüttmann A, Müller SP, Hanoun M, Boellaard R, Brinkmann M, et al. Dynamic risk assessment based on positron emission
414 tomography scanning in diffuse large B-cell lymphoma: Post-hoc analysis from the PETAL trial. *Eur J Cancer*. 2020 Jan;124:25–36.
- 415 33. Mac Manus MP, Hicks RJ, Matthews JP, Wirth A, Rischin D, Ball DL. Metabolic (FDG-PET) response after radical
416 radiotherapy/chemoradiotherapy for non-small cell lung cancer correlates with patterns of failure. *Lung Cancer*. 2005 Jul;49(1):95–108.
- 417 34. Sachpekidis C, Anwar H, Winkler J, Kopp-Schneider A, Larribere L, Haberkorn U, et al. The role of interim 18F-FDG PET/CT in prediction
418 of response to ipilimumab treatment in metastatic melanoma. *Eur J Nucl Med Mol Imaging*. 2018 Jul;45(8):1289–96.
- 419 35. Evangelista L, Guarneri V, Conte PF. 18F-Fluoroestradiol Positron Emission Tomography in Breast Cancer Patients: Systematic Review of
420 the Literature & Meta-Analysis. *Curr Radiopharm*. 2016;9(3):244–57.
- 421 36. Tan N, Oyoyo U, Bavadian N, Ferguson N, Mukkamala A, Calais J, et al. PSMA-targeted Radiotracers versus 18F Fluciclovine for the
422 Detection of Prostate Cancer Biochemical Recurrence after Definitive Therapy: A Systematic Review and Meta-Analysis. *Radiology*. 2020
423 Jul;296(1):44–55.
- 424 37. Wester H-J. Nuclear imaging probes: from bench to bedside. *Clin Cancer Res*. 2007 Jun 15;13(12):3470–81.
- 425 38. Boellaard R, O’Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, et al. FDG PET and PET/CT: EANM procedure
426 guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging*. 2010 Jan;37(1):181–200.
- 427 39. Graham MM, Wahl RL, Hoffman JM, Yap JT, Sunderland JJ, Boellaard R, et al. Summary of the UPICT Protocol for 18F-FDG PET/CT
428 Imaging in Oncology Clinical Trials. *J Nucl Med*. 2015 Jun;56(6):955–61.
- 429 40. Makris NE, Huisman MC, Kinahan PE, Lammertsma AA, Boellaard R. Evaluation of strategies towards harmonization of FDG PET/CT
430 studies in multicentre trials: comparison of scanner validation phantoms and data analysis procedures. *Eur J Nucl Med Mol Imaging*. 2013
431 Oct;40(10):1507–15.
- 432 41. Aide N, Lasnon C, Veit-Haibach P, Sera T, Sattler B, Boellaard R. EANM/EARL harmonization strategies in PET quantification: from daily
433 practice to multicentre oncological studies. *Eur J Nucl Med Mol Imaging*. 2017 Aug;44(Suppl 1):17–31.
- 434 42. Jun W, Cong W, Xianxin X, Daqing J. Meta-Analysis of Quantitative Dynamic Contrast-Enhanced MRI for the Assessment of Neoadjuvant
435 Chemotherapy in Breast Cancer. *Am Surg*. 2019 Jun 1;85(6):645–53.
- 436 43. Hirashima Y, Yamada Y, Tateishi U, Kato K, Miyake M, Horita Y, et al. Pharmacokinetic parameters from 3-Tesla DCE-MRI as surrogate
437 biomarkers of antitumor effects of bevacizumab plus FOLFIRI in colorectal cancer with liver metastasis. *Int J Cancer*. 2012 May
438 15;130(10):2359–65.

- 439 44. Rata M, Collins DJ, Darcy J, Messiou C, Tunariu N, Desouza N, et al. Assessment of repeatability and treatment response in early phase
440 clinical trials using DCE-MRI: comparison of parametric analysis using MR- and CT-derived arterial input functions. *Eur Radiol*. 2016
441 Jul;26(7):1991–8.
- 442 45. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, et al. Quantitative imaging biomarkers
443 alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. *J Magn Reson*
444 *Imaging*. 2019 Jun;49(7):e101–21.
- 445 46. Winfield JM, Tunariu N, Rata M, Miyazaki K, Jerome NP, Germuska M, et al. Extracranial Soft-Tissue Tumors: Repeatability of Apparent
446 Diffusion Coefficient Estimates from Diffusion-weighted MR Imaging. *Radiology*. 2017 Jul;284(1):88–99.
- 447 47. Health C for D and R. Artificial Intelligence and Machine Learning in Software as a Medical Device. FDA [Internet]. 2021 Nov 1 [cited
448 2021 Aug 26]; Available from: [https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-](https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device)
449 [learning-software-medical-device](https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device)
- 450 48. Avanzini S, Kurtz DM, Chabon JJ, Moding EJ, Hori SS, Gambhir SS, et al. A mathematical model of ctDNA shedding predicts tumor
451 detection size. *Sci Adv*. 2020 Dec;6(50).
- 452 49. Nabet BY, Esfahani MS, Moding EJ, Hamilton EG, Chabon JJ, Rizvi H, et al. Noninvasive Early Identification of Therapeutic Benefit from
453 Immune Checkpoint Inhibition. *Cell*. 2020 Oct 15;183(2):363–376.e13.
- 454 50. Dasari A, Morris VK, Allegra CJ, Atreya C, Benson AB, Boland P, et al. ctDNA applications and integration in colorectal cancer: an NCI
455 Colon and Rectal-Anal Task Forces whitepaper. *Nat Rev Clin Oncol*. 2020 Dec;17(12):757–70.
- 456 51. Tapia C, Aung PP, Roy-Chowdhuri S, Xu M, Ouyang F, Alshawa A, et al. Decrease in tumor content assessed in biopsies is associated with
457 improved treatment outcome response to pembrolizumab in patients with rare tumors. *J Immunother Cancer*. 2020 Apr;8(1).
- 458 52. Ryan R, Gibbons D, Hyland JMP, Treanor D, White A, Mulcahy HE, et al. Pathological response following long-course neoadjuvant
459 chemoradiotherapy for locally advanced rectal cancer. *Histopathology*. 2005 Aug;47(2):141–6.
- 460 53. Voskuilen CS, Oo HZ, Genitsch V, Smit LA, Vidal A, Meneses M, et al. Multicenter Validation of Histopathologic Tumor Regression Grade
461 After Neoadjuvant Chemotherapy in Muscle-invasive Bladder Carcinoma. *Am J Surg Pathol*. 2019 Dec;43(12):1600–10.
- 462 54. Jaraj SJ, Camparo P, Boyle H, Germain F, Nilsson B, Petersson F, et al. Intra- and interobserver reproducibility of interpretation of
463 immunohistochemical stains of prostate cancer. *Virchows Arch*. 2009 Oct;455(4):375–81.
- 464 55. Stoler MH, Schiffman M, Atypical Squamous Cells of Undetermined Significance-Low-grade Squamous Intraepithelial Lesion Triage Study
465 (ALTS) Group. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage
466 Study. *JAMA*. 2001 Mar 21;285(11):1500–5.
- 467 56. Stattaus J, Hahn S, Gauler T, Eberhardt W, Mueller SP, Forsting M, et al. Osteoblastic response as a healing reaction to chemotherapy
468 mimicking progressive disease in patients with small cell lung cancer. *Eur Radiol*. 2009 Jan;19(1):193–200.

- 469 57. Messiou C, Cook G, Reid AHM, Attard G, Dearnaley D, de Bono JS, et al. The CT flare response of metastatic bone disease in prostate
470 cancer. *Acta Radiol.* 2011 Jun 1;52(5):557–61.
- 471 58. Ciray I, Lindman H, Aström KG, Bergh J, Ahlström KH. Early response of breast cancer bone metastases to chemotherapy evaluated with
472 MR imaging. *Acta Radiol.* 2001 Mar;42(2):198–206.
- 473 59. Tombal B, Rezazadeh A, Therasse P, Van Cangh PJ, Vande Berg B, Lecouvet FE. Magnetic resonance imaging of the axial skeleton enables
474 objective measurement of tumor response on prostate cancer bone metastases. *Prostate.* 2005 Oct 1;65(2):178–87.
- 475 60. Padhani AR, Lecouvet FE, Tunariu N, Koh D-M, De Keyzer F, Collins DJ, et al. METastasis Reporting and Data System for Prostate Cancer:
476 Practical Guidelines for Acquisition, Interpretation, and Reporting of Whole-body Magnetic Resonance Imaging-based Evaluations of
477 Multiorgan Involvement in Advanced Prostate Cancer. *Eur Urol.* 2017 Jan;71(1):81–92.
- 478 61. Van Nieuwenhove S, Van Damme J, Padhani AR, Vandecaveye V, Tombal B, Wuts J, et al. Whole-body magnetic resonance imaging for
479 prostate cancer assessment: Current status and future directions. *J Magn Reson Imaging.* 2020 Dec 31;
- 480 62. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J*
481 *Nucl Med.* 2009 May;50 Suppl 1:122S-50S.
- 482 63. O JH, Lodge MA, Wahl RL. Practical PERCIST: A Simplified Guide to PET Response Criteria in Solid Tumors 1.0. *Radiology.* 2016
483 Aug;280(2):576–84.
- 484 64. Schmidkonz C, Cordes M, Schmidt D, Bäuerle T, Goetz TI, Beck M, et al. 68Ga-PSMA-11 PET/CT-derived metabolic parameters for
485 determination of whole-body tumor burden and treatment response in prostate cancer. *Eur J Nucl Med Mol Imaging.* 2018 Oct;45(11):1862–
486 72.
- 487 65. Schmidkonz C, Cordes M, Goetz TI, Prante O, Kuwert T, Ritt P, et al. 68Ga-PSMA-11 PET/CT derived quantitative volumetric tumor
488 parameters for classification and evaluation of therapeutic response of bone metastases in prostate cancer patients. *Ann Nucl Med.* 2019
489 Oct;33(10):766–75.
- 490 66. Zhou J, Li Q, Cao Y. Spatiotemporal Heterogeneity across Metastases and Organ-Specific Response Informs Drug Efficacy and Patient
491 Survival in Colorectal Cancer. *Cancer Res.* 2021 May 1;81(9):2522–33.
- 492 67. Pires da Silva I, Lo S, Quek C, Gonzalez M, Carlino MS, Long GV, et al. Site-specific response patterns, pseudoprogression, and acquired
493 resistance in patients with melanoma treated with ipilimumab combined with anti-PD-1 therapy. *Cancer.* 2020 Jan 1;126(1):86–97.
- 494 68. Osorio JC, Arbour KC, Le DT, Durham JN, Plodkowski AJ, Halpenny DF, et al. Lesion-Level Response Dynamics to Programmed Cell Death
495 Protein (PD-1) Blockade. *J Clin Oncol.* 2019 Dec 20;37(36):3546–55.
- 496 69. Schmid S, Diem S, Li Q, Krapf M, Flatz L, Leschka S, et al. Organ-specific response to nivolumab in patients with non-small cell lung
497 cancer (NSCLC). *Cancer Immunol Immunother.* 2018 Dec;67(12):1825–32.

- 498 70. Choi H, Charnsangavej C, de Castro Faria S, Tamm EP, Benjamin RS, Johnson MM, et al. CT evaluation of the response of gastrointestinal
499 stromal tumors after imatinib mesylate treatment: a quantitative analysis correlated with FDG PET findings. *AJR Am J Roentgenol.* 2004
500 Dec;183(6):1619–28.
- 501 71. Choi H, Charnsangavej C, Faria SC, Macapinlac HA, Burgess MA, Patel SR, et al. Correlation of computed tomography and positron
502 emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal
503 of new computed tomography response criteria. *J Clin Oncol.* 2007 May 1;25(13):1753–9.
- 504 72. Fournier L, Ammari S, Thiam R, Cuénod C-A. Imaging criteria for assessing tumour response: RECIST, mRECIST, Cheson. *Diagn Interv*
505 *Imaging.* 2014 Aug;95(7–8):689–703.
- 506 73. Benjamin RS, Choi H, Macapinlac HA, Burgess MA, Patel SR, Chen LL, et al. We should desist using RECIST, at least in GIST. *J Clin Oncol.*
507 2007 May 1;25(13):1760–4.
- 508 74. Maas M, Beets-Tan R, Gaubert J-Y, Gomez Munoz F, Habert P, Klompenhouwer LG, et al. Follow-up after radiological intervention in
509 oncology: ECIO-ESOI evidence and consensus-based recommendations for clinical practice. *Insights Imaging.* 2020 Jul 16;11(1):83.
- 510 75. Sabet A, Meyer C, Aouf A, Sabet A, Ghamari S, Pieper CC, et al. Early post-treatment FDG PET predicts survival after 90Y microsphere
511 radioembolization in liver-dominant metastatic colorectal cancer. *Eur J Nucl Med Mol Imaging.* 2015 Mar;42(3):370–6.
- 512 76. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. *Semin Liver Dis.* 2010 Feb;30(1):52–60.
- 513 77. Imseeh G, Giles SL, Taylor A, Brown MRD, Rivens I, Gordon-Williams R, et al. Feasibility of palliating recurrent gynecological tumors with
514 MRGHIFU: comparison of symptom, quality-of-life, and imaging response in intra and extra-pelvic disease. *Int J Hyperthermia.* 2021;38(1):623–
515 32.
- 516 78. Ji Y, Zhu J, Zhu L, Zhu Y, Zhao H. High-Intensity Focused Ultrasound Ablation for Unresectable Primary and Metastatic Liver Cancer: Real-
517 World Research in a Chinese Tertiary Center With 275 Cases. *Front Oncol.* 2020;10:519164.
- 518 79. Wolchok JD, Hoos A, O’Day S, Weber JS, Hamid O, Lebbe C, et al. Guidelines for the Evaluation of Immune Therapy Activity in Solid
519 Tumors: Immune-Related Response Criteria. *Clinical Cancer Research.* 2009 Dec 1;15(23):7412–20.
- 520 80. Tazdait M, Mezquita L, Lahmar J, Ferrara R, Bidault F, Ammari S, et al. Patterns of responses in metastatic NSCLC during PD-1 or PDL-1
521 inhibitor therapy: Comparison of RECIST 1.1, irRECIST and iRECIST criteria. *Eur J Cancer.* 2018 Jan;88:38–47.
- 522 81. Park HJ, Kim KW, Pyo J, Suh CH, Yoon S, Hatabu H, et al. Incidence of Pseudoprogression during Immune Checkpoint Inhibitor Therapy
523 for Solid Tumors: A Systematic Review and Meta-Analysis. *Radiology.* 2020 Oct;297(1):87–96.
- 524 82. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: guidelines for response criteria for use in trials
525 testing immunotherapeutics. *Lancet Oncol.* 2017 Mar;18(3):e143–52.
- 526 83. Russo M, Crisafulli G, Sogari A, Reilly NM, Arena S, Lamba S, et al. Adaptive mutability of colorectal cancers in response to targeted
527 therapies. *Science.* 2019 Dec 20;366(6472):1473–80.

- 528 84. Siravegna G, Lazzari L, Crisafulli G, Sartore-Bianchi A, Mussolin B, Cassingena A, et al. Radiologic and Genomic Evolution of Individual
529 Metastases during HER2 Blockade in Colorectal Cancer. *Cancer Cell*. 2018 Jul 9;34(1):148-162.e7.
- 530 85. Dong Z-Y, Zhai H-R, Hou Q-Y, Su J, Liu S-Y, Yan H-H, et al. Mixed Responses to Systemic Therapy Revealed Potential Genetic
531 Heterogeneity and Poor Survival in Patients with Non-Small Cell Lung Cancer. *Oncologist*. 2017 Jan;22(1):61–9.
- 532 86. de Castro J, Cobo M, Isla D, Puente J, Reguart N, Cabeza B, et al. Recommendations for radiological diagnosis and assessment of
533 treatment response in lung cancer: a national consensus statement by the Spanish Society of Medical Radiology and the Spanish Society of
534 Medical Oncology. *Clin Transl Oncol*. 2015 Jan;17(1):11–23.
- 535 87. Ferretti GR, Reymond E, Delouche A, Sakhri L, Jankowski A, Moro-Sibilot D, et al. Personalized chemotherapy of lung cancer: What the
536 radiologist should know. *Diagn Interv Imaging*. 2016 Mar;97(3):287–96.
- 537 88. Crabb SJ, Patsios D, Sauerbrei E, Ellis PM, Arnold A, Goss G, et al. Tumor cavitation: impact on objective response evaluation in trials of
538 angiogenesis inhibitors in non-small-cell lung cancer. *J Clin Oncol*. 2009 Jan 20;27(3):404–10.
- 539 89. Marom EM, Martinez CH, Truong MT, Lei X, Sabloff BS, Munden RF, et al. Tumor cavitation during therapy with antiangiogenesis agents
540 in patients with lung cancer. *J Thorac Oncol*. 2008 Apr;3(4):351–7.
- 541 90. Kong BY, Menzies AM, Saunders CAB, Liniker E, Ramanujam S, Guminski A, et al. Residual FDG-PET metabolic activity in metastatic
542 melanoma patients with prolonged response to anti-PD-1 therapy. *Pigment Cell Melanoma Res*. 2016 Sep;29(5):572–7.
- 543 91. Rubin DL, Willrett D, O'Connor MJ, Hage C, Kurtz C, Moreira DA. Automated tracking of quantitative assessments of tumor burden in
544 clinical trials. *Transl Oncol*. 2014 Feb;7(1):23–35.
- 545 92. Barash Y, Klang E. Automated quantitative assessment of oncological disease progression using deep learning. *Ann Transl Med*. 2019
546 Dec;7(Suppl 8):S379.
- 547 93. Kickingeder P, Isensee F, Tursunova I, Petersen J, Neuberger U, Bonekamp D, et al. Automated quantitative tumour response
548 assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. *Lancet Oncol*. 2019 May;20(5):728–40.
- 549 94. Baidya Kayal E, Kandasamy D, Yadav R, Bakhshi S, Sharma R, Mehndiratta A. Automatic segmentation and RECIST score evaluation in
550 osteosarcoma using diffusion MRI: A computer aided system process. *Eur J Radiol*. 2020 Dec;133:109359.
- 551 95. Tang Y, Yan K, Xiao J, Summers RM. One Click Lesion RECIST Measurement and Segmentation on CT Scans. In: Martel AL, Abolmaesumi
552 P, Stoyanov D, Mateus D, Zuluaga MA, Zhou SK, et al., editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*.
553 Cham: Springer International Publishing; 2020. p. 573–83. (Lecture Notes in Computer Science).
- 554 96. Moawad AW, Fuentes D, Khalaf AM, Blair KJ, Szklaruk J, Qayyum A, et al. Feasibility of Automated Volumetric Assessment of Large
555 Hepatocellular Carcinomas' Responses to Transarterial Chemoembolization. *Front Oncol*. 2020;10:572.
- 556 97. Iannessi A, Beaumont H, Liu Y, Bertrand A-S. RECIST 1.1 and lesion selection: How to deal with ambiguity at baseline? *Insights Imaging*.
557 2021 Mar 18;12(1):36.

- 558 98. Ruchalski K, Braschi-Amirfarzan M, Douek M, Sai V, Gutierrez A, Dewan R, et al. A Primer on RECIST 1.1 for Oncologic Imaging in Clinical
559 Drug Trials. *Radiol Imaging Cancer*. 2021 May;3(3):e210008.
- 560 99. Kantarjian HM, Fojo T, Mathisen M, Zwelling LA. Cancer drugs in the United States: Justum Pretium--the just price. *J Clin Oncol*. 2013 Oct
561 1;31(28):3600-4.
- 562 100. Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of evidence of benefits on overall survival and quality of life of
563 cancer drugs approved by European Medicines Agency: retrospective cohort study of drug approvals 2009-13. *BMJ*. 2017 Oct 4;359:j4530.
- 564 101. Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, et al. Recommendations for initial evaluation, staging, and response
565 assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol*. 2014 Sep 20;32(27):3059-68.
- 566 102. Younes A, Hilden P, Coiffier B, Hagenbeek A, Salles G, Wilson W, et al. International Working Group consensus response evaluation
567 criteria in lymphoma (RECIL 2017). *Ann Oncol*. 2017 Jul 1;28(7):1436-47.
- 568 103. van den Bent MJ, Wefel JS, Schiff D, Taphoorn MJB, Jaeckle K, Junck L, et al. Response assessment in neuro-oncology (a report of the
569 RANO group): assessment of outcome in trials of diffuse low-grade gliomas. *Lancet Oncol*. 2011 Jun;12(6):583-93.
- 570 104. Lin NU, Lee EQ, Aoyama H, Barani IJ, Barboriak DP, Baumert BG, et al. Response assessment criteria for brain metastases: proposal from
571 the RANO group. *Lancet Oncol*. 2015 Jun;16(6):e270-278.
- 572 105. Shafrin J, Brookmeyer R, Peneva D, Park J, Zhang J, Figlin RA, et al. The value of surrogate endpoints for predicting real-world survival
573 across five cancer types. *Curr Med Res Opin*. 2016;32(4):731-9.
- 574 106. Jacene HA, Leboulleux S, Baba S, Chatzifotiadis D, Goudarzi B, Teytelbaum O, et al. Assessment of interobserver reproducibility in
575 quantitative 18F-FDG PET and CT measurements of tumor response to therapy. *J Nucl Med*. 2009 Nov;50(11):1760-9.
- 576 107. Rothe JH, Grieser C, Lehmkuhl L, Schnapauff D, Fernandez CP, Maurer MH, et al. Size determination and response assessment of liver
577 metastases with computed tomography--comparison of RECIST and volumetric algorithms. *Eur J Radiol*. 2013 Nov;82(11):1831-9.
- 578 108. Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, et al. Measurement of clinical and subclinical tumour response
579 using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for
580 Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer*. 1999 Dec;35(13):1773-82.
- 581 109. de Langen AJ, Vincent A, Velasquez LM, van Tinteren H, Boellaard R, Shankar LK, et al. Repeatability of 18F-FDG uptake measurements
582 in tumors: a metaanalysis. *J Nucl Med*. 2012 May;53(5):701-8.
- 583 110. Burger IA, Huser DM, Burger C, von Schulthess GK, Buck A. Repeatability of FDG quantification in tumor imaging: averaged SUVs are
584 superior to SUVmax. *Nucl Med Biol*. 2012 Jul;39(5):666-70.
- 585 111. Kurland BF, Peterson LM, Shields AT, Lee JH, Byrd DW, Novakova-Jiresova A, et al. Test-Retest Reproducibility of 18F-FDG PET/CT Uptake
586 in Cancer Patients Within a Qualified and Calibrated Local Network. *J Nucl Med*. 2019 May;60(5):608-14.

587 112. Brendle C, Kupferschläger J, Nikolaou K, la Fougère C, Gatidis S, Pfannenber g C. Is the standard uptake value (SUV) appropriate for
588 quantification in clinical PET imaging? - Variability induced by different SUV measurements and varying reconstruction methods. *Eur J Radiol.*
589 2015 Jan;84(1):158–62.

590 113. Miyashita K, Takahashi N, Oka T, Asakawa S, Lee J, Shizukuishi K, et al. SUV correction for injection errors in FDG-PET examination. *Ann*
591 *Nucl Med.* 2007 Dec;21(10):607–13.

592 114. Boellaard R, Oyen WJG, Hoekstra CJ, Hoekstra OS, Visser EP, Willemsen AT, et al. The Netherlands protocol for standardisation and
593 quantification of FDG whole body PET studies in multi-centre trials. *Eur J Nucl Med Mol Imaging.* 2008 Dec;35(12):2320–33.

594 115. Zhuang M, García DV, Kramer GM, Frings V, Smit EF, Dierckx R, et al. Variability and Repeatability of Quantitative Uptake Metrics in 18F-
595 FDG PET/CT of Non-Small Cell Lung Cancer: Impact of Segmentation Method, Uptake Interval, and Reconstruction Protocol. *J Nucl Med.* 2019
596 May;60(5):600–7.

597 116. Laffon E, de Clermont H, Marthan R. A method of adjusting SUV for injection-acquisition time differences in (18)F-FDG PET imaging. *Eur*
598 *Radiol.* 2011 Nov;21(11):2417–24.

599 117. Kamimura K, Nagamachi S, Wakamatsu H, Higashi R, Ogita M, Ueno S, et al. Associations between liver (18)F fluoro-2-deoxy-D-glucose
600 accumulation and various clinical parameters in a Japanese population: influence of the metabolic syndrome. *Ann Nucl Med.* 2010
601 Apr;24(3):157–61.

602 118. Krak NC, Boellaard R, Hoekstra OS, Twisk JWR, Hoekstra CJ, Lammertsma AA. Effects of ROI definition and reconstruction method on
603 quantitative outcome and applicability in a response monitoring trial. *Eur J Nucl Med Mol Imaging.* 2005 Mar;32(3):294–301.

604 119. Lasnon C, Quak E, Le Roux P-Y, Robin P, Hofman MS, Bourhis D, et al. EORTC PET response criteria are more influenced by
605 reconstruction inconsistencies than PERCIST but both benefit from the EARL harmonization program. *EJNMMI Phys.* 2017 Dec;4(1):17.

606 120. Akamatsu G, Ikari Y, Nishida H, Nishio T, Ohnishi A, Maebatake A, et al. Influence of Statistical Fluctuation on Reproducibility and
607 Accuracy of SUVmax and SUVpeak: A Phantom Study. *J Nucl Med Technol.* 2015 Sep;43(3):222–6.

608 121. Paquet N, Albert A, Foidart J, Hustinx R. Within-patient variability of (18)F-FDG: standardized uptake values in normal tissues. *J Nucl*
609 *Med.* 2004 May;45(5):784–8.

610 122. O JH, Lim SJ, Wang H, Leal JP, Shu H-KG, Wahl RL, et al. Quantitation of cancer treatment response by 2-[18F]FDG PET/CT: multi-center
611 assessment of measurement variability using AUTO-PERCIST™. *EJNMMI Res.* 2021 Feb 12;11(1):15.

612 123. Apostolova I, Wiemker R, Paulus T, Kabus S, Dreilich T, van den Hoff J, et al. Combined correction of recovery effect and motion blur for
613 SUV quantification of solitary pulmonary nodules in FDG PET/CT. *Eur Radiol.* 2010 Aug;20(8):1868–77.

614 124. Lodge MA, Chaudhry MA, Wahl RL. Noise considerations for PET quantification using maximum and peak standardized uptake value. *J*
615 *Nucl Med.* 2012 Jul;53(7):1041–7.

- 616 125. Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. *AJR Am J*
617 *Roentgenol.* 2010 Aug;195(2):310–20.
- 618 126. Quak E, Le Roux P-Y, Hofman MS, Robin P, Bourhis D, Callahan J, et al. Harmonizing FDG PET quantification while maintaining optimal
619 lesion detection: prospective multicentre validation in 517 oncology patients. *Eur J Nucl Med Mol Imaging.* 2015 Dec;42(13):2072–82.
- 620 127. Gürses B, Altınmakas E, Böge M, Aygün MS, Bayram O, Balık E. Multiparametric MRI of rectal cancer-repeatability of quantitative data: a
621 feasibility study. *Diagn Interv Radiol.* 2020 Mar;26(2):87–94.
- 622 128. Lundsgaard Hansen M, Fallentin E, Axelsen T, Lauridsen C, Norling R, Svendsen LB, et al. Interobserver and Intraobserver Reproducibility
623 with Volume Dynamic Contrast Enhanced Computed Tomography (DCE-CT) in Gastroesophageal Junction Cancer. *Diagnostics (Basel).* 2016 Feb
624 1;6(1):E8.
- 625 129. Conte GM, Castellano A, Altabella L, Iadanza A, Cadioli M, Falini A, et al. Reproducibility of dynamic contrast-enhanced MRI and dynamic
626 susceptibility contrast MRI in the study of brain gliomas: a comparison of data obtained using different commercial software. *Radiol Med.* 2017
627 Apr;122(4):294–302.
- 628 130. Aronhime S, Calcagno C, Jajamovich GH, Dyvorne HA, Robson P, Dieterich D, et al. DCE-MRI of the liver: effect of linear and nonlinear
629 conversions on hepatic perfusion quantification and reproducibility. *J Magn Reson Imaging.* 2014 Jul;40(1):90–8.
- 630 131. Smits M, Bendszus M, Collette S, Postma LA, Dhermain F, Hagenbeek RE, et al. Repeatability and reproducibility of relative cerebral
631 blood volume measurement of recurrent glioma in a multicentre trial setting. *Eur J Cancer.* 2019 Jun;114:89–96.
- 632 132. Ng CS, Waterton JC, Kundra V, Brammer D, Ravoori M, Han L, et al. Reproducibility and comparison of DCE-MRI and DCE-CT perfusion
633 parameters in a rat tumor model. *Technol Cancer Res Treat.* 2012 Jun;11(3):279–88.
- 634 133. Newitt DC, Zhang Z, Gibbs JE, Partridge SC, Chenevert TL, Rosen MA, et al. Test-retest repeatability and reproducibility of ADC measures
635 by breast DWI: Results from the ACRIN 6698 trial. *J Magn Reson Imaging.* 2019 Jun;49(6):1617–28.
- 636 134. Kwee RM, Dik AK, Sosef MN, Berendsen RCM, Sassen S, Lammering G, et al. Interobserver reproducibility of diffusion-weighted MRI in
637 monitoring tumor response to neoadjuvant therapy in esophageal cancer. *PLoS One.* 2014;9(4):e92211.
- 638 135. Koh D-M, Blackledge M, Collins DJ, Padhani AR, Wallace T, Wilton B, et al. Reproducibility and changes in the apparent diffusion
639 coefficients of solid tumours treated with combretastatin A4 phosphate and bevacizumab in a two-centre phase I clinical trial. *Eur Radiol.* 2009
640 Nov;19(11):2728–38.
- 641 136. Weller A, Papoutsaki MV, Waterton JC, Chiti A, Stroobants S, Kuijter J, et al. Diffusion-weighted (DW) MRI in lung cancers: ADC test-retest
642 repeatability. *Eur Radiol.* 2017 Nov;27(11):4552–62.
- 643 137. Sorace AG, Elkassem AA, Galgano SJ, Lapi SE, Larimer BM, Partridge SC, et al. Imaging for Response Assessment in Cancer Clinical Trials.
644 *Semin Nucl Med.* 2020 Nov;50(6):488–504.

645 138. Galbán CJ, Ma B, Malyarenko D, Pickles MD, Heist K, Henry NL, et al. Multi-site clinical evaluation of DW-MRI as a treatment response
 646 metric for breast cancer patients undergoing neoadjuvant chemotherapy. PLoS One. 2015;10(3):e0122151.

647 139. Newitt DC, Tan ET, Wilmes LJ, Chenevert TL, Kornak J, Marinelli L, et al. Gradient nonlinearity correction to improve apparent diffusion
 648 coefficient accuracy and standardization in the american college of radiology imaging network 6698 breast cancer trial. J Magn Reson Imaging.
 649 2015 Oct;42(4):908–19.

650 140. Newitt DC, Amouzandeh G, Partridge SC, Marques HS, Herman BA, Ross BD, et al. Repeatability and Reproducibility of ADC Histogram
 651 Metrics from the ACRIN 6698 Breast Cancer Therapy Response Trial. Tomography. 2020 Jun;6(2):177–85.

652

653 **Table 1. RECIST categories of response.**

Overall response	Target lesions	Non target lesions	New lesions
Definition	<ul style="list-style-type: none"> ● Lesions with longest diameter ≥ 10 mm and limits that are sufficiently well defined for their measurement to be considered reliable ● Lymph nodes: measurement of short axis, target lesion if short-axis measures ≥ 15 mm ● Maximum number of selected target lesions 5/patient and 2/organ 	<ul style="list-style-type: none"> ● Lesions that are too small (< 10 mm) ● Lesions for which measurement is considered unreliable as their limits are difficult to define (bone or leptomeningeal lesions, ascites, pleural or pericardial effusion, lymphangitic carcinomatosis etc.) ● Measurable lesions not selected as target lesions ● Lymph nodes: measurement of short axis, non-target lesion if $10 \text{ mm} \leq$ short-axis diameter < 15 mm ● Levels of tumour markers > normal (if relevant and predefined) 	
Complete response (CR)	<ul style="list-style-type: none"> ● Disappearance of all target lesions and all nodes have short axis < 10 mm 	<ul style="list-style-type: none"> ● Disappearance of all non-target lesions and normalisation of tumour marker levels 	<ul style="list-style-type: none"> ● No
Partial response (PR)	<ul style="list-style-type: none"> ● ≥ 30 % decrease in the sum of target lesions taking as reference the baseline sum 	<ul style="list-style-type: none"> ● No progression 	<ul style="list-style-type: none"> ● No
Stable disease (SD)	<ul style="list-style-type: none"> ● Neither response nor progression 	<ul style="list-style-type: none"> ● Persistence of one or more non-target lesions and/or tumour marker levels > normal 	<ul style="list-style-type: none"> ● No
Progressive disease (PD) : response is PD if at least one category of lesions meets progression criteria	<ul style="list-style-type: none"> ● ≥ 20 % increase in the sum of target lesions taking as reference the smallest sum measured during follow-up (nadir) and ≥ 5 mm in absolute value 	<ul style="list-style-type: none"> ● 'Unequivocal' progression (assessed qualitatively) in lesion size (an increase in size of a single lesion is not sufficient) 	<ul style="list-style-type: none"> ● Yes (appearance of new unequivocally metastatic lesion(s))

In review

Biomarker	Reproducibility			Factors impacting reproducibility
	95% limits of agreement	Kappa	Other	
RECIST (measurement) CT (size)	<u>Per lesion</u> - Intra-obs: -18% to 16% - Inter-obs: -22% to 25% (1(17)) <u>Per sum of diameters</u> - Intra-obs: -10% to 13% - Inter-obs: -20% to 20% <u>Interval change in tumour burden</u> (% change between time points) - -31% to 30% <u>Repeatability</u> (same image on repeat CT taken within 15 minutes) - -4% to +4% (24)	<u>With target lesion selection</u> - Intra-obs: 0.957 (25) - Inter-obs 0.954 (25) <u>Target response classification</u> - Inter-obs: 0.48 (21) to 0.66 (20) <u>Non-target response classification</u> - Inter-obs: 0.58 (21)	<u>Lesion size ICC</u> (106) - Pre-treatment: 0.72 - Post-treatment: 0.85 - Interval change: 0.70	<ul style="list-style-type: none"> - Selection of target lesions differs in 21 to 33 % (17,18,30) - Practical training (ref 40) / expertise (20) - Same observer (17,21) - Well delineated lesions (17,25) - Lesions size (greater variability for smaller lesions) (24,25) - Adjudication could reduce easily avoidable inconsistencies (21,22)
RECIST (overall response)		<u>With target lesion selection</u> - Inter-obs: 0.97 (18) <u>Without target lesion selection</u> - Inter-obs: 0.51 (21), 0.53 (23,30) to 0.58 (18)	<ul style="list-style-type: none"> - 30% of patients classified differently in a cohort of 39 pts with 2 readers (23) 	<ul style="list-style-type: none"> - Arbitrary nature of CR/ PR/ SD/PD categories (10) - Inconsistencies mainly due to interpretation of new lesions (21,23) - Choice of target lesions
3D measurement	<ul style="list-style-type: none"> - Intra-obs: 0.4 to 33% according to automated volume measurement method (107) 	<u>Whole body volumetry</u> - Inter-obs: 0.95 (30)	<ul style="list-style-type: none"> - Discordant classification in overall response in 10 to 21% of patients according to automated volume measurement method (107) 	<ul style="list-style-type: none"> - Time consuming (31) - Do not resolve the discrepancies linked to the choice of target lesions (30)

Table 3: RECIST reproducibility and factors impacting it. 95% limits of agreement are derived from the Bland-Altman method comparing two measurements of the same variable. Kappa coefficients measure agreement between qualitative observations. ICC measures the reliability of measurements by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects.

Abbreviations: Intra-obs: intra-observer; Inter-obs: inter-observer; ICC: Intra-class coefficient; CR: complete response; PR: partial response; SD: stable disease; PD: progressive disease.

Biomarker	Reproducibility			Factors impacting reproducibility
	ICC	Coefficient of Variation	Other	
Metabolic activity (¹⁸ F-FDG PET) <i>Semiquantitative:</i> SUV (SUV _{max} , SUV _{mean} , SUV _{peak}), SUL (SUL _{max} , SUL _{mean} , SUL _{peak}); MTV, TLG <i>Response criteria:</i> PERCIST (62) /EORTC (108)	<u>SUV_{max} (4 observers) (106)</u> - Pre-treatment: 0.93 - Post-treatment: 0.91 - Interval change: 0.94 <u>SUV_{mean} repeatability (109)</u> - 0.91 (meta-analysis) <u>SUV_{peak}</u> - -31% to 30% -	<u>SUVmax (4 observers) (106)</u> - Pre-treatment: 6.3% - Post-treatment: 18.4% - Interval change: 16.7%	<u>Repeatability standard deviation (110)</u> - SUV _{max} : 1.01 - SUV _{mean} : 0.28	<i>Technical factors:</i> Scanner calibration/injected activity calibration (41,111) Incorrect decay correction (112) Tracer extravasation (113,114) Residual activity in syringe (41) Synchronization of clocks (41) <i>Biological factors:</i> Blood glucose levels (114) Inflammation (41) Patient preparation (114) Injection-acquisition interval (115,116) BMI/metabolic syndrome (117) Drug interaction/corticosteroids (114) <i>Physical factors:</i> Acquisition parameters/matrix size (41,112) Reconstruction algorithm (115,118,119) Partial volume effect (120) Normalization factor for SUV (41,121) Use of contrast agents (41) ROI/VOI definition (115,118) Semiautomated/manual contouring (122) Movement artifacts/respiratory movements (41) Recovery effect/motion blur(123) Image noise (120,124) Background activity/visual assessment (118,125) Lesion size/location (126)
Vascularity (DCE MRI)	<u>DCE-MRI ktrans</u> - Intra-obs: 0.98 (127) <u>DCE CT (arterial flow, blood volume, permeability)</u> - Intra-obs: 0.72-0.89 - Inter-obs: 0.70-0.91 (128)	<u>DCE MRI</u> - model-free parameters (ex: AUC ₆₀ , peak...): 12-24% - modeled parameters (ex: distribution volume, blood flow, mean transit time): 21-29% (130)		- Parameter extraction model (130) - Segmentation: 3D vs 2D regions of interest (128) - Software (129)

	<u>DCE and DSC-MRI intersoftware reproducibility</u> ICC 0.31 to 0.58 (129)	<u>DSC MRI normalised rCBVmax</u> - repeatability: 50%, - reproducibility: 6% (131) <u>DCE-CT</u> (blood flow, blood volume, mean transit time, permeability) - within subject: 18% to 25%; <u>DCE-</u> <u>MRI</u> (Ktrans, k(ep), v(e)) - within subject 16% to 23%, (132)		
Cellularity (MR) ADC	<u>ADC mean value</u> - Intra-obs: 0.91 (133) – 0.99 (127) - Inter-obs: 0.92 (133) <u>ROI segmentation method</u> (Inter-obs) - Manual method: 0.69 - Semi-automated volumetric method: 0.96 (134)	<u>Repeatability</u> - ADC total = 4.8% (133), 7.1% (142) to 13.3% (135) <u>Different post-processing platforms</u> - 2.8% (136) <u>Different sites</u> - multicentric: 9% (137) - ice-water phantom: 1.6% (143) - breast fibroglandular tissue: 7.0% (137)	<u>Repeatability</u> (single-centre) - $\leq \pm 0.1 \times 10^{-3} \text{ mm}^2/\text{s}$ (138)	- Field homogeneity gradient linearity (139) - QA procedure by trained operators assessing artifacts, fat suppression, and signal-to-noise ratio (133) - Segmentation: 2D vs. 3D, manual vs. semi- automatic (134) - Choice of measurement: mean/min/max/percentiles of ADC (140) - Lesion size (136)

Table 4 : Reproducibility and factors impacting it of other imaging biomarkers. SUV_{max} is measured as the maximum single voxel value of SUV, SUV_{mean} is the average value of SUV in all voxels above a threshold, SUV_{peak} (is the average value of SUV in a region of interest positioned so as to maximize the enclosed average).

Abbreviations: SUV = standardized uptake value; SUL = lean body mass corrected SUV; MTV = metabolic tumour volume; TLG = total lesion glycolysis; PERCIST = PET Response Criteria in Solid Tumours; EORTC = European Organization for Research and Treatment of Cancer; wCV = within-subject coefficient of variation; BMI = body mass index; ROI = region of interest; VOI = volume of interest; ICC inter correlation coefficient; DCE dynamic contrast enhanced; DSC-MRI dynamic susceptibility contrast magnetic resonance imaging; ADC: apparent diffusion coefficient; QA quality assurance; 3D: three-dimensional; 2D: bi-dimensional; AUC_{60} : area under the curve at 60s; rCBV: relative cerebral blood volume; Ktrans: transfer constant; k(ep): wash-out transfer constant; v(e): extracellular volume.

Acknowledgements

We would like to thank the numerous colleagues for their insightful discussions and comments that have facilitated the work in this manuscript.

Conflicts of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Funding

No specific funding was received to enable this work.

In review

Figure 1.JPEG

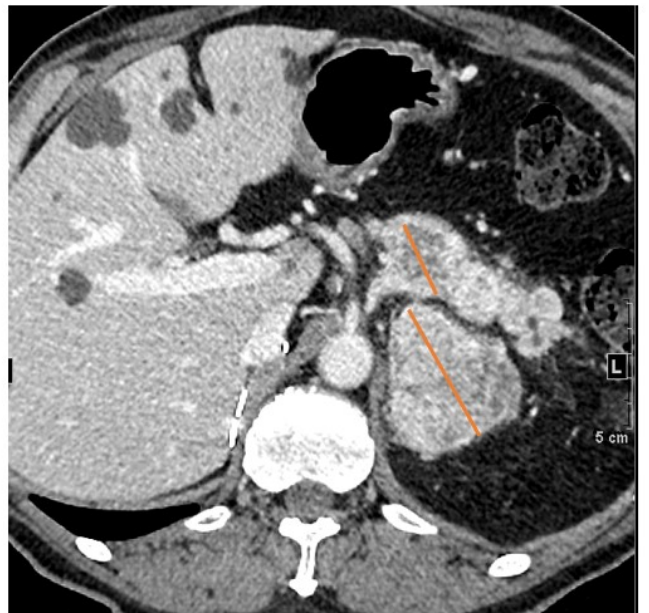
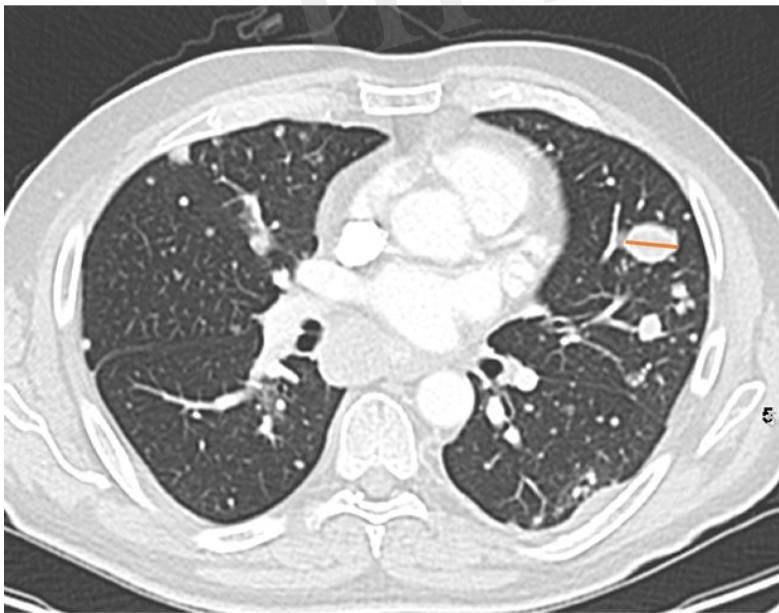


Figure 2.JPEG

In review



Figure 3.JPEG

In review

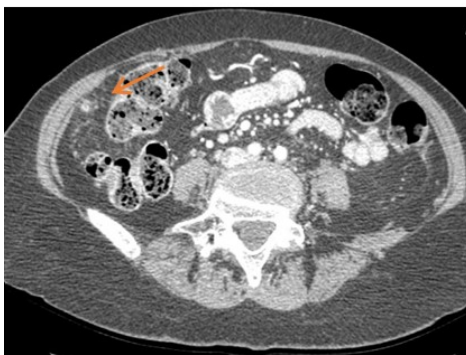
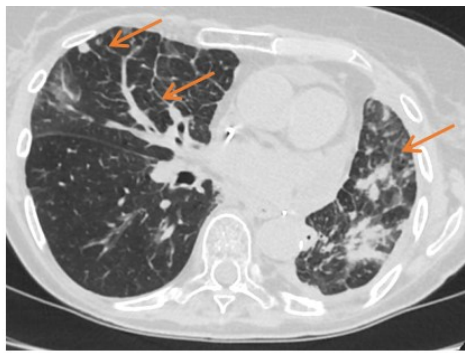
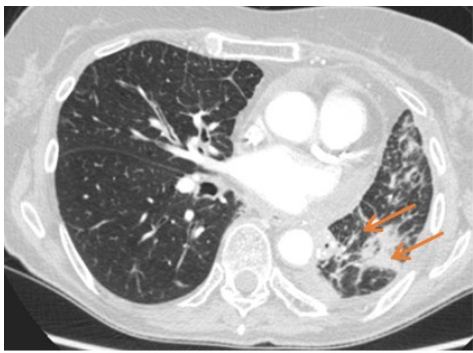


Figure 4.JPEG

