Interobserver Variability in CT-based Morphologic Tumor Response Assessment of Colorectal Liver Metastases

Nina J. Wesdorp, MD* • Ruby Kemna, MSc* • Karen Bolhuis, MD • Jan H. T. M. van Waesberghe, MD, PhD • Irene M. G. C. Nota, MD • Femke Struik, MD • Ikrame Oulad Abdennabi, MD • Saffire S. K. S. Phoa, MD, PhD • Susan van Dieren, PhD • Martinus J. van Amerongen, MD, PhD • Thiery Chapelle, MD, PhD • Cornelis H. C. Dejong, MD, PhD • Marc R. W. Engelbrecht, MD, PhD • Michael F. Gerhards, MD, PhD • Dirk Grünhagen, MD, PhD • Thomas M. van Gulik, MD, PhD • John J. Hermans, MSc, MD, PhD • Koert P. de Jong, MD, PhD • Joost M. Klaase, MD, PhD • Mike S. L. Liem, MD, PhD • Krijn P. van Lienden, MD, PhD • I. Quintus Molenaar, MD, PhD • Gijs A. Patijn, MD, PhD • Arjen M. Rijken, MD, PhD • Theo M. Ruers, MD, PhD • Cornelis Verhoef, MD, PhD • Johannes H. W. de Wilt, MD, PhD • Rutger-Jan Swijnenburg, MD, PhD • Cornelis J. A. Punt, MD, PhD • Joost Huiskens, MD, PhD • Jaap Stoker, MD, PhD** • Geert Kazemier, MD, PhD** • for the Dutch Colorectal Liver Expert Panel

Purpose: To evaluate interobserver variability in the morphologic tumor response assessment of colorectal liver metastases (CRLM) managed with systemic therapy and to assess the relation of morphologic response with gene mutation status, targeted therapy, and Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 measurements.

Materials and Methods: Participants with initially unresectable CRLM receiving different systemic therapy regimens from the randomized, controlled CAIRO5 trial (NCT02162563) were included in this prospective imaging study. Three radiologists independently assessed morphologic tumor response on baseline and first follow-up CT scans according to previously published criteria. Two additional radiologists evaluated disagreement cases. Interobserver agreement was calculated by using Fleiss κ. On the basis of the majority of individual radiologic assessments, the final morphologic tumor response was determined. Finally, the relation of morphologic tumor response and clinical prognostic parameters was assessed.

Results: In total, 153 participants (median age, 63 years [IQR, 56–71]; 101 men) with 306 CT scans comprising 2192 CRLM were included. Morphologic assessment performed by the three radiologists yielded 86 (56%) agreement cases and 67 (44%) disagreement cases (including four major disagreement cases). Overall interobserver agreement between the panel radiologists on morphology groups and morphologic response categories was moderate (κ = 0.53, 95% CI: 0.48, 0.58 and κ = 0.54, 95% CI: 0.47, 0.60). Optimal morphologic response was particularly observed in patients treated with bevacizumab (P = .001) and in patients with *RAS/BRAF* mutation (P = .04). No evidence of a relationship between RECIST 1.1 and morphologic response was found (P = .61).

Conclusion: Morphologic tumor response assessment following systemic therapy in participants with CRLM demonstrated considerable interobserver variability.

Clinical trial registration no. NCT02162563

Supplemental material is available for this article.

©RSNA, 2022

Abbreviations

CRLM = colorectal liver metastases, RECIST = Response Evaluation Criteria in Solid Tumors

Summary

Assessment of morphologic tumor response on CT scans of patients with initially unresectable colorectal liver metastases treated with systemic therapy resulted in considerable interobserver variability.

Key Points

- Morphologic tumor response assessment showed a moderate interobserver agreement (κ = 0.54, 95% CI: 0.47, 0.60) among three abdominal radiologists, and disagreement was observed in 67 of 153 (44%) participants.
- Optimal morphologic tumor response was particularly observed in participants treated with chemotherapy in combination with bevacizumab, namely 35 of 37 (95%) participants with optimal morphologic response (P = .001).

Keywords

- Tumor Response, Observer Performance, CT, Liver, Metastases, Oncology, Abdomen/Gastrointestinal

Colorectal cancer is currently the third most common form of cancer in men and women (1). More than half of patients with colorectal cancer develop colorectal liver metastases (CRLM), and for these patients, local treatment is considered the only curative strategy (2–5). However, at diagnosis of CRLM, only 20% of patients are eligible for local therapy with curative intent (ie, resection and/or ablation) (5,6). The remaining patients with CRLM most often receive systemic therapy to either induce downsizing of the tumor load for secondary local treatment or to prolong survival while maintaining quality of life in a palliative setting. Over the years, advancements in systemic therapy regimens have been introduced, such as the combination of cytotoxic agents with the addition of targeted therapies, and these have been associated with improved outcomes (7–9). Additionally, tumor mutation status and location of the primary tumor play an increasing role in assessing the prognosis and the selection of treatment for patients with CRLM. Primary tumor location and *RAS/BRAF* wild-type and mutant tumors are typically associated with different responses to specific targeted agents, including anti–epidermal growth factor receptor agents (10–14). Induction systemic therapy can downsize CRLM, thereby increasing the chance of successful secondary local therapy and survival (7,8,15). Accurate evaluation of tumor response to systemic therapy is therefore crucial.

Currently, the Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 are most widely applied to evaluate tumor response to treatment (16). These size-based radiologic criteria evaluate the tumor response by measuring the change in diameter of a maximum of two target lesions per affected organ (16). However, CRLM do not only undergo changes in size following systemic therapy but also tend to undergo morphologic changes (17,18). Consequently, assessing tumor response using the standard RECIST 1.1 might lead to an underestimation of the tumor response, as morphologic changes of CRLM can appear earlier than or without concomitant tumor lesion shrinkage (17–19).

These observed changes in morphologic condition correlated with pathologic response and overall survival in patients with CRLM treated with bevacizumab-containing systemic therapy, whereas RECIST 1.1 did not (17). Similarly, the association between morphologic tumor response and survival has been demonstrated in multiple other studies (18,20–23). However, morphologic tumor response assessment is based on visual observations and hence depends on observer perception. Therefore, it is essential to assess the level of agreement between observers for morphologic tumor response assessment (24). Furthermore, morphologic tumor response has particularly been observed in patients with CRLM treated with chemotherapy in combination with the anti–vascular endothelial growth agent bevacizumab (17,18,20–22,25). Information about morphologic tumor response in patients with CRLM treated with anti–epidermal growth factor receptor therapy (eg, panitumumab) is lacking, and the role of *RAS/BRAF* mutation status in morphologic tumor response to different agents has not previously been well established in the literature.

We hypothesized that morphologic tumor response assessment is vulnerable to subjectivity because it depends on visual observations. This study aims to evaluate interobserver variability between multiple observers performing morphologic tumor response assessment according to set criteria in participants with CRLM treated with systemic therapy in a prospective substudy of a therapeutic clinical trial. The secondary objective was to assess the relation of morphologic tumor response with gene mutation status and targeted therapy. Finally, we compared morphologic tumor response with those of the RECIST 1.1 measurements.

Materials and Methods

Study Sample

The study sample consisted of participants registered between November 2014 and April 2019 in the multicenter randomized clinical trial of the Dutch Colorectal Cancer Group, CAIRO5 (ClinicalTrials.gov identifier: NCT02162563) (26). The ongoing CAIRO5 trial aims to select the optimal induction systemic treatment strategy for patients with initially unresectable CRLM. Resectability is assessed by a panel of expert liver surgeons according to predefined criteria. At the moment of initiating this imaging study, 369 patients with colorectal cancer aged 18 years or older with initially unresectable, liver-only CRLM had been randomized for different systemic therapy regimen combinations on the basis of *RAS/BRAF* gene mutation status and primary tumor location (categorized as left or right of the splenic flexure). All participants signed a written consent form, and the study was conducted according to the ethical standards of the Helsinki Declaration of 1975 and has been approved by the medical ethical committee. Patients with *RAS/BRAF* mutated or right-sided primary tumors were randomized between doublet (5-fluorouracil with leucovorin and oxaliplatin or 5-fluorouracil with leucovorin and irinotecan) or triplet (5-fluorouracil with leucovorin, oxaliplatin, and irinotecan) chemotherapy, both in combination with bevacizumab. Participants with left-sided primary and *RAS/BRAF* wild-type tumors were randomized for
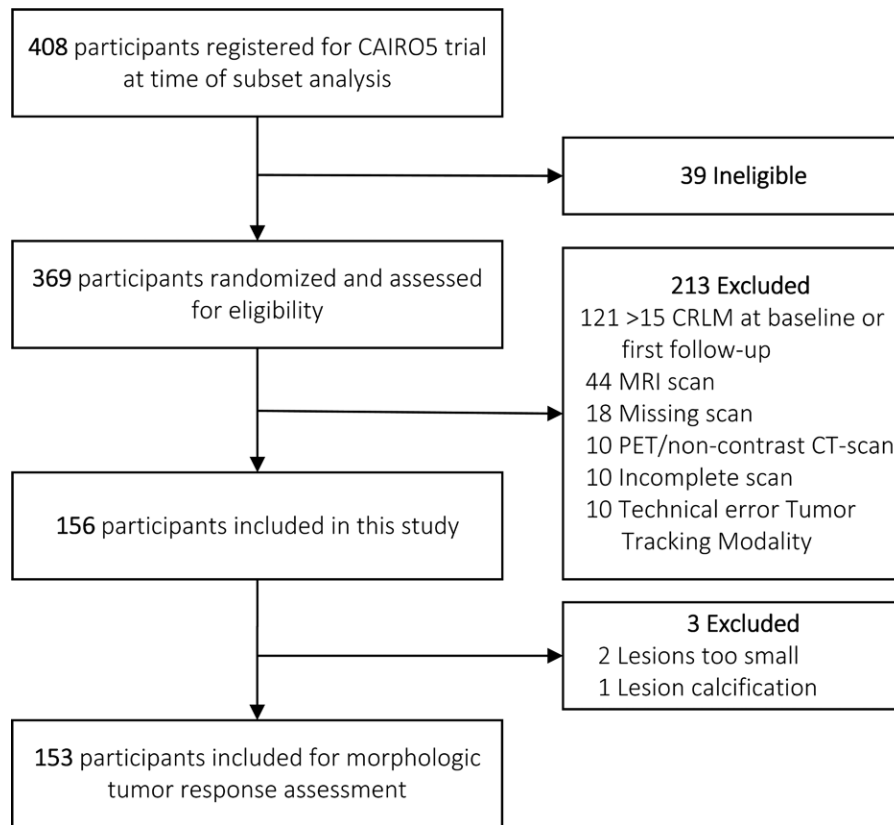
Figure 1: Flow diagram of selection study sample for morphologic tumor response assessment. CLRM = colorectal liver metastases.

doublet chemotherapy (5-fluorouracil with leucovorin and ox-aliplatin or 5-fluorouracil with leucovorin and irinotecan) with either bevacizumab or panitumumab. In the current study, only participants with a maximum of 15 CRLM either at baseline or first follow-up were included (Fig 1). This maximum number was determined in consultation with the research team before initiating this study to ensure study feasibility. Exclusion criteria included the following: more than 15 CRLM at baseline or first follow-up CT examination, missing baseline or first follow-up CT examination, incomplete CT examination, technical error in diagnostic software, and the use of MRI or PET with non–contrast-enhanced CT.

Imaging Protocol

Contrast-enhanced chest-abdomen CT scans of the CAIRO5 trial at baseline and, subsequently, every 2 months during systemic therapy were available. All CT examinations were performed in one of the 54 medical centers responsible for inclusion, resulting in different CT scanners and acquisition protocols. In the current study, only the baseline and first follow-up contrast-enhanced abdominal CT scans in the portal venous phase were included for morphologic tumor response assessment. Median values of CT parameters included section thickness of 3.0 mm (range, 1.00–5.00 mm), tube voltage of 120 kVp (range, 80–140 kVp), and total collimation width of 40 mm (range, 10–80 mm). The exposure ranged between 50 and 293 mAs, and the CT dose index ranged between 2 and 15 mGy.

Morphologic Criteria

Tumor response to systemic therapy was evaluated according to the morphologic criteria formulated by Chun et al (17), classifying CRLM into one of the three morphology groups. In patients with multiple metastases, morphologic response criteria were assigned on the basis of the response seen in the majority of tumors. Morphology group 1 was characterized by homogeneous and hypoattenuating attenuation, sharply defined tumor–liver interface, and, if initially present, complete disappearance of peripheral rim of enhancement. Morphology group 3 was characterized by heterogeneous attenuation and ill-defined tumor–liver interface, and peripheral rim of enhancement may be present. Morphology group 2 could not be assigned to morphology group 1 or morphology group 3 and was defined as mixed attenuation, variable tumor–liver interface, and, if initially present, partially resolved rim enhancement. Morphologic response was defined as *optimal response* if a change from either morphology group 2 or 3 to morphology group 1 was observed, *suboptimal response* if a change from morphology group 3 to 2 was observed, or *no response* if the morphology group or CRLM did not change or moved to a higher group (Fig 2).

Morphologic Tumor Response Assessment

Three abdominal radiologists (F.S., I.M.G.C.N., J.H.T.M.v.W., with 1, 2, and 18 years of experience, respectively), independently assessed morphologic treatment
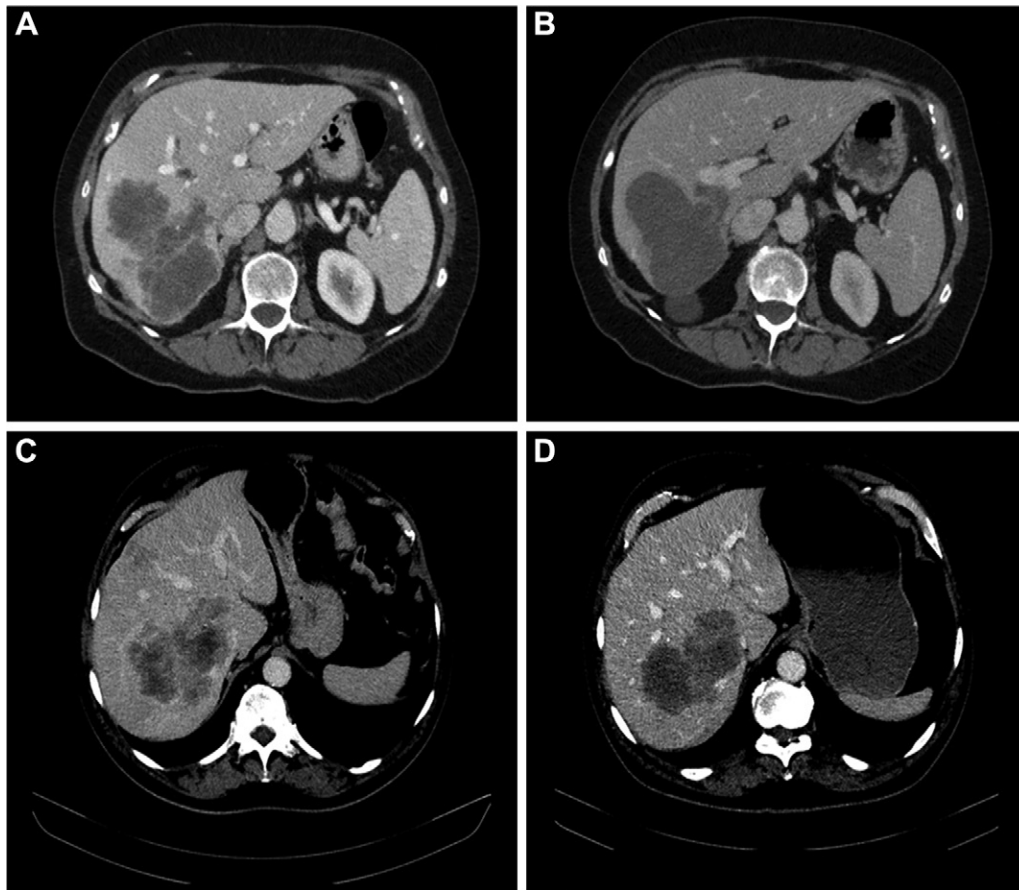
Figure 2: Morphologic response categories. Baseline and follow-up CT scans show **(A, B)** an optimal morphologic response of colorectal liver metastases with the transition of heterogeneous attenuation and ill-defined tumor–liver interface **(A)** to homogeneous and hypoattenuating attenuation and sharp tumor–liver interface **(B)** and **(C, D)** suboptimal morphologic response with the transition of heterogeneous attenuation and ill-defined tumor–liver interface **(C)** to homogeneous attenuation but ill-defined tumor–liver interface remaining after treatment **(D)**.

response of the same set of CT scans while blinded to participant characteristics and outcomes. All three radiologists assessed the CT scans in the same patient order, assessing the baseline scan first and subsequently the first follow-up scan. Disagreement cases were independently and blindly assessed by two additional abdominal radiologists (I.O.A., S.S.K.S.P., with 4 and 32 years of experience, respectively). These two additional radiologists were blinded to the assessments of the three radiologists. All five radiologists received identical information (see next paragraph) on the definition of the morphology groups and morphologic response prior to the individual assessments (17). Prior to the morphologic tumor response assessment, all radiologists reviewed six training CT scans depicting the different morphology groups; these six scans were not part of the study cohort. Morphologic tumor response assessments of the baseline and first follow-up CT examinations were performed in software (Tumor Tracking Modality of IntelliSpace Portal 9.0; Philips).

Determination of Final Morphologic Tumor Response

A final determination of morphologic response was assigned per patient to compare the morphologic tumor response with clinical prognostic parameters. Morphologic tumor response assessments of all individual participants evaluated by the three radiologists were compared. When all three radiologists assigned the same morphologic response category to a single patient, that particular morphologic response category was assigned as the definitive category. The disagreement cases were either minor or major disagreements. A minor disagreement was defined as an evaluation in which one of the three radiologists classified the CRLM as suboptimal response and another radiologist classified the CRLM as optimal response or no response. A major disagreement was defined as an evaluation in which at least one of the three radiologists assessed the CRLM as an optimal response and another radiologist classified the CRLM as no response.

All disagreement cases were read by the three radiologists and the two additional radiologists. Agreement on a morphologic response category for each disagreement case was based on the majority of votes on morphologic response among these five radiologists. If a majority of votes could not be reached between the five radiologists, the final decision was based on a consensus review by two of the three initial radiologists.

Table 1: Baseline Patient Characteristics

Parameter	Total (n = 153)
Age (y)	63 (56–71)
Men	101 (66)
Site of primary tumor	
Right colon	36 (24)
Left colon or rectum	117 (77)
Time to metastases	
Synchronous	127 (83)
Metachronous	26 (17)
Mutation status	
RAS	75 (49)
BRAF^{V600E}	6 (4)
RAS/BRAF wild-type	72 (47)
No. of liver metastases	8 (5–10)
Diameter of largest liver metastasis (mm)	40 (29–87)
No. of liver segments involved	5 (4–6)
Distribution of liver metastases	
Unilobar	20 (13)
Bilobar	133 (87)
Induction systemic therapy	
FOLFOX and FOLFIRI with bevacizumab	77 (50)
FOLFOX and FOLFIRI with panitumumab	31 (20)
FOLFOXIRI and bevacizumab	45 (30)

Note.—Values are shown as medians with IQRs in parentheses or numbers of participants with percentages in parentheses. FOLFIRI = 5-fluorouracil with leucovorin and irinotecan, FOLFOX = 5-fluorouracil with leucovorin and oxaliplatin, FOLFOXIRI = 5-fluorouracil with leucovorin, oxaliplatin, and irinotecan.

Radiologic Assessment of CAIRO5 Trial

The morphologic tumor response assessment performed in this study was compared with the radiologic assessment reported in the CAIRO5 trial. These radiology reports contained the tumor response evaluation according to RECIST 1.1 (16,17). This radiologic assessment was performed by one of five expert abdominal radiologists of the CAIRO5 trial on a digital platform (ALEA Forms Vision BV) (26). These CAIRO5 trial radiologists differed from the radiologists performing the morphologic assessment in current study. The RECIST 1.1 classification criteria for objective response were defined as complete response (disappearance of all target lesions), partial response (at least 30% decrease in the sum of diameters of target lesions), progressive disease (at least 20% increase in the sum of diameters of target lesions, including an absolute increase of 5 mm in diameter, or appearance of one or more new lesions), or stable disease (neither progressive disease nor partial or complete response) (16).

Statistical Analysis

Statistical analyses were performed by using software (SAS Studio version 5.2, SAS Viya release V.03.05, SAS Institute).

Categorical variables were reported as frequencies and percentages. Continuous variables were displayed as median with IQR (Q1–Q3), as the collected data were not normally distributed. Interobserver agreement was calculated among the three radiologists for the morphologic scores and the morphologic response categories by using the Fleiss κ. Interpretation of the Fleiss κ was categorized as having agreement that was either poor (κ < 0.20), fair (κ = 0.21–0.40), moderate (κ = 0.41–0.60), good (κ = 0.61–0.80), or very good (κ = 0.81–1.0) (27). Clinical parameters (gene mutation status and targeted therapy) were compared among the morphologic tumor response groups. In addition, RECIST 1.1 measurements of the CAIRO5 trial were compared with the morphologic tumor response. Categorical variables were compared by using the χ^2 test or Fisher exact test, when appropriate. Tests were considered statistically significant at P less than .05.

Results

Participant Characteristics

After eligibility assessment, a total of 156 participants with 312 CT scans from the CAIRO5 trial were included in this study. The most common reason for exclusion was exceeding the limit of 15 CRLM per patient either at baseline or first follow-up CT examination (n = 121) (Fig 1). Other reasons for exclusion were missing baseline or first follow-up CT examination (n = 18), incomplete CT examination (n = 10), technical error in diagnostic software (n = 10), and the use of MRI (n = 44) or the use of PET with non–contrast-enhanced CT (n = 10). During the morphologic tumor response assessment, three additional patients were excluded because lesions were either too small to assess (n = 2) and/or calcification after chemotherapy led to unevaluable lesions (n = 1). Finally, a total of 153 participants (median age, 63 years [IQR, 56–71]; 101 men) with 306 CT scans comprising 2192 CRLM were included for the morphologic tumor response assessment. Baseline characteristics of the included patients are shown in Table 1. The majority of participants had a left-sided primary colon tumor (117 of 153, 76%) and synchronous metastases (127 of 153, 83%). The median number of CRLM at baseline was eight (IQR, five to 10), and 133 of 153 (87%) of the participants had a bilobar distribution of the CRLM. More than half of participants (81 of 153, 53%) had a RAS/BRAF mutation, and 122 of 153 (80%) received doublet or triplet chemotherapy in combination with bevacizumab.

Morphology Group Assignment by the Three Radiologists

Each of the three radiologists independently evaluated all 306 CT scans (153 obtained at baseline and 153 obtained at first follow-up), thereby assigning one single morphologic group to each CT scan. The distribution of the morphology group assignments of the individual radiologists is depicted in Figure 3. The overall interobserver agreement among the three radiologists showed a moderate agreement (κ = 0.53, 95% CI: 0.48, 0.58) (Table 2). Regarding the single morphology groups, a good agreement for morphology group 1, a fair agreement for morphology group 2, and a good agreement
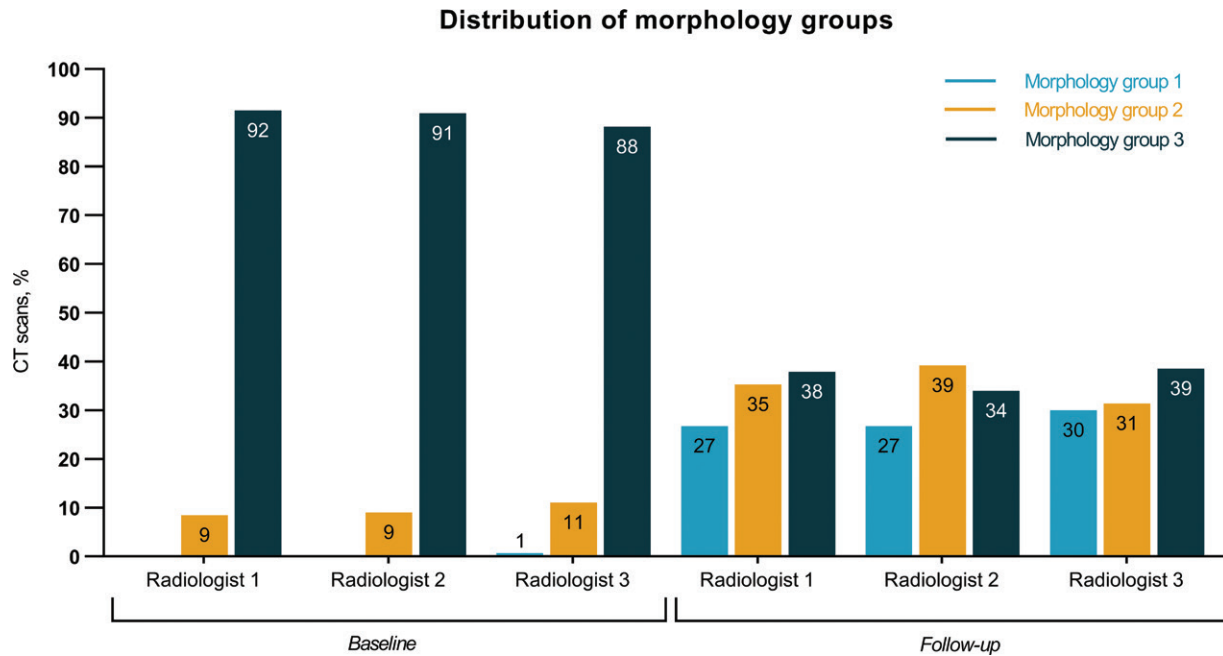
Distribution of morphology groups

Figure 3: Distribution of the morphology group assignments according to baseline CT scan (*n* = 153) and follow-up CT scan (*n* = 153) performed by the three radiologists.

for morphology group 3 were observed, with κ = 0.70 (95% CI: 0.64, 0.77), κ = 0.31 (95% CI: 0.25, 0.38), and κ = 0.61 (95% CI: 0.54, 0.67), respectively (Table 2). In addition, the overall interobserver agreement on the three different morphology groups among the individual radiologists varied between moderate and good (Table 3).

Morphologic Tumor Response Assessment by the Three Radiologists

The morphologic tumor response assessment was performed in 153 participants on the basis of the assigned morphology groups of the baseline and first follow-up CT scans. In most participants, the radiologists did not observe a mixed response in morphologic tumor response assessment within the same person. However, a mixed response was observed in some participants (11 of 153 [7%]) with small lesions, as these lesions were difficult to assign to one of the morphology groups. These mixed response cases were all considered as disagreement cases and were reevaluated by the two additional radiologists. The distribution of the morphologic tumor response assessments of the individual radiologists is shown in Figure 4. Participants were most frequently classified as nonresponders by all three radiologists (59 of 153 [39%], 63 of 153 [41%], and 65 of 153 [42%] participants). The overall interobserver agreement showed moderate agreement (κ = 0.54, 95% CI: 0.47, 0.60) (Table 2). Focusing on the single morphologic response categories, good agreement was observed for optimal responders, fair agreement was observed for suboptimal responders, and good agreement was observed for nonresponders, with κ = 0.66 (95% CI: 0.57, 0.75), κ = 0.33 (95% CI: 0.23, 0.42), and κ = 0.62 (95% CI: 0.53, 0.71), respectively (Table 2). In addition, the overall in-

Table 2: Overall Interobserver Agreement on Morphology Groups and Morphologic Response Categories among Three Radiologists

Parameter	Value
Morphology group	
Group 1	0.70 (0.64, 0.77)
Group 2	0.31 (0.25, 0.38)
Group 3	0.60 (0.54, 0.67)
Overall agreement	0.53 (0.48, 0.58)
Morphologic response	
Optimal response	0.66 (0.57, 0.75)
Suboptimal response	0.33 (0.23, 0.42)
No response	0.62 (0.53, 0.71)
Overall agreement	0.54 (0.47, 0.60)

Note.—Values are shown as κ statistics with 95% CIs in parentheses (lower bound, upper bound). Morphology groups: 1 = homogeneous and hypoattenuating attenuation, sharp defined tumor–liver interface, and (if initially present) complete disappearance of peripheral rim of enhancement; 2 = mixed attenuation, variable tumor–liver interface, and (if initially present) partially resolved rim enhancement; and 3 = heterogeneous attenuation, ill-defined tumor–liver interface, and peripheral rim of enhancement may be present. Morphologic responses: optimal response = change from group 2 or 3 to 1; suboptimal response = change from group 3 to 2; and no response = no change or change to higher group (eg, 1 to 2).

terobserver agreement on the different morphologic response groups among individual radiologists varied between moderate and good (Table 3).

Table 3: Overall Interobserver Agreement among Individual Radiologists

Parameter	Radiologist 1	Radiologist 2	Radiologist 3
Morphology group			
Radiologist 1	…	0.61 (0.53, 0.70)	0.42 (0.33, 0.50)
Radiologist 2	0.61 (0.53, 0.67)	…	0.56 (0.47, 0.64)
Radiologist 3	0.42 (0.33, 0.50)	0.56 (0.47, 0.64)	…
Morphologic response			
Radiologist 1	…	0.63 (0.52, 0.75)	0.43 (0.32, 0.55)
Radiologist 2	0.63 (0.52, 0.75)	…	0.54 (0.43, 0.66)
Radiologist 3	0.43 (0.32, 0.55)	0.54 (0.43, 0.66)	…

Note.—Values are shown as κ statistics with 95% CIs in parentheses (lower bound, upper bound).

Final Morphologic Tumor Response Assessment

The morphologic tumor response assessment was performed in 153 participants and resulted in 86 (56%) agreement cases and 67 (44%) disagreement cases (Fig E1 [supplement]). Disagreement cases could be subdivided into 63 minor disagreement cases and four major disagreement cases. In three of the four major disagreement cases, all three radiologists each assigned a different morphologic response category to the same participant. Subsequently, all 67 disagreement cases were evaluated by two additional radiologists (Fig 5). Figure 6 depicts the distribution of the morphologic tumor response assessment of the disagreement cases performed by the five radiologists. Based on the majority of votes, 58 cases could be solved, but nine cases remained unsolved. These nine cases were reassessed during a consensus meeting with two of the three initial radiologists. Of 153 total participants, CRLM of 37 (24%) were classified as optimal morphologic response, CRLM of 45 (29%) were classified as suboptimal morphologic response, and CRLMs of 71 (46%) were classified as no response.

Clinical Parameters and Morphologic Tumor Response

Gene mutation status and targeted therapy were compared among the different morphologic response groups (Table 4). The majority of participants with optimal and suboptimal morphologic response had a *RAS/BRAF* mutation (22 of 37 [59%] and 29 of 45 [64%], respectively), whereas the majority of participants (41 of 71 [58%]) with no morphologic response had a wild-type *RAS/BRAF* (*P* = .04). Targeted therapy also differed between the morphologic response groups (*P* = .001). In the optimal morphologic response group, 35 of 37 (95%) participants were treated with chemotherapy in combination with bevacizumab. Of the participants treated with panitumumab, the majority (23 of 31 [74%]) showed no morphologic response. In participants treated with chemotherapy in combination with panitumumab, CRLM were classified as optimal morphologic response in two participants and as suboptimal morphologic response in six participants.

RECIST 1.1 and Morphologic Tumor Response

Tumor response assessment according to RECIST 1.1 was compared with the different morphologic response groups (Table 4). According to RECIST 1.1, CRLM were classi-fied as having an objective response to treatment, stable disease, or progression of disease in 68 of 153 (44%), 83 of 153 (54%), and two of 153 (1%) participants, respectively. Of all participants classified as optimal morphologic responders using the morphologic response criteria, 17 of 37 (46%) were classified as response by RECIST 1.1, 19 of 37 (51%) as stable disease by RECIST 1.1, and one of 37 (3%) as progression by RECIST 1.1. Of all participants classified as suboptimal morphologic responders, 18 of 45 (40%) were classified as response by RECIST 1.1, 26 of 45 (58%) were classified as stable disease by RECIST 1.1, and one of 45 (2%) as progression by RECIST 1.1. Of all participants classified as nonresponders using the morphologic response criteria, 33 of 71 (46%) were classified as response by RECIST 1.1 and 38 of 71 (54%) as stable disease by RECIST 1.1. We found no evidence of a relation between the morphologic response groups and the RECIST 1.1 response groups (*P* = .61).

Discussion

CRLM not only tend to reduce in size, but also undergo morphologic changes in response to systemic therapy (17). Morphologic tumor response assessment, however, is based on visual observations and hence depends on observer perception. This study showed overall moderate agreement between abdominal radiologists in the assessment of morphologic tumor response on CT scans of patients with initially unresectable CRLM treated with systemic therapy. Disagreement in morphologic response assessment was observed in 67 of 153 (44%) participants, including four major disagreement cases. Moderate agreement among three radiologists was found for both morphology group and morphologic response assignments (respectively, κ = 0.53, 95% CI: 0.48, 0.58 and κ = 0.54, 95% CI: 0.47, 0.60). Moreover, the evaluation of the disagreement cases showed high variability in the assignments of the morphologic response categories within the same patients.

The study of Chun et al (17) reported a good interobserver agreement between the radiologists using the morphologic criteria in the assessment of morphologic tumor response to neoadjuvant systemic therapy. Two other studies assessed the level
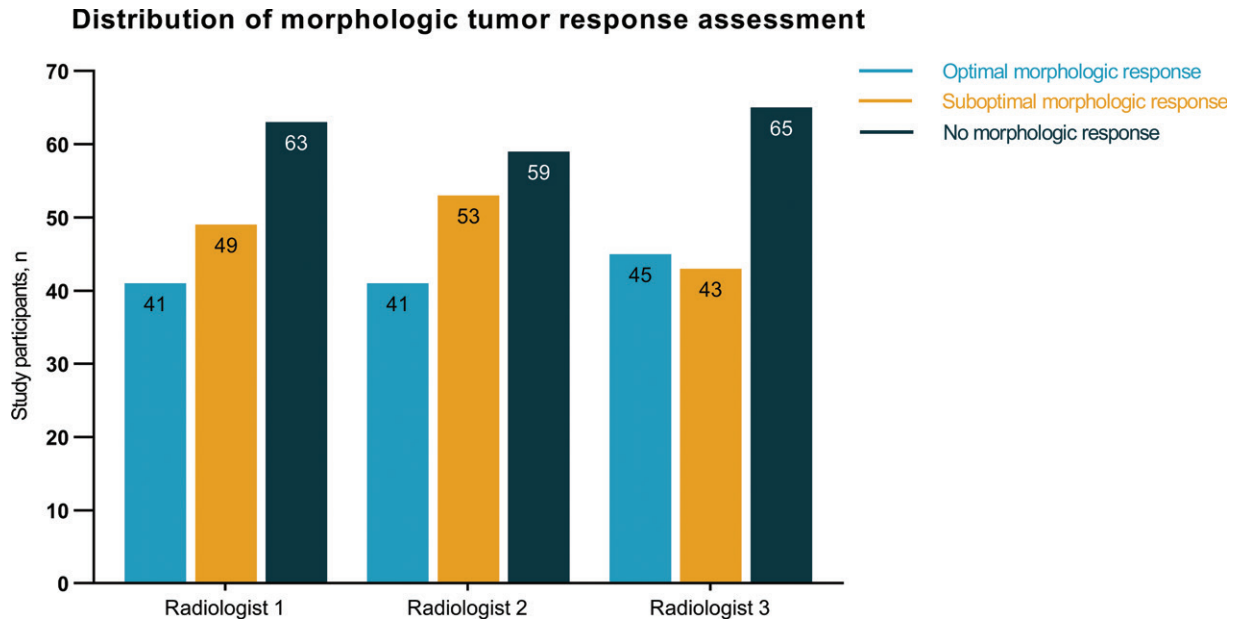
Distribution of morphologic tumor response assessment

Figure 4: Distribution of the morphologic tumor response assessments (*n* = 153) performed by the three radiologists.

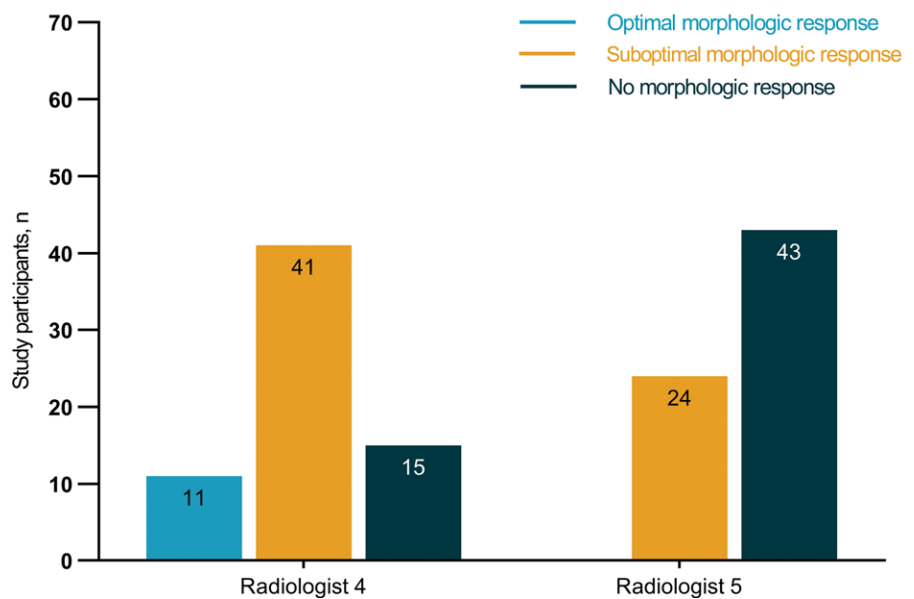## Distribution of morphologic tumor response assessment

Figure 5: Distribution of the morphologic tumor response assessments of the disagreement cases (*n* = 67) performed by the additional two radiologists.

of agreement of combined morphologic categories (17,20,25). However, all other studies investigating morphologic tumor response in patients with CRLM did not assess nor report the level of agreement between observers (18,21–23). Although promising results about the predictive value of this morphologic tumor response categorization at CT for overall survival and progression-free survival have been reported in previous studies, our data suggest that radiologists or physicians should be careful when implementing the proposed morphologic criteria in future studies (18,20–23).

High variability in our study could be explained by the lack of a clear definition of the morphology groups, as described by Chun et al (17). Particularly, morphology group 2 is imprecisely defined, using phrasings such as "mixed overall attenuation" and "variable liver interface." The present study supports this assumption, as the interobserver agreement on both the single morphology groups and morphologic response categories showed fair agreement ($\kappa < 0.41$) for morphology group 2 and suboptimal response, whereas good agreement ($\kappa > 0.61$) was observed for morphology groups 1 and 3 and for both optimal response and no morphologic tumor response.

To reduce variability, a more precise definition of the morphology response groups is warranted. Moreover, one might consider combining suboptimal and no responders to form one category and the optimal responders to form the other category. These proposed merged response groups have already been applied in two studies showing very good and moderate agreement among the observers (20,25). In the clinical setting, however, it would be relevant to distinguish the morphologic responders from the nonmorphologic responders.

Similar to other studies, our study showed that optimal morphologic response was particularly observed in participants

treated with chemotherapy in combination with bevacizumab, namely 35 of 37 (95%) participants (18,20,21). Optimal morphologic response was observed in two of 37 (5%) participants receiving chemotherapy plus panitumumab. Additionally, most participants with optimal morphologic response had a tumor with a *RAS/BRAF* mutation. Although these observations were statistically significant, results should be interpreted carefully, as

heterogeneity in location of the primary tumor, overall tumor burden, and chemotherapy regimen may all have affected the morphologic response. In addition, it should be noted that this study was not designed to investigate how genotype predisposes participants to a particular morphologic response.

The shift to optimal morphologic response may be caused by the mechanism of anti–vascular endothelial growth factor therapy, as inhibition of the vascular endothelial growth factor receptor leads to reduction of microvascular blood vessel growth and limits the blood supply to tumor tissue, eventually leading to necrosis (28). The change of heterogeneous attenuation to homogeneous hypoattenuating attenuation of CRLM likely reflects the replacement of tumor tissue to fibrotic tissue or tumor necrosis. Chun et al (17) showed that major pathologic response after systemic treatment was characterized by the replacement of tumor cells by fibrosis, rather than tumor necrosis in patients with CRLM. The relation between the morphologic changes at CT and pathologic tumor response should be further explored in future studies.

In the present study, we found no evidence of a relationship between RECIST 1.1 and morphologic tumor response. These findings are in line with the results of other studies and strengthen the hypothesis that use of the classic RECIST 1.1 might lead to an underestimation of the tumor response following systemic therapy, as morphologic changes of CRLM appear earlier
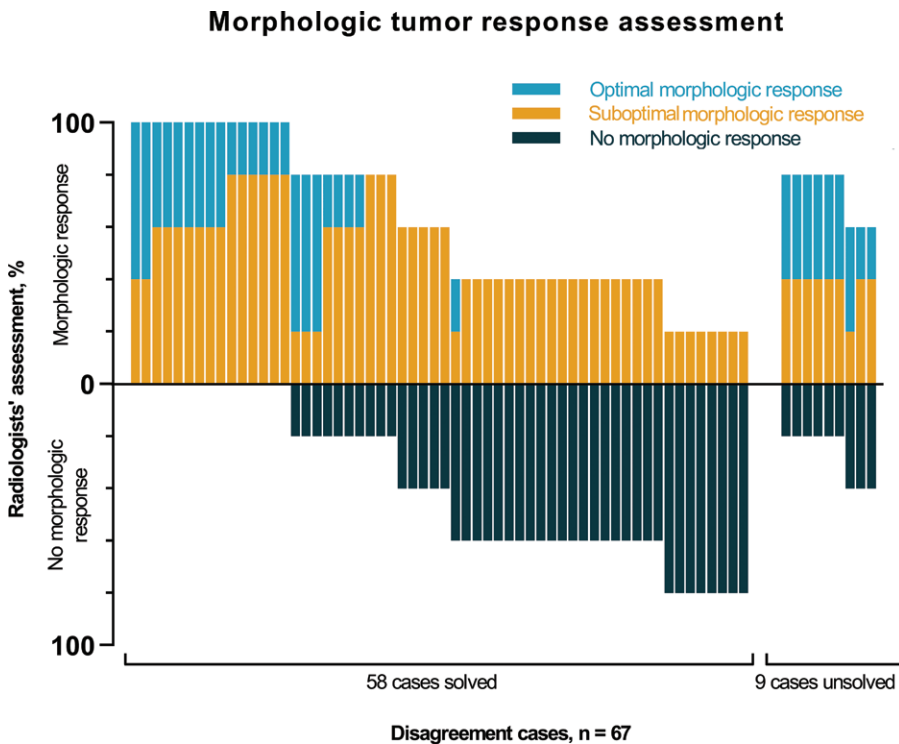

Figure 6: Bar chart depicts the morphologic tumor response assessment of the 67 disagreement cases of the three radiologists and the assessment by the two additional radiologists involved only in the disagreement cases' assessment. Each bar represents one participant. The chart depicts the percentages of radiologists' assessment for classifying participants as optimal morphologic response (blue), suboptimal morphologic response (orange), and no morphologic response (dark blue). Based on the majority of votes, 58 cases were solved and nine cases remained unsolved.

Table 4: Comparison of Clinical Parameters and RECIST 1.1 in Final Morphologic Tumor Response Groups

Parameter	Total Cohort (n = 153)	Optimal Morphologic Response (n = 37)	Suboptimal Morphologic Response (n = 45)	No Morphologic Response (n = 71)	P Value
Gene mutation status					.043
RAS/BRAF mutation	81 (53)	22 (60)	29 (64)	30 (42)	
RAS/BRAF wild-type	72 (47)	15 (41)	16 (36)	41 (58)	
Targeted therapy					.001
Anti-VEGF (bevacizumab)	122 (80)	35 (95)	39 (87)	48 (68)	
Anti-EGFR (panitumumab)	31 (20)	2 (5)	6 (13)	23 (32)	
RECIST 1.1					
Response	68 (44)	17 (46)	18 (40)	33 (47)	
Stable	83 (54)	19 (51)	26 (58)	38 (54)	
Progression	2 (1)	1 (3)	1 (2)	0 (0)	

Note.—Values are shown as numbers of participants with percentages in parentheses. EGFR = epidermal growth factor receptor, RECIST = Response Evaluation Criteria in Solid Tumors, VEGF = vascular endothelial growth factor.

than or without tumor lesion shrinkage (18,20,22). This study showed that in 19 of 83 (23%) participants with CRLM classified as stable disease by RECIST 1.1, an optimal morphologic response was observed. For these participants with CRLM, morphologic tumor response assessment could be of added prognostic value to RECIST 1.1 in helping to evaluate tumor response following systemic therapy. The direct clinical implications (ie, local treatment or no local treatment) for patients with borderline resectable and unresectable CRLM should be assessed in a prospective study.

There were some limitations within the present study that should be considered. First, the variation in CT scan quality across the 54 centers might have influenced study results. Because all scans were performed using different CT scanners and acquisition protocols, the quality of the CT scans varied. However, all scans were of adequate quality to be used for patient management. This variety of quality is a good representation of CT scans in daily practice and could be considered as a strength with respect to external validity. Additionally, when CRLM were too small, assessing lesions according to the proposed criteria was difficult and often resulted in cases being assigned to morphology group 3. The integration of size-based cutoff values might be needed to improve accuracy and reproducibility of morphologic tumor response assessment. Further, the selection of participants with initially unresectable CRLM might have influenced the results. As already reported by Dohan et al (25), the morphologic criteria might be more suitable for patients with primarily resectable CRLM rather than those with initially unresectable CRLM, as the former group often has fewer metastases. However, Dutch guidelines do not recommend neoadjuvant systemic treatment in patients with primarily resectable CRLM, and there is no international consensus on its use in these patients (7,29,30). Furthermore, it should be noted that the RECIST 1.1 assessment per participant was performed by one of the CAIRO5 radiologists (as part of the CAIRO5 protocol), whereas significant interobserver variability can be observed among RECIST 1.1 measurements (31,32). Even though the evaluation of RECIST 1.1 was not the main aim of this study, we still found it relevant to include this information. Finally, the relation between morphologic tumor response and histopathologic or survival outcomes could not be investigated yet, as the CAIRO5 trial is still recruiting patients. This is a future study objective to be explored when the CAIRO5 trial is completed.

Morphologic tumor response assessment could potentially be clinically relevant for patients with CRLM. However, the relevance and applicability of this assessment are difficult to determine because of interobserver variation. Apart from the possible improvement brought by a clearer definition of the morphology groups, an automated approach using artificial intelligence is expected to quantify changes in morphology with greater accuracy and reproducibility than the human eye. Automation of this assessment could eliminate the interobserver variation, allowing further exploration of its relevance and applicability in clinical practice. In future studies, the use of objective, standardized quantitative features to predict pathologic response to treatment in patients with CRLM should be advocated. With advanced analytics (eg, radiomics), hundreds of imaging features, including morphologic features, can be analyzed and used in predictive modeling through machine learning (33,34). One of these objective features could include CT attenuation, measured in Hounsfield units (35).

In summary, our findings demonstrated interobserver variability for morphologic tumor response assessment following systemic therapy in participants with CRLM. Results of the current study suggest that the proposed morphologic criteria should be used with caution in future studies or a clinical setting, and further research is needed to develop a more precise definition of morphologic tumor response, and a more objective, potentially automated approach of assessment.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019;69(1):7–34.
2. Elferink MAG, de Jong KP, Klaase JM, Siemerink EJ, de Wilt JHW. Metachronous metastases from colorectal cancer: a population-based study in North-East Netherlands. Int J Colorectal Dis 2015;30(2):205–212.
3. van der Geest LGM, Lam-Boer J, Koopman M, Verhoef C, Elferink MA, de Wilt JH. Nationwide trends in incidence, treatment and survival of colorectal cancer patients with synchronous metastases. Clin Exp Metastasis 2015;32(5):457–465.
4. Adam R, De Gramont A, Figueras J, et al. The oncosurgery approach to managing liver metastases from colorectal cancer: a multidisciplinary international consensus. Oncologist 2012;17(10):1225–1239.

5. de Ridder JAM, van der Stok EP, Mekenkamp LJ, et al. Management of liver metastases in colorectal cancer patients: a retrospective case-control study of systemic therapy versus liver resection. Eur J Cancer 2016;59:13–21.

6. Norén A, Eriksson HG, Olsson LI. Selection for surgery and survival of synchronous colorectal liver metastases; a nationwide study. Eur J Cancer 2016;53:105–114.

7. Adam R, Kitano Y. Multidisciplinary approach of liver metastases from colorectal cancer. Ann Gastroenterol Surg 2019;3(1):50–56.

8. Lam VW, Spiro C, Laurence JM, et al. A systematic review of clinical response and survival outcomes of downsizing systemic chemotherapy and rescue liver surgery in patients with initially unresectable colorectal liver metastases. Ann Surg Oncol 2012;19(4):1292–1301.

9. Abdalla EK, Vauthey JN, Ellis LM, et al. Recurrence and outcomes following hepatic resection, radiofrequency ablation, and combined resection/ablation for colorectal liver metastases. Ann Surg 2004;239(6):818–825; discussion 825–827.

10. Peeters M, Oliner KS, Price TJ, et al. Analysis of KRAS/NRAS mutations in a phase III study of panitumumab with FOLFIRI compared with FOLFIRI alone as second-line treatment for metastatic colorectal cancer. Clin Cancer Res 2015;21(24):5469–5479.

11. Eklöf V, Wikberg ML, Edin S, et al. The prognostic role of KRAS, BRAF, PIK3CA and PTEN in colorectal cancer. Br J Cancer 2013;108(10):2153–2163.

12. Peeters M, Kafatos G, Taylor A, et al. Prevalence of RAS mutations and individual variation patterns among patients with metastatic colorectal cancer: A pooled analysis of randomised controlled trials. Eur J Cancer 2015;51(13):1704–1713.

13. Loupakis F, Yang D, Yau L, et al. Primary tumor location as a prognostic factor in metastatic colorectal cancer. J Natl Cancer Inst 2015;107(3):dju427.

14. Petrelli F, Tomasello G, Borgonovo K, et al. Prognostic survival associated with left-sided vs right-sided colon cancer: a systematic review and meta-analysis. JAMA Oncol 2017;3(2):211–219.

15. Adam R, Delvart V, Pascal G, et al. Rescue surgery for unresectable colorectal liver metastases downstaged by chemotherapy: a model to predict long-term survival. Ann Surg 2004;240(4):644–657; discussion 657–658.

16. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45(2):228–247.

17. Chun YS, Vauthey JN, Boonsirikamchai P, et al. Association of computed tomography morphologic criteria with pathologic response and survival in patients treated with bevacizumab for colorectal liver metastases. JAMA 2009;302(21):2338–2344.

18. Shindoh J, Loyer EM, Kopetz S, et al. Optimal morphologic response to preoperative chemotherapy: an alternate outcome end point before resection of hepatic colorectal metastases. J Clin Oncol 2012;30(36):4566–4572.

19. Shindoh J, Chun YS, Loyer EM, Vauthey JN. Non-size-based response criteria to preoperative chemotherapy in patients with colorectal liver metastases: the morphologic response criteria. Curr Colorectal Cancer Rep 2013;9(2):198–202.

20. Nishioka Y, Shindoh J, Yoshioka R, et al. Radiological morphology of colorectal liver metastases after preoperative chemotherapy predicts tumor viability and postoperative outcomes. J Gastrointest Surg 2015;19(9):1653–1661.

21. Yoshita H, Hosokawa A, Ueda A, et al. Predictive value of optimal morphologic response to first-line chemotherapy in patients with colorectal liver metastases. Digestion 2014;89(1):43–48.

22. Mazard T, Boonsirikamchai P, Overman MJ, et al. Comparison of early radiological predictors of outcome in patients with colorectal cancer with unresectable hepatic metastases treated with bevacizumab. Gut 2018;67(6):1095–1102.

23. Primavesi F, Fadinger N, Biggel S, et al. Early response evaluation during preoperative chemotherapy for colorectal liver metastases: combined size and morphology-based criteria predict pathological response and survival after resection. J Surg Oncol 2019;121(2):382–391.

24. Benchoufi M, Matzner-Lober E, Molinari N, Jannot AS, Soyer P. Interobserver agreement issues in radiology. Diagn Interv Imaging 2020;101(10):639–641.

25. Dohan A, Gallix B, Guiu B, et al. Early evaluation using a radiomic signature of unresectable hepatic metastases to predict outcome in patients with colorectal cancer treated with FOLFIRI and bevacizumab. Gut 2020;69(3):531–539.

26. Huiskens J, van Gulik TM, van Lienden KP, et al. Treatment strategies in colorectal cancer patients with initially unresectable liver-only metastases, a study protocol of the randomised phase 3 CAIRO5 study of the Dutch Colorectal Cancer Group (DCCG). BMC Cancer 2015;15(1):365.

27. Altman DG. Practical statistics for medical research. London, England: Chapman & Hall, 1991.

28. Kazazi-Hyseni F, Beijnen JH, Schellens JHM. Bevacizumab. Oncologist 2010;15(8):819–825.

29. Adam R, de Gramont A, Figueras J, et al. Managing synchronous liver metastases from colorectal cancer: a multidisciplinary international consensus. Cancer Treat Rev 2015;41(9):729–741.

30. Lehmann K, Rickenbacher A, Weber A, Pestalozzi BC, Clavien PA. Chemotherapy before liver resection of colorectal metastases: friend or foe? Ann Surg 2012;255(2):237–247.

31. Beaumont H, Evans TL, Klifa C, et al. Discrepancies of assessments in a RECIST 1.1 phase II clinical trial - association between adjudication rate and variability in images and tumors selection. Cancer Imaging 2018;18(1):50.

32. Yoon SH, Kim KW, Goo JM, Kim DW, Hahn S. Observer variability in RECIST-based tumour burden measurements: a meta-analysis. Eur J Cancer 2016;53:5–15.

33. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14(12):749–762.

34. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48(4):441–446.

35. Froelich MF, Heinemann V, Sommer WH, et al. CT attenuation of liver metastases before targeted therapy is a prognostic factor of overall survival in colorectal cancer patients. Results from the randomised, open-label FIRE-3/AIO KRK0306 trial. Eur Radiol 2018;28(12):5284–5292.