Journal of Nuclear Medicine, published on June 29, 2018 as doi:10.2967/jnumed.118.210161

Metabolic Active Tumor Volume and Total Lesion Glycolysis by ^{18}F-FDG PET/CT Validated as Prognostic Imaging Biomarkers in Chemorefractory Metastatic Colorectal Cancer

Erwin Woff*[1], Alain Hendlisz[2], Lieveke Ameye[3], Camilo Garcia[1], Tarek Kamoun[1], Thomas Guiot[1], Marianne Paesmans[3], Patrick Flamen[1]

1 Nuclear Medicine Department, Institut Jules Bordet - Université Libre de Bruxelles (ULB), Brussels, Belgium.

2 Medical Oncology Department, Institut Jules Bordet - Université Libre de Bruxelles (ULB), Brussels, Belgium.

3 Data Centre, Institut Jules Bordet - Université Libre de Bruxelles (ULB), Brussels, Belgium.

Corresponding author: Patrick Flamen, MD, PhD

Nuclear Medicine Department,

Institut Jules Bordet - Université Libre de Bruxelles (ULB),

1 rue Héger-Bordet, 1000 Brussels, Belgium.

Tel: +322/5413717; Fax: +322/5413224; E-mail: patrick.flamen@bordet.be

First author: Erwin Woff, MD

Nuclear Medicine Department,

Institut Jules Bordet - Université Libre de Bruxelles (ULB),

1 rue Héger-Bordet, 1000 Brussels, Belgium.

Tel: +322/5417314; Fax: +322/5413224; E-mail: erwin.woff@bordet.be

Content and Word count: 5429; Abstract: 344; Text: 3420; Figures: 4; Tables: 3; Appendix: 5, References: 27

Running title: Prognostic Value of WB-MATV/TLG in mCRC

ABSTRACT

This study aimed to validate the prognostic value of baseline whole-body metabolic active tumor volume (WB-MATV) and total lesion glycolysis (WB-TLG) measured with [^{18}F]fluorodeoxyglucose positron emission tomography-computed tomography (^{18}F-FDG PET/CT) in a large cohort of chemorefractory metastatic colorectal cancer (mCRC) patients treated with multikinase inhibitors (MKI). The secondary objective of this study was to compare WB-MATV and WB-TLG respective prognostic values to commonly used clinical prognostic factors.

Methods

Out of 238 patients pooled from two successive prospective multicenter trials investigating MKI in chemorefractory mCRC, 224 were considered suitable for analysis. The patients were retrospectively randomly assigned to a development set (n = 155 patients) and a validation set (n = 69 patients). WB-MATV and WB-TLG optimal cutoffs for prediction of overall survival (OS) were determined by Contal and O'Quigley's method. Univariate analyses were performed to assess the prognostic values of WB-MATV and WB-TLG. Multivariate analyses were performed for WB-MATV and WB-TLG along with clinical factors to identify the independent prognostic factors of OS. The prognostic weight for each parameter was obtained from the Cox's model.

Results

WB-MATV and WB-TLG optimal cutoffs for OS prediction were 100 cm^3 and 500 g, respectively. Univariate analyses showed that WB-MATV and WB-TLG parameters were strongly related to outcome in both the development and validation sets. In the validation set, the median OS was 5.2 months vs 12.8 months for high vs low WB-MATV (hazard ratio [HR]: 3.12, $P < 0.001$), and 4.7 months vs 13.9 months for high vs low WB-TLG (HR: 3.67, $P < 0.001$). The multivariate analyses identified that both high WB-MATV and WB-TLG were independent

negative prognostic parameters for OS, with the highest prognostic weight among the well-known clinical prognostic factors (HR: 2.46 and 2.23, respectively, $P < 0.001$).

Conclusion

Baseline WB-MATV and WB-TLG parameters were validated as strong prognosticators of outcome in a large cohort of chemorefractory mCRC patients treated with MKI. These parameters were identified as independent prognostic imaging biomarkers with the highest prognostic values among the commonly used clinical factors. These biomarkers should therefore be used to support the optimal therapeutic strategy.

Keywords

FDG PET/CT; metabolic active tumor volume; total lesion glycolysis; metastatic colorectal cancer; multikinase inhibitors

INTRODUCTION

Colorectal cancer is the third leading cause of cancer death worldwide (*1*). Despite important improvements in the management of metastatic colorectal cancer (mCRC), the prognosis remains poor, with low 5-year survival rates (*1*). Nevertheless, there are wide variations in the overall survival (OS) of mCRC patients, and the factors explaining this heterogeneity in survival have not all been identified. So far, most of the independent prognostic factors that have been validated (e.g., Eastern Cooperative Oncology Group [ECOG] performance status [PS], age, body mass index [BMI], elevated levels of lactate dehydrogenase, and serum albumin) are related to the general medical condition of the patient and are not tumor-specific (*2-5*). Investigations of tumor-specific markers such as CEA, CA 19-9, pathological staging, and gene expression signatures have failed to accurately predict prognosis (*6,7*). Considering that biomarkers directly related to the tumor should be more specific than existing clinical factors, and could more accurately identify those patients at risk of shorter OS, there is an urgent need to investigate and validate new tumor-specific prognostic biomarkers.

Whole-body [18F]fluorodeoxyglucose positron emission tomography-computed tomography (^{18}F-FDG PET/CT) is now widely accepted as a powerful tool for the assessment and monitoring of oncologic disease (*8*). The volume-based metabolic parameters, whole-body metabolic active tumor volume (WB-MATV) and total lesion glycolysis (WB-TLG) have recently been studied as prognosticators of outcome in lymphoma (*9-11*) and some solid tumors (*12-14*). TLG was shown to be a predictor of outcome in mCRC in a recent study conducted on a small cohort of patients treated with regorafenib (*15*). However, this result should be confirmed and validated in a larger cohort of patients, and should also be tested with the MATV parameter; indeed, the respective prognostic values of TLG and MATV have never been compared in mCRC.

This study therefore aimed to validate baseline WB-MATV and WB-TLG as prognostic imaging biomarkers in a large cohort of chemorefractory mCRC patients included in two successive prospective trials with comparable inclusion criteria. The secondary objective of this study was to compare their respective prognostic values to commonly used clinical prognostic factors.

MATERIALS AND METHODS

Study Design and Participants

This retrospective pooled analysis measured WB-MATV and WB-TLG at baseline time point. The ^{18}F-FDG PET/CT data were extracted from two sequential prospective multicenter phase II non-randomized clinical trials investigating patients with unresectable chemorefractory mCRC treated with multikinase inhibitors (MKI). The SoMore study (EudraCT number: 2010-023695-91; NCT number: 01290926) investigated a combination of sorafenib (Nexavar, BAY 43-9006, Bayer Pharma AG, Berlin, Germany) and capecitabine (Xeloda, Roche Pharma, Basel, Switzerland), (*16*) while the RegARd-C study (EudraCT number: 2012-005655-16) investigated regorafenib (Stivarga, BAY 73-4506; Bayer Pharma AG, Berlin, Germany) (*17*). The main enrollment criteria for these two studies were: tumor refractory to all standard chemotherapy agents (fluoropyrimidines, irinotecan, and oxaliplatin) and anti-EGFR monoclonal antibodies in the case of RAS-wild type (cetuximab or panitumumab); age > 18; ECOG PS ≤ 1; life expectancy > 12 weeks; a baseline ^{18}F-FDG PET/CT exam performed within the 7 days previous to the day of inclusion in the trial with at least one measurable target lesion; ability to undergo the therapy; and provide signed informed consent (*17*). Both studies were conducted within the same Belgian hospital network and followed a similar study design, testing metabolic response after one course of treatment as a predictor of patient's outcome, with OS as the primary endpoint.

Twelve clinical centers and nine PET/CT centers, all located in Belgium, were involved in these trials. Each PET/CT center followed strict procedural guidelines for patient preparation and imaging (*18,19*). All but one center (n = 1 patient included) obtained EARL accreditation during the first trial. All examinations were performed with the locally available PET/CT model: General Electric Discovery 690 or LS (General Electric Company, Fairfield, Connecticut, USA), Philips Gemini TF (Philips, Amsterdam, the Netherlands), and Siemens Biograph 64 (Siemens, Munich, Germany). Quality assessment was assured by an independent dedicated academic PET/CT imaging core lab.

Approval from the institutional review board was obtained for this retrospective pooled analysis (CE2616), and all patients signed a written informed consent form for this study. All imaging data was anonymized.

^{18}F-FDG PET/CT Image Acquisition and Image Analysis Procedures

A quality control analysis was applied to all ^{18}F-FDG PET/CT. All patients who had not fasted for at least 6 h before FDG injection were rejected, as well as all exams with a delay between FDG injection and scanning outside the range of 55–75 minutes, and/or glycemia > 150 mg/dL at the time of FDG injection. All PET scans were acquired from skull to mid-thigh in three-dimensional mode with an acquisition time of 90 seconds per bed position. The PET images were corrected for attenuation and scatter using the data from the unenhanced low-dose CT.

Two experienced nuclear medicine physicians (EW, TK) assessed WB-MATV and WB-TLG for a set of 100 patients randomly selected from the two studies. In cases of discrepancy between the two observers of WB-MATV or WB-TLG (WB-MATV/TLG) values (defined as a > 10% absolute difference in WB-MATV/TLG values) implying or not a change of category between low and high tumor load, a consensus was reached by a third experienced physician

(PF). As there was good reproducibility in the WB-MATV/TLG measurements between the two observers (EW, TK), the WB-MATV/TLG of the remaining patients were assessed by the more experienced nuclear medicine physician (EW).

All nuclear medicine physicians involved in this study were blinded to the medical records and treatment outcomes. All WB-MATV/TLG measurements were computed on a dedicated workstation (Advantage Workstation; General Electric Company, Fairfield, Connecticut, USA) using the commercial PETVCAR 4.6 software, and were normalized to lean body mass. Target lesions were defined as follows: an unequivocal tumor origin, transverse diameter > 15 mm on a registered CT image, and an FDG standardized uptake value normalized to lean body mass (SUL) higher than 1.5 × the mean liver SUL + 2 × SD, or in the presence of liver metastasis, 2.0 × mean aorta SUL + 3 × SD, following PERCIST recommendations with a minor adaptation (3 SD instead of 2 SD in order to have comparable thresholds between liver and aorta reference background activities) (*20*). A volume of interest was drawn on each target lesion using segmentation with a fixed absolute threshold calculated from the patient's background liver or blood pool activity, as described above.

The MATV of a lesion was defined as the volume of tumor tissue demonstrating metabolic activity at or above the calculated threshold. TLG was calculated as MATV multiplied by SUL_{mean} ($TLG = MATV \times SUL_{mean}$). WB-MATV and WB-TLG were calculated as the sum of the MATV or TLG values of all target lesions, without a predefined limitation on their number. $WB\text{-}SUL_{mean}$ was extracted from the WB-TLG formula as follows: $WB\text{-}SUL_{mean}$ = WB-TLG / WB-MATV.

Statistical Analysis

The baseline clinical characteristics and survival data were collected prospectively and measured from the date of inclusion in the trials to death from any cause. The patients alive at last follow-up were censored.

Two-thirds of the patients remaining after application of the inclusion criteria (n = 224) were randomly assigned to a development set (n = 155 patients) to define the optimal WB-MATV, WB-TLG, and WB-SUL_{mean} cutoff values, while the other one-third were assigned to a validation set (n = 69 patients) to validate these cutoff values. Patients were stratified by medication (sorafenib/regorafenib), BMI ($\geq$ 25 vs < 25 kg/m^2), and documentation of progression (radiological or not).

Contal and O'Quigley's method was used to determine the optimal WB-MATV, WB-TLG, and WB-SUL_{mean} cutoff values for prediction of survival in the development set, with these values then being tested on the validation set (*21*). Survival analyses were performed using the Kaplan-Meier method and the Cox's proportional hazards model to estimate the hazard ratio (HR) with 95% confidence intervals (CI). A log-rank test was then performed to compare OS between groups. In the multivariate Cox's model, the following variables were considered for association with OS: WB-MATV, WB-TLG, age, gender, BMI, ECOG PS, number of years between diagnosis and inclusion in the respective trial, KRAS mutation status, medication (sorafenib vs regorafenib), and prior use of bevacizumab. The prognostic weight for each parameter was obtained from the Cox's model by dividing its estimate by the estimate in absolute value of the parameter with the smallest value. The obtained value was then rounded. P-values < 0.05 were considered as statistically significant. Interobserver agreement was assessed using the Cohen κ statistic (*22*). Inter-observer agreements in WB-MATV measurements, WB-TLG

measurements, and segmentation thresholds were represented on Bland-Altman plots. Statistical analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC) and Bland-Altman plots using GraphPad Prism 7 (GraphPad Software, La Jolla, CA).

RESULTS

A total of 238 mCRC patients were included in this pooled analysis: 97 from the SoMore trial and 141 from the RegARd-C trial. Seven patients from each trial were excluded for one of the following reasons: patients who did not meet the inclusion criteria (n = 4 in SoMore and n = 3 in RegARd-C), declined to participate (n = 1 in SoMore), had a too short follow-up (4 days; n = 1 in RegARd-C), or did not have baseline WB-MATV/TLG measurement due to absence of a target lesion or major violation to the imaging protocol (n = 2 in SoMore and n = 3 in RegARd-C). This left 224 patients who were considered suitable for WB-MATV/TLG measurements (Fig. 1).

The patient characteristics of the pooled population are summarized in Table 1. The survival characteristics of the pooled population, and of the development and validation sets, are summarized in Supplemental Table 1.

Baseline WB-MATV, WB-TLG, and WB-SUL$_{mean}$

The median baseline WB-MATV, WB-TLG, and WB-SUL$_{mean}$ in the development set were 166 cm^3 (5–95% percentile, 11 to 1524), 720 g (5–95% percentile, 31 to 7334), and 4.4 g/mL (5–95% percentile, 3.1 to 6.1), respectively. The optimal baseline WB-MATV, WB-TLG, and WB-SUL$_{mean}$ thresholds associated with OS were determined to be 100 cm^3, 500 g, and 4.5 g/mL, respectively. WB-MATV and WB-TLG measurements were highly correlated, with a Spearman correlation of 0.982 (95% CI, 0.976–0.987, P < 0.001). The contingency table for WB-

MATV and WB-TLG measurements after applying their respective cutoff values for categorization into low or high tumor load showed only 12/155 (8%) discrepancy cases in the development set (Supplemental Table 2).

The baseline WB-MATV/TLG measurements were highly reproducible between the two observers, with a substantial overall agreement in the categorization of patients (κ, 0.80) (*23*). Discrepancies between categorization (low versus high tumor load) were observed in only 4% (4/100) and 5% (5/100) of patients according to WB-MATV and WB-TLG parameters respectively (Supplemental Table 3). Discrepancies between values were observed for 23% (23/100) of the patients, with 87% (20/23) of these discrepancies being due to differences in the selection of target lesions (including all the discrepancy cases where the two observers assigned the patient to a different category), and 13% (3/23) being due to differences in the placement of the reference volume of interest. A consensus was achieved by recourse to a third reader for all these discrepancies. The Bland-Altman plots did not reveal any bias between the two observers in WB-MATV and WB-TLG measurements, or in the calculated segmentation thresholds (Supplemental Fig. 1).

Correlation between Baseline WB-MATV and OS

In the development set, patients with a high baseline WB-MATV (≥ 100 cm^3) had a significantly worse outcome, with a median OS of 4.5 months (95% CI, 3.5–5.7) versus 10.9 months (95% CI, 9.4–13.9) for patients with a lower WB-MATV (< 100 cm^3) (HR WB-MATV ≥ 100 vs < 100 cm^3, 2.64; 95% CI, 1.87–3.73; $P < 0.001$; Fig. 2A).

These results were confirmed in the validation set: patients with a high baseline WB-MATV had a significantly worse outcome, with a median OS of 5.2 months (95% CI, 3.7–7.5) versus 12.8 months (95% CI, 8.1–17.6) for patients with a lower WB-MATV (HR WB-MATV ≥

100 vs < 100 cm^3, 3.12; 95% CI, 1.77–5.50; $P < 0.001$; Fig. 2B). Examples of patients with a low and a high baseline WB-MATV/TLG are shown in Figure 3.

Correlation between Baseline WB-TLG and OS

In the development set, patients with a high baseline WB-TLG (≥ 500 g) had a significantly worse outcome, with a median OS of 4.7 months (95% CI, 3.5–5.7) versus 10.5 months (95% CI, 8.7–13.4) for patients with a lower WB-TLG (< 500 g) (HR WB-TLG ≥ 500 vs < 500 g, 2.16; 95% CI, 1.54–3.02; $P < 0.001$; Fig. 4A).

These results were confirmed in the validation set: patients with a high baseline WB-TLG had a significantly worse outcome, with a median OS of 4.7 months (95% CI, 3.4–6.8) versus 13.9 months (95% CI, 8.9–19.9) for patients with a lower WB-TLG (HR WB-TLG ≥ 500 vs < 500 g, 3.67; 95% CI, 2.07–6.50; $P < 0.001$; Fig. 4B).

Correlation between Baseline WB-SUL_{mean} and OS

There was no statistical correlation found between baseline WB-SUL_{mean} and OS in the development set (HR WB-SUL_{mean} ≥ 4.5 vs < 4.5 g/mL, 0.89; 95% CI, 0.65–1.24; $P = 0.50$). Therefore, this parameter was not evaluated in the validation set, nor was it included in the multivariate analyses.

Identification of Independent Predictors of OS among the PET and Clinical Parameters

The multivariate analysis identified high baseline WB-MATV as a significant independent predictor of OS (HR, 2.46; $P < 0.001$), together with the following clinical parameters: number of years since diagnosis (HR, 0.86 per 1-year increase; $P < 0.001$), ECOG PS 1 (HR, 1.67; $P = 0.003$), and BMI ≥ 25 (HR, 0.56; $P < 0.001$). The prognostic weights of the parameters were 6, −1, 3, and −4, respectively (Table 2).

Similarly, the multivariate analysis performed with WB-TLG also showed that high baseline WB-TLG was a significant independent predictor of OS (HR, 2.23; $P < 0.001$), together with the same clinical parameters as for WB-MATV: number of years since diagnosis (HR, 0.86 per 1-year increase; $P < 0.001$), ECOG PS 1 (HR, 1.81; $P < 0.001$), and BMI ≥ 25 (HR, 0.53; $P < 0.001$). The prognostic weights of the parameters were 5, −1, 4, and −4, respectively (Table 3).

DISCUSSION

In this pooled analysis of two prospective multicenter studies investigating a large cohort of chemorefractory mCRC patients, baseline WB-MATV and WB-TLG were validated as robust pre-treatment predictors of OS. Patients with a high WB-MATV or WB-TLG (≥ 100 cm^3 or ≥ 500 g, respectively) had a significantly worse clinical outcome than patients with a low WB-MATV or WB-TLG.

To the best of our knowledge, neither the WB-MATV nor the WB-TLG parameter have been validated as prognostic biomarkers in mCRC patients. The TLG parameter was recently investigated alongside other metabolic parameters in an exploratory cohort of 40 mCRC patients treated with regorafenib (*15*). Patients with a $TLG_{40\%}$ (TLG with a segmented threshold fixed at 40% of the lesion's SUV_{max} investigated on one or several target lesions) lower than the median $TLG_{40\%}$ value had a significantly longer median OS (14.2 months) than patients with a $TLG_{40\%}$ above the median (9.1 months) (*15*). These results are in complete accordance with our findings for both the WB-MATV and WB-TLG parameters investigated in our development and validation cohorts.

In this study, the WB-MATV and WB-TLG parameters were strongly consistent for categorizing patients into high or low tumor load according to their respective cutoff values. Only few discrepancy cases were found in the development set (8%). These cases occurred when the

WB-MATV or WB-TLG values were close to the cutoff. As multivariate analyses showed that the prognostic weight for WB-MATV was slightly higher than the one for WB-TLG (6 versus 5), we recommend choosing WB-MATV value in case of such a discrepancy.

The WB-MATV and WB-TLG parameters also did not have any clinically relevant difference in terms of outcome prediction. The difference between these two volume PET-based parameters is the inclusion of the SUV_{mean} in the TLG formula. The SUL_{mean} parameter (corresponding to SUV_{mean} normalized to lean body mass) investigated in our study did not show any statistical association with OS. Therefore, in our study, the MATV component of the TLG formula had a dominant impact on the prognostic value of the TLG parameter. However, this does not mean that the ^{18}F-FDG uptake intensity of the lesions, whether expressed as SUV_{mean} or SUL_{mean}, has no prognostic importance. Because of our chosen methodology for the selection of target lesions, lesions with a low ^{18}F-FDG uptake (below the PERCIST-based segmenting threshold) were not considered as target lesions, and therefore did not contribute to the WB-MATV/TLG. Therefore, it can only be concluded that differences in ^{18}F-FDG uptake intensity were not significantly related to outcome within the selected target lesions, which all showed intense ^{18}F-FDG uptake.

The WB-MATV and WB-TLG parameters were analyzed among well-known clinical prognostic factors in mCRC, and both were identified as strong independent predictors of OS with a significantly higher prognostic weight than the clinical factors. Unlike most of the currently accepted non-tumor-specific prognostic factors in mCRC such as BMI and ECOG PS, which all primarily represent the general medical condition of the patient, the WB-MATV and WB-TLG parameters represent the viable and aggressive tumor load, and are thus less influenced by non-tumoral factors.

These prognostic biomarkers accurately identify patients with a high or low risk of shorter OS, and could be particularly useful for determining the optimal therapeutic strategy. In daily clinical practice, knowledge of the estimated prognosis provides an opportunity for the oncologist and patient to reconsider the current risk-benefit balance, before initiating a novel line of therapy, especially as available therapeutic agents such as MKI are associated with a high toxicity and limited efficacy (*24*,*25*).

In this study, several factors were taken into consideration to allow the most precise and reliable WB-MATV/TLG measurements. To minimize the inclusion of false-positive non-tumoral (e.g., inflammatory) lesions, which would lead to overestimation of tumor load, the observers were selected for their experience in interpreting oncologic ^{18}F-FDG PET/CT images. For the selection of target lesions, observers were recommended to be more specific than sensitive, excluding all hypermetabolic lesions with uncertain origin. A fixed threshold for tumor delineation was determined for each patient based on the patient's background blood pool or liver FDG-activity, according to the PERCIST methodology, and not a threshold of 41% of the lesion SUV_{max} as recommended by the EANM (*18*). Such a fixed threshold relative to background applied to every target lesion of a patient limits the overestimation of the tumor volume in cases of low or moderately active lesions. This thresholding method also renders the delineation process easier to use in cases of multiple metastases.

In terms of interobserver variability, this analysis showed very good agreements on WB-MATV/TLG measurements (κ, 0.80) in a large subset of patients (n = 100). The two observers assigned patients to a different category in 4% and 5% of cases for WB-MATV and WB-TLG respectively. The Bland-Altman plots revealed no systemic bias in the WB-MATV and WB-TLG measurements. Discrepancies in WB-MATV or WB-TLG values between the two observers were mainly due to random errors such as the selection of target lesions as illustrated in Supplemental

Figure 2. The different placement of the reference volume of interest between observers was not shown to be a major cause of measurement variability (13% of our discrepancy cases between values, and no discrepancy case in the categorization of patients), which is in line with earlier reports (*26,27*).

Whether WB-MATV/TLG is purely prognostic or whether it can act as a predictive biomarker for MKI response remains uncertain. It would require a prospective trial with a treatment control arm with similar mCRC patients not treated with MKI to verify whether the treatment effect is identical in both high and low WB-MATV/TLG groups. In the meantime, further ongoing analysis on the current dataset will verify whether early metabolic (with PERCIST) and late morphologic (with RECIST) response rates are related to the tumor load.

Potential limitations of this study were that WB-MATV/TLG were not validated in an independent dataset and that the pooled cohort developed for this study used the data of two prospective clinical trials investigating mCRC patients treated with different targeted agents (sorafenib and regorafenib). However, as the multivariate analyses did not identify the use of regorafenib versus sorafenib as an independent prognostic factor, the results of this study can reliably be extrapolated to mCRC patients treated with currently used targeted agents.

To conclude, this study validated baseline WB-MATV and WB-TLG parameters as strong independent predictors of OS in a large cohort of chemorefractory mCRC patients treated with MKI. When compared with well-known clinical prognostic factors in mCRC, these parameters were shown to have the highest prognostic values. On this basis, the authors would recommend the use of one of these two imaging biomarkers to define the optimal care for chemorefractory mCRC patients.

Disclosure

Bayer Healthcare AG provided sorafenib and regorafenib, and a research grant for the SoMore and RegARd-C trials, but played no further role in the design and conduct of the study, data collection, management, analysis, and interpretation of the data, or preparation, review, or approval of the manuscript. The authors have no other conflicts of interest to disclose.

Acknowledgments

The authors thank all physicians who participated in the recruitment of patients for the SoMore and RegARd-C studies, notably, Amelie Deleporte, MD, Karen Geboes, MD, PhD, Thierry Delaunoit, MD, Gauthier Demolin, MD, Marc Peeters, MD, PhD, Lionel D'Hondt, MD, PhD, Jos Janssens, MD, Javier Carrasco, MD, Stéphane Holbrechts, MD, Jean-Charles Goeminne, MD, PhD, Jean-Luc Van Laethem, MD, PhD, and Philippe Vergauwe, MD; and Hugo Levillain, MSc, for his valuable advice on Bland-Altman plots.

The authors express their deep gratitude to the patients and their families who agreed to participate in this project. The authors thank all staff involved in the nine PET/CT centers who participated in this study. This academic work was supported and sponsored by the Jules Bordet Institute.

REFERENCES:

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2014;136:359–386.

2. Renfro LA, Loupakis F, Adams RA, et al. Body mass index is prognostic in metastatic colorectal cancer: pooled analysis of patients from first-line clinical trials in the ARCAD database. *J Clin Oncol*. 2016;34:144–150.

3. Lieu CH, Renfro LA, de Gramont A, et al. Association of age with survival in patients with metastatic colorectal cancer: analysis from the ARCAD clinical trials program. *J Clin Oncol*. 2014;32:2975–2984.

4. Shibutani M, Maeda K, Nagahara H, et al. The pretreatment albumin to globulin ratio predicts chemotherapeutic outcomes in patients with unresectable metastatic colorectal cancer. *BMC Cancer*. 2015;15:347.

5. Pietrantonio F, Miceli R, Rimassa L, et al. Estimating 12-weeks death probability in patients with refractory metastatic colorectal cancer: the colon life nomogram. *Ann Oncol*. 2017;28:555-561.

6. Kishiki T, Masaki T, Matsuoka H, et al. Modified Glasgow prognostic score in patients with incurable stage IV colorectal cancer. *Am J Surg*. 2013;206:234–240.

7. Van Cutsem E, Cervantes A, Adam R, et al. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann Oncol*. 2016;27:1386–1422.

8. Laurens ST, Oyen WJG. Impact of fluorodeoxyglucose PET/computed tomography on the management of patients with colorectal cancer. *PET Clin*. 2015;10:345–360.

9. Meignan M, Cottereau AS, Versari A, et al. Baseline metabolic tumor volume predicts outcome in high-tumor-burden follicular lymphoma: a pooled analysis of three multicenter studies. *J Clin Oncol*. 2016; 34:3618–3626.

10. Sasanelli M, Meignan M, Haioun C, et al. Pretherapy metabolic tumour volume is an independent predictor of outcome in patients with diffuse large B-cell lymphoma. *Eur J Nucl Med Mol Imaging*. 2014;41:2017–2022.

11. Cottereau AS, Becker S, Broussais F, et al. Prognostic value of baseline total metabolic tumor volume (TMTV0) measured on FDG-PET/CT in patients with peripheral T-cell lymphoma (PTCL). *Ann Oncol*. 2016;27:719–724.

12. Bazan JG, Duan F, Snyder BS, et al. Metabolic tumor volume predicts overall survival and local control in patients with stage III non-small cell lung cancer treated in ACRIN 6668/RTOG 0235. *Eur J Nucl Med Mol Imaging*. 2017;44:17–24.

13. Van de Wiele C, Kruse V, Smeets P, Sathekge M, Maes A. Predictive and prognostic value of metabolic tumour volume and total lesion glycolysis in solid tumours. *Eur J Nucl Med Mol Imaging*. 2013;40:290–301.

14. Gauthé M, Richard-Molard M, Fayard J, Alberini J-L, Cacheux W, Lièvre A. Prognostic impact of tumour burden assessed by metabolic tumour volume on FDG PET/CT in anal canal cancer. *Eur J Nucl Med Mol Imaging*. 2017;44:63–70.

15. Lim Y, Bang J-I, Han S-W, et al. Total lesion glycolysis (TLG) as an imaging biomarker in metastatic colorectal cancer patients treated with regorafenib. *Eur J Nucl Med Mol Imaging*. 2017;44:757–764.

16. Deleporte A, Hendlisz A, Garcia C, Delaunoit T. SoMore trial: Early metabolic response assessment of a sorafenib (SOR) and capecitabine (CAP) combination in chemorefractory metastatic colorectal cancer (mCRC). *J Clin Oncol*. 2014;32(suppl 3):524.

17. Hendlisz A, Deleporte A, Vandeputte C, et al. Regorafenib assessment in refractory advanced colorectal cancer: RegARd-C study protocol. *BMJ Open*. 2015;5:e007189.

18. Boellaard R, Delgado-Bolton R, Oyen WJG, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging*. 2015;42:328–354.

19. Delbeke D, Coleman RE, Guiberteau MJ, et al. Procedure guideline for tumor imaging with 18F-FDG PET/CT 1.0. *J Nucl Med*. 2006;47:885–895.

20. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med*. 2009;50(suppl 1):122S–150S.

21. Williams BA, Mandrekar JA, Mandrekar SJ et al. Finding optimal cutpoints for continuous covariates with binary and time-to-event outcomes. Rochester, MN, Mayo Foundation, Technical Report Series 79, 2006.

22. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37–46.

23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.

24. Krishnamoorthy SK, Relias V, Sebastian S, Jayaraman V, Saif MW. Management of regorafenib-related toxicities: a review. *Therap Adv Gastroenterol*. 2015;8:285–297.

25. Samalin E, Bouché O, Thézenas S, et al. Sorafenib and irinotecan (NEXIRI) as second- or later-line treatment for patients with metastatic colorectal cancer and KRAS-mutated tumours: a multicentre Phase I/II trial. *Br J Cancer*. 2014;110:1148–1154.

26. O JH, Jacene H, Luber B, et al. Quantitation of cancer treatment response by 18F-FDG PET/CT: multicenter assessment of measurement variability. *J Nucl Med*. 2017;58:1429–1434.

27. Tahari AK, Paidpally V, Chirindel A, Wahl RL, Subramaniam RM. Two-time-point FDG PET/CT: liver SULmean repeatability. *AJR*. 2015;204:402–407.

FIGURES:

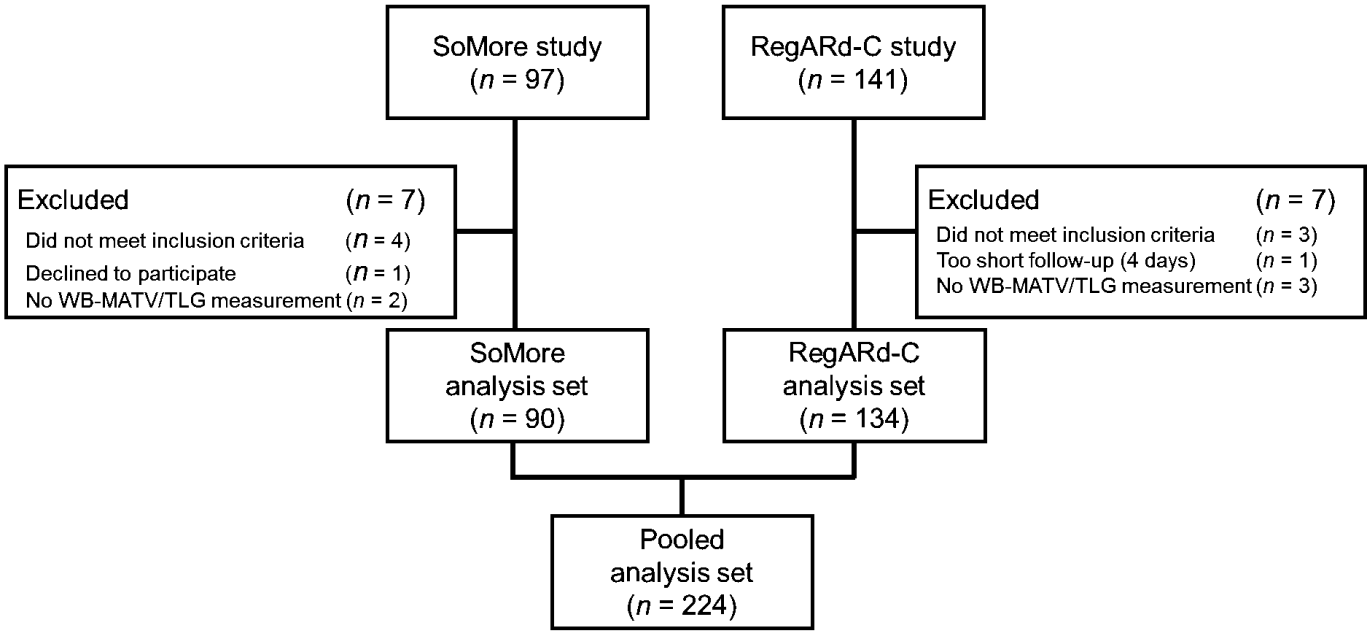
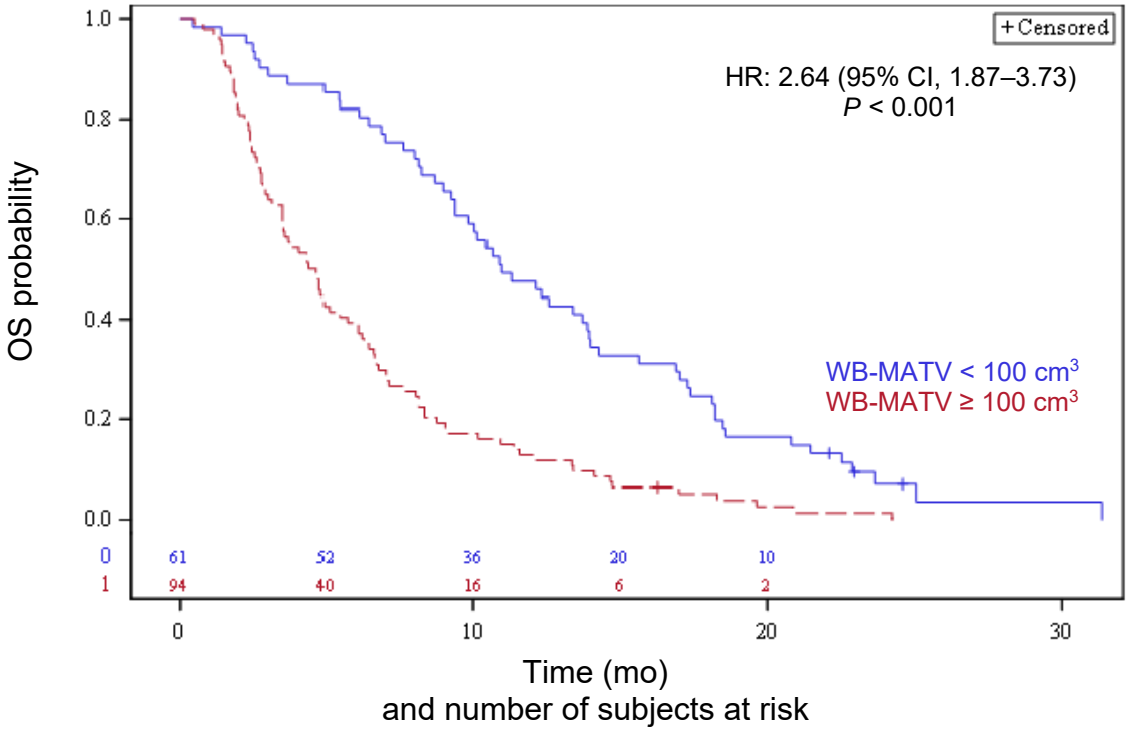


FIGURE 1. Flowchart of the pooled SoMore-RegARd-C population.

Abbreviations: WB-MATV/TLG, whole-body metabolic active tumor volume or total lesion glycolysis.

A


+Censored
HR: 2.64 (95% CI, 1.87–3.73)
P < 0.001
WB-MATV < 100 cm3
WB-MATV ≥ 100 cm3
OS probability
1.0
0.8
0.6
0.4
0.2
0.0
0 61 52 36 20 10
1 94 40 16 6 2
0 10 20 30
Time (mo)
and number of subjects at risk

B

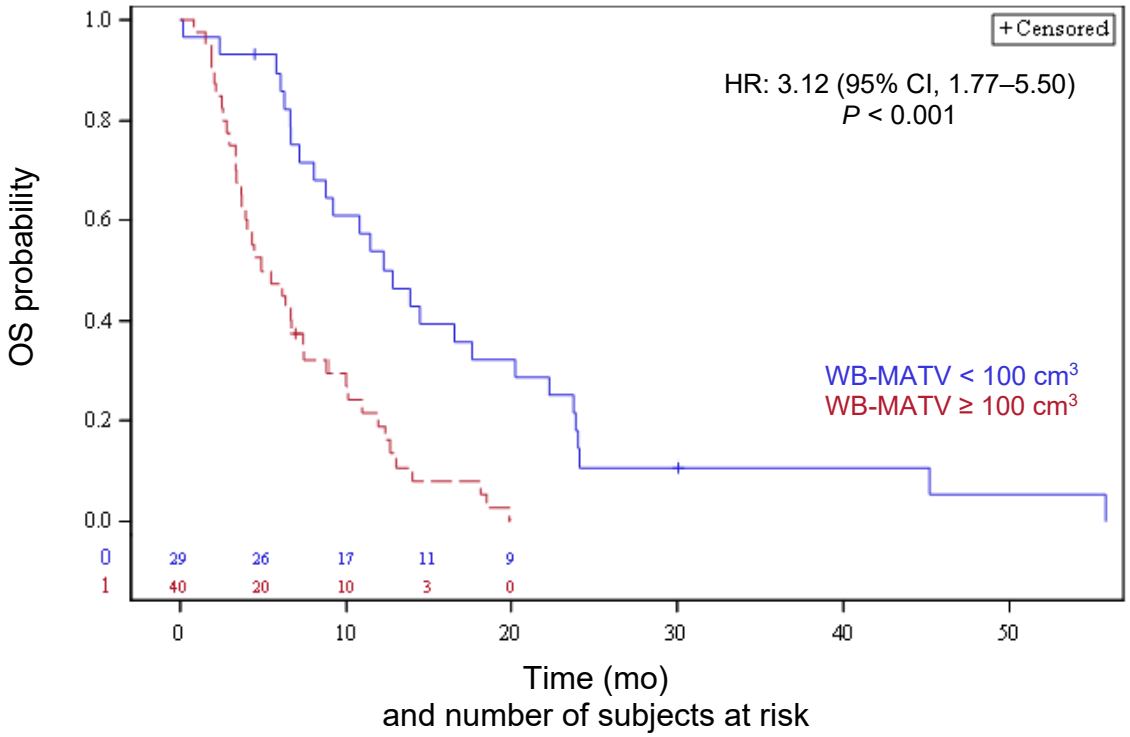

+Censored
HR: 3.12 (95% CI, 1.77–5.50)
P < 0.001
WB-MATV < 100 cm3
WB-MATV ≥ 100 cm3
OS probability
1.0
0.8
0.6
0.4
0.2
0.0
0 29 26 17 11 9
1 40 20 10 3 0
0 10 20 30 40 50
Time (mo)
and number of subjects at risk

FIGURE 2. Overall survival (OS) according to baseline whole-body metabolic active tumor volume (WB-MATV) with a cutoff of 100 cm^3 in the development set (A) and validation set (B).

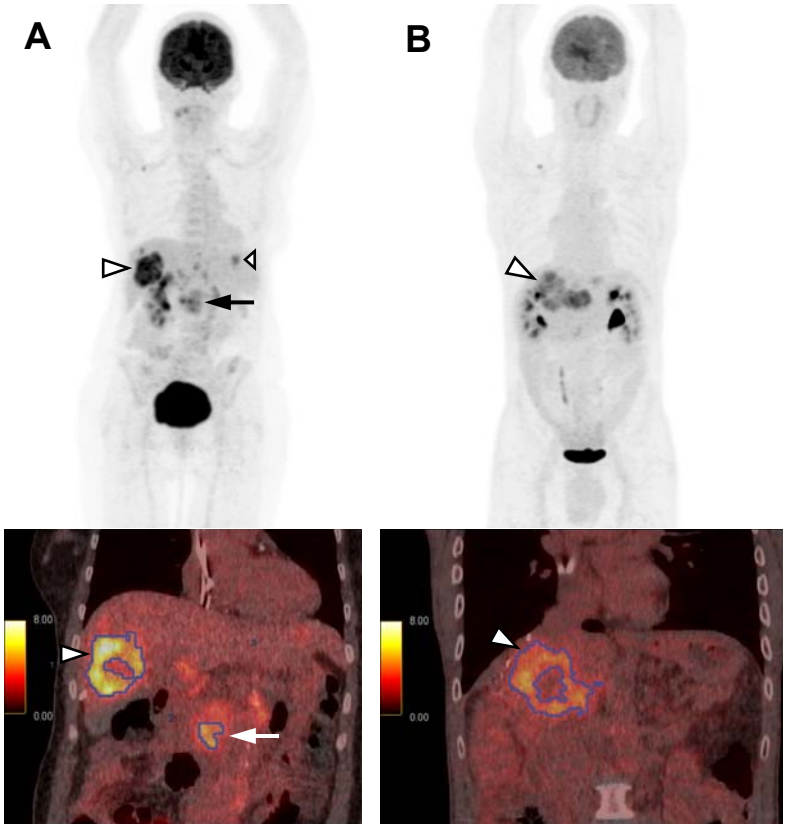
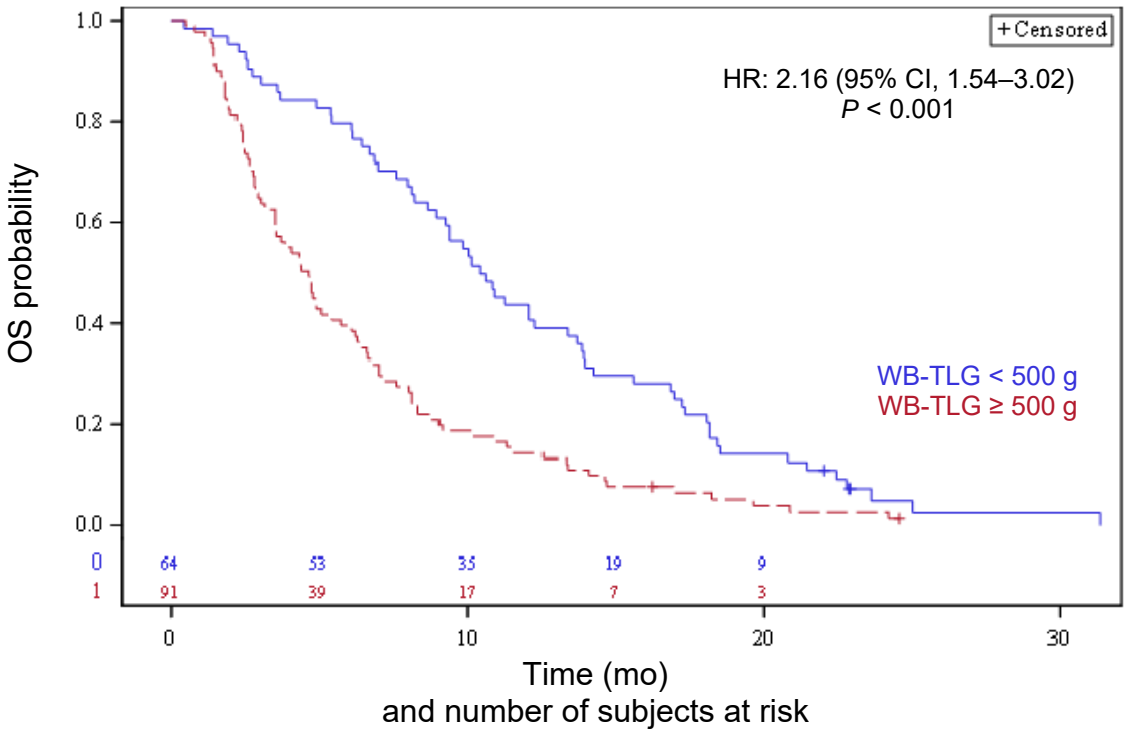


FIGURE 3. Illustration of a patient with a low (A) or high (B) baseline whole-body metabolic active tumor volume (WB-MATV) and total lesion glycolysis (WB-TLG) according to the cutoff values of 100 cm^3 and 500 g, respectively. PET maximum-intensity projection images (top) and coronal fused PET/CT images with delineation of lesions (blue contours) (bottom) of example cases. WB-MATV and TLG of the patient on the left side (A) were 94 cm^3 and 486 g, respectively (inferior to the cutoff values of MATV and TLG) and 147 cm^3 and 559 g (superior to both cutoff values), respectively for the patient on the right side (B). Liver lesions (arrowheads) for both patients and retro-peritoneal lymph nodes (arrows) of the left side patient were identified as target lesions and were taken into account for WB-MATV and TLG measurements.

A


+Censored
HR: 2.16 (95% CI, 1.54–3.02)
P < 0.001
WB-TLG < 500 g
WB-TLG ≥ 500 g
OS probability
1.0
0.8
0.6
0.4
0.2
0.0
0 64 53 35 19 9
1 91 39 17 7 3
0 10 20 30
Time (mo)
and number of subjects at risk

B

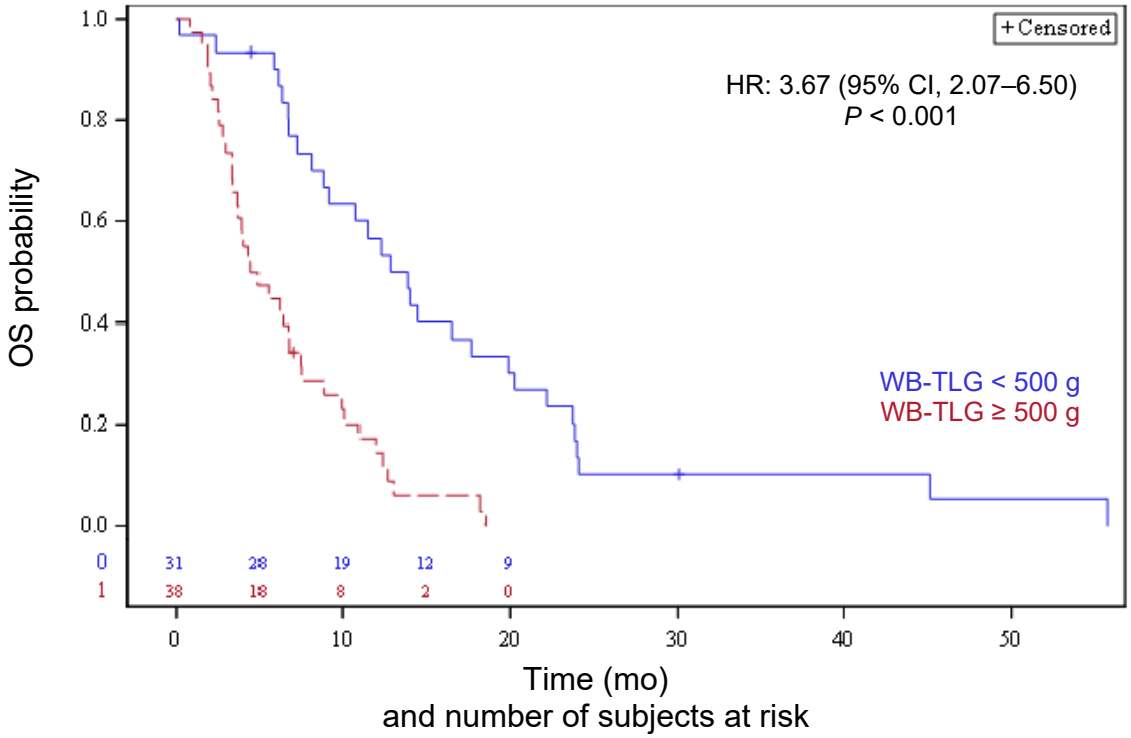

+Censored
HR: 3.67 (95% CI, 2.07–6.50)
P < 0.001
WB-TLG < 500 g
WB-TLG ≥ 500 g
OS probability
1.0
0.8
0.6
0.4
0.2
0.0
0 31 28 19 12 9
1 38 18 8 2 0
0 10 20 30 40 50
Time (mo)
and number of subjects at risk

FIGURE 4. Overall survival (OS) according to baseline whole-body total lesion glycolysis (WB-TLG) with a cutoff of 500 g in the development set (A) and validation set (B).

TABLES:

TABLE 1. Baseline patient characteristics of the overall population (*n* = 224)

Characteristic	Value	%
Age (years)		
Median (range)	65.0 (27.8–84.7)	
Gender		
Male	127	57%
Female	97	43%
BMI		
Median (range)	24.9 (14.1–48.3)	
ECOG PS		
0	112	50%
1	112	50%
Years between diagnosis and inclusion in the trial		
Median (range)	2.7 (0.1–14.9)	
Prior use of bevacizumab		
Yes	154	69%
No	70	31%
KRAS		
Wild type	103	46%
Mutant	120	54%
Missing	1	<1%
Targeted agent		
Sorafenib	90	40%
Regorafenib	134	60%
WB-MATV		
Median (range)	160 (2–5448)	
WB-TLG		
Median (range)	698 (6–19099)	
WB-SUL$_{mean}$		
Median (range)	4.4 (0.2–7.8)	

Abbreviations: BMI, body mass index; ECOG PS, Eastern Cooperative Oncology Group performance status; WB-MATV, whole-body metabolic active tumor volume; WB-TLG, whole-body total lesion glycolysis.

TABLE 2. Independent predictors of OS with baseline WB-MATV included in the multivariate analysis along with commonly used clinical factors

	OS			
Parameter	**Prognostic weight**	**Parameter estimate**	**P-value**	**HR (95% CI)**
WB-MATV ≥ 100 cm^3	6	0.90	<0.001	2.46 (1.71–3.52)
No. years since diagnosis (per 1-year increase)	−1	−0.15	<0.001	0.86 (0.80–0.93)
ECOG PS 1	3	0.52	0.003	1.67 (1.19–2.35)
BMI ≥ 25	−4	−0.57	<0.001	0.56 (0.40–0.79)

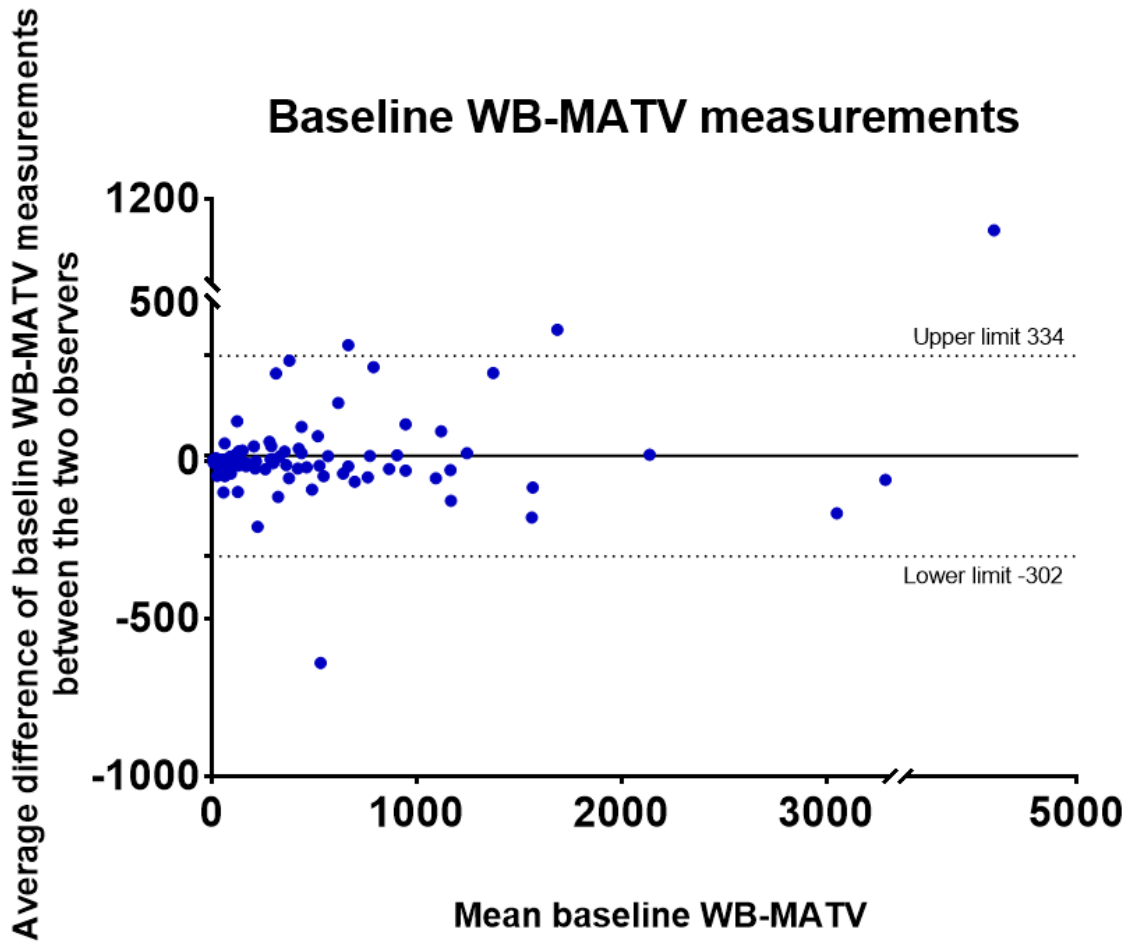
Abbreviations: OS, overall survival; HR, hazard ratio; WB-MATV, whole-body metabolic active tumor volume; ECOG PS, Eastern Cooperative Oncology Group performance status; BMI, body mass index.

TABLE 3. Independent predictors of OS with baseline WB-TLG included in the multivariate analysis along with commonly used clinical factors

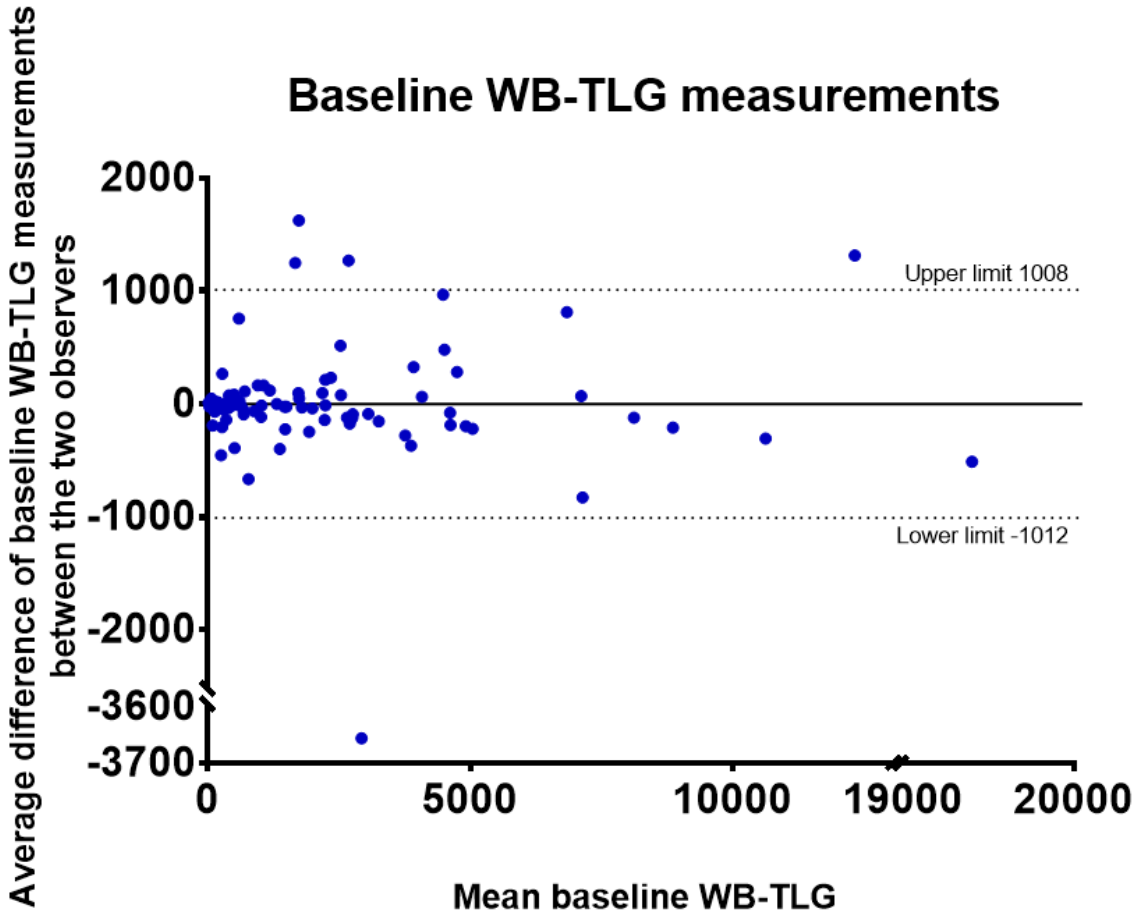
	OS			
Parameter	**Prognostic weight**	**Parameter estimate**	**P-value**	**HR (95% CI)**
WB-TLG ≥ 500 g	5	0.80	<0.001	2.23 (1.57–3.17)
No. years since diagnosis (per 1-year increase)	−1	−0.15	<0.001	0.86 (0.80–0.93)
ECOG PS 1	4	0.59	<0.001	1.81 (1.29–2.53)
BMI ≥ 25	−4	−0.63	<0.001	0.53 (0.38–0.75)

Abbreviations: OS, overall survival; HR, hazard ratio; WB-TLG, whole-body total lesion glycolysis; ECOG PS, Eastern Cooperative Oncology Group performance status; BMI, body mass index.

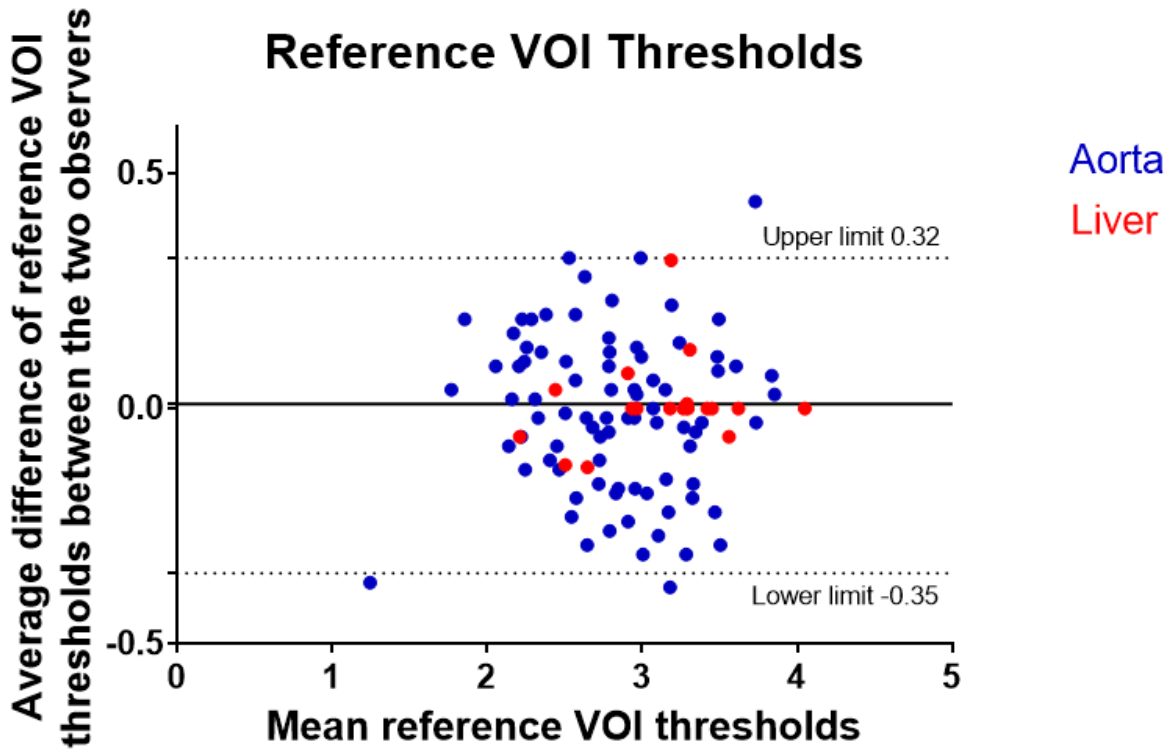
A


Baseline WB-MATV measurements
Average difference of baseline WB-MATV measurements between the two observers
1200
500
0
-500
-1000
Upper limit 334
Lower limit -302
0
1000
2000
3000
5000
Mean baseline WB-MATV

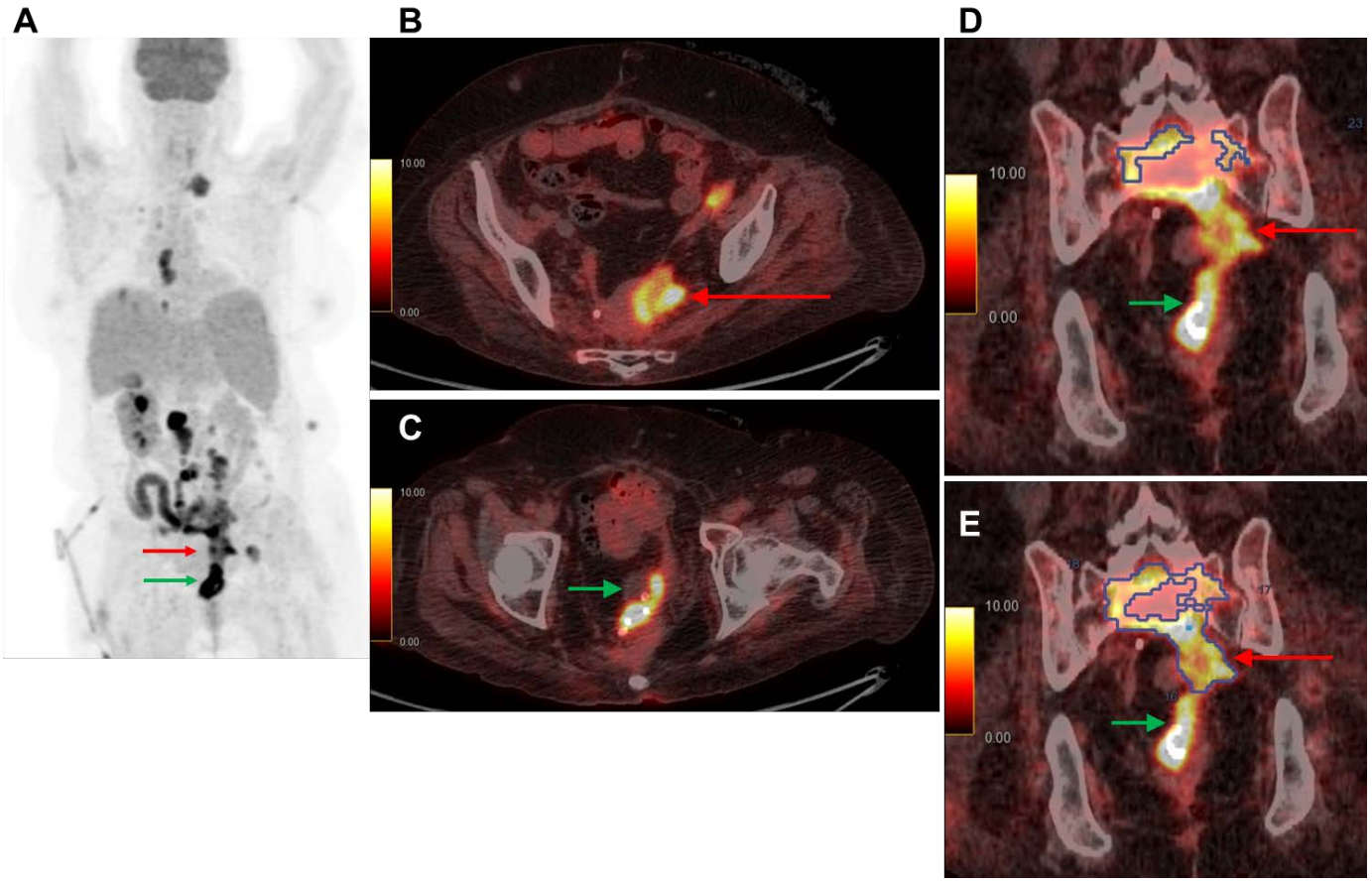
B


Baseline WB-TLG measurements
Average difference of baseline WB-TLG measurements between the two observers
2000
1000
0
-1000
-2000
-3600
-3700
Upper limit 1008
Lower limit -1012
0
5000
10000
19000
20000
Mean baseline WB-TLG

C



Supplemental FIGURE 1. Bland-Altman plots for baseline WB-MATV (A), WB-TLG (B), and reference VOI thresholds (C). Each dot represents a patient viewed by the two observers (total, 100 patients). x-axis represents the mean measurement of the two observers. y-axis represents average difference of the measurement between the two observers. Solid line represents average bias, and dashed lines represent corresponding bias ± 2 SDs.



Supplemental FIGURE 2. The main cause of variability between observers, who were completely unaware of clinical history for both WB-MATV and WB-TLG measurements, was the difference in selection of target lesions. PET maximum-intensity projection (A), axial fused PET/CT images (B, C), and coronal fused PET/CT images (D, E) of the example case. A metabolically active lesion located in the pelvic cavity (red arrow), with an equivocal origin (inflammatory/infectious or tumoral) and too close to urinary activity (green arrow), was not taken into consideration for WB-MATV/TLG measurements by observer 1 (D) (WB-MATV and TLG of observer 1: 78 cm^3 and 319 g; inferior to the cutoff values of MATV [100 cm^3] and TLG [500 g]) but was measured by observer 2 (E) (WB-MATV and TLG of observer 2: 176 cm^3 and 712 g, respectively).

Supplemental TABLE 1. Survival characteristics of the pooled, development, and validation sets from SoMore-RegARd-C population.

	Pooled population (n = 224)	Development set (n = 155)	Validation set (n = 69)
Median OS			
Months	6.9	6.7	7.5
95% CI	6.2–8.1	5.4–8.1	6.2–10.8
No. of deaths	217	151	66
Median PFS			
Months	3.3	3.2	3.3
95% CI	2.2–3.7	2.1–3.9	2.0–3.8
No. of progressive events	224	155	69

Supplemental TABLE 2. Agreement between the dichotomized WB-MATV and WB-TLG measurements for the development set.

	WB-MATV		
WB-TLG	**< 100**	**≥ 100**	**Total**
< 500	113	5	118
≥ 500	7	30	37
Total	120	35	155

Supplemental TABLE 3. Contingency tables showing the agreement between the two observers of WB-MATV and WB-TLG measurements and discrepancy cases between categorization (cases assigned to a different category [low versus high tumor load] by the two observers) amongst the set of 100 patients measured by the two observers.

WB-MATV		Observer 1		
		< 100	**≥ 100**	**Total**
Observer 2	**< 100**	34	1	35
	≥ 100	3	62	65
	Total	37	63	100

WB-TLG		Observer 1		
		< 500	**≥ 500**	**Total**
Observer 2	**< 500**	38	3	41
	≥ 500	2	57	59
	Total	40	60	100