

research article

Role of quantitative imaging biomarkers in an early FDG-PET/CT for detection of immune-related adverse events in melanoma patients: a prospective study

Nezka Hribernik^{1,2}, Katja Strasek³, Daniel T Huff^{4,5}, Andrej Studen^{3,6}, Katarina Zevnik⁷, Katja Skalic⁷, Robert Jeraj^{3,4,5,6}, Martina Rebersek^{1,2}

¹ Department of Medical Oncology, Institute of Oncology Ljubljana, Ljubljana, Slovenia

² Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

³ Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia

⁴ University of Wisconsin Carbone Cancer Centre, Madison, WI, USA

⁵ Department of Medical Physics, University of Wisconsin-Madison, Madison, WI, USA

⁶ Experimental Particle Physics Department, Jožef Stefan Institute, Ljubljana, Slovenia

⁷ Department of Nuclear Medicine, Institute of Oncology Ljubljana, Ljubljana, Slovenia

Radiol Oncol 2024; 58(3): 335-347.

Received 13 June 2024

Accepted 14 July 2024

Correspondence to: Assist. Prof. Martina Reberšek, M.D., Ph.D., Department of Medical Oncology, Institute of Oncology Ljubljana, Zaloška 2, SI-1000 Ljubljana, Slovenia. E-mail: mrebersek@onko-i.si

Disclosure: Robert Jeraj is a co-founder and CSO of AIQ Solutions, Madison, WI, USA. No potential conflicts of interest were disclosed for other co-authors.

This is an open access article distributed under the terms of the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

Background. To evaluate the role of the novel quantitative imaging biomarker (QIB) $SUV_{x\%}$ of ^{18}F -FDG uptake extracted from early ^{18}F -FDG -PET/CT scan at 4 weeks for the detection of immune-related adverse events (irAE) in a cohort of patients with metastatic melanoma (mM) patients receiving immune-checkpoint inhibitors (ICI).

Patients and methods. In this prospective non-interventional, one-centre clinical study, patients with mM, receiving ICI treatment, were regularly followed by ^{18}F -FDG PET/CT. Patients were scanned at baseline, early point at week four (W4), week sixteen (W16) and week thirty-two (W32) after ICI initiation. A convolutional neural network (CNN) was used to segment three organs: lung, bowel, thyroid. QIB of irAE - $SUV_{x\%}$ - was analyzed within the target organs and correlated with the clinical irAE status. Area under the receiver-operating characteristic curve (AUROC) was used to quantify irAE detection performance.

Results. A total of 242 ^{18}F -FDG PET/CT images of 71 mM patients were prospectively collected and analysed. The early W4 scan showed improved detection only for the thyroid gland compared to W32 scan ($p=0.047$). The AUROC for detection of irAE in the three target organs was highest when $SUV_{x\%}$ was extracted from W16 scan and was 0.76 for lung, 0.53 for bowel and 0.81 for thyroid. $SUV_{x\%}$ extracted from W4 scan did not improve detection of irAE compared to W16 scan (lung: $p = 0.54$, bowel: $p = 0.75$, thyroid: $p = 0.3$, DeLong test), as well as compared to W32 scan in lungs ($p = 0.32$) and bowel ($p = 0.3$).

Conclusions. Early time point ^{18}F -FDG PET/CT at W4 did not lead to statistically significant earlier detection of irAE. However, organ ^{18}F -FDG uptake as quantified by $SUV_{x\%}$ proved to be a consistent QIB of irAE. To better assess the role of ^{18}F -FDG PET/CT in irAE detection, the time evolution of ^{18}F -FDG PET/CT quantifiable inflammation would be of essence, only achievable in multi centric studies.

Key words: ^{18}F -FDG PET/CT; immune-checkpoint inhibitors; immune-related adverse effects; quantitative imaging biomarkers; SUV percentiles

Introduction

Immunotherapy with immune checkpoint inhibitors (ICI) is considered the standard of care for the treatment of unresectable stage III and IV metastatic melanoma (mM). It leads to high response rates and improvement of survival in this group of cancer patients.¹⁻⁴ ICI treatments can cause different immune-related adverse events (irAE), where an immune response is generated against healthy tissue.⁵⁻⁷ Nearly one quarter of ICI treated patients experience higher grades of irAE that require hospitalization, 26% due to immune-related pneumonitis (irPneumonitis) and 17% due to immune-related colitis (irColitis).⁸ Consequently, prolonged immune suppressive treatment is needed, which can significantly endanger patient health and reduce the efficacy of ICI treatment.

The immense progress of ICI has brought different challenges, one of them being assessment of treatment efficacy and management of variety of irAE. Positron emission tomography/computed tomography with [¹⁸F]fluoro-2-deoxy-D-glucose (¹⁸F-FDG PET/CT) is a non-invasive method, commonly used in diagnosis, staging, and treatment response monitoring in mM. Its role in the treatment process is being recognized through different guidelines and emerging evidence.^{9,10}

In addition to monitoring response, ¹⁸F-FDG PET/CT can detect certain irAE, some even before they clinically or biochemically manifest.¹¹⁻¹³ Early detection or even prediction of irAE is of great interest, as this could contribute to better treatment.^{11,14} To date, no study has been published investigating whether a systematic search for irAE with ¹⁸F-FDG PET/CT imaging as early as four weeks after the start of ICI treatment would help in early detection of irAE. However, several studies investigated utility of different ¹⁸F-FDG PET/CT timing. A small prospective study with 20 mM patients evaluated if ¹⁸F-FDG PET/CT at days 21-28 would help predicting response to ICI in mM patients. It showed that increased ¹⁸F-FDG uptake early during ICI treatment may be associated with immune activation and favourable outcome.¹⁵ No analysis of irAE was reported. A similar study investigated prognostic value of ¹⁸F-FDG PET/CT after two cycles of ICI administration.¹⁶ They looked also for patterns of ¹⁸F-FDG uptake suggestive of irAE and found that the emergence of PET signs of irAE was not associated with patient survival. Another recent study analysed the usefulness of ¹⁸F-FDG PET/CT performed even earlier, one week after starting the treatment with pembrolizumab

of mM patients. The study was aimed to evaluate metabolic changes in melanoma metastases one week after treatment start in order to predict the response to treatment. No evaluation of organs for irAE was reported.¹⁷

The segmentation of metastatic lesions and organs on ¹⁸F-FDG PET/CT with deep learning-based convolutional neural network (CNN) has recently been developed.^{18,19} In a previous work by our group of 58 mM patients receiving ICI, a novel quantitative imaging biomarker (QIB) of irAE development was proposed – percentiles of SUV distribution (SUV_{x%}) of ¹⁸F-FDG uptake within the target organs.¹¹ Three target organs were considered for identification of irAE: lung, bowel and thyroid. These target organs were segmented using a CNN, and SUV_{x%} of ¹⁸F-FDG uptake within the target organs were correlated with the clinical irAE status. It was shown that increased ¹⁸F-FDG uptake within irAE-affected organs provides predictive information about the development of irAE and represents a potential QIB for irAE. Some irAE were detected on ¹⁸F-FDG PET/CT before clinical symptoms appeared.¹¹

The aim of this prospective study was to evaluate the role of the QIB SUV_{x%} of ¹⁸F-FDG uptake extracted from early ¹⁸F-FDG PET/CT scans for the detection of irAE in a cohort of patients with mM, treated with ICI. ¹⁸F-FDG PET/CT imaging was performed at an early timepoint at week 4 (W4) and two consecutive scans at week 16 (W16) and week 32 (W23) after ICI initiation. We hypothesized that quantitative analysis of three organs of interest (lung, bowel and thyroid) on week 4 ¹⁸F-FDG PET/CT would give information about irAE detection that precede clinical diagnosis and outperform later scans.

Patients and methods

Patient population and study protocol

We have been conducting a noninterventional, prospective study with primary endpoint of quantitative analysis of ¹⁸F-FDG PET/CT scans in metastatic melanoma patients for detection of irAE. The patients were treated per standard of care with ICI with anti-cytotoxic T-lymphocyte-associated antigen 4 (anti-CTLA-4) and/or anti programmed death-1 (anti-PD-1) treatment in first and second line of systemic treatment at the Institute of Oncology Ljubljana (OIL), Slovenia. Eligible patients were older than 18 years and had a stage III.D unresectable melanoma or a stage IV mela-

noma. All patients had an Eastern Cooperative Oncology Group (ECOG) performance status score of 0 or 1. Patients were eligible if they had the baseline ^{18}F -FDG PET/CT performed within four weeks prior to ICI treatment initiation. Key exclusion criteria included symptomatic brain metastases and malignant diseases other than melanoma. Written consent was received from all participants.

Patients were monitored with regular ^{18}F -FDG PET/CT. A baseline scan was performed within four weeks prior to ICI initiation. After treatment initiation, ^{18}F -FDG PET/CT was performed at W4 ($\pm$ 5 days), W16 ($\pm$ 7 days), W32 ($\pm$ 7 days) and every 16 weeks after. Because the W4 ^{18}F -FDG PET/CT was performed for investigational purposes, the scan data was not necessarily used to guide treatment decisions. In case of clear progression such as hyperprogression²⁰, systemic treatment was allowed to be switched based on the decision of the medical oncologists leading the treatment. Clinical data and images included in this analysis were data obtained from disease diagnosis up to the W32 scan plus an additional 4 weeks, allowing for clinical detection of up to W32 clinically unobserved irAE in monitored organs, or 252 days after treatment start in cases where the patient did not have a W32 scan.

All ^{18}F -FDG PET/CT data acquired before and during ICI treatment and all clinical data was collected for review. Grading of irAE was assigned prospectively following Common Terminology Criteria for Adverse Events (CTCAE, v.5.0).²¹ Imaging and clinical data were stored in a secure LabKey database server.²²

The clinical protocol was approved by the Ethics Committee ERIDEK-0034/2020 and the Clinical Trials Protocol Review Committee ERID-KSOPKR-0032/2020 at the Institute of Oncology Ljubljana, and by Commission of the Republic of Slovenia for Medical Ethics (approval number: 0120-256/2020-14, September 15th 2020). It was conducted following the ethical standards defined by the Declaration of Helsinki and the International Conference on Harmonisation Guidelines for Good Clinical Practice. The study was registered with the ClinicalTrials.gov under the registration number NCT06207747.

The study was conducted with acknowledgment and consent of the subjects. All patients have signed informed consent for treatment and consent allowing the usage of their data for scientific purposes. For publication of images in Figure 6 and 7, research participants provided informed consent.

^{18}F -FDG PET/CT image acquisition

^{18}F -FDG PET/CT scans were primarily performed for immunotherapy treatment response evaluation in melanoma patients. All ^{18}F -FDG PET/CT scans were obtained on Biograph mCT PET/CT (Siemens, Knoxville, TN); one scan was obtained at University Clinical Centre Maribor (UKC MB), two at University Clinical Centre Ljubljana (UKC LJ) and the rest were acquired at the Institute of Oncology Ljubljana (OIL) following institutional protocols. Imaging protocol required patients to fast for 6 hours prior to injection of the radiotracer and have a blood glucose level below 10 mmol/L at the time of the scan. Patients were required to hold all diabetic medication, including metformin, for 6 hours prior to radiotracer injection. All scans were acquired per Standard of Care (SOC). CT that meets RECIST analysis needs was acquired according to adjusted protocol including SAFIR reconstruction to minimize dose. Following reconstruction, PET images were normalized by patient weight and injected dose to compute Standardized Uptake Values (SUV). More information on scanning parameters for each scanner can be found in Supplementary Table 1.

^{18}F -FDG PET/CT image analysis

Before image analysis, images were resampled to a cubic 2-mm grid (2mm $\times$ 2mm $\times$ 2mm) using linear resampling and normalized to have a mean of 0 and variance of 1 within patient to match DeepMedic model developed for Hribernik *et al.*¹¹

To quantify organ uptake, a 3D CNN was used to segment three target organs: lung, bowel and thyroid. The architecture used was nnUNet.²³ The network was trained to perform inference on CT component from the ^{18}F -FDG-PET/CT scans to produce contours of all three organs and was trained on same data as DeepMedic network, used in our previous pilot study.¹¹

The organ contours, produced by nnUNet were applied to PET component of the ^{18}F -FDG-PET/CT image to quantify ^{18}F -FDG uptake in organs of interest. Percentiles X% of the organ SUV distribution ($\text{SUV}_{X\%}$) were extracted, as they were found to be potential biomarkers of irAE in retrospective study of mM patients and were found to be more reliable than SUV_{max} .^{11,14,24} To quantify organ uptake, only the percentile, that was found to be optimal irAE biomarker in our pilot study in each organ was used: $\text{SUV}_{95\%}$ for lung and bowel, $\text{SUV}_{75\%}$ for thyroid.¹¹

TABLE 1. Patient characteristics

	Characteristics	Count (proportion %)
Number of patients	Total	71 (100)
Age; mean (+/-sd) (yr)		62 (12)
Gender	Male	43 (61)
	Female	28 (39)
ECOG performance status	0	30 (42)
	1	41 (58)
AJCC	III.D	1 (1)
	M1a	16 (23)
	M1b	10 (14)
	M1c	32 (45)
	M1d	12 (17)
Anatomic site of primary	Cutaneous	58 (82)
	Ocular	4 (6)
	Mucosal	3 (4)
	Unknown primary	6 (8)
Prior radical treatment	Surgical resection	38 (54)
	Surgical resection + adjuvant RT	7 (10)
	Surgical resection + adjuvant RT + adjuvant ICI	4 (6)
	Surgical resection + adjuvant RT + adjuvant TKI	2 (3)
	Surgical resection + adjuvant TKI	2 (3)
	Surgical resection + adjuvant interferon alpha	2 (3)
Line of systemic treatment for metastatic disease	None	16 (23)
	1 st line	63 (89)
Baseline LDH	2 nd line	8 (11)
	Elevated	22 (31)
Actionable mutation	Normal	49 (69)
	BRAF wild type	21 (30)
	BRAF V600E	28 (39)
	BRAF V600K	10 (14)
	BRAF V600 - others	1 (1)
Type of systemic treatment	NRAS	11 (16)
	PD-1 inhibitors	47 (66)
	Combination of PD-1 and CTLA-4 inhibitors	24 (34)

AJCC = American Joint Classification of Cancer; BRAF = V-Raf Murine Sarcoma viral oncogene homolog B; CTLA-4 = Cytotoxic T-lymphocyte-associated antigen 4; ECOG = Eastern Cooperative Oncology Group; ICI = Immune checkpoint inhibitors; No = number of patients; NRAS = Neuroblastoma RAS viral homolog; PD-1 = Programmed death-1; SD = standard deviation; RT = radiotherapy; TKI = Tyrosine kinase inhibitors

Statistical analysis

Patients were divided into two groups for each analysed target organ (lung, bowel, thyroid): patients, who experienced irAE from time on ICI start up until the cut-off and patients who did not experience irAE during this period.

All timing was calculated in days relative to start of treatment with ICI. In the group of patients experiencing irAE, statistics for all irAE events were calculated per type of irAE and per grade (any grade and high grade: grade 3-5 per CTCAE, v.5.0).²¹

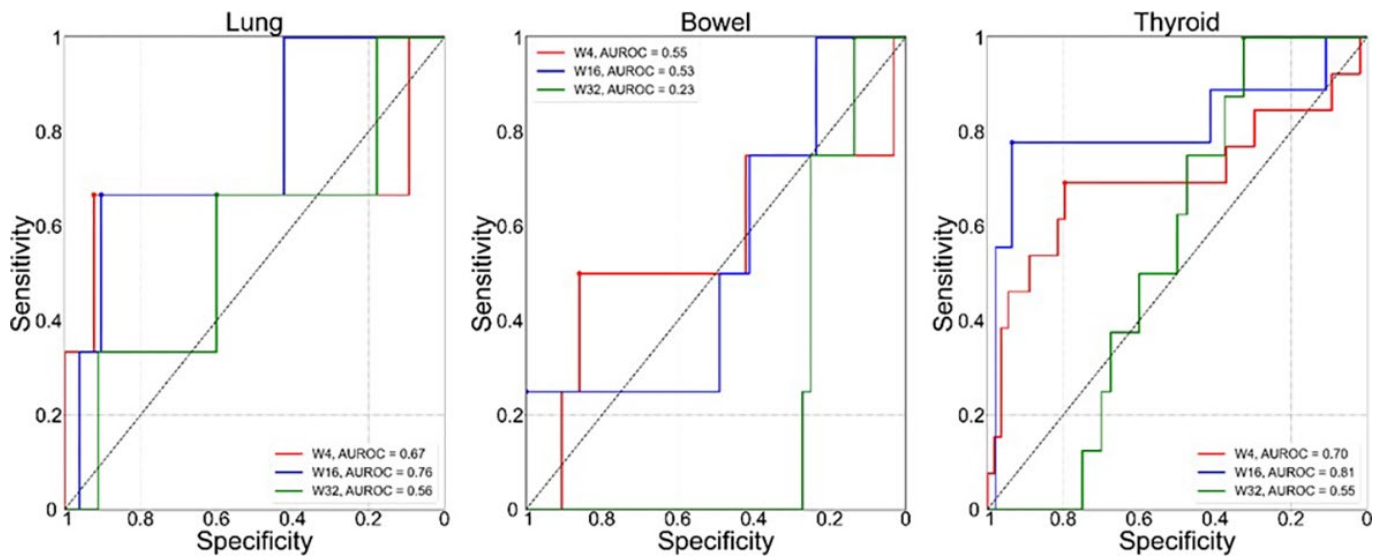


FIGURE 1. ROC curves for the optimal SUV percentile ($SUV_{x\%}$) for predicting irAE status in the three target organs: lung (left), bowel (middle), thyroid (right). Each plot shows comparison of ROC based on the value of $SUV_{x\%}$ extracted from W4 (red), W16 (blue) and W32 (green) ¹⁸F-FDG PET/CT. Values of corresponding area under the ROC curve (AUROC) are shown in the legends.

From all clinically detected irAE observed in our follow-up period, the following were included in the analysis: immune-related pneumonitis (irPneumonitis), immune-related colitis (irColitis) and immune-related thyroiditis (irThyroiditis). Receiver Operating Characteristic (ROC) analysis was performed to determine the ability of extracted $SUV_{x\%}$ in each organ to detect observed irAE on scans at W4, W16 and W32 individually. This was done by comparing the $SUV_{x\%}$ extracted from each timepoint scan with the clinical irAE status of the patient up until cut-off for all three analysed organs. Additionally, the maximum organ $SUV_{x\%}$ value, extracted from all available images (including baseline) up until W32, was investigated as a predictor of irAE. This ROC analysis was compared to ROC curves previously obtained in¹¹ using the same methodology. Area under the ROC curve (AUROC) was calculated. The performance of different models was compared using DeLong test.

In the following part of the analysis, the threshold to differentiate between patients without irAE and those experiencing irAE based on $SUV_{x\%}$ and the normal $SUV_{x\%}$ range from analysis in pilot study were used.¹¹ In that work, normal ranges of $SUV_{x\%}$ were for lung [0.7, 1.5] g/mL, bowel [1.8, 2.9] g/mL, and thyroid [0.9, 2.0] g/mL. Optimal thresholds used were: $SUV_{95\%} = 1.7$ g/mL in lung, $SUV_{95\%} = 2.7$ g/mL in bowel and $SUV_{75\%} = 2.1$ g/mL in thyroid.

Target organ ¹⁸F-FDG uptake was assessed in those patients without clinically diagnosed irAE which had $SUV_{x\%}$ above threshold at any imaging timepoint. Medical oncologist and nuclear medicine specialist identified causes of higher uptake of ¹⁸F-FDG in the organ of interest in these patients using patient chart review. $SUV_{x\%}$ values in each organ of interest were compared between causes using Mann-Whitney U test.

Statistical comparison of irAE detection using $SUV_{x\%}$ and clinical detection was done. As a measure of instance when irAE was detected by ¹⁸F-FDG-PET/CT, we identified the first scan on which the $SUV_{x\%}$ in the organ of interest was above the mentioned threshold. We compared this instance to the timing of clinical diagnosis of irAE using Wilcoxon signed-rank test.

Image analysis and statistical testing was done using Python programming language version 3.7.16 (Python Software Foundation, <https://www.python.org/>). Statistically significant differences were observed when p value was below 0.05.

Results

Patient characteristics

Prospective study started with patient inclusion in September 2020. While the study is still ongoing, patient inclusion stopped in September 2022. The inclusion period of the study was initially planned

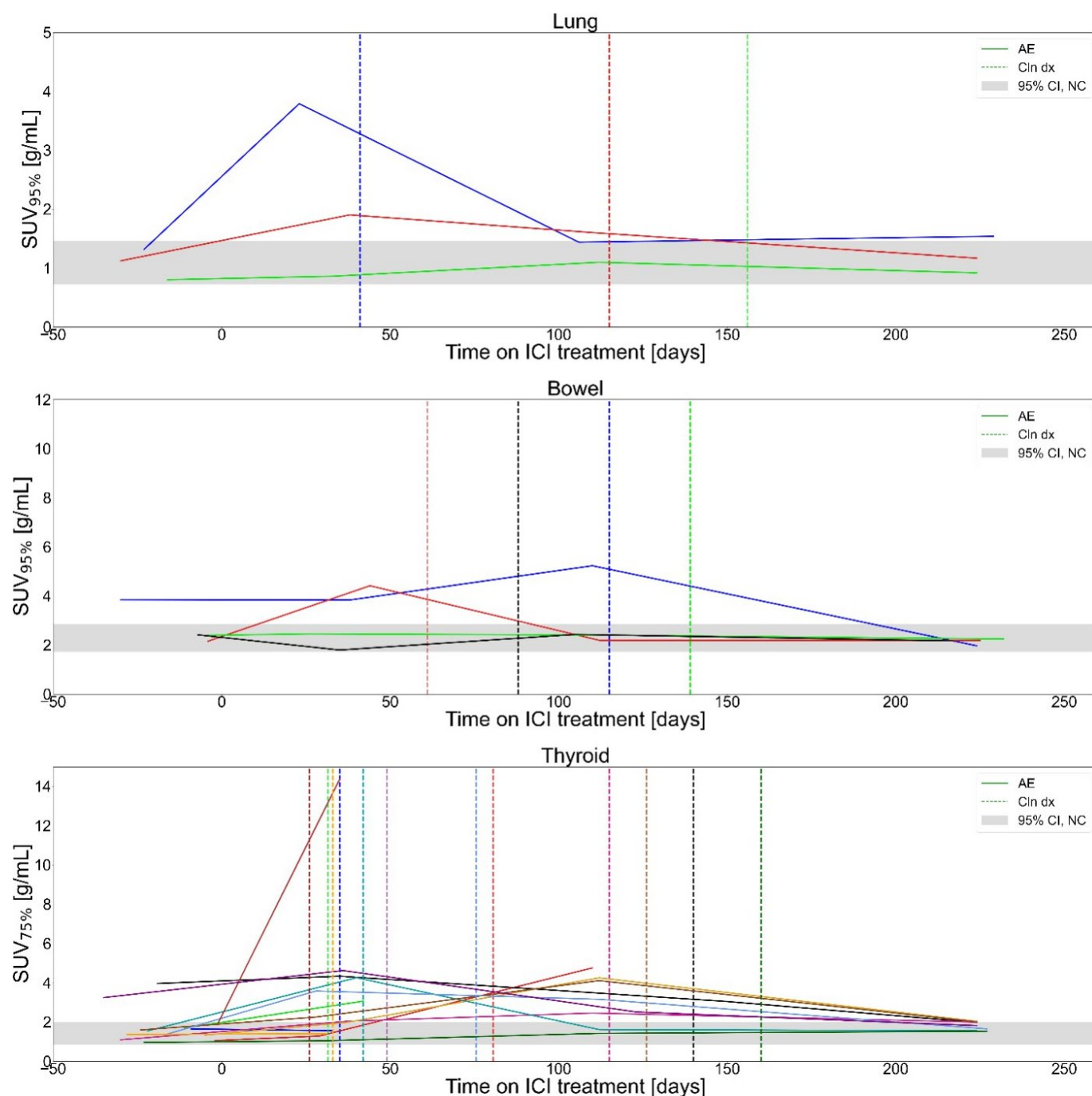


FIGURE 2. Longitudinal optimal SUV percentile (SUV_{95%}) for patients with clinically detected immune-related adverse events (irAE) in bowel, lung and thyroid, included in study. The grey band indicates the 95% confidence interval for organ SUV_{95%} of patients who did not experience irAE in¹¹. Colours were randomly selected for participants in each plot, vertical dashed lines of matching colour indicate dates of clinical irAE identification.

ICI = Immune checkpoint inhibitors

to last 18 months but was extended for 6 months due to the COVID-19 pandemic. Altogether, 71 patients were included in the study; patient demographics are summarised in Table 1.

Clinical detection of irAE

Dates and grades of clinically detected irPneumonitis, irColitis and irThyroiditis were systematical-

ly collected. Number of events for all three types of irAE and median time to first clinical diagnosis of all irAE during observational period are given in Table 2.

During our observational period, 10 patients (14%) developed hypothyroidism and 7 (10%) hyperthyroidism. We identified 13 patients (18%) to have developed irThyroiditis based on patterns of thyroid dysfunction.²⁵

TABLE 2. Clinical diagnosis of immune-related adverse events

irAE	Any Grade No (%)	Grade 3-5 No (%)	Time to onset of irAE (mean $\pm$ SD) [days]
irPneumonitis	3 (4)	0	104 $\pm$ 58
irColitis	4 (6)	2 (3)	101 $\pm$ 34
Hypothyroidism	10 (14)	0	91 $\pm$ 40
Hyperthyroidism	7 (10)	0	51 $\pm$ 50

irAE = immune-related adverse events; No = number of patients; SD = standard deviation

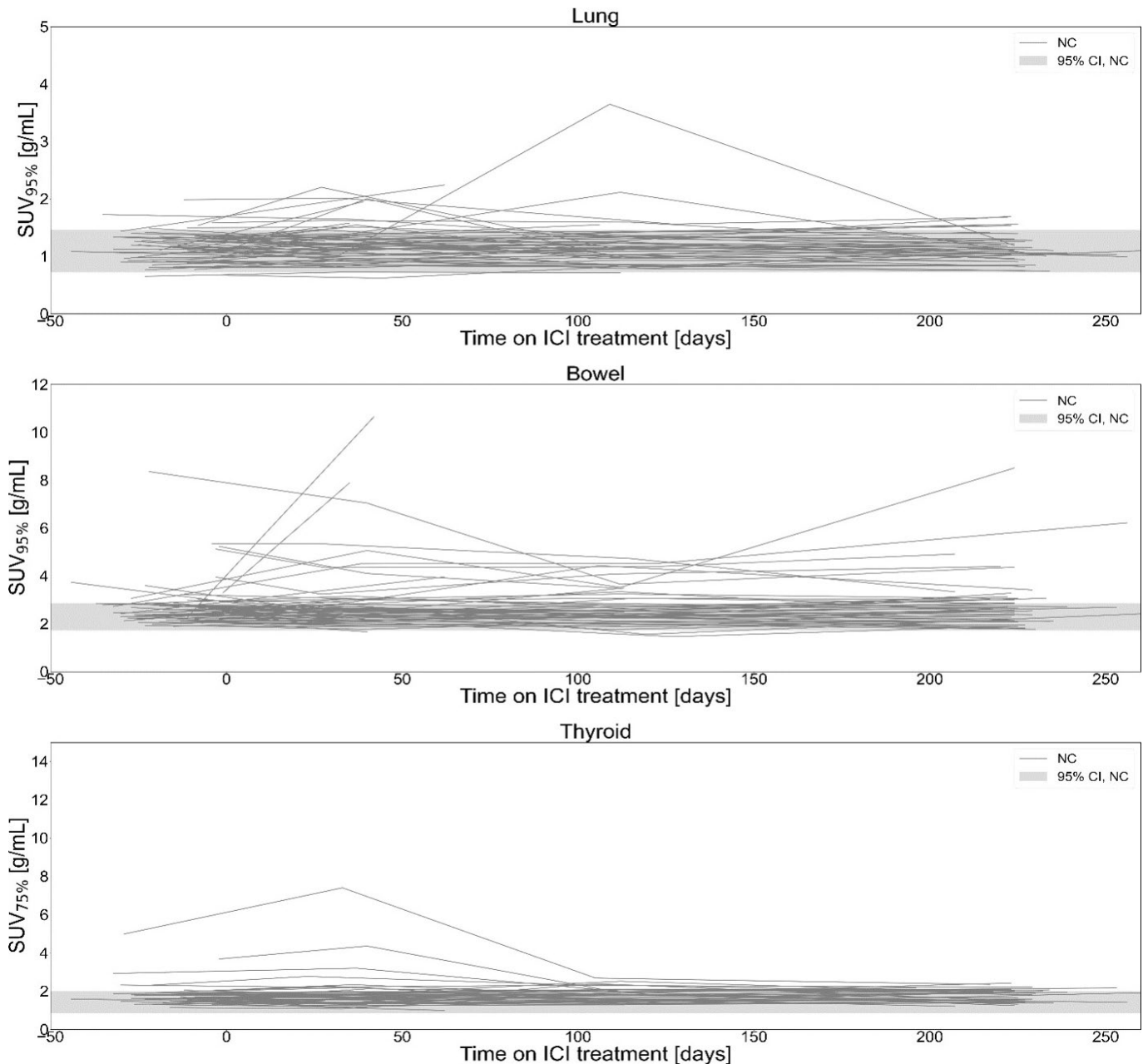


FIGURE 3. Longitudinal optimal SUV percentile (SUV_{75%}) for patients without clinically detected immune-related adverse events (irAE) in bowel, lung and thyroid, included in study. The grey band indicates the 95% confidence interval for organ SUV_{75%} of patients who did not experience irAE in¹¹.

ICI = Immune checkpoint inhibitor

TABLE 3. Timing of ¹⁸F-FDG PET/CT relative to immune-checkpoint inhibitors (ICI) treatment initiation

	Number of patients	Range (min-max) (days)	Mean (days)	Median (days)	SD (days)
Pre-Treatment	71	[-44, -1]	-16	-16	10
W4 (28 days)	68	[21, 62]	34	34	7
W16 (112 days)	55	[105,150]	115	112	8
W32 (224 days)	48	[196,280]	226	224	12

SD = standard deviation; W = week

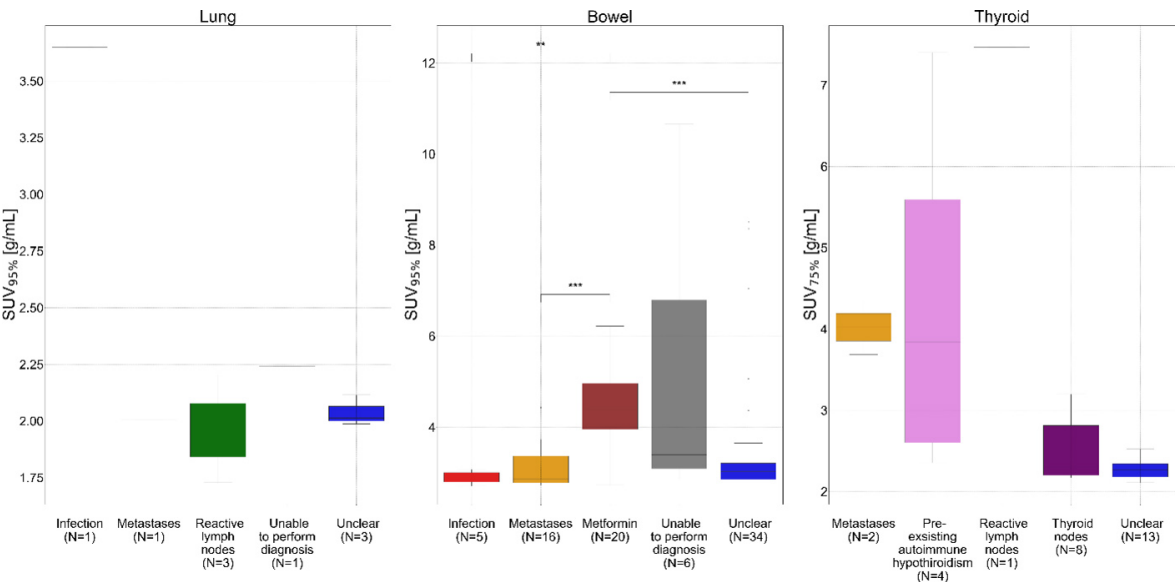


FIGURE 4. Boxplot shows values of optimal SUV percentile (SUV_{x%}) above thresholds for immune-related adverse events (irAE) diagnosis in patients without clinically diagnosed irAE in lung, bowel, and thyroid. Values of SUV_{x%} are separated based on reason for SUV_{x%} above threshold, identified from patient's chart. Each image timepoint and SUV_{x%} extracted from that image is considered as separate instance so multiple points can be contributed by a single patient. p-value was calculated using Mann-Whitney U test. P-value is shown only for pairings with statistically significant differences

* = shows p-value between 0.05 and 0.01; ** = for p-value between 0.01 and 0.001; *** = for p-value between 0.001 and 0.00001

¹⁸F-FDG PET/CT

A total of 242 ¹⁸F-FDG PET/CT scans were collected and analyzed. Table 3 shows the number of scans per scan timepoint and timing of the scans. Out of 71 patients, 3 patients had only baseline scan and no further scans due to rapid disease progression and death prior to next imaging appointment. 13 patients had only two scans (baseline and W4); 5 due to disease progression and deterioration of their performance status, 8 due to disease progression and death prior to the next imaging appointment. 7 patients had only three scans (baseline, W4 and W16) due to disease progression and death prior to next imaging appointment. 48 patients had all scans, including W32.

Detection of irAE from ¹⁸F-FDG PET/CT scan

To predict irAE status from maximum value of SUV_{x%} at any of the time points observed, the area under the receiver-operating characteristic curve (AUROC) was 0.57 in bowel, 0.7 in lung, 0.77 in thyroid. Compared to our previous analysis¹¹, the performance of the model did not change (DeLong test, p > 0.05) in all three organs (AUROC: 0.79 in bowel, 0.98 in lung and 0.88 in thyroid in ¹¹). To evaluate the role of SUV_{x%} extracted from early ¹⁸F-FDG PET/CT scan, the AUROC analysis of all time point scans was done. The AUROC for detection of adverse events in the three target organs was highest when SUV_{x%} was extracted from

W16 scan and was 0.76 for lung, 0.53 for bowel and 0.81 for thyroid. $SUV_{x\%}$ extracted from W4 scan did not improve detection of irAE compared to W16 scan (lung: $p = 0.54$, bowel: $p = 0.75$, thyroid: $p = 0.3$, DeLong test), as well as compared to W32 scan in lungs ($p = 0.32$) and bowel ($p = 0.3$). For thyroid gland, the W4 scan marginally improved detection compared to W32 scan ($p = 0.047$). Comparison of detection using W16 and W32 scan in thyroid showed, that W16 was more optimal for detection then W32 ($p = 0.02$), but this was not observed in lungs and bowel ($p = 0.09$). The ROC curves for selected $SUV_{x\%}$ extracted from W4, W16 and W32 scan for lung, bowel and thyroid are shown in Figure 1.

irAE detection by ^{18}F -FDG PET/CT versus clinical detection

Longitudinal time series of $SUV_{x\%}$ for all patients included in the study are shown in Figure 2 and Figure 3. Patients, clinically diagnosed with irPneumonitis, irColitis and irThyroiditis, are shown with patient specific colours (each plot) in Figure 2; dates of clinical diagnosis of irAE for each participant are indicated with dashed vertical lines of the matching color. Longitudinal series of $SUV_{x\%}$ for patients that did not have clinically diagnosed irAE in any of the three organs are shown in Figure 3. The gray band indicated the 95% confidence interval for $SUV_{x\%}$ of patients who did not experience irAE in¹¹.

We identified false positives cases where $SUV_{x\%}$ was above threshold without clinically confirmed irAE in any of the three organs. Values of $SUV_{x\%}$ above threshold, separated by reason, identified from patient's chart, in lung, bowel and thyroid are shown in boxplot in Figure 4. Of 9 excursions in lung, 3 (33%) remained unclear, 3 (33%) were attributed to reactive lymph nodes and 1 (11%) to infection, metastases in that area, and inability to assess causes of high uptake each. In bowel, 34 excursions of 81 (42%) remained unclear, 20 (25%) were attributed to metformin and 16 (20%), 6 (7%), 5 (6%) to metastases, inability to assess causes of high uptake and infection, respectively. In thyroid, 13 of 28 (46%) excursions were unclear, 8 (29%), 4 (14%), 2 (7%) and 1 (4%) were attributed to thyroid nodes, pre-existing autoimmune thyroiditis, metastases in that area and reactive lymph nodes respectively. Inability to assess causes for $SUV_{x\%}$ above threshold was attributed to cases where poor patient's condition and death prevented clinicians from assessing and confirming irAE. Cases where the cause could not be clarified by analysing patient's charts were assigned to the group unclear.

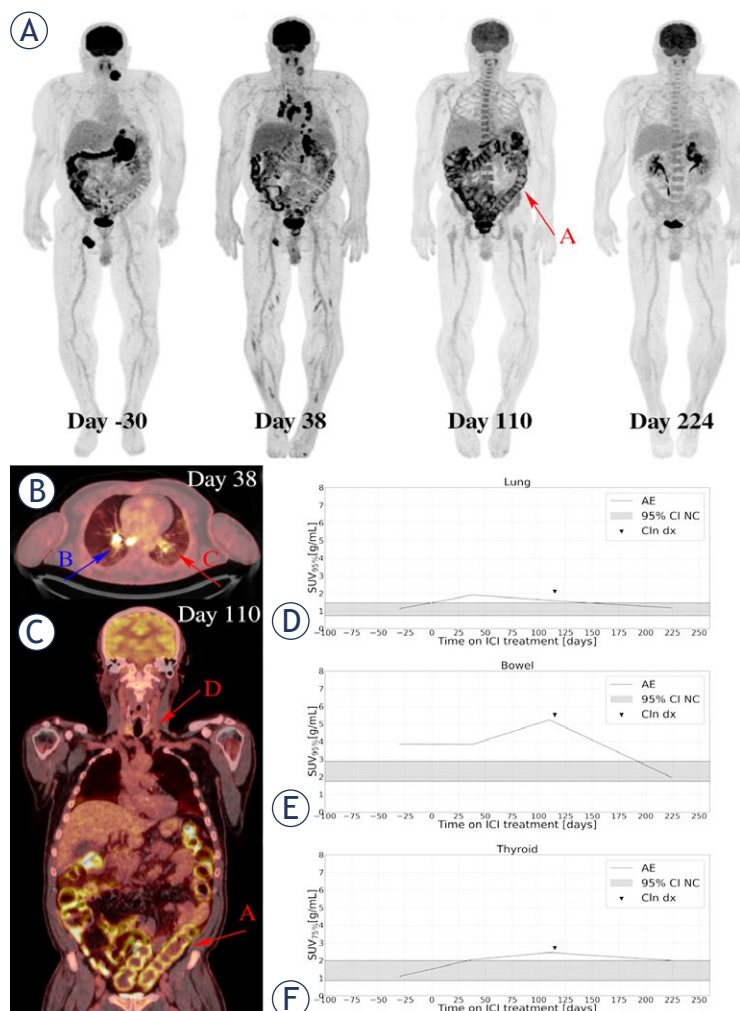


FIGURE 5. A 58-year-old patient, who was diagnosed in April 2022 with NRAS mutated melanoma of unknown primary was treated with ipilimumab/nivolumab combination. Serial ^{18}F -FDG PET maximum intensity projections were obtained per study protocol (A). The baseline image (day -30) showed sites of metastases in soft tissue, adrenal gland and spleen. Early time point (W4, day 38) image showed higher metabolic activity in mediastinal lymph nodes typical for granulomatous sarcoid like reaction, which is also seen on axial slices of the day 38 imaging study fused ^{18}F -FDG PET/CT (B, BLUE ARROW B), and diffuse lung opacities with mildly elevated ^{18}F -FDG uptake (B RED ARROW C). Patient reported of having no respiratory symptoms. Partial response in lymph nodes and in the spleen was described, and metabolic progression of the adrenal gland metastasis was observed. Day 110 (W16) image showed elevated ^{18}F -FDG uptake in the thyroid gland (C RED ARROW D), lung, and colon (C, RED ARROW A) reflecting multiple irAE. Coronal slices of the day 110 imaging study on fused ^{18}F -FDG PET/CT (C). On day 115 patient reported of having diarrhoea without respiratory or other symptoms, thyroid hormone laboratory test was pathological. Colonoscopy with biopsy was performed and examination by endocrinologist. IrColitis grade 2 and irThyroiditis grade 2 were confirmed. irPneumonitis grade 1 was confirmed based on imaging, no invasive procedure with bronchoscopy was performed for confirmation due to absence of respiratory symptoms and significant irColitis at the same time. After treatment with systemic corticosteroids, diarrhoea resolved. Images on day 224 (W32) were showing resolutions of all irAE and complete metabolic response in all metastatic lesions except progression in adrenal gland. This one metastatic lesion was surgical resected. Due to multiple irAE, the patient then stopped ICI treatment. Quantification of organ ^{18}F -FDG PET uptake demonstrated elevated uptake in lung and preceded clinical symptoms (D), elevated uptake in bowel corresponded with the time of clinical diagnosis of irColitis (E) as well as elevated uptake in thyroid corresponded with clinical diagnosis of irThyroiditis (F).

AE = adverse events(irAE)

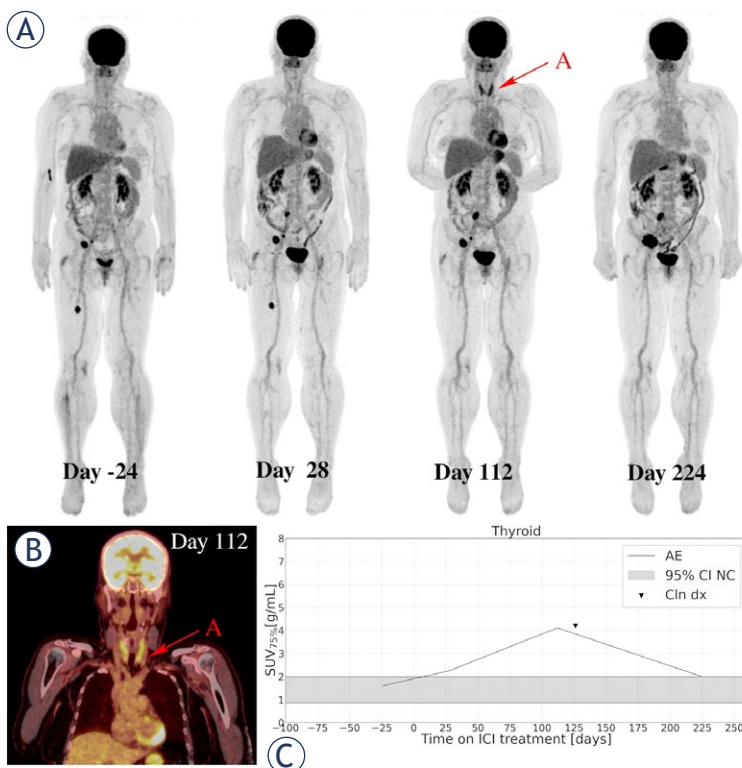


FIGURE 6. A 60-year-old female patient with metastatic *BRAF* V600E-mutated cutaneous melanoma, diagnosed in September 2021, was treated with pembrolizumab in the first-line setting. Serial ^{18}F -FDG PET maximum intensity projections were obtained per protocol (A). The baseline ^{18}F -FDG PET/CT image (day -24) showed metastatic disease present in soft tissue. Day 28 (W4) imaging study showed moderately increased colon uptake and heterogeneous response of all metastatic lesions. Patient had no symptoms of irColitis. Image at day 112 (W16) showed again a heterogeneous response of melanoma metastases and marked increase in ^{18}F -FDG uptake in the thyroid gland (RED ARROW A), which can also be observed on coronal slice of the day 112 imaging study on fused ^{18}F -FDG PET/CT (B). Patient was confirmed to have irThyroiditis on day 126, based on thyroid hormone laboratory test and examination of endocrinologist. Day 224 (W32) image showed decrease in ^{18}F -FDG uptake in the thyroid gland and a clear progression of melanoma metastases. Systemic treatment was changed to targeted treatment with BRAF and MEK inhibitor. Moderate increase in ^{18}F -FDG PET uptake in the thyroid gland was seen already on W4 ^{18}F -FDG PET, 98 days before clinical detected (C).

AE = adverse events(irAE)

Statistical analysis showed that patients on metformin had statistically significant higher $\text{SUV}_{95\%}$ in bowel compared to those identified with infection or metastases in bowel and unclear reason for higher uptake (Mann-Whitney-U test, $p < 0.05$; Figure 4). There were no statistically significant differences between other reasons and in other organs.

On average, the irAE were identified on ^{18}F -FDG-PET/CT using $\text{SUV}_{x\%}$ 38 days prior to clinical diagnosis; 48 days (range: 18, 77) for irPneumonitis, 81 (range: 17, 145) days for irColitis and 27 days

(range: -47, 159) for irThyroiditis. Statistical comparison of the timing of clinical irAE diagnosis and timing of first $\text{SUV}_{x\%}$ above the threshold for all three observed organs showed that there is no statistically significant difference ($N = 14$, $p = 0.1$, Wilcoxon signed-rank test).

Images of two patients who experienced irPneumonitis, irColitis, and irThyroiditis are highlighted in Figure 5 and Figure 6, respectively.

Discussion

This is the first study that aimed to prospectively evaluate QIB $\text{SUV}_{x\%}$ on a W4 ^{18}F -FDG PET/CT scan to detect irAE early during the systemic treatment with ICI. For now, there are only few published studies concerning the value of ^{18}F -FDG PET/CT in detecting irAE, all performed retrospectively and none of them used ^{18}F -FDG PET/CT performed at W4 for early evaluation.²⁶ Recently published retrospective study determined the organ-specific accuracy of ^{18}F -FDG PET/CT in detecting irAE in melanoma patients in adjuvant setting. They concluded that ^{18}F -FDG PET/CT has a moderate to high sensitivity and specificity for diagnosis of irAE in lung, intestines and thyroid gland.²⁷ Quantitative approach for the evaluation of ^{18}F -FDG uptake in the target organs was till now used only in our retrospective analysis.¹¹

We hypothesized that mM patients who were treated with ICI would demonstrate increased ^{18}F -FDG uptake in the organs of interest before the time of clinical irAE diagnosis, which could be detectable as early as on W4 ^{18}F -FDG PET/CT. We found that $\text{SUV}_{x\%}$ had no higher detection performance when extracted from W4 compared to W16 but had improved detection performance power when compared to W32. This might suggest that irAE are not yet visible or cannot be detected on W4 scan, but a later scan, possibly week 8, might be more appropriate for irAE detection. Interestingly enough, this is also the suggested time interval in guidelines for management of mM management.^{27,28}

For mM patients receiving ICI, ^{18}F -FDG PET/CT is primarily performed for disease response assessment. Data on optimum time of the first evaluation is scarce. In clinical practice it is usually performed after three to four cycles of ICI but can also be performed earlier or later during the course of the treatment, depending on the decision of medical oncologist and availability of ^{18}F -FDG PET/CT scans.²⁸ Of the three irAE investigated, the typical

onset times and times to resolution of irThyroiditis fall exactly within W4 to W16, making it highly likely for irThyroiditis to be detected by either of the two early scans.²⁹ For other two irAE, the onset of irAE can appear at any time during therapy, with a predominant period of 2-16 weeks from the commencement, depending on the type of ICI administered.^{6,7,30} It is hence unlikely to capture and detect irAE at its worse on routine ¹⁸F-FDG PET-CT images. Our data however shows that quantitative evidence of irAE can be seen in images as early as 80 days prior to clinical diagnosis of colitis. Due to rareness of occurrence in realistic study cohorts, multi-centre studies would be needed to describe the actual time-development curve of irAE. This could define the optimal timepoint of first on-treatment imaging for early diagnosis of irAE and earlier treatment intervention, which would in turn lower the possibility for detrimental impact of irAE on survival outcomes.³¹

In our study, there was no statistically significant difference in time of clinical diagnosis of irAE compared to its identification via $SUV_{x\%}$ extracted from ¹⁸F-FDG PET/CT. However, using $SUV_{x\%}$ irAE were on average identified as early as 81 days for irColitis, 48 days for irPneumonitis and 27 days for irThyroiditis prior to clinical diagnosis. In case of irThyroiditis, the small difference is primarily due to early clinical detection via regular laboratory control of thyroid hormones performed routinely before each ICI infusion. To a lesser extent, the short median time to onset of thyroid dysfunction, usually less than two months.²⁹ coincides well with W4 scans performed in our study. On the other hand, clinical detection of irColitis and irPneumonitis is driven by symptoms exhibited by the patient. Using $SUV_{x\%}$ irColitis and irPneumonitis could be first observed on ¹⁸F-FDG PET/CT while patient is still asymptomatic, which would enable closer monitoring and earlier treatment adjustment. This is illustrated in two cases of patients with $SUV_{x\%}$ in affected organs above threshold before clinical diagnosis of irAE in Figures 5 and 6.

The study also assessed $SUV_{x\%}$ for the detection of irAE, previously described in a pilot study from Hribernik *et al.* in a prospective setting.¹¹ Comparing AUROC curves of $SUV_{x\%}$ for predicting irAE status in the three target organs showed similar predictive power to that first reported in our past publication ($p > 0.05$). This indicates that $SUV_{x\%}$ is a consistent QIB for irAE detection, as was previously concluded in Huff *et al.*¹⁴

$SUV_{x\%}$ is not a perfect QIB for identification of irAE, as it captures any $SUV_{x\%}$ above the

threshold due to causes other than inflammation. Specifically, in the case of bowel, there are different patterns of physiological ¹⁸F-FDG uptake, and many different factors known and yet to be determined that possibly influence the physiological bowel ¹⁸F-FDG uptake, such as metformin and other antidiabetic therapy, gut microbiota, and bowel movement.³²⁻³⁴ Additionally, metastases present in organ of interest also impact ¹⁸F-FDG uptake and consequently, $SUV_{x\%}$. Although $SUV_{x\%}$ were hypothesized to be robust to metastases captured in the volume of interest, a more detailed studies, where segmented lesions would be excluded from organ uptake would help assess the impact of metastases uptake on $SUV_{x\%}$ robustness. Other factors that were found to impact value of $SUV_{x\%}$ and lead to false-positive results were also infections, reactive lymph nodes and overactive thyroid nodes before ICI treatment. Many of these possible false-positive cases were already described by Aide *et al.* in his review article.¹²

There are some limitations in this study. Variation in ¹⁸F-FDG PET/CT scan timing was present due to various reasons, mainly radiopharmaceutical outages, special interventions during the COVID-19 pandemic, and patient infections or deterioration of condition as the disease progressed. Also, not all patients were scanned on the same device, limiting quantification consistency across scanners. While potentially challenging, the inconsistencies were deemed acceptable due to comparable scanning protocols, low number of images performed elsewhere and similarity of scanners. For multicentric studies, scanner harmonization was shown to remove potential imaging bias. Another limitation is inconsistency in CNN for organ segmentation between previously published work and the present study. Our studies in consistency of AUROC are a strong indicator of $SUV_{x\%}$ robustness over segmentation methods.³⁵

Our future aim is to use prospectively gathered data from this clinical study to analyse the potential role of a W4 ¹⁸F-FDG PET/CT for prediction of response to ICI. We will evaluate the prediction of response using lesion by lesion analysis, with the possible use of lesion segmentation with CNN and the analysis of different lesion based QIB.

Conclusions

This is the first study to quantitatively analyse early timepoint ¹⁸F-FDG PET/CT using $SUV_{x\%}$ for detection of irAE in mM patients on ICI. While $SUV_{x\%}$

were found to be a consistent biomarker of irAE, its use on a W4 ^{18}F -FDG PET/CT had no higher performance compared to subsequent imaging. As the level of statistics exceeds a single-center studies, further larger studies are needed to investigate the use QIB of ^{18}F -FDG PET/CT.

Acknowledgement

The research was financially supported by The Slovenian Research Agency (ARIS), grant numbers P3-0321 and P1-0389.

References

1. Carlino MS, Larkin J, Long GV. Immune checkpoint inhibitors in melanoma. *Lancet* 2021; **398**: 1002-14. doi: 10.1016/S0140-6736(21)01206-X
2. Michielin O, van Akkooi ACJ, Ascierto PA, Dummer R, Keilholz U; ESMO Guidelines Committee. Cutaneous melanoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2019; **30**: 1884-901. doi: 10.1093/annonc/mdz411
3. Hribernik N, Boc M, Ocvirk J, Knez-Arbeiter J, Mesti T, Ignjatovic M, et al. Retrospective analysis of treatment-naïve Slovenian patients with metastatic melanoma treated with pembrolizumab - real-world experience. *Radiol Oncol* 2020; **54**: 119-27. doi: 10.2478/raon-2020-0003
4. Wolchok JD, Chiarion-Sileni V, Gonzalez R, Grob JJ, Rutkowski P, Lao CD, et al. Long-term outcomes with nivolumab plus ipilimumab or nivolumab alone versus ipilimumab in patients with advanced melanoma. *J Clin Oncol* 2022; **40**: 127-37. doi: 10.1200/JCO.21.02229
5. Ramos-Casals M, Brahmer JR, Callahan MK, Flores-Chávez A, Keegan N, Khamashta MA, et al. Immune-related adverse events of checkpoint inhibitors. *Nat Rev Dis Primers* 2020; **6**: 38. doi: 10.1038/s41572-020-0160-6
6. Postow MA, Sidlow R, Hellmann MD. Immune-related adverse events associated with immune checkpoint blockade. *N Engl J Med* 2018; **378**: 158-68. doi: 10.1056/NEJMr1703481
7. Haanen J, Obeid M, Spain L, Carbonnel F, Wang Y, Robert C, et al; ESMO Guidelines Committee. Management of toxicities from immunotherapy: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann Oncol* 2022; **33**: 1217-38. doi: 10.1016/j.annonc.2022.10.001
8. Balaji A, Zhang J, Wills B, Marrone KA, Elmariah H, Yarchoan M, et al. Immune-related adverse events requiring hospitalization: Spectrum of toxicity, treatment, and outcomes. *J Oncol Pract* 2019; **15**: e825-e834. doi: 10.1200/JOP.18.00703
9. Lopci E, Hicks RJ, Dimitrakopoulou-Strauss A, Dercle L, Iravani A, Seban RD, et al. Joint EANM/SNMMI/ANZSNM practice guidelines/procedure standards on recommended use of ^{18}F FDG PET/CT imaging during immunomodulatory treatments in patients with solid tumors version 1.0. *Eur J Nucl Med Mol Imaging* 2022; **49**: 2323-41. doi: 10.1007/s00259-022-05780-2
10. Iravani A, Hicks RJ. Imaging the cancer immune environment and its response to pharmacologic intervention, Part 1: the role of ^{18}F -FDG PET/CT. *J Nucl Med* 2020; **61**: 943-50. doi: 10.2967/jnumed.119.234278
11. Hribernik N, Huff DT, Studen A, Zevnik K, Klaneček Ž, Enamekhoo H, et al. Quantitative imaging biomarkers of immune-related adverse events in immune-checkpoint blockade-treated metastatic melanoma patients: a pilot study. *Eur J Nucl Med Mol Imaging* 2022; **49**: 1857-69. doi: 10.1007/s00259-021-05650-3
12. Aide N, Iravani A, Prigent K, Kottler D, Alipour R, Hicks RJ. PET/CT variants and pitfalls in malignant melanoma. *Cancer Imaging* 2022; **22**: 3. doi: 10.1186/s40644-021-00440-4
13. Wong A, Callahan J, Keyaerts M, Neyns B, Mangana J, Aberle S, et al. ^{18}F -FDG PET/CT based spleen to liver ratio associates with clinical outcome to ipilimumab in patients with metastatic melanoma. *Cancer Imaging* 2020; **20**: 36. doi: 10.1186/s40644-020-00313-2
14. Huff DT, Ferjancic P, Namias M, Enamekhoo H, Perlman SB, Jeraj R. Image intensity histograms as imaging biomarkers: application to immune-related colitis. *Biomed Phys Eng Express* 2021; **7**: 10.1088/2057-1976/ac27c3. doi: 10.1088/2057-1976/ac27c3
15. Cho SY, Lipson EJ, Im HJ, Rowe SP, Gonzalez EM, Blackford A, et al. Prediction of response to immune checkpoint inhibitor therapy using early-time-point ^{18}F -FDG PET/CT imaging in patients with advanced melanoma. *J Nucl Med* 2017; **58**: 1421-8. doi: 10.2967/jnumed.116.188839
16. Sachpekidis C, Kopp-Schneider A, Pan L, Papamichail D, Haberkorn U, Hassel JC, et al. Interim ^{18}F FDG PET/CT can predict response to anti-PD-1 treatment in metastatic melanoma. *Eur J Nucl Med Mol Imaging* 2021; **48**: 1932-43. doi: 10.1007/s00259-020-05137-7
17. Anderson TM, Chang BH, Huang AC, Xu X, Yoon D, Shang CG, et al. FDG PET/CT imaging 1 week after a single dose of pembrolizumab predicts treatment response in patients with advanced melanoma. *Clin Cancer Res* 2024; **24**: OF1-OF10. doi: 10.1158/1078-0432.CCR-23-2390
18. Taghanaki SA, Zheng Y, Kevin Zhou S, Georgescu B, Sharma P, Xu D, Comaniciu D, et al. Combo loss: Handling input and output imbalance in multi-organ segmentation. *Comput Med Imaging Graph* 2019; **75**: 24-33. doi: 10.1016/j.compmedimag.2019.04.005
19. Wang H, Zhang N, Huo L, Zhang B. Dual-modality multi-atlas segmentation of torso organs from ^{18}F FDG-PET/CT images. *Int J Comput Assist Radiol Surg* 2019; **14**: 473-82. doi: 10.1007/s11548-018-1879-3
20. Adashek JJ, Subbiah IM, Matos I, Garralda E, Menta AK, Ganeshan DM, et al. Hyperprogression and immunotherapy: Fact, fiction, or alternative fact? *Trends Cancer* 2020; **6**: 181-91. doi: 10.1016/j.trecan.2020.01.005
21. National Cancer Institute (NCI). NCI Common Terminology Criteria for Adverse Events (CTCAE). Version 5.0.2021. [internet]. [cited 2023 Oct 23]. Available at: https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/ctcae_v5_quick_reference_8.5x11.pdf
22. Nelson EK, Piehler B, Eckels J, Rauch A, Bellew M, Hussey P, et al. LabKey Server: an open source platform for scientific data integration, analysis and collaboration. *BMC Bioinformatics* 2011; **12**: 71. doi: 10.1186/1471-2105-12-71
23. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 2021; **18**: 203-11. doi: 10.1038/s41592-020-01008-z
24. Baiocco S, Matteucci F, Mezzenga E, Caroli P, Di Iorio V, Vittanti C, et al. SUV95th as a reliable alternative to SUVmax for determining renal uptake in ^{68}Ga PSMA PET/CT. *Mol Imaging Biol* 2020; **22**: 1070-7. doi: 10.1007/s11307-019-01451-1
25. Olsson-Brown A, Lord R, Sacco J, Wagg J, Coles M, Pirmohamed M. Two distinct clinical patterns of checkpoint inhibitor-induced thyroid dysfunction. *Endocr Connect* 2020; **9**: 318-25. doi: 10.1530/EC-19-0473
26. Nadaraja S, Helsing J, Naghavi-Bezhad M, Land LH, Ruhlmann CH, Gerke O, et al. Value of ^{2-18}F FDG-PET/CT in identifying immune-related adverse events in patients with melanoma or non-small cell lung cancer: a systematic scoping review. *Clin Transl Imaging* 2024; **12**: 187-95. doi: 10.1007/s40336-024-00618-3
27. Gideonse BM, Birkeland M, Vilstrup MH, Grupe P, Naghavi-Bezhad M, Ruhlmann CH, et al. Organ-specific accuracy of ^{18}F FDG-PET/CT in identifying immune-related adverse events in patients with high-risk melanoma treated with adjuvant immune checkpoint inhibitor. *Jpn J Radiol* 2024; **42**: 753-64. doi: 10.1007/s11604-024-01554-y
28. Lopci E, Hicks RJ, Dimitrakopoulou-Strauss A, Dercle L, Iravani A, Seban RD, et al. Joint EANM/SNMMI/ANZSNM practice guidelines/procedure standards on recommended use of ^{18}F FDG PET/CT imaging during immunomodulatory treatments in patients with solid tumors version 1.0. *Eur J Nucl Med Mol Imaging* 2022; **49**: 2323-41. doi: 10.1007/s00259-022-05780-2
29. Chieng JHL, Htet ZW, Zhao JJ, Tai ES, Tay SH, Huang Y, et al. Clinical presentation of immune-related endocrine adverse events during immune checkpoint inhibitor treatment. *Cancers* 2022; **29**: 2687. doi: 10.3390/cancers14112687

30. Martins F, Sofiya L, Sykietis GP, Lamine F, Maillard M, Fraga M, et al. Adverse effects of immune-checkpoint inhibitors: epidemiology, management and surveillance. *Nat Rev Clin Oncol* 2019; **16**: 563-80. doi: 10.1038/s41571-019-0218-0
31. Pan EY, Merl MY, Lin K. The impact of corticosteroid use during anti-PD1 treatment. *J Oncol Pharm Pract* 2020; **26**: 814-22. doi: 10.1177/1078155219872786
32. Hamidizadeh R, Eftekhari A, Wiley EA, Wilson D, Alden T, Bénard F. Metformin discontinuation prior to FDG PET/CT: A randomized controlled study to compare 24- and 48-hour bowel activity. *Radiology* 2018; **289**: 418-25. doi: 10.1148/radiol.2018180078
33. Moasses-Ghafari B, Fallahi B, Esfehiani AF, Eftekhari M, Rahmani K, Eftekhari A, et al. Effect of diet on physiologic bowel ¹⁸F-FDG uptake. *J Nucl Med Technol* 2021; **49**: 241-5. doi: 10.2967/jnmt.120.257857
34. Kang JY, Kim HN, Chang Y, Yun Y, Ryu S, Shin H, et al. Gut microbiota and physiologic bowel ¹⁸F-FDG uptake. *EJNMMI Res* 2017; **7**: 72. doi: 10.1186/s13550-017-0318-8
35. Strašek K, Huff DT, Hribernik H, Santoro Fernandez VMA, Vincent T., Klaneček Z, et al. Assessing the impact of CNN architectures for whole organ segmentation on predictive models of organ toxicity. [abstract]. APS March Meeting 2024. March 4–8, 2024; Minneapolis & Virtual. *Bulletin of the American Physical Society* 2024. Abstract Z18.00003