A Prospective Study to Determine Inter-observer Variability of Gross Tumor Volume with [18F] Fludeoxyglucose-PET/CT Compared to CT Alone in Stage III Non-Small Cell Lung Cancer Using Three-Dimensional Analysis

David J. Peterson¹, Naseer Ahmed², Ryan Rivest⁵, Bashir Bashir⁴, Shahida Ahmed⁵, Sandor Demeter⁶, Zoann Nugent¹, Pam Brechin⁸, Boyd McCurdy⁹, William M. Hunter¹⁰, Amit Chowdhury¹¹, Leylek Ahmet¹²

1. Cancer Care Manitoba, University of Manitoba, Winnipeg, Manitoba, Canada 2. Radiation Oncology, CancerCare Manitoba, University of Manitoba, Canada 3. Cancer Care Manitoba, Winnipeg, Manitoba, Canada 4. Cancer Care Manitoba, Winnipeg, Manitoba, Canada; University of Manitoba, Winnipeg, Manitoba, Canada 5. Cancer Care Manitoba, University of Manitoba, Winnipeg, Manitoba, Canada 6. Health Sciences Centre, University of Manitoba, Winnipeg, Manitoba, Canada 7. Cancer Care Manitoba, Winnipeg, Manitoba, Canada 8. Cancer Care Manitoba, Winnipeg, Manitoba, Canada 9. Cancer Care Manitoba, University of Manitoba, Winnipeg, Manitoba, Canada 10. Radiation Oncology, Tom Baker Cancer Centre 11. Cancer Care Manitoba, University of Manitoba, Winnipeg, Manitoba, Canada 12. Cancer Care Manitoba, University of Manitoba, Winnipeg, Manitoba, Canada

Corresponding author: Naseer Ahmed, naseer.ahmed@cancercare.mb.ca

Disclosures can be found in Additional Information at the end of the article

Abstract

Purpose: This study determined the interobserver variability of Gross Tumor Volume (GTV) with [18F] Fludeoxyglucose (FDG)-PET/CT compared to CT alone in Stage III Non-Small Cell Lung Cancer (NSCLC) using 3D analysis.

Materials and Methods: Twenty-nine patients underwent simultaneous co-registered CT and FDG-PET/CT for radiotherapy planning. GTV for lung tumor and mediastinal lymphadenopathy contoured by three different radiation oncologists was compared for changes in volume and position. Interobserver variability was determined with vector displacement and Dice Similarity Coefficient (DSC). Concordance for the number of lymph nodes was determined.

Results: Mean GTV for lung tumor with FDG-PET/CT and CT alone was 62.0 cm³ and 74.64 cm³, (p=0.0005) with 17% reduction in GTV by FDG-PET/CT. Mean GTV for mediastinal lymphadenopathy was 15.72 cm³ and 19.02 cm³ (p=0.084) with 17% reduction in GTV. Mean vector displacement of lung tumor was 2.0 mm with FDG-PET/CT versus 7.1 mm with CT alone (p = 0.0016) with 3.6 fold reduction in interobserver variability. Mean vector displacement of mediastinal lymphadenopathy was 1.53 mm with FDG-PET/CT versus 10.2 mm for CT alone (p=0.0005) with 6.7 fold reduction in interobserver variability. Median DSC for the primary GTV was 0.87 for FDG-PET/CT and 0.74 for CT alone. Median DSC for nodal GTV was 0.79 and 0.59 respectively. All physicians agreed on the number of lymph nodes on CT alone in 15/29 vs. 27/29 patients on PET/CT. Only two of the three physicians agreed on the number of lymph nodes contoured for CT alone in 12/29 versus only 2/29 patients for FDG-PET/CT (p=0.0018).

Conclusion: FDG-PET/CT compared to CT alone is more precise, reduces mean lung tumor and mediastinal nodal GTV and interobserver variability. There was greater agreement for the number of lymph nodes contoured on FDG-PET/CT compared to CT alone.
Introduction

Locally advanced NSCLC remains a lethal disease with a five-year survival of only 15%. Any potential improvement in the survival is related to better locoregional control with concurrent radiation and chemotherapy [1]. New radiation technologies, including Three Dimensional Conformal Radiotherapy (3DCRT) and Intensity-Modulated Radiation therapy (IMRT), rely on the assumption of accurate and reproducible anatomical target delineation. However, defining GTV purely based on anatomical imaging obtained through CT scans may result in geographical miss and significant interobserver variability [2,3].

Positron Emission Tomography (PET) using [18F] FDG is an imaging modality that characterizes malignant disease based on specific metabolic activity of glucose within the tumors. PET detects undiscovered distant metastasis in up to 30% of patients staged with CT alone in stage III NSCLC [4,5]. In studies with pathological confirmation, average sensitivities and specificities for FDG-PET are reported as 83% and 91% respectively compared to 64% and 74% for CT alone [6].

There are no randomized trials comparing CT versus FDG-PET/CT based radiotherapy planning for lung cancer or any other disease site. However based on phase 11 studies, a convincing body of data has emerged with in last ten years incorporating the use of PET scans for radiotherapy planning in NSCLC [7]. The most significant advantage has been the use of integrated FDG-PET and planning CT to reduce interobserver variability of GTV compared to conventional CT planning alone [8-11]. A review of published data comparing changes in volume measured with FDG-PET/CT to CT alone indicates that the magnitude of treatment volume changes with incorporation of PET in radiotherapy planning for lung cancer varies from 27% - 100% [12]. However, volumetric data provides only information on changes in size and does not account for potential changes in position and shape of the target, thereby affecting variability of the GTV in NSCLC.

In this study we describe the influence of FDG-PET/CT or CT alone for the primary and mediastinal nodal disease in radiation planning for stage III NSCLC in relation to changes in volume, position and overlap of the GTV. We report the interobserver variability between radiation oncologists for FDG-PET/CT and CT alone-derived GTV. In addition to volumetric measurements, we have used a vector displacement method for three-dimensional (3D) positional analysis of the volumes. We have confirmed our results by evaluating the overlap of the primary and nodal GTV with the DSC method. We also determined concordance among radiation oncologists for the number of lymph nodes contoured at similar nodal stations.

Materials And Methods

Statement of Ethics

The University of Manitoba, Research Ethics Board, approved this study. Informed consent was obtained from all patients for being included in the study. All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008.

Patients

Patients with stage III NSCLC (n=29) were recruited between April 2007 and August 2009. Mean patient age was 62 years (range 48-79). Patients were staged based on bronchoscopy, mediastinoscopy, diagnostic contrast enhanced CT, bone scan, pulmonary function tests and CT or MRI of the brain according to the AJCC 6th edition TNM classification. Inclusion criteria included pathologically proven and confirmed by FDG-PET/CT stage IIA or IIIB NSCLC with radical intent of treatment involving concurrent chemotherapy and radiotherapy, or radical radiotherapy alone, ECOG 0-2, age > 18, FEV1 > 1.0L, no prior or concurrent malignancies and no prior radiotherapy to
the thorax.

**FDG-PET/CT**

Our PET imaging center has a dedicated Siemens Biograph 16 HR PET/CT scanner. The FDG-PET/CT acquisition refers to the sequential acquisition of a low dose anatomic localization and attenuation correction non-contrast CT followed by a FDG-PET scan. FDG-PET/CT image acquisition parameters and patient dosing were defined in the protocol (Appendix 1). Patients were imaged on a flat radiation therapy planning bed in treatment position with immobilization using a Uvex shell. Calibrated external beam lasers were used for precise positioning of the patient during the FDG-PET/CT scans. A data transfer protocol from the PET center to the treatment planning application was used with Dicom and image fusion. Two data sets were transferred to Eclipse planning station and named according to the modality to facilitate automatic registration of the two images. Since the FDG-PET and CT images were obtained with the same device with the same patient orientation and shared the same DICOM coordinates, only minor adjustments were required during the image fusion process. The fused image sets were checked by the radiation therapist and radiation oncologist to ensure the accuracy of the fusion. Any fusions errors were manually corrected by the radiation therapist and the radiation oncologist.

**GTV Delineation**

Physicians were provided with a contouring protocol. In the first phase of the study, three thoracic radiation oncologists blinded to the results of the FDG-PET scan delineated the target volumes for the primary tumor and nodal GTV on the CT alone in the Eclipse planning station. The primary GTV was contoured using a lung window setting and the nodal volumes were contoured using a mediastinal window setting. Any modification based on bronchoscopy, mediastinoscopy or initial diagnostic scan was left to the discretion of the contouring radiation oncologist. Physicians were not allowed to access FDG-PET images or reports issued by the nuclear medicine physician while contouring on the CT images.

In the second phase of the study, FDG-PET/CT information was included with all other relevant clinical information. The same physicians now delineated the GTV on the fused FDG-PET/CT images available on the same Eclipse planning station. They also had access to the PET/CT report from Nuclear Medicine. Physicians were able to visualize both components of the FDG-PET/CT (PET and CT) by using a sliding bar available on the Eclipse software. The edge of the lung GTV was delineated using the anatomical edge defined by the CT component of the FDG-PET/CT. In situations where there was collapse, consolidation or non-specific changes seen in the lung parenchyma, contouring physicians were advised to use their best clinical judgment to define the edge of the GTV. Similarly for the mediastinum, the edge of the nodal GTV was defined by the CT. This process was carried out by switching between FDG-PET and CT windows on each slice separately. This resulted in four different contours for each of the 29 patients by each of the three physicians: GTVpCT, GTVpPET, GTVnCT and GTVnPET (p = primary lung tumor, n = nodal disease). This was not an interventional study and the volume used for treatment planning was left the up to the discretion of the radiation oncologist.

**Volume Analysis**

The volume of each GTV contour was evaluated using the built-in volume measurement tool in the Eclipse treatment planning system.

**Vector Displacement Analysis**

Vector displacement determines changes in position of the GTV in three dimensions by measuring the distance between the centers of the contoured volumes. The treatment planning computer initially calculated the position of the geometric center of each GTV contour. The distance between the centers of two different physician GTV contours was evaluated by calculating the magnitude of the 3D vector connecting the two central positions. Vector displacement measures positional change but is independent of the volume. If one volume is twice as large as another but their centers are in
the same position, the resulting vector displacement would still be zero. The greater the distance between the centers of the target volumes, the farther apart the tumors are.

**Dice Similarity Coefficient (DSC) Analysis**

The overlap of physician GTV contours was evaluated using DSC. The DSC of two contours delineated by two different physicians that have volumes $V_A$ and $V_B$, respectively, is defined as,

$$DSC = \frac{(V_A \cap V_B)}{V_A + V_B}$$

The metric has possible values ranging from 0 for no overlap to 1 for perfect agreement between the two contours. It is unique in that it is dependent on not only the magnitude of the GTV volumes, but also their intersecting or overlapping volume.

**Primary End Points**

The primary endpoints of the study included interobserver variability and modification of GTV for size (volume in cm$^3$), position in three-dimensional space (vector displacement) and overlap (DSC). Interobserver variability was calculated by comparing the GTV between the three radiation oncologists for changes in volume and position using DSC and vector displacement for the lung tumor and mediastinal nodal disease separately. The contours of physician A were compared to physician B, physician B to C and C to A and the mean of the three were reported. The physicians were blinded to each other’s contours. For some patients, the number of nodal stations included in the GTVnCT and GTVnPET contours differed amongst physicians. In these cases, only the similar nodal stations were included in the evaluation.

**Statistical Analysis**

The impact of FDG-PET/CT fusion on interobserver variability of GTV contouring for the primary and nodal disease for each patient was determined by comparing GTV contours and analyzed using a two sided paired t-test with a probability value of $p \leq 5\%$ for significance. Statistical analysis was done with SAS$^{TM}$ version 9.1.

**Results**

**Volume Analysis**

The mean lung tumor GTV for each physician (A through C) was 60.54 cm$^3$, 62.10 cm$^3$ and 63.39 cm$^3$ with FDG-PET/CT and 84.35 cm$^3$, 80.81 cm$^3$ and 58.75 cm$^3$ with CT alone. The FDG-PET/CT decreased the volume of the lung tumor compared to CT alone in 86%, 90% and 72% of patients for physician A, B and C respectively (mean = 83%). The mean lung tumor GTV of all three physicians with FDG-PET/CT was 62.01 cm$^3$ and 74.64 cm$^3$ with CT alone ($p = 0.0005$) (Fig. 1 & 2) resulting in 17% reduction in GTV using FDG-PET/CT compared to CT alone. Primary tumors displaying atelectasis had larger volumes (123.48cm$^3$ for CT and 102.91cm$^3$ for FDG-PET/CT) than those free of atelectasis (44.79cm$^3$ for CT and 37.01cm$^3$ for FDG-PET/CT). Both showed a 17% reduction in GTV with FDG-PET/CT. The mean mediastinal lymph node GTV for each physician was 16.95 cm$^3$, 15.25 cm$^3$ and 14.96 cm$^3$ with FDG-PET/CT and 22.24 cm$^3$, 20.18 cm$^3$ and 14.65 cm$^3$ with CT alone. The mean mediastinal lymph node GTV volume for all three physicians was 15.72 cm$^3$ with FDG-PET/CT and 19.02 cm$^3$ using CT alone ($p= 0.084$). This also resulted in a 17% reduction in the GTV for the mediastinal lymph nodes using FDG-PET/CT. (Fig. 1 & 3)
FIGURE 1: Mean GTV for both the Lung Tumor and Nodal disease is significantly smaller for FDG-PET/CT compared to CT alone.

FIGURE 2: Mean Lung Tumor Volumes by the individual patient, of all the patients
Vector Displacement Analysis

The mean vector displacement of the lung tumor was 2.0 mm with FDG-PET/CT and 7.1 mm with CT alone (p = 0.0016) (Fig. 4). This resulted in a 3.6 fold reduction in interobserver variability using FDG-PET/CT as compared to CT alone. The mean vector displacement of the mediastinal lymphadenopathy was 1.53 mm with FDG-PET/CT and 10.2 mm for CT alone (p= 0.0006) (Fig. 4) resulting in a 6.7 fold reduction in interobserver variability using FDG-PET/CT as compared to CT alone. The magnitude and variability in vector displacement between the GTV is much smaller for both the lung tumor and the mediastinal lymphadenopathy for the FDG-PET/CT data (Fig. 5). The GTV with FDG-PET/CT for the lung tumor and nodal disease is tightly grouped and smaller in magnitude compared to CT independent of the specific contouring physician. For the 11 patients with atelectasis, mean vector displacement was 11.6 mm for CT and 2.8mm for PET (P=0.012) resulting in 4.1-fold reduction in interobserver variability. For the 15 patients without atelectasis mean vector displacement was 3.8 mm for CT and 1.4 mm for PET and (P=0.017) resulting in 2.7 fold reduction in interobserver variability.
FIGURE 5:
Mean Vector Displacement of all 29 patients. Each symbol represents one patient. The first two lines demonstrate the difference in variability between nodal volumes with and without FDG-PET/CT. Note that in the top line (X = GTV Node PET/CT) all of the GTVs are grouped together with little variation between different patients. The last two lines demonstrate the difference in variability between the lung tumors with and without FDG-PET/CT. Again, note on the third line of the y-axis (□ = GTV Lung PET/CT) all of the GTVs are grouped together with little variation between different patients. This is in stark contrast to the GTVs contoured with CT alone where there is significant variation in Vector Displacement between patients.

**Dice Similarity Coefficient (DSC) Analysis**

When all patients and physician pairs were considered, the median DSC for the primary GTV contours was 0.87 for FDG-PET/CT and 0.74 for CT alone. Median values for the nodal GTV contours were 0.79 and 0.59, respectively (Fig. 6). Individual DSC values for each patient and observer combination are plotted in (Fig. 7) for the primary tumors. The vertical and horizontal axes represent the FDG-PET/CT and CT-alone values, respectively. The FDG-PET/CT DSC value was greater than its corresponding CT alone value in approximately 85% of cases. There was significantly (p < 0.0001) better agreement among the radiation oncologists for the FDG-PET/CT contours compared to CT alone.
Lymph Node Analysis

The number of involved mediastinal lymph nodes contoured on FDG-PET/CT was 49, 47 and 47 by physicians A, B and C respectively compared to 50, 49 and 34 by CT alone. The mean number of nodes contoured by CT alone was 1.53 versus 1.64 by FDG-PET/CT, (p=0.55) but the difference was not significant. However, the range of variation among physicians in contouring nodal disease is significantly less with FDG-PET/CT. All three physicians agreed on the number of lymph nodes contoured on CT in 15/29 patients and in 27/29 patients on PET/CT. Two of the three contouring physicians agreed on the number of lymph nodes contoured in 12/29 patients on CT alone versus only 2/29 patients on PET/CT (P=0.0018).
Discussion

This study demonstrates the utility of FDG-PET/CT in target delineation in radiotherapy treatment planning and the enhanced precision of GTV in NSCLC. In RTOG 0515 the mean GTV for PET/CT was 86.2 cm$^3$ vs. 98.7 cm$^3$ for CT alone with a 12.5% decrease in GTV with PET/CT [13]. Our results suggest greater variability in position, rather than size. We found a 17% decrease in GTV volume with FDG-PET/CT but a 72% reduction of interobserver variability in position for the lung tumor.

Previously reported small studies have shown that FDG-PET/CT modifies the GTV defined by CT [11, 14-24]. Most of these studies determined only the interobserver variability of target volumes for changes in size, not position and overlap. The only study to date that evaluated interobserver variability of target delineation with respect to changes in position using 3D analysis with FDG-PET/CT in lung cancer was by Steenbakkers [10]. The study involved eleven radiation oncologists who delineated the GTV of twenty-two patients. In the first phase they delineated the GTV on CT only. In the second phase the GTV was delineated on a matched FDG-PET/CT scan. The observer variation was computed in three dimensions by measuring the distance between the median GTV surfaces of each individual GTV. They found that the observer variation was reduced from 1.0 cm for CT only to 0.4 cm for matched FDG-PET/CT. Our results are consistent with the results of this investigation.

In our study, 11 of 29 (38%) patients had a significant amount of atelectasis. Tumors with and without associated atelectasis had similar reduction in GTV (17%) with PET/CT. But there was more significant reduction in interobserver variability for the tumors with atelectasis compared to the tumors with no atelectasis. (4.1 vs 2.7 fold). 38% of patients having these changes are similar to what is seen in clinical practice and therefore our results have good external validity.

Based on the European Organization for Treatment and Research of Cancer recommendations, standardized protocols with PET/CT acquired in treatment position with rigid immobilization are required for radiotherapy planning of lung cancer [25]. Our FDG-PET/CT simulator acquired simultaneous co-registered imaging in the treatment position with laser alignment and immobilization in every patient. FDG-PET/CT acquisition and image transfer with Dicom and image fusion parameters were strictly followed by a protocol. All patients were contoured with a standardized protocol, using Eclipse version 10.6. We used visual interpretation rather than SUV threshold method to define GTV on FDG-PET/CT, as so far this is the best contouring method described in the literature [26].

We have further evaluated our results using DSC. DSC is defined as the intersection volume between volume A and volume B divided by the mean of volumes of A and B [27-29]. This metric has possible values ranging from 0 for no overlap to 1 for perfect agreement between the two contours. This parameter initially described in ecologic studies has been used to analyze interobserver variation to delineate clinical target volume (CTV) and Organs at Risk (OAR) in breast cancer to quantify the effect of a consensus contouring protocol [28]. In our study (Fig. 7) for the primary lung and nodal GTV, DSC was higher for FDG-PET/CT as compared to CT alone further confirming reduced interobserver variability among the physicians with FDG-PET/CT.

As a secondary analysis for the contoured lymph nodes, there was a higher degree of agreement among physicians for the number of nodes contoured on FDG-PET/CT versus CT alone. There was no significant difference between the mean number of nodes for CT alone vs. PET/CT. We have not used CT IV contrast to delineate the GTV for nodal disease and it was difficult for the contouring physicians to differentiate between the nodal conglomerates from the individual nodes on CT alone. This may be the reason for slightly lower number of nodes contoured on CT as compared to PET/CT. But this is least likely to affect the results of our data considering the magnitude of GTV change by size and vector displacement influenced by PET component of the study.

Ideally, to compare the clinical effect of two imaging modalities one would create dose volume histograms, and have long term follow-up to demonstrate any impact on local control and survival. The authors acknowledge that we lack such data. However we believe that our FDG-PET/CT
imaging acquisition and contouring tool as described in this study are precise, reliable and reproducible and can be used for future clinical studies.

**Conclusions**

FDG-PET/CT significantly reduces GTV and interobserver variability, in primary lung tumors and mediastinal lymphadenopathy and has a higher degree of agreement among physicians for the number of nodes contoured. Thus FDG-PET/CT is significantly more precise both quantitatively and qualitatively compared to CT alone in defining target volumes for patients with stage III NSCLC.

**Appendix: 1**

**CT Parameters for PET (Siemens Biograph 16 HR)**

*Near Whole body / Whole body*

Topogram:
- mA = 50
- KV = 120
- Tube position = AP

CT:
- Eff mAs 100 for <100 kg
- 120 for >100 kg and 140 for WB
- Care Dose enabled
- KV = 120
- Rotation time = 0.5 sec
- Feed/rotation = 22.1 mm
- Collimation = 1.5 mm
- Pitch = 0.92 (Feed/rotation / collimation (1.5) / 16 slices)
- Slice = 5.0 mm & Kernel = B40f medium

**Positron Emission Scan Acquisition Parameters**

*Near Whole body / Whole body / Head & Neck*

- 3 min / bed position <100 kg
- 4 min/bed >100 kg (if tolerable)
- Image matrix = 168
- Zoom = 1.0
- Voxel size = 2.0 X 4.06 X 4.06 (Z, X, Y)
- FOV Diameter (mm) Whole body is Axial = 162 mm and Transaxial = 585
- 2.3 cm overlap in bed positions

*Recon Parameters*

- OSEM: Iterations = 4, Subsets = 8
- Filter FWHM = 5.0 mm
- XY and Z Filter = Gaussian
- PET Slice will match CT slice location

**Additional Information**

**Disclosures**
Conflicts of interest: The authors have declared that no conflicts of interest exist.

Acknowledgements
First authorship is shared between David Peterson and Naseer Ahmed. We would like to acknowledge Dr. Alia Norman for her contribution in the development of initial protocol.

References


