AI-based quantification of PET/CT lesions is associated with survival in patients with lung cancer

Pablo Borrelli
Sahlgrenska University Hospital: Sahlgrenska universitetssjukhuset
Jose Luis Loaiza Gongora
Sahlgrenska University Hospital: Sahlgrenska universitetssjukhuset
Reza Kaboteh
Sahlgrenska University Hospital: Sahlgrenska universitetssjukhuset
Johannes Ulen
Malmö Högskola: Malmo Universitet
Olof Enqvist
Chalmers University of Technology: Chalmers tekniska hogskola
Elin Trägårdh
Lund University: Lunds Universitet
Lars Edenbrandt (lars.edenbrandt@gu.se)
University of Gothenburg: Goteborgs Universitet https://orcid.org/0000-0002-0263-8820

Research Article

Keywords: Computer-assisted analysis, Tumor burden, Total lesion glycolysis, Prognosis

Posted Date: March 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-305562/v1

Abstract

Purpose

The use of modern artificial intelligence (AI) based methods offers new possibilities for automated and objective image analysis of PET/CT studies. We aimed to develop an AI tool for segmentation and quantification of tumor burden in [^{18}F]-fluorodeoxyglucose (FDG) PET/CT studies and to evaluate the association between the AI-based measurements and overall survival (OS) in patients with lung cancer.

Methods

A total of 320 consecutive patients referred for FDG PET/CT due to suspected lung cancer were retrospectively selected for the study. Two nuclear medicine specialists manually segmented abnormal FDG uptake in all PET/CT studies. One third of the patients were assigned to a test group. Survival data was collected for this group. An AI tool based on convolutional neural networks was trained to segment lung tumors and thoracic lymph nodes. The AI tool includes a previously trained convolutional neural network, which segments organs in the CT. Total lesion glycolysis (TLG) was calculated based on the AI-based and manual segmentations. Associations between TLG and OS were investigated using a univariate Cox proportional hazards regression model.

Results

Both AI-based TLG (Hazard ratio 1.64, 95% confidence interval 1.21–2.21; p = 0.001) and manual TLG (Hazard ratio 1.54, 95% confidence interval 1.14–2.07; p = 0.004) were significantly associated with OS.

Conclusion

Fully automated AI-based TLG measurements of PET/CT studies showed significant association with OS in patients with lung cancer. This type of measurements may be of value in the management of future patients with lung cancer. Our AI tool is available on reasonable request for research purposes at the RECOMIA platform (https://recomia.org).

Introduction

[^{18}F]fluorodeoxyglucose (FDG) positron emission tomography/computed tomography (PET/CT) has an important role for lung cancer, both for small cell and for non-small cell cancer diagnosis, staging, response assessment, and follow up [1–4]. Several studies have shown the prognostic value of different metabolic PET-parameters [5–10]. The methods to quantify tumor burden are, however, usually based on manually selected lesions by local imaging experts and easily accessible measurements such as

maximum or peak standardized uptake value (SUV). An objective way to analyze the PET/CT studies would make it easier to compare the results from different studies and thereby to facilitate the assessment of the value of FDG-PET/CT in patients with suspected lung cancer. The use of modern artificial intelligence (AI) based methods offers new possibilities for automated and objective image analysis [11]. AI-based methods can also be trained to assess the entire burden of disease by including both extent and activity of the tumor and not only maximum or peak SUV, which represents a very small volume of the tumor [12]. Manual methods to assess the total tumor burden, for example segmentations of all tumor lesions and estimation of total lesion glycolysis (TLG), are too time-consuming for clinical use and hampered by low reproducibility.

We have used convolutional neural networks (CNN) for the analysis of PET/CT studies in prostate cancer patients. Automatically calculated measurements of extent and activity of the tumor in the prostate gland [13] and in bone metastases [14] showed to be associated with overall survival (OS).

It has also been shown that physicians supported by an AI tool for automated analysis of PSMA-PET/CT studies showed significantly less inter-reader variability in the quantification of primary prostate tumors and bone metastases than when performing a completely manual analysis [15].

We have also used CNN for the detection of lung lesions in PET/CT from patients with lung cancer in a recent pilot study [16]. Based on the experience from previous studies we aimed to develop an AI tool for segmentation and quantification of tumor burden in FDG PET/CT studies and to evaluate the association between the AI-based TLG measurements and OS in patients with lung cancer.

Methods

Patients

Two groups of consecutive patients referred for FDG PET/CT due to suspected lung cancer were retrospectively selected to develop and evaluate the AI tool. One group of 113 patients underwent PET/CT between April 2008 and December 2010 at the Sahlgrenska University hospital, Gothenburg, Sweden. This group has been used in our pilot study to train and evaluate a CNN for detection of lung tumors [16]. The other group of 207 patients underwent PET/CT between November 2017 and November 2018 at the Skåne University Hospital in Lund/Malmö, Sweden.

The total study group of 320 patients was divided into a test group of 106 patients (33%) and a training group of 214 patients (67%). Only patients from the Skåne University Hospital were selected randomly for the test group, since the patients from the Sahlgrenska University Hospital already had been used to train and test an AI-tool in our previous study. Clinical information and survival data for the test group were collected from the local medical records and radiology information system up until November 2020.

The study was conducted according to the principles expressed in the Declaration of Helsinki, approved by the local research ethics committees at Gothenburg (#295-08) and Lund Universities (#2016/193 and

#2018/753).
All patients provided written informed consent.

Imaging

PET/CT scans were obtained using integrated PET/CT systems (Siemens Biograph 64 Truepoint, Siemens Healthineers, Erlangen, Germany and GE Discovery MI, GE Healthcare, Chicago, United States). The patients were injected with 4 MBq/kg (maximum of 400Mbq) of FDG, fasted for at least 4 h prior to the injection and had adequate glucose levels prior to injection. The accumulation time was 60 min. Images were acquired with 3 min per bed position (Sahlgrenska) or 1.5 min per bed position (Skåne) from the base of the skull to the mid-thigh.

PET images obtained from the Siemens Biograph 64 Truepoint PET/CT scanner were reconstructed with a slice thickness of 3 mm with an iterative ordered subset expectation maximization 3D algorithm (four iterations, eight subsets), matrix size 168 x 168. CT-based attenuation and scatter corrections were applied. A low-dose CT scan (64-slice helical, 120 kV, 30 mAs, 512 x 512 matrix) was obtained covering the same part of the patient as the PET scan. The CT was reconstructed using a filtered back projection algorithm with slice thickness and spacing matching the PET scan.

For images obtained from the GE Discovery MI systems, PET images were reconstructed using the commercially available block-sequential regularized expectation maximization (BSREM) algorithm Q.Clear (GE Healthcare, Milwaukee, WI, USA) with a beta factor of 550. The time-of-flight and point spread functions were used, with a 256 x 256 matrix (pixel size 2.7 x 2.7 mm2, slice thickness 2.8 mm). CT images were acquired for attenuation correction and anatomic correlation of the PET images. A diagnostic CT with intravenous and oral contrast or a low-dose CT without contrast was performed. In our clinical routine, a low-dose CT is performed if a previous diagnostic CT has been performed within 4 weeks. For diagnostic CTs, tube current modulation was applied by adjusting the tube current for each individual with a noise index of 42.25 and a tube voltage of 100 kV. For low-dose CT, the tube voltage was 120 kV with a noise index of 45. If a diagnostic CT was performed, it was used for attenuation correction (delayed venous phase of intravenous contrast). The adaptive statistical iterative reconstruction technique (ASiR-V) was applied for all CT reconstructions.

Manual segmentations

Two nuclear medicine specialists with > 6 and > 12 years of PET/CT experience together segmented abnormal FDG uptake in the PET/CT studies of the training and test groups. Uptakes were classified into one of the following groups: lung tumor, thoracic lymph node, extra-thoracic lymph node, adrenal, bone, or liver metastasis, inflammatory uptake, high pleura uptake, or other high uptake. No clinical data, only PET/CT images were available during the segmentation process. The cloud-based annotation tool (RECOMIA, https://www.recomia.org) was used for the manual segmentations [17].

AI-tool

The AI tool was trained to segment lung tumors and thoracic lymph nodes only. The other types of abnormal uptake were represented by an insufficient number of examples for CNN training.

The model is based on a CNN with the U-net architecture [18]. The final convolutional layer contains three channels with softmax activation, one for background, one for lung tumor and one for thoracic lymph node. The network has three separate inputs, the CT image, the PET image and a one-hot encoded organ mask constructed using the model from [17]. The purpose of the organ mask is to help the network with a rough anatomical localization of a given uptake. It uses one channel each for bone, liver, lung, heart, aorta, and adrenal gland.

The model is trained using patches of minimal size, where the patches are chosen with care to give a good balance between different classes.
All pixels are divided in four groups; background, lung tumor, thoracic lymph node and other abnormal uptake (this group included extrathoracic lymph nodes, metastases, inflammatory uptake, high pleura uptake, and other high uptake). The center point of each patch is chosen randomly with equal probably to be inside any of the four groups.

Training was done with 100 epochs using 10,000 patches per epoch. Categorical cross-entropy was used as the loss function, and the optimization was performed using the Adam method [19] with Nesterov momentum. After this phase, the resulting model was applied to the training set. The model was then retrained with 20% of the patches focusing on pixels incorrectly classified by the model. These steps were repeated for times. Finally lung tumors and throcaic lymph nodes with TLG below 0.1 was removed.

Statistical analysis

Associations between TLG and OS were investigated using a univariate Cox proportional hazards regression model. OS was calculated from the date of the PET/CT study to the date of death or last follow-up. Hazard ratios (HR) and 95% confidence intervals (CI) were estimated. The level of significance was set at 0.05. The TLG measurements had a skewed distribution and were log10 transformed after adding 1.0 to handle zeros.

The TLG measurements for the Kaplan-Meier analysis were categorized according to the corresponding median value, higher vs. lower than the median.
The statistical analysis was performed in R (version 4.0.3).

Results

The test group comprised 106 patients (59 female and 47 male), who had a median age of 76 years (IQR 61–79). A total of 51 patients died during the follow-up period, with a median survival time from the PET/CT study of 0.9 years (IQR 0.54–1.54). The group of 55 patients that were still alive had a median follow-up time from the PET/CT study of 2.6 years (IQR 2.5–2.7). Clinical information from the local medical records and radiology information system showed that 85 test patients had non-small cell lung

carcinoma, 11 lung cancer of unknown type, 5 lung metastases, and 5 other diagnoses (hamartoma, lymphoma, pneumoconiosis, schwannoma, unknown).

TLG measurements

Uptake from lung tumors and thoracic lymph nodes were used to compute a total TLG for each patient, both from the AI-based and manual segmentations. Figure 1 shows a Bland-Altman plot comparing AI-based and manual TLG. The mean difference between AI-based and manual TLG was 20 with 95% limits of agreement from − 327 to 366. Both AI-based TLG (HR 1.64, 95% CI 1.21–2.21; p = 0.001) and manual TLG (HR 1.54, 95% CI 1.14–2.07; p = 0.004) were significantly associated with OS in univariate proportional regression Cox analyses.

Kaplan-Meier curves are shown in Fig. 2. The 53 patients with AI-based TLG above the median value had a significantly shorter survival time than the 53 patients with values below the median. The median survival times for the two groups were 1.57 years vs. not reached after 3 years of follow-up. The logrank test showed a significant difference in survival times ($p < 0.001$). The corresponding median survival values for the two groups stratified using manual TLG were 1.75 years vs. not reached after 3 years of follow-up (p = 0.002).

Lesion classification results

The AI tool and the physicians classified 99/106 (93%) patients regarding presence of lung tumor in the same way, 94 as positive and 5 as negative. Four patients were classified as positive only by the AI tool. The AI detections in these four cases were classified as thoracic lymph nodes (2), high pleura uptake (1), and inflammatory uptake (1) by the physicians. Three patients were classified as positive only by the physicians. A total of 135 lung tumors were detected by both the AI tool and the physicians. These tumors had a median TLG of 13.9 (IQR 2.3-181.7). A total of 56 lesions with a median TLG of 0.4 (IQR 0.1–1.6) were detected as lung tumors by the physicians only. Seven of these lesions were classified as thoracic lymph nodes by the AI tool. A total of 153 lesions with a median TLG of 0.6 (IQR 0.3–2.3) were detected as lung tumors by the AI tool only. Thirty-five of these lesions were classified as thoracic lymph nodes by the physicians.

The AI tool and the physicians classified 71/106 (67%) patients regarding presence of thoracic lymph nodes in the same way, 59 as positive and 12 as negative. Thirty-five patients were classified as positive only by the AI tool. Both AI tool and physicians detected 170 thoracic lymph nodes. These lesions had a median TLG of 5.8 (IQR 1.5–27.9). A total of 70 lesions with a median TLG of 0.3 (IQR 0.1–0.9) were detected as thoracic lymph nodes by the physicians only. Five of these lesions were classified as lung tumors by the AI tool. A total of 274 lesions with a median TLG of 0.8 (IQR 0.3–2.4) were detected as thoracic lymph nodes by the AI tool only. Thirty-two of these lesions were classified as lung tumors and 23 as extra-thoracic lymph nodes by the physicians.

Discussion

The aim of this study was to assess if an AI tool can be used to calculate TLG measurements, which are clinically relevant in patients with lung cancer. The results show that an AI tool can be trained to automatically segment lung tumors and thoracic lymph nodes and calculate TLG measurements, which are significantly associated with OS. The AI-based TLG measurements showed similar prognostic performance as corresponding manual measurements.

In a clinical setting a physician would be able to check AI-based segmentations and dismiss false positive lesions and to add false negative lesions. The use of AI support for the calculation of total lesion uptake has shown to reduce inter-reader variability significantly for the analysis of prostate lesions and bone metastases in PSMA PET/CT [15]. This process is also more timesaving compared to a completely manual segmentation process and feasible also for a clinical setting.

In research, this type of AI tools may facilitate comparisons of studies from different centers, pooling data within multicenter trials and performing meta-analysis, when the objective evaluation is applied rather than local image readers.

Objective AI-based TLG measurements may be useful, not only for prognostic evaluation of patients with lung cancer at the time of diagnosis, but also for monitoring treatment response. FDG-PET/CT findings have shown to be associated with a clinical benefit in patients with non-small cell lung cancer receiving immunotherapy [20].

AI tools have been used as computer aided detection (CAD) systems to highlight potential lesions in chest x-ray and lung CT [21]. The aim is to decrease the possibility of a radiologist missing tumors with subsequent delay in diagnosis. Such an AI tool needs to have high sensitivity and low false positive rate to be of clinical value. The aim of our AI tool was not to be a CAD system, but a tool for automated quantification of tumor burden. The sensitivity and false positive rate of our AI tool was not sufficient to be used as a clinical CAD system. The disagreement between the physicians and the AI tool was in most cases found in lesions with low TLG. A common reason for disagreement was physician classifying a lesion as lung tumor and the AI tool as a mediastinal/hilar lymph node or vice versa. This is not always an easy decision even for experienced physicians.

The retrospective design of this study is on the one side a limitation of the study, but gave us the opportunity to assess performance of the AI tool not only compared to manual interpretations of the same images, but to the independent reference OS. The training material was too small to train CNN to detect other abnormal uptake than lung tumors and thoracic lymph nodes. Extra-thoracic lymph nodes and distant metastases will be added in future versions of our AI tool. Further development will also include localization of lymph nodes into established anatomical definitions and classifications into ipsilateral versus contralateral regions [22]. This type of information may improve the prognostic value of an AI analysis.

Conclusion

Fully automated AI-based TLG measurements of PET/CT studies showed significant association with OS in patients with lung cancer. This type of measurements may be of value in the management of future patients with lung cancer. Our AI tool is available on reasonable request for research purposes at the RECOMIA platform (https://recomia.org).

Declarations

Funding

The study was financed by grants from the Knut and Alice Wallenberg foundation, the Medical Faculty at Lund University, Region Skåne and the Swedish state under the agreement between Swedish government and the county councils, and the ALF-agreement (ALFGBG-720751).

Conflict of interest

The authors declare no conflict of interest.

Availability of data and material

Data is available on reasonable request.

Code availability

The AI tool presented in this study is available on reasonable request for research purposes at the RECOMIA platform (https://recomia.org).

Authors´contribution

Data collection was performed by RZ. Image segmentation was performed by PB, and JLLG. AI tool development was performed by JU and OE. ET and LE contributed to the design and interpretation of the data. All authors read and approved the final manuscript.

Ethics approval

The study was conducted according to the principles expressed in the Declaration of Helsinki, approved by the local research ethics committees at Gothenburg (#295-08) and Lund Universities (#2016/193 and #2018/753).

Consent to participate and publication

All patients provided written informed consent.

References

1 Volpi S, Ali JM, Tasker A, Peryt A, Aresu G, Coonar AS. The role of positron emission tomography in the diagnosis, staging and response assessment of non-small cell lung cancer. Ann Transl Med. 2018;6(5):95.

2 Carretta A, Bandiera A, Muriana P, Viscardi S, Ciriaco P, Samanes Gajate AM, et al. Prognostic role of positron emission tomography and computed tomography parameters in stage I lung adenocarcinoma. Radiol Oncol. 2020; 54(3): 278-284.

3 Martucci F, Pascale M, Valli MC, Pesce GA, Froesch P, Giovanella L, et al. Impact of 18F-FDG PET/CT in staging patients with small cell lung cancer: A systematic review and meta-analysis. Front Med (Lausanne). 2020;6:336.

4 Humbert O, Cadour N,Paquet M, Schiappa R, Poudenx M, Chardin D, et al. 18FDG PET/CT in the early assessment of non-small cell lung cancer response to immunotherapy: frequency and clinical significance of atypical evolutive patterns. Eur J Nucl Med Mol Imaging. 2020;47:1158–1167.

5 Chang H, Lee SJ, Lim J, Lee JS, Kim YJ, Lee WW. Prognostic significance of metabolic parameters measured by 18F-FDG PET/CT in limited-stage small-cell lung carcinoma. J Cancer Res Clin Oncol. 2019;145:1361–1367.

6 Chardin D, Paquet M, Schiappa R, Darcourt J, Bailleux C, Poudenx M, et al. Baseline metabolic tumor volume as a strong predictive and prognostic biomarker in patients with non-small cell lung cancer treated with PD1 inhibitors: a prospective study. J Immunother Cancer. 2020;8:e000645. doi:10.1136/ jitc-2020-000645

7 Jin F, Qu B, Fu Z, Zhang Y, Han A, Kong L, et al. Prognostic value of metabolic parameters of metastatic lymph nodes on 18F-FDG PET/CT in patients with limited-stage small-cell lung cancer with lymph node involvement. Clin Lung Cancer. 2017;19:e101-8.

8 Seban RD, Mezquita L, Berenbaum A, Dercle L, Botticella A, Le Pechoux C, et al. Baseline metabolic tumor burden on FDG PET/CT scans predicts outcome in advanced NSCLC patients treated with immune checkpoint inhibitors. Eur J Nucl Med Mol Imaging. 2020;47:1147–1157.

9 Zer A, Domachevsky L, Rapson Y, Nidam M, Flex D, Allen AM, et al. The role of 18F-FDG PET/CT on staging and prognosis in patients with small cell lung cancer. Eur Radiol. 2016;26:3155–3161.

10 Geredeli C, Artac M, Kocak I, Koral L, Sakin A, Altinok T, et al. The prognostic significance of the 18F-fluorodeoxyglucose positron emission tomography/computed tomography in early-stage nonsmall cell lung cancer. J Can Res Ther. 2020;16:816-821.

11 Sibille L, Seifert R, Avramovic N, Vehren T, Spottiswoode B, Zuehlsdorff S, et al. 18F-FDG PET/CT uptake classification in lymphoma and lung cancer by using deep convolutional neural networks. Radiology. 2020;294:445-452.

12 Høilund-Carlsen PF, Piri R, Gerke O, Edenbrandt L, Alavi A. Assessment of total-body atherosclerosis by PET/computed tomography. PET Clin. 2021;16:119-128.

13 Polymeri E, Sadik M, Kaboteh R, Borrelli P, Enqvist O, Ulén J, et al. Deep learning based quantification of PET/CT prostate gland uptake: association with overall survival. Clin Physiol Funct Imaging. 2020;40:106-113.

14 Lindgren Belal S, Sadik M, Kaboteh R, Hasani N, Enqvist O, Svärm L, et al. 3D skeletal uptake of 18F sodium fluoride in PET/CT images is associated with overall survival in patients with prostate cancer. EJNMMI Res 2017;7:1.

15 Edenbrandt L, Borrelli P, Ulén J, Enqvist O, Trägårdh E. Automated analysis of PSMA-PET/CT studies using convolutional neural networks. medRxiv preprint doi: https://medrxiv.org/cgi/content/short/2021.03.03.21252818v1

16 Borrelli P, Ly J, Kaboteh R, Ulén J, Enqvist O, Trägårdh E, et al. AI-based detection of lung lesions in18F-FDG PET-CT from lung cancer patients. EJNMMI Physics (Accepted for publication 2021)

17 Trägårdh E, Borrelli P, Kaboteh R, Gillberg T, Ulén J, Enqvist O, et al. RECOMIA-a cloud-based platform for artificial intelligence research in nuclear medicine and radiology. EJNMMI Phys. 2020;7:51.

18 Çiçek Ö., Abdulkadir A., Lienkamp S.S., Brox T., Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin S., Joskowicz L., Sabuncu M., Unal G., Wells W. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science, vol 9901. Springer, Cham. https://doi.org/10.1007/978-3-319-46723-8_49

19 Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.

20 Humbert O, Cadour N, Paquet M, Schiappa R, Poudenx M, Chardin D, et al. 18FDG PET/CT in the early assessment of non-small cell lung cancer response to immunotherapy: frequency and clinical significance of atypical evolutive patterns. Eur J Nucl Med Mol Imaging 2020;47:1158–1167.

21 Fazal MI, Patel ME, Tye J, Gupta Y. The past, present and future role of artificial intelligence in imaging. Eur J Radiol 2018;105:246-250.

22 UyBico SJ, Wu CC, Suh RD, Le NH, Brown K, Krishnam MS. Lung cancer staging essentials: The new TNM staging system and potential imaging pitfalls. RadioGraphics 2010;30:1163–1181.

Figures

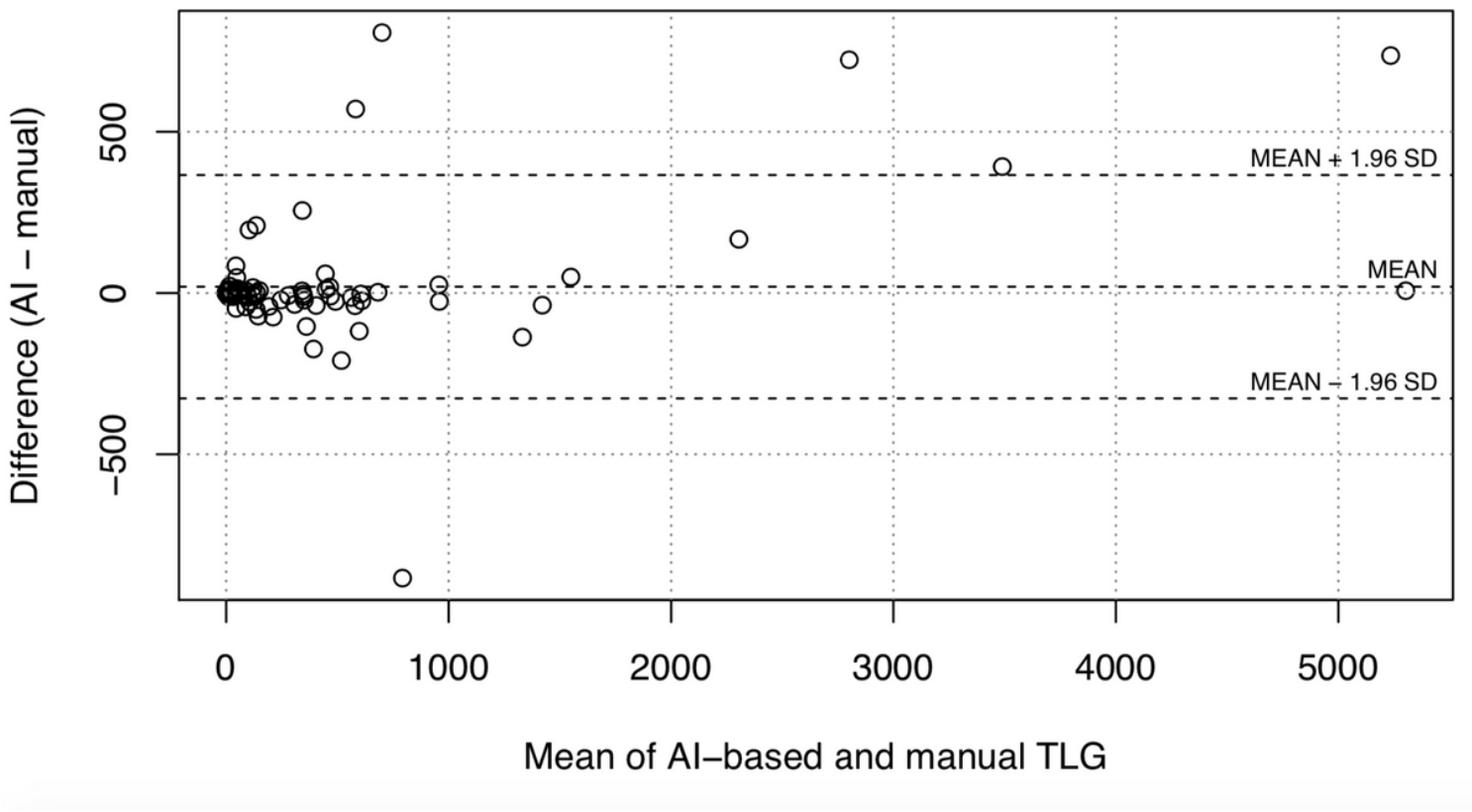

Figure 1

Bland-Altman plot comparing AI-based and manual TLG measurements.

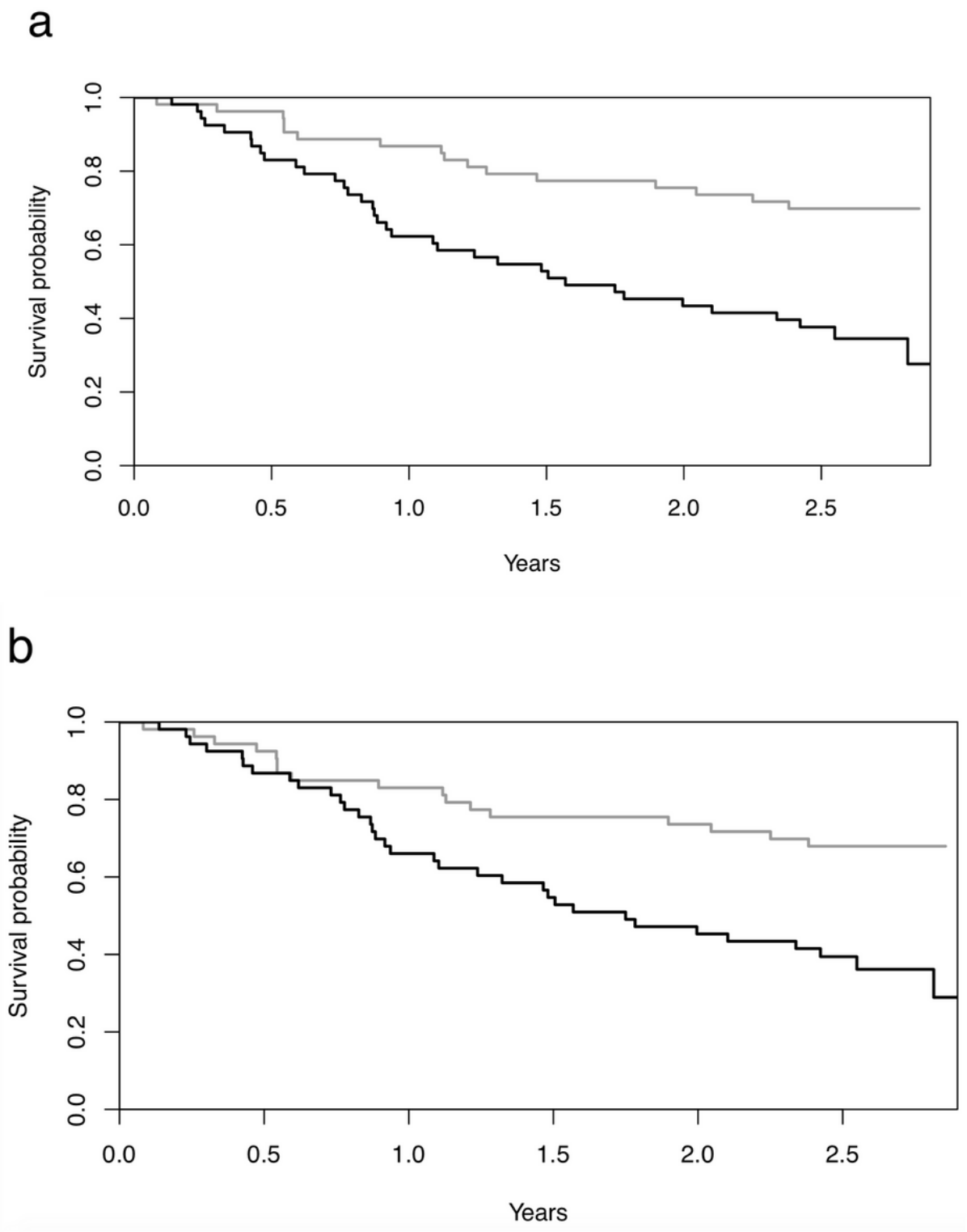

Figure 2

Kaplan-Meier curves for the groups higher (black line) vs. lower (grey line) than the median TLG values for AI-based TLG (a) and manual TLG (b).