QMENTA

QMENTA Imaging Hub for Medical Research:
A Complete Use Case from Tool Integration to Biomarker Evaluation

Whitepaper

QMENTA Imaging Hub for Medical Research: A Complete Use Case from Tool Integration to Biomarker Evaluation

Introduction

Approaching medical imaging research at a large-scale can seem like an obstacle course of challenges across all disciplines and modalities. Researchers must navigate a labyrinth of methodological crossroads, high computational hurdles, and convoluted data security regulations.

In this whitepaper, we present the QMENTA Imaging Hub as a solution to these challenges of medical imaging research. Using several positron emission tomography (PET) image analysis tools integrated into the QMENTA Imaging Hub as examples, we demonstrate how the platform can be used in practice to streamline the path from raw data to results and to ultimately enhance research processes.

Setting the Scene

A large proportion of medical imaging research is focused on the pursuit of biomarkers: objective indicators of biological or pathogenic processes or responses to an intervention. Imaging biomarkers in particular are lauded for their non-invasive nature, real-time capabilities, and cost-effectiveness, offering clinicians insights beyond those of visual assessment. However, the process of biomarker development and consequently the realization of these three benefits is complicated by another trio, in the need for precision, accuracy and reproducibility.

With its high sensitivity and specificity for biochemical processes in vivo, PET emerges as an important imaging modality for the study of neurological diseases and assessment of pharmacological interventions. Quantitative PET in particular shows great biomarker promise, but as foreshadowed, lacks accuracy and reproducibility in practice. Reproducibility issues are sweeping neuroimaging research more generally[1]. The problem lies partly with the vast number of methodological choices researchers are faced with when processing complex data such as PET. PET data analysis is a multi-step procedure involving image reconstruction, partial volume correction, kinetic modeling, and more, all of which can be implemented differently depending on the study design and context. Every choice is important for interpretation. Couple this with high computational demands, and replication and validation efforts become difficult.

To date, many attempts at validating PET quantification tools have involved in-house or commercially available software. One approach towards the validation of a software tool is to compare its results with those of other tools, particularly established ones such as PMOD[2]. Comparisons can be performed across multiple platforms[3,4], exclusively in a local environment[5] or through a combination of the two[6]. In any case, researchers may have to deal with clunky GUIs, limited or inflexible functionality, and computing infrastructures inadequate for big data analyses. Above all, in the absence of a centralized location in which to store data and run analyses, the road to results is a logistically disjointed one. So it is unsurprising that similar undertakings involving open-source analysis software - from which a wider research community could benefit - are few and far between.

[1]CM Bennett, MB Miller and GL Wolford. Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: an argument for multiple comparisons correction. Neuroimage. 2009. Vol. 47:S125. DOI: 10.1016/S1053-8119(09)71202-9

[2]Mikolajczyk K, Szabatin M, Rudnicki P, Grodzki M, Burger CA. JAVA environment for medical image data analysis: initial application for brain PET quantitation. Med Inform (Lond). 23:207–14.

[3]Colin Wilson, Reed Selwyn, Saeed Elojeimy. Comparison of PET/CT SUV Metrics Across Different Clinical Software Platforms at UNMH

[4]Roh, H.W., Son, S.J., Hong, C.H. et al. Comparison of automated quantification of amyloid deposition between PMOD and Heuron. Sci Rep 13, 9891 (2023). https://doi.org/10.1038/s41598-023-36986-5

[5]Schwarz, C. G., Therneau, T. M., Weigand, S. D. et al. (2021). Selecting software pipelines for change in flortaucipir SUVR: Balancing repeatability and group separation. NeuroImage, 238, 118259. https://doi.org/10.1016/j.neuroimage.2021.118259

[6]Tjerkaski, J., Cervenka, S., Farde, L., & Matheson, G. J. (2020). Kinfitr - an open-source tool for reproducible PET modelling: validation and evaluation of test-retest reliability. EJNMMI research, 10(1), 77. https://doi.org/10.1186/s13550-020-00664-8

The QMENTA Imaging Hub is well-poised for undertakings such as this. Researchers can implement image analysis plans with ease using a unified solution that:

- Executes analysis tools in a scalable **cloud** environment. No need to set up virtual machines or build/maintain local clusters.

- Efficiently stores and organizes data in projects that facilitate **collaboration**.

- Complies with strict industry **security standards** such as HIPAA, GDPR, ISO 27001 and FDA Part 11 & 820. The platform itself is FDA 510(k) cleared.

- **Readily documents** all analysis steps to simplify integration of code modifications or re-running analyses on updated data.

- Grants access to an existing **ecosystem of biomarker tools** for reproducible research, all discoverable in an extensive catalog.

Here we integrate three open-source tools for PET imaging analysis, and in doing so illustrate the power of the QMENTA Imaging Hub to support biomarker development and imaging research more generally. Secondarily, we emulate a typical validation exercise, comprising a quantitative evaluation of the three tools by comparing the outputs of a PET kinetic modeling analysis obtained using real-life and simulated data.

The QMENTA Imaging Hub in Action

I. Tool Integration

QMENTA's Medical Imaging and Research Team set out to integrate several open-source brain PET-MRI analysis tools into the platform. Numerous tools with varying capabilities are available, but we limited the search only to those offering the possibility of quantification with kinetic modeling (KM). This comprises a series of mathematical models for estimating the concentration of a target molecule (i.e. the radioligand receptor) in a region of interest (ROI). With this in mind, the list was narrowed down to three:

1. **PETSurfer**
   A module developed within the popular software package FreeSurfer for end-to-end PET analysis, directly using the outputs of a prior recon-all segmentation to support steps including motion correction, registration and partial volume correction. Three choices of kinetic model are available: MRTM1, MRTM2 and Logan.

2. **Dynamic PET**
   A Python-dependent command line-based tool offering denoising and SUVR calculation, plus several variations of the Simplified Reference Tissue Model (SRTM) KM method. Additional models may be accessed by installing the R package *kinfitr* [7], but these were not applied here.

3. **NiftyPAD**
   A Python package featuring a range of quantitative approaches to dynamic PET analysis with both linear and non-linear KM methods, including those listed in the tools above, as well as dual-time window analyses.

Integration to the QMENTA Imaging Hub began by developing the main script for running each processing software in Python according to a conventional format: download the input data, process the data by calling binaries/scripts, upload the results data. Each script was packaged into a Docker container configured with the necessary environment which, subject to testing, was then ready to be deployed in the cloud. This 'dockerization' step is designed to facilitate 'plug-and-play' with the platform, ensuring reproducibility. Each tool was fully integrated and ready to be executed within the platform in a matter of weeks. For researchers outside of QMENTA, development and deployment can be streamlined using the QMENTA SDK.

---

[7]Matheson, G. J. (2019). Kinfitr: Reproducible PET Pharmacokinetic Modelling in R. bioRxiv: 755751. https://doi.org/10.1101/755751

QMENTA®

II. Data Collection

Next, we focused on the data. Any data used for the proposed tool comparison needed to be suitable for analysis by KM. Specifically, the inputs to kinetic models are time-activity curves which, in the context of PET, are derived from dynamic imaging acquisitions. It is generally agreed that the more frames in an acquisition, the more completely one can capture physiologic changes, and therefore more accurately derive kinetic parameters. We therefore only considered datasets with a higher number of frames. For example, ADNI PET data[8] of four, five and six frames were deemed inadequate. Of note, larger datasets also require (potentially a lot) more resources to store and process; implementing our research plan entirely within the QMENTA Imaging Hub, rather than a resource-limited local cluster, meant that this need not be a concern.

We identified and uploaded three publicly available dynamic PET-MRI datasets to the platform:

Dataset A was obtained from the NRM 2018 PET Grand Challenge[9]. It is a simulated dataset revealing the binding properties of the fictional radiotracer "[11C]-LondonPride" to an unspecified receptor, but generated using [11C]-Ro154513 scans. Consists of five 'subjects' each scanned twice, totalling 10 scans with 23 frames each.

Dataset B is taken from an investigation into levels of the enzyme COX-1 in the brain using the radioligand [11C]-PS13[10]. There are 15 subjects, each with two sessions, totalling 30 scans with 33 frames each.

Dataset C was gathered during a study aimed at quantifying rates of cerebral protein synthesis during sleep[11] using the tracer [11C]-L-Leucine. Each of the 18 participants were scanned three times during various stages of wakefulness/sleep, totalling 54 scans with 42 frames each.

III. Kinetic Modeling Analysis

The three integrated software tools offer multiple methods for KM analysis. The characteristics of a dataset are among several important considerations to make when choosing which to employ. This includes the pharmacokinetic properties of the particular radioligand, whether arterial blood data has been acquired, and the research question at hand. In our case, choice was limited by the differing sets of models available in each software tool and the focus on optimizing similarity to improve comparisons. It is argued that for the modeler there is "no right choice"; nonetheless we acknowledge the impact of the model on the results presented and reiterate their illustrative purpose in the context of presenting the QMENTA Imaging Hub.

We used what are referred to as 'reference tissue models', with binding potential (BP) as the outcome measure of interest. For both PETSurfer and NiftyPAD, Ichise's Multilinear Reference Tissue Model[12] (MRTM1) was used. Conversely, since at the time of writing MRTM1 was not available in the standalone Dynamic PET software, we opted for the SRTM model with the most successful fit to the data: Zhou's 2003 implementation of the SRTM[13].

All PET-MRI analyses in the present exercise were conducted as ROI-based analyses. To segment the ROI inputs for each of the three tools in a consistent manner, we executed FreeSurfer's recon-all as the first step in the pipeline. This ability to access and seamlessly incorporate existing, well-recognized medical image tools into any workflow within a single platform in this way is an invaluable feature of the QMENTA Imaging Hub.

[8] https://adni.loni.usc.edu/methods/pet-analysis-method/pet-analysis/

[9] Veronese M, Rizzo G, Belzunce M, et al. Reproducibility of findings in modern PET neuroimaging: insight from the NRM2018 grand challenge. Journal of Cerebral Blood Flow & Metabolism. 2021;41(10):2778-2796. doi:10.1177/0271678X211015101

[10] Kim, M. J., Lee, J. H., Juarez Anaya, F. et al. First-in-human evaluation of [11C]PS13, a novel PET radioligand, to quantify cyclooxygenase-1 in the brain. European journal of nuclear medicine and molecular imaging. 2020; 47(13), 3143–3151. https://doi.org/10.1007/s00259-020-04855-2

[11] Dante Picchioni and Kathleen C Schmidt and Inna Loutaev and Adriana J Pavletic and Carrie Sheeler and Shrinivas Bishu and Thomas J Balkin and Carolyn B Smith (2023). Rates of cerebral protein synthesis in stages of sleep. OpenNeuro. [Dataset] doi: doi:10.18112/openneuro.ds004733.v1.0.0

[12] Ichise M, Liow JS, Lu JQ, Takano A, Model K, Toyama H, Suhara T, Suzuki K, Innis RB, Carson RE. Linearized Reference Tissue Parametric Imaging Methods: Application to [11C]DASB Positron Emission Tomography Studies of the Serotonin Transporter in Human Brain. Journal of Cerebral Blood Flow & Metabolism. 2003 Sep 1;23(9):1096-112.

[13] Zhou, Y., Endres, C. J., Brašić, J. R., Huang, S. C., & Wong, D. F. (2003). Linear regression with spatial constraint to generate parametric images of ligand-receptor dynamic PET studies with a simplified reference tissue model. NeuroImage, 18(4), 975–989. https://doi.org/10.1016/s1053-8119(03)00017-x

We included data from 10 ROIs, plus a reference region: cerebellar cortex, striatum, brainstem, thalamus, hippocampus, amygdala, frontal lobe, temporal lobe, parietal lobe, occipital lobe and cingulate cortex[14]. Reference regions - from which reference tissue models get their name - should have low specific binding, as they are used as a baseline against which to measure changes in specific binding. The cerebellar cortex was chosen as the reference region in all analyses. For dataset A, the creators artificially constrained the cerebellum to be the ideal reference region. For dataset B, the cerebellum was chosen based on the assumption that COX-1 is found almost exclusively in microglia, and the density distribution of microglia is lowest in the cerebellum[15]. On the other hand, leucine has a widespread distribution throughout the brain, meaning it is likely that there is no true reference region for dataset C and a model based on blood analysis is preferable. Thus, in favoring consistency over deviations arising from the exact appropriateness of the model, we proceeded with the cerebellum, the same reference region as in datasets A and B.

As a final step in this phase of the work, we executed the KM analyses. Preprocessing was kept as minimal and consistent as possible in each case, meaning no motion or partial volume correction. The QMENTA Imaging Hub allows users to easily and flexibly adjust a tool's input settings such as the reference region, as well as the choice of model and any other arguments (e.g. PET scanner point-spread function) directly via its web interface (Figure 1). This eliminates the need to go back into your code and make specific modifications for each new analysis case. The selected settings are also recorded for future reference. Executing each of the three tools on each PET imaging session in a reproducible manner is therefore very simple. Though each analysis instance is a standalone unit, we were able to run them all in a controlled way.

In a matter of hours, the results were available in the QMENTA Imaging Cloud for further statistical analyses.

Figure 1. An example 'Settings' pop-out in QMENTA's web interface. Users can flexibly adjust tool input settings. This example is taken from the Dynamic PET tool integrated as part of the current work.

IV. Results Download & Statistics

Unified access to the results of all three integrated PET tools - alongside the input data and analysis specifications - in the Imaging Cloud made comparisons more straightforward. The PET analysis outputs were in the form of easily manipulable CSV and NIfTI files, which could be readily downloaded. QMENTA offers two methods for file download: per file in one click directly via the web, or in batch using the QMENTA Core Python library. The latter facilitates direct integration of the download with code for statistical analyses, accelerating final result acquisition.

Our statistical analysis plan to assess the correspondence between kinetic modeling results obtained via the three software tools was as follows:

Step 1: Two-way repeated measures ANOVA to test for differences between tools.

Step 2: Pearson correlation coefficient to assess the extent of a linear relationship between each pair of tools.

Step 3: Intraclass correlation coefficient (ICC) (ICC(3,k)) to evaluate the consistency of the relationship between all three tools.

Step 4: Bland-Altman analysis to assess the agreement between each pair of tools, as well as bias.

All statistical analyses were performed using Python. Statistical significance was set at $p < 0.01$, and the Bonferroni method was used to correct for multiple comparisons.

---

[14] https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation
[15] Tan, YL., Yuan, Y. & Tian, L. Microglial regional heterogeneity and its role in the brain. Mol Psychiatry 25, 351–367 (2020). https://doi.org/10.1038/s41380-019-0609-8

QMENTA®

Your Results, All in One Place

Dataset A

- We found no significant differences (α = 0.01) between BP values derived from each tool, in any of the ten ROIs, following corrections for multiple comparisons. Figure 2 compares mean BP values per tool per ROI.
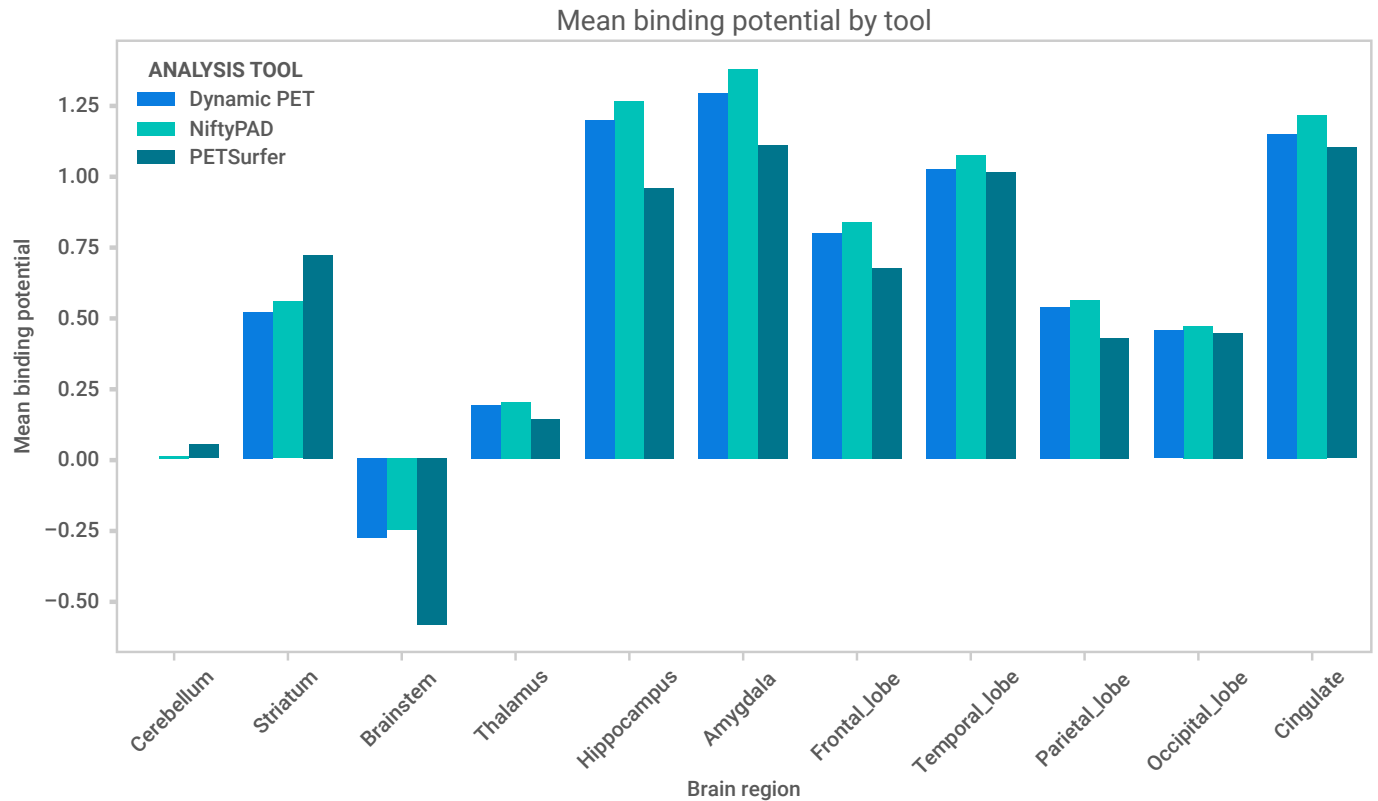

Figure 2. Bar plot depicting the mean binding potential derived from each analysis tool for each brain region in Dataset A. Ten brain regions plus the reference region are shown; in all cases, the reference region corresponds to the Cerebellum.

- There was a strong linear correlation between each pair of tools for all brain regions. Across ten ROIs, the median Pearson's r of the PETSurfer vs NiftyPAD comparison was 0.925 [0.845-0.991]. For PETSurfer vs Dynamic PET the median r was 0.907 [0.789-0.987]. For Dynamic PET vs NiftyPAD the median r was 0.994 [0.987-0.999]. The results for the frontal lobe are shown as an example in Figure 3.
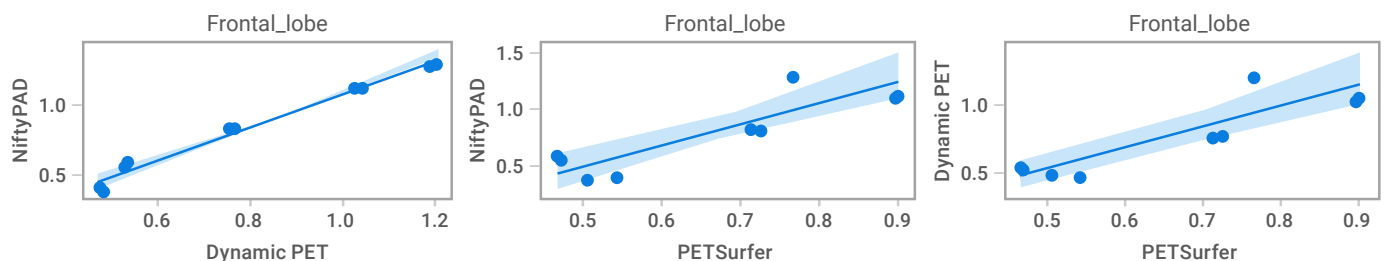

Figure 3. Scatter graphs for each of the three tool comparisons in Dataset A. Only the results of the frontal lobe are shown as an example of the correspondence between binding potential values.

- The ICC showed excellent consistency between the three tools across ROIs, with a median value of 0.947 [0.919-0.982].

- Bland-Altman plots (Figure 4) suggest a close agreement between each pair of tools for all ROIs (mean difference line close to zero), though there appears to be a very small degree of bias whereby BP values derived from PETSurfer are smaller than the other tools.
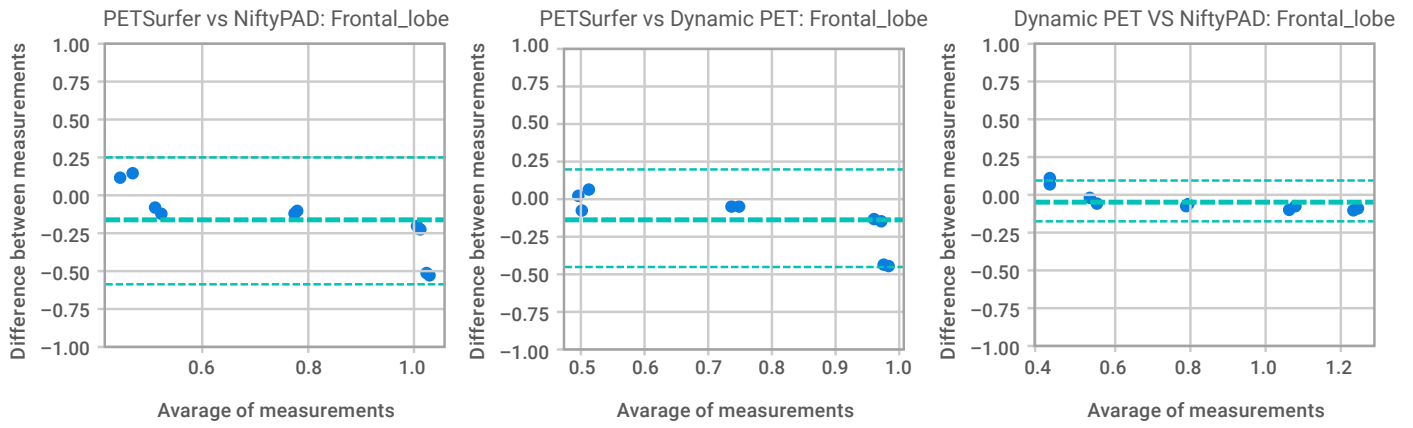
QMENTA®

Figure 4. Bland-Altman plots depicting the agreement between binding potential values for each pair of tools, taken from an example brain region in Dataset A: the frontal lobe. The thicker dashed line represents the mean difference between measurements, while the thinner dashed lines either side represent the limits of agreement, or ±1.96 standard deviations.

Dataset B

- The results of the ANOVA revealed a statistically significant main effect of the tool on BP values in dataset B in five of the 10 ROIs, as shown in Figure 5.
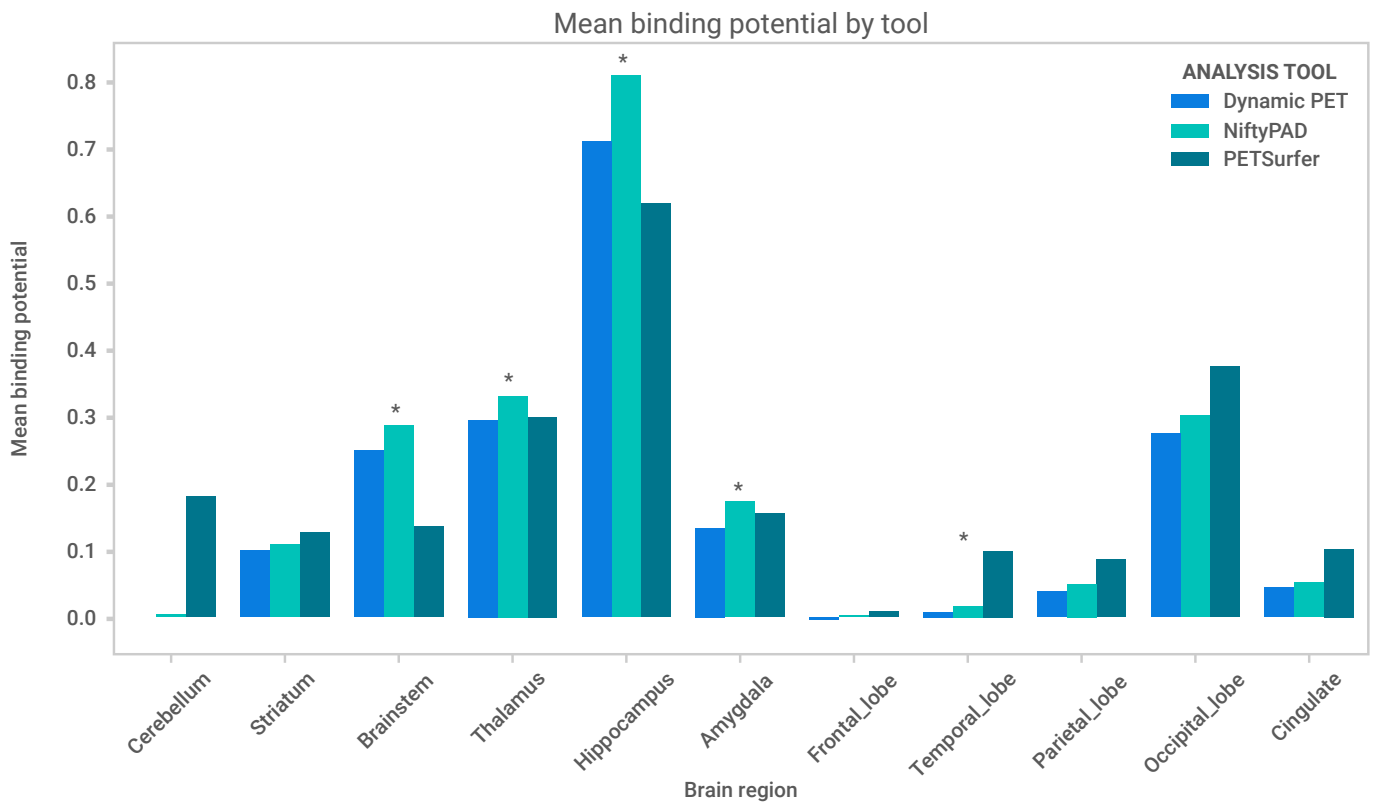

Figure 5. Bar plot depicting the mean binding potential derived from each analysis tool for each brain region in Dataset B. Ten brain regions plus the reference region are shown; in all cases, the reference region corresponds to the Cerebellum. * indicates brain regions in which the ANOVA revealed a statistically significant main effect of analysis tool.

QMENTA®

- The linear correlation between each pair of tools varied among different brain regions, though it was generally positive and high. The Dynamic PET vs NiftyPAD comparison showed the strongest correlation, evident in Figure 6, with a median Pearson's r across ROIs of 0.947 [0.857-0.978]. For PETSurfer vs Dynamic PET the median was 0.699 [-0.019-0.959]. For PETSurfer vs NiftyPAD the median was 0.667 [0.050-0.895].
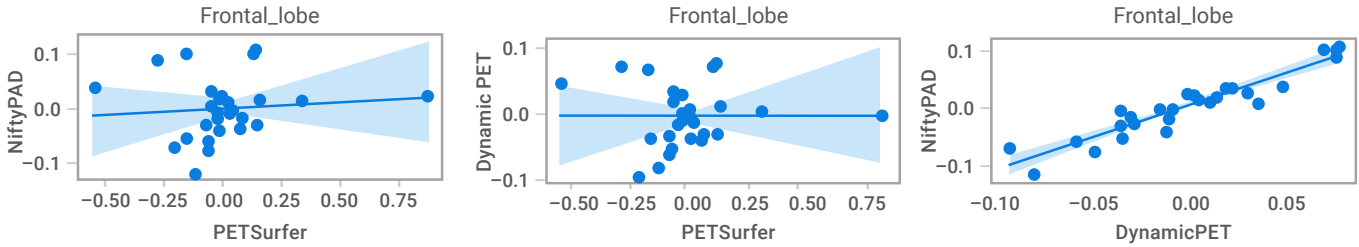

Figure 6. Scatter graphs for each of the three tool comparisons in Dataset B. Only the results of the frontal lobe are shown as an example of the correspondence between binding potential values.

- The ICC showed variable consistency between the three tools across ROIs, ranging from excellent in the hippocampus to poor in the frontal lobe, with a median value of 0.838 [0.086-0.980].
-
- Bland-Altman plots (Figure 7) suggest that the agreement between Dynamic PET and NiftyPAD BP values is close for all brain regions. Comparing both of these tools with PETSurfer, the extent of agreement appears to be more variable, with some brain regions, including the frontal lobe and cingulate cortex, showing a proportional bias. The limits of agreement are wide in these regions, suggesting poor agreement.
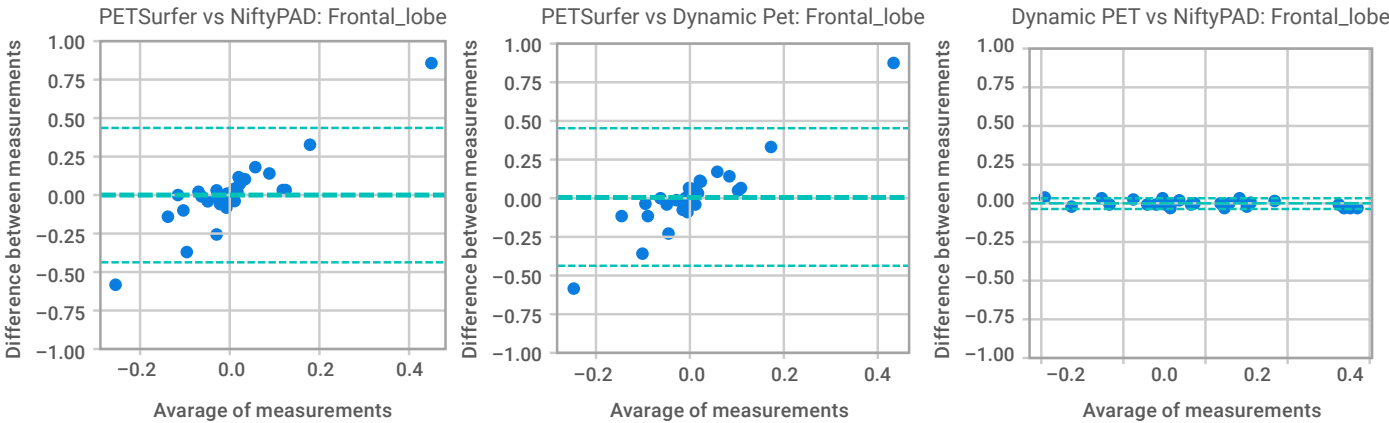

Figure 7. Bland-Altman plots depicting the agreement between binding potential values for each pair of tools, taken from an example brain region in Dataset B: the frontal lobe. The thicker dashed line represents the mean difference between measurements, while the thinner dashed lines either side represent the limits of agreement, or ±1.96 standard deviations.

QMENTA®

Dataset C

- There was a statistically significant main effect of analysis tool on BP in four out of 10 ROIs in dataset C after Bonferroni correction: the brainstem, temporal lobe, parietal lobe and occipital lobe, depicted in Figure 8.
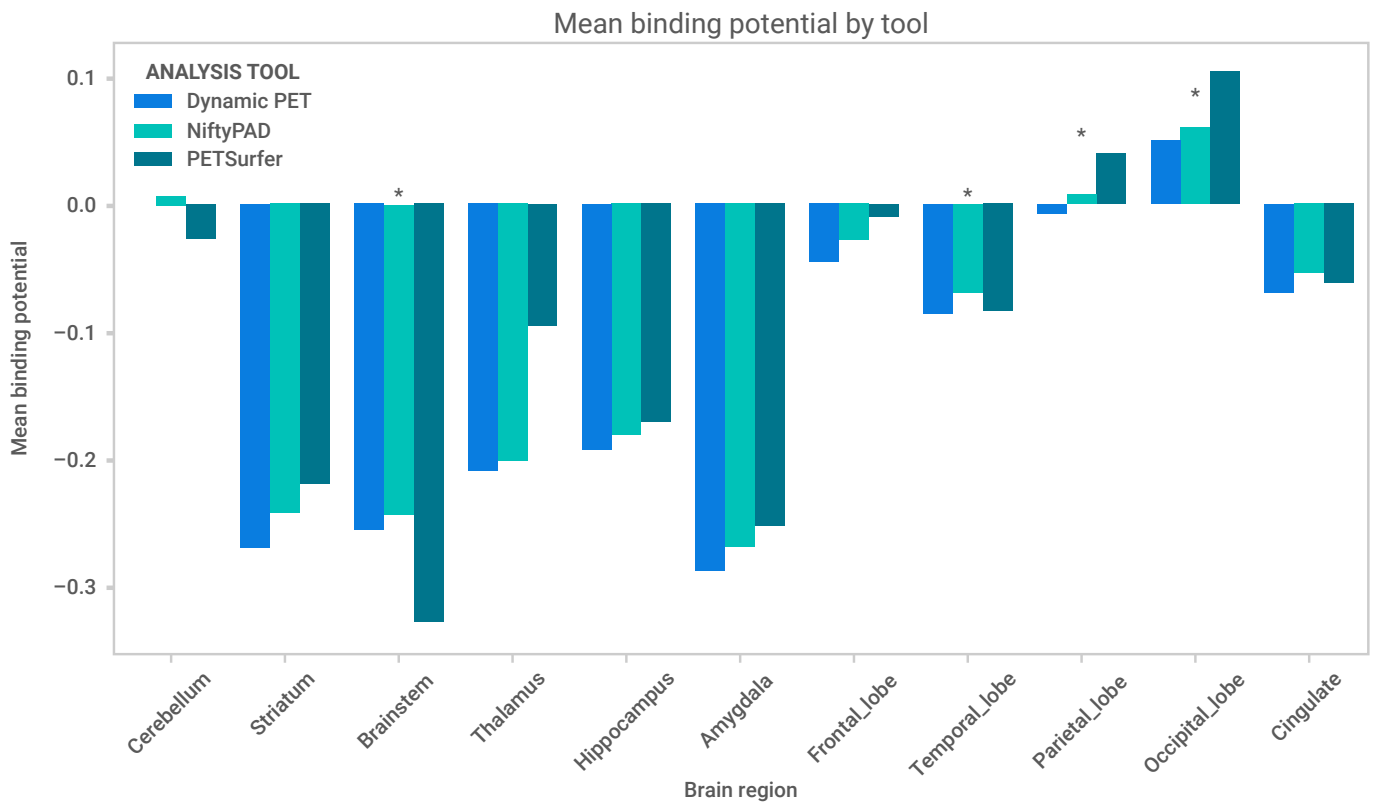

Figure 8. Bar plot depicting the mean binding potential derived from each analysis tool for each brain region in Dataset C. Ten brain regions plus the reference region are shown; in all cases, the reference region corresponds to the Cerebellum. * indicates brain regions in which the ANOVA revealed a statistically significant main effect of analysis tool.

- Linear correlation was highly variable across brain regions in the PETSurfer vs NiftyPAD and PETSurfer vs Dynamic PET comparisons, with a median Pearson's r of 0.327 [-0.179-0.648] and 0.334 [-0.143-0.640] respectively, ranging from weaker in subcortical regions to stronger in cortical ROIs. By contrast, it was consistently stronger in the Dynamic PET vs NiftyPAD comparison, with a median r of 0.971 [0.939-0.985]. Scatter plots for the frontal lobe are shown in Figure 9.
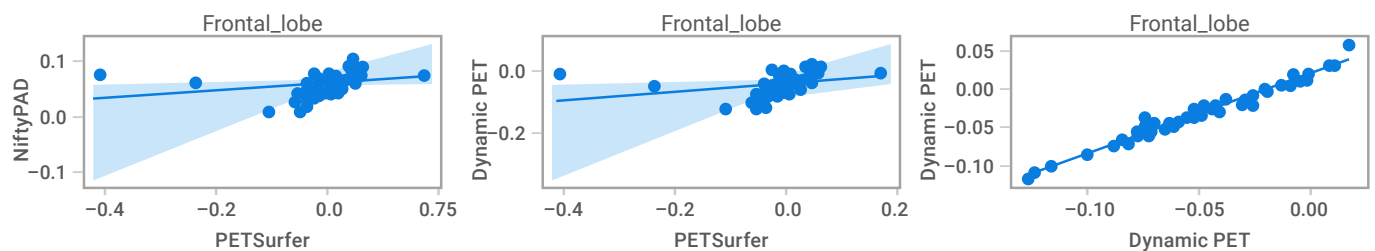

Figure 9. Scatter graphs for each of the three tool comparisons in Dataset C. Only the results of the frontal lobe are shown as an example of the correspondence between binding potential values.

QMENTA

- As in dataset B, consistency among results varied across brain regions, with stronger ICC scores in cortical ROIs than subcortical ROIs. The median ICC was 0.617 [-0.045-0.890].
- Bland-Altman plots (Figure 10) demonstrate excellent agreement between BP values derived from NiftyPAD and Dynamic PET analyses, with mean difference lines close to zero and very narrow limits of agreement. Comparisons involving PETSurfer show somewhat poorer agreement, particularly in subcortical ROIs, where a degree of bias is evident.
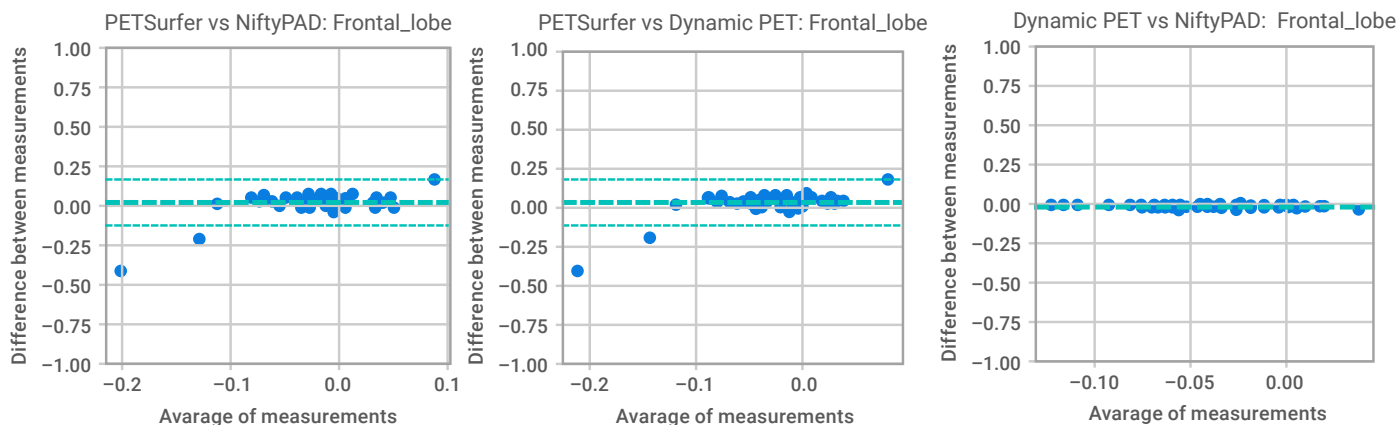

Figure 10. Bland-Altman plots depicting the agreement between binding potential values for each pair of tools, taken from an example brain region in Dataset C: the frontal lobe. The thicker dashed line represents the mean difference between measurements, while the thinner dashed lines either side represent the limits of agreement, or ±1.96 standard deviations.

Overall, the kinetic modeling results from the three chosen PET analysis tools show varying degrees of correspondence across datasets. The highest correspondence across tools is evident in dataset A, the only simulated dataset. In datasets B and C, poorer correspondence in some ROIs appears to be driven by discrepancies with PETSurfer BP values which require further investigation. Lower agreement may also be a reflection of suboptimal KM model choices for datasets that use tracers without established reference regions, as discussed. For researchers asking whether there are differences in quantification using the same data, these comparisons demonstrate that this can depend on the data itself, the brain regions studied, and the analysis software and methods used. Above all, our results emphasize the crucial importance of methodological choices in PET analyses for determining reproducibility.

Conclusion

In this whitepaper, we've showcased the QMENTA Imaging Hub as an indispensable device for navigating the intricacies of medical imaging research. It offers an uncomplicated option to integrate image processing tools of your own, and store and analyze data in one centralized place. Furthermore, by applying this solution to PET imaging analysis, we demonstrated the varying degrees of correspondence between the results of three open-source software tools. It can thus be extrapolated that our platform has all the features necessary to develop your own analysis tool, perform a validation study, or similar, all while fostering reproducible research practices.

The QMENTA Imaging Hub can support your medical imaging research by increasing efficiency and scientific throughput while cutting down on costs.

QMENTA®

QMENTA®

IMAGING TO INSIGHT

LEARN MORE

US Headquarters

75 State Street, Suite 100
Boston, MA 02109
+1 339 368 8040

EU Offices

C/ Roger de Llúria 46, Pral. 1ª
08009 Barcelona, Spain
+34 933 282 007

qmenta_inc

@QMENTA_Inc

www.qmenta.com