Published in final edited form as:
Neuroimage. 2014 February 15; 87: 89–95. doi:10.1016/j.neuroimage.2013.10.068.

Retest imaging of [^{11}C]NOP-1A binding to nociceptin/orphanin FQ peptide (NOP) receptors in brain of healthy humans

Talakad G. Lohith[a,*], **Sami S. Zoghbi**[a], **Cheryl L. Morse**[a], **Maria D. Ferraris Araneta**[a], **Vanessa N. Barth**[b], **Nancy A. Goebl**[b], **Johannes T. Tauscher**[b], **Victor W. Pike**[a], **Robert B. Innis**[a], and **Masahiro Fujita**[a]

[a] Molecular Imaging Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA

[b] Eli Lilly & Co. Indianapolis, IN, USA

Abstract

[^{11}C]NOP-1A is a novel high-affinity PET ligand for imaging nociceptin/orphanin FQ peptide (NOP) receptors. Here, we report reproducibility and reliability measures of binding parameter estimates for [^{11}C]NOP-1A binding in brain of healthy humans.

After intravenous injection of [^{11}C]NOP-1A, PET scans were conducted twice on eleven healthy volunteers on the same (10/11 subjects) or different (1/11 subjects) days. Subjects underwent serial sampling of radial arterial blood to measure parent radioligand concentrations. Distribution volume (V_T; a measure of receptor density) was determined by compartmental (one- and two-tissue) modeling in large regions and by simpler regression methods (graphical Logan and bilinear MA1) in both large regions and voxel data. Retest variability and intraclass correlation coefficient (ICC) of V_T were determined as measures of reproducibility and reliability, respectively.

Regional [^{11}C]NOP-1A uptake in brain was high, with a peak radioactivity concentration of 4 – 7 SUV (standardized uptake value) and a rank order of putamen > cingulate cortex > cerebellum. Brain time-activity curves fitted well in 10 of 11 subjects by unconstrained two-tissue compartmental model. The retest variability of V_T was moderately good across brain regions except cerebellum, and was similar across different modeling methods, averaging 12% for large regions and 14% for voxel-based methods. The retest reliability of V_T was also moderately good in most brain regions, except thalamus and cerebellum, and was similar across different modeling methods averaging 0.46 for large regions and 0.48 for voxels having gray matter probability > 20%. The lowest retest variability and highest retest reliability of V_T was achieved by compartmental modeling for large regions, and by the parametric Logan method for voxel-based methods.

Moderately good reproducibility and reliability measures of V_T for [^{11}C]NOP-1A make it a useful PET ligand for comparing NOP receptor binding between different subject groups or under different conditions in the same subject.

Corresponding author (post-publication): Robert B. Innis, MD, PhD Molecular Imaging Branch National Institute of Mental Health 10 Center Drive, MSC-1026 Bldg. 10, Rm. B1D43 Bethesda, MD 20892-1026 Tel: 301-594-1368 robert.innis@nih.gov. First author and corresponding author (pre-publication): Talakad G. Lohith, MD, PhD Molecular Imaging Branch National Institute of Mental Health 10 Center Drive, MSC-1026 Bldg. 10, Rm. B1D43K5 Bethesda, MD 20892-1026 lohithhere@gmail.com.
[*]This work was performed during Dr. Lohith's postdoctoral fellowship at NIH.

Keywords

NOP receptors; nociceptin; test-retest imaging; PET; retest variability; intraclass correlation coefficient

1. Introduction

Positron emission tomography (PET) is used to measure binding site occupancy of medications and differences in the receptor density between groups by comparing the measurements within or between subjects. The sensitivity and specificity of the PET studies are influenced by variation in quantification. In this regard, the test-retest imaging, where in the same subject undergoes two identical scans, is useful to assess both within subject variations as reproducibility or between subject variations as reliability of the outcome measures (Laruelle, 1999).

Our laboratory recently developed carbon-11-labeled NOP-1A ([^{11}C]NOP-1A) as a promising PET radioligand for *in vivo* imaging of nociceptin/orphanin FQ peptide (NOP) receptors (Pike et al., 2011). [^{11}C]NOP-1A has high affinity, binds selectively to the NOP receptor as an antagonist, and has appropriate lipophilicity (log D = 3.41) for blood-brain barrier permeability. After [^{11}C]NOP-1A injection in monkeys, about 60% of brain radioactivity reflects specific (i.e., displaceable) binding to NOP receptors (Kimura et al., 2011). We used [^{11}C]NOP-1A to visualize NOP receptors in human brain for the first time, and quantified them as total distribution volume (V_T), which is proportional to receptor density (Lohith et al., 2012). V_T values were measured both in large brain regions by compartmental modeling and in individual voxels by simpler regression analyses. In addition, V_T values were well identified across brain regions and stable over time, which is consistent with radiometabolites not entering brain.

However, by doing a single scan in each subject (Lohith et al., 2012), the precision of measuring binding can only be estimated mathematically based on standard errors (i.e. identifiability) to measure V_T. In the current study, we sought to determine the reproducibility and reliability of V_T by scanning each subject twice, i.e. test and retest scans. Reproducibility was measured as retest variability, and reliability was measured as intraclass correlation coefficient (ICC) (Laruelle, 1999). Retest variability and reliability were studied not only in large brain regions but also at the voxel level, because such parametric images are useful for localizing brain regions with altered binding in patient and control groups. Because voxel-wise analyses are prone to underestimate V_T, we compared two parametric methods (i.e., graphical Logan and MA1) with different sensitivities to underestimation. Furthermore, based on the results obtained we sought to determine the necessary sample size for alterations in a prospective between-subject receptor density studies.

2. Materials and Methods

2.1 Radioligand preparation

[^{11}C]NOP-1A was labeled by [^{11}C] methylation of an *N*-desmethyl precursor, as previously described (Pike et al., 2011). The radioligand was prepared according to our Investigational New Drug Application (114,313), which was submitted to the U.S. Food and Drug Administration; a copy is available at http://pdsp.med.unc.edu/snidd/IND/nop1a.html. The radioligand was obtained with high radiochemical purity (> 99%) and a specific activity of 128 ± 34 GBq/μmol at the time of injection (n = 22 batches).

2.2 Subjects

Eleven healthy volunteers (8 males, 3 females) participated in the brain PET scans (mean age = 29 years (range: 22 – 42 years); mean weight = 74 kg (range: 59 – 99 kg)). All subjects were free of current medical or psychiatric illnesses, as determined by medical history, physical examination, electrocardiogram, urinalysis including drug screening, and laboratory blood tests (complete blood count, serum chemistries, and thyroid function test). Subjects' vital signs were recorded before [^{11}C]NOP-1A injection and at 15, 30, 90, and 120 minutes after injection. Repeat urinalysis and blood tests were conducted within two hours of PET scan completion. The protocol was approved by the Institutional Review Board of the National Institutes of Health. All subjects signed a written informed consent form.

2.3 PET scans and measurement of [^{11}C]NOP-1A in arterial plasma

All PET scans were performed on an Advance tomograph (GE Medical Systems, Waukesha, WI). Each subject underwent test and retest scans after bolus injection of [^{11}C]NOP-1A along with arterial blood sampling for metabolite corrected input function. Test and retest scans were performed on the same day separated by three hours between radiotracer injections except for one subject whose scans occurred 10 days apart. After an 8-minute brain transmission scan using ^{68}Ge rod source, dynamic three-dimensional emission scans were acquired for 120 minutes as previously described (Lohith et al., 2012). Arterial blood samples were drawn manually after radioligand injection with 1.5 mL samples at 15 s intervals until 150 s, followed by 3 mL samples at 3, 4, 6, 8, 10, 15, 20, 30, 40, and 50 min, and 5 mL samples at 60, 75, 90, and 120 min. The concentration of parent radioligand and the metabolite-corrected plasma input function were obtained as previously described (Lohith et al., 2012; Pike et al., 2011). The plasma free fraction (f_p) was measured for each scan by ultrafiltration, as previously described (Gandelman et al., 1994). Assay-to-assay variation in f_P measurement was corrected based on f_P measured from a standard plasma sample along with the subject's sample (Abi-Dargham et al., 1999). Radiochemical purity was measured by incubating [^{11}C]NOP-1A from all 22 syntheses in whole blood and plasma for 30 minutes at room temperature.

PET images were analyzed by applying a template of 10 pre-set volumes-of-interest in Montreal Neurologic Institute space after coregistration to same subject's magnetic resonance (MR) image using Statistical Parametric Mapping, SPM (Version 8 for Windows, Wellcome Department of Cognitive Neurology, UK) as previously described (Lohith et al., 2012). As a measure of receptor binding, V_T was calculated using brain and arterial input function from each of the test and retest scans by compartmental modeling and graphical (Logan_{VOI} and bilinear MA1_{VOI}) analyses on large regions as well as by parametric methods (Loganvoxel and bilinear MA1voxel) on voxel-wise data as described before (Lohith et al., 2012). V_T/f_P was also calculated as another measure of receptor binding, because only free ligand enters the brain. Kinetic analyses and generation of parametric images were performed using pixelwise modeling software (PMOD 3.16, PMOD Technologies Ltd, http://www.pmod.com/).

2.4 Measurement of retest variability and reliability of radioligand binding

The optimal compartment model (i.e., one- vs two-tissue compartments) to determine V_T was chosen based on Akaike information criterion, model selection criterion (proposed by Micromath, Saint Louis, Missouri USA, http://www.micromath.com/products.php?p=scientist&m=statistical_analysis), and *F*-tests (Hawkins et al., 1986). V_T values measured between models or between test and retest scans in the same subject were compared using factorial repeated measures analysis of variance (rmANOVA) with Bonferroni adjustment. The retest variability of V_T was calculated as the absolute difference between test and retest V_T divided by the average between the two, expressed as a percentage. Retest variability

under 10% was considered excellent; and over 10% but under 20% as moderate. The retest reliability of V_T was the intraclass correlation coefficient (ICC) calculated as follows:

$$ICC = \frac{BSMSS - WSMSS}{BSMSS + (n-1)\ WSMSS}$$

where BSMSS and WSMSS are between- and within-subject mean sum of squares, respectively, and *n* is the number of within-subject observations (in this case, *n* = 2). ICC values between 0 and 1 indicated higher variability between subjects than within subjects; values close to 1 suggested good reliability. Values between −1 and 0 indicated that variability was higher within subjects than between subjects and suggested poor reliability (Landis and Koch, 1977; Shrout and Fleiss, 1979). Retest variability and ICC values were also calculated at the voxel level by SPM8 using voxelwise parametric images of V_T, yielding 3-dimensional spatial maps of reproducibility and reliability.

2.5 Sample size estimation

Because of variability in PET imaging results, it is necessary to estimate the sample size required to detect a significant effect for between-subject studies (e.g., differences in NOP density between patients and controls). The test scan results from the 11 healthy subjects in this study were pooled with single scans previously obtained by our group in 7 healthy subjects (Lohith et al., 2012). The sample size was calculated for a 10, 15 and 20% change in V_T assuming that similar variance exists between healthy subjects and patients. A 10, 15 or 20% change for the between-subject studies was chosen based on the premise that a change of at least 10 – 20% in outcome parameter is necessary and sufficient for establishing group differences by PET studies (Deschwanden et al., 2011; Hirvonen et al., 2012).

2.6 Statistical Analysis

All statistical analyses were performed with SPSS (Version 17 for Windows, SPSS Inc. Chicago, IL). Test-retest study parameters were compared with paired *t*-tests after testing for normality with the Shapiro-Wilk's test. $P < 0.05$ was considered statistically significant. Values represent mean ± standard deviation (SD). For parameters that were not normally distributed (injected radioactivity and plasma f_P), non-parametric Mann-Whitney *U* tests were used.

3. Results

3.1 Pharmacologic Effects

The injected radioactivity (*n* = 22 injections in 11 subjects) of [^{11}C]NOP-1A was 691 ± 126 MBq (range, 228 – 760 MBq). The injected mass dose was 81 ± 32 pmol/kg (range, 25 – 155 pmol/kg). The injected radioactivity and mass dose did not differ statistically between the test and retest scans (Table 1). There were no adverse or clinically detectable pharmacologic effects in any subject during test or retest scans. No significant changes were observed in vital signs or electrocardiograms or the results of laboratory studies.

3.2 Plasma Analysis

The arterial plasma concentration of parent radioligand peaked at ~10 SUV within one minute of injection of [^{11}C]NOP-1A, followed by a rapid decline and slow terminal clearance (Figure 1A). The area-under-the-curve (AUC) of parent radioligand in plasma did not show statistically significant difference between the two scans (Table 1). In all subjects,

and similar to our previous study (Lohith et al., 2012), the radioactivity curves in whole blood and total plasma curves were well fit with a triexponential function, and the curve of parent fraction (i.e., percentage of parent radioligand in total plasma radioactivity) was well fit with a Hill function. The average plasma clearance was similar for test and retest scans, with no statistically significant differences (Table 1). Although plasma clearance within each subject showed high retest variability (21 ± 18%), the high ICC (+0.89) indicated good reliability—i.e., higher variability between subjects than within subjects. Therefore, clearance measurements effectively detected differences between subjects. Of the three associated half-lives, the first two (1 ± 2 min and 31 ± 74 min, $n = 22$) largely reflected distribution, and the last (125 ± 100 min) reflected clearance (i.e. metabolism and elimination). The long value of this terminal half-life accounted for ~62% of the total AUC integrated to infinity.

The average plasma free fraction (f_P) was similar for test and retest scans, with no statistically significant difference (Table 1). However, f_P showed high retest variability (21 ± 11%), indicating poor reproducibility, and low ICC (− 0.17), indicating poor reliability—i.e., higher variability within subjects than between subjects.

The radiochemical purity of [^{11}C]NOP-1A in whole blood and plasma from both test and retest scans was 99 ± 3% and 99 ± 2% (mean ± SD, $n = 22$), respectively, after 30 minutes of incubation.

3.3 Brain Radioactivity and Kinetics Analysis

After intravenous injection of [^{11}C]NOP-1A, all subjects showed high concentrations of radioactivity in brain (4–7 SUV) followed by quick washout. Radioactivity distribution was widespread in different brain regions known to express NOP receptors with a rank order of putamen > cingulate cortex > cerebellum, consistent with previously results (Lohith et al., 2012). The putamen AUC for test and retest scans was 359 ± 92 SUV · min and 329 ± 73 SUV · min, respectively, with no statistically significant difference between the two scans (Figure 1B).

The brain time-activity curves fitted well in 10 of 11 subjects by unconstrained two-tissue compartment model based on F test, Akaike information criterion, and model selection criterion scores, consistent with the presence of distinct specific and non-specific compartments in brain (Figure 2). In one subject, the two-tissue compartment model did not converge for both test and retest scans in 4 and 2 of the 10 regions, respectively; hence, V_T from the one-tissue compartment model was chosen for this subject. Regional values of two-tissue V_T (mL · cm^{-3}) ranged from 13.1 in parietal cortex to 4.6 in cerebellum. The average value of V_T in the retest scan was slightly higher than the average test V_T, with differences of less than 5% for the compartmental model, and less than 10% for the non-compartmental models (Figure 3, Table 2). However, no statistically significant difference was noted in compartmental or non-compartmental V_T values between test and retest scans in any region (Table 2). Both region- and voxel-based Logan and MA1 methods significantly (all $P < 0.05$) underestimated V_T values compared to compartmental modeling in all subjects (rmANOVA $F_{1.27,12.72} = 45.96$), similar to previously published results (Table 2) (Lohith et al., 2012). The mean percent difference in V_T was highest (12.6%) between compartmental and Logan$_{voxel}$ methods. In addition, V_T estimated by bilinear MA1$_{voxel}$ was significantly higher (6.4%, $P < 0.001$) than that estimated by Logan$_{voxel}$ model (Figure 4).

3.4 Retest variability and reliability of radioligand binding

Retest variability was moderately good, and similar for both region-based (averaging 12%) and voxel-based (averaging 14%) methods (Table 3). Retest variability was consistent

across models and across regions (except cerebellum), with the lowest variability observed in occipital cortex by compartmental model and the highest variability observed in cerebellum by voxel-wise MA1 method. However, the coefficient of variance (SD/mean) for retest variability was higher for voxel-wise compared to large regions (35 and 25% increased mean COV for $Logan_{voxel}$ and $MA1_{voxel}$, respectively) suggesting that parametric methods are vulnerable to noise from the voxels. Parametric maps of retest variability from both $Logan_{voxel}$ (Figure 5) and $MA1_{voxel}$ (not shown) methods indicated similar retest variability across voxels in gray matter of cerebral cortices. The average variability in voxels with gray matter probability > 20% in the entire brain was 13% and 14% for $Logan_{voxel}$ and $MA1_{voxel}$, respectively. Moreover, 44% of voxels with gray matter probability of >20% had retest variability of <12% by $Logan_{voxel}$ method.

Retest reliability (measured as ICC) was moderately good and similar for both region-based (average ICC of 0.46) and voxel-based (average ICC of 0.48) methods (Table 4). ICC values were higher in larger brain areas (e.g., cortical regions) than in smaller brain areas (e.g., thalamus) or low NOP density regions (e.g., cerebellum). Among the region-based methods, compartmental model showed inferior reliability in thalamus than in other regions and other models. Parametric maps of retest reliability from both $Logan_{voxel}$ (Figure 5) and $MA1_{voxel}$ methods indicated similar ICC values across voxels in gray matter of cerebral cortices. The average ICC in voxels with gray matter probability > 20% in the entire brain was 0.52 and 0.49 for $Logan_{voxel}$ and $MA1_{voxel}$, respectively. Moreover, 59% of voxels with gray matter probability of >20% had ICC values of >0.5 by $Logan_{voxel}$ method.

Because only free radioligand in plasma can cross the blood-brain barrier, V_T should be corrected for (i.e. divided by) plasma free fraction f_P to more accurately measure NOP receptor availability. However, V_T/f_P showed a high retest variability of 22% and poor reliability of ICC = −0.27. We interpret these results to mean that V_T/f_P is vulnerable to the measurement noise in the added variable f_P, although it is theoretically more accurate than V_T alone.

3.5 Sample size required for between-subject studies

The combined group of 18 subjects (11 subjects from this study plus 7 subjects from our previous study (Lohith et al., 2012)) had V_T values in ten regions of 8.8 ± 1.5 mL · cm^{-3} (mean ± SD). A 20% change in V_T would correspond to 1.8 mL · cm^{-3}. Assuming the variance is similar in patients and healthy subjects, a sample of at least 45, 20 and 12 subjects per group would be necessary to detect a 10, 15 and 20% difference, respectively, in mean V_T for between-subject studies with 80% power and a P value of 0.05.

4. Discussion

The present study sought to determine the reproducibility and reliability of measuring NOP binding by comparing V_T values from 2 [^{11}C]NOP-1A PET scans in the same subject. Reproducibility (measured as retest variability) and reliability (measured as ICC) of V_T were moderately good, but not excellent. We confirmed these measures both at regional-level using compartmental modeling and at voxel-level using simpler regression methods such as Logan and bilinear MA1. The Logan voxel-wise method tends to underestimate V_T because of the relatively high noise in the time-activity curves of individual voxels (Slifstein and Laruelle, 2000). Because of this underestimation (i.e., bias), other methods, including MA1, have been developed to be more robust to this noise (Ichise et al., 2002). The current study found that, compared to MA1, voxel-based Logan method showed greater underestimation of the `gold standard' V_T measured with compartmental modeling. Despite its lower accuracy at the voxel level, the Logan method may be more sensitive than MA1 method for detecting small changes between groups, because it has slightly smaller variability and better

reliability than MA1 method. Furthermore, a power analysis based on these results indicated that for measuring precise changes, a sample size of at least 12 subjects was required in receptor density studies.

After [^{11}C]NOP-1A injection, the distribution of radioactivity in brain was similar to that observed in our previous study (Lohith et al., 2012), with cortical and subcortical structures such as putamen, caudate and thalamus having higher uptakes, and cerebellum having the lowest uptake. Brain kinetics were better described by a two-tissue (rather than a one-tissue) compartmental model in all but one subject, consistent with the kinetic identifiability of both a fast (non-specific) and a slow (specific) compartment. In addition, V_T values estimated by region-based compartmental modeling were highest compared to region- or voxel-based Logan and bilinear MA1 methods, which underestimated V_T to a variable extent in test and retest scans. These underestimates could be due to differences in data noise during fitting (Ichise et al., 2002; Slifstein and Laruelle, 2001). Although absolute V_T values were underestimated by Logan and MA1 parametric modeling, these methods showed retest variability and reliability comparable to those obtained via the gold standard compartmental modeling in most brain areas (Tables 3, 4 and Figure 5).

Retest variability and reliability of V_T in the current study were moderately good and similar both by region-wise (average variability of 12% and ICC of 0.46) and voxel-wise (average variability of 14% and ICC of 0.48) analysis methods. Compared to some radioligands such as ^{11}C-carfentanil (Hirvonen et al., 2009) and [^{11}C]DASB (Kim et al., 2006) that showed low retest variability (< 10%) and a high ICC (> 0.8) for most regions, a large number of radioligands such as [^{11}C] (R)-rolipram (Zanotti-Fregonara et al., 2011), [^{11}C] (R)-PK11195 (Jucaite et al., 2012), and [^{11}C]ABP688 (DeLorenzo et al., 2011) showed a moderate to high retest variability of > 10% and moderate to low ICC of < 0.7 for most regions. At least 3 factors may have caused [^{11}C]NOP-1A to have only moderate reproducibility and reliability.

First, measurement errors in brain or plasma activity could account for the variability because V_T is, in theory, the AUC of brain divided by that of [^{11}C]NOP-1A in arterial plasma from time zero to infinity (Terry et al., 2009). Although uptake in the brain (for instance AUC for putamen) did not differ significantly between test and retest scans, retest and inter-subject variability (SD/mean across subjects) were moderately high ~13 and 26%, respectively. Similarly, although no significant difference between test and retest scans was observed in the AUC for plasma, retest and inter-subject variability were moderately high (~21 and 25%, respectively). Because we drew all blood samples manually, measurements at early time-points may include small errors caused by the manual drawings over a few seconds instead of instantaneous sampling and dispersion in the artery. However, error contributions from such measurements are negligible for plasma AUC and calculation of V_T. Accurate measurement of plasma parent may be difficult at later time-points because of low levels (~20%) of fraction of parent in plasma and low levels of total plasma activity. The short half-life of C-11 (20.4 minutes) may limit accurate measurement of parent levels in plasma, especially at late time-points. Labeling with longer-lived ^{18}F may be more useful in reducing such measurement errors, because radioactive counts will be high and can be reliably measured even at late time-points. In fact in a recent study of imaging NOP receptors by a new [^{18}F]-labeled PET ligand ^{18}F-MK-0911 (Hostetler et al., 2013), the retest variability of V_T averaged 3% across brain regions, although this result was obtained from only a small sample size of 3 subjects and also from plasma parent fraction measured at only 6 time points. The authors note that such a low retest variability of their radiotracer could be due to obtaining higher and plasma counts with low noise levels at later time points because of [^{18}F] label. Thus, the major source of high retest variability in V_T of [^{11}C]NOP-1A could be from plasma rather than brain data. These measurement errors may together contribute to

higher retest variability in V_T although, individually, neither makes a markedly disproportionate contribution.

Second, physiological changes in both receptor density and affinity—rather than measurement errors per se—can also cause variability. For instance, stress from the scanning procedure may cause subtle changes in NOP receptor levels. A recent study found that acute and repeated restraint stress in rats reduced NOP mRNA levels by ~15% in the mediodorsal forebrain and hypothalamus ~4 hrs after restraint (Delaney et al., 2012). Being an antagonist, [^{11}C]NOP-1A binding is supposedly insensitive to the affinity states of the NOP receptor; and thus the binding getting affected by the endogenous NOP agonist is a remote possibility, but it cannot be ruled out. Any significant diurnal changes in endogenous ligand levels between the test and retest scans (particularly because all but one test-retest scan were done on same day) could affect the specific binding of [^{11}C]NOP-1A and consequently cause higher retest variability. A good measure of the physiological changes in receptor availability caused by diurnal variation could be V_T/f_P. The retest variability of V_T/f_P was 23% for the 10 subjects who underwent morning and afternoon scans, although values of V_T/f_P did not significantly differ between the two scans in any brain region (Wilcoxon signed rank $Z = -1.274$ to -1.682, $P = 0.093$ to 0.203). This variation should be interpreted carefully as f_P measurements could contain small errors. However, intersubject variability (SD/mean) of V_T/f_P for morning and afternoon scans in 10 subjects was similar in most brain regions (15% and 17%, respectively) indirectly indicating consistent f_P measurements across the subjects. For these reasons, we are uncertain whether actual receptor availability could be measured as V_T/f_P, and we do not recommend its use for future studies with [^{11}C]NOP-1A unless solid data exist to support differences in f_P between comparison groups.

Third, both the reproducibility and reliability could be affected by region size and regional receptor density. Small regions have lower counting statistics and are more prone to head movement or coregistration errors (Parsey et al., 2000). However, retest variability was marginally higher (~ 14%) for small regions such as caudate and thalamus, and ICC was moderately good (> 0.5) for caudate, whereas the thalamus had relatively poor ICC (0.3) for unknown reasons. Regions with low receptor density are statistically noisy, which accounts for relatively poor reproducibility measures for cerebellum, a region with low NOP density, but the poor reproducibility was seen only for voxel but not clearly for VOI data (Tables 3 and 4).

Fourth, the amount of non-radioactive ligand between test and retest injections could affect the test-retest reproducibility, although it is least likely, because the absolute mass dose in PET studies is so low as a result of high specific activity of the radioligand. As the mass dose in this study had a high variability of 44% (although specific activity was high) on an average (Table 1), we examined the correlation between mass dose and V_T values in all regions among all subjects for both test and retest scans (data not shown) and found no significant correlation between mass dose and V_T values (Average Pearson's r = 0.14). Therefore, the variability in V_T was not affected by mass dose of NOP-1A.

The current study used arterial sampling, an invasive technique but provides accurate measurement of input function. To reduce invasiveness in clinical studies, image-derived input function (IDIF) and population-based input function (PBIF) have been developed and used in retest studies (Zanotti-Fregonara et al., 2012; Zanotti-Fregonara et al., 2011). An image-derived input function could not be reliably obtained because the carotid artery is not well visible on [^{11}C]NOP-1A brain images, even using the early summed frames. By contrast, using PBIF, which does not require indentifying the carotid artery on images, the retest variability of V_T significantly increased compared to that obtained with blood

sampling (12 vs. 19%; data not shown). This larger variability of PBIF could still be justified if PBIF requires markedly reduced invasiveness of the procedure. However, PBIF does require arterial sampling because the PBIFs were individually scaled with two arterial blood samples and [^{11}C]NOP-1A concentrations differed between artery and vein. Therefore we think that there is no acceptable alternative to full arterial sampling for [^{11}C]NOP-1A studies.

Despite the moderate retest variability and reliability observed in large brain regions, at the voxel-level, 44 and 59% of gray matter voxels showed a retest variability of < 12% and ICC of > 0.5, respectively by parametric Logan analysis. This suggests that voxel-based analyses of [^{11}C]NOP-1A studies could be sensitive enough to measure NOP receptor binding with good precision. In addition, a sample size of at least 12 subjects needed to detect a group-level difference of 20% in V_T could be reasonable for PET studies that compare binding measurements between patient and control groups. These initial estimates of reproducibility, reliability, and power justify the use of [^{11}C]NOP-1A as a promising PET radioligand in further studies of NOP receptor binding between different subject groups.

5. Conclusion

The present study confirms that [^{11}C]NOP-1A has high brain uptake and regional distribution in human brain consistent with that of NOP receptors. [^{11}C]NOP-1A can quantify NOP receptor binding in terms of distribution volume both at the regional level using compartmental model and at the voxel level using simpler regression analyses. The absolute value, reproducibility, and reliability of V_T were comparable among different models, with compartmental model providing the most reproducible and reliable V_T values among VOI-based methods and Logan analysis providing slightly better values between voxel-wise methods. In conclusion, [^{11}C]NOP-1A is a promising PET radioligand for comparing NOP receptors between different subject groups.

Acknowledgments

This study was supported by the Intramural Research Program of the National Institute of Mental Health, National Institutes of Health (IRP-NIMH-NIH) and by a Cooperative Research and Development Agreement with Eli Lilly & Co. We thank Denise Rallis-Frutos, Gerald L. Hodges, Kimberly Jenko, David Clark, Jeih-San Liow, Robert L. Gladding, and the staff of the PET Department for assistance completing the studies, and PMOD Technologies (Zurich, Switzerland) for providing its image analysis and modeling software. Francois Vandenhende (Eli Lilly & Co.) provided valuable statistical guidance. Ioline Henter (NIMH) provided excellent editorial assistance.

Disclosures: This study was supported by the Intramural Research Program of the National Institute of Mental Health, National Institutes of Health (IRP-NIMH-NIH) and by a Cooperative Research and Development Agreement with Eli Lilly & Co.

Abbreviations

AUC	area-under-the-curve
BSMSS	between-subject mean sum of squares
f_P	plasma free fraction
ICC	intraclass correlation coefficient
NOP	nociceptin/orphanin FQ peptide
PET	positron emission tomography
WSMSS	within-subject mean sum of squares

V_T	total distribution volume

References

Abi-Dargham A, Laruelle M, Krystal J, D'Souza C, Zoghbi S, Baldwin RM, Seibyl J, Mawlawi O, de Erasquin G, Charney D, Innis RB. No evidence of altered in vivo benzodiazepine receptor binding in schizophrenia. Neuropsychopharmacology. 1999; 20:650–661. [PubMed: 10327433]

Delaney G, Dawe KL, Hogan R, Hunjan T, Roper J, Hazell G, Lolait SJ, Fulford AJ. Role of nociceptin/orphanin FQ and NOP receptors in the response to acute and repeated restraint stress in rats. J Neuroendocrinol. 2012; 24:1527–1541. [PubMed: 22835008]

DeLorenzo C, Kumar JS, Mann JJ, Parsey RV. In vivo variation in metabotropic glutamate receptor subtype 5 binding using positron emission tomography and [11C]ABP688. J Cereb Blood Flow Metab. 2011; 31:2169–2180. [PubMed: 21792244]

Deschwanden A, Karolewicz B, Feyissa AM, Treyer V, Ametamey SM, Johayem A, Burger C, Auberson YP, Sovago J, Stockmeier CA, Buck A, Hasler G. Reduced metabotropic glutamate receptor 5 density in major depression determined by [(11)C]ABP688 PET and postmortem study. Am J Psychiatry. 2011; 168:727–734. [PubMed: 21498461]

Gandelman MS, Baldwin RM, Zoghbi SS, Zea-Ponce Y, Innis RB. Evaluation of ultrafiltration for the free-fraction determination of single photon emission computed tomography (SPECT) radiotracers: beta-CIT, IBF, and iomazenil. J Pharm Sci. 1994; 83:1014–1019. [PubMed: 7965658]

Hawkins RA, Phelps ME, Huang SC. Effects of temporal sampling, glucose metabolic rates, and disruptions of the blood-brain barrier on the FDG model with and without a vascular compartment: studies in human brain tumors with PET. J Cereb Blood Flow Metab. 1986; 6:170–183. [PubMed: 3485641]

Hirvonen J, Aalto S, Hagelberg N, Maksimow A, Ingman K, Oikonen V, Virkkala J, Nagren K, Scheinin H. Measurement of central mu-opioid receptor binding in vivo with PET and [11C]carfentanil: a test-retest study in healthy subjects. Eur J Nucl Med Mol Imaging. 2009; 36:275–286. [PubMed: 18779961]

Hirvonen J, Goodwin RS, Li CT, Terry GE, Zoghbi SS, Morse C, Pike VW, Volkow ND, Huestis MA, Innis RB. Reversible and regionally selective downregulation of brain cannabinoid CB1 receptors in chronic daily cannabis smokers. Mol Psychiatry. 2012; 17:642–649. [PubMed: 21747398]

Hostetler ED, Sanabria-Bohorquez S, Eng W, Joshi AD, Patel S, Gibson RE, O'Malley S, Krause SM, Ryan C, Riffel K, Bi S, Okamoto O, Kawamoto H, Ozaki S, Ohta H, de Groot T, Bormans G, Depre M, de Hoon J, De Lepeleire I, Reynders T, Cook JJ, Burns HD, Egan M, Cho W, van Laere K, Hargreaves RJ. Evaluation of [(1)(8)F]MK-0911, a positron emission tomography (PET) tracer for opioid receptor-like 1 (ORL1), in rhesus monkey and human. Neuroimage. 2013; 68:1–10. [PubMed: 23238431]

Ichise M, Toyama H, Innis RB, Carson RE. Strategies to improve neuroreceptor parameter estimation by linear regression analysis. J Cereb Blood Flow Metab. 2002; 22:1271–1281. [PubMed: 12368666]

Jucaite A, Cselenyi Z, Arvidsson A, Ahlberg G, Julin P, Varnas K, Stenkrona P, Andersson J, Halldin C, Farde L. Kinetic analysis and test-retest variability of the radioligand [11C](R)-PK11195 binding to TSPO in the human brain - a PET study in control subjects. EJNMMI Res. 2012; 2:15. [PubMed: 22524272]

Kim JS, Ichise M, Sangare J, Innis RB. PET imaging of serotonin transporters with [11C]DASB: test-retest reproducibility using a multilinear reference tissue parametric imaging method. J Nucl Med. 2006; 47:208–214. [PubMed: 16455625]

Kimura Y, Fujita M, Hong J, Lohith TG, Gladding RL, Zoghbi SS, Tauscher JA, Goebl N, Rash KS, Chen Z, Pedregal C, Barth VN, Pike VW, Innis RB. Brain and whole-body imaging in rhesus monkeys of 11C-NOP-1A, a promising PET radioligand for nociceptin/orphanin FQ peptide receptors. J Nucl Med. 2011; 52:1638–1645. [PubMed: 21880575]

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159–174. [PubMed: 843571]

Laruelle M. Modelling: when and why? Eur J Nucl Med. 1999; 26:571–572. [PubMed: 10369941]

Lohith TG, Zoghbi SS, Morse CL, Araneta MF, Barth VN, Goebl NA, Tauscher JT, Pike VW, Innis RB, Fujita M. Brain and whole-body imaging of nociceptin/orphanin FQ peptide receptor in humans using the PET ligand 11C-NOP-1A. J Nucl Med. 2012; 53:385–392. [PubMed: 22312136]

Parsey RV, Slifstein M, Hwang DR, Abi-Dargham A, Simpson N, Mawlawi O, Guo NN, Van Heertum R, Mann JJ, Laruelle M. Validation and reproducibility of measurement of 5-HT1A receptor parameters with [carbonyl-11C]WAY-100635 in humans: comparison of arterial and reference tisssue input functions. J Cereb Blood Flow Metab. 2000; 20:1111–1133. [PubMed: 10908045]

Pike VW, Rash KS, Chen Z, Pedregal C, Statnick MA, Kimura Y, Hong J, Zoghbi SS, Fujita M, Toledo MA, Diaz N, Gackenheimer SL, Tauscher JT, Barth VN, Innis RB. Synthesis and evaluation of radioligands for imaging brain nociceptin/orphanin FQ peptide (NOP) receptors with positron emission tomography. J Med Chem. 2011; 54:2687–2700. [PubMed: 21438532]

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979; 86:420–428. [PubMed: 18839484]

Slifstein M, Laruelle M. Effects of statistical noise on graphic analysis of PET neuroreceptor studies. J Nucl Med. 2000; 41:2083–2088. [PubMed: 11138696]

Slifstein M, Laruelle M. Models and methods for derivation of in vivo neuroreceptor parameters with PET and SPECT reversible radiotracers. Nucl Med Biol. 2001; 28:595–608. [PubMed: 11516703]

Terry GE, Liow JS, Zoghbi SS, Hirvonen J, Farris AG, Lerner A, Tauscher JT, Schaus JM, Phebus L, Felder CC, Morse CL, Hong JS, Pike VW, Halldin C, Innis RB. Quantitation of cannabinoid CB1 receptors in healthy human brain using positron emission tomography and an inverse agonist radioligand. Neuroimage. 2009; 48:362–370. [PubMed: 19573609]

Zanotti-Fregonara P, Hines CS, Zoghbi SS, Liow JS, Zhang Y, Pike VW, Drevets WC, Mallinger AG, Zarate CA Jr. Fujita M, Innis RB. Population-based input function and image-derived input function for [(1)(1)C](R)-rolipram PET imaging: methodology, validation and application to the study of major depressive disorder. Neuroimage. 2012; 63:1532–1541. [PubMed: 22906792]

Zanotti-Fregonara P, Zoghbi SS, Liow JS, Luong E, Boellaard R, Gladding RL, Pike VW, Innis RB, Fujita M. Kinetic analysis in human brain of [11C](R)-rolipram, a positron emission tomographic radioligand to image phosphodiesterase 4: a retest study and use of an image-derived input function. Neuroimage. 2011; 54:1903–1909. [PubMed: 21034834]

Highlights

- We report reproducibility and reliability for [^{11}C]NOP-1A binding in human brain
- Reproducibility was moderately good across most brain regions and modeling methods
- Reliability was moderately good across most brain regions and modeling methods
- [^{11}C]NOP-1A is useful to compare NOP receptor binding within and between subjects

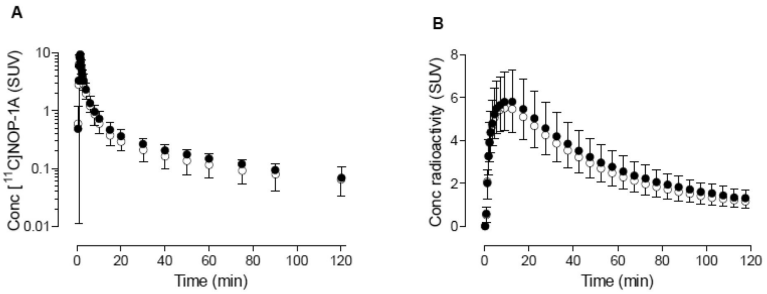

Figure 1.

Time course of parent radioligand in arterial plasma (A) and the radioactivity concentration in putamen (B) from test (●) and retest (o) scans after [^{11}C]NOP-1A injection. Data represent mean ± SD from 11 subjects.

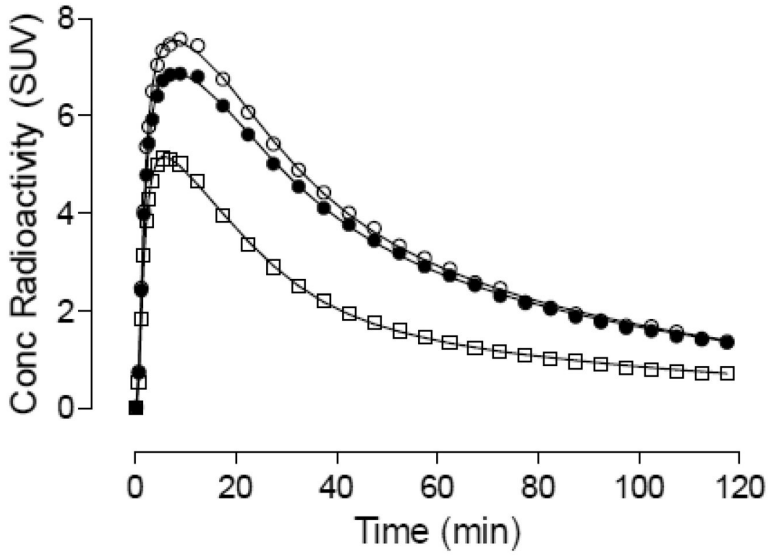

Figure 2.

Representative brain uptake with unconstrained two-tissue compartmental fitting from a 22-year-old healthy woman injected with 747 MBq of [^{11}C]NOP-1A. Concentration of radioactivity from 3 regions is shown: putamen (o), with highest uptake; cingulate cortex (●), with medium uptake; and cerebellum (□), with lowest uptake. Line represents unconstrained two-tissue compartmental fitting.

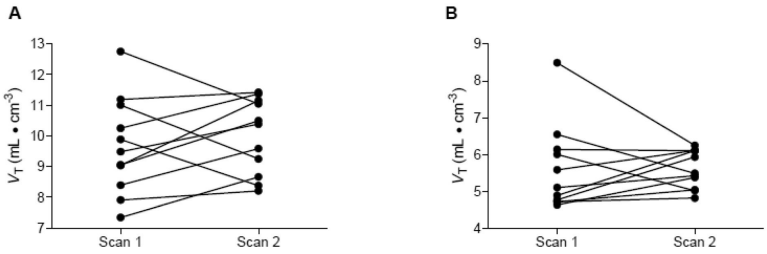

Figure 3.

Individual test and retest scan V_T values from compartmental modeling for putamen (A) and cerebellum (B), which have high and low density NOP receptor, respectively. Values from same subject are connected by a line.

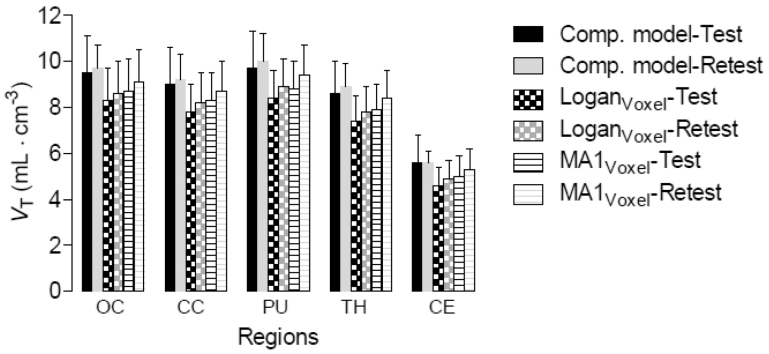

Figure 4.

Comparison of V_T in various brain regions from test (black) and retest (gray) scans. V_T values were calculated in 11 subjects injected with [^{11}C]NOP-1A using unconstrained two-tissue compartment model (except for 1 subject fitted by one-tissue compartment model) for large regions, and Logan and MA1 methods for voxel data (Logan$_{voxel}$ and MA1$_{voxel}$). Data represent mean ± SD from 11 subjects. No statistically significant difference in V_T was observed for any method between test and retest scans in any region ($P > 0.05$ by factorial repeated measures ANOVA). Both Logan$_{voxel}$ and MA1$_{voxel}$ showed significantly smaller V_T ($P < 0.05$). OC = occipital cortex; CC = cingulate cortex; PU = putamen; TH = thalamus; CE = cerebellum.

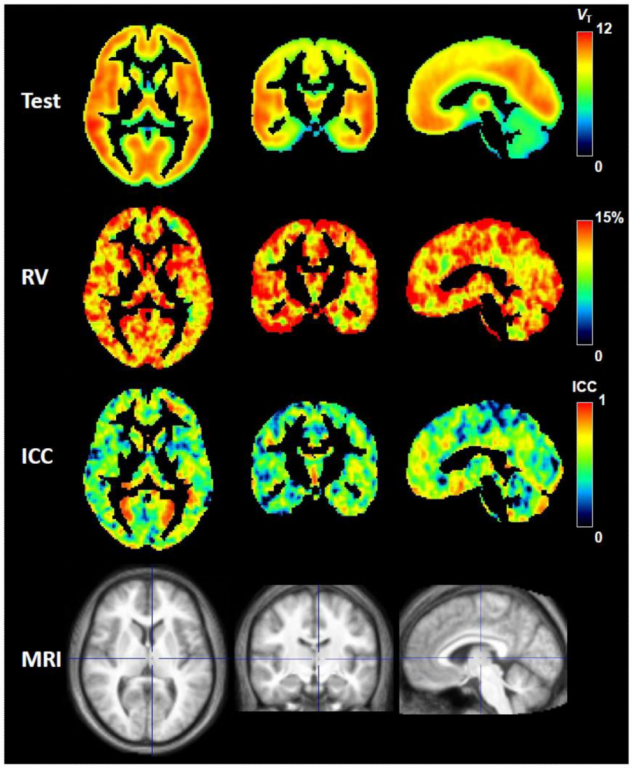

Figure 5.
Voxel-wise reproducibility measures of V_T from spatially normalized and gray matter segmented [^{11}C]NOP-1A PET images in three orthogonal planes. Gray matter probability of 20% was used as the threshold. Mean parametric Logan images from test scans (n = 11) where in each voxel represents V_T is shown in top row. Retest variability (RV) images where in each voxel represents % variance is shown in second row. Retest reliability (ICC) images where in each voxel represents ICC is shown in third row. Spatially normalized mean MR image in three orthogonal planes is shown in bottom row. Crosshairs on MR image indicate the slicing for three orthogonal planes common for all images.

Table 1

Comparison of different parameters from test-retest scans in healthy subjects (n = 11)

Parameters	Test	Retest	Retest Variability (%)	*P*-value
Injected activity (MBq)	713 ± 79	669 ± 161	18	0.699
Injected mass dose (pmol/kg)	80 ± 33	81 ± 32	44	0.933
Putamen $AUC_{0–120}$ (SUV · min)	359 ± 92	329 ± 73	13	0.160
Plasma [^{11}C]NOP-1A $AUC_{0–120}$ (SUV · min)	45 ± 10	39 ± 10	21	0.091
Plasma Clearance (L/min)	0.82 ± 0.38	0.89 ± 0.45	21	0.246
Plasma free fraction, f_p (%)	9.5 ± 0.7	8.8 ± 1.9	21	0.047

Test and retest values represent mean ± SD. Retest variability is the absolute difference between test and retest parameters divided by the average between the two, expressed as a percentage. *P*-values are from student *t*-tests for all parameters except for injected activity and plasma f_p, which were from non-parametric Mann-Whitney tests.

Abbreviations: AUC: area-under-the-curve; SUV: standardized uptake value.

Table 2

Comparison of V_T values by Compartmental model, Logan, and MA1 methods

Model	Test V_T (mL · cm^{-3})	Retest V_T (mL · cm^{-3})	Change (%)	*P* value (Repetition)	*P* value (Region×Repetition)	Model underestimation of V_T (%)	*P* value (Model underestimation)
Compartmental	8.7 ± 1.9	8.9 ± 1.6	+ 4.2	0.616	0.189	-	-
$Logan_{VOI}$	8.1 ± 1.5	8.5 ± 1.5	+ 6.4	0.186	0.198	− 6.0	0.008
$MA1_{VOI}$	8.0 ± 1.5	8.5 ± 1.5	+ 6.6	0.177	0.164	− 6.3	0.010
$Logan_{voxel}$	7.5 ± 1.4	7.9 ± 1.4	+ 5.9	0.191	0.456	−12.6	< 0.001
$MA1_{voxel}$	8.0 ± 1.5	8.4 ± 1.5	+ 6.7	0.159	0.399	− 7.0	0.005

V_T values are mean ± SD from 11 subjects in all 10 regions. Change is (Retest V_T - Test V_T)/Test V_T averaged across regions and across subjects, and expressed as a percentage. Model underestimation of V_T values is with respect to compartmental model. V_T values were compared with factorial repeated measures ANOVA with Bonferroni adjustment using the results in individual regions for all subjects.

Table 3

Retest variability of V_T in different models across brain regions

	Retest variability (%) of V_T				
Brain region	**Compartmental**	**Logan$_{VOI}$**	**MA1$_{VOI}$**	**Logan$_{Voxel}$**	**MA1$_{Voxel}$**
Occipital cortex	11 ± 6	12 ± 5	12 ± 6	12 ± 9	13 ± 8
Cingulate cortex	12 ± 7	11 ± 5	12 ± 5	12 ± 8	12 ± 8
Putamen	13 ± 6	12 ± 6	12 ± 6	12 ± 8	13 ± 8
Thalamus	14 ± 7	12 ± 6	13 ± 6	13 ± 9	14 ± 8
Cerebellum	14 ± 9	12 ± 6	12 ± 6	18 ± 17	19 ± 17

Values are mean ± SD from 11 subjects. The results of the voxel-based analyses were obtained by applying the volumes-of-interest (VOIs) to the parametric images.

Table 4

Retest reliability (ICC) of V_T in different models across brain regions

	ICC				
Brain region	**Compartmental**	**Logan$_{VOI}$**	**MA1$_{VOI}$**	**Logan$_{Voxel}$**	**MA1$_{Voxel}$**
Occipital cortex	0.60	0.50	0.45	0.58	0.54
Cingulate cortex	0.51	0.52	0.47	0.55	0.51
Putamen	0.51	0.48	0.43	0.46	0.43
Thalamus	0.23	0.33	0.27	0.41	0.34
Cerebellum	0.39	0.36	0.31	0.43	0.35