Turku PET Centre Modelling report TPCMOD0012 2002-11-03 DRAFT
Vesa Oikonen

Removing bias in the graphical analysis methods

This document reviews and studies with simulations the strategies for removing the bias caused by measurement errors (noise) to the Patlak and Logan plots, which has been documented previously [Slifstein and Laruelle 2000, Logan et al 2001, Price et al 2002]. The simulations were run in a PIII/Windows 2000 computer, using data and batch files that are included in companion files as specified later.

Demonstrating the effect of a single outlier in TAC

Together with this file, is a collection of data and batch files necessary for this simulation, packed into file tpcmod0012_add1.zip.

Input function was created by fitting rational function to an arterial plasma curve (to which, is not important in this simulation) to remove noise and to interpolate the curve to short time intervals to120 min. One reference tissue region and five regions were simulated using program p2t_3c, one set for irreversible uptake (patlak.dft, Fig. O1) and one set for reversible uptake (logan.dft, Fig. O2). The three-compartment model rate constants can be seen from the batch file test1.bat and test2.bat, that were used to run the simulations.

The effect on four graphical analysis methods is demonstrated: Gjedde-Patlak plot (Fig. O3), Modified Gjedde-Patlak plot (Fig. O4), Logan plot (Fig. O5) and Yokoi plot (Fig. O6). A single outlier in the tissue curve causes only one outlier also in the first two plotting methods (irreversible uptake). The effect on the estimated uptake rate (slope in Gjedde-Patlak plot and y axis intercept in modified Gjedde-Patlak plot) is dependent on in which section of the plot the outlier resides.

However, the plotting methods for reversible uptake require also the integral of tissue curve. Thus, an outlier leads to a bias in all plot data after the outlier itself. Largest effect is naturally seen in the plot point calculated from the outlier point, and although the point is still close to the Logan plot line, it may easily change order with the other plot points. Again, its effect on the distribution volume (DV) or distribution volume ratio (DVR), determined as the Logan plot slope or the Yokoi plot x axis intercept, depends on its place in the plot. The rest of the plot is moved to the direction of the outlier. Similar results were obtained by Slifstein and Laruelle (2000) for Logan plot.

The effect of the noise (i.e. all data points are more or less “outliers”) is studied next.

Demonstrating the effect of noise in TAC

Together with this file, is a collection of data files necessary for this simulation, packed into file tpcmod0012_add2.zip.

The variance-free tissue curves were simulated as before: input function was created by fitting rational function to an arterial plasma curve (to which, is not important in this simulation) to remove noise and to interpolate the curve to short time intervals to120 min. One reference tissue region and five regions were simulated using program p2t_3c, one set for irreversible uptake (patlak.dft) and one set for reversible uptake (logan.dft). The three-compartment model rate constants can be seen from the batch file test1.bat and test2.bat, that were used to run the simulations.

The variance-free curves were cloned to form datafiles, which contained 512 similar sets of curves. Gaussian variance was added to these datafiles using the noise model represented in TPCMOD0008. Proportionality constant was set to 2 and isotope to F-18 for irreversible model and to C-11 for reversible model. The noisy data is contained in datafiles patlakn.dft and logann.dft, respectively. The noisy TACs were also averaged and saved in files patlaka.dft and logana.dft, to make sure that the addition of noise does not produce any bias in the TACs. All these steps are made with batch files test1.bat and test2.bat. The variance-free TACs, TACs averaged from noisy curves, and 5 first noisy curves from selected "regions" are shown in Fig. N1. (irreversible model) and Fig. N2. (reversible model).

Plots calculated from noisy TACs contained often outliers, which were highly outside the normal plot data range. Thus it was not possible to draw meaningful graphs on the plots, showing the effect of noise, but the results of simulations are represented in tables (Tables O1-O4) instead. These calculations with plasma input were made using batch files test1b.bat and test2b.bat. As can be seen in the tables, modified Gjedde-Patlak plot produced essentially equal results as traditional Gjedde-Patlak plot. Variance in modified plot was little smaller, thus it may be worth the effort to test it in pixel-by-pixel compution, where variance may be critical. Yokoi plot is an alternative to Logan plot, and although it produced much more variable results than the Logan plot, especially in low DV, the much smaller bias, and (opposite) direction of bias, favours the Yokoi plot over Logan plot. It may be easier to reduce the variance of Yokoi plot results than to reduce the bias in Logan plots.

Similar results were obtained by Slifstein and Laruelle (2000) for Logan plot.

In addition, the noisy TACs simulated before were used to compare different line fitting methods.

Linear regression models

There exists several methods to fit a straigth line to a *X,Y*-data.

Traditional linear regression, predicting *Y* from *X*. Estimates of the regression parameters

$$y = m \cdot x + b \qquad (1)$$

are searched for by minimizing the sum of squared differences ss_c between the measured and estimated values in n measurements:

$$ss_c = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad (2)$$

The minimal value of ss_c is reached when the regression parameters, slope m_c, and y axis intercept b_c, are [Varga & Szabo 2002]

$$\begin{cases} m_c = Q_{xy} / Q_{xx} \\ b_c = \bar{y} - m_c \cdot \bar{x} \end{cases} \qquad (3)$$

where

$$\begin{cases} \bar{x} = \sum_{i=1}^{n} x_i \Big/ n \\ \bar{y} = \sum_{i=1}^{n} y_i \Big/ n \\ Q_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \\ Q_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \end{cases} \qquad (4)$$

Perpendicular linear regression, to treat the variables symmetrically, when both the x and y variables are subject to measurement errors [Adcock 1878, Pearson 1901, Varga & Szabo 2002]. The sum of Euclidean (perpendicular) distances, ss_p, between the measured points and the fitted straight line are minimized:

$$ss_p = \sum_{i=1}^{n} d_i^{\,2} \qquad (5)$$

Distance d_i from point (x_i, y_i) to line $Ax + By + C = 0$ is [Spiegel & Liu]:

$$d_i = \frac{Ax_i + By_i + C}{\pm\sqrt{A^2 + B^2}} \qquad (6)$$

where the sign is chosen so that the distance is non-negative.

The regression parameters can be calculated from the equations

$$\begin{cases} m_p^2 \cdot Q_{xy} + m_p \cdot (Q_{xx} - Q_{yy}) - Q_{xy} = 0 \\ b_p = \bar{y} - m_p \cdot \bar{x} \end{cases} \qquad (7)$$

where Q_{yy} is calculated from y variables in the same way as Q_{xx} in Eq. (4). The slope mp is solved as the only real root, only positive root, or the root producing smaller ss_p, of the quadratic equation.

Iterative linear least-squares methods, originally put forth by York [York 1966], and further developed, among others, by Lybanon [Lybanon 1984] and Reed [Reed 1992], agree as to the correct values of the slope and intercept. In contrast, there is considerable disagreement as to the variances of slope and intercept between the various algorithms. The method represented in [Reed 1992] is applied in the C-function *llsqwt()*. Since the weighting factors in X and Y for data points (x_i, y_i) are not known, weights are set to 1 in this work. The slope and intercept computed using this method are presented as m_r and b_r.

Least Median Squares (LMS) and **Least Median (Absolute) Deviance (LMD)** estimation are less sensitive to outliers in data [Rousseeuw 1984]. Unfortunately, there exists no closed-form solution to these methods. The slope and intercept need to be determined with iterative optimization methods [Edelsbrunner & Souvaine 1990], and the optimum surface is full of local minima. Therefore, these methods have not yet been applied. C source code for Least Median Absolute Deviation is available in the Numerical Recipes (section 15.7).

Median line estimation methods do not assume Gaussian errors (distribution-free estimation of slope), do not need weighting of data, and are not sensitive to outliers. The most simple method, Theil-Sen estimator [Sen 1968] is applied in this study. In this method, the slope of the line of fit is taken to be the median of the set of slopes that result by passing a line through each pair of distinct points in the plot. In this work, the y axis intercept is determined in similar way. Several other methods do exist, e.g. Siegel's repeated median estimator (RM) [Siegel 1982], but are not applied here.

Comparison of different line fitting methods with noisy plot data

The same simulated noisy TACs that were produced before were used to compare the different line fitting methods. The calculations were included in the same batch files test1b.bat and test2b.bat. The results of the iterative linear least-squares method is not shown here, because without weights the results do not differ from the perpendicular linear regression method.

Gjedde-Patlak plot results were equal when line was fitted using perpendicular and traditional regression method. Also median lines were similar, except the SD was little smaller. The perpendicular regression and iterative methods were not applicable to the modified Gjedde-Patlak plot, but produced highly variable results. The median method resulted into a bit higher variance than traditional regression method.

The bias in Logan plot results was decreased when perpendicular regression methods were used, but variance was increased (Tables N3 and F1). The median method resulted into both decreased bias and decreased variation (Tables N3 and F2). Perpendicular regression method did not change the Yokoi plot results (Tables N4 and F3). Median method decreased the positive bias and variance in low DV (Tables N4 and F4), which supports the idea that after further method refinement, Yokoi plot could be used instead of Logan plot.

Logan's proposition for removing the bias

The slope obtained from the Logan plot with the traditional linear regression model is biased and has been shown to result in increasing underestimation of DVs with increasing noise. Possible reason for for the bias is that the errors add up during integration, and to compensate for this effect, the application of generalized linear (iterative) least-squares methods has been proposed by Logan et al. (2001). This method was applied to Logan plot with both plasma and reference tissue input, and it was found to effectively reduce the bias. However, also the variance of results was increased somewhat [Logan et al. 2001].

This method has not yet been applied in Turku PET Centre. It is uncertain, whether receptor binding potential measurements will benefit from this fairly complicated method, as compared to the RPM (SRTM) method, which is computationally fast and validated for a few tracers, even in cases when the model requirements are not strictly met.

Likelihood approach for removing bias

Recently, Ogden et al (2002) presented a method which removes the noise from the data in the linear range of the Logan plot. This method is model independent and as such more reasonable to be applied with intrinsically model independent Logan analysis than the Logan's proposition. This method has no general analytic solution either, and its use requires a non-linear optimization, which makes also its usefulness questionable.

Automatic detection of linear plot range

In yokoi program version 1.1 (2002-10-31) and logan 1.6 (2002-10-31) an option for automatic detection of plot range that produces the least coeffcient of variation for the DV was added. This applies only to traditional regression line fit (option –c), and to iterative line fir method (option –r), in which the plot range is determined based on the best line fit.

The results calculated from the previously simulated noisy datasets, using batch file test2bx.bat, are represented in Tables B1 and B2. This method reduced the variance of DV values, even so, that standard deviations were lower than when using Logan plot with any line fit method. The bias was smaller in all regions, although surprisingly, in Tis3 (DV=6) the bias was still too high (-7%), which is higher than the bias from fixed-range Yokoi plot, but much lower than from Logan plot. This "region" was simulated with the smallest k_4 value, leading into relatively slow kinetics and late equilibrium.

Summary and directions for future development

Gjedde-Patlak plot was found to not to produce bias with noisy data with plasma (noiseless) input. Modified Gjedde-Patlak was found to provide equal results as the traditional Gjedde-Patlak plot in the simulations. Since in modified plot the input values are not used as dividers, it may be possible that in cases where the input curve falls very rapidly close to zero, the modified plot might provide more reliable results;

this should be investigated with further simulations and actual PET data, e.g. from [^{11}C]choline studies.

In contrast to the results of Varga and Szabo (2002), different line fitting method was not found to remove most of the bias in Logan plot. Instead, the variance was increased. If other line fitting methods are to be considered, the median line methods seem to be more promising in removing the bias caused by outliers. However, none of the line fitting methods is assumed to totally remove the bias caused by additive noise errors which affect the plot as a whole, not only causing outliers in the plot.

Yokoi plot seemed to produce lower bias to DV values than Logan plot and it even decreased the variance, unlike the other proposed bias reducton methods [Logan et al. 2001, Varga and Szabo 2002]. However, Yokoi plot seems to be more sensitive to the fit range: high positive bias and variance in low DV is caused by the small range of x values in Yokoi plot (tissue integral / input integral), if the same fit range is used as for high DV regions. Noise in this case can easily turn the fitted line so that x axis intercept (DV) may get very high values. Thus, in low DV, the fit must be started earlier than in high DV. The determination of the best range to fit is difficult for noisy data, and it can lead to negative bias in DV, especially with slow kinetics.

The effect of noisy reference tissue input was not simulated yet, although the noisy reference tissue curves were produced. The noise model has been widely used in literature, but is not validated. It also seems that biases are dependent on regional kinetics. Therefore, it is necessary to make the simulations with realistic noise levels and kinetic parameters, derived from actual PET data. And this must be done separately for different tracers.

The effect of tissue heterogeneity was not studied. It is known not to affect the linearity of the traditional Gjedde-Patlak plot, but it may cause nonlinearity and positive bias (end of plot is curved up) in Yokoi plot [Yokoi et al. 1993].

References

1. Adcock RJ. A problem in least squares. *Analyst* 1878; 5:53-54.
2. Edelsbrunner H, Souvaine DL. Computing least median of squares regression lines and guided topological sweep. *J. Am. Stat. Assoc.* 1990; 85: 115-119.
3. Logan J, Fowler JS, Volkow ND, Ding YS, Wang G-J, Alexoff DL. A strategy for removing the bias in the graphical analysis method. *J. Cereb. Blood Flow Metab.* 2001; 21:307-320.
4. Lybanon M. A better least squares method when both variables have uncertainties. *Am. J. Phys.* 1984; 52: 22-26 and 276-278.
5. Ogden RT, Parsey RV, Mann JJ. Likelihood approach to parameter estimation in Logan graphical analysis. *NeuroImage* 2002; 16: S73.
6. Pearson K. On lines and planes of closest fit to systems of points in space. *Philosophical Mag.* 1901; 2:559-572.
7. Price JC, Xu L, Mazumdar S, Meltzer CC, Drevets WC, Mathis CA, Kelley DE, Ryan CM, Reynolds III CF. Impact of graphical analysis bias on group comparisons of regional [carbonyl-^{11}C]WAY binding potential measures. *NeuroImage* 2002; 16: S72.
8. Reed BC. Linear least-squares fits with errors in both coordinates. II. Comments on parameter variances. *Am. J. Phys.* 1992; 60: 59-62.
9. Rousseeuw PJ. Least median of squares regression. *J. Am. Stat. Assoc.* 1984; 79: 871-880.
10. Sen PK. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* 1968; 63: 1379-1389.
11. Siegel AF. Robust regression using repeated medians. *Biometrika* 1982; 69: 242-244.
12. Slifstein M, Laruelle M. Effects of statistical noise on graphical analysis of PET neuroreceptor studies. *J. Nucl. Med.* 2000; 41: 2083-2088.
13. Spiegel MR, Liu J. *Mathematical handbook of formulas and tables.* 2nd ed. McGraw-Hill, 1999. ISBN 0-07-038203-4
14. Varga J, Szabo Z. Modified regression model for the Logan plot. *J. Cereb. Blood Flow Metab.* 2002; 22:240-244.
15. Yokoi T, Iida H, Itoh H, Kanno I. A new graphic plot analysis for cerebral blood flow and partition coefficient with iodone-123-iodoamphetamine and dynamic SPECT validation studies using oxygen-15-water and PET. *J. Nucl. Med.* 1993; 34: 498-505.
16. York D. Linear-squares fitting of a straight line. *Can. J. Phys.* 1966; 44: 1079-1086.

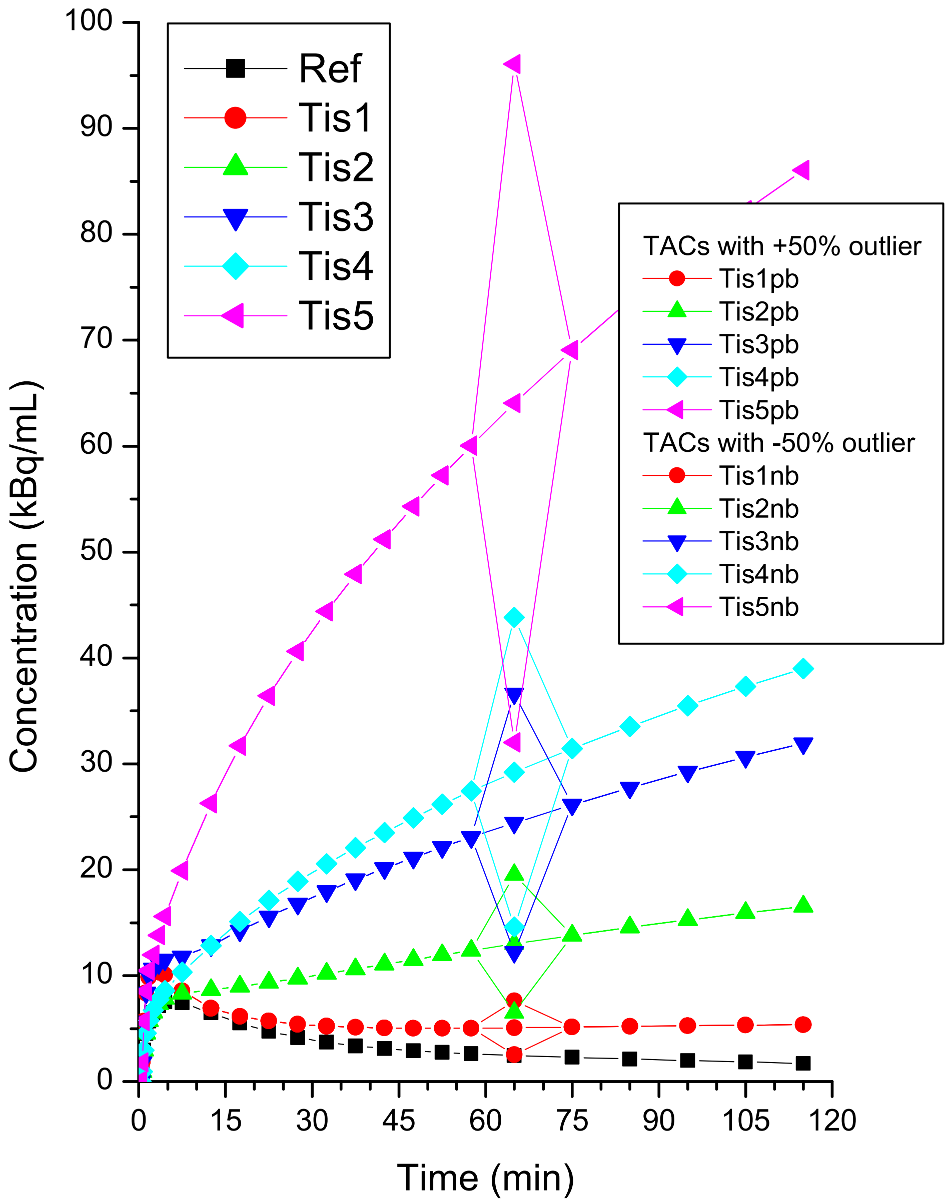

Fig. O1. Simulated data with single outliers. Irreversible 3-compartment model.

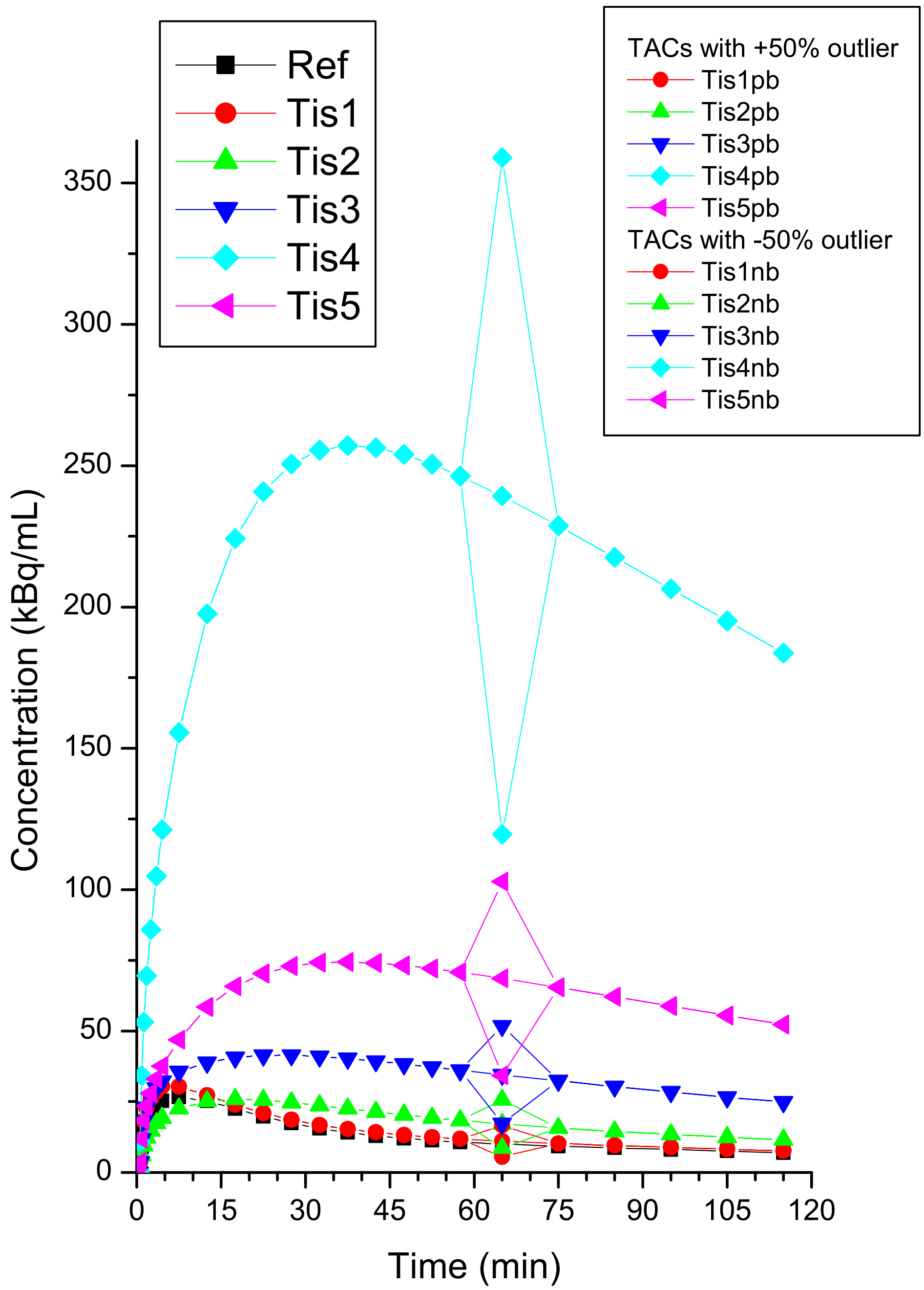

Fig. O2. Simulated data with single outliers. Reversible 3-compartment model.

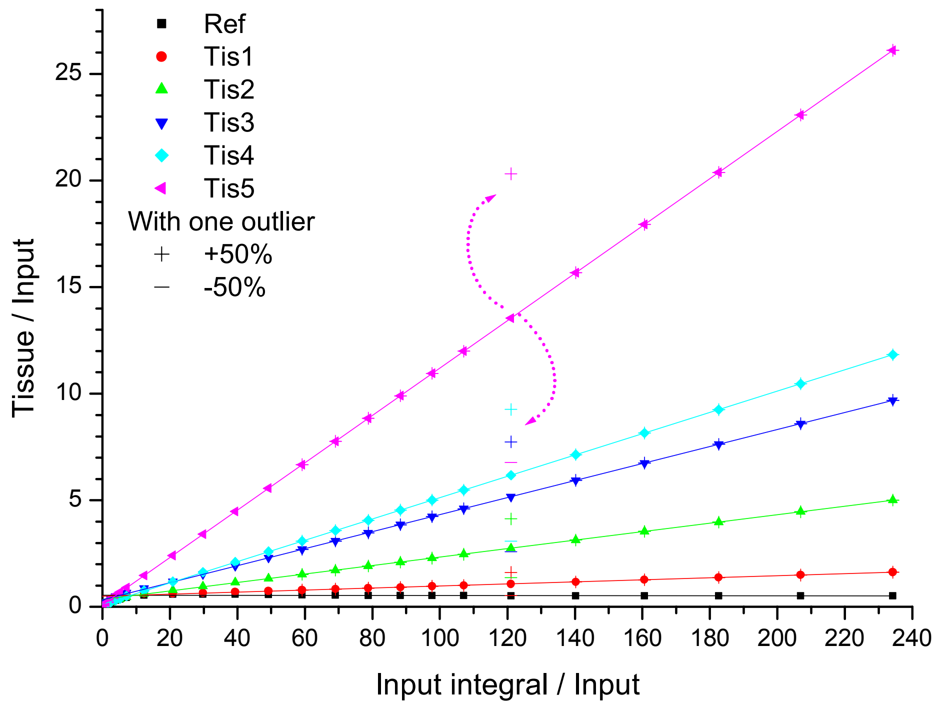

Fig. O3. The effect of single outlier on Gjedde-Patlak plots.

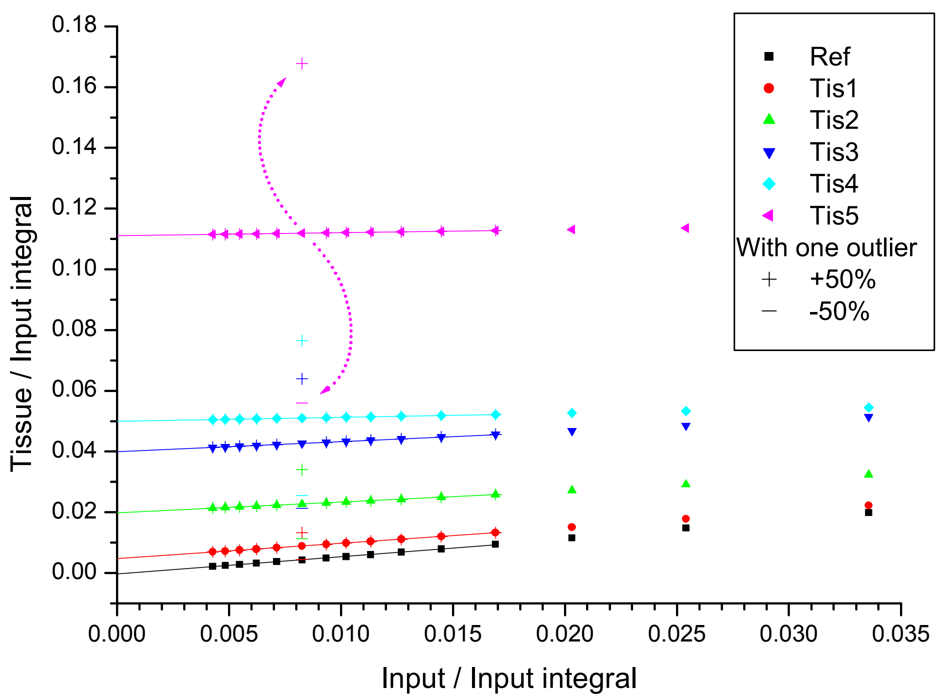

Fig. O4. The effect of single outlier on modified Gjedde-Patlak plots.

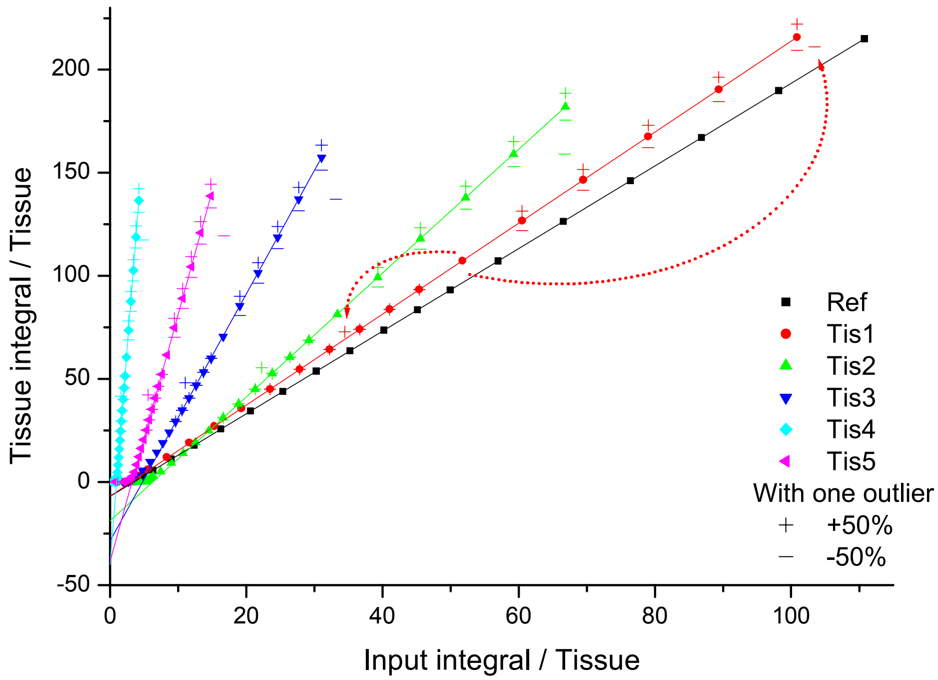

Fig. O5. The effect of single outlier on Logan plots.

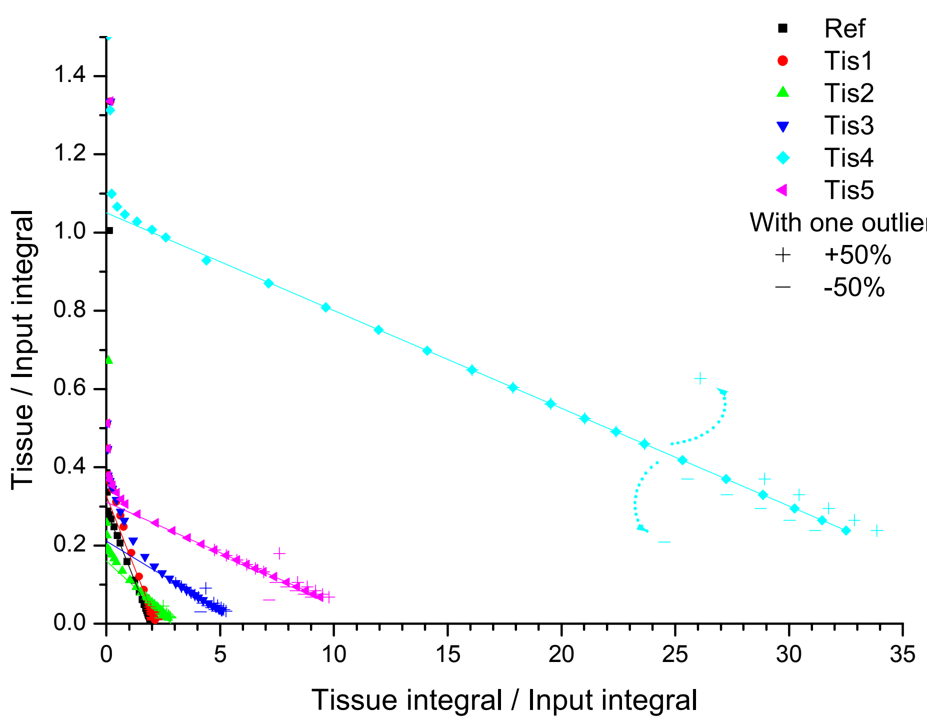

Fig. O6. The effect of single outlier on Yokoi plots.

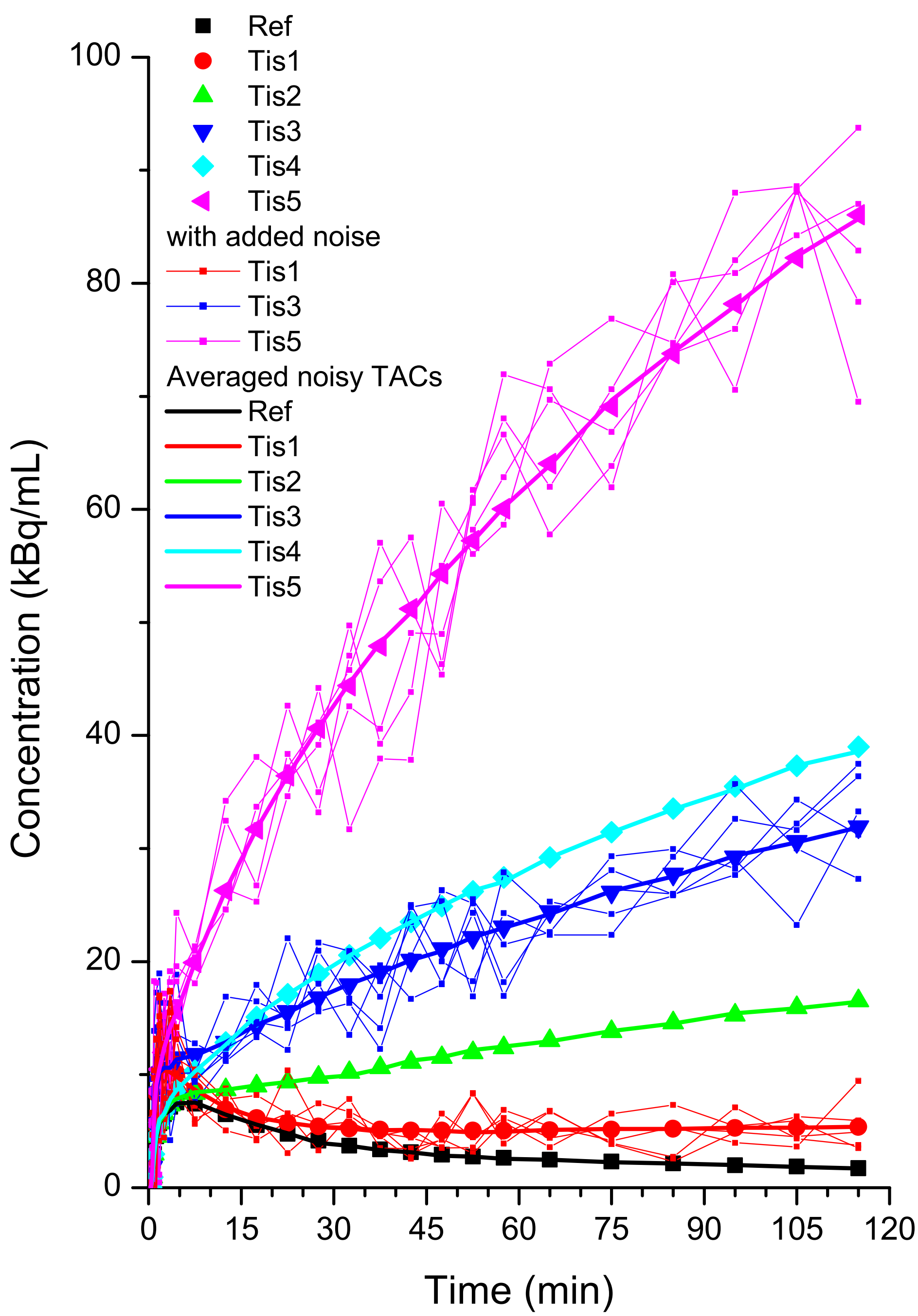

Fig. N1. Simulated TACs from irreversible model. Graph contains the first five noisy curves from selected simulations, and the average curve calculated from all 512 noisy TACs.

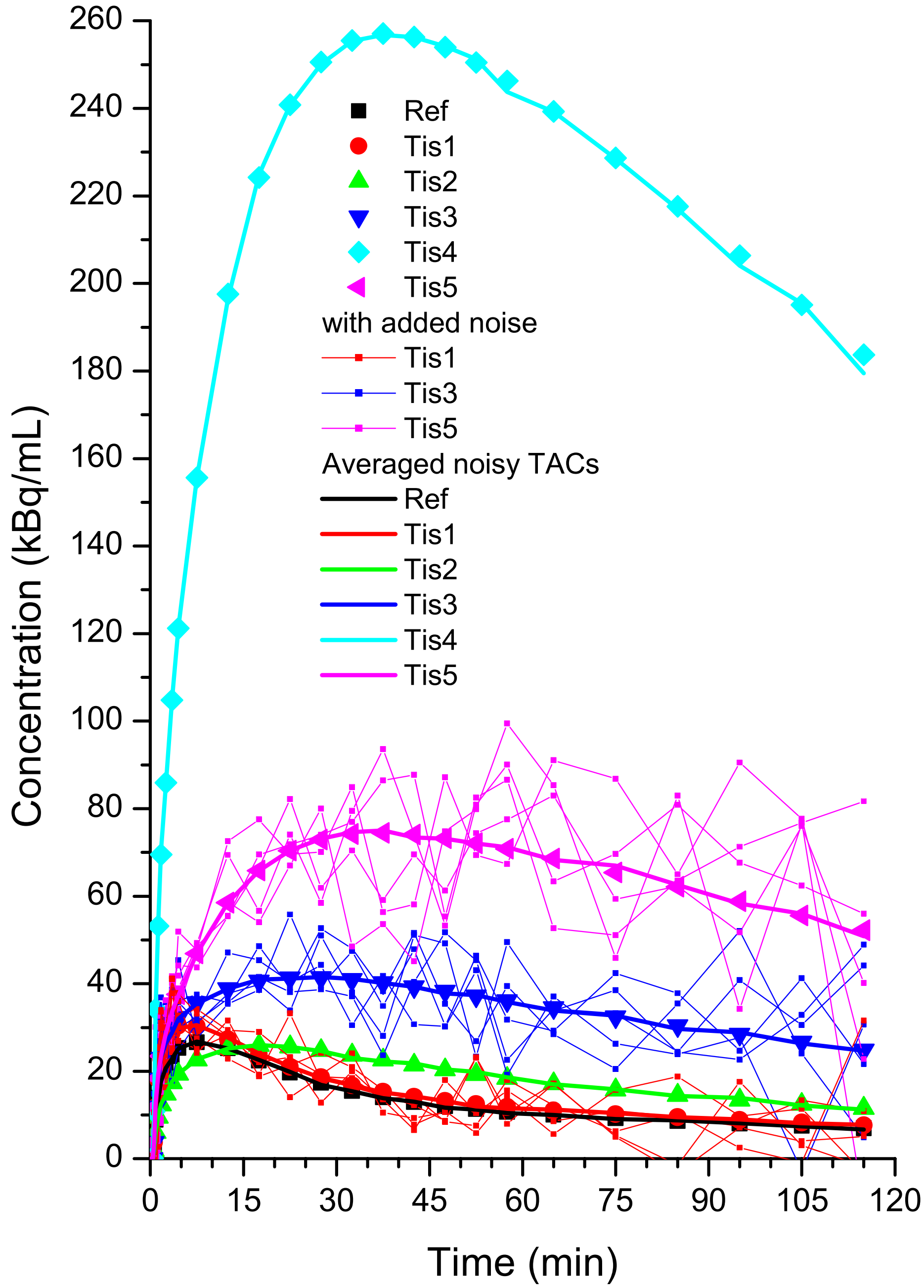

Fig. N2. Simulated TACs from reversible model. Graph contains the first five noisy curves from selected simulations, and the average curve calculated from all 512 noisy TACs.

Table O1. Gjedde-Patlak plot results (*K_i*) from simulated noisy TACs, as compared to the results obtained from TACs averaged from the noisy TACs.

	K_i without noise	Noise-indused bias (%) in mean	SD
Ref	-0.0002	0	0.0013
Tis1	0.0049	0	0.0020
Tis2	0.0199	0	0.0033
Tis3	0.0399	0	0.0046
Tis4	0.0496	0	0.0050
Tis5	0.1107	0	0.0073

Table O2. Modified Gjedde-Patlak plot results (*K_i*) from simulated noisy TACs, as compared to the results obtained from TACs averaged from the noisy TACs.

	K_i without noise	Noise-indused bias (%) in mean	SD
Ref	-0.0003	0	0.0016
Tis1	0.0048	0	0.0020
Tis2	0.0201	0	0.0031
Tis3	0.0398	0	0.0042
Tis4	0.0497	0	0.0046
Tis5	0.1108	0	0.0068

Table O3. Logan plot results (*DV*) from simulated noisy TACs, as compared to the results obtained from TACs averaged from the noisy TACs.

	DV without noise	Noise-indused bias (%) in mean	SD
Ref	1.9888	-3	0.1839
Tis1	2.2108	-4	0.2028
Tis2	3.0077	-9	0.3092
Tis3	5.9480	-11	0.5726
Tis4	41.708	-5	2.8875
Tis5	11.979	-12	1.1192

Table O4. Yokoi plot results (*DV*) from simulated noisy TACs, as compared to the results obtained from TACs averaged from the noisy TACs.

	DV without noise	Noise-indused bias (%) in mean	SD
Ref	1.9869	-14	7.3444
Tis1	2.2075	+37	14.388
Tis2	3.0212	+10	2.6712
Tis3	5.9301	+4	0.9225
Tis4	41.780	+1	2.8791
Tis5	12.006	+3	1.7196

Table F1. Logan plot results (*DV*) from simulated noisy TACs, as compared to the results obtained from TACs averaged from the noisy TACs. Instead of traditional regression, line is fitted with a method which treats both plot variables symmetrically.

	DV without noise	Noise-indused bias (%) in mean	SD
Ref	1.9880	-3	0.1825
Tis1	2.2108	-3	0.2085
Tis2	3.0078	-8	0.3394
Tis3	5.9485	-9	0.6563
Tis4	41.710	-3	3.1032
Tis5	11.981	-9	1.3430

Table F2. Logan plot results (*DV*) from simulated noisy TACs, as compared to the results obtained from TACs averaged from the noisy TACs. Instead of traditional regression line fit, median method is used to estimate plot slope and intercepts.

	DV without noise	Noise-indused bias (%) in mean	SD
Ref	1.9866	-3	0.1839
Tis1	2.2098	-3	0.1970
Tis2	3.0130	-7	0.3032
Tis3	5.9380	-9	0.5508
Tis4	41.889	-2	2.6899
Tis5	11.996	-8	1.0929

Table F3. Yokoi plot results (*DV*) from simulated noisy TACs, as compared to the results obtained from TACs averaged from the noisy TACs. Instead of traditional regression, line is fitted with a method which treats both plot variables symmetrically.

	DV without noise	Noise-indused bias (%) in mean	SD
Ref	1.9869	-13	7.2975
Tis1	2.2075	+36	13.900
Tis2	3.0212	+10	2.6705
Tis3	5.9301	+4	0.9223
Tis4	41.780	+1	2.8790
Tis5	12.006	+3	1.7193

Table F4. Yokoi plot results (*DV*) from simulated noisy TACs, as compared to the results obtained from TACs averaged from the noisy TACs. Instead of traditional regression line fit, median method is used to estimate plot slope and intercept.

	DV without noise	Noise-indused bias (%) in mean	SD
Ref	1.9849	+11	1.3000
Tis1	2.2103	+2	3.5778
Tis2	3.0195	+7	0.9691
Tis3	5.9374	+4	1.0393
Tis4	41.835	+1	3.1145
Tis5	12.020	+2	1.7056

Table B1. Logan plot results (*DV*) from simulated noisy TACs, as compared to the results obtained from TACs averaged from the noisy TACs. Traditional regression line fit, finding automatically plot range with the smallest CV of DV.

	DV without noise	Noise-indused bias (%) in mean	SD
Ref	1.9885	-4	0.1696
Tis1	2.2082	-4	0.1865
Tis2	3.0046	-9	0.2851
Tis3	5.9480	-12	0.4819
Tis4	41.796	-4	2.5490
Tis5	11.972	-11	1.0769

Table B2. Yokoi plot results (*DV*) from simulated noisy TACs, as compared to the results obtained from TACs averaged from the noisy TACs. Traditional regression line fit, finding automatically plot range with the smallest CV of DV.

	DV without noise	Noise-indused bias (%) in mean	SD
Ref	1.9901	+0	0.1451
Tis1	2.1887	+0	0.1505
Tis2	3.0146	-3	0.2632
Tis3	5.9301	-7	0.5004
Tis4	41.818	+0	2.1916
Tis5	11.963	-2	0.9853