ORIGINAL RESEARCH

Multivariate analysis of PET pharmacokinetic parameters improves inferential efficiency

Granville J. Matheson[1,2,3]* and R. Todd Ogden[1,2]

*Correspondence:
granville.matheson@nyspi.
columbia.edu

[1] Department of Psychiatry, Columbia University, New York, NY 10032, USA
[2] Department of Biostatistics, Columbia University Mailman School of Public Health, New York, NY 10032, USA
[3] Department of Clinical Neuroscience, Center for Psychiatry Research, Karolinska Institutet and Stockholm County Council, SE-171 76 Stockholm, Sweden

Abstract

Purpose: In positron emission tomography quantification, multiple pharmacokinetic parameters are typically estimated from each time activity curve. Conventionally all but the parameter of interest are discarded before performing subsequent statistical analysis. However, we assert that these discarded parameters also contain relevant information which can be exploited to improve the precision and power of statistical analyses on the parameter of interest. Properly taking this into account can thereby draw more informative conclusions without collecting more data.

Methods: By applying a hierarchical multifactor multivariate Bayesian approach, all estimated parameters from all regions can be analysed at once. We refer to this method as Parameters undergoing Multivariate Bayesian Analysis (PuMBA). We simulated patient–control studies with different radioligands, varying sample sizes and measurement error to explore its performance, comparing the precision, statistical power, false positive rate and bias of estimated group differences relative to univariate analysis methods.

Results: We show that PuMBA improves the statistical power for all examined applications relative to univariate methods without increasing the false positive rate. PuMBA improves the precision of effect size estimation, and reduces the variation of these estimates between simulated samples. Furthermore, we show that PuMBA yields performance improvements even in the presence of substantial measurement error. Remarkably, owing to its ability to leverage information shared between pharmacokinetic parameters, PuMBA even shows greater power than conventional univariate analysis of the true binding values from which the parameters were simulated. Across all applications, PuMBA exhibited a small degree of bias in the estimated outcomes; however, this was small relative to the variation in estimated outcomes between simulated datasets.

Conclusion: PuMBA improves the precision and power of statistical analysis of PET data without requiring the collection of additional measurements. This makes it possible to study new research questions in both new and previously collected data. PuMBA therefore holds great promise for the field of PET imaging.

Keywords: Positron emission tomography, Bayesian statistics, Precision, Power, Pharmacokinetic modelling

Introduction

Positron emission tomography (PET) is an in vivo neuroimaging method with high biochemical sensitivity and specificity. It is an essential tool for the study of the neurochemical pathophysiology of psychiatric and neurological disease, as well for pharmaceutical research. However, PET is a very costly and invasive procedure that involves exposing participants to radioactivity, thereby limiting the feasibility of large studies. As a result, low statistical power is a common obstacle encountered for studying clinically relevant research questions. Efforts to improve the power of PET imaging have typically focused on the development of new radiotracers with improved sensitivity as well as new pharmacokinetic (PK) models with greater accuracy; more recently, there have been data standardisation and sharing initiatives to foster inter-group collaboration and increase sample sizes [18, 19, 28]. However, there has been comparatively little attention paid to the development of more nuanced statistical analysis of PET data for the same purpose.

PET quantification involves fitting PK models to a series of radioactivity concentrations in a region of the brain over time, called a time activity curve (TAC), most often using nonlinear least squares (NLS) optimisation. These models typically consist of between 1 and 5 parameters of which one (or a function of two or more parameters) is used as a measure of the binding of the radioligand to the target protein. Once the TAC data from all regions and all subjects have been fit using the selected model, the parameter estimates reflecting target binding for each region and subject are then entered into a subsequent statistical model, e.g., a *t* test comparing patients and control subjects, while the other estimated parameters are not taken into account in the analysis.

We recently introduced SiMBA [23], which makes use of Bayesian hierarchical multifactor modelling to fit PET TAC data and perform statistical analysis simultaneously across both individuals and regions. The primary disadvantage of this technique is that it is highly computationally intensive, and currently only implements the two-tissue compartment model [15]. However, the model improves the estimation of binding parameters, and yields substantial advantages in terms of increased precision and statistical power for statistical comparisons. Intriguingly, in simulation studies we found that the statistical power of SiMBA for detecting group differences was even greater than for univariate statistical analysis performed on the "true" binding values from which the TACs were generated. This suggests that even if binding measures could be measured exactly for each subject, it would still not be possible to attain the statistical power that we observed with SiMBA.

Upon further inspection, we discovered that this performance gain could be explained by the multivariate modelling strategy employed in SiMBA. In other words, instead of extracting only a single parameter as a measure of binding, statistical analysis was performed using all estimated parameters simultaneously, thereby allowing the model to exploit shared information among all the PK parameters. This general concept is demonstrated in Fig. 1: the shape of the 2-dimensional density plot of the two variables is highly dependent upon their correlation with one another. If two parameters are highly correlated with one another, then the conditional variance of each parameter at any given value of the other is considerably smaller. Hence, if the estimation of both variables and their correlation with one another are sufficiently precise, then the *conditional* variance of estimated parameters can be reduced to below that of the marginal true values. In
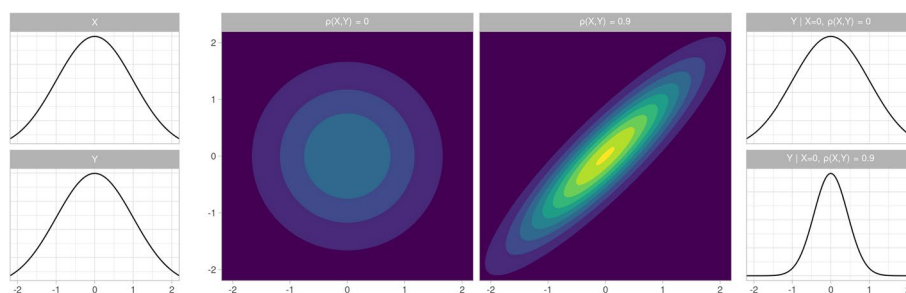
Fig. 1 Comparisons of marginal and conditional densities for multivariate normal distributions. Left: Marginal densities of variables X and Y after standardisation. Middle: Multivariate contour plots of the two-dimensional densities of X and Y, with either no correlation or a strong correlation between them. Right: Conditional densities of variable Y conditional on X when there is either no correlation or a strong correlation between variables X and Y

other words, by exploiting shared information between parameters, even if those parameters are not directly relevant to the statistical contrast of interest, the performance of the statistical model can be improved.

In this study, we evaluate whether applying a multivariate statistical analysis to PET pharmacokinetic outcome parameters estimated in the conventional manner using NLS estimation can also provide inferential advantages, without needing to fit the full SiMBA model to the dynamic TAC data. We refer to this approach, by analogy with SiMBA, as Parameters undergoing Multivariate Bayesian Analysis: PuMBA. The computational requirements for this modelling strategy are on the order of minutes on a single core, compared with days for SiMBA, and can readily be adapted to a wider range of pharmacokinetic models, thereby facilitating its application to a broader range of research questions. PuMBA may therefore serve as a convenient intermediate substitute for a full SiMBA analysis.

Methods

Model specification

PuMBA can be described as a multivariate hierarchical multifactor model. It is multivariate in that there are multiple dependant variables estimated at once—in contrast with a multivariable model in which there are multiple *independent* variables. It is hierarchical in that it makes use of "partial pooling". This means that parameters are modelled as originating from a common distribution, and are therefore shrunk towards the global mean in an adaptive regularisation process. This shrinkage allows the model to take advantage of similarities between individuals within the dataset to improve its inferences [3, 24, 25]. Finally, PuMBA is multifactor in that there are multiple hierarchies at once within which we perform partial pooling [5].

For PuMBA, as for SiMBA, linear models are defined for each of the PK parameters, defined by an intercept, covariates and partially pooled deviations from the expectation value for each individual j and region k, for each of the m PK parameters. We define a global mean intercept (α) for each parameter, representing the mean value for that parameter. For each PK parameter i, the influence of covariates for individual j are expressed by a covariate vector (β_i) multiplied by a covariate matrix ($X_{i,j}$, or more specifically its transpose, $X_{i,j}^T$ within the linear model). These covariate matrices are defined

independently for each PK parameter, and can include variables such as age, sex or group membership, for instance. Lastly, we define an additive sequence of differences for each of the separate hierarchies [5]: across individuals (τ_j), across regions (υ_k), as well as a final term for residual variation ($\epsilon_{j,k}$). These deviations are drawn from multivariate normal distributions, from which each draw is an m-dimensional vector.

$$\theta_{i,j,k} = \alpha_i + X_{i,j}^T \beta_i + \tau_{i,j} + \upsilon_{i,k} + \epsilon_{i,j,k}$$
$$[\tau_{1,j}, \ldots, \tau_{m,j}]^T \sim \text{MVNormal}([\mathbf{0}], \Sigma_{\text{Subject}})$$
$$[\upsilon_{1,k}, \ldots, \upsilon_{m,k}]^T \sim \text{MVNormal}([\mathbf{0}], \Sigma_{\text{Region}})$$
$$[\epsilon_{1,j,k}, \ldots, \epsilon_{m,j,k}]^T \sim \text{MVNormal}([\mathbf{0}], \Sigma_{\text{residual}})$$

This defines the generalised model framework, in which estimation is performed using partial pooling for all parameters across all hierarchies, resulting in some degree of shrinkage towards the mean. In practice, shrinkage of most parameters towards a shared mean is desirable; however, regional differences in certain parameters are so heterogeneous that a common distribution cannot be assumed owing to regional neuroanatomical differences. For this reason, blood delivery and binding parameters are estimated independently from one another without pooling, i.e. using fixed effects. More details are provided in the following section.

Model implementation

Firstly, all PK parameters are transformed to their natural logarithms. This serves several purposes. Firstly, this naturally constraints all parameters to be positive, corresponding to their theoretical range as biological quantities and rate constants. Secondly, this serves to define additive differences within the linear model as proportional differences in the original quantity, since biological differences or changes in PET are typically assumed to exhibit similar proportional, as opposed to absolute, differences between different regions or individuals. Lastly, this serves to stabilise the variance between regions: in PET, we typically make the assumption that the proportional variance between regions is relatively similar.

The input parameters for PuMBA are the PK parameters estimated by the kinetic model from the TACs using NLS. Importantly, this means that PuMBA does not improve the estimation of PK model parameters at the individual TAC level, but rather that it improves subsequent statistical inferences drawn from these parameters.

For the two-tissue compartment (2TC) model [17], the model parameters were K_1, V_{ND}, BP_{P} and k_4. For the 1TC [17], we used K_1 and V_{T}. Finally, for the SRTM [20], we used R_1, k_2' and BP_{ND}. We selected parameterisations of the model parameters in such a way as to improve the ease by which priors could be defined. To this end, for each model we defined a binding parameter and a blood delivery parameter, and defined the remaining PK parameters in such a way as to maximise the extent to which shrinkage towards a common mean value is most theoretically motivated, i.e. which can be considered as originating from a common distribution. For the 2TC, BP_{P} was selected over BP_{ND}, as the former is more identifiable using NLS and in simulations, estimated values show stronger correlations with the true values compared to BP_{ND}. For SRTM, we used k_2' rather than k_2 as the former parameter is a property of the reference region and

should theoretically be fairly consistent between regions within each individual, similar to how this parameter is set to a global estimate using SRTM2 [39].

For the definition of covariate matrices for each parameter, this task can depend both on the tracer and the sample itself. For instance, age might be a predictor for both blood delivery and binding. On the other hand, patient status might only be included as a predictor for binding—unless the condition is also thought to affect regional blood delivery, in which case patient status might also be included as a predictor for blood delivery for example. Careful judgement should be applied to this task, although model comparison methods can also be helpful [12, 24, 38].

Model fitting

We make use of multivariate Bayesian hierarchical multifactor modelling to fit the model described above using Markov Chain Monte Carlo (MCMC) sampling. We defined the model using the STAN probabilistic programming language [8], which applies Hamiltonian Monte Carlo (HMC), with code generated using `brms` 2.15.0 [7] using R version 4.0.5 (*Shake and Throw*) [33].

Priors were specified in such a way as to exclude parameter values which could be deemed as unlikely *a priori* based on domain knowledge, but not to greatly inform the model. We used moderately informative normal priors for the intercept (α) terms, and zero-centred half-normal regularising priors for the standard deviation of all pooled parameters. LKJ [21] priors were defined for correlation matrices. More details are provided in Additional file 1: S1.

NLS parameter estimation was performed using `kinfitr` [22, 37] for the one-tissue compartment model (1TC) and the simplified reference tissue model (SRTM) [20]. For the two-tissue compartment model (2TC), the model was fitted directly using NLS using an analytical convolution of the arterial input function with the impulse response function, as previously described [23], solving for K_1, V_{ND}, BP_P and k_4. In all cases, weights were estimated using the default `kinfitr` weighting scheme.

Model assumptions

There are several major assumptions underlying PuMBA. First, PuMBA assumes that the estimated PK parameters are an approximately unbiased representation of the true underlying biological quantities. Second, PuMBA assumes that the distribution of these true underlying biological quantities can be described by a statistical distribution such as the normal distribution. Third, PuMBA assumes that it is possible to model correlations between these PK parameters between individuals and regions if they exist, and that these correlations can be exploited to inform inference.

When it comes to implementation, we usually make the assumption that a logarithmic transformation is justified, and that parameters exhibit similar proportional variance between regions, and exhibit approximately proportional relationships with covariates. Similarly, we tend to assume that the multivariate correlation structure is similar between groups—although this assumption can be relaxed at the likely cost of some reduction in inferential efficiency.
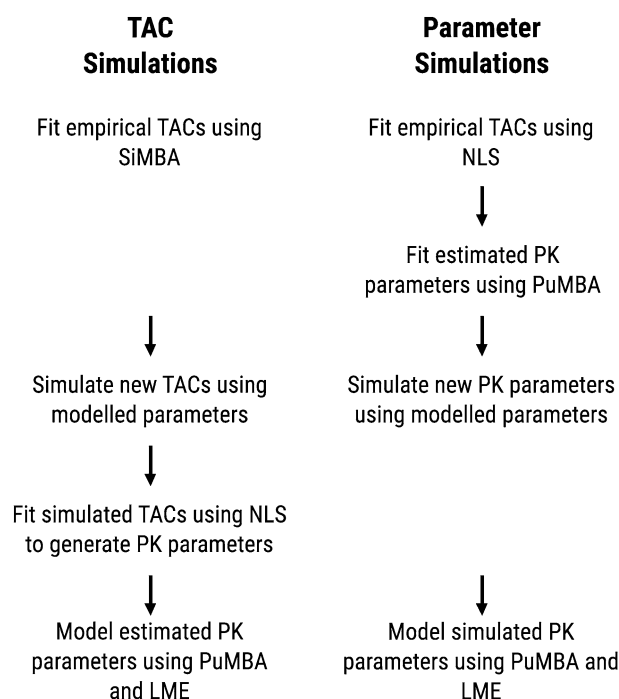
Simulations

For the purpose of assessing the performance of this modelling approach, we generated simulated datasets to compare the proposed methodology with that of a more conventional approach, i.e. univariate analysis of the estimated binding parameter. Simulation parameters were generated based on the posterior mean values of parameters estimated from empirical data by fitting the relevant model to the data and simulating from the estimated parameters. The datasets used were as follows: 97 individuals measured with $[^{11}C]$WAY100635 [9], 16 measurements from 8 individuals measured with $[^{11}C]$ABP688 [10], 47 measurements with $[^{11}C]$DASB from 33 individuals [29, 32] and 23 individuals measured with $[^{11}C]$GR103545 [26]. Simulated datasets had between 10 and 100 individuals in each of a patient and a control group, i.e. between 20 and 200 individuals in total with equal sample sizes in each group. Data from nine regions were included for each ligand (more details in Additional file 1: S2). For $[^{11}C]$WAY100635, the regions were the dorsolateral prefrontal cortex (DLPFC), medial prefrontal cortex (MPFC), hippocampus, amygdala, parahippocampus, insula, anterior cingulate cortex (ACC), posterior cingulate cortex (PCC) and dorsal raphe nucleus (DRN). For $[^{11}C]$ABP688, the regions were the DLPFC, MPFC, amygdala, hippocampus, dorsal putamen, ventral striatum, insula, PCC and ACC. For $[^{11}C]$DASB, the regions were ACC, dorsal putamen, ventral striatum, amygdala, thalamus, hippocampus, PCC, insula and midbrain. For $[^{11}C]$GR103545, the regions were DLPFC, amygdala, dorsal putamen, hippocampus, insula, MPGC, parahippocampus, DRN and ventral striatum.

We simulated new sets of individuals by simulating from the estimated multivariate and univariate normal distributions describing variation between individuals and regions within individuals. When simulating regional variation, we used the estimated values for each region, rather than simulating from the estimated distributions. In this way, we simulate from the same set of regions, but within a unique set of individuals, with a unique set of individual variations at the regional level.

We set global group differences in the natural logarithm of the binding parameter to be equal to 0.1, corresponding to a 10.5% difference between groups, and separately to zero to assess the false positive rate. Simulated data were generated without any effects of age or sex, and so these covariates were not included in the PuMBA models applied to the simulated data. The univariate analyses performed included *t* tests and linear mixed effects (LME) models with the natural logarithm of the parameter representing ligand binding as the dependent variable, considering group and region as fixed effects and subject as the only random effect. The *t* tests were performed independently for each region of the dataset. LME analysis was performed across all regions using `lme4` [1].

We evaluated power and false positive rates by fitting logspline density functions [36] to the upper and lower bounds of the 95% confidence/credible intervals of the estimated difference between groups. We then assessed the cumulative density of these fitted density functions above and below zero, which we could use to estimate the proportion of simulated datasets for which the 95% confidence/credible interval would overlap with (or not overlap with) zero. We have shown previously that this method closely aligns with empirical estimates, and allows for the estimation of power in small numbers of trials [23].

TAC Simulations **Parameter Simulations**

Fit empirical TACs using SiMBA

Fit empirical TACs using NLS

↓

Fit estimated PK parameters using PuMBA

↓ ↓

Simulate new TACs using modelled parameters

Simulate new PK parameters using modelled parameters

↓

Fit simulated TACs using NLS to generate PK parameters

↓ ↓

Model estimated PK parameters using PuMBA and LME

Model simulated PK parameters using PuMBA and LME

Fig. 2 Summary of the approaches used for the TAC and parameter simulations

TAC simulations

For TAC simulations (Fig. 2), we first fit the TAC data using SiMBA, and then used the posterior mean values of the SiMBA fit to simulate new TAC data. We used SiMBA to simulate TACs because SiMBA is fitted to TACs directly, and can therefore generate TACs corresponding to the estimated population parameters directly. To model these data, the TACs were first fitted using NLS to generate PK parameters, which served as the input to the LME and PuMBA models. Data were generated with the standard deviation of the measurement error equal to that estimated in the original sample. We also considered half, double and quadruple the original measurement error to assess the sensitivity of the different approaches to the magnitude of the measurement error. We assessed the performance of these approaches using 500 simulated datasets for each condition.

We also assessed the performance of PuMBA in the same simulated TAC datasets for each condition as examined previously with SiMBA [23], in order to compare the performance of PuMBA and SiMBA in the same data. In this dataset, group differences were equal to 0.182 in the natural logarithm of BP_P, corresponding to a group difference of 20%. This included 50 simulated datasets for each condition owing to the much-greater computational burden of SiMBA.

The simulation parameters are provided in Additional file 1: S2. The binding parameter for which group differences were simulated and estimated, was BP_P.

Parameter simulations

While the TAC simulations above were based on parameters estimated using SiMBA, SiMBA only currently applies the 2TC. For SRTM and 1TC, we therefore simulated data from estimated PuMBA parameters instead. To this end, the empirical TAC data were first fitted using NLS to generate PK parameters, which were then modelled using PuMBA (Fig. 2). The posterior mean values of the PuMBA model fit were used to simulate new sets of parameter estimates. PuMBA parameters, however, do not allow for the discrimination of variance originating from error in the estimation of the PK parameters using NLS, and variation at the region-within-individual level: both of these sources of variation are part of the residual variance, ϵ. Hence, by generating PK parameters from PuMBA, the estimation inaccuracy resulting from the use of NLS is already included in the generated PK parameters For this reason, simulating TAC data using the simulated PuMBA PK parameters, and subsequently estimating PK parameters from these TACs using NLS, would effectively amount to doubling the influence of estimation error in the resulting PK parameters estimated from the simulated data. It would be present both in the estimated ϵ matrix from the PuMBA model fitted to the empirical data, as well as from the estimation of the PK parameters from the generated TACs themselves. Removal of the ϵ term, however, implies the removal of both biological region-within-individual variance as well as estimation error, which is also not appropriate. For this reason, for SRTM and 1TC we estimated only parameter data from the PuMBA parameters and not full TACs generated from these parameters. To model these data, PuMBA, LME and t tests were applied to the simulated parameter estimates, using 250 simulated datasets for each condition.

We made use of data from two radioligands for each model: for the one-tissue compartment (1TC) model, we used [11C]DASB and [11C]GR103545; and for the simplified reference tissue model (SRTM), we used [11C]WAY100635 and [11C]DASB with cerebellar white and grey matter, respectively, as reference region corresponding with previous recommendations [16, 31, 32, 34]. The simulation parameters are described in Additional file 1: S2. The binding parameter used for each model in which group differences were simulated and estimated, were BP_{ND} for SRTM and V_T for the 1TC model.

Data and code availability statement

The R code used to apply this method is provided in an open repository (https://github.com/mathesong/PuMBA_Materials), including sample simulated datasets. The measured data used for generating the simulation parameters were drawn from previous studies [9, 10, 26, 29, 32].

Results

The results of the simulations are presented below. The metrics by which the methods were evaluated and compared are summarised in Fig. 3.

TAC simulations

For the TAC simulations, we compare the performance of PuMBA to that of LME models applied to the estimated BP_P values. For comparison, we also applied LME to the "true" BP_P values from which the TACs were simulated, i.e. representing the "ideal" case
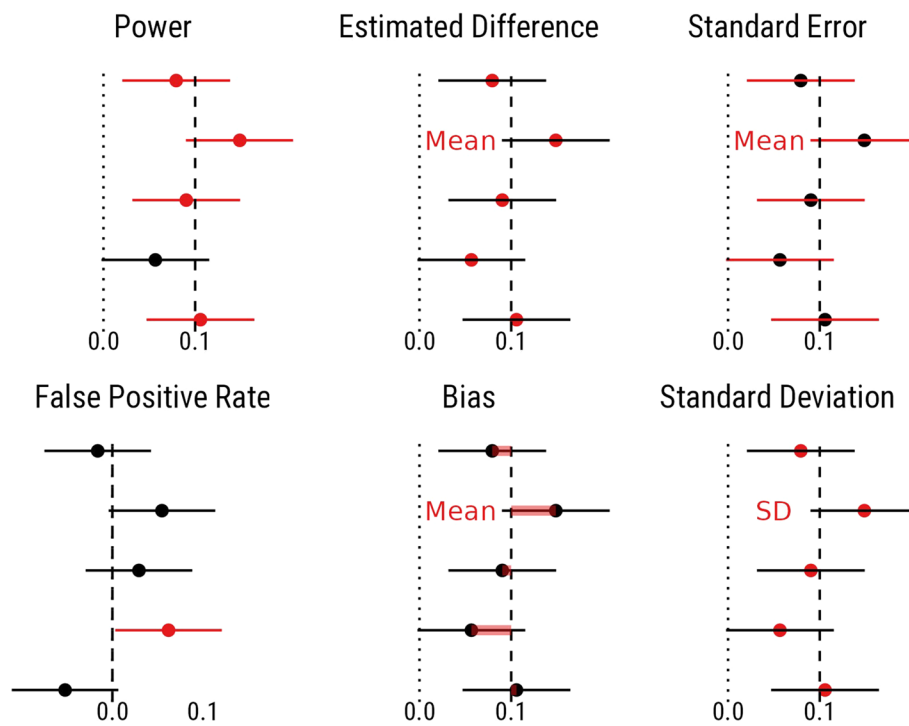
Fig. 3 Summary of the different metrics for evaluating the methods. In each case, the model estimates are depicted for each simulated dataset along the y axis, with the posterior mean and its uncertainty represented by the point and the error bars, respectively, represented on the x axis. Dotted lines represent zero, and dashed lines represent the true values. Red is used to highlight the focus of each metric, of which the mean or standard deviation (SD) is assessed, depicted by red text

in which binding parameters are perfectly estimated, incorporating only individual-level and regions-within-individual-level variation. The results of these simulations are shown in Fig. 4. Naturally, we see that the power of both LME and PuMBA is decreased with increase in measurement error, although these decreases are of less consequence for LME compared to PuMBA. Although concerns have previously been raised about the accuracy of direct quantification of BP_P [35], i.e. without the use of a reference region, we observe that LME applied to the estimated BP_P values with the original measurement error exhibits similar or only marginally reduced power compared to when LME is applied to the true values. This supports the use of BP_P as a sufficiently good index of specific binding for these two tracers.

In all cases, the power of PuMBA exceeds that of the LME model for the same degree of measurement error. In most circumstances, the power of PuMBA even exceeds that of applying LME to the "true" BP_P values, i.e. assuming perfect quantification. PuMBA analysis yielded lower standard error (Fig. 4B) as well as lower standard deviation between simulated datasets (Fig. 4C and Additional file 1: S3) of the estimated group differences between simulations relative to LME. These imply, respectively, that PuMBA estimates exhibit greater precision, or certainty, of the magnitude of the group differences; and that PuMBA estimates are more consistent between simulated samples, i.e. differ less from sample to sample. Decreases in the standard error of estimates of group differences without correspondingly large decreases in the standard deviation of these
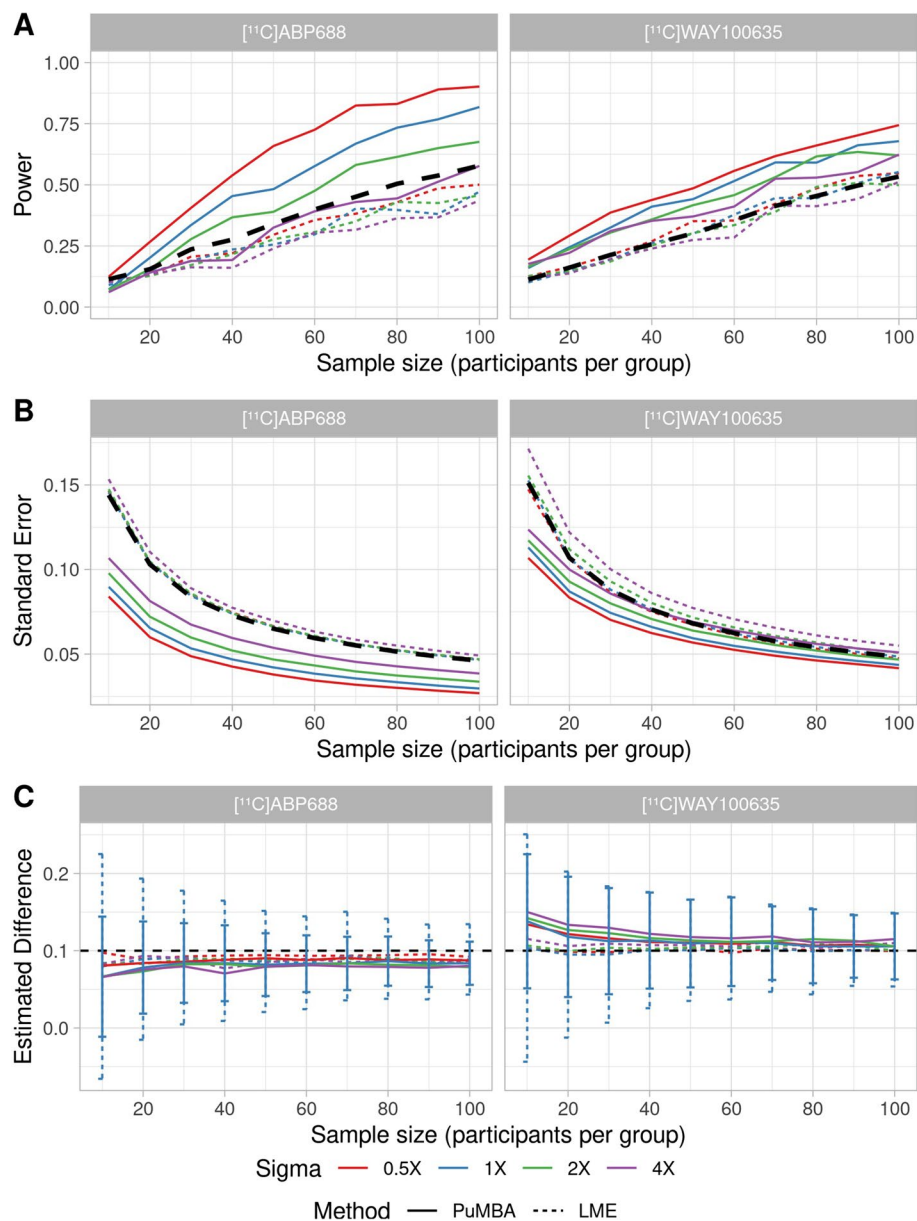
Fig. 4 TACs were simulated with the standard deviation of measurement error equal to the same, half, double or quadruple the estimated measurement error in the original data, represented with colours and modelled with LME or with PuMBA. **A** Power as a function of sample size for a 0.1 difference between patients and controls. The black dashed line represents the power of the LME analysis applied to the true values of BP_P from which the simulations were generated. **B** Mean standard error of the estimated group differences across the simulated datasets. The black dashed line represents the mean standard error of the LME analysis applied to the true values of BP_P. **C** Mean estimated group differences across simulated datasets, with the true difference of 0.1 shown with a dashed black line. The error bars represent the standard deviation of estimated group differences across simulations, shown only for the original measurement error

estimates across samples would result in an increase in the false positive rate in the presence of no group differences. However, we see no evidence for any increase in the false positive rate when PuMBA is applied to data simulated with no group differences: in

fact PuMBA exhibits a lower false positive rate on average for both tracers, and for every level of measurement error (Additional file 1: S3).

Both LME and PuMBA exhibit a small degree of bias in the estimated group differences (Fig. 4C) presumably owing to inaccuracies in the PK parameters estimated using the NLS model fitting procedure, i.e. prior to the parameters being entered into the statistical models. For $[^{11}C]$ABP688, there was a tendency to underestimate group differences, while for $[^{11}C]$WAY100635 there was a tendency to overestimate group differences. In all cases, this bias was greater for PuMBA than for LME, suggesting that PuMBA exacerbates this bias. The bias was also more pronounced both in cases of smaller sample sizes and greater measurement error. For $[^{11}C]$ABP688 with the original measurement error, mean estimates across the simulated datasets of the true difference of 0.1 were 0.066 and 0.080 for PuMBA and LME, respectively, for $n = 10$; while for $n = 100$ the mean estimates were 0.084 and 0.089. For $[^{11}C]$WAY100635 with the original measurement error, the mean estimates were 0.138 and 0.104 for PuMBA and LME, respectively, for $n = 10$, while for $n = 100$ they were 0.105 and 0.101, respectively. In all cases however, the bias of the estimates was smaller than the sample-to-sample variation: the bias of the group difference estimates relative to the true value was never greater than 65% of the standard deviation of these estimates across simulated datasets (median: 41% for $[^{11}C]$ABP688 and 23% for $[^{11}C]$ WAY100635; see Additional file 1: S3). This implies that the combined effects of sampling error and measurement error are of greater consequence for the estimation of group differences compared to the bias in the PuMBA estimate for any given sample.

Comparison with SiMBA

In order to directly compare the performance of PuMBA with SiMBA using the same priors, we applied both models to the same simulated datasets as described in our previous report [23]. Furthermore, in order to maximise the comparability of the outcomes, we estimated the same binding parameter, BP_{ND} (as opposed to BP_P as before) using both PuMBA and SiMBA with exactly the same priors on all shared parameters. Firstly, MCMC sampling time is much more rapid for PuMBA compared to SiMBA (Fig. 5A), showing that fitting a PuMBA model is completed approximately 4000 times more quickly compared to SiMBA for the same number of iterations. We show that SiMBA outperforms PuMBA, with greater power (Fig. 5B), lower standard error (Fig. 5C), lower standard deviations of estimated group differences (Fig. 5D) and a smaller degree of bias (Additional file 1: S4). Lastly, while PuMBA and SiMBA both outperform univariate LME analysis of the true values of BP_{ND} underlying the simulations in terms of power, standard error and standard deviation across simulations, they do not outperform a PuMBA model fit to the true values of all four of the PK parameters. In other words, multivariate (i.e. PuMBA or SiMBA) analysis of all noisy estimated parameters together is more efficient than univariate (i.e. LME) analysis of noiseless parameter values for the parameter of interest; but multivariate analysis of all noiseless parameters together shows the greatest efficiency of all.
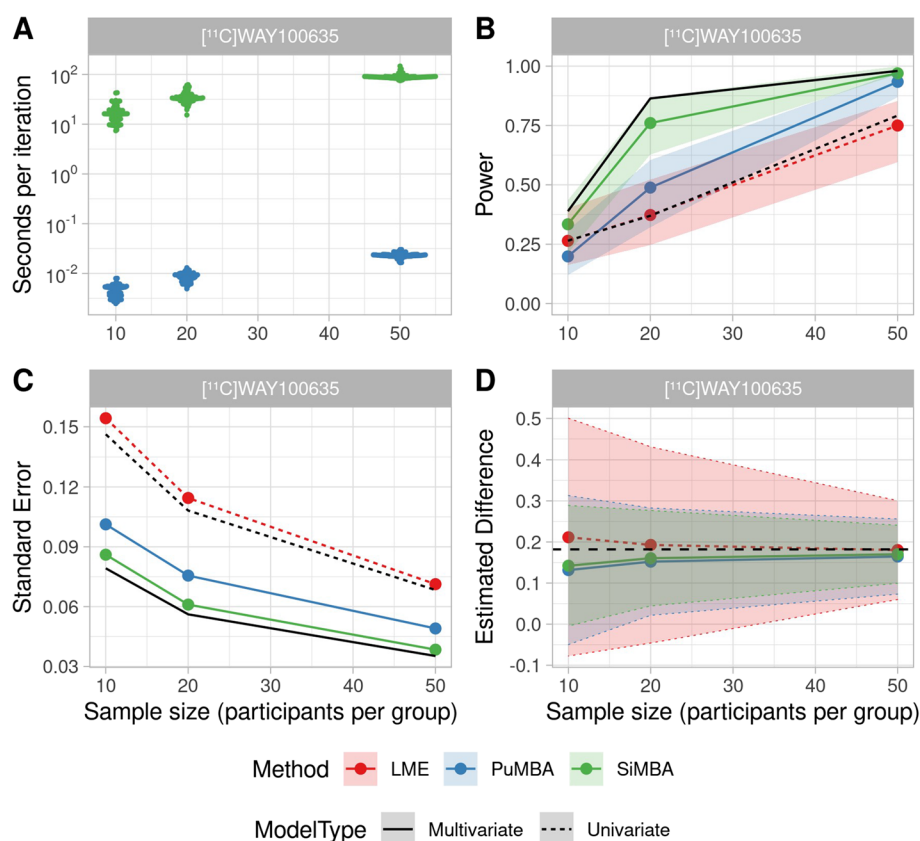
Fig. 5 PuMBA and SiMBA were applied to the same simulated datasets in each condition, with identical priors on all shared parameters to compare their performance. Note that the measurement error sigma is equal to approximately double the original measurement error [23]. Black lines refer to the performance of these models applied to the true values underlying the simulations: univariate LME applied to the binding parameter, and multivariate PuMBA analysis applied to the true values of all the pharmacokinetic parameters. **A** MCMC sampling times for PuMBA and SiMBA per iteration. **B** Power as a function of sample size for a 0.182 (20%) difference between patients and controls. The shaded area represents the upper and lower bounds of the 95% confidence interval obtained using bootstrap resampling. **C** Standard error of group difference estimates as a function of sample size. **D** Mean estimated group differences across simulated datasets, with the true difference of 0.182 shown with a dashed black line. The shaded area represents the standard deviation of estimated group differences across simulations for each approach, with their individual boundaries emphasised with dotted lines

Correlation matrix recovery

Since PuMBA exploits the correlations between PK parameters, it is important to consider how well these correlations are estimated and, as a result, to what extent the bias or variance in these estimates affect the quality of PuMBA inferences, especially the false negative rate. To this end, we extracted the estimated correlation coefficients from the TAC simulations described in "TAC simulations" section to compare with the corresponding matrices that were set for the simulation. We also simulated additional uncorrelated data in which the parameters were generated using multivariate distributions but with the correlations between parameters removed, i.e. with diagonal variance–covariance matrices. Using the uncorrelated simulation parameters, we simulated both TAC data and parameter data to examine the influence of PK parameter estimation from TACs. More details are provided in Additional file 1: S5.

For simulated uncorrelated parameter data, estimated correlations were centred around zero for all parameter pairs, demonstrating that PuMBA itself is capable of recovering the parameter intercorrelations accurately in the absence of any bias introduced during PK parameter estimation from TACs. However in simulated TAC data, the recovery of the parameter intercorrelations was reasonably poor for most pairs of PK parameters in the individual (τ) correlation matrices, both for the correlated and uncorrelated data, in contrast to SiMBA [23]. Together, these results imply that the poor recovery of the true parameter correlations is primarily due to bias in the estimation of the PK parameters from TACs using NLS.

Despite the poor recovery of parameter intercorrelations, application of PuMBA to uncorrelated data resulted in higher mean standard error and standard deviation of group difference estimates, and reductions in statistical power relative to the original correlated data. In the correlated simulations, estimated correlation coefficients were closer to the true simulated values for $[^{11}C]$ABP688 than for $[^{11}C]$WAY100635, which may explain the greater improvements performance observed with PuMBA relative to LME with $[^{11}C]$ABP688. Together, these results suggest that the power and precision of PuMBA estimates are influenced by the true underlying parameter intercorrelations, and are likely improved when these intercorrelations are more accurately estimated.

PuMBA therefore exploits parameter intercorrelations to improve its inferences, yet these correlations tend to be estimated relatively poorly. We were concerned that this might imply that, in the absence of true correlations, the artefactual correlations arising from parameter estimation bias might yield a higher risk of false positive conclusions. However, for both $[^{11}C]$ABP688 and $[^{11}C]$WAY100635 we observed no apparent increase in the false positive rate in either of the uncorrelated datasets relative to either PuMBA applied to the correlated datasets or to LME (Additional file 1: S5).

Prior sensitivity

Next, we wanted to assess the sensitivity of the PuMBA model performance to ill-defined priors. We examined the effect of prior misspecification on simulated $[^{11}C]$ABP688 data, for which PuMBA yielded the greatest improvements relative to LME in order to examine the extent to which this improvement would be diminished. Priors are defined over many parameters, but the only *informative* priors are defined over the global intercepts, as opposed to the other priors which merely regularise estimates towards zero. For this reason, we randomly halved or doubled all four of the global intercept priors in each simulation by adding or subtracting 0.69 to the prior mean (in its natural logarithm), where the randomisation was for each parameter within each simulated dataset, i.e. K_1 might be doubled while V_{ND} might be halved in the prior specification for one dataset. These misspecifications appear to make little-to-no difference to the performance of the model across all of the evaluation metrics (Additional file 1: S6). The only apparent change is that the SD of the group difference estimates appears to be slightly elevated in small sample sizes, which is to be expected, since the influence of priors will be greater when the influence of the likelihood is weaker, i.e. when the data itself are less informative.
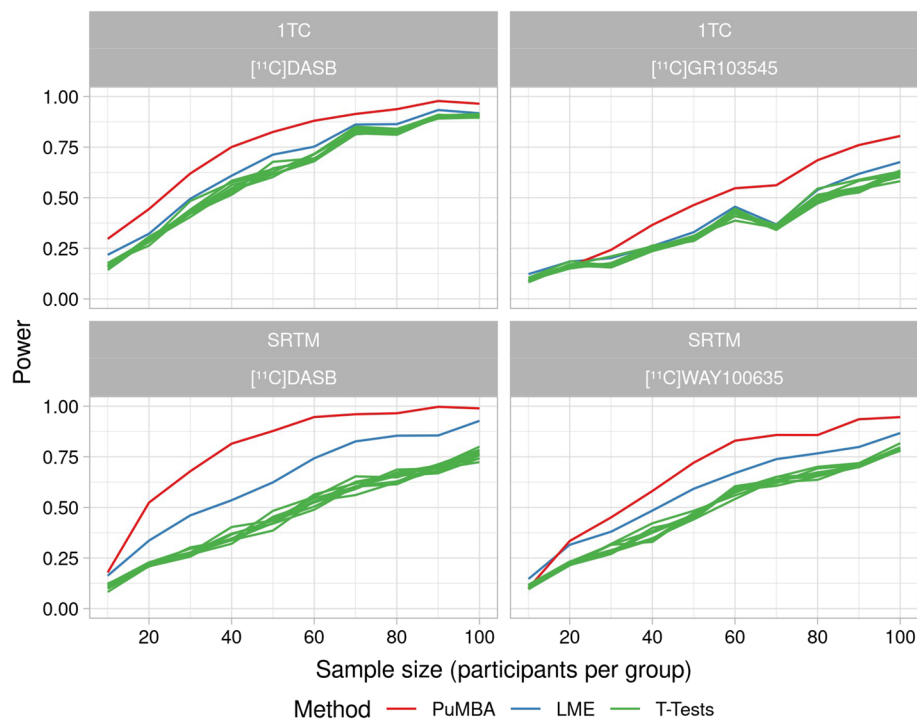
Fig. 6 Power as a function of sample size for a 0.1 difference between patients and controls. The models are the one-tissue compartment model (1TC) and the simplified reference tissue model (SRTM)

Parameter simulations

In order to test whether PuMBA can also be applied for other kinetic models which cannot yet be modelled using SiMBA, we performed additional parameter-only simulations using the 1TC and SRTM, as described in "Parameter simulations" Section. The results of these simulations are shown in Fig. 6, in which we observe increases in power for linear mixed effects (LME) modelling relative to t tests, and improved power for PuMBA relative to LME. We observe larger improvements in power for SRTM, while improvements in power for the 1TC were more modest. In all cases, these improvements in power were not associated with any increase in false positive rate (Additional file 1: S7).

When examining the estimated group differences, we observed lower standard error as well as lower standard deviation of the estimated group differences between simulated datasets for LME relative to t tests, as well as for PuMBA relative to LME. While we observed no bias in the mean estimated group differences for t tests or LME, PuMBA showed slightly biased estimates in all cases, with more bias in smaller sample sizes (Additional file 1: S7).

Application in real data

Lastly, in order to demonstrate the application of PuMBA in real data, we applied PuMBA to study patient–control differences in an empirical dataset. For this purpose, we examined $[^{11}C]$GR103545 data, consisting of 13 healthy controls and 10 patients with major depressive disorder (MDD) (described in more detail in [26]), with TACs from the nine regions used in the simulations.

Fitting the 2TC to this data using NLS resulted in good fits to the data, but highly variable parameter estimates. For instance, we would not expect V_{ND} to vary greatly between individuals and regions. For V_{ND} the median was 0.51, while the 5% and 95% quantiles were equal to 4% and 3900% of this figure, respectively. To assess identifiability, we examined the condition number of the variance–covariance matrix (i.e. the ratio of the maximum to the minimum eigenvalues of the matrix after rescaling the columns to unit Euclidean norms). Of the 207 NLS TAC fits, only 40% had a condition number lower than the threshold of 10^6 recommended by [6]. This indicates severe ill-conditioning of the resulting fits, and thereby high sensitivity to small perturbations in the data. For this reason, we conclude that the 2TC model applied to [^{11}C]GR103545 data is insufficiently identifiable. In contrast, for [^{11}C]ABP688, estimates were less variable (5% and 95% quantiles of estimated V_{ND} values were equal to 57% and 151% of the median), and 100% of the TAC fits yielded acceptable condition numbers, together implying much better identifiability of 2TC fits. Hence, for [^{11}C]GR103545, we fit the 1TC model instead, producing fits of which all produced an acceptable condition number.

For fitting PuMBA to the [^{11}C]GR103545 data, we used identical priors to those used for the analysis of the simulated [^{11}C]GR103545 data above, but also including age (centred, in decades) and sex as covariates for both $\log(K_1)$ and $\log(V_T)$, with priors defined by normal distributions centred at zero with standard deviations of 0.1. Because we do not expect there to be differences between MDD patients and controls in blood flow, we did not include patient status as a predictor for $\log(K_1)$. Estimating $\log(V_T)$ using the same parameters for both LME and PuMBA resulted in group-difference estimates with greater precision for PuMBA (SE = 0.097) compared to LME (SE = 0.132). The magnitude of patient–control difference estimates in $\log(V_T)$ were similar too, with both models yielding estimates which were within one SE of one another (LME 0.163; PuMBA 0.086)—although both models' 95% CI intervals for this parameter overlapped with zero.

Discussion

In this study, we show that a multivariate statistical analysis of all estimated PET pharmacokinetic parameters using PuMBA yields significant inferential advantages over univariate analysis of the parameter of interest performed without considering the other parameters. We also show that PuMBA can be fruitfully applied even when there is a substantial degree of measurement error. As would be expected, because PuMBA is applied to parameter estimates, it cannot outperform a similar analysis which is conducted on the full TAC data in which quantification and statistical analysis are both able to benefit from the hierarchical multivariate framework, i.e. SiMBA. However, PuMBA

Table 1 Systematic comparison of the different methods applied

	LME	PuMBA	SiMBA
Modelling and statistical analysis	2 consecutive steps	2 consecutive steps	1 step, simultaneously
Input data	PK parameters	PK parameters	TACs
Partial pooling	Univariate	Multivariate	Multivariate
Inferential paradigm	Frequentist	Bayesian	Bayesian
Computation time	Less than a second	Minutes	Hours to days

models can be fitted in minutes, in contrast to days required by SiMBA (Table 1). Furthermore, PuMBA can be directly applied to data from more PET pharmacokinetic models as a substitute for conventional statistical analysis approaches, while SiMBA requires the user to incorporate the pharmacokinetic model itself into the Bayesian model, which can be challenging.

While PuMBA can more easily be applied to a wider range of pharmacokinetic models than SiMBA, this does not necessarily imply that it can be applied when the identifiability of PK parameters for a given model and tracer is poor. For instance, the identifiability of the 2TC is reasonably good for [^{11}C]WAY100635 and [^{11}C]ABP688, in contrast with [^{11}C]GR103545 for which the 2TC is rather poorly identified [27]. On the other hand, SiMBA stabilises the fitting of the PK model itself using the hierarchical multifactor framework, and therefore has the additional advantage of improving model identifiability: this is not a property of PuMBA. Hence, while SiMBA is more generalisable in the sense that it makes the application of the more complex models possible for a wider variety of PET tracers for which they are otherwise insufficiently identifiable, PuMBA is more generalisable in the sense that it can, without any substantive modifications, be applied to data originating from a wider assortment of PK models, including reference tissue models—provided that the parameters are sufficiently identifiable.

Since PuMBA, in contrast to SiMBA, cannot improve the estimation of the PK parameters themselves from the NLS model, its function is only to take advantage of information shared between individuals, regions and parameters. Both LME and PuMBA make use of partial pooling, and both make use of the same predictors for the binding parameter. Thus, the only difference between the performance of these models is due to the multivariate partial pooling strategy applied in PuMBA, in contrast to the univariate partial pooling applied in LME. Moreover, by exploiting the relationships between PK parameters, PuMBA exhibits greater precision and power compared to a univariate LME analysis of even the *true* binding parameters from which the simulations were generated, i.e. as if binding could be estimated perfectly. This can be described with an analogy. Consider that we wish to compare binding potential values for two groups of individuals. We have the option of either measuring all of our participants in a noisy PET system and modelling their resulting TACs or of being informed of each individual's true binding potential value by an omniscient oracle. Counter-intuitively, the present results suggest that we ought to shun the oracle, and perform the PET study as planned using PuMBA, or SiMBA, for the statistical analysis. If, however, the oracle could be persuaded to provide us with the true values of *all* of the PK parameters, only then would it be advantageous to accept the oracle's aid. This corresponds to the comparisons of SiMBA and PuMBA with the solid and dashed black lines in Figs. 4A, B and 5B, C. Historically, one important focus of PET PK modelling method development has been to improve the accuracy of estimated PK parameters in order to improve their inferential efficiency. In contrast, SiMBA and PuMBA, by incorporating the otherwise irrelevant PK parameters which are conventionally discarded during univariate statistical analysis, exhibit better inferential properties compared to even what had previously been considered optimal. The present results imply that this previous floccinaucinihilipilification regarding the additional parameters has been detrimental.

It cannot be assumed that a multivariate analysis will improve the power of any given statistical comparison. With reference to Fig. 1, it is clear that with sufficiently strong correlations between parameters, and sufficiently precise estimates of the parameters and their correlations with one another, the variance of the conditional distribution of an estimated parameter can be reduced to below that of the marginal distribution of the true parameter. However, in the presence of high uncertainty, small samples or low correlation between variables, this potential gain is likely to be reduced: we see some evidence of this with $n = 10$ in Figs. 4A and 5B. We also see that PuMBA yielded greater improvements in precision and statistical power when the correlation matrices were more accurately estimated in $[^{11}C]$ABP688 compared to $[^{11}C]$WAY100635.

In all applications, PuMBA exhibited bias in its group difference estimates, particularly for smaller sample sizes and with larger measurement error, for instance, in Fig. 4C. We find that this bias was never greater than 65% of the standard deviation of sample-to-sample estimate variability, which implies that even in the worst case scenario, the estimated differences for $[^{11}C]$ABP688 will still be greater than the true value approximately 25% of the time given infinite replications of the same study. Although the magnitude of this bias is not large, it is still important to understand its source. One factor is that PuMBA makes use of a regularising prior for the estimation of group differences, which makes the model sceptical of large differences between groups, effectively shrinking the estimated group differences towards zero. However, for $[^{11}C]$WAY100635, we observe that the bias is positive, which is incompatible with shrinkage being solely responsible for the observed bias. This corresponds with the observation that the LME group differences are also biased in the same direction as the PuMBA estimates in the TAC simulations, but to a smaller degree in each case. This suggests that the bias is also partially explained by identifiability issues in the PK parameter estimation from the TAC data using NLS, and that bias in the estimation of the binding parameter is also accompanied by bias in the estimation of the other model parameters in a systematic fashion. This corresponds with the results of the uncorrelated data simulations, in which correlations were partially induced between parameters by the PK parameter estimation from TACs. It is noteworthy, however, that PuMBA exacerbates this bias, presumably owing to its estimation of the correlations between the PK parameter estimates with potentially induced correlations. Finally, for 1TC parameter simulations of $[^{11}C]$DASB, there is positive bias in small sample sizes, which cannot be explained by either of the above reasons. This suggests that the identifiability of the PuMBA model itself may also play a role in the observed bias, particularly in small sample sizes. For these reasons, the precise magnitude of PuMBA estimates should be interpreted with some degree of care particularly in applications for which the identifiability of the individual parameters is poor, or in small sample sizes.

While PuMBA is much faster to estimate and easier to implement compared to SiMBA, it does still require additional effort from the analyst compared to conventional approaches. Firstly, we would advise paying more attention to the estimated parameters generated by the NLS procedures prior to analysis: even when estimates of binding potential, for example, may appear reasonable, estimates for some of the other parameters could be problematic. To this end, we recommend estimating the parameters for the PuMBA analysis (i.e. V_{ND} and BP_P instead of k_2 and k_3 in the context of the 2TC) directly

in the NLS procedure, and using reasonably conservative upper and lower limits, coupled with estimating the fit with multiple starting points [30]. In this way, overall fits are often similar, but we observed that we less frequently obtained biologically inconsistent values for the combined parameters. We thereby minimise the chances of our NLS model yielding parameter estimates obtained from a local minimum. We also advise examining distributions of parameters as well as assessing the condition number of the variance–covariance matrices, after standardisation of the Euclidean norms of the columns [2], of NLS fits to assess identifiability. For the condition number, [6] recommends a threshold of 10_6. If V_{ND} or BP_P, for instance, show orders of magnitude more variation than would be expected biologically, then this may either call for a simpler model (such as the 1TC used in "Application in real data" Section), or perhaps PuMBA may not be appropriate for the particular application. Another additional requirement of PuMBA compared to conventional approaches is that, because all parameters are included in the PuMBA model, the analyst must define linear model specifications for each of them. Age, for instance, might affect blood delivery, or the kinetics of the radioligand in a reference tissue, even if it does not affect the binding potential for a particular target. Analysts should also carefully consider whether the examined condition itself might also affect any of the other parameters. For instance, while most psychiatric conditions would probably not be expected to affect blood flow, there may be neurological conditions which would be expected affect other parameters. Furthermore, because PuMBA is a Bayesian model, prior distributions need to be defined not only for all covariates, but for all parameters in the model. Using PuMBA therefore requires more care and consideration than conventional approaches in order to be applied most effectively, and we recommend collaboration between researchers with domain expertise and technical expertise. This allows for the incorporation what is already known about the relevant clinical and biological constraints into the specification and priors of the model for a given research question. While we did not observe a large effect of prior misspecification on the results, this is likely caused by the fact that none of our prior definitions were particularly informative, and because none of these targets exhibit a particularly high degree of heterogeneity across individuals such as observed in tau PET imaging, nor does this heterogeneity follow non-normal distributions. Domain expertise can be exploited for, among other things, testing more specific hypotheses using highly informative priors, making use of other statistical distributions, allowing different groups to express different covariance patterns, or for incorporating patient-group effects in other PK parameters other than binding if justified by the relevant biology. From a technical perspective, a Bayesian workflow [4, 13] incorporating model comparison [12, 24, 38] and posterior predictive distribution visualisation [4, 11] can also be instructive for optimally defining complex models as a complement domain expertise.

In order to facilitate the application and extension of this methodology in novel settings by other researchers, we have provided a repository including simulated datasets as well as analysis notebooks including code and outputs to demonstrate the application of this approach for each of the PK models described above. All software is freely available, easily downloaded and open-source, and the provided code can easily be used as a template from which to modify this approach for use in new applications. While we have demonstrate the use of PuMBA for four PET radioligands, and for three different PK

Table 2 Blood flow and binding parameters were positively correlated across all radiotracers and PK models in their inter-individual correlation matrices estimated by PuMBA

Model	Radiotracer	Estimated correlation
2TC	[^{11}C]ABP688	0.89
2TC	[^{11}C]WAY100635	0.81
1TC	[^{11}C]GR103545	0.53
1TC	[^{11}C]DASB	0.39
SRTM	[^{11}C]DASB	0.64
SRTM	[^{11}C]WAY100635	0.54

models, we cannot guarantee that PuMBA can be fruitfully applied to any novel application in PET. Prior sensitivity analysis and posterior predictive checking are helpful for more closely examining the validity of inferences obtained using PuMBA, and Bayesian modelling in general.

One interesting observation which emerges from the fitting of all of the datasets is the fact that inter-individual variation (i.e. τ) in blood delivery (K_1 and R_1) and binding (BP_P, V_T and BP_{ND}) were positively correlated with one another across all models and tracers, using both PuMBA and SiMBA (Table 2 and Additional file 1: S2). Importantly, these correlations refer to the correlation of partially pooled estimates (i.e. random effects) of the PK parameters at the individual level, and not to the parameters estimated for each individual TAC. While PuMBA may be more sensitive to issues of poor identifiability when fitting the PK model, SiMBA ought to be more robust to this possibility. The exact meaning of this observation is beyond the scope of the current investigation, and perhaps cannot be understood using pharmacokinetic modelling alone. However, this raises questions about how independent the parameters of these models are, or ought to be between individuals; as well as whether this has implications for how these parameters can be understood.

Potential avenues for future research include examining the similarity of estimated parameter intercorrelations between different datasets collected by different groups, assessing the factors which contribute to the differential performance of PuMBA in different settings, as well as whether PuMBA could be applied simultaneously to data collected using multiple radioligands in the same individuals to estimate and take advantage of similarities in blood delivery or in target protein levels within individuals. Another promising direction for future work might be to incorporate the spatial distributions of these parameters throughout the brain or along the cortical surface [14] with parametric imaging.

In conclusion, PuMBA allows researchers to make more precise and powerful statistical inferences, and thereby extract more information from a given dataset without needing to collect any additional information. PuMBA takes on the order of minutes to fit on a single computer core and can more easily be applied to a wider range of pharmacokinetic models than SiMBA. PuMBA may therefore serve as a convenient

intermediate substitute for a full SiMBA analysis, and a useful tool for statistical analysis of PET data in general.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40658-023-00537-8.

> **Additional file 1**. Supplementary Materials SX.

Declarations

Ethics approval and consent to participate
This study did not involve collection of any new data. Ethical approval was granted for all original studies. Informed consent was obtained from all individual participants included in the original studies.

Competing interests
The authors have no relevant financial or non-financial interests to disclose.

References
1. Bates D, Mächler M, Bolker B, et al. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48. https://doi.org/10.18637/jss.v067.i01.
2. Belsley DA. Conditioning diagnostics: collinearity and weak data in regression, vol. 262. Wiley series in probability and statistics. Wiley-Interscience; 1991.
3. Betancourt M. Hierarchical modeling. 2020a; Retrieved from https://github.com/betanalpha/knitr_case_studies, commit 27c1d260e9ceca710465dc3b02f59f59b729ca43.
4. Betancourt M. Towards a principled bayesian workflow (RStan). 2020b; Retrieved from https://github.com/betanalpha/knitr_case_studies, commit aeab31509b8e37ff05b0828f87a3018b1799b401.
5. Betancourt M. Factor modeling. 2021; Retrieved from https://github.com/betanalpha/ knitr_case_studies, commit 6e4566309163ee79f8b7c907e2efce969a96bc54.
6. Bonate PL. Nonlinear models and regression. In: Bonate PL, editor. Pharmacokinetic-pharmacodynamic modeling and simulation. Boston: Springer; 2011. p. 101–30. https://doi.org/10.1007/978-1-4419-9485-1_3.
7. Bürkner PC. Brms: an r package for bayesian multilevel models using stan. J Stat Softw. 2017. https://doi.org/10.18637/jss.v080.i01.
8. Carpenter B, Gelman A, Hoffman MD, et al. Stan: a probabilistic programming language. J Stat Softw. 2017. https://doi.org/10.18637/jss.v076.i01.
9. Chen Y, Goldsmith J, Ogden RT. Nonlinear mixed-effects models for PET data. IEEE Trans Biomed Eng. 2019;66(3):881–91. https://doi.org/10.1109/TBME.2018.2861875.
10. DeLorenzo C, Kumar JSD, Mann JJ, et al. In vivo variation in metabotropic glutamate receptor subtype 5 binding using positron emission tomography and [11C]ABP688. J Cereb Blood Flow Metab. 2011;31(11):2169–80. https://doi.org/10.1038/jcbfm.2011.105.
11. Gabry J, Simpson D, Vehtari A, et al. Visualization in Bayesian workflow. J R Stat Soc A Stat Soc. 2019;182(2):389–402. https://doi.org/10.1111/rssa.12378.
12. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. Stat Comput. 2014;24(6):997–1016. https://doi.org/10.1007/s11222-013-9416-2.
13. Gelman A, Vehtari A, Simpson D, et al. Bayesian Workflow. 2020. arXiv:2011.01808 [stat], arXiv: 2011.01808.

14. Greve DN, Svarer C, Fisher PM, et al. Cortical surface-based analysis reduces bias and variance in kinetic modeling of brain PET data. NeuroImage. 2014;92:225–236. https://doi.org/10.1016/j.neuroimage.2013.12.021, , arXiv: NIHMS 150003 ISBN: 1095-9572 (Electronic) r1053-8119 (Linking) Publisher: Elsevier Inc. tex.arxivid: NIHMS150003.

15. Gunn RN, Gunn SR, Cunningham VJ. Positron emission tomography compartmental models. J Cereb Blood Flow Metab. 2001;21(6):635–52. https://doi.org/10.1097/00004647-200106000-00002.

16. Hirvonen J, Kajander J, Allonen T, et al. Measurement of serotonin 5-HT1A receptor binding using positron emission tomography and [carbonyl-(11)C]WAY-100635-considerations on the validity of cerebellum as a reference region. J Cereb Blood Flow Metab Off J Int Soc Cereb Blood Flow Metab. 2007;27(1):185–95. https://doi.org/10.1038/sj.jcbfm.9600326.

17. Innis RB, Cunningham VJ, Delforge J, et al. Consensus nomenclature for in vivo imaging of reversibly binding radioligands. J Cereb Blood Flow Metab. 2007;27(9):1533–9. https://doi.org/10.1038/sj.jcbfm.9600493.

18. Knudsen GM, Jensen PS, Erritzoe D, et al. The center for integrated molecular brain imaging (cimbi) database. NeuroImage. 2016;124:1213–9. https://doi.org/10.1016/j.neuroimage.2015.04.025.

19. Knudsen GM, Ganz M, Appelhoff S, et al. Guidelines for the content and format of pet brain data in publications and archives: a consensus paper. J Cereb Blood Flow Metab. 2020;40(8):1576–85. https://doi.org/10.1177/0271678X20905433.

20. Lammertsma AA, Hume SP. Simplified reference tissue model for PET receptor studies. Neuroimage. 1996;4(3):153–8. https://doi.org/10.1006/nimg.1996.0066.

21. Lewandowski D, Kurowicka D, Joe H. Generating random correlation matrices based on vines and extended onion method. J Multivar Anal. 2009;100(9):1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008.

22. Matheson GJ. Kinfitr: reproducible PET pharmacokinetic modelling in R. Bioinformatics. 2019. https://doi.org/10.1101/755751.

23. Matheson GJ, Ogden RT. Simultaneous multifactor bayesian analysis (SiMBA) of PET time activity curve data. NeuroImage. 2022. https://doi.org/10.1016/j.neuroimage.2022.119195.

24. McElreath R. Statistical rethinking: a bayesian course with examples in r and stan. Boca Raton: CRC Press; 2016.

25. McElreath R. Multilevel regression as default. 2017.

26. Miller JM, Zanderigo F, Purushothaman PD, et al. Kappa opioid receptor binding in major depression: a pilot study. Synapse. 2018;72(9): e22042. https://doi.org/10.1002/syn.22042.

27. Naganawa M, Jacobsen LK, Zheng MQ, et al. Evaluation of the agonist PET radioligand [11C]GR103545 to image kappa opioid receptor in humans: kinetic model selection, test–retest reproducibility and receptor occupancy by the antagonist PF-04455242. Neuroimage. 2014;99:69–79. https://doi.org/10.1016/j.neuroimage.2014.05.033.

28. Norgaard M, Matheson GJ, Hansen HD, et al. Pet-bids, an extension to the brain imaging data structure for positron emission tomography. 2021. bioRxiv https://doi.org/10.1101/2021.06.16.448390.

29. Ogden RT, Ojha A, Erlandsson K, et al. In vivo quantification of serotonin transporters using [(11)C]DASB and positron emission tomography in humans: modeling considerations. J Cereb Blood Flow Metab Off J Int Soc Cereb Blood Flow Metab. 2007;27(1):205–17. https://doi.org/10.1038/sj.jcbfm.9600391.

30. Padfield D, Matheson GJ. Nls.multstart: robust non-linear regression using AIC scores. R package version 1.0.0. 2018.

31. Parsey RV, Arango V, Olvet DM, et al. Regional heterogeneity of 5-HT 1A receptors in human cerebellum as assessed by positron emission tomography. J Cereb Blood Flow Metab. 2005;25(7):785–93. https://doi.org/10.1038/sj.jcbfm.9600072.

32. Parsey RV, Kent JM, Oquendo MA, et al. Acute occupancy of brain serotonin transporter by sertraline as measured by [11C]DASB and positron emission tomography. Biol Psychiat. 2006;59(9):821–8. https://doi.org/10.1016/j.biopsych.2005.08.010.

33. R Core Team. R: A language and environment for statistical computing. https://www.r-project.org/, tex.address: Vienna, Austria tex.institution: R Foundation for Statistical Computing. 2022.

34. Shrestha S, Hirvonen J, Hines CS, et al. Serotonin-1A receptors in major depression quantified using PET: controversies, confounds, and recommendations. Neuroimage. 2012;59(4):3243–51. https://doi.org/10.1016/j.neuroimage.2011.11.029.

35. Slifstein M, Laruelle M. Models and methods for derivation of in vivo neuroreceptor parameters with PET and SPECT reversible radiotracers. Nuclear Med Biol. 2001;28(5):595–608. https://doi.org/10.1016/S0969-8051(01)00214-1.

36. Stone CJ, Hansen MH, Kooperberg C, et al. Polynomial splines and their tensor products in extended linear modeling. Ann Stat. 1997;25(4):1371–425.

37. Tjerkaski J, Cervenka S, Farde L, et al. Kinfitr: an open source tool for reproducible PET modelling: Validation and evaluation of test-retest reliability. 2020. https://doi.org/10.1101/2020.02.20.957738.

38. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput. 2017;27(5):1413–32. https://doi.org/10.1007/s11222-016-9696-4.

39. Wu Y, Carson RE. Noise reduction in the simplified reference tissue model for neuroreceptor functional imaging. J Cereb Blood Flow Metab. 2002;22(12):1440–52. https://doi.org/10.1097/01.WCB.0000033967.83623.34.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.