Psychopharmacology
DOI 10.1007/s00213-015-3968-0

REVIEW

The role of machine learning in neuroimaging for drug discovery and development

Orla M. Doyle[1] · Mitul A. Mehta[1] · Michael J. Brammer[1]

Received: 25 February 2015 / Accepted: 11 May 2015

Abstract Neuroimaging has been identified as a potentially powerful probe for the in vivo study of drug effects on the brain with utility across several phases of drug development spanning preclinical and clinical investigations. Specifically, neuroimaging can provide insight into drug penetration and distribution, target engagement, pharmacodynamics, mechanistic action and potential indicators of clinical efficacy. In this review, we focus on machine learning approaches for neuroimaging which enable us to make predictions at the individual level based on the distributed effects across the whole brain. Crucially, these approaches can be trained on data from one study and applied to an independent study and, unlike group-level statistics, can be readily use to assess the generalisability to unseen data. In this review, we present examples and suggestions for how machine learning could help answer fundamental questions spanning the drug discovery pipeline: (1) Who should I recruit for this study? (2) What should I measure and when should I measure it? (3) How does the pharmacological agent behave using an experimental medicine model?, and (4) How does a compound differ from and/or resemble existing compounds? Specifically, we present studies from the literature and we suggest areas for the focus of future development. Further refinement and tailoring of machine learning techniques may help realise their tremendous potential for drug discovery and drug validation.

✉ Orla M. Doyle
orla.doyle@kcl.ac.uk

[1] Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF, UK

Keywords Machine learning · Probabilistic models · Drug discovery · Neuroimaging · Personalised medicine · Stratification · Experimental medicine models

Introduction

Advancing technological developments have greatly increased the amount of information that can be ascertained about the effect of neuropharmacologically active compounds. This information can be highly disparate and collected at various stages of central nervous system (CNS) drug development. However, despite these technological advances, many of the new drugs treatments that have come to the market have been derived from the pharmacology of well-established targets with development in this case focused on increased potency, tolerability and ease of administration. Drug development for CNS disorders, and psychiatric disorders in particular, is limited by the lack of fundamental understanding of the pathophysiology of these disorders (Insel 2012), the use of simple diagnostic labels which represent heterogeneous symptom profiles (Kapur et al. 2012), the lack of accessible surrogate endpoints for therapeutic response, and the difficulty in translation from preclinical studies to clinical phases (Mak et al. 2014). This had led to an increased interest in objective biological measures (biomarkers) of the effects of pharmacological treatments. Neuroimaging has been identified as a potentially powerful probe for studying pharmacological effects on the brain with utility across several phases of drug development spanning preclinical and clinical (Borsook et al. 2006; Medhi et al. 2014; Wise and Preston 2010; Wong et al. 2009). Specifically, neuroimaging can provide insight into drug penetration and distribution, target engagement, pharmacodynamics, mechanistic action and potentially, proof of clinical efficacy (Wong et al. 2009). Magnetic resonance

Published online: 28 May 2015

imaging (MRI)-based techniques hold great potential offering high spatial and temporal resolution in a non-invasive manner and obviating the need for radiolabeled ligands for the specific target of interest. However, several operational and analytical issues need to be considered and clarified before realising the potential of neuroimaging for drug development (Borsook et al. 2010; Schwarz et al. 2011a, b; Wise and Preston 2010).

In this review, we focus on the analytical aspects of utilising neuroimaging data. Traditional mass-univariate or region-of-interest approaches perform statistical tests at each voxel and have been of great value in understanding brain function and its perturbation in the presence of pharmacological treatments, providing a straightforward method for localising brain changes at the group level (Mehta and O'Daly 2011). These studies have demonstrated sensitivity to different compounds (Honey and Bullmore 2004), and indicated that simple one-to-one mapping of drugs and brain changes are unlikely. That is, different imaging modalities may be sensitive to different aspects of drug effect. Even within a single functional MRI (fMRI) task, univariate studies have already established that different components of task performance (e.g. correct responses or errors) show different modulatory effects of drugs in the same subjects during the same scanning session (Dodds et al. 2008; Pauls et al. 2012). Integration of such effects represents an extant challenge.

An alternative approach for the analysis of neuroimaging data is provided by machine learning methods whereby predictions can be made for the individual based on the distributed pattern of effects across the whole brain, i.e. a multivariate approach at the single-subject level. Crucially, a machine learning model can be trained on data from one study and applied to an independent study. Unlike group-level statistics, we can readily use these methods to assess the generalisability of our findings to new, unseen data. Combining machine learning with neuroimaging data is well-suited to make inference on questions pertaining to drug discovery, such questions include the following: What should I measure and when should I measure it? Who should I recruit to this study? How does the pharmacological agent behave using an experimental medicine model? How does a compound differ from and/or resemble existing compounds? To answer these questions, we can build models that range from agnostic (no prior hypothesis) to models that are tailored to the question and incorporate mechanistic information (Doyle et al. 2013c). This flexibility enables the exploration of the data to answer a particular question and where appropriate incorporate prior knowledge which may uncover previously unknown associations and thus ultimately contribute to hypothesis generation (Oquendo et al. 2012).

The field of machine learning spans many paradigms but in the context of this review, we refer mainly to discriminative learning whereby the aim is to learn the mapping between a data source (e.g. neuroimaging) and a set of labels, which could be binary (e.g. placebo or drug) or real-valued (symptom scores). The primary outcome measure for these analyses is how well the discriminative model generalises to new data. To ideally assess generalisation, the model would be trained using one cohort and tested using an independent cohort. However, in neuroimaging, particularly pharmacological imaging, it is rare to have access to sufficiently similarly acquired cohorts; therefore, we employ cross-validation whereby models are often trained using a subset of the entire dataset and then tested on the remaining 'unseen' data from the dataset. This process is then iterated to achieve an estimate of the models' generalisation to unseen data. Another consideration for neuroimaging is the relationship between the number of subjects (N) and the number of features (P); it is often the case that $P \gg N$ which is an ill-posed problem. To circumvent this, we can employ sophisticated machine learning methods which incorporate regularisation (Cortes and Vapnik 1995; Rasmussen and Williams 2006). Both regularisation and cross-validation help to alleviate overfitting so that the model generalises well to new data.

In this review, we will discuss specific studies that have utilised machine learning methods with neuroimaging data. We specifically focus on studies which present methods that have potential for inclusion in drug discovery and development pathways. Several of the studies highlighted here arose from the recently completed Innovative Medicines Initiative Joint Undertaking consortium, NEWMEDS. The majority of methods can be accessed via an open-source toolbox (PIPR) developed as part of the NEWMEDS project which can be downloaded from http://www.kcl.ac.uk/ioppn/depts/neuroimaging/research/imaginganalysis/Software/PIPR.aspx.

Nomenclature

Data This is the information that we can collect from the participants, i.e. neuroimaging, genetics, cognitive test scores, symptoms, etc. The data are collected in a matrix which has dimensions $N \times P$ where N is the number of samples and P is the number of features. For neuroimaging data, a single volume could be used as a sample where all voxels are collected in a single vector. In this review, we will include examples of machine learning applied to structural MRI, fMRI and perfusion imaging. Structural MRI is a non-invasive technique for examining the anatomy of the brain. There are several methodological approaches for structural imaging which highlight different aspects of normal and abnormal brain tissue. fMRI is a collection of techniques that aim to increase our understanding of brain function. fMRI indirectly measures changes in neural activity by detecting associated changes in blood oxygenation levels in microcirculation. Perfusion imaging refers to a collection of techniques measuring blood flow and blood haemodynamic properties. A particular example is the use of arterial spin labelling (ASL) to quantify regional cerebral blood flow (rCBF).

Labels The target that we want to discriminate or predict. The labels are binary [−1, 1] for traditional classification tasks. For example, binary labels may be used to indicate the placebo and active compounds groups. The labels are real valued for regression tasks; these could represent symptom severity scores. For ordinal regression, the labels are ranking [1, 2, 3, …, *R*] where *R* is the number of labels; these could represent drug doses as low, medium and high or a more abstract representation of symptom severity as mild, moderate and severe.

Overfitting This occurs when the model is overly complex (e.g. too many parameters are used to relate the data to the label) and is not likely to generalise well to new data, i.e. the predictions for the test data will have poor accuracy.

Kernel methods Kernel functions are employed within machine learning algorithms as a means of quantifying the similarity between data samples (Shawe-Taylor and Cristianini 2004). Algorithms involving kernel methods are particularly advantageous in neuroimaging where the number of voxels greatly exceeds the number of samples (i.e. $N \ll P$) as most operations can be performed directly on the kernel representation of the data whose dimensions are determined by the number of samples, i.e. the kernel matrix is $N \times N$. Kernel functions can also perform non-linear mappings from data input space to a feature space which is usually of higher dimensions. In some cases, these non-linear mappings may provide increased predictive power. For neuroimaging data, linear kernel functions are generally preferred as the data are already very high dimensional, therefore further increasing the dimensionality is not likely to benefit the model.

Regularisation This is an important concept in ill-posed ($N \ll P$) machine learning problems. Regularisation involves incorporating a penalty for complexity, hence helping to alleviate overfitting.

Cross-validation This is utilised when a completely independent cohort are not available to test the model. Instead, the data are partitioned in training and test sets whereby the training data are used for model selection and the performance of the model on the test set is used an estimate of generalisability. Cross-validation (CV) can be implemented using two main approaches: leave-one-out CV (LOOCV) or *k*-fold CV. For LOOCV, the data from *N*-1 subjects is used for training and the data from the *N*th subject is used for testing. This process is then iterated until each subject has been used as a test case. For *k*-fold CV, the original dataset is partitioned into *k* equally sized samples. Note that the partitioning can be random or stratified (e.g. in the case of multisite data, it may be sensible to ensure that data from different scanners in spread evenly across the folds). Data from *k*-1 folds are used to train the model and data from the *k*th fold are used to test the model. The process is then iterated until each fold has been used as test data.

Support vector machine The support vector machine (SVM) is the most common approach for binary classification of neuroimaging data, i.e. the discrimination between two groups (Burges 1998). SVMs employ kernel methods to efficiently represent high-dimensional data. SVMs incorporate regularisation which helps this method to be less susceptible to overfitting. The SVM provides categorical predictions on the test data. When using a linear kernel function for neuroimaging data, the brain regions driving classification can be visualised as a multivariate map by extracting the weight vector. This aids the neurobiological interpretation of the model.

Gaussian process learning This methodology can be used to perform classification or regression based on high-dimensional data (Rasmussen and Williams 2006). Similarly to SVMs, Gaussian process learning employs kernel functions for efficient data representation. Learning is performed using a Bayesian framework which provides probabilities predictions, hence quantifying the uncertainty of the predictions. Moreover, the Bayesian framework provides an elegant solution for model selection and model comparison.

Weight vector For both SVMs and GP learning methods, the weight vector can be extracted to aid interpretation of the predictive model. A general misconception about the weight vector is that we can directly relate large weights at particular voxels to a large change in magnitude of the signal across groups. Given that these multivariate maps are sensitive to spatial correlations and variance in the data, directly linking weights with magnitude is not appropriate. Nonetheless, these maps can provide an insight into the spatial pattern of brain regions driving the predictive model. The interpretation of the weight vector is an active topic in machine learning for neuroimaging data with a potential solution for linear models presented by Haufe et al. (2014).

Who should I recruit?—personalised medicine approaches

As highlighted by Borsook et al. (2013), misclassification of study subjects (i.e. incorrect diagnosis, a mild form of the disease under investigation, challenging comorbities) is a major cause of failure in CNS trials. Therefore, it may be advantageous to stratify patients according to a particular biological and/or symptomatic aspect/s of the disease. Ideally, translating mechanistic knowledge (genetics, molecular biology, neurobiology, behaviour, etc.) may lead to finer grain diagnostic categories which could, in theory, help us to better predict treatment response and to develop more targeted interventions. To move closer to this paradigm, several challenges need to be overcome and particularly so in psychiatry as highlighted by Kapur et al. (2012): lack of biological gold standard, an overabundance of underpowered studies, the findings of which are difficult to translate and rarely replicated and disproportionate, focus on the comparison of healthy controls and prototypical patients which is a poor proxy for the clinical setting. A potentially exciting shift would be to move away from simple diagnostic labels which group patients presenting with highly heterogeneous symptom profiles and instead focus on developing and validating interventions on stratified patient groups identified by homogenous biological, cognitive and/or behavioural profiles; a particular example of this rationale in the community is the NIH-funded research domain criteria project (Insel 2014). We have illustrated a simplistic patient stratification scheme in Fig. 1 where the stratified groups could relate to subtypes of the disorder, those with varying risk for disease progression, those with varying likelihood of treatment response and so on.

The identification of homogenous groups could be a more appropriate entry point for studying the disease and treatment response. These subgroups could be mined from the available data using unsupervised learning. Unsupervised learning can be used for exploratory analyses to discover hidden groupings in the data and is conceptually similar to ‘clustering’. Latent class (LC) analysis has been used in many psychiatric studies to uncover symptomatic groupings within a disorder. Kendler et al. (1996) applied LC analysis to 14 disaggregated DSM-III-R symptoms for major depression in order to uncover different subtypes of depression. The analysis revealed seven classes, of which three represented clinically significant depressive syndromes: mild typical depression, severe typical depression and atypical depression. These findings were later replicated by Sullivan et al. (1998) who applied LC analysis to the National Comorbidity Survey data. Their analysis revealed that typical depression is clustered by severity and the atypical subtype is also clustered within a class. These latent subgroups could be a useful classification for selection of patients for trials.

It may also be advantageous to enrich a trial with patients who are likely to remain stable or likely to progress to a more pathological state during the course of the trial. Numerous machine learning approaches have been developed to predict the transition from mild cognitive impairment to Alzheimer’s disease (AD) (Cuingnet et al. 2011). In our work, we applied multivariate ordinal regression in a Gaussian process framework to baseline imaging data in order to predict the disease state at 12 months from baseline (Doyle et al. 2014). We applied this methodology to baseline structural MRI data from 1023 participants from two studies: the US-based Alzheimer’s Disease Neuroimaging Initiative (ADNI) and the European-based AddNeuroMed programme. Volumetric segmentation, cortical surface reconstruction and cortical parcellation, based on the FreeSurfer package (4.5.0, http://surfer.nmr.mgh.harvard.edu/), were used to quantify baseline cortical thicknesses and volumes of subcortical brain regions. Right and left hemisphere measures were averaged. In total, this results in 57 measures to be used as input features for ordinal regression, 34 regional cortical thickness measures and 23 regional volumes. The ADNI data were used in a *k*-fold cross-validated manner and the AddNeuroMed data were used as a completely independent test set. Areas under the receiver operator curve (AUC) of 0.75 and 0.81 were achieved for the ADNI and AddNeuroMed data, respectively. Ordinal regression was found to outperform binary classification. This study illustrates that machine learning applied to imaging data can provide accurate predictions for disease trajectories at the individual level and these models can generalise with a similar level of accuracy to a completely independent cohort.

Knowing which patients are likely to respond to a particular treatment could greatly benefit the patient’s recovery and the financial costs associated with treatment. For example, in psychiatry, treatment is usually chosen on an empirical basis informed by the clinical characteristics such as comorbidity

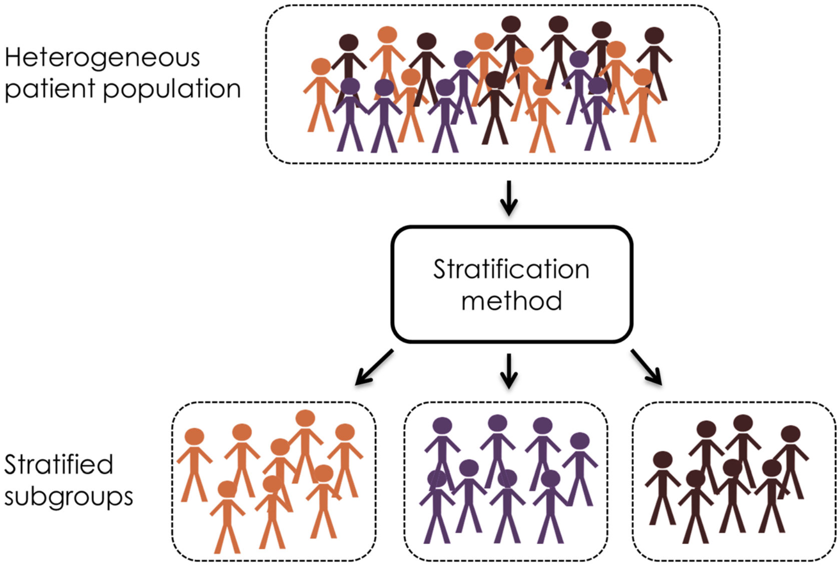

Fig. 1 Illustration of patient stratification. Patients can be stratified by likelihood of response to treatment (predictive model trained on longitudinal data); risk of disease progression (predictive model trained on longitudinal data); subtypes of the illness defined by symptoms, cognition, neurobiology, genetics, etc. (data driven). The underlying assumption is that patient stratification leads to more homogenous subtypes which is advantageous for clinical trials and drug discovery

and treatment history (Fu et al. 2013). Often, the clinical efficacy of treatment is evaluated after 6 to 12 weeks, and those who do not respond to treatment may have experienced persistent or transient symptoms throughout this period. Response prediction could ideally be used to identify non-responders that may require a different treatment at an earlier stage than is currently possible, and thus, these patients can be spared ineffective treatment and their associated side effects.

In neuroimaging to date, most machine learning studies for predicting treatment response have focused on cohorts diagnosed with psychiatric disorders. Using a sample cohort (N=18, 9 per group), Costafreda et al. (2009) applied SVMs to voxel-based morphometry data extracted from structural brain images in order to predict clinical response to antidepressant medication with an accuracy of 88.9 %. Using a similar rationale but in a larger cohort (N=46), Gong et al. (2011) used baseline structural images acquired before treatment with a single antidepressant drug to predict response. An accuracy of 69.6 % was achieved for predicting those that would later respond to treatment and those that would not. Khodayari-Rostamabad et al. (2010) acquired electroencephalography (EEG) data in patients diagnosed with schizophrenia prior to treatment with clozapine. Using an independent test cohort (N=14), they achieved 85 % accuracy in predicting response to clozapine. Hahn et al. (2015) acquired an fMRI paradigm in medication-free patients with panic disorder with agoraphobia after which cognitive behavioural therapy was conducted over several weeks. Gaussian process classification was used to predict response in 46 patients with an accuracy of 79 % on combining predictions from two aspects of the fMRI paradigm.

What data should I collect?

The data collected from participants or patients in order to study the effect of a compound or disease process is often driven by hypotheses based on the pharmacology or pathology of the illness. This can help narrow the options for data collection but nonetheless, a number of options may still be available for consideration including neuroimaging, genetics, behaviour, cognition and environment. Multimodal machine learning could be applied to data spanning these options to infer, in a data-driven manner, which modality was most sensitive to the research question. In Fig. 2, we illustrate how this could be achieved using multiple kernel learning (MKL). For MKL, data sources of widely different dimensionality can be combined using computationally cheap operations via the kernel trick, i.e. each modality is represented by a kernel whose dimensions are determined by the number of samples rather than features. A weighting across data sources can be learned within the MKL framework providing an indication of the contribution of each modality to the predictive performance. Combining data-driven frameworks such as MKL with information such as financial cost and/or tolerance of the participants to data acquisition may produce a more informed choice of marker. Similarly, removing the most expensive or difficult to acquire data and observing the cost to sensitivity would provide an important insight into how data collection could be prioritised.

When should I collect the data?

A primary consideration in neuropharmacology is whether the drug crosses the blood brain barrier and if so, what is the pharmacodynamic profile. Often, the effects of the drug in question can be hypothesised to have distributed changes across the brain or, in some cases, the mechanism of action is not fully known.

In Doyle et al. (2012), a Gaussian process framework was developed which simultaneously ‘learned’ the discrimination between the blood oxygen level-dependent (BOLD) response to a saline versus ketamine infusion and also the shape of the BOLD response to ketamine which afforded maximal discrimination. This approach could discriminate between ketamine and saline with an accuracy of 91 % and predicted that the peak BOLD response occurred at 282 s (141 volumes) after the infusion commenced, and then slowly decayed after this peak. This study illustrates how machine learning methods can be tailored to a particular question to produce a model that is highly accurate, but also provides information regarding the dynamics of the time series phMRI data, i.e. that the BOLD response to ketamine peaks on average 282 s after intravenous infusion. Moreover, we obtain a predictive probability of belonging to the ketamine class. These probabilities could help identify non-responders or related to behavioural or pharmacokinetic data.

Oxytocin plays an important role in the development of mammalian social behaviour (Donaldson and Young 2008). Experimental paradigms involving oxytocin are limited by the absence of data relating to the pharmacodynamics of oxytocin in the human brain. Consequently, Paloyelis et al. (2014) investigated the effects of the administration of intranasal oxytocin on resting-state regional cerebral blood flow in healthy male volunteers. Using arterial spin labelling, rCBF volumes were acquired in 15 min before and up to 78 min after administration of either placebo or intranasal oxytocin. Gaussian process classification was implemented in a leave-one-out cross-validated framework to discriminate between baseline rCBF maps (prior to treatment) and post treatment rCBF maps. The results indicated that oxytocin-induced changes in rCBF were sustained over the entire post treatment period with a peak change observed between 39 and 51 min.

These exemplar studies (Doyle et al. 2012; Paloyelis et al. 2014), illustrate that machine learning methods can be

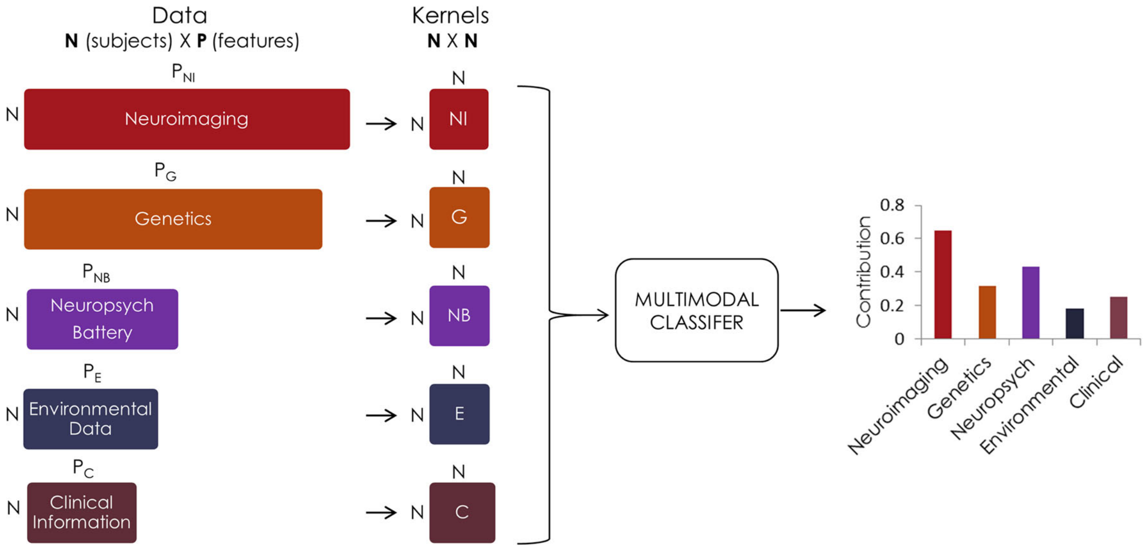

Fig. 2 A pipeline for performing classification using multiple disparate data source. The scheme represents a typical multiple kernel learning approach. Each modality is represented by a kernel which is in the dimensions of subjects (*N*). Data can be combined using a set of rules to create a single kernel from several kernels for, e.g. a weighted sum where the weightings are optimised to increase or decrease the contribution of each modality. In this graphical illustration we observe that neuroimaging followed by the neuropsychological battery and so on, had the greatest contribution to the predictive model

implemented to not only provide discriminative information but also to model the pharmacodynamic profile of the compound's effects in the brain. This profile provides an important insight into experiment design and when to ideally perform tasks which are likely to be altered by the compound. For, e.g. in the case of oxytocin, our work indicates that approximately 40 min after administration would be the optimal time to conduct a behavioural/fMRI paradigm.

How does the pharmacological agent behave using an experimental medicine model?

Experimental medicine models of brain disorders provide an approach to study the action of existing and novel compounds in a controlled setting and crucially in the absence of confounds typically present in the disease population—chronic effects of the illness, medication, environment, etc. A particular example of an experimental medicine model would be the administration of a compound known to induce symptoms reminiscent of the disorder of interest and the subsequent administration of a compound which is hypothesised to reduce these particular symptoms. Often for this paradigm, the main outcome is to assess if the treatment affected the induced disease-like state. In the context of neuroimaging, we can employ machine learning methods to assess the effect of treatment. We illustrate the methodology using two compounds often used in experimental medicine models—ketamine and scopolamine. Fundamentally, we want to extract a single outcome measure from brain imaging data. In Fig. 3, we illustrate a pipeline for investigating the effect of a pre-treatment with compound 'A' on the response to compound 'Z'. In this example, the classifier is trained using data collected from participants on placebo or compound Z. This classifier is then tested using data collected from an independent participant on placebo, both compounds 'A + Z' and compound Z producing a probability of having received compound Z. Using these probabilities, we can infer the effect of A on Z. These data could also be analysed using a multivariate ordinal regression framework (Doyle et al. 2013a). Multivariate ordinal regression inherently models the natural ordering in the data labels for example discrete states along a continuum of disease progression *healthy* → *at risk* → *early disease state* → *chronic disease state*. Using this approach, training data spanning all ordinal states are modelled rather than only the extremes. In addition, we can also extract a weight map which enables us to visualise the brain regions strongly driving the predictions on the test data.

Ketamine is an NMDA receptor antagonist, which when administered at sub-anaesthetic doses, induces symptoms resembling schizophrenia and exacerbates symptoms in patients diagnosed with schizophrenia (Krystal et al. 1994). While not producing a precise phenocopy, acute administration of ketamine has been used to model the pathophysiological features of schizophrenia in healthy human participants, with a particular interest in glutamatergic contributions. In our work, we investigated the effect of acute ketamine administration on the blood oxygen level-dependent (BOLD) response and the effect of pre-treatment on this response using a randomised cross-over design in 16 healthy male volunteers. Previously, we have shown that acute administration of ketamine produces a BOLD response that is reproducible and reliable (De

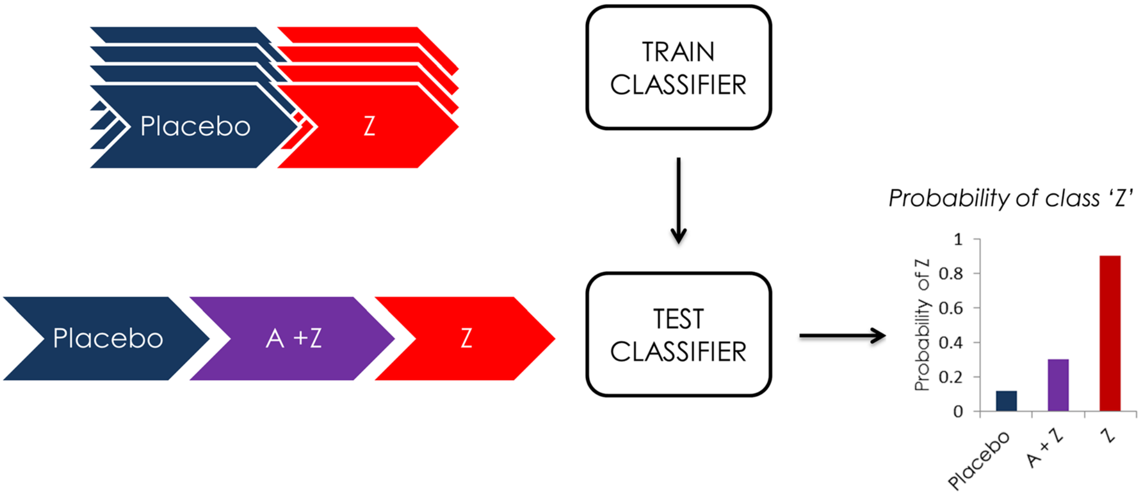

Fig. 3 A pipeline for quantifying attenuation of response to compound 'Z' by pre-treatment compound 'A'. The classifier is trained using data from participants on placebo and Z. This classifier is then tested using data from an independent participant in a repeated measures design treated with either placebo, 'A + Z' and Z producing a probability of belonging to treatment group Z. Using these probabilities, we can infer the extent of attenuation. The arrows denote the increasing response in to treatment from the placebo session, to the A + Z session and finally, the Z session

Simoni et al. 2013). To perturb the ketamine effect, two pre-treatment compounds were administered: the anticonvulsant, lamotrigine and the antipsychotic, risperidone (Doyle et al. 2013b). The key outcome measure in this case is an assessment of the extent to which these pre-treatments attenuate the BOLD response to ketamine. Initially, we trained a binary classifier to discriminate between the placebo condition and the ketamine condition and then tested this classifier on the placebo, ketamine and the presumed attenuated states produced by lamotrigine + ketamine and risperidone + ketamine from an independent participant; this process was iterated until each of the volunteers was used as a test case. We found that the probability of belonging to the ketamine class was significantly reduced following pre-treatment with lamotrigine and risperidone.

Scopolamine is a non-selective muscarinic receptor antagonist that induces impairments across multiple cognitive domains in healthy human volunteers (Bymaster et al. 1993). Therefore, acute administration of scopolamine can be used to model impairment and provide an opportunity to improve cognitive performance in participants who would otherwise perform at a neurotypical cognitive level (Lenz et al. 2012). The model is a commonly used translational tool to model cholinergic deficits in cognition as might be present in a range of neurodegenrerative conditions. In our work, we investigated the effect of scopolamine on regional cerebral blood flow (rCBF) and its potential reversal by pre-treatment with donepezil (Doyle et al. 2013a), an acetylcholinesterase inhibitor which has been previously shown to improve cognitive deficits in Alzheimer's disease (Di Santo et al. 2013) and reverses scopolamine-induced cognitive impairment (Cho et al. 2011). We modelled the ordinal trend in cerebral blood flow in 15 healthy male volunteers as placebo − donepezil + scopolamine − scopolamine. However, because scopolamine is a non-selective muscarinic acetylcholine receptor antagonist and donepezil will enhance cholinergic transmission to both nicotinic and muscarinic receptors, donepezil is not expected to simply reverse the effects of scopolamine when co-administered alongside the antagonist. Independent effects of donepezil are also expected. Thus, we would not predict the effects to be globally ordinal; that is, the effects of donepezil may attenuate scopolamine effects in some areas but not others. Therefore, we perform ordinal regression on three regions of interest previously indicated in neuroimaging studies—the thalamus and the occipital lobe and additionally, the anterior cingulate cortex (ACC) which receives cholinergic innervation from the basal forebrain. Ordinal regression could discriminate the placebo, pre-treated scopolamine and scopolamine sessions with high accuracy (recalling random chance level is 33.3 % for a three-class problem) in the thalamus (80 %) and the ACC (73.3 %) with more modest accuracy in the occipital lobe (64.4 %). This study used machine learning methods to confirm that pre-treatment with donepezil can reduce the scopolamine-induced effects on rCBF. The approach was applied in a multivariate manner to anatomically defined regions of interest which provides a more locally oriented interpretation of the model's performance and validating its potential use in profiling other compounds which may directly or indirectly attenuate scopolamine effects.

Here, we have provided examples of how machine learning can provide a framework for validating pharmacological models of disease and their perturbation using existing treatments. We observed that both the ketamine and scopolamine models were attenuated using existing compounds with machine learning providing an interpretable marker of attenuation. The combination of state-of-the-art statistical tools with pharmacological models is an exciting avenue for investigating the repurposing of existing compounds and the mechanisms of action of either existing or novel compounds. The use of well-characterised models such as ketamine or scopolamine also invites dose-finding studies, where minimally effective doses can be selected to be taken forward into further experimental studies or early phase trials.

How does a compound differ from and/or resemble existing compounds?

The comparison of pharmacological agents is a natural question in drug discovery. In general, it may be advantageous to compare a novel or existing compound to a library of relevant compounds across a wide variate of attributes. This has gained traction to profile and screen for attributes including absorption, distribution, metabolism and excretion properties (Lavecchia 2014; Liao et al. 2009). Automated profiling of compounds could help to gain a richer understanding of the compounds and help to decide which to prioritise for screening. Therefore, there are two questions: how are they different and how are they similar? Of course, we note here that these questions are not simply the inverse of each other. The first question pertaining to differentiation is the more conventional setting for machine learning where the classes that we are learning to discriminate represent different compounds. This setting is exemplified in a neuroimaging study in healthy volunteers by Marquand et al. (2012). In this study, a multi-class classifier was trained to discriminate between placebo, acute administration of atomoxetine and acute administration of methylphenidate using regional cerebral blood flow (rCBF) measured using arterial spin labelling. Sparse multinomial logistic regression was used to perform multi-class classification. This method is formulated to discriminate between two or more classes and incorporates penalties that are optimised using the training data to produce a sparse weight vector (i.e. weight vector for some voxels is set to zero). This study illustrated that multi-class classification was sensitive to differential effects of atomoxetine and methylphenidate on rCBF providing highly accurate predictions about class membership and also an insight into the brain regions driving the discrimination. The use of arterial spin labelling is attractive as a quantitative, translational tool (Bruns et al. 2009; Marquand et al. 2012; Wang et al. 2011) and additional insights may also be provided when coupled with task-dependent effects.

Duff et al. (2015) applied ranking support vector machines to eight fMRI studies investigating the effects of six different analgesic compounds on brain responses to painful stimuli. First, the SVMs were trained to distinguish between placebo and analgesic compound to assess the pharmacodynamic effect. Moderate to strong evidence (accuracies ranging from 70 to 91 %) was found for an analgesic effect in five out of six compounds tested. Second, SVMs were used to assess evidence for clinical efficacy by training on brain responses to pain in the presence of either placebo or an analgesic compound. Successful discrimination was reported for five of the compounds in the range of 69 to 83 %. Finally, the authors investigated whether a framework based on a limited number of existing compounds could be effective. To achieve this, they trained the machine learning algorithm using data from a single study and then tested this algorithm on the remaining studies. While the discriminative performance of the model for identifying an analgesic effect was reduced, nonetheless in many cases an analgesic effect could be accurately identified. In this study, the authors have shown that machine learning and imaging data from multiple studies can identify drug effects on brain activity and clinical efficacy. This type of multi-study, multi-drug paradigm could leverage existing data to optimise drug discovery.

A consideration for these approaches would be how to assess similarity of the effect of compounds (e.g. A and B). An obvious suggestion would be to train a model to discriminate between placebo and compound A and to test this model on compound B. For compound B, we could assign the probability of belonging to the class represented by compound A; if this probability is high, then we can infer that they are likely to be similar, but what can we infer if the probability is not high? In this latter case, we may not be able to say anything definitive about compound B in the context of compound A. This situation could arise when compound B has modes of action which do not completely overlap with compound A. Then, it may be more appropriate to analyse these compounds using techniques which relate two independent multivariate patterns, i.e. modelling the relationship between induced changes in brain activity as a result of compound A and a result of compound B. This can be achieved using techniques such as canonical correlation analysis (Cherry 1996) which is closely related to partial least squares (Sun et al. 2009). Canonical correlation analysis aims to find a set of linear transformation variables for each class (compounds A and B) so that the data are maximally correlated. Therefore, it does not produce a consensus map which we can interpret as a type of 'similarity heat map'. Further development is required to produce a multivariate measure of similarity that can be interpreted at the regional level.

Discussion

In this review, we have detailed several machine learning approaches, which help to answer questions around drug characterisation in the context of imaging data. These approaches could be readily used to help enhance drug discovery and development. Here, we have chosen to focus on neuroimaging as a marker of pharmacodynamic effects in the context of personalised/stratified medicine. We note that this review does not exhaustively cover different feature extraction approaches for neuroimaging data, i.e. the inputs to the machine learning method. For example, surogates for functional connectivity can be extracted from fMRI (Bullmore and Sporns 2009; Smith et al. 2011). Models based on these metrics could provide insight into the effect of pharmacological agents on brain function at a network level. For a detailed illustration of this approach in the context of pharmacological neuroimaging, we refer readers to the work of Joules et al. (2015).

The methods presented here are highly flexible and could therefore be also used for different data types and their combination which may ultimately lead to more accurate predictive models. These models could help to answer fundamental questions spanning the drug discovery pipeline: (1) Who should I recruit into this trial (Fig. 1)? (2) What should I measure (Fig. 2) and when should I measure it? (3) How does the compound behave using an experimental medicine model (Fig. 3)? (4) How does this compound differ from and/or resemble existing compounds?

To realise the potential of these tools , further development and tailoring of the methodology is necessary. One of the most common pitfalls in machine learning studies is overfitting. To help alleviate overfitting, techniques which penalise complexity can be employed and ideally models should be tested on independent data. In neuroimaging, the models are often tested using cross-validation. It is crucial to ensure that this cross-validation structure is preserved throughout the *entire* pipeline to avoid circularity (Kriegeskorte et al. 2010). A basic example could be using the entire dataset to select regions of the brain that are strongly affected by either the compound or a pathology and then performing a machine learning analysis using only these regions in the same dataset. This would result in a machine learning model with inflated performance that is not likely to generalise well to new data.

It may be beneficial to move away from the ‘black box’ biomarker approach (i.e. only the inputs and outputs are available, the internal processes are known) and instead move towards models that are interpretable and tailored to the particular question and perhaps incorporate mechanistic aspects of the mode of action of the drug (Doyle et al. 2013c). Development of methods that are robust to missing data and data acquired using different protocols or scanners are also an important consideration. While machine learning can exploit existing data to help inform several aspects of the study, an unresolved question is how to determine the sample size for these multivariate methods. Several methods, primarily for genetic data, have been proposed in the literature (Figueroa et al. 2012; Guo et al. 2010); however, a consensus has not yet been reached. Further work is required to assess these methods for machine learning in neuroimaging data. Moreover, in the absence of prior knowledge of the properties of the data, it may be difficult to reach a useful estimate for how many subjects will be required. Throughout the entire pipeline, it is essential to be mindful of the potential confounds that could produce misleading results. For example, if a compound is a stimulant or a sedative, then a systematic difference in motion in the scanner or performance on a cognitive task could be present and this difference could drive discrimination between placebo and drug. As with all studies, the potential confounds should ideally be identified before the study and appropriate measures should be taken to help alleviate their effects.

We have highlighted several studies which demonstrate the utility of machine learning to answer important questions for drug discovery and development. Further refinement and tailoring of these techniques may hold tremendous potential for drug discovery and drug validation.

Acknowledgements OMD gratefully acknowledges support from the EPSRC grant EP/L000296/1. OMD, MAM and MJB gratefully acknowledge support from the Innovative Medicines Initiative Joint Undertaking under Grant Agreement No 115008 (NEWMEDS). The Innovative Medicines Initiative Joint Undertaking is a public-private partnership between the European Union and the European Federation of Pharmaceutical Industries and Associations.

Conflict of interest Over the past 5 years, Mitul A. Mehta has received research funding from Eli Lilly, Roche and Takeda; has acted as a consultant for Cambridge Cognition and Lundbeck and received fees from Shire for contributions towards education. All others declare no conflicts of interest.

References

Borsook D, Becerra L, Hargreaves R (2006) A role for fMRI in optimizing CNS drug development. Nat Rev Drug Discov 5:411–424

Borsook D, Beccera L, Bullmore E, Hargreaves R (2010) Imaging in CNS drug discovery and development. Springer, New York

Borsook D, Becerra L, Fava M (2013) Use of functional imaging across clinical phases in CNS drug development. Transl Psychiat 3

Bruns A, Kunnecke B, Risterucci C, Moreau JL, von Kienlin M (2009) Validation of cerebral blood perfusion imaging as a modality for quantitative pharmacological MRI in rats. Magnetic Reson Med : Off J Soc Magn Reson Med/ Soc Magn Reson Med 61:1451–8

Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10:186–98

Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 2:121–167

Bymaster FP, Heath I, Hendrix JC, Shannon HE (1993) Comparative behavioral and neurochemical activities of cholinergic antagonists in rats. J Pharmacol Exp Ther 267:16–24

Cherry S (1996) Singular value decomposition analysis and canonical correlation analysis. J Clim 9:2003–2009

Cho W, Maruff P, Connell J, Gargano C, Calder N, Doran S, Fox-Bosetti S, Hassan A, Renger J, Herman G, Lines C, Verma A (2011) Additive effects of a cholinesterase inhibitor and a histamine inverse agonist on scopolamine deficits in humans. Psychopharmacology 218:513–24

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20: 273–297

Costafreda SG, Chu C, Ashburner J, Fu CH (2009) Prognostic and diagnostic potential of the structural neuroanatomy of depression. PLoS One 4:e6353

Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehericy S, Habert MO, Chupin M, Benali H, Colliot O (2011) Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. NeuroImage 56:766–81

De Simoni S, Schwarz AJ, O'Daly OD, Stephenson S, Zelaya FO, Williams SCR, Mehta MA (2013) Test-retest reliability of the BOLD pharmacological MRI response to ketamine in healthy volunteers. NeuroImage 64:75–90

Di Santo SG, Prinelli F, Adorni F, Caltagirone C, Musicco M (2013) A meta-analysis of the efficacy of donepezil, rivastigmine, galantamine, and memantine in relation to severity of Alzheimer's disease. J Alzheimers Dis 35:349–61

Dodds CM, Muller U, Clark L, van Loon A, Cools R, Robbins TW (2008) Methylphenidate has differential effects on blood oxygenation level-dependent signal related to cognitive subprocesses of reversal learning. J Neurosci : Off J Soc Neurosci 28:5976–82

Donaldson ZR, Young LJ (2008) Oxytocin, vasopressin, and the neurogenetics of sociality. Science 322:900–4

Doyle OM, Mehta MA, Brammer MJ, Schwartz AJ, De Simoni S, Marquand AF (2012) Data-driven modeling of BOLD drug response curves using Gaussian process learning. Springer Lecture Notes in Artificial Intelligence: 7263: In Press.

Doyle OM, Ashburner J, Zelaya FO, Williams SC, Mehta MA, Marquand AF (2013a) Multivariate decoding of brain images using ordinal regression. NeuroImage 81C:347–357

Doyle OM, De Simoni S, Schwarz AJ, Brittain C, O'Daly OG, Williams SCR, Mehta MA (2013) Quantifying the attenuation of the ketamine phMRI response in humans: a validation using antipsychotic and glutamatergic agents. . J Pharmacol Exp Ther (In Press)

Doyle OM, Tsaneva-Atansaova K, Harte J, Tiffin PA, Tino P, Diaz-Zuccarini V (2013c) Bridging paradigms: hybrid mechanistic-discriminative predictive models. IEEE Trans Bio-Med Eng 60:735–42

Doyle OM, Westman E, Marquand AF, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Lovestone S, Williams SC, Simmons A (2014) Predicting progression of Alzheimer's disease using ordinal regression. PLoS One 9:e105542

Duff EP, Vennart W, Wise RG, Howard MA, Harris RE, Lee M, Wartolowska K, Wanigasekera V, Wilson FJ, Whitlock M, Tracey I, Woolrich MW, Smith SM (2015) Learning to identify CNS drug action and efficacy using multistudy fMRI data. Sci Trans Med 7:274ra16

Figueroa RL, Zeng-Treitler Q, Kandula S, Ngo LH (2012) Predicting sample size required for classification performance. BMC Med Inform Decis Making 12:8

Fu CH, Steiner H, Costafreda SG (2013) Predictive neural biomarkers of clinical response in depression: a meta-analysis of functional and structural neuroimaging studies of pharmacological and psychological therapies. Neurobiol Dis 52:75–83

Gong Q, Wu Q, Scarpazza C, Lui S, Jia Z, Marquand A, Huang X, McGuire P, Mechelli A (2011) Prognostic prediction of therapeutic response in depression using high-field MR imaging. NeuroImage 55:1497–503

Guo Y, Graber A, McBurney RN, Balasubramanian R (2010) Sample size and statistical power considerations in high-dimensionality data settings: a comparative study of classification algorithms. BMC Bioinforma 11:447

Hahn T, Kircher T, Straube B, Wittchen HU, Konrad C, Strohle A, Wittmann A, Pfleiderer B, Reif A, Arolt V, Lueken U (2015) Predicting treatment response to cognitive behavioral therapy in panic disorder with agoraphobia by integrating local neural information. JAMA Psychiatry 72:68–74

Haufe S, Meinecke F, Gorgen K, Dahne S, Haynes JD, Blankertz B, Biessmann F (2014) On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage 87:96–110

Honey G, Bullmore E (2004) Human pharmacological MRI. Trends Pharmacol Sci 25:366–74

Insel TR (2012) Next-generation treatments for mental disorders. Science translational medicine 4

Insel TR (2014) The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. Am J Psychiatry 171:395–7

Joules R, Doyle OM, Schwarz AJ, O'Daly OG, Brammer MJ, Williams SCR, Mehta MA (2015) Ketamine induces a robust whole-brain connectivity pattern that can be differentially modulated by drugs of different mechanism and clinical profile. In submission

Kapur S, Phillips AG, Insel TR (2012) Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Mol Psychiatry 17:1174–1179

Kendler KS, Eaves LJ, Walters EE, Neale MC, Heath AC, Kessler RC (1996) The identification and validation of distinct depressive syndromes in a population-based sample of female twins. Arch Gen Psychiatry 53:391–399

Khodayari-Rostamabad A, Hasey GM, Maccrimmon DJ, Reilly JP, de Bruin H (2010) A pilot study to determine whether machine learning methodologies using pre-treatment electroencephalography can predict the symptomatic response to clozapine therapy. Clin Neurophysiol 121:1998–2006

Kriegeskorte N, Lindquist MA, Nichols TE, Poldrack RA, Vul E (2010) Everything you never wanted to know about circular analysis, but were afraid to ask. J Cereb Blood Flow Metab : Off J Int Soc Cereb Blood Flow Metab 30:1551–7

Krystal JH, Karper LP, Seibyl JP, Freeman GK, Delaney R, Bremner JD, Heninger GR, Bowers MB Jr, Charney DS (1994) Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans. Psychotomimetic, perceptual, cognitive, and neuroendocrine responses. Arch Gen Psychiatry 51:199–214

Lavecchia A (2014) Machine-learning approaches in drug discovery: methods and applications. Drug discovery today

Lenz RA, Baker JD, Locke C, Rueter LE, Mohler EG, Wesnes K, Abi-Saab W, Saltarelli MD (2012) The scopolamine model as a pharmacodynamic marker in early drug development. Psychopharmacology 220:97–107

Liao Q, Wang J, Webster Y, Watson IA (2009) GPU accelerated support vector machines for mining high-throughput screening data. J Chem Inf Model 49:2718–25

Mak IWY, Evaniew N, Ghert M (2014) Lost in translation: animal models and clinical trials in cancer treatment. Am J Transl Res 6:114–118

Marquand AF, O'Daly OG, De Simoni S, Alsop DC, Maguire RP, Williams SC, Zelaya FO, Mehta MA (2012) Dissociable effects of methylphenidate, atomoxetine and placebo on regional cerebral blood flow in healthy volunteers at rest: a multi-class pattern recognition approach. NeuroImage 60:1015–24

Medhi B, Misra S, Avti PK, Kumar P, Kumar H, Singh B (2014) Role of neuroimaging in drug development. Rev Neurosci 25:663–673

Mehta M, O'Daly O (2011) Pharmacological Application of fMRI. In: Modo M, Bulte JWM (eds) Magnetic resonance neuroimaging (methods in molecular biology). Humana Press, pp 551–565

Oquendo MA, Baca-Garcia E, Artes-Rodriguez A, Perez-Cruz F, Galfalvy HC, Blasco-Fontecilla H, Madigan D, Duan N (2012) Machine learning and data mining: strategies for hypothesis generation. Mol Psychiatry 17:956–9

Paloyelis Y, Doyle OM, Zelaya FO, Maltezos S, Williams SC, Fotopoulou A, Howard MA (2014) A spatiotemporal profile of in vivo cerebral blood flow changes following intranasal oxytocin in humans. Biol Psychiatry

Pauls AM, O'Daly OG, Rubia K, Riedel WJ, Williams SC, Mehta MA (2012) Methylphenidate effects on prefrontal functioning during attentional-capture and response inhibition. Biol Psychiatry 72:142–9

Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning. MIT Press, Cambridge, Mass

Schwarz AJ, Becerra L, Upadhyay J, Anderson J, Baumgartner R, Coimbra A, Evelhoch J, Hargreaves R, Robertson B, Iyengar S, Tauscher J, Bleakman D, Borsook D (2011a) A procedural framework for good imaging practice in pharmacological fMRI studies applied to drug development #1: processes and requirements. Drug Discov Today 16:583–93

Schwarz AJ, Becerra L, Upadhyay J, Anderson J, Baumgartner R, Coimbra A, Evelhoch J, Hargreaves R, Robertson B, Iyengar S, Tauscher J, Bleakman D, Borsook D (2011b) A procedural framework for good imaging practice in pharmacological fMRI studies applied to drug development #2: protocol optimization and best practices. Drug Discov Today 16:671–82

Shawe-Taylor J, Cristianini N (2004) Kernel methods for pattern analysis. Cambridge University Press, Cambridge

Smith SM, Miller KL, Salimi-Khorshidi G, Webster M, Beckmann CF, Nichols TE, Ramsey JD, Woolrich MW (2011) Network modelling methods for FMRI. NeuroImage 54:875–91

Sullivan PF, Kessler RC, Kendler KS (1998) Latent class analysis of lifetime depressive symptoms in the national comorbidity survey. Am J Psychiatry 155:1398–406

Sun LA, Ji SW, Yu SP, Ye JP (2009) On the equivalence between canonical correlation analysis and orthonormalized partial least squares. 21st International Joint Conference on Artificial Intelligence (Ijcai-09), Proceedings: 1230–1235.

Wang DJ, Chen Y, Fernandez-Seara MA, Detre JA (2011) Potentials and challenges for arterial spin labeling in pharmacological magnetic resonance imaging. J Pharmacol Exp Ther 337:359–66

Wise RG, Preston C (2010) What is the value of human FMRI in CNS drug development? Drug Discov Today 15:973–980

Wong DF, Tauscher J, Grunder G (2009) The role of imaging in proof of concept for CNS drug discovery and development. Neuropsychopharmacol 34:187–203