



**HAL**  
open science

# Learning to identify CNS drug action and efficacy using multi-study fMRI data

Duff Eugene

► **To cite this version:**

Duff Eugene. Learning to identify CNS drug action and efficacy using multi-study fMRI data. ICML Workshop on Statistics, Machine Learning and Neuroscience (Stamfins 2015), Bertrand Thirion, Lars Kai Hansen, Sanmi Koyejo, Jul 2015, Lille, France. hal-01225256

**HAL Id: hal-01225256**

**<https://inria.hal.science/hal-01225256>**

Submitted on 4 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# Learning to identify CNS drug action and efficacy using multi-study fMRI data

---

## Abstract

We summarise recently published work developing and validating machine-learning based protocols that permit functional magnetic resonance brain imaging (fMRI) to be used in drug discovery. To develop such protocols required strategies for identifying neuroimaging markers of drug efficacy that are robust to variability in the actions of effective compounds, and other factors, without requiring extensive new experimental data acquisition. Our approach is to learn signature brain activation modulations indicative of drug efficacy across databases of heterogeneous fMRI studies associated with the disease state of interest. Activity signatures are learnt through training classifiers to distinguish effective drugs from placebo. We validated the approach by developing a protocol for the identification of novel analgesics. This protocol successfully identified evidence for efficacy in 8 separate studies of established analgesic compounds. Further work has extended approaches to resting-state imaging. This work demonstrates a framework in which machine learning can facilitate translational applications of brain imaging. More broadly, this work shows how combining meta-analytic and machine learning approaches can broaden the inferences and applications possible with imaging technology.

## 1. Introduction

There is an urgent need to improve the success rates of clinical trials in drug development. Biomarkers that screen candidate compounds prior to large-scale clinical trials have the potential to substantially improve the productivity and cost-effectiveness of drug development. fMRI, which is capable of identifying signature effects of disease and treatment on brain activity, shows considerable promise for such applications (Upadhyay et al., 2011). However, de-

---

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

spite this promise, progress towards validated fMRI-based procedures for assessing the potential efficacy of new CNS compounds has been slow (Wise & Preston, 2010). In part, this is due to the difficulty of achieving validated inferences regarding the likelihood that an unseen drug will be successful (clinical efficacy). fMRI does not quantify physiological variables directly associated with drug action, so evidence for the efficacy of compounds must be derived from empirically established associations between brain activity patterns and measurable clinical variables such as treatment outcome. Modulation of brain activity alone (pharmacological effect) does not imply that a drug is efficacious. What is required is the identification of a signature of clinical improvement that is robust to auxiliary effects of specific compounds, and heterogeneity in patients, acquisitions and other factors. Furthermore, it is necessary that imaging methods provide predictive capabilities beyond what can be obtained from available clinical measures alone.

To achieve this, we use a machine-learning based meta-analytic strategy. This provides the sensitivity required to identify consistent signatures of the action of efficacious compounds from existing experimental datasets, which are typically highly heterogeneous.

## 2. Methods

### 2.1. General protocol

The fMRI protocols we describe here are intended to be applied to potential candidates for stage IIa clinical trials. Here, fMRI might be used as a screening procedure, rapidly identifying drugs unlikely or very likely to succeed, using a modest number of patients. The procedures utilize a database of studies of established compounds or other treatments.

The assessment protocol has stages, Quality Assurance (QA), Assessment for Pharmacodynamic Effect, and Assessment for Clinical Efficacy (Fig 1A-C). Meta-analytic and machine learning are used across multiple stages. After QA, Pharmacodynamic effect assessment determines whether the compound modulates brain activity in any way, which can provide important evidence for the promise of a compound. This assessment is achieved by determining whether drug states can be discriminated from placebo in

000  
001  
002  
003  
004  
005  
006  
007  
008  
009  
010  
011  
012  
013  
014  
015  
016  
017  
018  
019  
020  
021  
022  
023  
024  
025  
026  
027  
028  
029  
030  
031  
032  
033  
034  
035  
036  
037  
038  
039  
040  
041  
042  
043  
044  
045  
046  
047  
048  
049  
050  
051  
052  
053  
054

055  
056  
057  
058  
059  
060  
061  
062  
063  
064  
065  
066  
067  
068  
069  
070  
071  
072  
073  
074  
075  
076  
077  
078  
079  
080  
081  
082  
083  
084  
085  
086  
087  
088  
089  
090  
091  
092  
093  
094  
095  
096  
097  
098  
099  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109

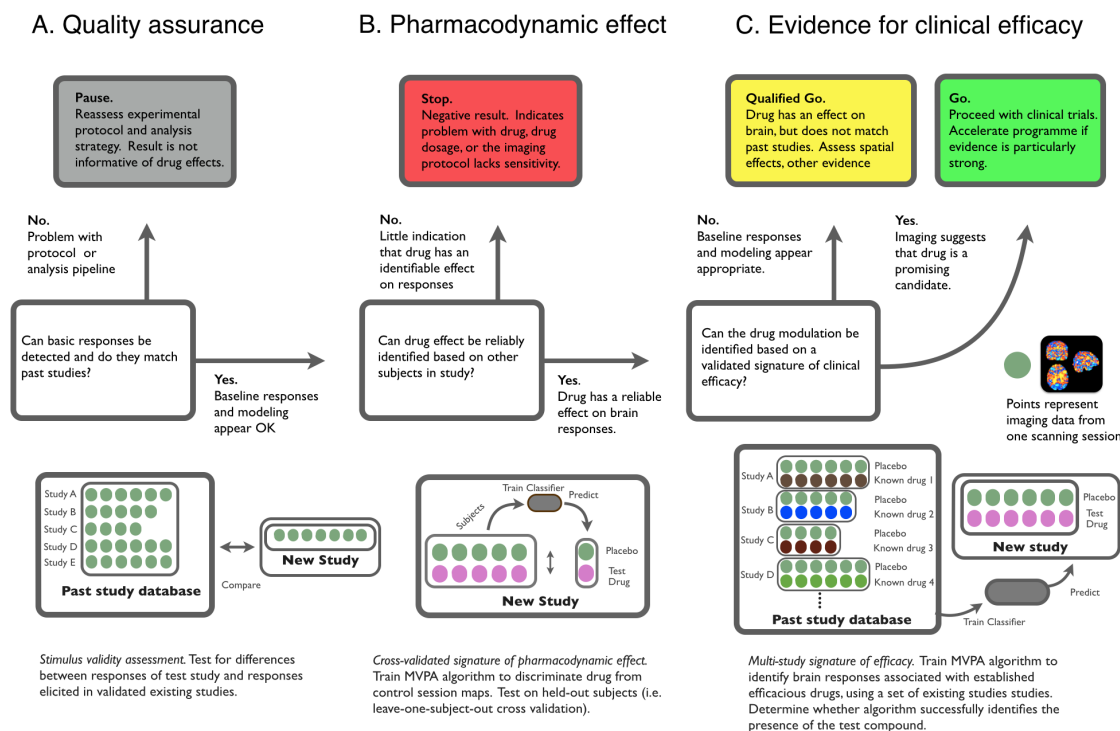


Figure 1. Protocol for assessing candidate CNS compounds using fMRI. Top row shows the general protocol. Bottom row shows a schematic of the proof-of-concept analgesic assessment protocol. (A) Quality-assurance. Prior to the assessment of drug effects, quality-assurance procedures test that the experimental protocol, preprocessing, and modeling identify expected brain responses. A comparison was performed between the effects of the painful stimulus in the test study and effects consistently observed in studies using similar stimuli. A study with lower responses in relevant brain areas was flagged as potentially problematic (t-test, corrected for multiple comparisons). (B) Test for pharmacodynamic effect. This stage tests whether the compound reliably modulates brain activity. Consistent modulation of brain responses is evidence for a pharmacodynamic effect. In the proof-of-concept procedure, a pharmacodynamic effect was flagged if an algorithm could be trained to distinguish drug from placebo responses in held-out subjects. (C) Test for evidence of clinical efficacy. The final stage tests whether observed modulations of responses correspond to established signs of efficacious action. The proof-of-concept approach determined whether a classifier trained to detect the effects of established analgesics could identify the test compound.

held out subjects. Evidence for clinical efficacy provides a similar assessment, but here the discriminative classifier is trained on a set of studies of established compounds, so that it identifies common features of efficacious drug action. This approach could identify brain activity modulations suggestive of efficacy that occur earlier or more reliably than changes in external clinical measures (Wise & Preston, 2010).

## 2.2. Proof-of-concept validation

The proof-of-concept protocol was tested on eight placebo-controlled crossover fMRI studies of the effects of established analgesics on responses to pain. Each study was assessed as if a novel compound, with the remaining studies used to identify a signature of analgesic efficacy. Across studies, the stimuli were always painful, but varied in their

timing, location, and quality. The studies investigated analgesics from a variety of compound classes, and varied in the MRI scanners, disease states, dosing regimens, and numbers of subjects used. Being capable of using heterogeneous datasets ensures that a high proportion of existing studies can be used to inform and validate assessments, and that assays can be developed in the context of evolving acquisition and experimental protocols.

A forced-choice linear support vector machine (SVM) was used for classification, using features derived from an Independent Component Analysis (ICA) based decomposition of the task response maps, and a c-parameter determined by cross-validation. This classifier was trained to predict which of the sessions was associated with the test analgesic. A leave-one-subject-out cross-validation scheme generated predictions for all subjects. Reliable identifica-

Study	Test compound	QA	Pharmacodynamic effect		Clinical efficacy		Decision
		Area of reduced response (%)	Accuracy (range)	P	Accuracy (range)	P	
<b>Analgesic drug study assessments</b>							
a	Gabapentin	0	92% (70–100)	0.0002	83% (60–94)	0.003	Go
b	Pregabalin	0	70% (52–83)	0.017	61% (44–76)	0.105	Go
c	Pregabalin	0	81% (61–92)	0.002	69% (42–79)	0.038	Go
d	Tramadol	0	56% (36–74)	0.22	75% (55–74)	0.01	Go (Q)
e	Remifentanyl	0	86% (70–94)	0.000	82% (65–92)	0.000	Go
f	Remifentanyl	0	83% (60–94)	0.003	75% (51–90)	0.003	Go
g	THC	0	71% (49–86)	0.028	57% (36–76)	0.22	Go (Q)
h	Naproxen	0	73% (55–83)	0.01	73% (55–83)	0.01	Go
<b>Control study assessments</b>							
i	Placebo (study b)	0	52% (22–54)	0.339	48% (32–64)	0.5	Stop
j	Placebo (study e)	0	27% (15–45)	0.97	45% (29–62)	0.584	Stop
k	Baseline (study f)	0	58% (36–78)	0.194	33% (16–57)	0.806	Stop
l	Visual stimulus (study f)	79	74% (51–90)	0.019	33% (16–57)	0.019	Reassess
m	Auditory stimulus (study f)	6	100% (82–100)	0.000	58% (36–78)	0.194	Reassess
n	Visual stimulus (study h)	48	73% (55–87)	0.009	47% (30–65)	0.5	Reassess

Figure 2. Results of the proof-of-concept drug assessment protocol applied to analgesic and control studies. QA identified the percentage of voxels showing responses that were significantly lower than responses in regions consistently responding to similar stimuli in the other studies (unpaired t-test, corrected for multiple comparisons). The pharmacodynamic effect assessment determined whether the stimulus responses associated with the test compound could be distinguished from those associated with placebo, evidence of a drug effect. The clinical efficacy assessment tested whether studies of established compounds identified patterns of effects that could be used to distinguish the test compound from placebo. Accuracies give the proportion of subjects for which the test compound session was correctly identified. P-values indicate the probability of achieving this accuracy or better given no drug effect (binomial test, chance = 50%). Sample sizes correspond to the number of subjects reported in Table 1. 90% Wilson-score confidence intervals are also shown. Colors indicate whether compounds passed a specific assessment phase based on the predefined decision rules (Fig. 1): green pass; yellow provisional pass; red fail; gray problem with QA, reassess.

tion of the drug session indicated evidence for pharmacodynamic effect or clinical efficacy.

### 3. Results

The results of applying the proof-of-concept procedure to the fMRI test datasets is shown in Figure 2. Moderate to strong evidence for an analgesic drug like effect was identified for all analgesic compounds (a-h), while a set of assessments of placebo control conditions identified no evidence for efficacy (i-n).

Examination of the brain activity contrast maps (Fig. 3B), and weight maps of the classifiers trained to discriminate drug from placebo sessions across all studies (Fig. 3C) identified a pattern of positive and negative weightings driving predictions, corresponding to established pain regions, including the insula and anterior cingulate cortex. Signal from a smaller selection of expert defined regions of interest did not provide the sensitivity to reliably identify held-out analgesics that the wholly data-driven signatures provided.

To determine whether approaches using existing compounds could be effective when only a limited number of studies of efficacious compounds were available, we investigated the performance of algorithms trained on just one study to identify other analgesics (Fig. 4). These assessments were less reliable, producing a maximum mean accuracy of 68% compared to a mean accuracy of 74% when classifiers were trained on all remaining studies. Transfer of discrimination capabilities across studies was good: when the same compounds was tested in two studies, training on one of the two studies, and testing on the second, produced high drug-placebo discrimination rates (minimum accuracy 74%,  $p < 0.02$ ).

### 4. Discussion

Early-stage testing of candidate CNS compounds has often been inadequate. Here, we have demonstrated and validated an fMRI-based drug assessment protocol, that utilises published datasets and machine learning methods to inform the assessment of a novel drug candidate. The approach was able to identify moderate to strong evidence for drug

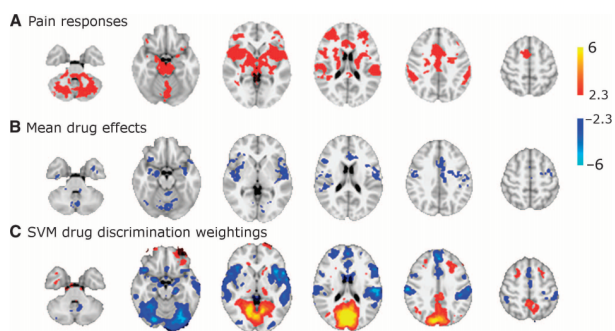


Figure 3. Multi-study effect maps. (A) Thresholded z-statistic maps of a mixed effects analysis of the average pain response across all 8 studies. Red corresponds to brain regions showing significant response to the painful stimuli (one-group t-test for nonzero response,  $p < 0.05$ , cluster-corrected). (B) Mixed effects analysis of modulation of pain responses by analgesics across all 8 analgesic studies. Blue corresponds to regions showing a significantly reduced response to the painful stimuli in the analgesic sessions (paired two-group t-test,  $p < .05$ , cluster-corrected). (C) Weight map of the SVM trained across all 8 studies (arbitrary threshold). Red corresponds to positive weights; blue, to negative weights.

effects for all clinically proven analgesics, with no false-positives for control data. These results suggest that fMRI can be of immediate value in clinical areas such as drug discovery. As it is designed to employ existing data sets, the general protocol described is useful for rapid, early-stage drug assessment. The approach would benefit considerably from experiments being performed in a co-ordinated manner, with matched acquisitions and systematic exploration of confounds, established compounds, and non-efficacious compounds (Wager et al., 2013).

The ICA/SVM-based forced choice classification approach described here is relatively simple but effective. The approach could be refined for the multi-study context in a variety of ways. Approaches that take into account the covariance structure of samples, variations in SNR across subjects and studies, and multiple modes of efficacious drug action, are likely to be beneficial. We are currently testing classification schemes that provide probabilistic estimates of the presence of a drug, which will allow the identification of drug effects in smaller samples. We have recently developed a pipeline for studies of drug modulations of the resting-state.

## References

Upadhyay, Jaymin, Anderson, Julie, Schwarz, Adam J., Coimbra, Alexandre, Baumgartner, Richard, Pendse, G., George, Edward, Nutile, Lauren, Wallin, Diana,

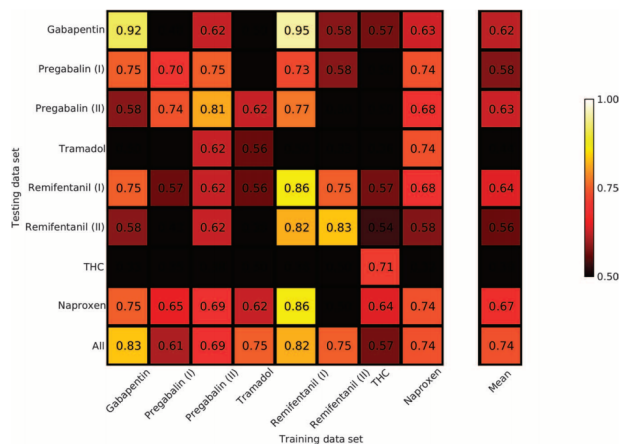


Figure 4. Identifying evidence for clinical efficacy using data from a single training study. Classifiers trained to discriminate a single analgesic from placebo were tested on other studies to determine the extent to which discriminative capabilities transferred. All corresponds to the training set comprising of all other studies, excluding studies of the same compound. The diagonal corresponds to within-study prediction rates measured using leave-one-subject-out cross-validation. The All row shows the prediction accuracies for predictors trained on multiple studies (Fig. 2). The mean column shows the average performance of each training dataset when applied to other studies. Black boxes indicate discrimination was not greater than 0.5.

Bishop, James, Neni, Saujanya, Maier, Gary, Iyengar, Smriti, Evelhoch, Jeffery L., Bleakman, David, Hargreaves, Richard, Becerra, Lino, and Borsook, David. Imaging Drugs with and without Clinical Analgesic Efficacy. *Neuropsychopharmacology*, 36(13):2659–2673, December 2011. ISSN 0893-133X. doi: 10.1038/npp.2011.156. URL .

Wager, Tor D, Atlas, Lauren Y, Lindquist, Martin A, Roy, Mathieu, Woo, Choong-Wan, and Kross, Ethan. An fMRI-based neurologic signature of physical pain. *The New England journal of medicine*, 368(15):1388–97, April 2013. ISSN 1533-4406. doi: 10.1056/NEJMoa1204471. URL .

Wise, Richard G. and Preston, Cliff. What is the value of human FMRI in CNS drug development? *Drug discovery today*, 15(21-22):973–980, November 2010. ISSN 1878-5832. doi: 10.1016/j.drudis.2010.08.016. URL .