



The relative importance of imaging markers for the prediction of Alzheimer's disease dementia in mild cognitive impairment – Beyond classical regression



Stefan J. Teipel^{a,b,*}, Jens Kurth^c, Bernd Krause^c,
Michel J. Grothe^a, for the Alzheimer's Disease Neuroimaging Initiative¹

^aGerman Center for Neurodegenerative Diseases (DZNE), Rostock, Germany

^bDepartment of Psychosomatic Medicine, University Medicine Rostock, Rostock, Germany

^cDepartment of Nuclear Medicine, University Medicine Rostock, Rostock, Germany

ARTICLE INFO

Article history:

Received 17 November 2014

Received in revised form 7 May 2015

Accepted 14 May 2015

Available online 21 May 2015

ABSTRACT

Selecting a set of relevant markers to predict conversion from mild cognitive impairment (MCI) to Alzheimer's disease (AD) has become a challenging task given the wealth of regional pathologic information that can be extracted from multimodal imaging data.

Here, we used regularized regression approaches with an elastic net penalty for best subset selection of multiregional information from AV45-PET, FDG-PET and volumetric MRI data to predict conversion from MCI to AD. The study sample consisted of 127 MCI subjects from ADNI-2 who had a clinical follow-up between 6 and 31 months. Additional analyses assessed the effect of partial volume correction on predictive performance of AV45- and FDG-PET data.

Predictor variables were highly collinear within and across imaging modalities. Penalized Cox regression yielded more parsimonious prediction models compared to unpenalized Cox regression. Within single modalities, time to conversion was best predicted by increased AV45-PET signal in posterior medial and lateral cortical regions, decreased FDG-PET signal in medial temporal and temporobasal regions, and reduced gray matter volume in medial, basal, and lateral temporal regions. Logistic regression models reached up to 72% cross-validated accuracy for prediction of conversion status, which was comparable to cross-validated accuracy of non-linear support vector machine classification. Regularized regression outperformed unpenalized stepwise regression when number of parameters approached or exceeded the number of training cases. Partial volume correction had a negative effect on the predictive performance of AV45-PET, but slightly improved the predictive value of FDG-PET data. Penalized regression yielded more parsimonious models than unpenalized stepwise regression for the integration of multiregional and multimodal imaging information. The advantage of penalized regression was particularly strong with a high number of collinear predictors.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Predicting the conversion from mild cognitive impairment (MCI) into Alzheimer's disease (AD) dementia is among the clinically most relevant diagnostic tasks in the field of AD (Jack, 2012). Several imaging measures have demonstrated promising accuracy of prediction, including cortical amyloid load based on ¹¹C- and ¹⁸F-PET tracers (Chen et al., 2014), regional cerebral hypometabolism based on ¹⁸F-fluorodeoxyglucose (FDG)-PET (for a recent review see (Cohen and Klunk, 2014)), and regional grey matter volume derived from volumetric MRI (for review see (Teipel et al., 2013)). In addition to prediction accuracy, researchers are interested in the interpretation of a prediction model. In general, a more parsimonious model will provide more insight into the relationship between the response and the predictor variables than a more complex model including

* Corresponding author at: Department of Psychosomatic Medicine, University of Rostock, and DZNE Rostock, Gehlsheimer Str. 20, 18147 Rostock, Germany. Tel.: +11 49 381 494 9470; fax: +11 49 381 494 9472.

E-mail address: stefan.teipel@med.uni-rostock.de (S.J. Teipel).

¹ Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu/). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

all available covariates ((Hastie et al., 2009), page 57). Efficient selection of the subset of the most relevant predictor variables from a rich set of available data modalities and parameters is therefore an important step in identifying clinically relevant prognostic markers.

An approach typically used for this selection process is multiple linear regression with stepwise selection. With highly collinear predictor variables, however, these models tend to inflate the variance of estimated regression coefficients so that coefficient estimates may change in response to small differences in the model. In the words from the seminal paper of Farrar and Glauber: “The mathematics, in its brute and tactless way, tells us that explained variance can be allocated completely arbitrarily between linearly dependent members of a completely singular set of variables, and almost arbitrarily between members of an almost singular set” (page 7 (Farrar and Glauber, 1967)).

Non-linear approaches based on machine learning can partly overcome this problem, but may over-fit data in the presence of a large number of features relative to the available number of cases ((Hastie et al., 2009), page 431f, (Duda et al., 2001), page 221 (Dormann et al., 2013)). In addition, they do not easily lend themselves to the analysis of survival data with censoring effects, and do not offer an easily interpretable prediction model. Penalized regression models provide an attractive alternative when the number of predictive features is high relative to the number of observed cases (Zou and Hastie, 2005; Zou and Zhang, 2009) and features are collinear (Hoerl and Kennard, 1970; Tibshirani, 1996). Penalization shrinks the regression coefficients by imposing a penalty on the size of the correlation strength, eventually, depending on the penalization term used, setting some of them to zero. In recent years these models have extended to cover models for survival analysis (Friedman et al., 2010), including proportional hazards models, such as Cox regression (Cox and Oakes, 1984), to take censoring of observations into account. The latter point is important for the prediction of conversion into dementia within a time frame that is relevant for a clinical trial, i.e. between 6 months and 3 years, where the likelihood of censored observations is relatively high in an MCI population (22%–34% conversion rate between 2 to 3 years of follow-up (Duara et al., 2011)).

In the present study, we determined the isolated and combined accuracy of structural MRI, FDG-PET and amyloid-sensitive AV45-PET to predict rapid to moderately fast conversion into dementia in 127 MCI subjects retrieved from the ADNI database. We further determined prediction accuracies before and after partial volume correction of PET data. We hypothesized that amyloid and FDG-PET measures would be more sensitive to conversion of MCI individuals into AD dementia than measures of atrophy and that correction for partial volume effects would alter the predictive accuracy of both PET modalities. Since the number of features used for prediction in the combined models was almost as high as the number of available cases and the regional measures were expected to be collinear within and across modalities, we compared penalized regression analysis, extending the Cox and logistic regression approach in an elastic net framework (Friedman et al., 2010), with more classical unregularized regression models in respect to prediction accuracy as well as parsimony of subset selection.

2. Material and methods

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a \$60 million, 5-year public–private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment

(MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55–90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow-up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see <http://www.adni-info.org>.

2.1. Participants

AV45-PET, FDG-PET and structural MRI scans were retrieved from the ADNI-2 extension of the ADNI project and included imaging data of 127 subjects in a late stage of MCI (LMCI) who had a clinical follow-up after at least 6 months. Status of conversion (MCI_c) or non-conversion (MCI_{nc}) was determined at follow-up (average follow-up time was 17.3 (SD 6.6) months, ranging between 6 and 31 months).

Detailed inclusion criteria for LMCI can be found at the ADNI web site (<http://adni.loni.usc.edu/methods/>). Briefly, LMCI subjects have MMSE scores between 24–30 (inclusive), a subjective memory concern reported by subject, informant, or clinician, objective memory loss measured by education adjusted scores on delayed recall (one paragraph from Wechsler Memory Scale Logical Memory II; education adjusted scores: ≥ 16 years: ≤ 8 ; 8–15 years: ≤ 4 ; 0–7 years: ≤ 2), a CDR = 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia.

2.2. Imaging data acquisition

ADNI-GO/-2 MRI data were acquired on multiple 3 T MRI scanners using scanner-specific T1-weighted sagittal 3D MPRAGE sequences. In order to increase signal uniformity across the multicentre scanner platforms, original MPRAGE acquisitions in ADNI undergo standardized image pre-processing correction steps.

AV45- and FDG-PET data were acquired on multiple instruments of varying resolution and following different platform-specific acquisition protocols. Similar to the MRI data, PET data in ADNI undergo standardized image pre-processing correction steps aimed at increasing data uniformity across the multicentre acquisitions.

More detailed information on the different imaging protocols employed across ADNI sites and standardized image pre-processing steps for MRI and PET acquisitions can be found on the ADNI web site (<http://adni.loni.usc.edu/methods/>).

The average acquisition delays between corresponding AV45-PET, FDG-PET and MRI scans used in this study were 35 (SD 32) days between MRI and FDG-PET, 11 (SD 20) days between FDG-PET and AV45-PET, and 41 (SD 36) days between MRI and AV45-PET.

2.3. Imaging data processing

Imaging data were processed by using statistical parametric mapping (SPM8, Wellcome Trust Center for Neuroimaging) and the VBM8-toolbox (<http://dbm.neuro.uni-jena.de/vbm/>) implemented in Matlab R2013b (MathWorks, Natick, MA).

2.3.1. MRI processing

First, MRI scans were automatically segmented into GM, white matter (WM) and cerebrospinal fluid (CSF) partitions of 1.5 mm isotropic voxel-size using the tissue prior-free segmentation routine of the VBM8-toolbox. The resulting GM and WM partitions of each subject in native space were then high-dimensionally registered to an aging/AD-specific reference template from a previous study (Grothe et al., 2013) using DARTEL (Ashburner, 2007). Structural brain characteristics change considerably in advanced age and AD and spatial registration accuracy worsens with deviance from the template characteristics, rendering the MNI standard space template inappropriate for high-dimensional image normalization of aged and demented populations. Therefore, the reference template in this study was derived by DARTEL—aligning 50 healthy elderly subjects and 50 subjects with very mild, mild and moderate AD retrieved from an open access MRI database (<http://www.oasis-brains.org>), and thus reflects unbiased aging/AD-specific structural characteristics. Individual flow-fields resulting from the DARTEL registration to the reference template were used to warp the GM segments, and voxel-values were modulated for volumetric changes introduced by the high-dimensional normalization, such that the total amount of GM volume present before warping was preserved. Finally, for voxel-based analyses modulated warped GM segments were smoothed with a Gaussian smoothing kernel of 8 mm full-width at half maximum (FWHM). All preprocessed GM maps passed a visual inspection for overall segmentation and registration accuracy.

2.3.2. PET data processing

Each subject's AV45- and FDG-PET scans were rigidly coregistered to a skull-stripped version of the corresponding structural MRI scan and corrected for partial volume effects (PVE). PVE correction followed the algorithm proposed by Müller-Gärtner (Müller-Gärtner et al., 1992) and was implemented using in-house written Matlab scripts based on SPM8's image processing routines. Briefly, the PET signal within WM and CSF compartments was assumed to be homogeneous and was measured as average signal within the individual WM and CSF partitions, thresholded at 99% tissue probability. Subsequently, WM and CSF tissue maps were multiplied by the PET activity estimate for the respective compartment and convolved by the point spread function of the PET scan. Spill-in effects of WM and CSF signal into the GM compartment were corrected by subtracting the convolved maps of WM and CSF PET activity from the original PET scan. Spill-out effects of GM PET signal into WM and CSF compartments were corrected by dividing the spill-in corrected PET scan by a convolved version of the GM probability map. Tissue segments were determined using smoothed probability maps of GM and WM, respectively. Finally, only regions with a GM probability of at least 50% were retained in the PVE corrected AV45- and FDG-PET scans. GM-specific PVE corrected PET scans in subject space were warped to the aging/AD-specific reference space (without modulation of voxel-values) using the DARTEL flow-fields derived from the registration of the corresponding MRI scans. For voxel-based analyses the warped PET scans were smoothed with a Gaussian smoothing kernel of 8 mm full-width at half maximum (FWHM).

2.3.3. Extraction of regional imaging features

For regional analysis, we used regions of interest (ROIs) derived from the Hammers Maximum Probability atlas (Hammers et al., 2003). From the 83 regions of the atlas, we disregarded regions that are known to be not prominently involved in AD, such as the cerebellum, not assessable with PET, such as the ventricles, or belonging to white matter regions, such as the corpus callosum. This left us with a total of 42 bilateral cortical and subcortical brain regions. The corresponding atlas labels were high-dimensionally warped into the reference space of this study based on a DARTEL registration of the MNI152 template (the template space of the Hammers Maximum Probability atlas) to the aging/AD-specific reference template. The warped atlas

labels were then multiplied with a binary GM mask of the reference template, thresholded at 50% GM probability. Fig. 1 provides an illustration of the full set of ROIs in the reference space.

Individual GM volumes of the ROIs were extracted automatically from the warped GM segments (before smoothing) by summing up the modulated GM voxel values within the respective ROI masks in the reference space. For further analyses, the extracted regional GM volumes were scaled by the total intracranial volume, calculated as the sum of total volumes of the GM, white matter and cerebrospinal fluid partitions.

Individual AV45- and FDG-PET uptake values within the ROIs were extracted from the warped PET maps (before smoothing) by averaging the voxel values within the respective ROI masks in the reference space. Regional FDG- and AV45-PET uptake means were converted to standard uptake value ratios (SUVRs) by scaling to the mean uptake value within a mask of cerebellar GM, also derived from the Hammers Maximum Probability atlas (Hammers et al., 2003).

2.4. Statistical analysis

2.4.1. Demographic characteristics

Baseline demographic characteristics were compared between MCI_c and MCI_{nc} using parametric and non-parametric tests as required: age and years of education were compared between groups using Student's *t* test, gender distribution using Chi² test, and neuropsychological test results using non-parametric Mann-Whitney U test.

2.4.2. Voxel-based analyses

Amyloid load as determined using AV45-PET, cortical metabolism as determined using FDG-PET, and regional reductions of grey matter volumes were compared between MCI_c and MCI_{nc} using a linear model within each modality based on proportionally scaled voxel based data (relative to cerebellar grey matter values for the PET data and to total intracranial volume for the grey matter maps, respectively). Clusters of at least 50 contiguous voxels passing an FDR corrected *p*-value of 0.05 were considered significant. Analyses were conducted using SPM8.

2.4.3. Models to predict time to conversion

To check for multicollinearity, we determined the variance inflation factor (VIF) (Belsley, 1991) for each dependent variable on the set of the remaining dependent variables using the Matlab function `colldiag.m`, available at (<http://www.subcortex.net/research/code/collinearity-diagnostics-matlab-code>).

Penalized regression models were calculated using the package `glmnet` (available at <http://cran.r-project.org/web/packages/glmnet/index.html>) in R version 3.1.0 (The R Foundation for Statistical Computing). Elastic net regression is controlled by two parameters, (i) α , which sets the degree of mixing between two extremes of regularized regression, namely ridge regression ($\alpha = 0$) and the Lasso (Least Absolute Shrinkage and Selection Operator; $\alpha = 1$), and (ii) λ , defining the strength of regularization (Friedman et al., 2010). After α had been selected according to the minimization of the partial likelihood deviance of the model (see Supplementary Fig. 1 for an example), λ was determined using grid search with 10-fold cross-validation. The optimal λ was determined as the mean across 100 iteratively determined λ values minimizing the deviance of the model.

2.4.3.1. Single modality prediction of time to conversion. For each modality, we selected the predictor variables that contributed to prediction of time to conversion within a penalized Cox regression framework among 42 regional values together with episodic memory performance, MMSE score, sex, age, handedness, and education. Since statistical inference tests are not yet developed for elastic net regression, we used two complementary approaches to assess the relevance of the estimated coefficients. First, we used bootstrapping with 1000 repeated random

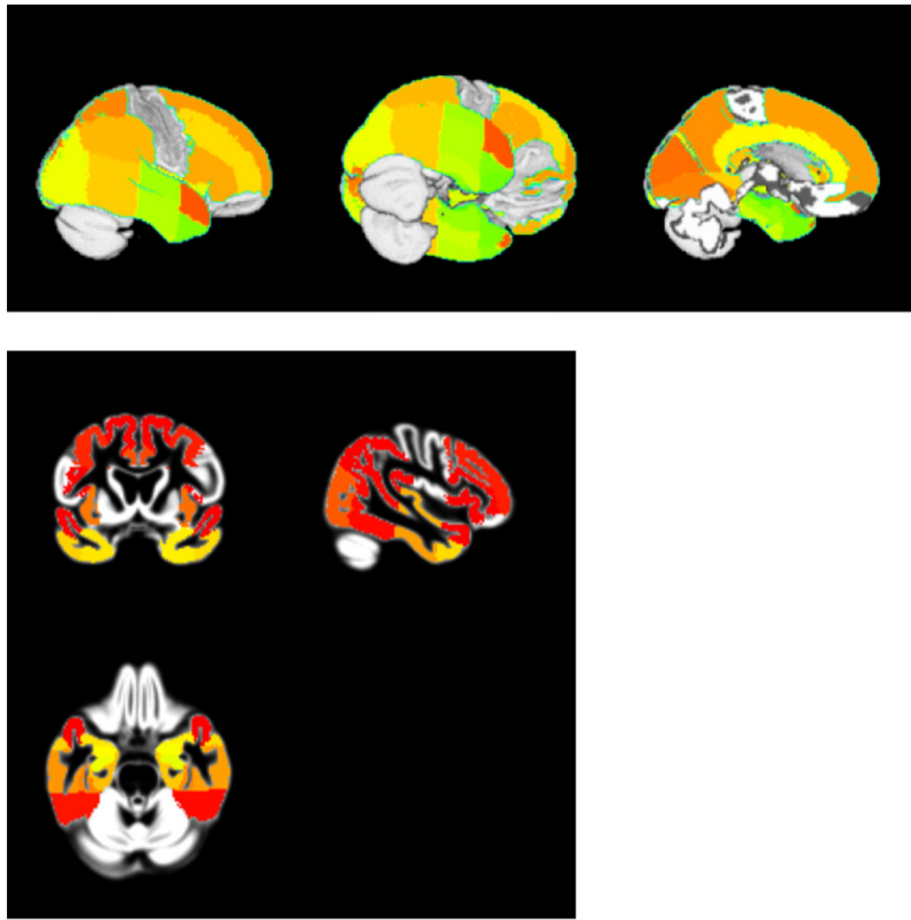


Fig. 1. Selected regions. Regions from the Hammers Maximum Probability atlas (Hammers et al., 2003) that were selected for the classification models projected onto the reference template in standard space. Upper row: surface view and midsagittal view. Lower row: Coronar, sagittal and axial sections focusing on subregions in temporal lobe.

selections from 63.2% of the data to determine the frequency of the selection of each single variable. In addition, we used permutation sampling of the data to determine the frequency distribution of the elastic net coefficients across 100 iterations to determine z-scores for coefficient estimates. Only variables with a selection frequency of at least 50% and a z-score larger than 1.96 in absolute value, corresponding to a two-tailed error rate of $p < 0.05$, are being reported.

For comparison we also calculated bidirectional (backward and forward) stepwise unpenalized Cox regressions using the function *step* in R. The function weights the choices via the AIC criterion, which takes account of the total number of fitted parameters.

2.4.3.2. Effect of partial volume correction of PET data. Given the heterogeneous use of PVE correction across PET imaging studies and the largely unknown effect of PVE correction on the predictive value of AV45- and FDG-PET data, voxel-wise analyses of group differences and predictive regression models were also performed using the AV45- and FDG-PET data without PVE correction.

2.4.3.3. Multiple modality prediction of time to conversion. Variables with a selection frequency of at least 50% and a z-score larger than 1.96 in absolute value, corresponding to a two-tailed error rate of $p < 0.05$, were included into a cross-modality penalized Cox regression model to determine the relative importance of each modality to predict conversion from MCI to AD dementia.

2.4.3.4. Prediction accuracy for conversion status. Cross-validated prediction sensitivity, specificity, overall accuracy, and area under receiver operating characteristics curve (AUC) for multiple modality models were

compared between bidirectional stepwise unpenalized logistic regression and logistic regression with an elastic net penalty (optimal alpha was 0.5, and lambda minimizing the misclassification error). Random samples of 63.2% of the data were drawn 100 times to train the prediction models, the prediction accuracy of the resulting models was determined using the remaining 37.8% as test data. The set of variables eligible for the models came from two different selection pathways: (i) combination of all variables that were identified as relevant predictors in single modality penalized Cox regression, and (ii) combination of all variables that were identified as relevant predictors in single modality unpenalized bidirectional stepwise Cox regression.

As a non-linear comparison approach we used a support vector machine with a radial kernel and cross-validation with optimization of gamma and cost function parameter in the training sample through a grid search and application of the resulting support vector machine model to the test sample consisting of a random selection of one third of the cases. The cross-validation was iterated 100 times and levels of accuracy were determined and averaged across iterations. The support vector machine was implemented using function “svm” and related functions in package “e1071” in R (available at <http://cran.r-project.org/web/packages/e1071/index.html>) (Karatzoglou et al., 2006).

3. Results

3.1. Demographic characteristics

At follow-up, 39 individuals were MCI converters (MCI_c), and 88 individuals were MCI non-converters (MCI_{nc}). Four of the 88 nonconverters reverted to the status normal cognitive control, 84 remained MCI. Groups

did not differ with regard to sex distribution, age, years of education, time of follow up, overall cognitive performance as determined using the Mini Mental Status Examination (MMSE) (Folstein et al., 1975) score and episodic memory performance as determined using the delayed recall of the logical memory subtest of the Wechsler Memory Scale–Revised (Wechsler, 1987) and the Auditory Verbal Learning Test (Query and Berger, 1980) (Table 1).

3.2. Voxel based analysis

Voxel based analysis for AV45-PET data revealed higher amyloid load in MCI_c compared to MCI_{nc} in lateral temporal, prefrontal and parietal association cortex, with sparing of primary sensory motor and visual cortex as well as medial temporal lobes. The effects of AV45-PET with PVE correction were spatially much more restricted than the effects for PVE uncorrected AV45-PET data (Fig. 2a). We used family wise error correction with these data, because at FDR corrected level of significance almost the entire cortex was involved in the effects in the PVE uncorrected AV45-PET data. FDG-PET data showed a significant reduction of glucose consumption bilaterally in parietotemporal association cortex, posterior cingulate and precuneus as well as medial temporal lobe areas in MCI_c compared to MCI_{nc} (Fig. 2b) at an FDR corrected level of significance of $p < 0.05$, with more extended effects in the PVE corrected compared to the PVE uncorrected data. Cortical grey matter volume reductions in MCI_c compared to MCI_{nc} showed significant reductions in bilateral hippocampus and adjacent medio-temporal lobe areas, as well as left predominant parietotemporal lobe association cortex (Fig. 2c) at an FDR corrected level of significance of $p < 0.05$.

The coefficient of variation (CV) across all 42 grey matter regions was 0.50 for PVE corrected amyloid PET data, but only 0.22 for uncorrected PET data. In contrast, coefficients of variation were very similar in FDG-PET across the 42 grey matter regions before and after PVE correction (0.19 and 0.21, respectively).

Before PVE correction, voxel-wise mean CV was 0.11 for AV45-PET, but 0.21 for FDG-PET data across the entire white matter, and 0.27 for AV45-PET and 0.21 for FDG-PET across the entire grey matter.

3.3. Prediction models

The VIF was >14 for the PVE corrected and uncorrected AV45-PET, >8 for the uncorrected FDG-PET data, >5 for the PVE corrected FDG-PET data, and >2 for the regional grey matter volumes, suggesting that multicollinearity was a relevant issue with most of the data. As an example, Supplementary Fig. 2 shows the cross-correlation matrix and the vector of VIF values for the AV45-PET regional values with PVE correction.

All models showed minimal partial likelihood deviance at an alpha of 0.3 (Supplementary Fig. 1), suggesting that the penalty of the regression models was close to an elastic-net penalty ($\alpha = 0.5$). For each

single modality and the combined model, we determined the mean of the bootstrapped lambda values yielding minimum cross-validated error.

3.3.1. Single modality prediction of time to conversion

We determined frequency of selection, z-scores and cross-validated prediction accuracy across 42 regional values together with age, sex, education, and cognitive measures within each single modality. Results are summarized in Fig. 3. Coefficients were in the expected direction for all predictors in the AV45-PET data (higher amyloid load in the converters), FDG-PET data (smaller metabolism in the converters), and grey matter volumes (smaller volumes in converters).

For comparison, we used unpenalized Cox-regression with stepwise selection of predictors. These models obviously overfitted the data. For example, with PVE corrected AV45-PET data, the unpenalized Cox-regression model selected 23 predictor variables, including sex and education as well as 21 regional amyloid values (Fig. 4). Eleven of 21 regional amyloid predictors carried a negative sign, ten carried a positive sign (for example left and right amygdala carried opposite signs).

3.3.2. Multimodality prediction of time to conversion

In a subsequent analysis, we entered the relevant predictors of the single modality analyses into a combined model across modalities. Results are summarized in Fig. 3. Coefficients with significant z-scores came from PVE uncorrected AV45-PET data (z-score up to 7.5), followed by grey matter volumes (z-score up to 4.3), PVE corrected AV45-PET data (z-score up to 3.6) and corrected FDG-PET (z-score up to 3.0). Uncorrected FDG-PET did not contribute a significant region.

3.3.3. Prediction accuracy for conversion status

Results for penalized and unpenalized binary logistic regression models are shown in Fig. 5. For the combined relevant predictors from penalized single modality Cox regressions, prediction accuracy was higher in the test set with penalized compared to the unpenalized logistic regression model, with overall classification accuracy of 72% (SD 7) compared to 66% (SD 8), respectively.

For the more extensive set of combined relevant predictors from unpenalized single modality stepwise Cox regression, prediction accuracy was 70% (SD 6) for penalized logistic regression in the test set, but only 60% (SD 6) with unpenalized stepwise logistic regression.

For comparison, cross-validated support vector machine classification with a radial kernel yielded accuracy of 70% (SD 0.07), equal for both subsets of pre-selected predictor variables.

For the training data (i.e. no cross-validation), the unpenalized regression reached 100% accuracy, indicating an overfit of the training data that was, however, not present in the penalized regression (last row of panel in Fig. 5).

Table 1
Subjects' characteristics.

	Gender (F/M) ^a	Age (SD)[years] ^b	Education (SD) [years] ^c	Follow-up (SD) [months] ^d	MMSEMedian (25th and 75th percentile) ^e	LM-DRMedian (25th and 75th percentile) ^f	AVLTMedian (25th and 75th percentile) ^g
MCI_{nc}	47/41	72.4 (8.3)	16.6 (2.7)	17.9 (6.2)	29 (24.5, 29)	6 (1, 9)	2.5 (0, 8.5)
MCI_c	22/17	72.6 (8.1)	16.4 (2.8)	15.8 (7.3)	28 (25, 29)	8 (3.11)	3 (1, 10)

MCI_c – MCI converter, MCI_{nc} – MCI non-converter, F – female, M – male, MMSE – mini-mental status examination, LM-DR – delayed recall of the logical memory subtest of the Wechsler Memory Scale–Revised, AVLT – delayed recall from the Auditory Verbal Learning Test.

^a Not significantly different between groups, $\chi^2 = 0.10$, 1 df, $p = 0.75$.

^b Not significantly different between groups, Student's $T = -0.15$, 125 df, $p = 0.88$.

^c Not significantly different between groups, Student's $T = 0.43$, 125 df, $p = 0.67$.

^d Not significantly different between groups, Student's $T = 1.67$, 125 df, $p = 0.10$.

^e Not significantly different between groups, Mann–Whitney U test, $p = 0.72$.

^f Not significantly different between groups, Mann–Whitney U test, $p = 0.11$.

^g Not significantly different between groups, Mann–Whitney U test, $p = 0.29$.

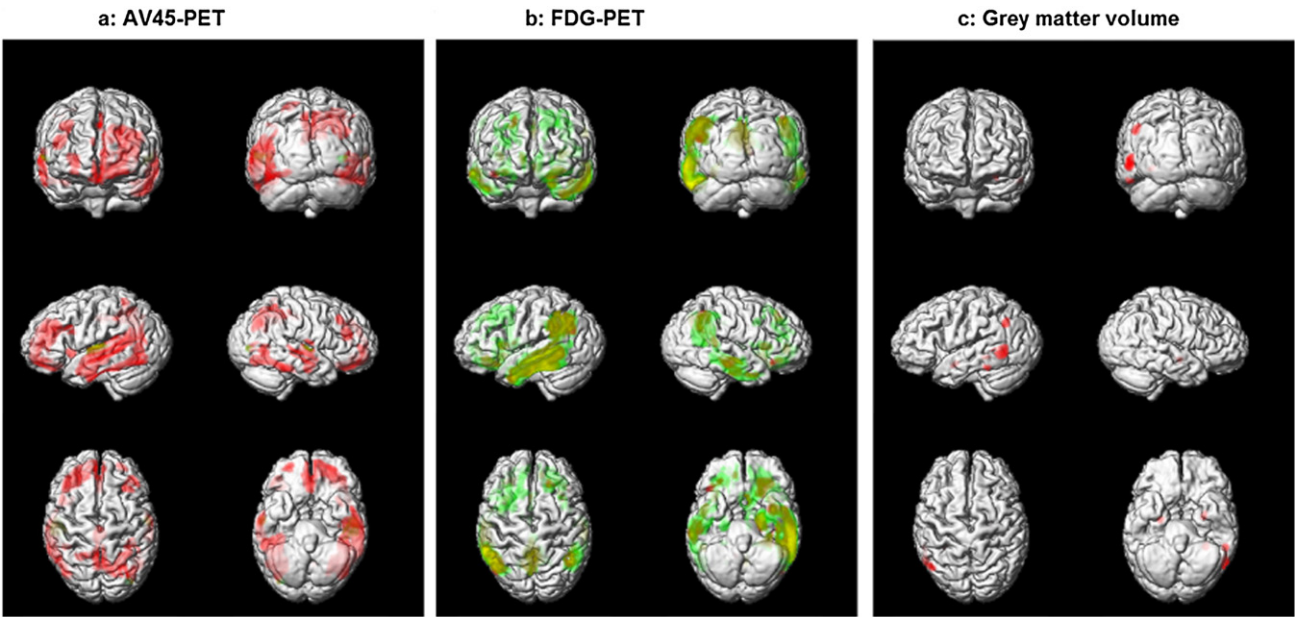


Fig. 2. Regional pattern of amyloid load, hypometabolism and grey matter atrophy in MCI_c vs. MCI_{nc}. Clusters of significant increase in AV45 uptake (Fig. 2a), significant decrease in ¹⁸F-FDG uptake (Fig. 2b) or significant decrease in grey matter volume (Fig. 2c) in MCI_c vs. MCI_{nc} projected onto the reference template in standard space. For PET data, red represents PVE uncorrected PET data, green represents PVE corrected PET data, and yellow indicates overlap. Please note that clusters are shown with at least 50 voxels passing the significance threshold of $p < 0.05$, FDR corrected, for FDG-PET and grey matter volume data, but with at least 50 voxels passing the significance threshold of $p < 0.05$, FWE corrected, for AV45-PET data, because at $p < 0.05$, FDR corrected, almost the entire supratentorial grey matter showed significant effects for AV45-PET uptake.

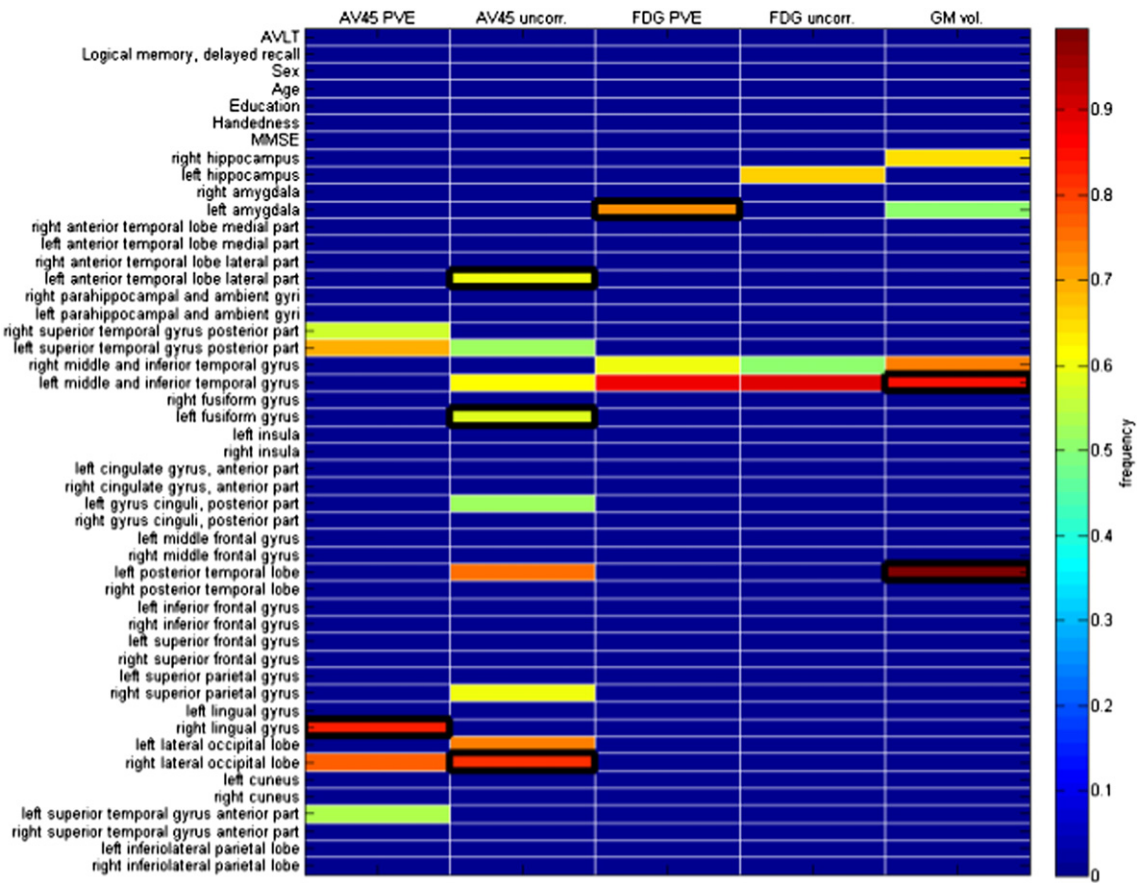


Fig. 3. Frequency of variable selection from penalized Cox regression. Colour coded intensity map for frequency of selection of regional values with the modalities of PVE corrected AV45-PET (AV45 PVE), non-PVE corrected AV45-PET (AV45 uncorr.), PVE corrected FDG-PET (FDG PVE), non-PVE corrected FDG-PET (FDG uncorr.), and grey matter volume (GM vol.). Squares with a black margin indicate those parameters that were selected as predictors in a cross-modality prediction model.

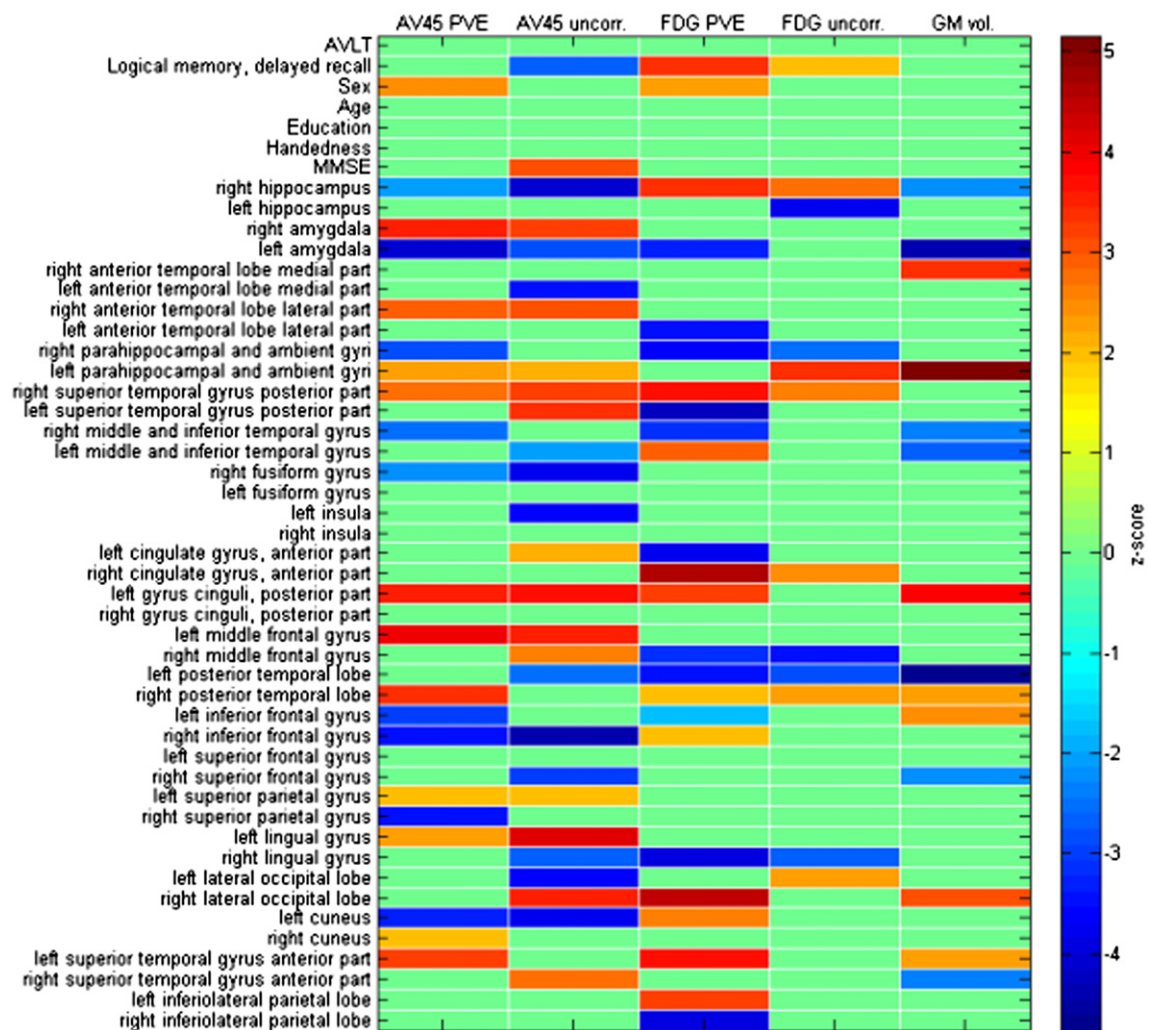


Fig. 4. Results of stepwise Cox regression. Colour coded intensity map for z-scores of selected regional values with the modalities of PVE corrected AV45-PET (AV45 PVE), non-PVE corrected AV45-PET (AV45 uncorr.), PVE corrected FDG-PET (FDG PVE), non-PVE corrected FDG-PET (FDG uncorr.), and grey matter volume (GM vol.) from stepwise Cox regression. Z-scores have been thresholded at an absolute value of $z > 1.96$, corresponding to a two-tailed error of $p < 0.05$.

4. Discussion

We compared penalized and unpenalized regression approaches to identify imaging parameters from amyloid PET, FDG-PET, and structural MRI that predict rapid to moderately fast conversion to dementia in MCI subjects. We had two major expectations: first, penalized regression would select a more parsimonious set of predictor variables than stepwise unpenalized regression. Secondly, prediction accuracy would be at least as high for penalized regression as for unpenalized regression.

Imaging studies on longitudinal follow-up of MCI individuals typically aim at two different endpoints: (i) maximization of prediction accuracy by a combination of parameters, and (ii) defining a model that provides insight into the relevance of different predictor variables.

In respect to *prediction accuracy*, with a large number of predictor variables it has been shown that by shrinking or setting to 0 some coefficients, prediction accuracy can be improved (Tibshirani, 2011). Using stepwise selecting procedures with unpenalized regression is an established approach to reduce the number of predictor variables. With a number of predictor variables that was lower than the number of cases, the prediction accuracy of bidirectional stepwise unpenalized regression was similar to the prediction accuracy of penalized regression despite high collinearity of predictors. When the number of predictors was higher than the number of cases used for training, stepwise regression yielded only 60% accuracy in the test data compared to 70%

for penalized regression; without bootstrapping, the bidirectional stepwise regression overfitted the outcome if the number of predictors was close to the number of cases (Fig. 5, left of last row).

For the multiple modalities model, the penalized Cox regression selected predictors primarily from AV45-PET data, less from FDG-PET data and grey matter atrophy. One previous study had used penalized logistic regression to determine the relative contribution of imaging modalities to prediction accuracy for MCI to AD conversion based on 50 MCI subjects from ADNI-1 (Trzepacz et al., 2014). Due to the low number of cases this previous study could only use leave one out cross-validation. In addition, the study did not take censoring of observations into account. Still, prediction accuracy levels were similar between our and this previous study, with up to 69% accuracy using cross-validation and 78% accuracy without cross-validation. Another study with ADNI-1 data used logistic regression to identify an optimum prediction model from a set of regressor variables, including structural MRI values, CSF Aβeta and tau levels, and neuropsychological test performance (Ewers et al., 2012). A correlation between predictors below $r = 0.9$ was considered sufficient to exclude collinearity between variables which is very liberal compared to the general recommendations from the literature (r below 0.7) (Booth et al., 1994; Dormann et al., 2013). Again, prediction accuracies were similar between this and our study, reaching up to 69% accuracy with cross-validation. Gaussian process classification, a machine learning approach, yielded a cross-

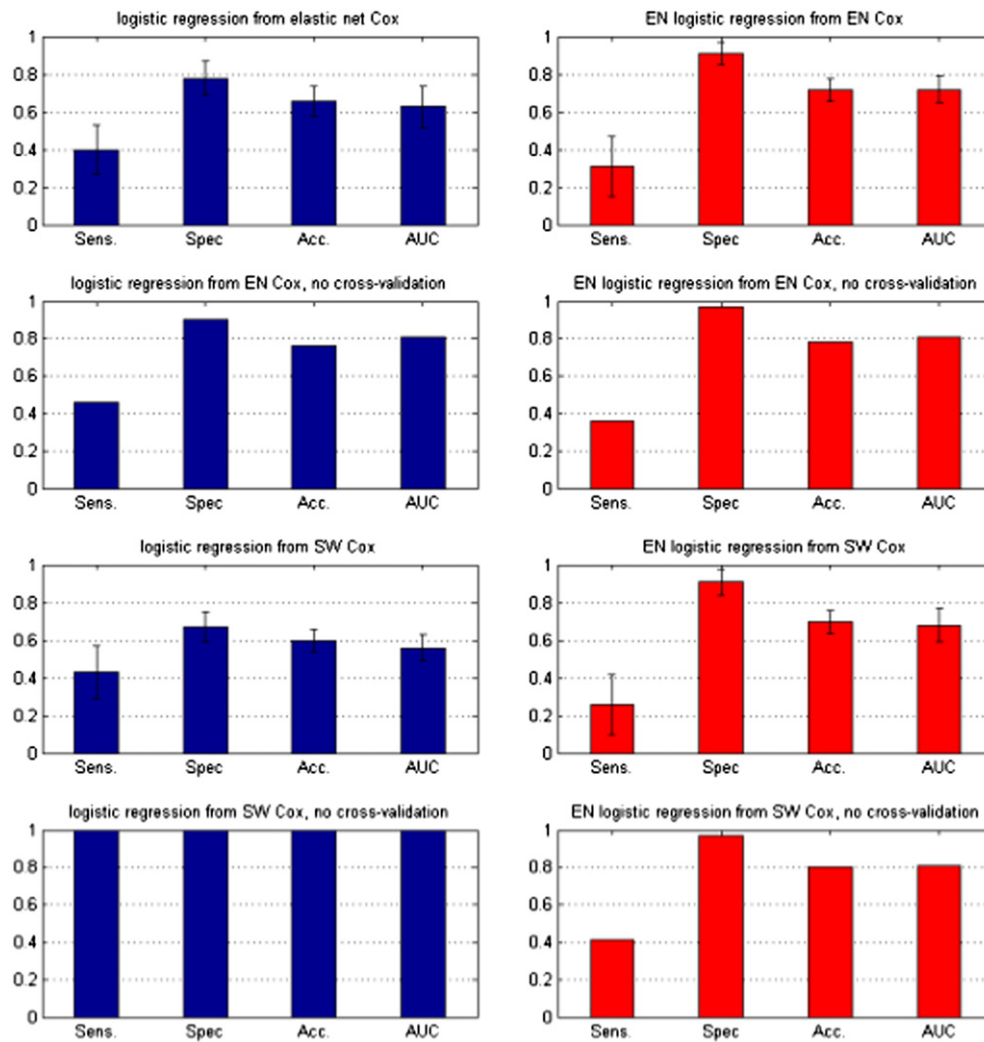


Fig. 5. Prediction accuracies without and with elastic net penalty. Bar diagram of sensitivity, specificity and overall accuracy to predict conversion status, MCI_c vs. MCI_{nc}. Bars on the left indicate results for unpenalized logistic regression, on the right from penalized logistic regression with an elastic net (EN) of $\alpha = 0.5$. In the first row, predictor variables have been identified from the within modality penalized Cox regression models and combined across modalities, using bootstrapped cross-validation. In the second row, predictor variables have been identified from the within modality penalized Cox regression models and combined across modalities without cross-validation (i.e. fit in the training sample). In the third row, predictor variables have been identified from the within modality unpenalized stepwise (SW) Cox regression models and combined across modalities, using bootstrapped cross-validation. The stepwise bidirectional selection process is not possible when the number of parameters (p) is larger than the number of cases (n). In the fourth row, predictor variables have been identified from the within modality unpenalized stepwise (SW) Cox regression models and combined across modalities without cross-validation (i.e. fit in the training sample). AUC — area under the receiver operating characteristics (ROC) curve.

validated prediction accuracy of up to 69% based on an ADNI-1 sample with FDG-PET, structural MRI, and CSF Aβeta and tau levels as predictors (Young et al., 2013). In contrast to our penalized Cox regression, the Gaussian process classification cannot consider time to conversion as endpoint. Within our study, unpenalized and penalized logistic regression analysis yielded similar prediction accuracy when using a smaller set of preselected predictors and bootstrap-based validation. These data suggest that penalized regression becomes more efficient than unpenalized regression when the number of predictor variables approximates or is larger than the number of cases in the training sample. The level of accuracy of penalized logistic regression was also comparable with the accuracy of cross-validated support vector machine classification using a non-linear radial kernel.

The aim to identify which modality and which regional values within a modality may be *relevant predictors* is strongly influenced by the collinearity of predictor variables. For all modalities, the penalized Cox regression yielded more parsimonious prediction models than the unpenalized Cox regression with stepwise feature selection. Although parsimony of a model is an attractive feature in itself, this alone does not guarantee that the model is neurobiologically more plausible than

a less parsimonious model. The signs of the regression coefficients varied largely between corresponding regions from both hemispheres in the unpenalized regression, but were homogeneous within the penalized regression model. These findings suggest that penalized regression analysis provides a neurobiologically plausible set of predictor variables in the presence of a high number of collinear predictor variables; at the same time it achieves at least equal levels of prediction accuracy compared to unpenalized regression. However, the final test of neurobiological plausibility requires replication of findings in independent cohorts and consistency of findings with neurobiological models of disease pathogenesis, which needs to be explored in future studies.

When combining prediction variables across modalities, PVE uncorrected amyloid PET came out as the single most prominent modality, both in terms of numbers of contributing regions as well as z-scores, followed by grey matter volume, with only a minor contribution from FDG-PET. Previous studies on multimodal prediction of MCI conversion have investigated different marker combinations, including FDG-PET, amyloid PET, atrophy, CSF tau and Aβeta, as well as neuropsychological scores. A recent Swedish study showed cognitive performance in the trail making test (TMT)-B as best single predictor for conversion into

dementia in MCI (Eckerström et al., 2014). Total tau levels in CSF were superior to Abeta42 levels in predicting AD dementia. This previous study, however, used no cross-validation of prediction accuracy, rendering the findings uninformative with respect to prediction accuracy in a test sample. Another study, based on ADNI-1 data, used cross-validation of the findings and showed hippocampus volume as well as episodic memory performance as best predictors of time to conversion in a stepwise Cox regression (Ewers et al., 2012). Also based on ADNI-1 data, a Gaussian process classification found FDG-PET and structural MRI as superior predictors compared to CSF tau and Abeta levels (Young et al., 2013), but this study did not take censoring into account. The study by Trzepacz et al. (Trzepacz et al., 2014), using elastic net logistic regression with leave one out cross-validation, found that amyloid imaging using PIB-PET and MRI based volumetry were the most relevant predictors with only a marginal effect of FDG-PET. This result is similar to our findings in the ADNI-2 sample after taking censoring into account.

The superior effect of amyloid PET compared to the other imaging modalities agrees with the role of amyloid positivity for disease progression in MCI. Particularly, as shown in the high specificity, absence of amyloid accumulation was an accurate predictor for MCI subjects to remain stable during clinical follow-up. This agrees with previous studies that showed a very low risk for conversion into AD dementia in amyloid negative MCI individuals (Grimmer et al., 2013; Hatashita and Yamasaki, 2013; Okello et al., 2009; Wolk et al., 2009). In addition, markers of neuronal injury, such as regional atrophy or CSF tau levels increase accuracy of prediction of time to conversion in amyloid positive MCI subjects (van Rossum et al., 2012).

In respect to regional distribution of findings, the penalized regression model retrieved regional amyloid load not only in early affected regions, such as posterior cingulate (Camus et al., 2012), but also in lateral temporal cortex, fusiform gyrus, superior parietal cortex and occipital cortex, areas that begin to build up amyloid relatively late in the course of AD (Braak and Braak, 1991). This agrees with the assumption that areas with early build-up of amyloid changes, already reaching a plateau in the MCI stage, would add little to the discrimination of early from advanced stages of MCI, whereas areas with later build-up of amyloid load such as posterior brain areas would be more relevant to detect advanced stages of MCI with possibly imminent conversion, because the build-up of amyloid in these regions is only at its beginning before this stage of disease. In respect to hypometabolism, among others, changes in medial temporal lobes contributed to the penalized prediction model. This finding agrees with a range of previous studies as summarized in a recent review (Mosconi, 2013). At the same time, however, as pointed out in the same review, hypometabolism of medial temporal lobe dementia is less consistently reported than hypometabolism of posterior cingulate gyrus in MCI and AD, possibly related to a mixture of biological, technical and image analysis issues that render this region more susceptible to measurement variability than cortical brain areas.

To further explore the influence of partial volume effects on the predictive value of AV45- and FDG-PET data, we assessed voxel-wise group differences and predictive regression models separately for PVE corrected and non-corrected PET data. Although based on theoretical considerations PVE correction of PET data should significantly increase the correspondence of the image signal with the true neuronal tracer uptake, MRI-based PVE correction methods also depend on a range of model assumptions that may not always hold true or may only be roughly approximated in the imaging data (Erlandsson et al., 2012; Thomas et al., 2011). Thus, PVE correction of AV45- and FDG-PET data is not uniformly being employed across imaging studies, and the effect of PVE correction on the predictive value of these imaging modalities has received little attention so far.

Non-corrected AV45-PET data showed more precise group separation between converters and non-converters than PVE corrected AV45-PET, both in the Cox regression as well as the voxel-based comparison. This agrees with a previous study, showing an increase of variability of amyloid values after PVE correction (Thomas et al., 2011). The

authors of the previous study argued that the observed increase of variability of amyloid values in clinical amyloid PET scans may reflect true variability of amyloid uptake. In our data, the coefficient of variation across all 42 regions was twice as high for PVE corrected compared to uncorrected amyloid PET data. These results suggest that the PVE correction related increase of variability degrades prediction accuracy of AV45-PET data. One could speculate that PVE correction-induced increases in variability of amyloid values may be due to the removal of the relatively high and homogeneous signal of unspecific binding of the ^{18}F -labelled amyloid tracer in the white matter (Clark et al., 2011) from the measured signal, leaving the more variable grey matter signal. This assumption agrees with our observation that the white matter coefficient of variation (CV) was much smaller than the grey matter coefficient of variation (CV 0.11 vs. 0.27) for AV45-PET data, and also much smaller than the variability of the FDG-PET signal in the white matter (CV = 0.21). Direct comparisons of PVE corrected and non-corrected amyloid PET imaging studies are still rare so that our findings warrant further replication.

In contrast, FDG-PET showed larger between group differences after PVE correction than before PVE correction. PVE correction of FDG-PET data has been shown to affect group differences between AD dementia subjects and healthy elderly controls to a limited degree (Bokde et al., 2001; Ibanez et al., 1998), but to increase the specificity of regional distribution of hypometabolism for cognitive deficits (Bokde et al., 2005). In addition, PVE correction of FDG-PET has been shown to increase the sensitivity to detect early hippocampal hypometabolism in MCI subjects (Mevel et al., 2007). Coefficients of variation were very similar before and after PVE correction (0.19 and 0.21, respectively) in our data. Therefore, the higher sensitivity for hypometabolism in predementia disease stages and the specificity for cognitive impairment of PVE corrected FDG-PET may contribute to its higher precision to discriminate between MCI converters and stable MCI compared to non-corrected FDG-PET data.

There are limitations associated with our study. First, as stated in the [Material and methods section](#), no statistical testing has been established for coefficient estimates from penalized regression with exception of a recently discovered significance test for the Lasso with continuous endpoints (Lockhart et al., 2014). Use of bootstrapping to assess z-scores of parameter distribution is not taking bias into account which is, however, higher in penalized regression (providing a trade-off between variance and bias) than in unpenalized regression. Counting the frequency of coefficient selection is an ad hoc solution to estimate the generalizability of the model and serves as an indicator of the relevance of each parameter. Statistical testing for coefficients from classical stepwise regression approaches, however, is not finally resolved either. Although software packages for stepwise regression print out tables on standard errors and p-values of selected variables, these values are not validly defined, since the standard errors are not being adjusted for the search process ((Hastie et al., 2009), page 60). This leads to too narrow estimates of confidence intervals in stepwise regression (Altman and Andersen, 1989).

Secondly, the number of converters was only half the number of non-converters. Therefore, the selection of the classification threshold (here 50%) for group classification in the logistic regression framework may influence the achievable levels of accuracy. To control for such an effect, we estimated areas under the receiver operating characteristic curves (that consider all possible classification thresholds) that suggested no major effect of threshold selection on accuracy levels.

Thirdly, different to a range of previous studies (Eckerström et al., 2014; Ewers et al., 2012; Trzepacz et al., 2014) global cognition as measured using the MMSE score and episodic memory performance were not different between MCI converters and non-converters in our sample. The criterion for 'late MCI' in the ADNI framework, as used in our analyses, is a more advanced delayed recall impairment compared with 'early MCI' subjects. This likely leads to a floor effect of delayed recall performance in the late MCI subjects. The lack of a difference,

however, offers the opportunity to assess the contribution of imaging markers to prediction accuracy independently of baseline differences in global cognition or episodic memory performance between converters and non-converters.

In summary, we found accurate cross-validated prediction of conversion status and time to conversion in a relatively large sample of MCI converters and non-converters between 6 to 31 months of follow-up. Penalized Cox and logistic regression yielded more parsimonious models than unpenalized stepwise regression with maintained or even more precise prediction performance. The advantages of penalized regression models became particularly strong with a high number of predictors (relative to the number of cases) and high collinearity between predictors. PVE correction, a widely discussed but still rarely used preprocessing approach for PET data, showed detrimental effects on group discrimination for AV45-PET, but beneficial effects for FDG-PET both in penalized logistic regression and simple linear models. From a clinical point of view, the differential effect of PVE correction on AV45-PET and FDG-PET data should be considered when deciding about the use of PVE correction in the context of predictive PET studies.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann–La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (<http://www.fnih.org>). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuroimaging at the University of Southern California.

Appendix A. Supplementary data

Supplementary material for this article can be found online at <http://dx.doi.org/10.1016/j.nicl.2015.05.006>.

References

Altman, D.G., Andersen, P.K., 1989. Bootstrap investigation of the stability of a Cox regression model. *Stat. Med.* 8 (7), 771–783. <http://dx.doi.org/10.1002/sim.47800807022672226>.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. *Neuroimage* 38 (1), 95–113. <http://dx.doi.org/10.1016/j.neuroimage.2007.07.00717761438>.

Belsley, D.A., 1991. *Conditioning Diagnostics: Collinearity and Weak Data in Regression*. John Wiley & Sons, Chichester, UK.

Bokde, A.L., Pietrini, P., Ibáñez, V., Furey, M.L., Alexander, G.E., Graff-Radford, N.R., Rapoport, S.I., Schapiro, M.B., Horwitz, B., 2001. The effect of brain atrophy on cerebral hypometabolism in the visual variant of Alzheimer disease. *Arch. Neurol.* 58 (3), 480–486. <http://dx.doi.org/10.1001/archneur.58.3.48011255453>.

Bokde, A.L., Teipel, S.J., Drzezga, A., Thissen, J., Bartenstein, P., Dong, W., Leinsinger, G., Born, C., Schwaiger, M., Moeller, H.J., Hampel, H., 2005. Association between cognitive performance and cortical glucose metabolism in patients with mild Alzheimer's disease. *Dement. Geriatr. Cogn. Disord.* 20 (6), 352–357. <http://dx.doi.org/10.1159/00008855816192725>.

Booth, G.D., Niccolucci, M.J., Schuster, E.G., 1994. Identifying proxy sets in multiple linear regression — an aid to better coefficient interpretation. *USDA Forest Serv. Intermountain Res. Stn. Res. Pap.* 1–13.

Braak, H., Braak, E., 1991. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* 82 (4), 239–259. <http://dx.doi.org/10.1007/BF003088091759558>.

Camus, V., Payoux, P., Barré, L., Desgranges, B., Voisin, T., Tauber, C., La Joie, R., Tafani, M., Hommet, C., Chételat, G., Mondon, K., de la Sayette, V., Cottier, J.P., Beauvais, E., Ribeiro, M.J., Gissot, V., Vierron, E., Vercoillie, J., Vellas, B., Eustache, F., Guilleaume, D., 2012. Using PET with 18F-AV-45 (florbetapir) to quantify brain amyloid load in a clinical environment. *Eur. J. Nucl. Med. Mol. Imaging* 39 (4), 621–631. <http://dx.doi.org/10.1007/s00259-011-2021-822252372>.

Chen, X., Li, M., Wang, S., Zhu, H., Xiong, Y., Liu, X., 2014. Pittsburgh compound B retention and progression of cognitive status — a meta-analysis. *Eur. J. Neurol.* 21 (8), 1060–1067. <http://dx.doi.org/10.1111/ene.1239824612390>.

Clark, C.M., Schneider, J.A., Bedell, B.J., Beach, T.G., Bilker, W.B., Mintun, M.A., Pontecorvo, M.J., Hefti, F., Carpenter, A.P., Flitter, M.L., Krautkramer, M.J., Kung, H.F., Coleman, R.E., Doraiswamy, P.M., Fleisher, A.S., Sabbagh, M.N., Sadowsky, C.H., Reiman, E.P., Zehntner, S.P., Skovronsky, D.M., 2011. Use of florbetapir-PET for imaging beta-amyloid pathology. *JAMA* 305 (3), 275–283. <http://dx.doi.org/10.1001/jama.2010.200821245183>.

Cohen, A.D., Klunk, W.E., 2014. Early detection of Alzheimer's disease using PiB and FDG PET. *Neurobiol. Dis.* 72, 117–122. <http://dx.doi.org/10.1016/j.nbd.2014.05.00124825318>.

Cox, D.R., Oakes, D., 1984. *Analysis of Survival Data*. Chapman and Hall, London, New York.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36 (1), 27–46. <http://dx.doi.org/10.1111/j.1600-0587.2012.07348.x>.

Duara, R., Loewenstein, D.A., Greig, M.T., Potter, H., Barker, W., Raj, A., Schinka, J., Borenstein, A., Schoenberg, M., Wu, Y., Banko, J., Potter, H., 2011. Pre-MCI and MCI: neuropsychological, clinical, and imaging features and progression rates. *Am. J. Geriatr. Psychiatry* 19 (11), 951–960. <http://dx.doi.org/10.1097/JGP.0b013e3182107c6921422909>.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. *Pattern Classification second edition*. John Wiley & Sons, New York, NY.

Eckerström, C., Olsson, E., Klasson, N., Berge, J., Nordlund, A., Bjerke, M., Wallin, A., 2015. Multimodal prediction of dementia with up to 10 years follow up: the Gothenburg MCI Study. *J. Alzheimers Dis.* 44 (1), 205–214. <http://dx.doi.org/10.3233/JAD-14105325201779>.

Erlandsson, K., Buvat, I., Pretorius, P.H., Thomas, B.A., Hutton, B.F., 2012. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. *Phys. Med. Biol.* 57 (21), R119–R159. <http://dx.doi.org/10.1088/0031-9155/57/21/R11923073343>.

Ewers, M., Walsh, C., Trojanowski, J.Q., Shaw, L.M., Petersen, R.C., Jack Jr., C.R., Feldman, H.H., Bokde, A.L., Alexander, G.E., Scheltens, P., Vellas, B., Dubois, B., Weiner, M., Hampel, H., North American Alzheimer's Disease Neuroimaging Initiative, 2012. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based on biomarkers and neuropsychological test performance. *Neurobiol. Aging* 33 (7), 1203–1214. <http://dx.doi.org/10.1016/j.neurobiolaging.2010.10.01921159408>.

Farrar, D.E., Glauber, R.R., 1967. Multicollinearity in regression analysis: the problem revisited. *Rev. Econ. Stat.* 49 (1), 92–107. <http://dx.doi.org/10.2307/1937887>.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. Mini-mental-state: a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12 (3), 189–198. [http://dx.doi.org/10.1016/0022-3956\(75\)90026-61202204](http://dx.doi.org/10.1016/0022-3956(75)90026-61202204).

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33 (1), 1–2220808728.

Grimmer, T., Wutz, C., Drzezga, A., Förster, S., Förstl, H., Ortner, M., Perneczky, R., Kurz, A., 2013. The usefulness of amyloid imaging in predicting the clinical outcome after two years in subjects with mild cognitive impairment. *Curr. Alzheimer Res.* 10 (1), 82–8523036071.

Grothe, M., Heinsen, H., Teipel, S., 2013. Longitudinal measures of cholinergic forebrain atrophy in the transition from healthy aging to Alzheimer's disease. *Neurobiol. Aging* 34 (4), 1210–1220. <http://dx.doi.org/10.1016/j.neurobiolaging.2012.10.01823158764>.

Hammers, A., Allom, R., Koeppe, M.J., Free, S.L., Myers, R., Lemieux, L., Mitchell, T.N., Brooks, D.J., Duncan, J.S., 2003. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum. Brain Mapp.* 19 (4), 224–247. <http://dx.doi.org/10.1002/hbm.1012312874777>.

Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning — Data Mining Inference, and Prediction second edition*. Springer, New York.

Hatashita, S., Yamasaki, H., 2013. Diagnosed mild cognitive impairment due to Alzheimer's disease with PET biomarkers of beta amyloid and neuronal dysfunction. *PLOS One* 8 (6), e66877. <http://dx.doi.org/10.1371/journal.pone.006687723799136>.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression — biased estimation for nonorthogonal problems. *Technometrics* 12 (1), 55–67. <http://dx.doi.org/10.1080/00401706.1970.10488634>.

Ibáñez, V., Pietrini, P., Alexander, G.E., Furey, M.L., Teichberg, D., Rajapakse, J.C., Rapoport, S.I., Schapiro, M.B., Horwitz, B., 1998. Regional glucose metabolic abnormalities are not the result of atrophy in Alzheimer's disease. *Neurol.* 50 (6), 1585–1593. <http://dx.doi.org/10.1212/WNL.50.6.15859633698>.

Jack Jr., C.R., 2012. Alzheimer disease: new concepts on its neurobiology and the clinical role imaging will play. *Radiology* 263 (2), 344–361. <http://dx.doi.org/10.1148/radiol.1211043322517954>.

Karatzoglou, A., Meyer, D., Hornik, K., 2006. Support vector machines in R. *J. Stat. Softw.* 15.

Lockhart, R., Taylor, J., Tibshirani, R.J., Tibshirani, R., 2014. A significance test for the lasso. *Ann. Statist.* 42 (2), 413–468. <http://dx.doi.org/10.1214/13-AOS1175>.

- Mevel, K., Desgranges, B., Baron, J.C., Landeau, B., De la Sayette, V., Viader, F., Eustache, F., Chételat, G., 2007. Detecting hippocampal hypometabolism in mild cognitive impairment using automatic voxel-based approaches. *Neuroimage* 37 (1), 18–25. <http://dx.doi.org/10.1016/j.neuroimage.2007.04.04817566762>.
- Mosconi, L., 2013. Glucose metabolism in normal aging and Alzheimer's disease: methodological and physiological considerations for PET studies. *Clin. Trans. Imaging* 1.
- Müller-Gärtner, H.W., Links, J.M., Prince, J.L., Bryan, R.N., McVeigh, E., Leal, J.P., Davatzikos, C., Frost, J.J., 1992. Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects. *J. Cerebr. Blood Flow Metab.* 12 (4), 571–583. <http://dx.doi.org/10.1038/jcbfm.1992.811618936>.
- Okello, A., Koivunen, J., Edison, P., Archer, H.A., Turkheimer, F.E., Nägren, K., Bullock, R., Walker, Z., Kennedy, A., Fox, N.C., Rossor, M.N., Rinne, J.O., Brooks, D.J., 2009. Conversion of amyloid positive and negative MCI to AD over 3 years: an 11C—PIB PET study. *Neurology* 73 (10), 754–760. <http://dx.doi.org/10.1212/WNL.0b013e3181b2356419587325>.
- Query, W.T., Berger, R.A., 1980. AVLT memory scores as a function of age among general medical, neurologic and alcoholic patients. *J. Clin. Psychol.* 36 (4), 1009–1012. [http://dx.doi.org/10.1002/1097-4679\(198010\)36:4<1009::AID-JCLP2270360433>3.0.CO;2-N7440727](http://dx.doi.org/10.1002/1097-4679(198010)36:4<1009::AID-JCLP2270360433>3.0.CO;2-N7440727).
- Teipel, S.J., Grothe, M., Lista, S., Toschi, N., Garaci, F.G., Hampel, H., 2013. Relevance of magnetic resonance imaging for early detection and diagnosis of Alzheimer disease. *Med. Clin. North Am.* 97 (3), 399–424. <http://dx.doi.org/10.1016/j.mcna.2012.12.01323642578>.
- Thomas, B.A., Erlandsson, K., Modat, M., Thurfjell, L., Vandenberghe, R., Ourselin, S., Hutton, B.F., 2011. The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease. *Eur. J. Nucl. Med. Mol. Imaging* 38 (6), 1104–1119. <http://dx.doi.org/10.1007/s00259-011-1745-921336694>.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B Stat. Methodol.* 58, 267–288.
- Tibshirani, R., 2011. Regression shrinkage and selection via the lasso: a retrospective. *J. R. Stat. Soc. B Stat. Methodol.* 73 (3), 273–282. <http://dx.doi.org/10.1111/j.1467-9868.2011.00771.x>.
- Trzepacz, P.T., Yu, P., Sun, J., Schuh, K., Case, M., Witte, M.M., Hochstetler, H., Hake, A., Alzheimer's Disease Neuroimaging Initiative, 2014. Comparison of neuroimaging modalities for the prediction of conversion from mild cognitive impairment to Alzheimer's dementia. *Neurobiol. Aging* 35 (1), 143–151. <http://dx.doi.org/10.1016/j.neurobiolaging.2013.06.01823954175>.
- van Rossum, I.A., Vos, S.J., Burns, L., Knol, D.L., Scheltens, P., Soininen, H., Wahlund, L.O., Hampel, H., Tsolaki, M., Minthon, L., L'Italien, G., van der Flier, W.M., Teunissen, C.E., Blennow, K., Barkhof, F., Rueckert, D., Wolz, R., Verhey, F., Visser, P.J., 2012. Injury markers predict time to dementia in subjects with MCI and amyloid pathology. *Neurology* 79 (17), 1809–1816. <http://dx.doi.org/10.1212/WNL.0b013e318270405623019259>.
- Wechsler, D., 1987. *Wechsler Memory Scale — Revised*. The Psychological Corporation, New York.
- Wolk, D.A., Price, J.C., Saxton, J.A., Snitz, B.E., James, J.A., Lopez, O.L., Aizenstein, H.J., Cohen, A.D., Weissfeld, L.A., Mathis, C.A., Klunk, W.E., De-Kosky, S.T., DeKoskym, S.T., 2009. Amyloid imaging in mild cognitive impairment subtypes. *Ann. Neurol.* 65 (5), 557–568. <http://dx.doi.org/10.1002/ana.2159819475670>.
- Young, J., Modat, M., Cardoso, M.J., Mendelson, A., Cash, D., Ourselin, S., Alzheimer's Disease Neuroimaging Initiative, 2013. Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neuroimage Clin.* 2, 735–745. <http://dx.doi.org/10.1016/j.nicl.2013.05.00424179825>.
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *J. R. Stat. Soc.* 67 (2), 301–320. <http://dx.doi.org/10.1111/j.1467-9868.2005.00503.x>.
- Zou, H., Zhang, H.H., 2009. On the adaptive elastic-net with a diverging number of parameters. *Ann. Stat.* 37 (4), 1733–1751. <http://dx.doi.org/10.1214/08-AOS62520445770>.