Integrative EEG biomarkers predict progression to Alzheimer’s disease at the MCI stage

Poil, Simon-Shlomo; de Haan, Willem; van der Flier, Wiesje M; Mansvelder, Huibert D; Scheltens, Philip; Linkenkaer-Hansen, Klaus

Abstract: Alzheimer’s disease (AD) is a devastating disorder of increasing prevalence in modern society. Mild cognitive impairment (MCI) is considered a transitional stage between normal aging and AD; however, not all subjects with MCI progress to AD. Prediction of conversion to AD at an early stage would enable an earlier, and potentially more effective, treatment of AD. Electroencephalography (EEG) biomarkers would provide a non-invasive and relatively cheap screening tool to predict conversion to AD; however, traditional EEG biomarkers have not been considered accurate enough to be useful in clinical practice. Here, we aim to combine the information from multiple EEG biomarkers into a diagnostic classification index in order to improve the accuracy of predicting conversion from MCI to AD within a 2-year period. We followed 86 patients initially diagnosed with MCI for 2 years during which 25 patients converted to AD. We show that multiple EEG biomarkers mainly related to activity in the beta-frequency range (13-30 Hz) can predict conversion from MCI to AD. Importantly, by integrating six EEG biomarkers into a diagnostic index using logistic regression the prediction improved compared with the classification using the individual biomarkers, with a sensitivity of 88% and specificity of 82%, compared with a sensitivity of 64% and specificity of 62% of the best individual biomarker in this index. In order to identify this diagnostic index we developed a data mining approach implemented in the Neurophysiological Biomarker Toolbox (http://www.nbtwiki.net/). We suggest that this approach can be used to identify optimal combinations of biomarkers (integrative biomarkers) also in other modalities. Potentially, these integrative biomarkers could be more sensitive to disease progression and response to therapeutic intervention.

DOI: https://doi.org/10.3389/fnagi.2013.00058

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-84961

Originally published at:
Poil, Simon-Shlomo; de Haan, Willem; van der Flier, Wiesje M; Mansvelder, Huibert D; Scheltens, Philip; Linkenkaer-Hansen, Klaus (2013). Integrative EEG biomarkers predict progression to Alzheimer’s disease at the MCI stage. Frontiers in Aging Neuroscience, 5:58.
DOI: https://doi.org/10.3389/fnagi.2013.00058
Integrative EEG biomarkers predict progression to Alzheimer’s disease at the MCI stage

Simon-Shlomo Poil1, Willem de Haan2,3, Wiesje M. van der Flier2,4, Huibert D. Mansvelder1, Philip Scheltens2 and Klaus Linkenkaer-Hansen1*

1 Department of Integrative Neurophysiology, Center for Neurogenomics and Cognitive Research, VU University Amsterdam, Amsterdam, Netherlands
2 Department of Clinical Neurophysiology and MEG, VU University Medical Center, Amsterdam, Netherlands
3 Department of Neurology, Alzheimer Center, VU University Medical Center, Amsterdam, Netherlands
4 Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, Netherlands

Alzheimer’s disease (AD) is a devastating disorder of increasing prevalence in modern society. Mild cognitive impairment (MCI) is considered a transitional stage between normal aging and AD; however, not all subjects with MCI progress to AD. Prediction of conversion to AD at an early stage would enable an earlier, and potentially more effective, treatment of AD. Electroencephalography (EEG) biomarkers would provide a non-invasive and relatively cheap screening tool to predict conversion to AD; however, traditional EEG biomarkers have not been considered accurate enough to be useful in clinical practice. Here, we aim to combine the information from multiple EEG biomarkers into a diagnostic classification index in order to improve the accuracy of predicting conversion from MCI to AD within a 2-year period. We followed 86 patients initially diagnosed with MCI for 2 years during which 25 patients converted to AD. We show that multiple EEG biomarkers mainly related to activity in the beta-frequency range (13–30 Hz) can predict conversion from MCI to AD. Importantly, by integrating six EEG biomarkers into a diagnostic index using logistic regression the prediction improved compared with the classification using the individual biomarkers, with a sensitivity of 88% and specificity of 82%, compared with a sensitivity of 64% and specificity of 62% of the best individual biomarker in this index. In order to identify this diagnostic index we developed a data mining approach implemented in the Neurophysiological Biomarker Toolbox (http://www.nbtwiki.net/). We suggest that this approach can be used to identify optimal combinations of biomarkers (integrative biomarkers) also in other modalities. Potentially, these integrative biomarkers could be more sensitive to disease progression and response to therapeutic intervention.

Keywords: Neurophysiological Biomarkers, Alzheimer’s disease, mild cognitive impairment (MCI), electroencephalography, predictive analysis, time series analysis, eyes closed resting state

INTRODUCTION

Caused by an increasing average age of the population in the developed world, dementia is becoming a major healthcare problem. Alzheimer’s disease is the most common form of dementia and the golden standard for diagnosis is the post-mortem identification of Amyloid Beta 42 depositions and tangles (Blennow et al., 2006; Herrup, 2010). It has been suggested that Alzheimer’s disease begins years, maybe even decades before actual cognitive symptoms appear (Sperling et al., 2011). However, normal aging is also characterized by a slow decline of cognitive functions, which means it can be difficult to disentangle normal ageing from Alzheimer at a very early stage.

Patients with mild cognitive impairment (MCI) are at high risk of developing Alzheimer’s disease. The label MCI is given when there is a cognitive complaint (mostly memory), which can also be demonstrated on formal testing, while general cognitive functioning is relatively intact and a patient is still living independently (Flicker et al., 1991; Gauthier et al., 2006; Albert et al., 2011). Therapies that stop the conversion to Alzheimer’s disease unfortunately remain to be developed, but it is likely that these drugs or therapies will appear in the future (Prins et al., 2010; Huang and Mucke, 2012). It is plausible that these therapies will be most effective before major brain damage has occurred and it is, therefore, important to develop biomarkers sensitive of this very early stage (Sperling et al., 2011). Early-stage identification may also help the development of new treatments that are more effective at this stage as it can facilitate monitoring of the response to the intervention.

We here focus on biomarkers obtained from electroencephalography (EEG) recordings in the eyes-closed resting state (ECR). EEG biomarkers are optimal for screening purposes because the EEG recording can be obtained using relative cheap and non-invasive equipment, which is widely available and fast to use. Several previous EEG studies of conversion from mild cognitive impairment to Alzheimer’s disease have been conducted (Flicek et al., 1996, 2000; Huang et al., 2000; Stam et al., 2003; Schoonenboom et al., 2004; Rombouts et al., 2005; Babiloni et al., 2006, 2011; Kwak, 2006; Rossini et al., 2006, 2008; Lehmann et al., 2007; Moretti et al., 2007a,b, 2008, 2011; Luckhaus et al., 2008) mainly using biomarkers such as spectral measures and...
synchronization between brain regions. Machine-learning techniques have been used to explore differences between MCI and AD with varying success (Huang et al., 2000; Bennys et al., 2001; Prichep et al., 2006; Buscema et al., 2007; Lehmann et al., 2007; Prichep, 2007; Rossini et al., 2008), however, only few studies have tried to predict the conversion from MCI to AD (Prichep et al., 2006; Prichep, 2007; Antila et al., 2013). Many studies typically focus on a small number of biomarkers (on the order of 15 marker values), and some do not have adequate validation of their results on independent groups. We perform large-scale data mining of multiple biomarkers (Figure 1A) and validate our results on an independent group of subjects.

Our focus is on the EEG measured as part of the initial hospital intake test, combined with longitudinal recordings measured 1 year after the initial intake test. We have mapped several classical EEG biomarkers, such as frequency and power, but also non-classical biomarkers such as detrended fluctuation analysis and oscillation burst analysis (Poil et al., 2008; Montez et al., 2009). By combining several biomarkers, it is often possible to find better separation boundaries between two groups (Figure 1C), because each biomarker gives additional information (Lehmann et al., 2007). In this longitudinal study we show that EEG biomarkers from the initial hospital in-take test retrospectively can be used in a classifier algorithm to predict the diagnosis that the patient obtained within the subsequent 2 years.

METHODS AND MATERIALS

SUBJECTS

The study involved 86 mild cognitive impairment (MCI) subjects who were referred to the Alzheimer Center at the VU University Medical Center in Amsterdam, the Netherlands (Figure 1B). Upon the first visit at the Alzheimer Center, all subjects underwent a thorough 1-day examination consisting of history taking, physical, and neurological assessment, neuropsychological testing including the Mini Mental State Examination (MMSE) (Folstein et al., 1975), laboratory tests, structural magnetic resonance imaging (MRI), and a routine electroencephalogram (EEG). After reviewing the clinical and ancillary imaging data, a multidisciplinary team established a consensus-based final diagnosis for each patient. The initial diagnosis of MCI was based on the criteria set by (Petersen et al., 1999), consisting of (a) objective memory impairment as seen during neuropsychological evaluation, defined by performances \( \geq 1.5 \) standard deviation below the mean value of education—and that of age matched controls, (b) normal activities of daily living, and (c) a rating score of 0.5 in clinical dementia (Hughes et al., 1982).

All MCI subjects were followed up clinically during an average period of 709 ± [537:779] days (1.9 years) (median ± 95% confidence interval). The clinical follow up included medical history and functional status assessment re-examination in order to measure potential changes in the cognitive domain. MCI subjects who showed steady or enhanced cognitive functioning (but still fulfilled the criteria for MCI) during re-assessment were considered as MCI-stable, while MCI subjects who showed impoverished cognitive functioning, and fulfilled the NINDS-ADRDA criteria (McKhann et al., 1984) to be diagnosed with Alzheimer’s disease, were considered to belong to the AD-converter group. Exclusion criteria were previous head trauma, history of neurological or psychiatric disease or use of psychotropic medications. Patients progressing from MCI to other disorders than Alzheimer’s disease (\( n = 22 \)) were excluded from the analyses reported here. These patients progressed to; “Subjective complaints” (\( n = 9 \)), possible Alzheimer’s disease (\( n = 1 \)), frontal lobe dementia (\( n = 1 \)), vascular dementia (\( n = 3 \)), Lewy body dementia (\( n = 1 \)), dementia other (\( n = 2 \)), psychiatric (\( n = 2 \)), or another neurological disorder (\( n = 3 \)). The measurements were approved by the Ethics Committee of the VU University Medical Center, and were in accordance to the Helsinki declaration. All subjects signed an informed consent.

EEG RECORDINGS

Twenty-one channel EEGs were recorded in a sound attenuated, electrically shielded, and dimly lit room. These recordings were performed with OSG digital equipment (Brainlab®) at the following locations of the international 10–20 system: Fp2, Fp1, FT9, FT10, F8, F7, F4, F3, A2, A1, T4, T3, C4, C3, T6, T5, P4, P3, O2, O1, Fz, C2, and Pz. The recording was referenced to the common
average of all electrodes, excluding Fp1 and Fp2. Sampling frequency was 500 Hz and analogue-digital precision was 16 bit. The impedance of all electrodes was less than 5 kΩ. Recordings were made with a 70 Hz low-pass filter (time constant 1 s). Subjects sat in a reclined chair for approximately 20 min. During this period the subjects kept their eyes closed most of the time, however, at irregular intervals, they were asked to open their eyes when drowsiness was noticed. Approximately 15 min into the recording a memory task, which consisted of remembering pictogram images for 1 min was performed.

EEG CLEANING
The recordings during task and eyes-open were not analyzed. The EEG was viewed in windows of 5 s, and sharp transient artifacts were cut out. On average 17.8 [range (12.4:24.1)] minutes of eyes-closed rest EEG was left. The JADE ICA algorithm was then used to separate the signal into 23 components (Cardoso and Souloumiac, 1993). Eye movements, eye blinks, muscle artifacts, and heartbeat components were rejected, based on abnormal topography, component activation, activity distribution, and spectrum.

BIOMARKERS AND PROCESSING FLOW
The Neurophysiological Biomarker Toolbox (NBT) (http://www.nbtwiki.net/) was used to organize, analyse, and calculate all biomarkers in this study (Hardstone et al., 2012). An EEG biomarker is a quantitative measure derived from the EEG, e.g., the dominant frequency of the beta frequency band (13–30 Hz), to be used as a diagnostic or prognostic predictor of disease (Figure 2).

We extracted 177 biomarkers from each EEG trace. We decided to focus on biomarkers we have had good experiences with in other studies, and acknowledge that many more biomarkers could have been selected.

Based on the broadband signal, we computed 28 biomarkers, namely: Hjorth’s activity, mobility and complexity parameters (Hjorth, 1970); Time domain Parameters (Goncharova and Barlow, 1990), Wackermann’s global field strength, global frequency, and spatial complexity (Wackermann, 1999), Barlow’s amplitude, frequency and spectral purity (Goncharova and Barlow, 1990). Alpha peak frequency, peak width, power corrected for 1/f baseline (Poil et al., 2011), when applicable the same parameters where found for double alpha peaks. Alpha-theta transition point (Klimesch, 1999), Beta peak frequency (Figure 2), width, power corrected for 1/f baseline (Van Aerde et al., 2009), same for second beta peak if present; Frequency stability was evaluated using different methods, by the standard deviation and interquartile range of the central frequency and maximum wavelet frequency calculated in windows, and by, the distribution parameters of the phase values above zero, and of the number of oscillation cycle peaks per window.

For each of the classical frequency bands—delta (1–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–45 Hz)—we computed 13 biomarkers; namely: The amplitude envelope was extracted using Hilbert transform and characterized extensively. We calculated the spearman correlations of amplitude envelopes in different channels. The distribution of amplitude values was characterized by kurtosis, skewness, interquartile range, median, range, and variance. Furthermore, detrended fluctuation analysis characterizing long-range temporal correlations (Linkenkaer-Hansen et al., 2001; Hardstone et al., 2012; Poil et al., 2012), multifractality spectral width (Kantelhardt et al., 2002; Ihlen, 2012) and oscillation bursts 95th percentile durations and sizes (Montez et al., 2009; Poil et al., 2011) were calculated on the amplitude envelope. The instantaneous phase was also extracted using Hilbert transform, and the 95th percentile duration and size of the stable phase bursts (a phase bursts is defined as the period between phase slips) were calculated. In addition, we computed for all frequency bands and individualized frequency bands, defined as Alpha1 (APF = individually defined Alpha peak frequency): (APF–4 to APF–2) Hz, Alpha2: (APF–2 to APF) Hz, Alpha3: (APF to APF+2) Hz;
Beta: (APF+2 to 30) Hz (Klimesch, 1999), 7 biomarkers: absolute, relative power, and power ratios, furthermore, the central frequency, power in central frequency, bandwidth and spectral edge (Vural and Yildiz, 2010; O’Gorman et al., 2013). In total, we extracted 177 biomarker values from each EEG trace (Table 1).

Next, we performed data mining on these biomarkers based on the second EEG recording (Figure 1A), to identify biomarkers that reached a significance level of \( p < 0.05 \) (student’s \( t \)-test) for the comparison of stable MCI vs. AD-converters (based on the diagnosis after 2 years). We here use student’s \( t \)-test because this test has best statistical power in most cases under the assumption of normal distributed biomarker values. The biomarkers were tested per channel, and a binomial multiple-comparison correction was performed (Poil et al., 2011). The binomial multiple-comparison correction tests whether a significant number of channels are found (i.e., 3 or more channels, \( p < 0.05 \)). The performance of two different classification algorithms (see below for details) in integrating significant biomarkers into a diagnostic index was then tested using their median values across significant channels.

DEVELOPMENT OF A DIAGNOSTICS INDEX

To move beyond single-biomarker classification we aimed to integrate several EEG biomarkers in a diagnostic index that would classify the AD-converter group from the MCI-stable group better than each individual biomarker. Using one dataset for development and testing is not recommended, because it is theoretically possible to find a perfect separation of two groups if enough biomarkers are included (so-called over-fitting). To counteract this issue we build our classification model based on the second EEG recording (which was obtained in 34 out of a total of 64 subjects that were either MCI-stable or AD-converters), and tested the classification accuracy retrospectively on the first EEG recording. Thirty subjects were not included in the training (22 MCI-stable, 8 AD-converters), because these subjects did not have any second-year recording. These subjects serve as our ultimate classification test. We also used half-split cross-validation to evaluate the stability and lower bound of the solution (see below).

**STATISTICS: LOGISTIC REGRESSION WITH GENETIC SEARCH**

Binary classification was performed using logistic regression. In logistics regression the binary outcome either AD-converter (1) or MCI-stable (0) is regressed with a linear combination of biomarkers. More specifically we fit a function \( f(z) \) using maximum likelihood.

\[
f(z) = \frac{1}{1 + e^{-z}}
\]

with

\[
z = \beta_0 + \sum_{i=1}^{k} \beta_i x_i,
\]

and \( x \) are the \( k \) biomarkers included in the regression (included as medians across significant channels), and \( \beta_i \) are the regression coefficients. The function \( f \) represents the probability of Alzheimer’s disease. We use the 50% probability as our classification threshold, i.e., if \( f \geq 0.5 \), the patient belong to the AD-converter group, otherwise the patient belongs to the MCI-stable group. We used a genetic search method to identify biomarkers that combined (using logistic regression) would give the best classification of the outcome MCI-stable vs. AD-converters. Genetic search is considered an efficient method for searching large data sets, instead of the computationally demanding alternative of testing all possible combinations (Koza and Poli, 2005; Zviling et al., 2005). The genetic approach is based around an evolutionary idea where the combined set of biomarkers is “mutated” by different mutation rules; addition of a random biomarker, removal of a biomarker, random selection of a new set of four biomarkers, and random substitution of a biomarker. Each rule was applied 5 times in each generation, leading to 20 new sets of biomarkers. The classifications of these new sets were then compared with the previous optimal set. Only the best biomarker set survived and was used as the base for next generation of mutations. We did not set limits on the maximum or minimum number of biomarkers in each set.

The genetic algorithm was seeded with an initial set of five biomarkers with the highest Matthew correlation coefficient (see outcome evaluation below). The genetic algorithm ran for 100

<table>
<thead>
<tr>
<th>Spatial biomarkers</th>
<th>Temporal biomarkers</th>
<th>Spectral biomarkers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman correlations of the amplitude envelope across channels</td>
<td>Detrended fluctuation analysis</td>
<td>Absolute and relative power</td>
</tr>
<tr>
<td></td>
<td>Multifractal spectral width</td>
<td>Central frequency</td>
</tr>
<tr>
<td></td>
<td>Oscillation bursts duration and size</td>
<td>Power in central frequency</td>
</tr>
<tr>
<td></td>
<td>Stable phase bursts duration and size</td>
<td>Bandwidth and spectral edge</td>
</tr>
<tr>
<td></td>
<td>Frequency stability; standard deviation, interquartile range of central frequency, maximum wavelet frequency; distribution parameters of the phase values above zero; number of oscillation cycles per window</td>
<td>Hjorth’s activity, complexity, and mobility</td>
</tr>
<tr>
<td></td>
<td>Amplitude envelope parameters; kurtosis, skewness, interquartile range, median, range, and variance</td>
<td>Wackerman’s Global Field strength, global frequency, and spatial complexity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Barlov’s amplitude, frequency, and spectral purity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Alpha peak frequency, peak width</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Alpha peak power corrected for 1/f baseline</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beta peak frequency, peak width</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beta peak power corrected for 1/f baseline</td>
</tr>
</tbody>
</table>
generations. At each generation the biomarker set with maximal positive likelihood ratio (see outcome evaluation below) survived. In all cases the logistic regression model was fitted using the second EEG recording, and the classification outcome was measured using the first EEG recording.

**STATISTICS: ELASTIC NET LOGISTIC REGRESSION**

As an alternative to genetic optimization of biomarkers included in the logistic regression, we employed an elastic net logistic regression algorithm (Zou and Hastie, 2005) as implemented in the glmnet package for Matlab (http://www-stat.stanford.edu/~tibs/glmnet-matlab/) (Friedman et al., 2010). This algorithm promises a build-in selection of features that optimally can perform much better than the less stable genetic optimization. The elastic net optimizes the number of biomarkers included in the diagnostic index by minimizing both the L1 and L2 norm of the regression coefficients by minimizing the equation

\[ L(\lambda_1, \lambda_2, \beta) = |z - X\beta|^2 + \lambda_1 |\beta| + \lambda_2 |\beta|^2 \]

where the first term is similar to the logistic regression, and the second and third are the penalizing terms (the elastic net) (Zou and Hastie, 2005). The parameters \( \lambda_1 \) and \( \lambda_2 \) determines the influence of either the L1 or L2 norm penalty. We define a new combined parameter

\[ \alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2} \]

which we optimized in 5-split cross-validation based on the best classification by training on second-year data, and testing on the 1/5 left-out subject group on first-year EEG (note that subjects which did not have a second-year EEG were not included, and, therefore, serve as our ultimate test group (see Results) (data not shown). We found the best classification with \( \alpha = 0.8 \).

**STATISTICS: CLASSIFICATION OUTCOME EVALUATION**

To evaluate the outcome of our classification we use five different measures:

- Sensitivity (SE): defined as the (number of correctly classified AD-converter patients)/(number of AD-converter patients).
- Specificity (SP): defined as the (number of correctly classified MCI-stable subjects)/(number of MCI-stable subjects).
- Positive predictive value (PPV): defined as (number of correctly classified AD-converter patients)/(number of patients classified as AD-converter).
- Positive likelihood ratio (PLR): defined as (Sensitivity)/(1-Specificity).
- Matthew correlation coefficient (MCC): explains the correlation between the outcome and the expected outcome (Baldi et al., 2000).

A Matthew correlation coefficient higher than 0.20, sensitivity higher than 65%, specificity higher than 65%, positive predictive value higher than 65%, and a positive likelihood ratio higher than 1.6 means that the classification is significantly different from a random classification (Monte Carlo simulation, 5000 iterations, \( n = 65 \), note these results depends on the sample size making the threshold levels lower for larger sample sizes, \( p < 0.05 \)). Perfect classification would give a Matthew correlation coefficient (MCC) of 1, sensitivity of 100%, specificity of 100%, positive predictive value of 100%, and an infinite positive likelihood ratio.

An issue with these outcome measures is that they only tell how well the classification fits the given subgroup of subjects, but not how well the classification generalizes to other subject populations. We counteract this by three approaches; (1) classification was performed on the second EEG recording, whereas the prediction was tested on the first EEG recording, (2) as the ultimate test we evaluated the prediction on subjects not included for classifier training (because not all subjects had a second EEG recording), and (3) we performed a half-split cross-validation. In the half-split cross-validation the sample was divided randomly in half several times (1000 iterations); the classifier was then trained on the first half, and the outcome was evaluated on the second half. We report the median outcome measures over these splits. Cross-validation gives an estimate of the classification performance on an “unknown” sample (Witten et al., 2011). However, cross-validation also suffers from lower \( n \) numbers, which means their outcome should be viewed as a conservative estimate of the average outcome.

**STATISTICS: GROUP DIFFERENCES AND CORRELATIONS**

We use non-parametric permutation tests based on median (Box and Andersen, 1955; Ernst, 2004) to test for differences between groups. Non-parametric tests are more robust toward non-normal data, but also often have lower power than parametric such as student’s t-test. Confidence intervals (95%) were found using non-parametric bias corrected and accelerated bootstrap (\( n = 5000 \)) (DiCiccio and Efron, 1996).

**STATISTICS: 2 × 2 TABLE INDEPENDENCE TESTS**

To test for dependence of genotype, gender, and patient group we used Barnard’s exact test, which is appropriate for low sample statistics compared with Chi-square test, and has better power compared with Fisher’s exact test (Barnard, 1947).

**STATISTICS: MULTIPLE COMPARISONS**

Because we do large-scale mapping of biomarkers, we employ a lenient approach to multiple comparisons correction at the first level of analysis. This means that in the initial mapping of potential difference between the stable MCI and AD-converter groups, we only perform a binomial correction for the number of significant channels in each biomarker (Poil et al., 2011). We do not correct the \( p \)-values across different biomarkers. This approach is appropriate since this mapping of potential difference is only used to identify candidate biomarkers for the genetic search algorithm.

**RESULTS**

**PATIENT GROUPS—AGE AND GENDER**

Initially 86 subjects (Age: 68.7 [66.5:71.3] years, median [95% confidence interval], age at first EEG, 58 males) were diagnosed with mild cognitive impairment (MCI). After 415 ± [393:478] days, 17 patients (9 males) had converted to Alzheimer’s disease. After 709 ± [537:779] days (1.9 years) a total of 25 patients
In MCI in the second measurement (MCI: 16.9 ± 1st measurement (age: 67 ± 61.74 years) (including frontal lobe and Vascular dementia). No significant difference was found in age and gender between stable MCI and AD-converters (Gender: Barnard's test, p = 0.16; Age: permutation test, p = 0.49) (Table 1). We only focus on the patients diagnosed with AD, and subjects remaining stable MCI. In the following we use the last diagnosis of the subjects for the definition of the MCI-stable and AD-converters groups.

MMSE RESULTS

The MMSE score of the MCI-stable group (28 ± [27:29]) was not significantly different from the score from AD-converter group (27 ± [26:28]) at the intake test (permutation test, p = 0.8). At the follow up approximately 1 year later the stable MCI subjects remained at a stable MMSE score of 28 ± [26:29], whereas the MMSE score of the AD-converter group changed to 24 ± [22:24] (permutation test, p = 0.0044), which is also lower than the stable MCI group's MMSE scores (permutation test, p = 0.0002) (Table 1).

APOE STATUS

We observed a significantly higher frequency of E4 allele vs. no E4 allele in AD-converter vs. stable MCI (Barnard test, p < 0.01). Only 38% of MCI-stable compared to 64% of AD-converter group had more than one E4 allele (Table 2).

SINGLE-BIMARKER LOGISTIC REGRESSION MODEL OF AD-CONVERTER vs. MCI-STABLE

To show the principle of logistic regression modeling on a single biomarker, we chose the beta peak frequency, because this biomarker showed significantly lower values in MCI (MCI: 17.6 ± [16.8:18.2] Hz, n = 39) compared with the AD-converter group (AD: 19.6 ± [18.1:21.0] Hz, n = 25) (p < 0.0005) in the first measurement (Figure 3A), and also significantly lower values in MCI in the second measurement (MCI: 16.9 ± [16.0:17.8] Hz, n = 17; AD: 19.3 ± [18.6:20.6] Hz, n = 17, p < 0.005) (frequency values are averages across the significant channels) (Figure 3A).

We fitted a logistic regression model to the second EEG measurement (n = 17 in both groups, Figure 3B). The model classified the second measurements with a sensitivity (SE) of 76%, 76% specificity (SP), 76% positive predictive value (PPV), 0.5 Matthew correlation coefficient (MCC), and a positive likelihood ratio (PLR) of 3.3. Next, we used this logistic model to retrospectively classify the first EEG measurement (Figure 3C). The classification had a SE of 72%, 59% SP, 53% PPV, 0.3 MCC and a PLR of 1.8; thus, as expected, a worse classification power (MCI n = 39, AD n = 25) (Figure 3D).

MULTIPLE-BIMARKER LOGISTIC REGRESSION MODEL OF AD-CONVERTER vs. MCI-STABLE

By combining several biomarkers it may be possible to obtain better classification power than the individual biomarkers alone (Schoonenboom et al., 2004; Buscema et al., 2007; Lehmann et al., 2007). However, it is not trivial which combinations of biomarkers are optimal, because of the high number of possible combinations. Here, we employ a genetic search approach and elastic net penalization to assist us in finding these optimal combinations (see Methods and Materials section).

The best set of biomarkers identified by the genetic search was (six biomarkers): Amplitude correlations with Cz in Beta (13–30 Hz), Bandwidth of subject-specific Beta frequency, Peak width of dominant beta peak, range of amplitude values in Beta (13–30 Hz), Ratio between theta and alpha power, and alpha relative power (normalized with 1–45 Hz broadband). The logistic regression training on this biomarker set using the second EEG data yielded a SE of 100%, 94% SP, 94% PPV, 0.94 MCC, and PLR of 17 (n = 17 in both groups).

The retrospective testing on first-year data using the classifier model trained on the second-year data gave a SE of 92%, 85% SP, 79% PPV, 0.75 MCC, and PLR of 6 (MCI-stable, n = 39; AD-convert, n = 25) (Figures 3F; G, Table 3), which indicates that even at this very early stage differences between AD-converters and MCI-stable can be identified. However, since second-year and first-year data from the same subjects may be strongly correlated we also performed a classification test using only subjects that were not used for training the model (i.e., the subjects without a second EEG recording). We obtained a good classification with a SE of 88%, 82% SP, 64% PPV, 0.64 MCC and a PLR of 4.8 (MCI-stable, n = 22; AD-convert, n = 8), suggesting the diagnostic index can generally be used for these patient groups. Furthermore, we performed a half-split cross-validation (1000 iterations), with a SE of 75%, 63% SP, 52% PPV, 0.37 MCC, and a PLR of 2, an indication of the average outcome. As expected, the classification powers decrease; however, this is at least partly explained by the lower n number. However, the combined classification is still much better than prediction obtained on the individual biomarkers in the set (Figures 3E, 4). The best single biomarker in the biomarker set (based on sensitivity and specificity) was the peak width of the dominant Beta peak, with a SE of 64%, 62% SP, 52% PPV, 0.24 MCC, and a PLR of 1.7 (MCI-stable, n = 22; AD-convert, n = 8) (Table 3). The logistic regression fitting coefficients for the combined solution were: −2.9 for Amplitude correlations with Cz in Beta, 0.5 for bandwidth of subject specific Beta, 3.4 for Peak width of dominant beta peak, −0.6 for range of amplitude values in Beta, −2.3 for ratio between theta and alpha power, and −0.2 for alpha relative power. This means that the peak width of the dominant beta peak had the greatest

Table 2 | Overview of patient groups.

<table>
<thead>
<tr>
<th>Patient group</th>
<th>Age [years]</th>
<th>MMSE 1st year</th>
<th>MMSE 2nd year</th>
<th>Number of APOE E4</th>
</tr>
</thead>
</table>

\[ p = 0.49 \quad p = 0.8 \quad p = 0.0002 \quad p < 0.01 \]
FIGURE 3 | Integration of multiple biomarkers using logistic regression improves the prediction of Alzheimer’s disease at the MCI stage. (A) A significant higher Beta peak frequency is observed in Alzheimer’s disease converter group (AD) (red) compared with mild cognitive impairment stable group (MCI) (blue), in both first (left) and second (right) year EEG recording. (permutation test on median, binomial corrected, **p < 0.005, ***p < 0.0005) (B) The logistic model is fitted to the second-year EEG recording. (C) The logistic model is used to predict outcome on the first year EEG recording. Separation plot of AD vs. MCI. (D) Outcome evaluation of beta peak frequency using five measures of classification power (warmer is better). SE, Sensitivity; SP, Specificity; PPV, Positive predictive value; MCC, Matthews Correlation Coefficient; AUC, area under the receiver operator curve. (E) Outcome evaluation as in (C), but for the “optimal” biomarker set found using genetic search. The first six columns are for classification of the individual biomarkers separately. The last column is the combined classification outcome. We clearly see that the combined outcome is better than the classification using the individual biomarkers. 1, Peak width of dominant beta peak; 2, range of amplitude values in Beta (13–30 Hz); 3, Bandwidth of subject-specific Beta frequency; 4, Ratio between theta and alpha power; 5, alpha relative power (normalized with 1–45 Hz broadband); 6, Amplitude correlations with Cz in Beta (13–30 Hz). (C) Combined logistic classification using the biomarkers 1, 2, 3, 4, 5, and 6. (F) Logistic curve for combined classification based on first-year EEG. (G) Separation plot of MCI vs. AD in first EEG recording using combined classification based on second-year logistic regression coefficients. Note that the recordings used for training in (F) are different from those used for testing in (G).

Table 3 | Overview of classification results [classification based on testing subjects that were not used for training the classifier (MCI-stable, n = 22; AD-convert, n = 8)].

<table>
<thead>
<tr>
<th>Model</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Positive predictive value (%)</th>
<th>Matthew correlation coefficient</th>
<th>Positive likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic search 6 biomarkers</td>
<td>88</td>
<td>82</td>
<td>64</td>
<td>0.64</td>
<td>4.8</td>
</tr>
<tr>
<td>Single best biomarker</td>
<td>64</td>
<td>62</td>
<td>52</td>
<td>0.24</td>
<td>1.7</td>
</tr>
<tr>
<td>Elastic-net 12 biomarkers</td>
<td>75</td>
<td>86</td>
<td>67</td>
<td>0.59</td>
<td>5.5</td>
</tr>
</tbody>
</table>

Taken together, our results show that it is possible to obtain a substantial synergistic effect from the integration of several biomarkers; however, they also show that it is not trivial to identify which combination of biomarkers is most optimal. The major issue with our genetic search is that from run to run we do not obtain the same solution, because the algorithm finds local maxima. We, therefore, employed an elastic net penalized logistic regression algorithm. This algorithm uses a penalization of the weights to optimize the set of biomarkers used for classification. The classification outcome from this algorithm is worse than genetic search optimized logistic regression, with a SE of 75%, 86% SP, 67% PPV, 0.59 MCC, and a PLR of 5.5 (MCI-stable, n = 22; AD-convert, n = 8) (Table 3) based on training on the second-year EEG and testing on the first-year recording of subjects (the test subjects were not used for training). The elastic net logistic regression combined 12 biomarkers (non-zero weights), namely; the amplitude correlations from Cz in Alpha (8–13 Hz) and Beta (13–30 Hz), the range of the generalized multifractal

Frontiers in Aging Neuroscience www.frontiersin.org October 2013 | Volume 5 | Article 58 | 7
We addressed the challenge of predicting whether an MCI subject would convert to AD within 2 years. To this end, we explored the added value of integrating multiple EEG biomarkers into a diagnostic index using logistic regression in combination with either a genetic search or elastic-net penalization for biomarker selection. From an initial cohort of 86 subjects with mild cognitive impairment, 25 converted to Alzheimer’s disease within 2 years. We showed how data mining of 177 EEG biomarkers could be used to identify a set of biomarkers that form a diagnostic index. The analysis was performed using the Neurophysiological Biomarker Toolbox (NBT, http://www.nbtwiki.net/) (Hardstone et al., 2012), which is specifically developed to support data mining and integration of large sets of biomarkers. We found that particularly biomarkers sensitive to changes in the beta frequency (13–30 Hz) band were optimal for classifying the very early EEG recordings of yet to be diagnosed AD patients.

**CLASSIFICATION BASED DIAGNOSTICS**

Previous studies have shown promise in using machine-learning algorithms to classify between MCI and AD based on EEG recordings (Huang et al., 2000; Bennys et al., 2001; Prichep et al., 2006; Buscema et al., 2007; Lehmann et al., 2007; Prichep, 2007; Rossini et al., 2008). A sensitivity of 89% and specificity of 95% were, e.g., found using the so-called IFAST model (Buscema et al., 2007; Rossini et al., 2008). However, these studies were based on training and testing on the same data, which makes it more difficult to judge the performance. Uniquely to the present study, we performed classification training on the second EEG recording, and retrospectively used this to perform prediction based on the first EEG recording from subjects not used for the training. We note, however, that the drawback of the present procedure is the low number of patients in the smallest patient group (i.e., the eight patients converting to AD) produced a fairly high error margin to the classification estimates (12.5%).

**OSCILLATIONS ARE INVOLVED IN COGNITION**

Empirical and theoretical evidence suggest that oscillations provide important systems-level mechanisms for normal brain function (Engel and Singer, 2001; Buzsáki and Draguhn, 2004; Axmacher et al., 2006; Klimesch et al., 2007; Palva and Palva, 2007, 2012; Lisman, 2010). For example, oscillations are involved in memory encoding (Raghavachari et al., 2001; Jensen et al., 2002), and are thought to provide a timing mechanism for spike-time dependent plasticity (Engel and Fries, 2010). It, therefore, seem plausible that if oscillations are abnormal in disorders such as MCI and AD, then cognition is also affected. Apart from relative Alpha power and the theta-alpha power ratio, which may reflect early changes toward the well-known slowing of the EEG in AD (Bennys et al., 2001; Rossini et al., 2006), our optimal set of biomarkers is derived from the Beta frequency band (13–30 Hz).

Beta-band changes have previously been observed in Alzheimer’s disease, e.g., by a more anterior distribution (Huang et al., 2000). The larger width of the beta peak and bandwidth could potentially be linked with a less stable beta frequency, and, therefore, also a less efficient working memory (Kopell et al., 2011). Beta oscillations are believed to maintain the current sensorimotor and cognitive state (Engel and Fries, 2010). Activity in the beta-frequency range has also traditionally been linked with motor function. Interestingly, it has been found that motor performance is impaired in early-stage Alzheimer’s disease but not in mild cognitive impairment (Sheridan et al., 2003; Pettersson et al., 2005), which is a potential explanation of the prominent role of beta-frequency changes in our data. Motor function, e.g., gait control, is a higher cognitive function requiring integration of several cognitive functions, as attention, planning (Hausdorff et al., 2005; Scherder et al., 2007), albeit unrelated to performance in memory tests (Hausdorff et al., 2005). Hyperexcitability of the motor cortex has also been observed in AD (Di Lazzaro et al., 2004), which our finding of higher beta frequency also suggests.
EEG BIOMARKERS AS POTENTIAL INDICATORS OF INFLAMMATION

The standard hypothesis of Alzheimer’s disease is the amyloid cascade hypothesis stating that the cause of Alzheimer’s should be found in the build up of amyloid and tangles (Hardy and Selkoe, 2002; Huang and Mucke, 2012). It has been hypothesized that Alzheimer’s disease is initiated by a micro injury, presumably a vascular event, in the brain with subsequent activation of inflammatory responses that further leads to initiation of the amyloid deposition cycle (De la Torre, 2004; Herrup, 2010).

The theta/(lower alpha) power ratio has previously been associated with vascular damage in AD (Moretti et al., 2007b), and the delta (2–4 Hz) power has been associated with inflammation (Babiloni et al., 2009). EEG power and frequency in general has also been correlated with cerebral perfusion (O’Gorman et al., 2013), which is known to be reduced in Alzheimer’s disease (De la Torre, 1999; Kogure et al., 2000; Murray et al., 2011). If we could detect early-stage changes using EEG, we would have a powerful tool that could detect Alzheimer’s disease at a point where a possible therapy would be most efficient. Mouse models, e.g., show that Aβ42 modifying therapy has limited effect after neurodegeneration has begun (Dubois et al., 2007; Sperling et al., 2011). Thus, meaning that diagnosing a patient based on neurodegeneration and cognitive decline may already be too late for a good treatment outcome because the brain damage has already occurred.

It has also been shown that the build-up of Aβ42 influences synaptic transmission, and thus, potentially also give rise to further effects in the EEG (Palop and Mucke, 2009; Verret et al., 2012). Further hippocampal injections of amyloid β in rats have been shown to induce impaired memory performance combined with reduced hippocampal theta oscillations and less activity in GABAergic neurons (GABA, gamma-aminobutyric acid) (Villette et al., 2010). A recent suggestion for a potential improvement of Alzheimer’s disease symptoms is transcranial direct current stimulation (tDCS) (Hansen, 2012). This method increased theta and alpha oscillations together with improved working memory performance (Zaehle et al., 2010). Interestingly, it has been suggested these effects may be caused by altered GABA concentration within the stimulated cortex, and potentially by an adjustment of the excitatory/inhibitory balance, which is disturbed in Alzheimer’s disease (Di Lazzaro et al., 2004; Rossini et al., 2007; Stagg et al., 2009). This balance may be directly linked to EEG biomarkers that have been shown sensitive to Alzheimer’s disease (Montez et al., 2009; Poil et al., 2011, 2012). It thus seems that EEG biomarkers may be sensitive to underlying pathophysiology of AD.

OUTLOOK

We here showed that exploratory data mining and integration of multiple biomarkers might yield many exciting results on the large databases of neuroscience data build up over the years. These studies may identify hidden structures (see schematic Figure 1C) and be beneficial for both pre-clinical and clinical research. With recent developments in automatic cleaning of EEG this analysis may potentially be performed immediately after the recording (Nolan et al., 2010; Mognon et al., 2011). This together with the non-invasive character of EEG could make a diagnostic index using EEG biomarkers a powerful tool to support the early-stage clinical assessment. EEG biomarkers, apart from being non-invasive and relatively inexpensive, have the advantage of monitoring brain activity in real time, and thus potentially able to identify tiny changes in ongoing cognition. However, we believe the best diagnostic/prognostic performance is achieved if EEG biomarkers are combined with information from other modalities. Future studies should specifically study how the synergistic information of integrative biomarkers can be improved further by the incorporation of different classes of biomarkers, which could range from cognitive markers (Tabert et al., 2006), functional connectivity markers (Stam et al., 2006, 2007), coherence, synchronization, and topographical location markers (Huang et al., 2000; Stam et al., 2005; Rossini et al., 2006) to questionnaire data providing quantitative data on the mental state of the patients during the resting-state EEG recording (Diaz et al., 2013). Improvement in algorithms used for pre-selecting biomarkers could, e.g., be based on measures of interrelatedness between biomarkers or taking scalp topographies into account as opposed to the averaged channel biomarker values used here. We believe the Neurophysiological Biomarker toolbox provides a promising framework for these studies. This could give rise to a better integrative understanding of biomarkers involved with Alzheimer’s disease and brain disorders in general (Sears, 2005; Dubois et al., 2007; Schneider, 2010).

ACKNOWLEDGMENTS

This work was supported by Netherlands Organization for Scientific Research (NWO), with a Toptalent grant (Grant nr. 021.002.080) to Simon-Shlomo Poil. The authors would like to acknowledge Emi Saliasi for help with data collection. The authors would like to thank C. J. Stam for sharing the clinical data.

REFERENCES


**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 July 2013; paper pending published: 22 August 2013; accepted: 11 September 2013; published online: 03 October 2013.


This article was submitted to the journal Frontiers in Aging Neuroscience.

Copyright © 2013 Poil, de Haan, van der Flier, Mansvelder, Scheltens and Linkenkaer-Hansen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.